Real-Time Registration of Acoustic Images

U. Castellani, L. Tao, A. Fusiello and V. Murino

Dipartimento di Informatica, University of Verona
Strada Le Grazie 15, 37134 Verona, Italy
{castellani,tao,fusiello, murino}@sci.univr.it

Abstract

This paper proposes a technique for online three-dimensional reconstruction of underwater environments from multiple range views acquired by an acoustic camera. 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. 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 in real time. Due to the absence of control of the sensor position, no information is available about the degree of overlapping between the range images; further, speckle noise and low resolution make more difficult the registration process. In order to allow the real time registration we introduce a method for speed up the Iterative Closest Point (ICP) basing on a reverse calibration approach. Examples on real images have been presented to show the promising performances of the proposed algorithm.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Object Modeling

1. Introduction

Underwater exploration is nowadays growing due to both industrial and scientific needs. Fortunately, even technology is improving with the advent of smart sensors able to provide data with high visual quality, differently from only few years ago. Recently, computer vision and computer graphics scientists have also approached underwater scene understanding issues [1].

This work have been carried out in the context of a project aimed at the 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, namely an offshore rig, composed by pipes connected each others by joints (see Fig. 1). The underwater environment is undoubtly a complex scenario for both the implicit limited accessibility and the difficulty to retrieve good quality data. Typically, acoustic systems are used to sense underwater scenes as they allow to achieve a larger visibility and to measure range distance, unlike the more used optical sensors. In the present case, our data are obtained by a high frequency acoustic camera, called Echoscope [2]. These data are affected by speckle noise, due to the coherent nature of the acoustic signals, which corrupts sensibly the visual quality and decreases the reliability of the estimated 3D measures.

In this paper our goal consists of building a 3D mosaic of the surrounding scene in real-time. We introduce a method

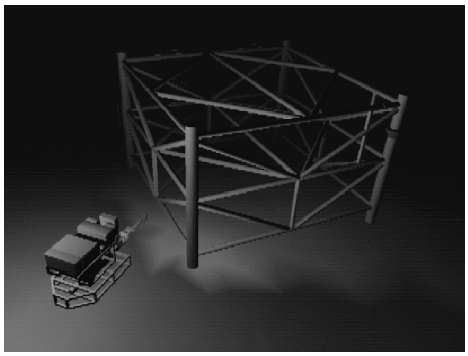


Figure 1: *Rendering of the model of an oil rig with the ROV.*

for the improvement of the speed of the registration between a pair of sequential frames. This process occurs while new frames come from the sensor as the vehicle carrying it moves. The proposed method is based on a reverse calibration approach [3]. By using information coming from the sensors parameters we are able to re-project each 3D points onto the range image and viceversa. In particular each 3D point of the source data is projected to the destination model described as a range image by finding immediately its corresponding point. In this way a dramatic improvement of the speed of closest points computation is obtained that allows the registration of the views in real time. Furthermore, in order to improve the accuracy of alignment, a statistical method for avoiding the outliers conditioning is implemented. The method is based on the so called X84 rule [4] from which a threshold is automatically defined. The main contribution of the paper consists in finding the best trade-off between accuracy and speed in a fully automatic context. The rest of the paper is organized as follow. In Section 2, we shortly describe the acoustic imaging process, in Section 3 the registration process is defined, Section 4 describes some preliminary results, and finally, conclusions are drawn.

2. Image Acquisition

Three-dimensional acoustic data are obtained with a high resolution acoustic camera, the *Echoscope* 1600 [2]. The scene is insonified by a high-frequency acoustic pulse, and a two-dimensional array of transducers gathers the backscattered signals (Figure 2). The whole set of raw signals is then processed in order to form computed signals whose profiles depend on echoes coming from fixed steering directions (called *beam signals*), while those coming from other directions are attenuated. The distance of a 3-D point can then be measured by detecting the time instant at which the maximum peak occurs in the beam signal (Figure 2) [5].

The 3-D image provided by this sensor is formed by 64×64 points ordered according to a polar reference system, as adjacent points correspond to adjacent beam signals. There is a trade-off between range resolution and field of view. Resolution depends on the frequency of the acoustic signal (it is about 5 cm at 500 KHz): roughly speaking, the higher the frequency, the higher the resolution, the narrower the field of view.

The acoustic image is affected by false reflections, caused by secondary lobes, and by acquisition noise, which is modelled as speckle noise. Moreover, the intensity of the maximum peak can be used to generate another image, representing the reliability of the associate 3-D measures, so that, in general, the higher the intensity, the safer the associate distance. A dramatic improvement of the range image quality is obtained by discarding points whose associated intensity is lower than a threshold depending on the secondary lobes [6, 7].

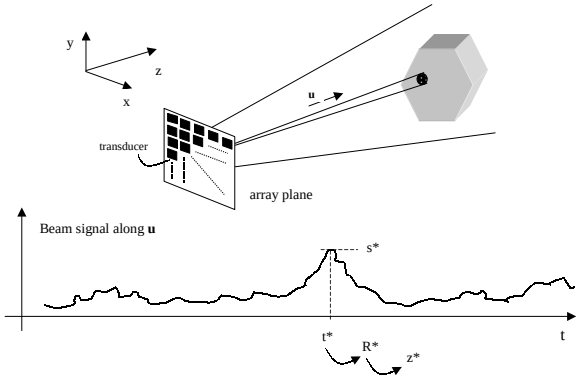


Figure 2: *Functioning of the acoustic camera: after the insonification of the scene, backscattered echoes are acquired and processed to form signals coming from specific directions (beam signals); the maximum peak of the beam signals identifies the scene distance in the related direction.*

3. Registration

In order to align two 3D views it is sufficient to compute a rigid transformation (translation $\mathbf{t}$ and 3D rotation $\mathbf{R}$). 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.

Assuming that for each point in the data set the corresponding point in the model set is known, the *point set registration problem* consists of finding a 3-D transformation which, when applied to the data set Y, minimizes the distance 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 [8], and the one based on Singular Value Decomposition (SVD) was found to be the best, in terms of accuracy and stability.

3.1. Iterated Closest Point

In general, when point correspondences are unknown, the Iterated Closest Point (ICP) algorithm may be used [9, 10]. 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. 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 $(\mathbf{R}, \mathbf{t})$;
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.

Besl and McKay [9] proved that this algorithm is guaranteed to converge monotonically to a local minimum of the Mean Square Error (MSE). Many variants to ICP have been proposed to cope with partially overlapping views and false matches in general, including the use of closest points in the direction of the local surface normal [10], the use of thresholds to limit the maximum distance between points [11], disallowing matching on the surface boundaries [12], and the use of robust regression [13, 14]. In [15] a survey on the main ICP variations is presented focusing both on the accuracy of results and speed.

We implemented a variation [16] similar to the one proposed by Zhang [11], 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 [4] 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. This is an improvement over [11], because the X84 threshold is independent on fine tuned parameters by allowing the on-line implementation. Moreover, experiments suggests that X84 achieves a larger basin of attraction (see [16] for a deepened explanation of the so called X84-ICP).

3.2. Acceleration techniques

In order to be able to work on-line, ICP needs to be modified. In general, the speed enhancement of ICP algorithm can be achieved by: i) reducing the number of iterations necessary to converge and ii) reducing the time spent in each iteration (i.e., time spent for the calculation of the correspondences)[15].

Basically, the first and the simplest step is to restrict the number of points to be registered, which is the trade-off between the registration speed and the measurement precision. However, this technique cannot be applied to the registration of the underwater images heavily since the already poor quality of the images.

In order to speed up calculation of the correspondences a number of classical approaches have been presented. For example it is customary to use k-D tree [17] [18] which allow to organize the points in a structure that decreases the complexity of the corresponences computation. Another approach is based on the substitution of the point-to-point distance metric with the point-to-surface distance, which [15] report to yield a faster convergence. Although new techniques [19, 20, 21, 22] have been proposed recently, the acceleration of the extraction of corresponding points is still open to be addresse.

In this paper we propose an acceleration method based on the so-called reverse calibration technique [3]. The method is based on the fact that from the sensor we obtain data stored in both the structures of unorganized cloud of point $x_i = (x, y, z)$ and range image $r(i, j)$ [23](i.e., a 2.5 dimensional image). Given a 3D point $y_i \in Y$ of data set and given the camera parameters it is possible to project y_i onto the range image of the model set $r_m(i, j)$. The 3D point $x_i \in X$ associated to $r_m(i, j)$ will be the *hypothetic* corresponding point of y_i. In order to improve the accuracy of finding correspondences it is possible to use the information of the connectivity given by the range image. We define the neighborhood N_i^x of x_i as:

$$N_i^x = \left\{ x_{k,h} \in X \quad s.t. \quad x_{k,h} = RP(r_m(i+k, j+h)) \right\} \qquad (2)$$

where $k, h = -w, \ldots + w$ (w is the dimension of a window centered on $r(i, j)$), and $RP(r(i, j))$ is the operator that reprojects the range point $r(i, j)$ onto the Euclidean 3D space. It is worth noting that the range image is not dense since after the filtering a lot of points are discarded. For each range point $r(i, j)$ there is a flag that indicates if the associated point is survived or not. More precisely, the operator $RP(r(i, j))$ gives 0 if the corresponding 3D point was discarded. Finally, the *definitive* corresponding point of y_i will be the closest x_i' of the points belonging to the neighborhood N_i^x of x_i. If the projection of the point y_i falls onto an empty area, this point remain without correspondence.

As we pointed out before, the main important step is the projection of the 3D point onto the range image. This process can be carried out by the following equation:

$$i = \frac{\alpha - I_{OFF}}{\alpha_{inc}}; \qquad j = \frac{\beta - J_{OFF}}{\beta_{inc}} \qquad (3)$$

where α_{inc} and β_{inc} are increments of the angles of the beams on the tilt and pan direction respectively, I_{OFF} and J_{OFF} are offsets and finally α and β are given by:

$$\alpha = arctg\frac{y}{z}; \qquad \beta = arctg\frac{x}{z} \qquad (4)$$

The parameters α_{inc}, β_{inc}, I_{OFF} and J_{OFF} are fixed by the sensor and they determine the aperture of the acquisition (i.e., field of view and resolution).

3.3. Summary of the algorithm

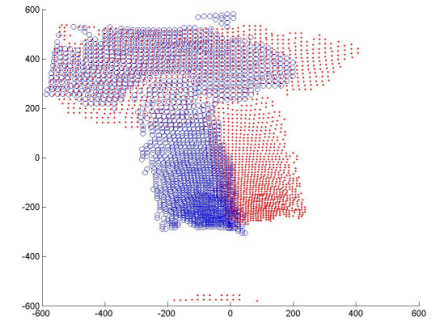
In summary, the algorithm for speed up the finding of corresponding points is given by the following steps:

For each 3D data-point $y_i \in Y$

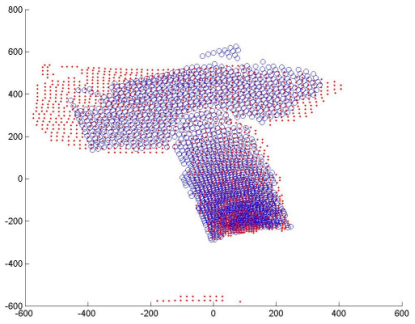
- find i and j by using eq. 3 and 4
- project y_i on to the model range image $r_m(i, j)$
- find the *hypothetic* corresponding point x_i
- find the neighborhood N_i^x of x_i
- find the *definitive* corresponding point x_i' (if it exists)

4. Experiments

In this section some experiments are presented. We test the algorithm on a sequence of real images in which the scene is composed of a joint of a pipes structure. For this sequence $\alpha_{inc} = \beta_{inc} = 1.4$ and $I_{OFF} = J_{OFF} = -44.8$. Considering that the beams are 64 for both the polar directions (tilt and pan) the aperture of this set-up is about 90×90. The sequence is composed of a set of 10 frames. Each frame is registered to the previous one. Figure 3 shows the registration of a pair of images. Figure 3.a shows the images before the registration. Figure 3.b shows their alignment by allowing the enlargement of the field of view.



(a)



(b)

Figure 3: *Images before (a) and after (b) the registration*

Finally, figure 4 shows the entire mosaic of the observed area in which all the views are represented on the same reference system. The transformations that bring each view on to the mosaic are computed just combining the sequential pairwise matrices. It is worth noting that the visible part of the observing object is increased.

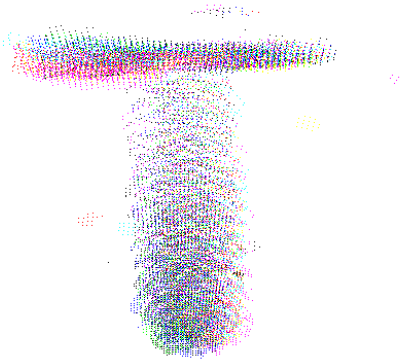


Figure 4: *Final mosaic after global registration*

Table 1 shows the performance of registration in terms of accuracy and speed. The speed of registration is very fast (about 7 frames per second) and it is sufficient to observe a mosaic of the scene in real time. Furthermore, the obtained accuracy is reasonable since it is little higher than the image resolution.

Registered Pairs	Time (sec.)	Accurancy
frame 2 to 1	0.110	10.9528
frame 3 to 2	0.140	10.9528
frame 4 to 3	0.140	10.9528
frame 5 to 4	0.140	10.9528
frame 6 to 5	0.130	10.9528
frame 7 to 6	0.151	10.9528
frame 8 to 7	0.110	10.9528
frame 9 to 8	0.161	10.9528
frame 10 to 9	0.130	10.9528

Table 1: *Performance of registration. The accuracy is given by the residual of the last ICP iteration*

5. Conclusions

In this paper we tackled the problem of automatically registering two clouds of points in real time. We introduce a method to speed up the calculation of the closest point of the well known ICP algorithm. The proposed algorithm allows the realization of an application that aims at improving the visual perception of a pilot driving an underwater ROV.

Preliminary results are satisfactory, since the speed is enough fast for the real time scene mosaicing. Furthermore, the accuracy of the alignment guarantees to obtain a global mosaic just combining the pairwise registration.

Further experiments will be needed for a deepened testing of the system. Future work will address the improvement of the rendering phase by extracting a global mesh from the set of aligned points and by introducing efficiently the colors.

Acknowledgments

This work was supported by the European Commission under the project no. GRD1-2000-25409 ARROV (Augmented Reality for Remotely Operated Vehicles based on 3D acoustical and optical sensors for underwater inspection and survey). Echoscope images are courtesy of Dr. R.K. Hansen of CodaOctopus Omnitech (Norway).

References

1. V. Murino and A. Trucco, Eds., *Special Issue on Underwater Computer Vision and Pattern Recognition*, Computer Vision and Image Understanding, July 2000. 1

2. R. K. Hansen and P. A. Andersen, "A 3-D underwater acoustic camera - properties and applications," in *Acoustical Imaging*, P.Tortoli and L.Masotti, Eds., pp. 607–611. Plenum Press, 1996. 1, 2

3. G. Blais and M. D. Levine, "Registering multiview range data to create 3-D computer objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 820–824, 1995. 2, 3

4. F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti, and W.A. Stahel, *Robust Statistics: the Approach Based on Influence Functions*, Wiley Series in probability and mathematical statistics. John Wiley & Sons, 1986. 2, 3

5. R. J. Urik, *Principles of Underwater Sound*, McGraw-Hill, 1983. 2

6. V. Murino, A. Trucco, and C.S. Regazzoni, "A probabilistic approach to the coupled reconstruction and restoration of underwater acoustic images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 9–22, January 1998. 2

7. A. Trucco V. Murino, "Three-dimensional image generation and processing in underwater acoustic vision," *Proceeding of the IEEE*, vol. 88, no. 12, pp. 1903–1946, December 2000. 2

8. A. Lorusso, D. W. Eggert, and R. B. Fisher, "A comparison of four algorithms for estimating 3-D rigid transformations," *Machine Vision and Applications*, vol. 9, pp. 272–290, 1997. 2

9. P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, February 1992. 2, 3

10. Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992. 2, 3

11. Z. Zhang, "Iterative point matching of free-form curves and surfaces," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 119–152, 1994. 3

12. Greg Turk and Marc Levoy, "Zippered polygon meshes from range images," in *Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994)*, Andrew Glassner, Ed. ACM SIGGRAPH, July 1994, Computer Graphics Proceedings, Annual Conference Series, pp. 311–318, ACM Press, ISBN 0-89791-667-0. 3

13. T. Masuda and N. Yokoya, "A robust method for registration and segmentation of multiple range images," *Computer Vision and Image Understanding*, vol. 61, no. 3, pp. 295–307, May 1995. 3

14. E. Trucco, A. Fusiello, and V. Roberto, "Robust motion and correspondence of noisy 3-D point sets with missing data," *Pattern Recognition Letters*, vol. 20, no. 9, pp. 889–898, September 1999. 3

15. M. Levoy S. Rusinkiewicz, "Efficient variants of the icp algorithm," in *IEEE Int. Conf. on 3-D Imaging and Modeling, 3DIM '01*, Quebec City (Canada), 2001. 3

16. U. Castellani, A. Fusiello, and V. Murino, "Registration of multiple acoustic range views for underwater scene reconstruction," *Computer Vision and Image Understanding*, vol. 87, no. 3, pp. 78–89, September 2002. 3

17. J. L. Bentley, "Multidimensional binary search trees used for associative searching," *CACM*, vol. 19, September 1975. 3

18. R. Rivest, "the optimality of elias's algorithm for performing best-match searches," 1974. 3

19. I. Pitas C. A. Kapoutsis, C.P. Vavoulidis, "Morphological iterative closest point algorithm," *IEEE Transaction on Image Processing*, vol. 8, no. 11, 1999. 3

20. G. Godin M. Greenspan, "A nearest neighbor method for efficient icp," in *IEEE Int. Conf. on 3-D Imaging and Modeling, 3DIM '01*, Quebec City (Canada), 2001. 3

21. C. Dorai, J. Weng, and A. Jain, "Optimal registration of object views using range data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1131 – 1138, 1997. 3

22. D.A. Simon, H. Hebert, and T. Kanade, "Techniques for fast and accurate intra-surgical registration," *Journal of Image Guided Surgery*, pp. 17–29, April 1995. 3

23. Paul J. Besl, "Active, optical imaging sensors," *Machine Vision and Applications*, pp. 127–152, 1988. 3