Proceedings of Mirage 2005, INRIA Rocquencourt, France, March, 1-2 2005

UNDERWATER ENVIRONMENT MODELLING BY FAST 3D MOSAICING

Umberto Castellani, Andrea Fusiello, Vittorio Murino

castellani@sci.univr.it
University of Verona, Department of Computer Science
Strada Le Grazie, 15 - 37134 Verona, Italy

ABSTRACT

This paper describes a technique for fast 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 speckle noise and low resolution a robust approach based on statistics is proposed in order to increase the registration accuracy. 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 performances of the proposed algorithm in terms of both speed and accuracy.

1. INTRODUCTION

Underwater exploration is nowadays growing due to both industrial and scientific needs. Fortunately, also technology is improved with the advent of smart sensors able to provide data with high visual quality, unlike 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, such as offshore rigs, pipelines, wreckage, etc.
The underwater environment is undoubtedly a complex scenario for both the implicit limited accessibility and the difficulty to retrieve good quality data. Therefore, very few systems are addressed to the reconstruction of underwater environments and, for the best of our knowledge, none of them are able to operate in real-time [2]. Kamgar-Parsi [3] proposed an *acoustic lens* technique for 3D data acquisition from which 3D models are recovered by using standard volumetric approach. Negahdaripour [4] focused on some computer visions techniques such as *shape from stereo and video, optical flow estimation, 2D mosaicing* by using optical camera(s).
In the present case, our data are obtained by a high frequency acoustic camera, called Echoscope [5]. 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 order to build a 3D mosaic of the surrounding scene in real-time, we introduce a method for the registration between a pair of sequential frames by focusing on the search of the best tradeoff between speed and accuracy. This process occurs while new frames come from the sensor as the vehicle moves. The proposed method is based on a reverse calibration approach [6]. By using information from the internal parameters of the acoustic camera, 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 corresponding 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 has been introduced in [7] and it is based on the so called X84 rule [8] from which a threshold is automatically defined. The rest of the paper is organized as follows. In Section 2, the acoustic imaging process is described and in Section 3 fundamental concepts of the registration are briefly introduced. The proposed method is detailed in Section 4, and Section 5 describes experimental results. Finally, in Section 6, conclusions are drawn.

2. IMAGE ACQUISITION

Three-dimensional acoustic data are obtained with a high resolution acoustic camera, the *Echoscope* 1600 [5]. The scene is insonified by a high-frequency acoustic pulse,

and a two-dimensional array of transducers gathers the backscattered signals. 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*), whereas those coming from other directions are attenuated. Successively, the distance of a 3D point can be measured by detecting the time instant at which the maximum peak occurs in the beam signal [9]. In particular, acoustic image is formed by the use of the Beamforming (BF) technique. It is a spatial filter that linearly combines temporal signals spatially sampled by a discrete antenna. In this way, if a scene is insonified by a coherent pulse, the signals, representing the echoes backscattered from possible objects in a specific direction, contain attenuated and degraded replicas of the transmitted pulse.

Let us denote by $\mathbf{v_k}$ the position of the k-th sensor (transducer), by c the sound velocity, and by $x_k(t)$ the signal received by the k-th sensor. Beamforming can form a *beam signal*, $bs_{\mathbf{u}}(t)$, steered in the direction of the vector $\mathbf{u}$, defined as:

$$bs_{\mathbf{u}}(t) = \sum_{k=0}^{M-1} \omega_k \cdot x_k(t - \theta_k) \qquad (1)$$

where ω_k are the weights assigned to each sensor, M is the number of transducers, and $\theta = (\mathbf{v}_k \cdot \mathbf{u})/c$ are the delays applied to each signal.

A common method to detect the scattering objects distances is to look for the maximum peak of the beam signal envelope [9]. Denoting by t^* the time instant at which the maximum peak occurs, the related distance, r^* (i.e., range value) is easily derivable (i.e., $r^* = c \cdot t^*/2$ if the pulse source is placed in the coordinate origin).

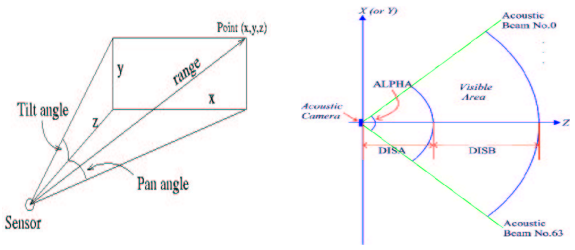
According to the spherical scanning technology, range values are measured from each steering direction $\mathbf{u}(i, j)$ where i and j are indices related to the elevation *(tilt)* and azimuth *(pan)* angles respectively. Figure 1(a) shows a schema of the scanning principle. Figure 1(b) shows a projection of the acquiring volume to the ZX plane, on which the sectors associated to each beam are marked. The minimum and maximum distances are also depicted. Going into the details, the Echoscope carries out 64 measures for both tilt and pan by defining a 64×64 range image $r_{i,j}$.

The conversion from spherical coordinates to usual Cartesian coordinates (both the coordinates system are centered to the camera) is recovered by the use of the following equations [10]:

$$x = r_{i,j} \tan(js_\alpha)/\sqrt{1 + \tan^2(is_\alpha) + \tan^2(js_\beta)} \qquad (2)$$

$$y = r_{i,j} \tan(is_\beta)/\sqrt{1 + \tan^2(is_\alpha) + \tan^2(js_\beta)} \qquad (3)$$

$$z = r_{i,j} \sqrt{\tan^2(is_\alpha) + \tan^2(js_\beta)} \qquad (4)$$



(a) (b)

Figure 1: Spherical scanning principle (a) and subdivision of the beams onto the acquiring volume (b). Each beam is associated to a (i, j) coordinate of the range image

where s_α and s_β are elevation and azimuth increments respectively. Figure 2 shows a range image and the related 3D points cloud. Both of them are provided by the Echoscope. There is a tradeoff between range resolution and field of view.

Resolution depends on the frequency of the acoustic signal (it is about 5 cm at 500KHz): roughly speaking, the higher is the frequency, the higher is the resolution, and the narrower is the field of view.

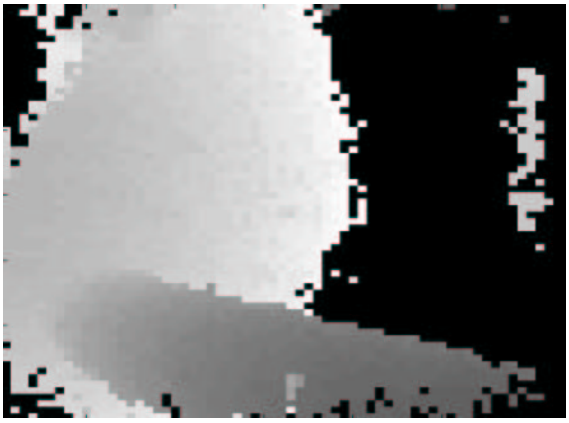
Unfortunately, due to secondary lobes and acquisition noise, the acoustic image is affected by false reflections which is modelled as speckle. Moreover, the intensity of the maximum peak can be used to generate another image, representing the reliability of the associate 3D measures, so that, in general, higher is the intensity, safer is the associate distance. A dramatic improvement of the range image quality can be obtained by discarding points whose related intensity is lower than a threshold, depending on the secondary lobes [11].

3. REGISTRATION

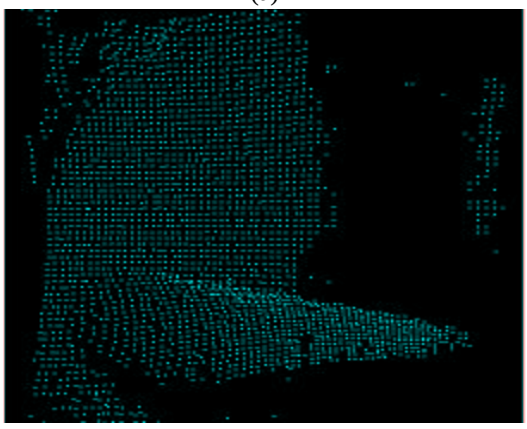
In order to build a 3D mosaic of the scene it is necessary to focus on the registration between each pair of views. Let us suppose that we have two sets of 3-D points which correspond to a single shape but are expressed in different reference frames. We will call one of these sets the *model* set X, 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 (translation $\mathbf{t}$ and rotation $\mathbf{R}$) 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:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \|\mathbf{x}_i - (\mathbf{R}\mathbf{y}_i + \mathbf{t})\|^2, \qquad (5)$$



(a)



(b)

Figure 2: Range images (a) and cloud of points (b). The scene consists of a pipe in underwater

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 [12], 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 [13, 14]. 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 [13] 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 [14], the use of thresholds to limit the maximum distance between points [15], disallowing matching on the surface boundaries [16], and the use of robust regression [17]. In [18] a survey on the main ICP variations is presented focusing both on the accuracy of results and speed.

4. PROPOSED METHOD

As we mentioned before, this work aim at finding the best trade-off between speed and accuracy. In this section we describe the main features of the proposed algorithm for both the improvement of speed, in order to permit the performance in real time, and the accuracy, in order to manipulate very noise data (i.e. acoustic data).

4.1. 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)[18].
A number of classical approaches, such as k-D tree [19], exist for speed up the calculation of the correspondences, which has been reputed to account for the bulk of the computational complexity of ICP algorithm. Another approach is based on the substitution of the point-to-point distance metric with the point-to-surface distance, which [18] reports to yield a faster convergence. The acceleration of the extraction of corresponding points is still an open issue to be addressed and different innovative techniques [20, 21, 22, 23] have been proposed recently.
In this paper, we propose an acceleration method based on a variation of the so-called *reverse calibration* technique [6]. The method is based on the fact that from the sensor we obtain data stored in both the structures of unorganized cloud of point $\mathbf{x}_i = (x, y, z)$ and range image $r(i, j)$ [10]. Given a 3D point $\mathbf{y}_i \in Y$ of data set and given the camera parameters it is possible to project $\mathbf{y}_i$ onto the range image of the model set $r_m(i, j)$. The 3D point $\hat{\mathbf{x}}_i \in X$ associated to $r_m(i, j)$ will be the *hypothetic* corresponding point of $\mathbf{y}_i$. 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 $N^{\mathbf{x}_i}$ of $\mathbf{x}_i$ as:

$$N^{\mathbf{x}_i} = \{\mathbf{x}' \in X \quad | \quad \mathbf{x}' = B(r_m(i+k, j+h)); \\ k, h = -w, ...+w\} \tag{6}$$

where w is the dimension of a window centered on $r(i, j)$ and $B(\cdot)$ is the operator that re-projects 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 (i.e., after thresholding on intensity signals) 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 $B(r(i, j))$ returns 0 if the corresponding 3D point was discarded. Finally, the closest point to $\mathbf{y}_i$ in $N^{\hat{\mathbf{x}}_i}$ is taken as *definitive* corresponding point of $\mathbf{y}_i$. If the projection of the point $\mathbf{y}_i$ falls onto an empty area, this point remains without correspondence.
As we pointed out above, 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:

$$i = \frac{\alpha - I_{OFF}}{s_\alpha}; \qquad j = \frac{\beta - J_{OFF}}{s_\beta} \tag{7}$$

where s_α and s_β are introduced in Section 2 , I_{OFF} and J_{OFF} are offsets and finally α and β are given by:

$$\alpha = arctg\frac{y}{z}; \qquad \beta = arctg\frac{x}{z} \tag{8}$$

The parameters s_α, s_β, I_{OFF} and J_{OFF} are fixed by the acquisition sensor and they determine the aperture of the acquisition (i.e., field of view and resolution)(Figure 1b). By considering the high computational cost of the *arctg* operator a variation of i and j indices extraction is introduced. From equation (8) we easily find that:

$$tg\alpha = \frac{y}{z}; \qquad tg\beta = \frac{x}{z} \tag{9}$$

Because the possible angles α_i and β_j on the model image are *a priori* known (i.e., according to the spherical scanning principle) the values $tg\alpha_i$ and $tg\beta_j$ are stored into a table. When a 3D point (x, y, z) from data set is coming, the values $\frac{x}{z}$ and $\frac{y}{z}$ are used to index the table and retrieve angles without calculating arc-tangent. In this way, by avoiding the arc-tangent computation, a dramatic improvement of the speed is obtained.
In the proposed method, a further improvement of the speed is obtained by reducing the number of points used for registration. We tested two different subsampling methods: *random* and *uniform*. In both of them the size of the sub-set is a priori fixed. The random method computes the subset by generating different random number. The uniform method simply calculates the subsampling rate (Sn) and select one point every Sn, according to the acquiring order. Although the random method is more accurate the uniform method is quite faster. For this reason we decide to implement the uniform method.

4.2. Accuracy techniques

In order to reduce the influence of the outliers for the computation of correspondences, we introduce a variation of the ICP algorithm according to the so called $X84$ *rule* [8]. We introduce a modified cost function based on robust statistics to limit the maximum distance between closest points. As pointed out by Zhang [15], 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 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 [15], because the X84 threshold is independent on fine tuned parameters by allowing the online implementation. Moreover, experiments suggests that X84 achieves a larger basin of attraction [7].
A further improvement of the accuracy is obtained by introducing a pre-alignment before the application of the fast registration. This is because we verified that the alignment based on the reverse projection could fail when the

two views are not enough close. In order to increase the robustness of the registration, we apply the classical ICP method without reverse projection for a certain number of iterations (a parameter NPI is introduced for manually fixing this number). This permits to obtain a good pre-alignment from which the reverse-calibration alignment converge fast and correctly to the optimal solution.

4.3. Parameters definition

In this section we summarize the parameters of the algorithm that are crucial for the performance:

- *Negligible Progress Threshold (NPT)*: this parameter defines the stop criteria of the ICP iterations. If the *residual* is less than this threshold the alignment is stopped. By reducing the *NPT* the accuracy is improved but more iteration must be carried out and, consequently, more time is spent. Although this parameter could be tuned by a user, we have found experimentally its best estimation [24].

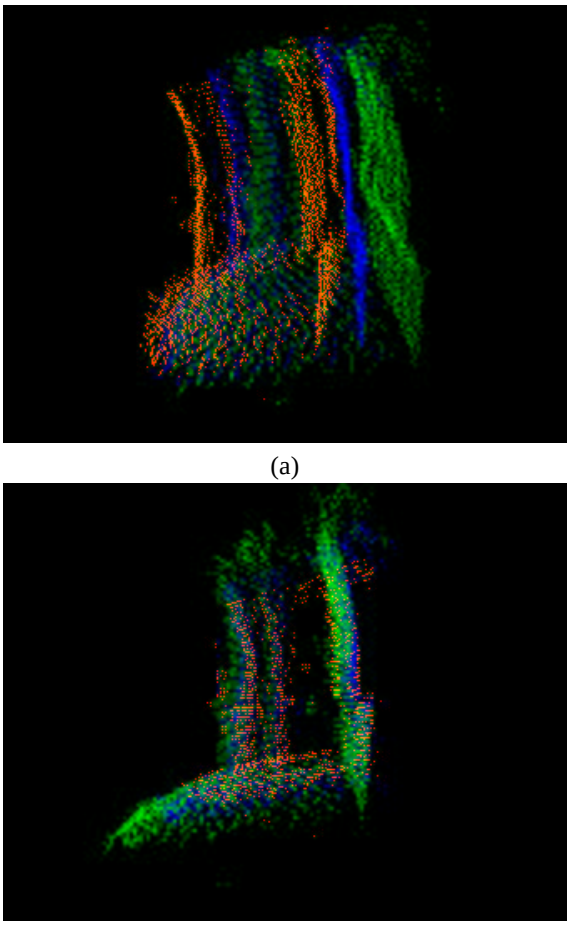
- *Automatic outliers rejection*: This procedure makes the algorithm robust to the outliers that rise when the two views are poorly overlapped. Because this computation is based on statistic a lot of time is spent for it. Especially when the motion of the ROV is very slow, it could be possible to disable this feature. Even if there is a flag in the algorithm for its disable, for the most of the cases this procedure is recommended.

- *Subsampling level*: As we mentioned before, the number of points of the subsampled image influences the performance of the alignment. This parameter is tunable by the user that can adapt the subsampling to the environment conditions. Also for this parameter we have find a reasonable estimation that can be used as a default value [24].

- *Number of pre-aligning iterations (NPI)*: Some iterations of the classic method for computing the closest points (without re-projection) are needed when the two images are not close enough even if this phase is very computational expensive. By testing different sequences of image pairs we verified that two iterations are sufficient for a good pre-alignment [24].

5. RESULTS

In this section some experiments are presented. We test the algorithm on two sequences of real images. For those sequences $\alpha_{inc} = \beta_{inc} = 1.4$ 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. After a deeper parameters estimation process, the *NPT* value is fixed to 10^{-3} and the number of points for sumbsamplig is 400. The number of pre-aligning iterations (*NPI*) are 2 and the *outliers rejection* procedure is enable [24].



(a)

(b)

Figure 3: In (a) three range images before the registration procedure is applied. In (b) the result of registration.

The first sequence (30 frames) is composed of an underwater wall and a pillar. Each frame has been registered with respect to the previous one. Figure 3 shows the registration of three images. Figure 3(a) shows the images before the registration and Figure 3(b) shows their alignment.

Figure 4 shows the points from all the range images represented in the same reference system after registration. 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 by highlighting both the pillar

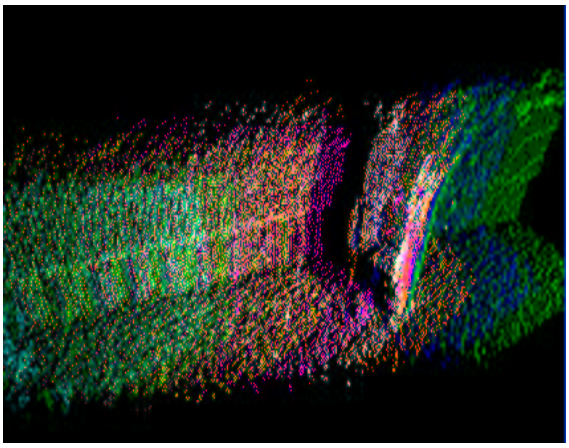


Figure 4: Input data from the whole system after registration.

and the wall.
In the second sequence (10 frames) the scene is composed by a part of a wreck. Figure 5 shows some images of the sequence. The images are very noise and it is very difficult to understand the scene. Furthermore the registration process is more challenging with such data.
After pairwise registration a global mosaic is obtain and the wreak is reconstructed. Figure 6 shows the goodness of the global alignment. Table 1 and 2 show the performances of registration in terms of accuracy and speed for both the sequences. The proposed method has been also compared with the classic ICP.
The speed of the proposed registration is about 4.5 frames per second, for both the sequences, and it is sufficient to observe a mosaic of the scene on-line (timings have been computed on a laptop $P3\ 1GHz$, with 128Mb Ram). Therefore, for the first sequence, the obtained accuracy is reasonable since it is little higher than the image resolution. For the second sequence, the measured accuracy is lower but it is still acceptable since the images are quite degraded and also they are poorly overlapped so that the registration is more likely to decrease the performance. It is worth noting that the accuracy obtained with classic ICP is little better than the accuracy obtained with the proposed method. Furthermore, the speed measured with classic ICP is very slow and, as we expected, it is not acceptable for on-line applications.

6. 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 using reverse calibration.

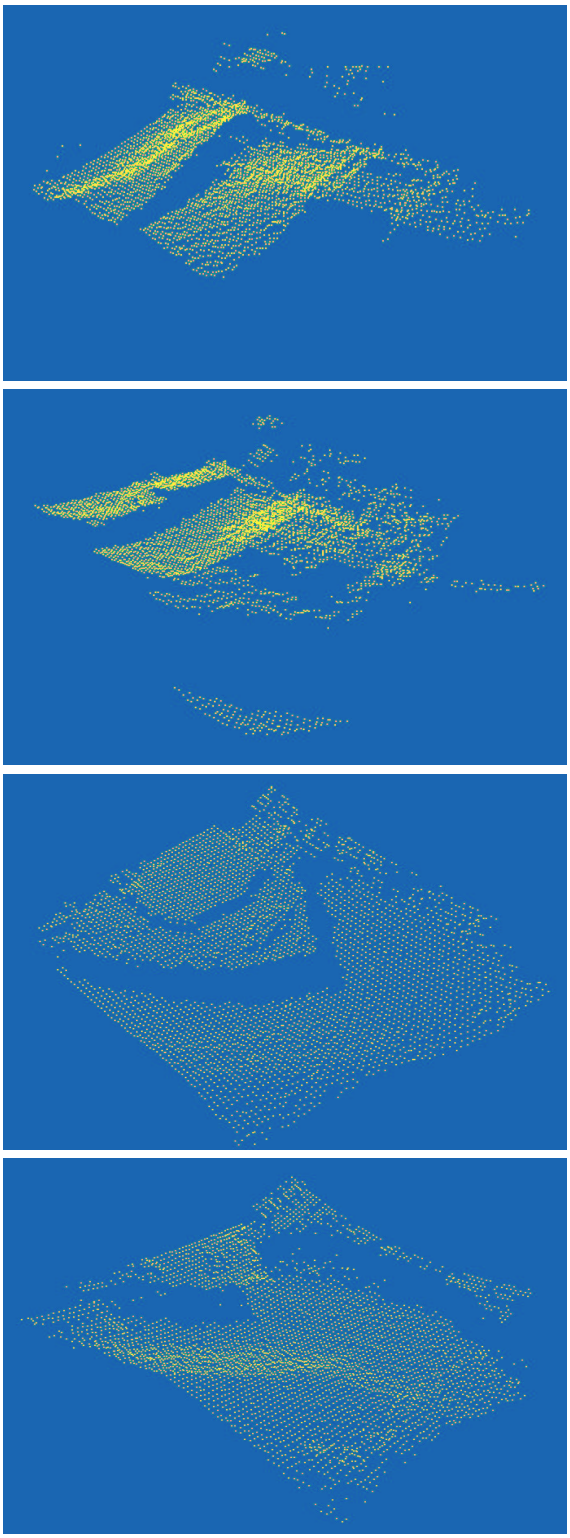


Figure 5: Some acoustic images of the second sequence.

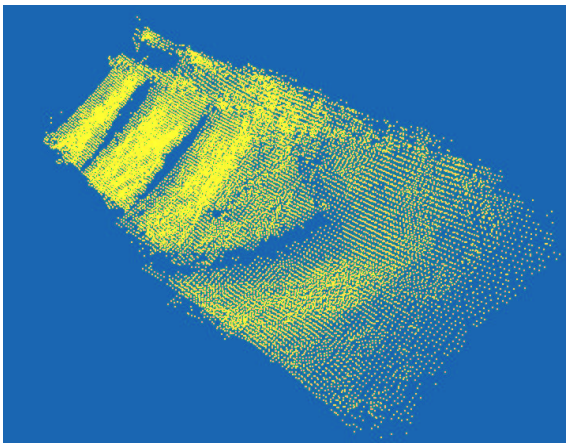


Figure 6: Mosaic of the second sequence.

The proposed algorithm allows the realization of an application that aims at improving the visual perception of a pilot driving an underwater ROV. Due to the very noisy nature of the images the mosaicing is made more challenging than for optical 3D images, and a robust approach is necessary, aiming at extracting more reliable information from the observed data.
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 are needed for a deeper 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.

	Classic ICP		Proposed ICP	
Frames	Time *(sec.)*	Accuracy *(cm.)*	Time *(sec.)*	Acuracy *(cm.)*
2to1	11.3070	12.4766	0.2210	21.5045
3to2	11.5160	11.6318	0.2910	14.7148
4to3	9.3940	11.7897	0.1600	15.2436
5to4	10.5650	10.8963	0.1510	13.8571
6to5	9.7740	10.4791	0.1900	13.6295
7to6	7.0100	10.5349	0.1410	13.5072
8to7	4.5170	11.1803	0.2100	13.0104
9to8	9.2930	10.7437	0.2700	12.7634
10to9	5.4180	10.8844	0.2000	13.4346
11to10	4.9370	11.1270	0.3700	13.3243
12to11	7.1600	11.3224	0.1010	12.5554
13to12	1.3520	14.4718	0.1200	13.4227
14to13	2.6640	12.7011	0.1000	13.8782
15to14	2.3230	12.5924	0.1200	13.2948
16to15	3.0740	12.4927	0.1900	12.3509
17to16	5.0980	9.8158	0.4910	12.9975
18to17	6.5090	9.4384	0.1600	13.5488
19to18	1.3420	9.7802	0.1100	12.8926
20to19	5.7690	10.0135	0.1200	13.6677
21to20	3.2250	11.0800	0.4410	13.7265
22to21	6.3690	10.1116	0.1700	14.3126
23to22	1.4720	8.8831	0.3110	14.7219
24to23	5.5690	7.8530	0.1700	15.2393
25to24	2.5030	8.9167	0.2200	14.3487
26to25	2.3330	8.8723	0.1400	12.8435
27to26	1.9630	11.3272	0.4000	11.6142
28to27	2.9440	13.1907	0.1910	11.8351
29to28	2.9340	12.9173	0.4410	13.4840
30to29	11.0560	9.7192	0.1100	16.0238
mean	5.7166	10.9025	0.2177	13.9614

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

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). The implementation of the ICP is partially due to Linmi Tao.

7. REFERENCES

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

[2] H. Singh, Xiaoou Tang, E. Trucco, and D. Lane, "Guest editorial special issue on underwater image and video processing," *IEEE Transactions on Oceanic Engineering*, vol. 28, no. 4, pp. 569 – 776, 2003.

[3] B. Kamgar-Parsi, L.J. Rosenblum, and E.O. Belcher, "Underwater imaging with a moving acoustic lens," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 91 – 99, 1998.

[4] S. Negahdaripour and H. Madjidi, "Stereovision imaging on submersible platforms for 3-d mapping of benthic habitats and sea-floor structures," *IEEE Transactions on Oceanic Engineering*, vol. 28, no. 4, pp. 625 – 650, 2003.

[5] R. K. Hansen and P. A. Andersen, "A 3-D underwater acoustic camera - properties and applications,"

	Classic ICP		Proposed ICP	
Frames	Time *(sec.)*	Accuracy *(cm.)*	Time *(sec.)*	Acuracy *(cm.)*
2to1	7.8510	19.6864	0.1800	23.4500
3to2	12.3770	21.4652	0.1200	24.9500
4to3	14.0600	17.4993	0.1200	22.3400
5to4	16.6540	18.1653	0.1100	25.4100
6to5	23.6440	17.4390	0.1100	28.4300
7to6	16.3440	22.5279	0.4410	24.0700
8to7	12.9490	21.6455	0.2700	25.2800
9to8	13.8100	21.5650	0.4200	21.2300
10to9	10.1640	22.6400	0.3400	19.5400
mean	14.2059	20.2926	0.2346	23.8556

Table 2: Performance of registration for sequence 2. The accuracy is given by the residual of the last ICP iteration

in *Acoustical Imaging*, P.Tortoli and L.Masotti, Eds., pp. 607–611. Plenum Press, 1996.

[6] 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.

[7] U. Castellani, 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, July 2002.

[8] 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.

[9] R. J. Urik, *Principles of Underwater Sound*, McGraw-Hill, 1983.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

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

[20] M. Greenspan and M. Yurick, "Approximate kd tree search for efficient icp," in *In Proceedings of 3-D Digital Imaging and Modeling (3DIM 2003)*, 2003, pp. 442– 448.

[21] R. Sagawa, T. Masuda, and K. Ikeuchi, "Effective nearest neighbor search for aligning and merging range images," in *In Proceedings of 3-D Digital Imaging and Modeling (3DIM 2003)*, 2003, pp. 79–86.

[22] Soon-Yong Park and M. Subbarao, "A fast point-to-tangent plane technique for multi-view registration," in *In Proceedings of 3-D Digital Imaging and Modeling (3DIM 2003)*, 2003, pp. 276– 283.

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

[24] U. Castellani, L. Tao, A. Fusiello, and V. Murino, "On-line icp based on a reverse calibration approach," Tech. Rep., Dipartimento di Informatica, Verona University, March 2003.