Proceedings of the Institute of Acoustics

ULTRASONIC IMAGING IMPROVEMENT BY A CO-OPERATIVE APPROACH

Vittorio Murino, Andrea Trucco

Department of Biophysical and Electronic Engineering (DIBE), University of Genoa
Via Opera Pia 11a, I - 16145 Genova, Italy

ABSTRACT

The goal of this paper is the construction of an accurate scene model to be used by an underwater vehicle inspect and navigate in the environment. To this end, a co-operative approach aimed at improving the quality of short-range acoustic images created by a beamforming process is proposed, which is mainly useful for human understanding and robotic applications. Our approach is essentially a voting method: it consists in the integration of two methods devoted to the estimation of the angle of arrival for each time instant and the estimation of the time of arrival for each fixed direction. Each sampled time instant of each beam signal contributes with an opportune quantity on a voting space. At the end of the process, maxima recovery in the voting space and post-processing methods are applied to recover the precise silhouette of a scene. Actual linear-scan sonar data and angular-scan simulated sonar data were used to verify the validity of the method. Results and error measures are reported in order to compare our method with commonly used methods, showing advantages and drawbacks of the co-operative approach.

1. INTRODUCTION

Many works present in the literature address the problem of scene reconstruction and recognition, especially for what concerns the recovery of the structure and understanding of aerial scenes. The proposed techniques dealt with 3D or 2D scene configuration retrieval by using optical (cameras, laser range finders, etc.) or acoustical sensors. The reconstruction is an important task in many applications related to both terrain vehicles and autonomous underwater vehicles moving in structured or, more often, in unstructured environments. Segmentation methods, feature extraction, matching of image sequences, transformation methods [1,2] are used to restore or reconstruct the operative scenario. In underwater environments this task is made more difficult due to the limited use of optical cameras whose range is reduced to a few tens of meters in the best conditions (i.e., pure water) which leads to the almost exclusively use of acoustical sensors. Sonars have the advantages of a longer range with respect to optical cameras and the possibility to estimate 3D information directly. In contrast, they present the drawback of a poorer resolution. Moreover, sonar maps are not easily understandable by non-expert users.

In this paper, we would like to propose a method for the accurate recovery of an underwater scene structure by using sonar data. The precise estimation of scene profile can be useful to an autonomous vehicle for navigating in and inspecting the environment. More precisely, our approach aims at the construction of good-quality short-range acoustic images generated by a beamforming process [3]. This would have to facilitate the scene understanding and improve the results of image-processing algorithms (e.g., aimed at recognition) to be possibly applied on the image. To this end, we propose the integration and the co-operation of two conventional methods widely applied for echo detection in multibeam bathymetric systems [4].

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

The imaging systems taken into account are those composed by a linear array connected to a beamformer in order to form a 2D image in the angle-range plane, i.e., representing a scene section. Generally, the envelopes of the computed beam signals are directly used to set the brightness of image pixels, resulting in an apparent loss of resolution due to halos present around imaged objects [3]. Therefore, the images obtained in this way are not well suited for the understanding of the scene.

The integration of two different methods, both widely applied to the bottom echo detection in sonar multibeam systems aimed at sea-floor bathymetry, is proposed. The scene structure is extracted by considering the beam signals collection along fixed directions, and specifically, 1) fixing the angular direction a priori and estimating the time of arrival for that direction, and 2) estimating the angles of arrival for all the echoes received at each time sample. In other words, our approach is essentially a voting method, i.e., available observations are transformed in a different representation space in which they *vote* according to a precise rule. The novelty of this approach lies in the simultaneous use of the two methods (i.e. a co-operative approach), which generates a representation space with potentially more information to be extracted to build up a more robust and reliable scene structure. This transformed space operates as interface in which the information selected by the two methods are added before the final scene profile extraction and visualisation on the image plane have been carried out. Therefore, it is possible to extract information about the structure of a scene: post-processing methods, like maximum point recovery, morphological filters, edge and contour extraction, can be applied to recover the precise silhouette of a scene. These techniques are commonly used in conventional image processing, but they are not feasible for information recovery in acoustic imaging.

Linear-scan actual sonar data and angular-scan simulated sonar data were used to verify the validity of the idea. Results and error measures are reported in order to compare our method with commonly used methods, showing advantages and drawbacks of our approach.

The paper is organised as follows. Section II describes the voting method exploiting beam signals' information. Section III presents how to extract the scene structure from the voting space. Results on actual and synthetic sonar data are shown in Section IV and, finally, conclusions are drawn in Section V.

2. THE CO-OPERATIVE APPROACH

Despite many techniques have been applied to determine the angle and time of arrival of the bottom echo, we have chosen and generalised for our purpose two major types of procedures: (1) fixing the angular direction and estimating the times of arrival for all the echoes coming from that direction [4,5]; (2) estimating the angles of arrival for all the echoes received at each time sample [4,6]. Then, unlike bottom detection procedures, in the proposed approach two subsequent analysis are performed: one for each possible beam signal (searching for the echo time of arrival) and one for each possible time instant (searching for the echo beam direction). Both analyses are sequentially applied on the beam signals' collection in order to select some beam samples (representing with high probability the echoes from a scene) that are added to a two-dimensional space (i.e., a matrix), called the voting space. There are many different procedures of selection; for instance, it is possible to add to the voting space the beam samples that: (1) exceed a fixed threshold, (2) are local maxima (peaks), (3) exceed a fixed threshold and are local maxima, (4) are the weighted average of a contiguous group of samples exceeding a fixed threshold [4,6]. Once a procedure for the selection at fixed beam and a procedure for the selection at fixed time (also setting the possible thresholds) have been chosen, the voting space can be generated. In some cases, more than two analysis can be performed and their results added to the voting space. For instance, we

could add the samples selected by a weighted average performed at fixed beam, those selected by peak detection performed at fixed time, and, finally, those selected by a weighted average performed at fixed time.

Since an object profile should be visualised in Cartesian co-ordinates whereas, in general, the beamformer operates in polar co-ordinates (i.e., it performs an angular-scan of the image instead of a linear-scan), a scan conversion operation is needed [7]. Some of the above-mentioned selection methods have been designed to operate on data organised in polar co-ordinates. Therefore, the scan conversion will be performed after the sample selection and before the addition on the voting space. The voting space is so organised in Cartesian co-ordinates and this supports the analysis of the voting space by methods designed to operate on data organised in Cartesian co-ordinates. It can be noticed that the voting space is a virtual interface between methods of analysis mainly used in sonar signal processing (in polar co-ordinates) and methods of analysis used in optical image processing (in Cartesian co-ordinates).

When a sample of the beam signals' collection is selected, a contribution is added to the voting space. This operation can be performed in two different ways. In the first case, the contribution of the selected sample vote is proportional to the magnitude of its value. In other words, we add to the correspondent cell of the voting space a number equal to the amplitude of the selected sample, eventually multiplied by a weighting factor. In second case, the importance of the sample vote is not dependent on the magnitude of its value, and we always add a unit value to the correspondent cells of the voting space. The adoption of one of these two options is dependent on the work conditions. The first option is desirable when the object brightness (on the conventional image) is enough uniform and we use only two selection methods. The second option is desirable when major differences are present in the brightness of the objects and we use more than two selection methods.

The voting space combines information extracted from at least two different methods and represents a more efficient organisation of information; this is a useful step toward the image generation and visualisation. The basic philosophy is that significant information (i.e., those representing the actual scene profile) is extracted by both adopted methods and, due to their geometric overlapping, is added to the voting space. Differently, spurious information due to the finite resolution and specific defects of the two methods are, in general, different and not overlapped. As a consequence, the co-operative use of more than one method allows a sort of image averaging visually resulting in an improvement of the signal to noise ratio (SNR), concerning the scene profile. In this sense, the higher the number of the methods used to add information to the voting space, the higher the SNR, resulting in an improved quality of the image.

3. INFORMATION EXTRACTION FROM THE VOTING SPACE

Starting from the voting space, it is possible to extract the significant information about the shape of objects by means of several techniques widely used in image analysis (e.g., information extraction from the Hough space in the context of the Hough-based methods) [1,2], which are not useful when directly applied on an image formed by envelope amplitudes of beam signals. The techniques we tested to move from the voting space to the image plane are: (1) local maxima extraction by means of 2D masks having variable dimension and threshold [8], (2) skeleton extraction by morphological filters [9], (3) edge extraction by Canny procedure [10].

The extraction of local maxima on a 2D matrix is performed by using a mobile square mask having a dimension set by the user (e.g., 5×5). Only the pixels having a value larger than a fixed threshold are

taken into account as possible maxima. The mask evaluates the number of pixel which value is smaller than that of the central pixel. If this number is larger than a predetermined output threshold, the central pixel is considered a local maximum.

The skeleton extraction is an operation of mathematical morphology performed, in this case, by erosion masks [9]. First, attention is focused on image patches containing useful information (i.e., the imaged object blobs). Then, by successive erosions, the skeleton of the object contained inside each image patch is extracted.

While skeleton extractor tries to determine the internal structure of an object blob on the basis of its local characteristics, the edge extractor tries to determine the edges of an object on the basis of its brightness distribution (gradient information) inside the blob. Although the erosion and Canny techniques are well suited for images that are affected by a large amount of noise [10], they are commonly applied on aerial optical images having better quality than acoustic images. Therefore, only in a little number of fortunate cases, the direct application of the edge or the skeleton extraction processes on the image obtained by the beam signals' envelopes (after the scan conversion, obviously) gives good results. In general, the image obtained by the beam signals' envelopes, without the use of some echo detection method, is too much confuse for this kind of operation, giving rise to results of poor quality.

However, the exclusive use of a single echo detection method could be restrictive. The methods for bottom detection presented in the previous section were designed to determine the profile of the sea floor that is present for every beam steering angle. In short-range robotic applications, several beams can be steered in directions along which nothing is present (e.g., consider a forward looking acoustic camera for obstacle avoidance and object inspection). Moreover, if the analysis at fixed beam (along the time) often provides a good accuracy in range determination but a poor lateral object resolution, vice versa, the analysis at fixed time (along the beam steering direction) provides a good lateral resolution but a poor accuracy in range determination. As a consequence, the joint use of these two methods followed by an information extraction appears as an interesting opportunity to be explored.

4. RESULTS

Real and simulated experiments were performed to test the effectiveness of the co-operative approach. The first experiment was actually performed by means of a linear-scan sonar configuration and a scene composed by points. Synthetic data were obtained by the second experiment, performed by means of an angular-scan sonar configuration and a scene composed by flat and continuous objects.

In the former case, the array was obtained by the synthetic-aperture technique by translating linearly a single transducer along 250 different positions. The scene consisted of a set of 65 parallel nylon strings stretched and arranged in such a way that their perpendicular section formed the word "SEA". The carrier frequency was 460 kHz, the pulse duration was 15 μs, and a base-band sampling frequency (i.e., quadrature scheme) equal to 1 Mhz was adopted. The number of formed parallel beam signals was equal to 250 and, for each of them, 250 time samples are considered. The visualised area represented a 19 cm × 38 cm rectangle, 50 cm far from the array. Each image column was formed starting from a given beam signal: the output values provided by the method considered were assigned to the image, point by point. Figure 1 was built by using the amplitude of the beam signal envelopes. The darkness is proportional to the beam amplitude and the beam values were normalised in order to exploit all the dynamics (8 bits, 256

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

grey levels). It is worth noting that the resulting image is characterised by low precision and high presence of halos.

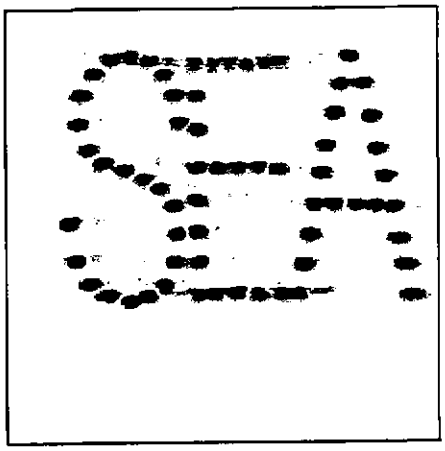
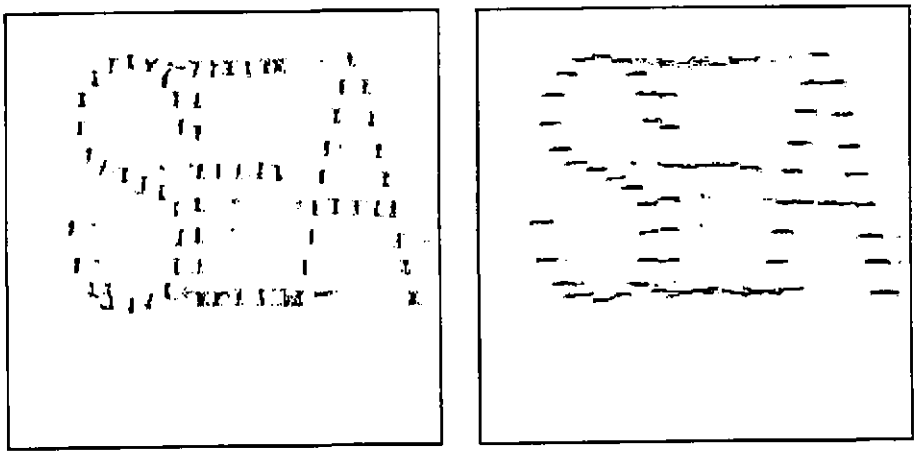


Figure 1. The original image obtained by the envelopes of the beam signal.

Figure 2 shows the results of two selection procedures used to form the voting space: in Fig. 2(a) one can see a local maxima extraction performed at fixed time (i.e., along the beam steering directions) after a 15% threshold, whereas Fig. 2(b) shows the same selection procedure performed, in this case, at fixed beam (i.e., over the time).



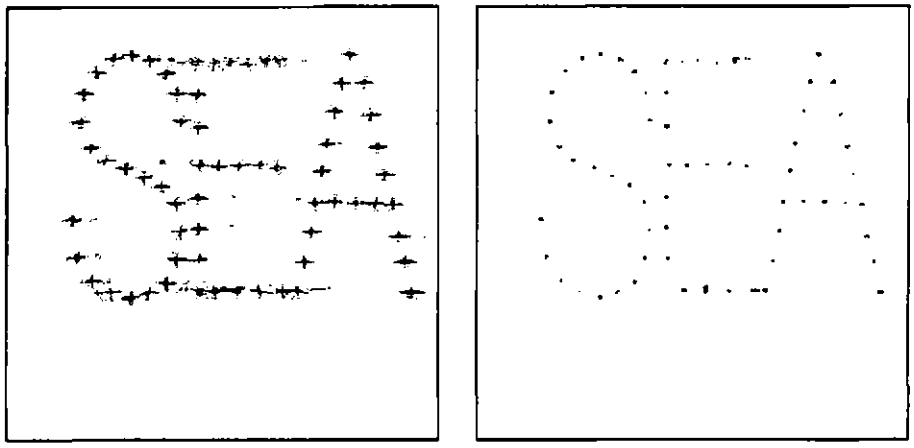
(a) (b)

Figure 2. Results from two selection procedures performed on the original image. (a) A local maxima detection performed at fixed time with a 15% threshold and, (b) the same local maxima detection performed at fixed beam with the same threshold.

In Fig. 3(a) one can see the voting space obtained by the sum with unit weight of the two previous selection procedures. Differently from the original image in Fig. 1, about all points have a cross structure indicating their real position. From this voting space, a local maxima extraction by a 5×5 mask (input threshold equal to 45% and output threshold equal to 16 pixels) was performed giving the result reported

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

in Fig. 3(b). Although all the 65 points are not displayed perfectly (for instance, some of them are visualised as two separate and very close points), in general their position is more precise and the halos are sharply reduced.



(a) (b)

Figure 3. (a) The voting space resulting from the addition of the two selections shown in Fig. 2 and (b) the result of a local maxima extraction performed on such a voting space by a 5×5 mask.

To verify the efficiency of the proposed method when the scene is composed by an arrangement of continuous objects, a simulated experiment was performed by an angular-scan ultrasonic system based on a linear array consisting of 200 $\lambda/2$-equispaced elements, a carrier frequency equal to 500 kHz, a rectangular pulse envelope having a time duration equal to 10 μs, and a base-band sampling frequency (i.e., quadrature scheme) equal to 1 Mhz. The scene was composed by three flat and thin objects (having different lengths) placed parallel in front of the array and whose distance ranged from 57 cm to 62 cm. A dynamic focused beamformer computed 149 beam signals steered from -0.4 rad to 0.4 rad. Figure 4(a) shows the envelope amplitude of the 149 beam signals (for each signal, 250 time samples are considered) after the scan conversion. As in the previous case, the darkness is proportional to the normalised beam amplitude and the resulting image is characterised by low precision and large presence of halos.

Figure 4(b) shows the skeleton extraction performed on the original image, after a 15% threshold operation, to mean that only the samples of the beam signal's collection exceeding the 15% of the maximum value among all the samples are considered. Although the thresholding process allows an acceptable skeleton extraction, the precision of this result is very poor. Figure 5 shows the results of three selection procedures used to form the voting space. In Fig. 5(a) one can see the result of a 30% threshold operation. Figure 5(b) presents the weighted average with a threshold equal to 10% performed at fixed beam, i.e., along time. One can note that, as in Fig. 4(b), the precision of the result is quite poor. Figure 5(c) displays the result of local maxima detection performed at fixed time (i.e., along the beam steering directions) after a 10% threshold. All these images are presented after a scan conversion and use all the dynamics of the image.

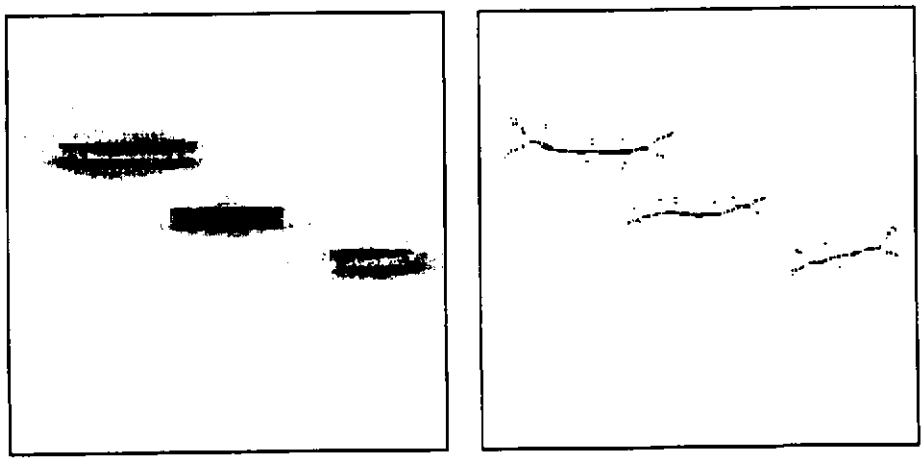
In Fig. 5(d) one can see the voting space obtained by the sum with unit weight of the three previous selection procedures. Differently from the original image of Fig. 4(a), the three objects have a dark kernel indicating their profiles. From this voting space, a skeleton extraction was performed resulting in the

image shown in Fig. 6(a). Although the silhouette is not perfectly flat, the profile of the three objects is easy understandable, like in Figs. 4(b) and 5(b), but in this case a better precision has been achieved.

Finally, Fig. 6(b) displays the local maxima extraction performed by a 3×3 mask (input threshold equal to 45% and output threshold equal to 5 pixels) operating on a different voting space, which was generated by the addition of a weighted average with a threshold equal to 15% performed at fixed beam and a 15% threshold operation. Also in this case, the previous consideration concerning precision and flatness are valid. This fact is confirmed by a measure of the mean square error (MSE) between one of the current images and the ideal image. The ideal image is available as the experiment was simulated and, therefore, we have a precise knowledge of the scene. Since the images visualises a section of the scene, the computation of the MSE was performed as follows: (a) for each pixel of the current image different from white (background) we add the square distance (in square millimetres) between it and the narrower pixel different from white (background) of the ideal image to the total distance, (b) for each pixel of the ideal image different from white we add the square distance between it and the narrower pixel different from white of the current image to the total distance, (c) the total distance is divided by the number of sums performed.

We summarised the MSE values in Table A using as current images those obtained by a single elaboration on the original image (skeleton extraction and weighted average, Figs. 4(b) and 5(b)) and the two images obtained (see Figs. 6(a) and 6(b)) by the information extraction operated on the voting space visualised in Fig. 5(d). It is worth noting that the resulting MSE computed on images extracted by the co-operative method is one order of magnitude lower than the MSE computed on the images extracted with the other methods considered.

As shown by the images and the error measures, the obtained images (see Figs. 3(b) and 6) visualise in a better way the acoustic information, limiting the halos around the objects and the effects due to the finite resolution, thus improving the understanding by an automatic or human operator.



(a) (b)

Figure 4. (a) The original image obtained by the direct scan conversion of the beam signal amplitude and (b) a skeleton extraction operated on such an image after a 15% thresholding process.

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

5. CONCLUSIONS

In this paper, a co-operative approach is proposed to extract the scene structure in 2D acoustic images for underwater robotic applications. Co-operative means that the collection of the beam signals is not directly visualised, but two methods, commonly used in quite different field (i.e., bottom detection), are applied and their results are combined, after scan conversion, in a different representation space called voting space. Such a space is a sort of interface in which the information extracted by the two methods interacts and on the basis of which the scene structure can be recovered by methods commonly applied in classical image analysis.

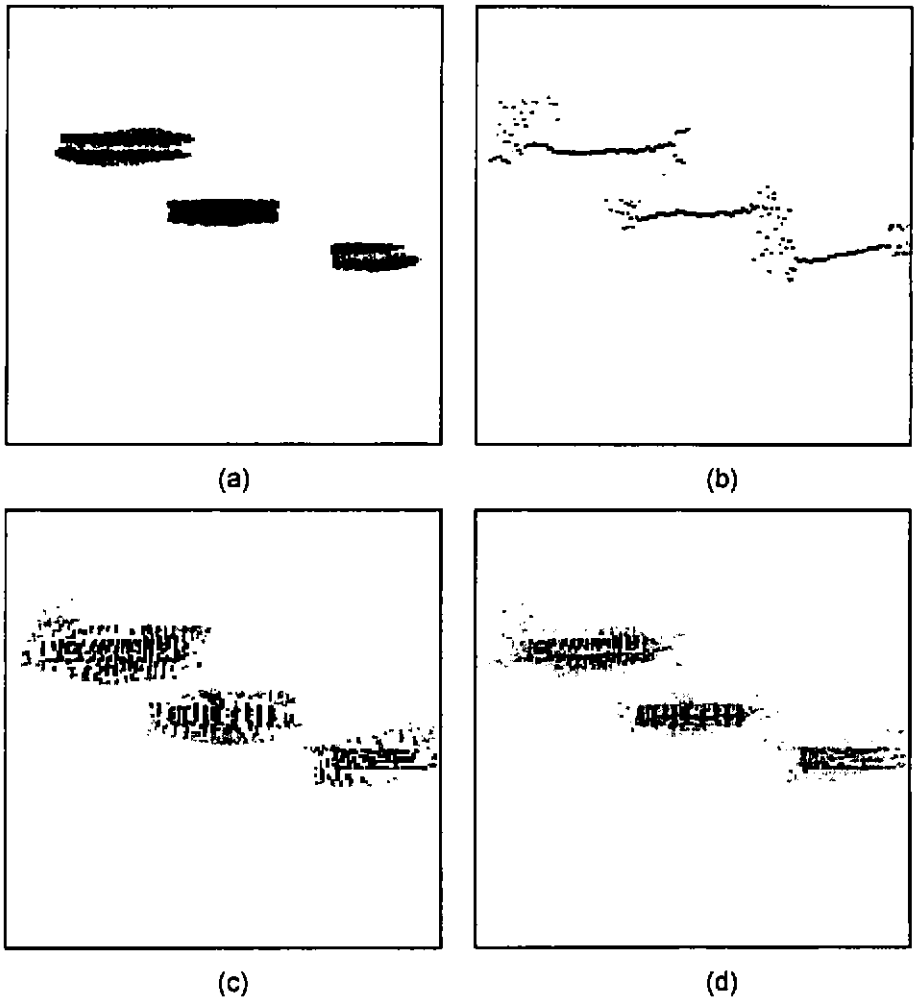
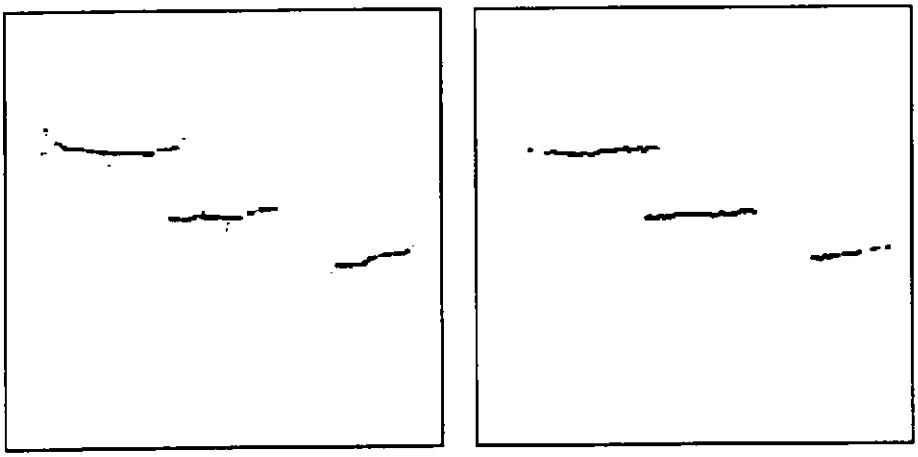


Figure 5. Results from three selection procedures performed on the original image in Fig. 4 and the resulting voting space. (a) A 30% threshold operation, (b) a weighted average operation performed at fixed beam with a 10% threshold, (c) a local maxima detection performed at fixed time with a 10% threshold and, (d) the voting space resulting from the addition of the three previous operations.

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

Current image	Figure	MSE [mm^2]
Skeleton extraction on the original image	4(b)	20.2
Weighted average on the original image	5(b)	66.7
Skeleton extraction on the voting space	6(a)	7.2
Maxima extraction on the voting space	6(b)	5.1

Table A. Mean square error (MSE) for some images representing the scene structure obtained with or without the co-operative approach.



(a) (b)

Figure 6. (a) The skeleton extraction performed on the voting space shown in Fig. 5 and (b) the result of a local maxima extraction performed by a 3×3 mask.

The mixture of two acoustical methods and some techniques applied in optical image processing appears a general framework to avoid drawbacks that such methods present, if singularly applied in short-range acoustic imaging. Future work is needed to test on real data the effectiveness of the proposed approach, to formalise it inside a more mathematical context, and to design a sort of automatic procedure able to choose the better operation sequence among the several possibilities.

6. REFERENCES

[1] B K P HORN, *Robot Vision*, The MIT Press, Cambridge, Massachusets, (1987)
[2] D VERNON, *Machine Vision*, Prentice Hall, New York, (1991)
[3] C F SCHUELER, H LEE, G WADE, 'Foundamentals of Digital Ultrasonics Imaging', *IEEE Trans. Sonics Ultrason*, 31 p195 (1984).
[4] C DE MOUSTIER, 'Signal Processing for Swath Bathymetry and Concurrent Seafloor Acoustic Imaging', *Acoustic Signal Processing for Ocean Exploration*, J M F MOURA and I M G LOURTIE (eds.), p329 (1993).

ULTRASONIC IMAGING IMPROVEMENT BY A COOPERATIVE APPROACH

[5] C DE MOUSTIER, F N SPIESS, D PANTZARTZIS, *et al.*, 'First Results from a Deep Tow Mulribeam Echo-Sounder', *IEEE Int. Conf. Oceans 94 Osates*, Brest (F), III p244 (1994).
[6] C DE MOUSTIER, D ALEXANDROU, 'Angular dependence of 12-kHz seafloor acoustic backscatter', *Journ. of Acoust. Soc. of America*, 90 p522 (1991).
[7] A DRUKAREV, K KOSTANTINIDES, G SEROUSSI, 'Beam Transformation Techniques for Ultrasonic Medical Imaging', *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, 40 p717 (1993).
[8] V MURINO, G L FORESTI, C S REGAZZONI, G VERNAZZA, 'Grouping of Rectilinear Segments by the Labeled Hough Transform', *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59 p22 (1994)
[9] J SERRA, *Image Analysis and Mathematical Morphology*, Academic Press, London, (1982).
[10] J CANNY, 'A Computational Approach to Edge Detection', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8 p679 (1986).