#

Three-dimensional underwater acoustical imaging

Andrea Trucco[1], Maria Palmese[1], Andrea Fusiello[2], and Vittorio Murino[2]
[1] *Department of Biophysical and Electronic Engineering, University of Genova, Italy*
[2] *Department of Computer Science, University of Verona, Italy*

1. INTRODUCTION

Acoustic imaging is an active research field that aims to study techniques for the formation and processing of images generated by raw signals acquired by an acoustic system [1]. Our purpose is to present a brief survey concerning the generation and processing of acoustic images for underwater applications [2,3], especially focusing on algorithms for three-dimensional (3-D) imaging. Like optical systems, acoustic systems can generate an image by processing the waves backscattered from the objects of a scene. The relative easiness of measuring the time-of-flight of an acoustic signal makes it possible to generate not only acoustic 2-D images similar to optical ones but also range estimates that can be used to produce a real 3-D map.

The main advantage of acoustic imaging systems over optical ones is that they properly work also beyond the optical visibility range. Underwater optical vision provides images with finer resolution, but its range is limited to a few tenths of meters in very good water conditions, and the situation is even worse when mud in suspension raised by underwater tasks sharply reduces visibility [2]. Although larger than the range of optical visibility, the range of applicability of a 3-D acoustic imaging system varies according to the specific sensor adopted and the signal frequency characterizing the sensor. Generally, high frequencies are utilized (from about a hundred kilohertz to few megahertz) over a range going from some centimeters to a hundred meters.

In general, the scene under investigation is first insonified by an acoustic signal, then the backscattered echoes acquired by the system are processed to create an image of the scene. This process can be performed by two different approaches [3]:

use of an acoustic lens followed by a retina of acoustic sensors, or acquisition of echoes by a two-dimensional (2-D) array of sensors and subsequent processing by adequate algorithms, thus avoiding the need for a physical lens. Such algorithms belong to the beamforming or the holography class.

To date, unlike the applications of two-dimensional imaging, the practical utilization of 3-D acoustic imaging is limited because of several scientific and technological issues that make it difficult and expensive to produce imaging systems working in real time. However, some recent achievements, in both hardware and software areas, allowed the production of a few prototypes and commercially available systems.

Once an acoustic image is generated, several image processing algorithms can be designed for the detection and reconstruction of the objects contained in the observed scene. Depending on the type of the generated image, i.e., whether it embeds three-dimensional (3-D) information together with intensity data or not, several approaches can be followed for the image segmentation. Actually, image filtering and segmentation are the first and fundamental phases after the generation of an image, and constitute the base for any subsequent processing and interpretation, and segmentation and reconstruction are natural subsequent steps. In the paper, such approaches will be surveyed, especially those devoted to 3-D image processing, leaving aside other methods specifically devoted to high-level image interpretation. Finally, a complete system for segmentation and reconstruction of underwater tubular structures is described, also showing the related high-level applications.

In this chapter, a model of the interaction of the acoustic energy with the scene to be imaged is presented in Section 2. In Section 3, specific methods (i.e., beamforming, holographic, and lens-based techniques) are outlined, focusing attention on the 3-D beamforming approach. In Section 4, a real system that exploits a matrix approach based on the holographic method is detailed, and in Section 5, the arrangement and representation of a 3-D image are described.

Section 6 deals with the state of the art literature about the 3-D image processing, focusing especially on acoustic 3-D image segmentation and reconstruction. A specific method for the segmentation and reconstruction of 3-D acoustic images is detailed in Section 7. In this part, a complete image processing system is presented, focusing on several segmentation algorithms, and showing possible applications which inspired this work. Finally, conclusions are drawn in Section 8.

2. DATA MODEL

In underwater 3-D imaging, a typical scene is composed of several solid and continuous objects. One of the most powerful methods for computing the field returned by a complex and realistic underwater object is to represent its surface by a collection of densely packed point scatterers or small facets [4,5,6].

We assume that the imaged scene is made up of M point scatterers; the *i*-th scatterer is placed at the position $\mathbf{r}_i = (x_i, y_i, z_i)$ and its distance from the coordinate origin is equal to $r_i = |\mathbf{r}_i|$, as shown in Fig. 1a. We can define the plane z = 0 as the plane that receives the backscattered field (see Fig. 1a). If an acoustic pulse q(t) is

.

emitted by an ideal point source placed in the coordinate origin, and if we assume that a spherical propagation occurs inside an isotropic, linear, absorbing medium, then the Fourier transform of the pressure measured at the position **p** = (x, y, 0) and due to the action of the M point scatterers [3] is equal to:

$$S(\omega, \mathbf{p}) = \sum_{i=1}^{M} D_i \frac{\omega^2 e^{-2\alpha(\omega) r_i}}{c^2 r_i^2} Q(\omega) e^{-j\frac{\omega}{c}(r_i + |\mathbf{p} - \mathbf{r}_i|)} \quad (1)$$

where $Q(\omega)$ is the Fourier transform of the emitted pulse $q(t)$, $\alpha(\omega)$ is the absorption coefficient of water, $\omega = 2\pi f$ is the angular frequency, f is the frequency, c is the sound velocity in the medium, and D_i is a coefficient dependent on the radius of the i-th scatterer and on the densities and the compressibilities of the propagating medium and of the i-th scatterer [3,5].

For typical frequencies and distances involved in 3-D underwater imaging, it has been shown [3] that the factor ω^2 in (1) is compensated for by water absorption. Therefore, we can introduce a constant factor σ_i (dependent on the distance r_i) to replace the term $\omega^2 \exp\{-2\alpha(\omega) r_i\}$, and we can rewrite equation (1) as:

$$S(\omega, \mathbf{p}) \cong \sum_{i=1}^{M} C_i \, Q(\omega) e^{-j\frac{\omega}{c}(r_i + |\mathbf{p} - \mathbf{r}_i|)} \quad (2)$$

$$C_i = D_i \frac{\sigma_i}{c^2 r_i^2} . \quad (3)$$

Moreover, if the cone whose vertex is in the coordinate origin and that contains the scene volume to be imaged has an angular aperture that is not too wide, we can use the Fresnel approximation for the Green function [7], [8] to write the term $|\mathbf{p} - \mathbf{r}_i|$ in the phase of (2) as follows:

$$|\mathbf{p} - \mathbf{r}_i| \approx r_i - \hat{\mathbf{r}}_i \cdot \mathbf{p} + \frac{p^2}{2 r_i} \quad (4)$$

where p is the modulus of **p,** and $\hat{\mathbf{r}}_i$ is a unitary vector equal to $\mathbf{r}_i / r_i$.

Finally, if we restrict our reasoning to the plane y = 0 (see Fig. 1b), then **p** = (x, 0, 0), $\mathbf{r}_i = r_i(\sin\beta_i, 0, \cos\beta_i)$, β_i being the angle between the vector $\mathbf{r}_i$ and the z-axis, and (2) can be simplified as follows:

$$S(\omega, x) = \sum_{i=1}^{M} C_i \, Q(\omega) e^{-j\frac{\omega}{c}\left(2 r_i - x \sin\beta_i + \frac{x^2}{2 r_i}\right)} . \quad (5)$$

The angle β_i refers to the arrival angle as it indicates the direction of the echo of the i-th scatterer.

The approximation for the term $|\mathbf{p} - \mathbf{r}_i|$ that has been discussed is of great practical importance as dealing with all combinations of **p** and $\mathbf{r}_i$ is sometimes very difficult (even in modern imaging systems exploiting digital technology).

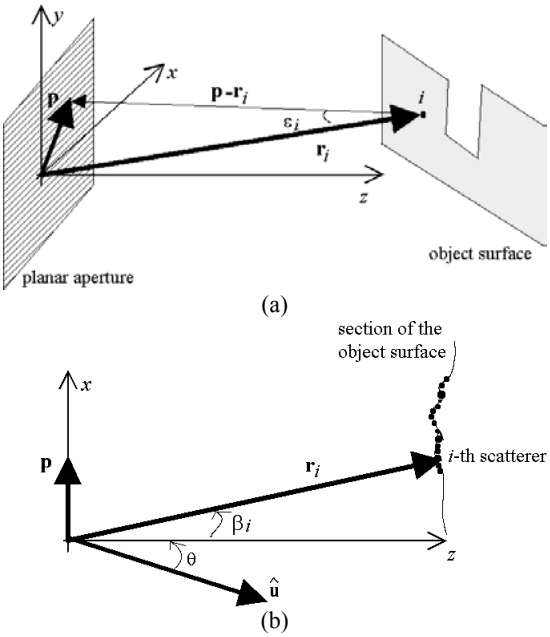


Figure 1. Notation and geometry of the data model. (a) 3-D representation. (b) 2-D projection on the plane $y = 0$.

3. GATHERING 3-D INFORMATION

In 3-D underwater imaging, a scene is typically illuminated by the emission of an acoustic pulse, and the backscattered echoes are collected over a 2-D aperture and processed to create an image of the scene. We can choose between two different approaches [3,1] to perform the operation of echo processing: use of acoustic lens followed by a retina of acoustic sensors, or acquisition of the echoes impinging on a planar array of sensors and processing of such echoes by appropriate algorithms. Beamforming and holographic methods are two processing approaches that can be successfully exploited in 3-D imaging systems.

Acoustic lenses work like optical ones: backscattered echoes are focused on an image plane where a 2-D retina of sensors transforms the acoustic image into electrical signals. Thanks to the facility of measuring the time-of-flight of an acoustic pulse, one can generate range estimates that can be utilized to produce a real 3-D map. Each sensor of the retina placed behind a lens receives a signal that represents the scene response coming from a well-defined direction. By collecting the signals of all the retina sensors, one can obtain complete information about the 3-D structure of the scene.

Beamforming systems collect backscattered echoes by a 2-D array of sensors only once; then, they arrange the echoes in such a way as to amplify the signal coming from a fixed direction (steering direction) and to reduce all the signals coming from any other directions. As the output signal gives information about the

.

scene structure in the steering direction, it is possible to create a 3-D image by repeating the beamforming process after fixing many adjacent steering directions, as in a raster-scan operation.

Also holographic systems start from the echoes acquired by a 2-D array of sensors, but they aim to reconstruct the 3-D structure of a scene by back-propagating the received signals. Acoustic holography is a special case of inverse diffraction and is performed through the inversion of the propagation and scattering equations. An image is not generated by a raster-scan operation, but the holographic algorithm produces the whole image at the same time.

In the rest of this work, we focus on the beamforming approach and, in a less detailed way, on the holographic approach.

3.1 Beamforming

Beamforming is a spatial filter that linearly combines the temporal signals spatially sampled by a discrete antenna, i.e., an array of sensors placed according to a known geometry.

Let us consider a set of N point-like and omnidirectional sensors that constitute a receiving 2-D array, numbered by the index l, from 0 to N-1. Denoting by $\mathbf{p}_l$ the position of a given sensor of the set on the plane z = 0 and by $s_l(t)$ the signal received by the sensor and linearly proportional to the pressure field, one can compute the beam signal [9], b(t, $\hat{\mathbf{u}}$), steered in the direction of the unitary vector $\hat{\mathbf{u}}$ by using the following definition:

$$b(t,\hat{\mathbf{u}}) = \sum_{l=0}^{N-1} \varphi_l \, s_l\left(t - \tau(\hat{\mathbf{u}}, r_0, l)\right) \tag{6}$$

$$\tau(\hat{\mathbf{u}}, r_0, l) = \frac{r_0 - |\mathbf{p}_l - r_0\hat{\mathbf{u}}|}{c} = \frac{r_0 - \sqrt{{r_0}^2 + |\mathbf{p}_l|^2 - 2r_0\hat{\mathbf{u}}\cdot\mathbf{p}_l}}{c} \tag{7}$$

where φ_l are the weights assigned to each sensor and r_0 is the focusing distance. The net result is the formation of a temporal signal in which the contributions coming from the direction $\hat{\mathbf{u}}$ and the distance r_0 are amplified, whereas those coming from other directions and distances are attenuated.

Now, let us restrict our reasoning to the plane y = 0 and assume all the weights φ_l to be unitary; then the steering direction can be indicated by the angle θ measured with respect to the z axis (see Fig. 1b). Beginning from equations (5) and (6), by applying the Fresnel approximation to the delay in (7) and by assuming the distance of all the scatterers making up the scene to be equal to r_0, one can obtain [3] the following expression for the beam signal:

$$b(t,\theta) = \sum_{i=1}^{M} q\left(t - \frac{2r_i}{c}\right) C_i \, BP_{BMF}(\overline{\omega}, \beta_i, \theta) \cdot \tag{8}$$

$$BP_{BMF}(\omega,\beta,\theta) = \frac{\sin\left[\omega N d(\sin\beta - \sin\theta)/2c\right]}{\sin\left[\omega d(\sin\beta - \sin\theta)/2c\right]} \tag{9}$$

where $BP_{BMF}(\omega, \beta, \theta)$ is a reception diagram commonly called beam pattern, which depends on the arrival angle β, the steering angle θ, and the angular frequency ω. We have assumed the array to be equispaced and centred in the coordinate origin, d

to be the inter-element spacing, and the beam pattern to have a constant profile over the signal bandwidth centred in the angular frequency $\overline{\omega}$.

With reference to an array composed of 40 1.5 mm-spaced elements, Fig. 2 shows some beam patterns as a function of the arrival angle (visualized on a logarithmic scale normalized to 0 dB, provided that the absolute values are considered) for different frequency values and steering angles.

Equation (8) clearly shows that each scatterer contributes to the beam signal by adding a replica of the acoustic pulse q(t), delayed on the basis of its distance r_i, weighted by its constant C_i (strictly related to the scatterer reflectivity) and by the beam pattern value that depends on the discrepancy between the arrival angle β_i of the scatterer and the steering angle θ. Owing to the profile of the beam pattern, the contributions of the scatterers characterized by arrival angles very close to the steering angle have predominant magnitudes.

The beam pattern of an imaging system is very useful to evaluate the system performances. A conventional beam pattern presents a main lobe in the steering direction and sidelobes of minor magnitudes in other directions. The width of the main lobe is the measure of the angular resolution (also called lateral resolution) of the imaging system, whereas the generation of artefacts degrading useful information depends on the level of the sidelobes. When the weight coefficients are unitary, the following equation [9] provides the arrival angles at which the main lobe is reduced to –3 dB, thus giving a measure of the main lobe's width:

$$\sin\beta_{-3dB} = \sin\theta \pm 0.44\frac{\lambda}{Nd}. \tag{10}$$

To improve the angular resolution, one can increase the number of elements, the frequency, or the inter-spacing. Nevertheless, if the frequency increases, a grating lobe appears in the beam pattern (see Fig. 2d). This aliasing effect is due to the spatial under-sampling that occurs when the array elements are equispaced and the inter-element spacing is larger than $\lambda/2$. To avoid ambiguity effects resulting from the presence of grating lobes in the beam pattern, it is necessary to limit both the insonification and steering operations inside a narrower angular sector [1]. In greater detail, the maximum steering angle to avoid ambiguities is the following:

$$\theta_{max} = \pm\arcsin\frac{\lambda}{2d}. \tag{11}$$

The beam patterns have been developed and discussed for a plane space and taking into consideration a linear array, although this configuration does not allow one to obtain 3-D images but only a section of a scene profile (see Fig. 1b). To steer the beam inside a 3-D space, two steering angles should be used and a planar array is mandatory. Figure 3a shows the geometry and the notation for a planar array, where β_a and β_e are the azimuth and elevation arrival angles, respectively, and θ_a and θ_e are the azimuth and elevation steering angles, respectively. Figure 3b shows the beam pattern for an array composed of 20×20 elements that are $\lambda/2$ apart, when the steering angles are $\theta_a = 40°$ and $\theta_e = 0°$.

.

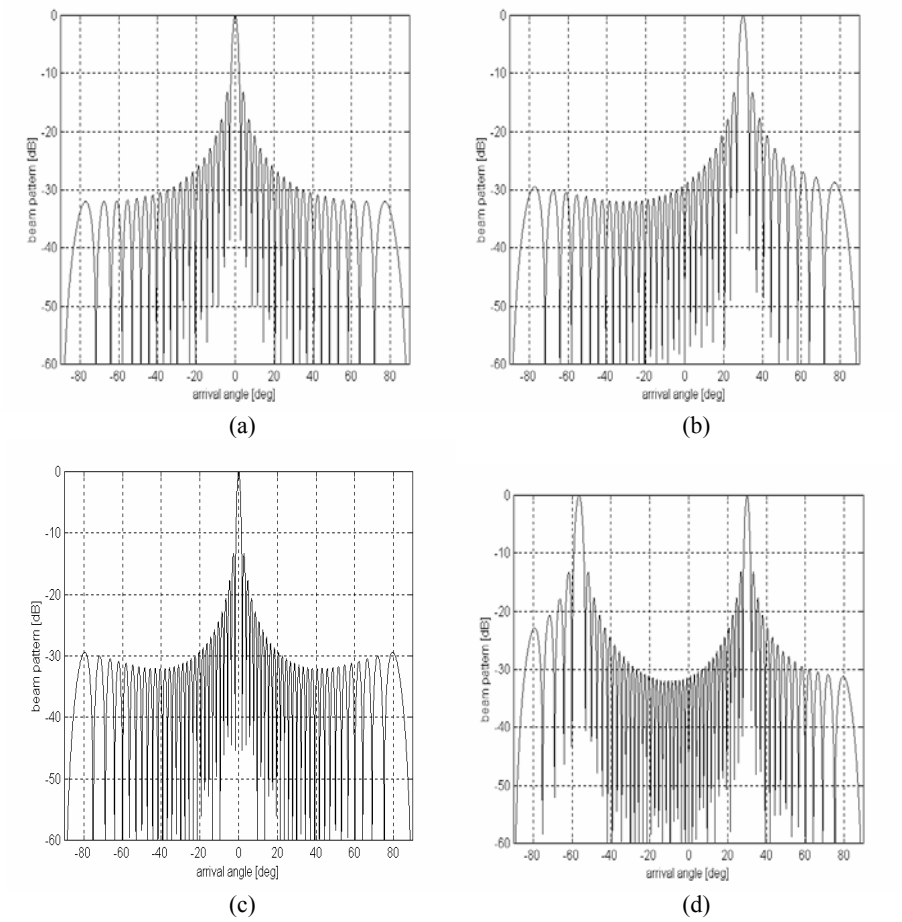


Figure 2. Beam power patterns of a 40-element array with 1.5 mm spacing and unitary weights. (a) Frequency = 500 kHz (i.e., inter-element spacing = $\lambda/2$), steering $\theta = 0°$. (b) Frequency = 500 kHz (i.e., inter-element spacing = $\lambda/2$), steering $\theta = 30°$. (c) Frequency = 750 kHz (i.e., inter-element spacing = $3\lambda/4$), steering $\theta = 0°$. (d) Frequency = 750 kHz (i.e., inter-element spacing = $3\lambda/4$), steering $\theta = 30°$.

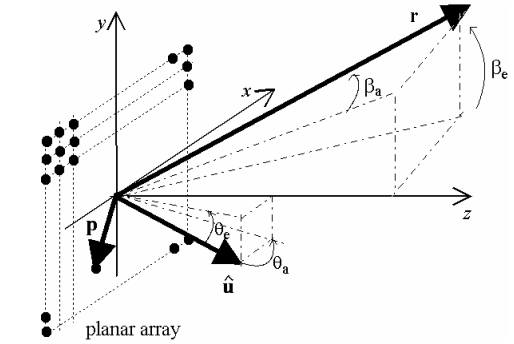
3.2 Resolution cell

At this point, it may be useful to clarify how beam signals can be exploited to generate a 3-D image. To this end, it is necessary to define the range resolution and to recall the angular resolution. Inside a wide interval called depth of field, centred at the focusing distance, the range resolution is defined as the minimum distance between two equal scatterers (placed in the same beam direction) that is needed to resolve their responses. The range resolution [6,10] is typically inversely proportional to the bandwidth of the emitted pulse q(t). Analogously, the angular resolution is the minimum angular spacing that allows two equal scatterers, placed at

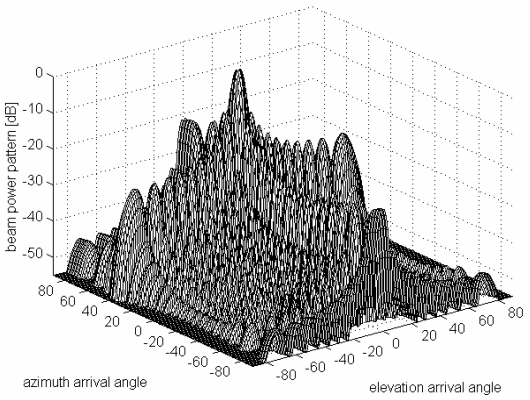
the same distance from the array centre, to be resolved. A resolution cell can be defined as a volume whose dimensions can be derived from the knowledge of the system's angular and range resolutions and inside which it is not possible to separate scatterer contributions. To arrange a grid of resolution cells that covers the whole volume to be imaged without leaving free holes one should plan the number of beam signals to be computed and their angular spacing and choose the sampling frequency of each beam signal in agreement with the range resolution [9].

The acoustic responses of the scatterers contained inside a cell are a function of the reflectivity and of the relative position of each scatterer; the reflectivity value to be assigned to each cell can be derived from the amplitude of the beam signal over the time interval related to such cell.

Resolution cells span the whole volume to be imaged, whereas scatterers are mainly placed on object surfaces; then, starting from this collection of cells, one can organize the effective information in more compact ways, discarding useless cells or directly extracting object surfaces.



(a)



(b)

Figure 3. (a) Notation and geometry of a 2-D array. (b) Beam power patterns of an array composed of 20×20 elements that are λ/2 far from one another, when the steering angles are $\theta_a = 40°$ and $\theta_e = 0°$ and the weights are unitary.

.

4. MATRIX APPROACH AND REAL SYSTEMS

The holographic approach is characterized by many possible approximations that provide different practical algorithms. Here we present a holographic method able to avoid the Fresnel approximation (whose validity region is often too restricted) and to process wide-band signals (very important to obtain a good range resolution). This method is based on a matrix formulation that is used by a few real 3-D systems [11,12].

To this end, it is necessary to express the data model presented in (2) in a different manner. Let us denote by $\mathbf{s}(\omega,\mathbf{p})$ the N×1 column vector of the field received by N sensors placed at $\mathbf{p}_l$ (l being an integer ranging from 1 to N), and by $\mathbf{c}(\omega,\mathbf{r})$ the M'×1 column vector of the reflectivity of each resolution cell contained in the scene volume to be imaged. The number of cells, M', is different from the number of scatterers, M, used in the previous data model, and if the i-th cell placed at $\mathbf{r}_i$ (i being an integer ranging from 1 to M') does not contain any object, its reflectivity may be null. The data model presented in (2) can be rewritten as follows:

$$\mathbf{s}(\omega,\mathbf{p}) = \mathbf{U}(\omega,\mathbf{p},\mathbf{r})\,\mathbf{c}(\omega,\mathbf{r}) \tag{12}$$

where $\mathbf{U}(\omega,\mathbf{p},\mathbf{r})$ is an N×M' transfer matrix whose element u_{li} is defined as:

$$u_{li} = Q(\omega)\, e^{-j\frac{\omega}{c}\left(r_i + |\mathbf{p}_l - \mathbf{r}_i|\right)} \tag{13}$$

The imaging process lies in estimating the vector $\mathbf{c}$ starting from the knowledge of $\mathbf{U}$ and the experimental measure of $\mathbf{s}$.

As, in a 3-D imaging system, the total number of resolution cells is too large to be handled by a single matrix, one can rewrite (12) in the following way:

$$\mathbf{s}_n(\omega,\mathbf{p}) = \mathbf{U}_n(\omega,\mathbf{p},\mathbf{r})\,\mathbf{c}_n(\omega,\mathbf{r}) \tag{14}$$

where $\mathbf{c}_n$ is a regular grid of cells on a spherical surface of radius r_n, $\mathbf{s}_n$ is the Fourier transform of the field received over a brief time interval starting at $t_n = 2r_n/c$, and $\mathbf{U}_n$ is the related transfer matrix. As a result, the resolution cells of the 3-D volume are organized as a sequence of concentric spherical layers, and a specific transfer matrix is defined for each layer. The best estimate, $\tilde{\mathbf{c}}_n$, of the vector $\mathbf{c}_n$ [11,12], assuming M' > N, is given by the minimum-norm solution of (14), that is:

$$\tilde{\mathbf{c}}_n = \mathbf{U}_n{}^H \left(\mathbf{U}_n \mathbf{U}_n{}^H\right)^+ \mathbf{s}_n \tag{15}$$

where ^{H} is the complex conjugate and transpose, $^+$ is the pseudoinverse, and the matrix arguments are neglected to simplify the notation.

Among 3-D underwater imaging systems that exploit the above matrix approach the EchoScope 1600 designed and produced by OmniTech in Norway [12,13,14] is the only off-the-shelf 3-D acoustic camera available.

The number of acoustic sensors is 1600, which make up a 40×40 array with a 19.5-cm side, able to work at three different frequencies: 150, 300, and 600 kHz. The value of the angular frequency ω is fixed, thus the Fourier transform is not necessary as the amplitude and phase data on a received signal are obtained by a

quadrature reception [9]. For each spherical surface of radius r_n, a grid of 64×64 partially overlapped resolution cells is determined, and the best estimate of the related vector $\mathbf{c}_n$ is obtained by means of (15), where the spectral theorem is exploited to decompose the matrix ($\mathbf{U}_n \mathbf{U}_n^H$), as described in [12]. As the inter-sensor spacing is fixed, to avoid grating-lobe effects the viewing and insonification angles are bounded as follows: 90°×90° at 150 kHz, 50°×50° at 300 kHz, and 25°×25° at 600 kHz, and the lateral resolutions (measured at –3 dB, with unitary weights) are 2.5°, 1.3°, and 0.6°, respectively. The allowed detection range spans from 1 to 100 m, and, thanks to the powerful computation architecture, it is possible to achieve on average five 3-D images per second with a range resolution of 5 cm. Owing to the above-mentioned features, the EchoScope 1600 represents a powerful way of obtaining real-time 3-D information over a very wide range of distances, thus allowing one to choose the best tradeoff among resolution, viewing angle, and side-lobe level.

5. 3-D IMAGE REPRESENTATION

Independently of the imaging system used, the obtained 3-D information is organized into a dense lattice of resolution cells of different dimensions to cover the whole volume of interest. Essential information about each cell lies in the coordinates of the cell centre and in the acoustic amplitude or intensity. Sometimes, to achieve an initial cleaning of the imaged volume, a first process discards all the resolution cells whose acoustic amplitudes do not exceed a given threshold. The aim is to remove the effects of electronic noise and, in particular, of the sidelobes in the beam pattern of the acoustic system. The resolution cells can be projected onto a regular 3-D grid of voxels (volume elements) of constant dimensions in a Cartesian coordinate system. This operation is called "scan conversion." To avoid loss of resolution, each voxel should be smaller than the smallest resolution cell to allow one or more voxels to be contained in a given resolution cell. To this end, there are two possible approaches: (1) searching for the voxel containing the centre of a given resolution cell, assigning the acoustic amplitude of the cell to such a voxel, repeating this procedure for all the cells, and, finally, interpolating in order to assign a value also to each voxel that does not contain a cell centre; (2) computing how many voxels are inside a given cell on the basis of the cell dimensions, assigning the acoustic amplitude of the cell to the voxels, repeating this operation for all the cells, and, finally, checking if some voxel is unassigned. The resulting regular grid of voxels may be a useful starting point for many post-processing purposes, and also facilitates the following operations: (1) extracting a sequence of 2-D images representing the acoustic responses of successive planar slices of a scene, (2) extracting several 2-D images cut with different orientations and projecting the 3-D volume by performing an orthogonal or perspective projection integrating the acoustic amplitudes of the cells met along the projection line.

If we are interested in the first object surface that is met starting from the array centre and following the steering direction, a more compact method to exploit the information contained in the set of resolution cells is to use a couple of 2-D images,

.

one for the amplitude and the other for range information. In this case, we expect to extract only two values from each beam: the distance of the object surface and the amplitude of the acoustic response. The distance of a scattering object can be estimated by searching for the maximum peak (or the first peak that exceeds a given threshold) of the envelope of the beam signal and, starting from the time instant at which that peak occurs, by deriving the distance from simple geometrical considerations. Therefore, for each steering direction $\hat{\mathbf{u}}$, a triplet ($\hat{\mathbf{u}}$, distance, amplitude) can be extracted, so 3-D data reduce to a set of triplets, the number of which is equal to the number of beam signals. Then, the collection of triplets can be projected onto the two images, according to the kind of information contained in each image.

6. ACOUSTIC 3-D IMAGE PROCESSING

A large literature is available concerning the segmentation of range images, but a few papers deal with acoustic range image processing. In general, looking at the state of the art of three-dimensional image processing, three possible research lines can be identified for range image segmentation: edge-based, region-based, and hybrid methods.

Edge-based methods [15,16,17,18,19] aim at determining step and roof edge points using separate procedures, mainly based on the analysis of local depth discontinuity, and the discontinuity of first derivative, respectively. Due to the low resolution and high noise level typical of acoustic images, edge-based approaches are not pursued in this application field, and the most of the methods are devoted to the design of region-based algorithms.

The goal of region-based methods is to determine image areas with similar differential geometrical properties. Typically, all these approaches follows the so-called *local-to-global* scheme composed by two phases. A pre-processing filter is preliminary applied to reduce noise effects. In the first *local* phase, an over-segmentation is first determined, labelling each point on the basis of the local differential features. The second *global* phase consists in a fusion stage, where regions detected in the 1st phase are merged if leading to a better segmentation. In practice, the first important step lies in the choice of the type of surfaces (i.e., models) one would like to classify. The models typically used are the six fundamental surfaces of differential geometry, i.e., plane, peak, pit, roof, valley, saddle. The 2nd phase is normally an iterative procedure aimed at refining the first segmentation. Although good results are typically obtained, they present convergency problems, so that computational complexity cannot be estimated a priori. This approach is the most common found in literature [20,21,22,23,24,25], and mostly used to 3-D acoustic image processing.

Hybrid methods combine the search for edge points with the determination of actual regions to get a real true segmentation of an image [24,26,27].

Actually, it can be noticed that region-based methods are more accurate, while presenting a higher complexity, also in computational terms, with respect to the other approaches. Among the region-based algorithms, surface fitting approaches have proved higher performances, but, to have a large representational accuracy, a

high degree of the polynomial should be used, at the expense of the computational cost.

Past work in acoustical imaging concerned mainly with restoration and segmentation of acoustic data, typically acquired by sidescan sonar. Speckle noise reduction from sidescan sonar images was addressed by using Simulated Annealing and Iterative Conditional Modes [28]. The image restoration is performed assuming speckle noise modelled as Rayleigh or Gaussian probability density function. The Graduated Non-Convexity [29] minimization method has been in used for the reconstruction of sea bed bathymetric data acquired by a multibeam sonar [30]. Specifically, evaluation of the error of the reconstruction process is performed using a weak membrane or a thin-plate model [31] together with a line process, included to remove the smoothness constraint at surface boundaries. MRF segmentation of images acquired by a multibeam echosounder is also addressed by Dugelay *et al*. [32]. By arranging the acoustical map on the basis of the acquisition parameters, they devise a classical MRF process modelling the prior energy as weighting coefficients depending on its geometrical location and the observation energy term as a χ^2 distribution. The segmentation of simulated high frequency 3-D sonar images was presented by Subramaniam and Bahl [33]. This work is essentially a surface fitting procedure [25], used to recover 3-D surface structures from sparse data. Mignotte *et al.* [34] address the segmentation of sonar images by using an unsupervised hierarchical MRF algorithm. Sonar images of the sea bed are segmented in 2 classes, shadow and seabed reverberation, useful for detection and classification of entities lying on the seabed. Multiresolution MRF models are devised at each scale: the model is hierarchical as local neighbours are devised at each resolution level and between levels. Gaussian and Rayleigh laws are assumed for the two classes and used in the model parameters' estimation phase performed during the segmentation process without using prior knowledge. Calder *et al*. [35] present another MRF-based technique for the segmentation of sidescan sonar images. The approach here is to segment an image on the basis of its textural content. The adopted MRF models are the auto-binomial model and the multinomial model [36], also including a method for the estimation of the model parameters. Carmichael *et al.* [37] performs segmentation and classification of sidescan sonar images by using fractal measures of roughness of sediments. A multiresolution directional operator is proposed able to discriminate between several types of seafloor sediments having different acoustic scattering. Locally estimated fractal measures provide a set of discriminant features that can be easily separated by typical classification algorithms. A further work of Linnet *et al.* [38] presents another texture-based approach to segment sidescan images by using fractal and spatial-point approaches. The former method utilizes a maximum likelihood classificator trained with the fractal dimension of characteristic sample textures, which is directly linked to local grey-level distribution. The latter method analyses the spatial distributions of the grey levels to characterize each image pixel, still utilizing the same training samples' set. Since objects are regarded as a perturbation over the background texture, they are detected just as small unclassified image areas sunk in a larger classified zone, practically, building an associated probability map indicating the reliability of each pixel to belong to a certain texture.

.

Zerr and Stage presents a method to estimate the 3-D structure of an object from several sonar images [39]. The method assumes that a sector-scan sonar acquire a set of images around the object of interest. Then, a classification of the images in 3 types of regions (echo, shadow, background) is performed by using a couple of thresholds fixed on the basis of the sonar performances a-priori known. Therefore, shadow information in an image is used to compute the cross section of the object from that direction, and then the whole 3-D structure is obtained by volumetric reconstruction exploiting occluding contours, by combining all the estimated cross sections. The reflectivity map is estimated by computerized tomography of the echo information by using the 2D Radon transform. Unfortunately, this approach assumes that the sonar images were taken at a constant distance from the object and a constant height from the seabed, so that it can be affected by positional uncertainties typical of underwater scenarios. Many works of Auran *et al*. [40] addressed the reconstruction problem from sonar data. These works utilizes the occupancy grid approach [41] to support several types of AUV tasks. More specifically, sector scan data are processed to build a dynamic 3-D occupancy map where some useful information is stored by using an adequate data structure representation. This map is actually a volumetric representation in polar coordinates where present cells denote the presence of an echo at that location and a set of associated information useful for subsequent processing. After the map formation, clusters are detected by the connected component analysis in order to identify the main objects present in the scene, and useful features (i.e., moments, radial and angular sizes, bounding box, area, volume, etc.) are extracted from the clusters. The subsequent steps aim at modelling and visualizing the detected objects by a surface fitting procedure and, if different views are available, a complete reconstruction can be performed after individual surface estimation by merging surface patches.

From the above summary, one can notice that several works address the problem of segmentation or reconstruction of acoustic images using statistical or geometrical approaches, but only a few of them exploits different information sources directly provided by the sensor.

The problem of reconstruction of acoustic data extracted by an acoustical multibeam system was already addressed by the authors in [42,43,44,45]. In [42], a first formulation of the energy function was devised, simply assuming reliable those 3-D measures associated to high echo returns (high confidence) and discarding the low confidence range points. In [43], this mechanism is evolved in the definition of a complex functional which takes into account the physical significance of the coupling term between 3-D and confidence images. Another paper [45] presents a generalization of the latter work where theoretical justifications in probabilistic terms are provided for the functional form proposed in [43] as well as for the other novel energy formulations.

A novel approach has been recently investigated in the field of acoustic image reconstruction. For example, in survey missions, autonomous or remotely operated vehicles (AUVs, ROVs) are typically equipped by acoustic devices sensing the environment. During vehicle navigation acoustic 3-D images are acquired but, due to the limited field of view and the implicit inaccuracy of its positioning, it is difficult to obtain a complete and accurate investigation of the surrounding environment. For these reasons, in [46] 3-D acoustic image mosaicing has been addressed. Image mosaicing is a recent technique used to align optical (2-D) images

in order to reconstruct a panoramic view of the environment. In this context, the term has been used to name a 3-D reconstruction of the environment from multiple 3-D acoustic images.

This work should considered in the research line devoted to the registration of image pairs or the integration of a set of range images. Typically, the term *registration* is used for the geometric alignment of a couple or more 3-D data point sets, while the term *fusion* is utilized when one would like to get a single surface representation from registered 3-D data sets. Most of the work present in the literature assume range images from a laser range finder looking at a single, even complex, object.

Among these ones, works related to registration, the Iterative Closest Point (ICP) procedure [47,48] and its variants [49] are seminal papers worth to be mentioned. They are all based on the original iterative algorithm based on the search of pairs of nearest points in the two sets, and estimating the rigid transformation which align them. Then, the rigid transformation is applied to the points of one set, and the procedure is iterated until convergence. Variants include the use of closest points in the direction of the local surface normal [48], and the use of a robust statistics technique [49] to limit the maximum distance between points. Originally, these works assumes that one point set is a subset of the other and in not too distant positions. In this line, several approaches have been investigated, and extended to register multiple 3-D images [50,51,52,53,54,55]. The underwater environment implies to deal 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. The work in [46] is the only dealing with acoustic data and aims at reconstructing a 3-D environment from a sequence of clutter, noisy, and low resolution data, in order to produce a 3-D panoramic mosaic of the scene.

7. SEGMENTATION AND RECONSTRUCTION OF UNDERWATER TUBULAR STRUCTURES

In this section, we address the use of the acoustic image processing to design a complete system for segmentation, reconstruction, and final augmented rendering, of underwater tubular structures.

The acoustic camera utilized [56] is formed by a two-dimensional array of transducers sensible to signals backscattered from the scene previously insonified by a high-frequency acoustic pulse. The whole set of raw signals is then processed to estimate signals coming from fixed steering directions (beamsignals) while attenuating those coming from other directions. Assuming that beamsignals represent the responses of a scene from a 2D set of (steering) directions, a 3-D point set can be extracted detecting the time instant t^* at which the maximum peak occurs in each beamsignal. Besides, the intensity of the maximum peak can be used to generate another image, registered with the former, representing the reliability of the associated 3-D measures. In other words, the higher the intensity, the safer the 3-D measure associated. Images are formed by 64×64 3-D points ordered according to an angular relation, as adjacent points correspond to adjacent beam signals. Their coordinates are expressed in a 3-D reference frame attached to the sensor.

.

7.1 3-D data processing

A preliminary low level processing phase is performed on the raw data obtained by the acoustic camera, both to clean the images from noise and to extract the 3-D edges of the objects observed.

Noise Filtering

Since acoustic raw images are typically quite noisy, due to both the environment conditions and the physical limitations of the camera, it is mandatory a pre-processing phase to reduce spurious information from such images. The used acoustic camera directly performs a preliminary low level stage: it provides together with the 3-D measures, the related intensity and other useful information associated to each point as well, like a connection matrix and normals. More specifically, data are arranged in a 64×64 array $\mathbf{C}_{ij}$ of 3-D coordinates, estimated from the time instants at which the main peak of the beamsignals are detected. An intensity value is also associated to each 3-D point, representing the reliability of the peak of the detected responses. Consequently, adjacent points in the image correspond to neighboring beamsignals, and connected components are formed by neighbouring points whose Euclidean distance is below a certain threshold T_r, set on the basis of sensor resolution and a-priori knowledge of the scene.

Using this procedure, we define two points as connected if it is possible to find a chain of neighbouring points connecting them. In such a way, it is possible to subdivide the image in a certain number of connected components, while discarding those components formed by a small number of points, likely not representing interesting physical objects.

This technique is usually called “size” filtering in the two-dimensional image processing literature. Finally, “reliable” connected components are formed by the points whose associated intensity is above a certain threshold, still depending on the camera properties.

Smoothing by line fitting

Smoothing of range data can typically be 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 [46,57], we use one-dimensional (linear) windows swept along several directions on the image plane, and integrate the results obtained 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 neighbourhood 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 neighbourhood per pixel always exists whose data are homogeneous. For every point ***P***, the algorithm fits one line per window on all windows containing ***P***, basing on Haralick's facet model. 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.

Experiments and results

Several experiments have been performed to test pre-processing phase of the acoustic images, and we obtained good results both in the cleaning of the images and the general understandability of them.

In Fig. 4, a typical raw image acquired by the acoustic camera and related the filtered data are shown. Such images have been obtained using with T_r = 20 cm and the threshold on the intensity equal to 5000 (set on the basis of the histogram distribution). We can see that the procedure is quite good and the cleaned images have been used in all the subsequent step of the recognition module.

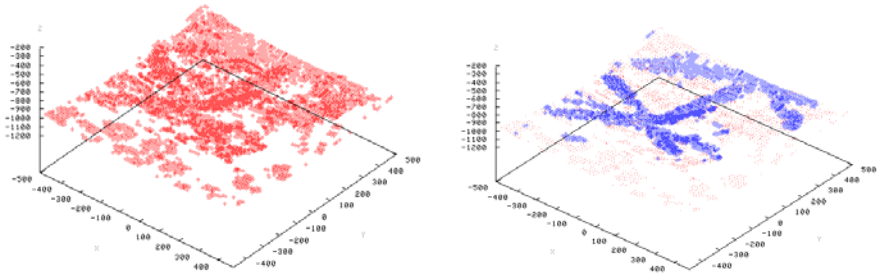


Figure 4. Raw 3-D data (left) and filtered 3-D data (right).

.

7.2 Image segmentation and reconstruction

After the pre-processing of the acoustic images, we obtain an ensemble of 3-D points representing, with a high confidence, physical points derived from objects in the scene. The main problem is now to segment such ensemble in well distinguished regions corresponding to different entities present in the observed scene. Since the underwater scene is composed by a structure of tubular elements, we consider mainly two kind of objects, i.e., pipe-like and not pipe-like ones.

If we are interested in the construction of an augmented representation of the underwater structure the segmentation phase to identify the pipes may suffice. Conversely, if we would like to build a virtual representation, a recognition of the joints formed by pipes intersection is necessary in order to identify them by applying a matching phase using a model reference structure. Moreover, such a matching phase is necessary for the pose and position determination of the ROV from the acoustic data interpretation.

Segmentation

Three possible techniques have been investigated to segment the acoustic images. The first one is based on a region growing algorithm guided by the fitting of the 3-D points with a set of quadrics. The second one is a morphological approach useful in find out rectangular shapes in the range images. The third one is based on the skeleton extraction from the distribution of points and a subsequent analysis of its branches. We briefly describe these methods with emphasis on the skeleton method that is the one used for the final tests.

Geometrical method

The geometrical method, that has been extensively discussed in [58], have been tested both on real and synthetic acoustic images. It consists in the following steps:

- Points Labelling (fitting with expansion)
  1. A point with its 25 neighbours points is randomly selected R_i set
  2. Points of R_i are fitted with a quadric, if fit error is too large choose another point (go to 1)
  3. Add to R_i all of the points (expansions) with error less than a threshold and go to 2, else if there are no points to be added, start with a new region (i=i+1 and go to 1)
  4. Repeat until all points are labelled
- Refinement
  1. Points not labelled are assigned to the closest region
  2. Elimination (fusion) of smaller regions
  3. Fit repeated in all regions
- Merging of regions
  1. All pairs of regions are considered
  2. Surface of fit of the union region is computed
  3. We merge regions comparing the error of such a fit with the mean error of the fit of the separate regions; if it is less the two regions are merged.
- Labelling of isolated points
  1. Detecting isolated points using a compactness measure

2. Reassignment of these points to the nearest region that contains them.

This method is very reliable for synthetic images and, in general, for clean images where the curvature of the objects (pipes) is evident; unfortunately this is not the case for real images captured with the acoustic camera. The geometrical algorithm, in these cases, tends to under-segment the image and to fit pipes with planes. Nevertheless it remains a valid method for sharp images and an interesting one to be studied for further development.

In Fig. 5, two examples of 3-D data as obtained from the filtering stage, and segmented data are displayed. One can notice a certain degree of error in the resulting segmentation.

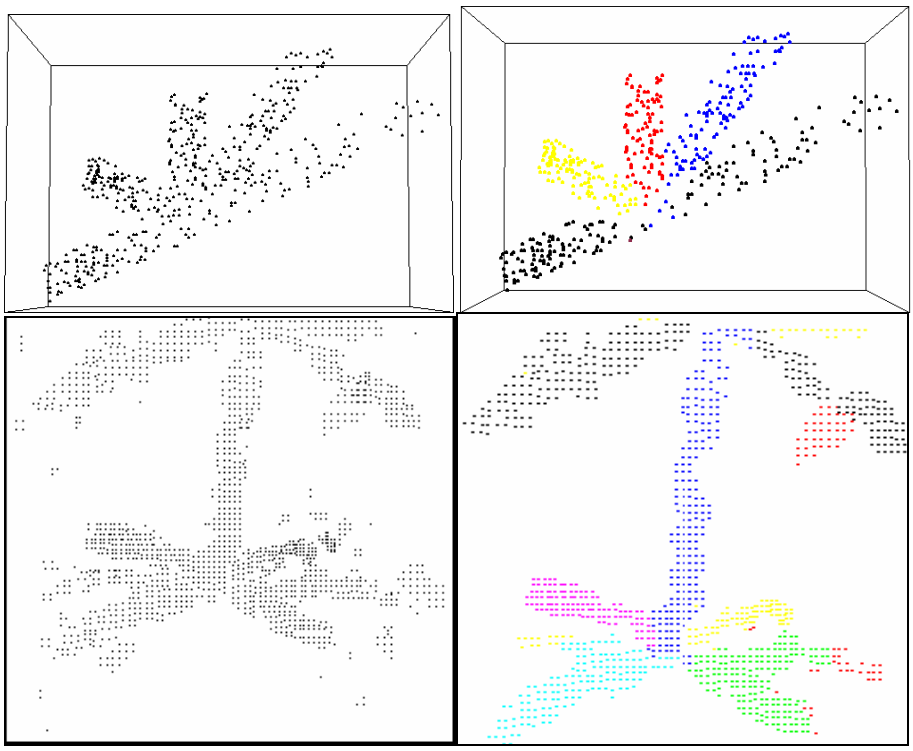


Figure 5. Two examples of filtered 3-D data (left column), and segmentation results using the geometric method (right column).

Morphological method

The second method starts from the idea that the projection of pipes on a plane gives mostly rectangular shapes. The search of pipe-like structures in the observed scene is then reduced to the search for rectangular shapes in the angular range images. The decomposition of an image in elementary shapes is a well known problem of the mathematical morphology [59]. Using standard morphological techniques we have studied the possibility to decompose the range images in elemental rectangular shapes and the subsequent back projection of these shapes in the 3-D space. Although the results are very good in some cases, this method is not

.

robust to noise and tends to over-segment the real images; furthermore, morphological operators, in general, tends to be very CPU intensive procedures and so are not very suitable for a real time application.

In Fig. 6, a couple of examples of 3-D data as obtained from the filtering stage, and segmented data are displayed. A certain degree of over-segmentation is evident, resulting in decomposing single pipes in more regions.

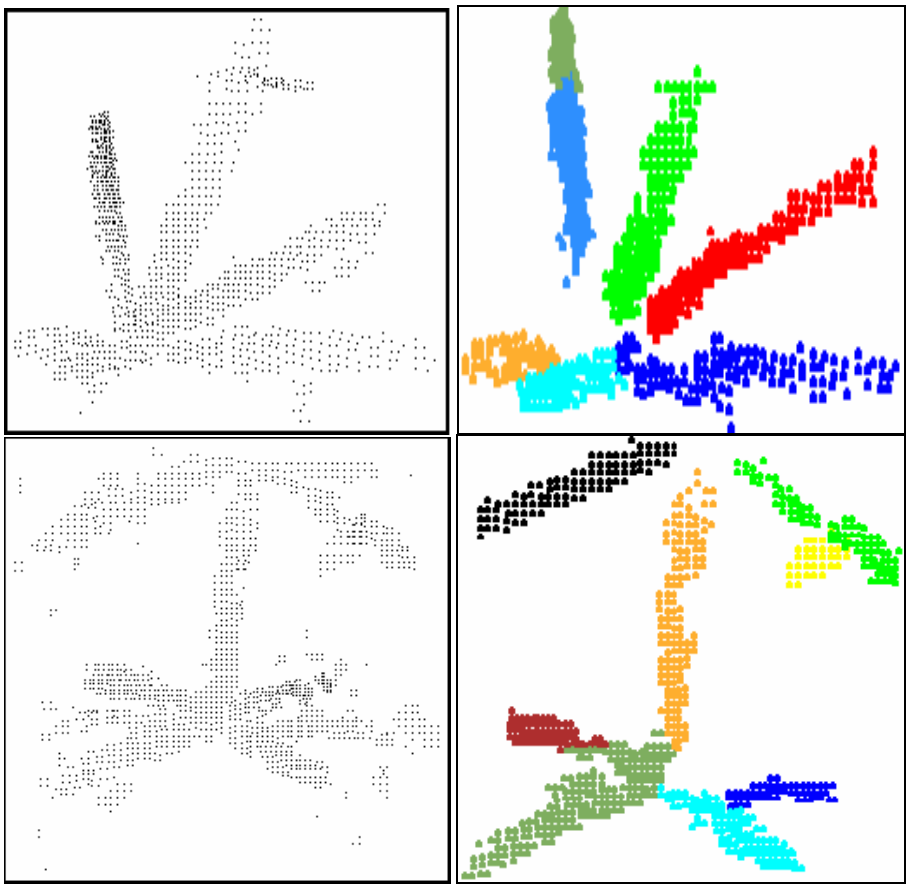


Figure 6. Two examples of filtered 3-D data (left column), and segmentation results using the morphological method (right column).

Skeleton-based method

The third algorithm is very simple and customized to deal with the extraction of tubular structures, even if it is not prevented its application to different kind of objects. A 3-D skeleton is first extracted and then used to subdivide the image in different convex components. Since this method is quite robust to noise and gives a good segmentation with respect to the other two methods, although it is possible to improve it in many ways, it has been chosen as the one utilized for the final tests and experiments [60].

To extract the skeleton, we apply to the image the following procedure: for every point ***A*** in the image we consider all the points that are in a sphere of radius R centred on ***A***. We then shift ***A*** from its actual position to the centre of mass of such distribution of points. We apply this procedure in a parallel way on all the points of the image.

In short, the skeleton extraction is performed along the following lines. Let us define the 3-D image as *I*, i.e., the ensemble of points in $\mathbf{R}^3$.

$$I = \{\mathbf{x}_i\} = \{x_i, y_i, z_i\}, \quad i = 1,\ldots,N \tag{16}$$

We also define for each point *i* and each *R* in $\mathbf{R}^+$ (ray of a sphere about the point *i*) the subset O_i^R of *I* and the 3-D point $\mathbf{b}_i^R$ defined in the following way:

$$O_i^R \equiv \left\{\mathbf{x}_j \in I : \left|\mathbf{x}_i - \mathbf{x}_j\right| < R\right\} \tag{17}$$

$$\mathbf{b}_i^R \equiv \frac{\sum_{\mathbf{x}_j \in O_i^R} \mathbf{x}_j}{\dim\left\{O_i^R\right\}} \tag{18}$$

Let us define the *interior* λ_i^R for the point *i* in the following way:

$$\lambda_i^R \equiv \frac{R - \left|\mathbf{x}_j - \mathbf{b}_i^R\right|}{R} \tag{19}$$

One can easily verify that:

$$0 \le \lambda_i^R \le 1 \quad \forall i \tag{20}$$

The interior is a measure of how much a point is "inside" the object specified by *I*. If $\lambda_i^R \approx 1$, the point is inside a homogeneous zone of radius at least equal to *R*; on the contrary, for $\lambda_i^R \approx 0$, the point is very near to a border of the three-dimensional distribution of points. Defining the following image transformation:

$$I \to I^R = \left\{\mathbf{b}_i^R\right\} \tag{21}$$

and indicating with the symbol $I^{R,n}$ the iterative application of it for *n* times, our skeleton extraction is simply the construction of the image $I^{R,n}$ for a suitable choice of *R* and *n*. The overall effect of this transformation is to shift points on the border, i.e., points with a low λ^R, toward the centre while leaving points well inside an object, i.e., points with a almost unitary value of λ^R, unaltered. The iterative application of such a procedure tends to shift all the points of the distribution towards the skeleton. Such an algorithm may be seen from a physical point of view like a short range interaction between physical points.

The choice of the free parameter *R* is very important. If it is too small, with respect to the average mutual distances of the points, the distribution would likely collapse in a certain number of disconnected punctual regions, preventing the extraction of the skeleton. If it is too large the value of λ may be small also for points well inside the object and again the skeleton will not be properly extracted. The parameter *R* plays a similar role played by the dimension of the structuring element in the classical mathematical morphology. It is straightforward to verify that the skeleton so extracted is invariant for three dimensional rotation (condition that is never exactly satisfied for skeletons of a 2D bitmap image. Preliminary experimental results show also that the homotopy group is preserved by the algorithm. Using a priori knowledge of the structure represented by the 3-D image and of the sensor resolution, it is possible to estimate proper values for *R* and *n* aimed at improving the extraction of a good skeleton. In the results, R = 70 and n = 2 are used.

.

A possible improvement for this algorithm is to find a procedure that is automatically able to set the free parameter R depending on 3-D distribution characteristics (like points density, moments, etc.). Actually, such selection should be local, so that the skeleton will be extracted in different ways in different image regions, so that the method could be applied to any kind of data without prior knowledge of the scene investigated. Moreover, we are also planning the implementation of an estimator to evaluate the "quality" of the skeleton obtained, in order to have an automatic criterion of stop of the algorithm. Several ideas are currently investigated to solve these two problems. We suppose now that every branches of the skeleton correspond to a component object in the scene. Therefore, the segmentation is obtained classifying every point of the skeleton in either *branch point* or *joint point*. Such a classification depends on the neighbourhood around the point: if all the neighbours (e.g., the ones contained in a small sphere) belong to a quite straight segment, it is classified as a branch point, otherwise it is a joint point. If we now consider only the branch points, we obtain well separated connected components of the skeleton; we can segment them with a percolation techniques and associate to each of them an object of the scene. In other words we can search for the connected components of the skeleton minus its joints in the same way we obtained the connected components for the 3-D distribution of points in the pre-processing phase; these regions forms the branches of the skeleton and are the candidates to be classified as pipes.

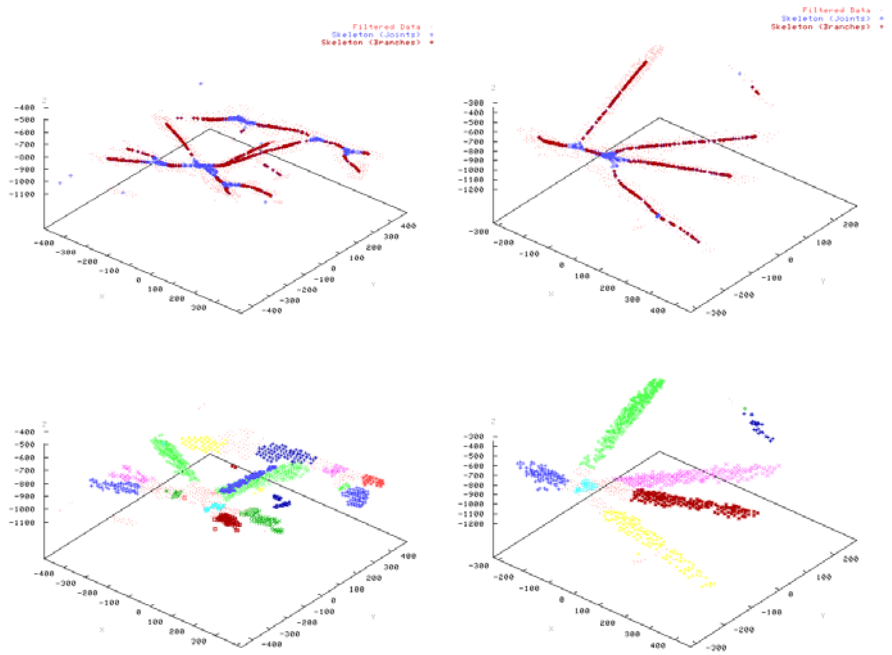


Figure 7. Skeleton extraction and branches' classification (top), and segmentation results using the skeleton-based method (bottom).

In Fig. 7, the results of the skeleton extraction algorithm are presented related to two 3-D images of a tubular structure (top), and the resulting segmented images are shown in the bottom part of the same figure. This method, although not perfect, appears more accurate and robust with respect to the above-mentioned approaches.

Classification

The segmentation procedure does not yet classify the regions in pipe-like and not-pipe-like elements. To perform this last step, we need to analyse the 3-D geometric shape of these regions. To obtain information on the shape of a 3-D distribution of points, it is possible to use a particular tensor associated to such distribution, the Inertial Tensor; such technique is deeply related, as we will show, to the so called Principal Component Analysis.

If ρ is a continuous distribution in R^3, we define the Inertial Tensor I of ρ as a 3x3 matrix given by the following formulas:

$$\begin{aligned}
IT_{xx} &= \int \rho(x,y,z)\left(y^2+z^2\right)dV \\
IT_{yy} &= \int \rho(x,y,z)\left(x^2+z^2\right)dV \\
IT_{zz} &= \int \rho(x,y,z)\left(x^2+y^2\right)dV \\
IT_{xy} &= IT_{yx} = -\int \rho(x,y,z)\,xy\,dV \\
IT_{yz} &= IT_{zy} = -\int \rho(x,y,z)\,yz\,dV \\
IT_{xz} &= IT_{zx} = -\int \rho(x,y,z)\,xz\,dV
\end{aligned} \tag{22}$$

where the coordinates are set so that the centre of mass of the distribution is in the origin. The eigenvalues of this tensor corresponds to the rotation inertia of the distribution around the principal axis defined by the directions of the corresponding eigenvectors of I.

For a discrete distribution we indicate with **o** the centre of mass of the distribution:

$$\mathbf{o} = \frac{\sum_i \mathbf{b}_i}{\mathbf{N}} \quad i \in C_k \tag{23}$$

where $\mathbf{b}_i$ are the position vectors of the distribution (the apex R is omitted for simplicity), and C_k is the k-th connected component. For the inertial tensor we have:

$$IT = \sum_i \left(\mathbf{b}_i - \mathbf{o}\right)\Theta\left(\mathbf{b}_i - \mathbf{o}\right) \tag{24}$$

where we use the symbol Θ for the following operator:

$$a\Theta b = \begin{bmatrix} \left(a_z b_z + a_y b_y\right) & -a_x b_y & -a_x b_z \\ -a_y b_x & \left(a_z b_z + a_x b_x\right) & -a_y b_z \\ -a_z b_x & -a_z b_y & \left(a_y b_y + a_x b_x\right) \end{bmatrix} \tag{25}$$

In the case of a symmetric distribution, the eigenvectors of such a matrix, being the principal inertial axes, are symmetry axes too. For example, for a pipe-like distribution, one of the eigenvector, the one corresponding to the lowest eigenvalue, lies along the pipe axis. In this way, we can use eigenvalues and eigenvectors of I to extract useful information on the shape of the discrete distribution.

.

The Principal Component Matrix [60] is essentially defined as:

$$\mathrm{PCM} = \sum_i (\mathbf{b}_i - \mathbf{o}) \otimes (\mathbf{b}_i - \mathbf{o}) \tag{26}$$

where the symbol $\otimes$ stands for the following operation:

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_x b_x & a_x b_y & a_x b_z \\ a_y b_x & a_y b_y & a_y b_z \\ a_z b_x & a_z b_y & a_z b_z \end{bmatrix} \tag{27}$$

One can see that from the preceding definitions we have:

$$\mathrm{PCM} + \mathrm{IT} = \left[\sum_i (\mathbf{b}_i - \mathbf{o}) \cdot (\mathbf{b}_i - \mathbf{o})\right] \cdot \mathbf{1} = \alpha \cdot \mathbf{1} \tag{28}$$

where **1** is the identity matrix. From this relation it follows that every eigenvector **v** of PCM with eigenvalue λ is also an eigenvector of I with eigenvalue (λ-α). In fact:

$$\mathrm{PCM} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}. \tag{29}$$

Hence:

$$(\mathrm{PCM} + \mathrm{IT}) \cdot \mathbf{v} = \lambda \cdot \mathbf{v} + \mathrm{IT} \cdot \mathbf{v} = \alpha \cdot \mathbf{v} \tag{30}$$

from which:

$$\mathrm{IT} \cdot \mathrm{v} = (\alpha - \lambda) \cdot \mathbf{v}. \tag{31}$$

Consequently, it is clear that one can use both methods to obtain the same results. We have chosen to use the Inertial Tensor: we calculate I for every segmented region in the scene and obtain three eigenvalues. If one is smaller with respect to the others, and these are of the same magnitude, we classify the region as a tubular one. Obviously, we have also to introduce a threshold: if it is too small, it is probable to classify as a pipe something that is only elongated, whereas, if it is too high, it is probable to lose some pipes from the scene.

$$\begin{aligned} & e_j \quad j = 1,..,3 \\ & e_m << e_k \quad m \neq k \\ & e_n \cong e_h \quad n,h \neq m \end{aligned} \tag{32}$$

From the value of the minimum eigenvalue it is possible to extract also the order of magnitude of the radius of the tubular region. In fact, in the case of a complete cylindrical distribution the following relation holds:

$$e_m = 0.5 \cdot \mathrm{M} \cdot \mathrm{r}^2 \tag{33}$$

where e_m is the minimum eigenvalue, M is the total number of points in the distribution and r is the radius. As in the range images of underwater pipes, data are not distributed on the surface of a cylinder, but only on a little portion of it (the portion "visible" from the camera), this relation is only approximately true, but it is sufficient to give the right order of magnitude for the radius, as we will see in the following. For a more precise extrapolation of this parameter, it is necessary a more sophisticated algorithm that is based on a robust fitting procedure. Another problem that is actually addressed in our research work is a fast and reliable refinement of the segmentation. Because of noise and the particular skeleton reconstruction procedure, it is quite probable to have an oversegmentation of the image, i.e. a single object may be segmented in more than one region. This problem is partially solved considering every pair of regions and calculating the inertia tensor of their union: if it is "more" pipe-shaped (by using the previous criterion), the two regions are merged.

We have applied our classification algorithm to the segmented images previously obtained and we have extracted the pipes as shown in Fig. 8. From these examples, it is clear the necessity, at least for Joint 1 (Fig. 7, bottom-left), of the merging phase sketched above; in fact some of the pipes in this scene have not been recognized because they were oversegmented in the classification step. This problem will be solved after the merging phase will be completed. It should be noted, however, that the Joint 1 is quite an extreme case that we present here only for completeness; in most of the underwater images we have analysed the results are good and similar to the ones obtained for Joint 2 (Fig. 7, bottom-right).

Figure 8. Augmented reality examples.

It is possible now to proceed in two different, but not excluding, ways: an augmented reality or a complete virtual reality representation [61]. The first one is simply obtained by projecting the virtual pipes, corresponding to the 3-D segmented regions, on the image plane. The resulting image is much more readable by a human operator, and hence, useful for navigation tasks or simply to realize what the camera is looking at. This procedure does not require any prior knowledge of the structure investigated (apart the generic knowledge on the objects' shape). As an example, we have rendered a scene with the data (raw and filtered) of the underwater acoustic images together with the pipes extracted with our algorithm (see Fig. 8); it is interesting to note that the radius of the pipes are quite good although obtained with the approximated formula.

The complete virtual environment requires a preliminary recognition phase. The idea is to build, from a database, a virtual scene of the structure investigated (e.g., an off-shore platform) and to use the segmented images to recognize the position and the pose of the acoustic camera in the scene by a suitable matching with the database information. Therefore, a human operator could not observe the real image, but the virtual one, possibly navigating in the virtual scenario reconstructed from real data.

8. CONCLUSIONS

In this work, we have presented a general overview of the approaches and methods aimed at the generation and processing of acoustic images. In particular, in the first part, methods for the generation of acoustic images are presented,

.

describing holographic, and lens-based methods, and, in detail, the 3-D beamforming technique. In the second part, the emphasis was on 3-D image processing techniques. After a general survey about the methods devoted to range image processing, a specific overview of the methods for acoustic image analysis and segmentation is reported, and finally, a technique for 3-D acoustic image segmentation and interpretation is presented, aimed at obtaining a synthetic representation of the observed scene, useful for augmented reality and virtual reality applications.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. R.K. Hansen of Omnitech A/S (Norway) for kindly providing the images acquired by the Echoscope acoustic camera, and Dr. Riccardo Giannitrapani, who partly contributes to this project. This work is supported by the European Commission under the BRITE-EURAM III project no. BE-2013 VENICE (Virtual Environment Interface by Sensory Integration for Inspection and Manipulation Control in Multifunctional Underwater Vehicles).

REFERENCES

[1] A. Macovski, "Ultrasonic Imaging Using Arrays," *IEEE Proceedings*, vol. 67, pp. 484-495, April 1979.

[2] J.L. Sutton, "Underwater Acoustic Imaging," *IEEE Proceedings*, vol. 67, pp. 554-566, April 1979.

[3] V. Murino and A. Trucco, "Three-Dimensional Image Generation and Processing in Underwater Acoustic Vision," *IEEE Proceedings*, vol. 88, pp. 1903-1946, December 2000.

[4] O. George and R. Bahl, "Simulation of Backscattering of High Frequency Sound From Complex Objects and Sand Sea-Bottom," *IEEE Jour. Ocean. Engin*., vol. 20, pp. 119-130, April 1995.

[5] T.L. Henderson and S.G. Lacker, "Seafloor Profiling by a Wideband Sonar: Simulation, Frequency-Response, Optimization, and Results of a Brief Sea Test," *IEEE Jour. Ocean. Engin.*, vol. 14, pp. 94-107, January 1989.

[6] B.D. Steinberg and H.M. Subbaram, *Microwave Imaging Technique*, J. Wiley & Sons, New York, 1991.

[7] J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.

[8] L.J. Ziomek, "Three Necessary Conditions for the Validity of the Fresnel Phase Approximation for the Near-Field Beam Pattern of an Aperture," *IEEE Journal of Oceanic Engineering*, vol. 18, pp. 73-75, January 1993.

[9] R.O. Nielsen, *Sonar Signal Processing*, Artech House, Boston, 1991.

[10] J.S. Jaffe, P.M. Cassereau, and D.J. Glassbrenner, "Code Design and Performance Characterization for Code Multiplexed Imaging," *IEEE Trans. Acoustics, Speech, Sig. Proc.*, vol. 38, pp. 1321-1329, August 1990.

[11] J.C. Bu, C.J.M. van Ruiten, and L.F. van der Wal, "Underwater Acoustical Imaging Algorithms," *Proc. Europ. Conf. on Underwater Acoustics*, Luxembourg, pp. 717-720, September 1992.

[12] R.K. Hansen and P.A. Andersen, "3D Acoustic Camera for Underwater Imaging," in *Acoustical Imaging*, Yu Wei and Benli Gu, Eds., vol. 20, pp. 723-727, New York, 1993, Plenum Press.
[13] R.K. Hansen and P.A. Andersen, "A 3D Underwater Acoustic Camera Properties and Applications," in *Acoustical Imaging*, P. Tortoli and L. Masotti, Eds., vol. 22, 22nd Int. Symposium, pp. 607-611, New York, 1996, Plenum Press.
[14] R.K. Hansen and P.A. Andersen, "The Application of Real Time 3D Acoustical Imaging," *IEEE/OES Int. Conf. Oceans '98*, Nice, France, pp. 738-741, September 1998.
[15] R. Hoffman and A.K. Jain, "Segmentation and Classification of Range Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 9, n. 5, pp. 608-619, September 1987.
[16] Y. Shirai, "*Three-Dimensional Computer Vision*," Springer-Verlag, Germany, 1987.
[17] M.A. Wami and B.G. Batchelor, "Edge-Region-Based segmentation of Range Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 16, n. 3, pp. 314-319, 1994.
[18] O.R.P. Bellon and C.L. Tozzi, "An Integrated Approach for Segmentation and Representation of Range Image," *ICAP'97*, pp. 279-286, 1997.D. Zhao, X. Zhang, "Range-Data-Based Object Surface Segmentation via Edges and Critical Points," *IEEE Trans. on Image Processing*, vol. 6, n. 6, pp. 826-820, June 1997.R.W. Taylor, M. Savini, and A.P. Reeves, "Fast Segmentation of Range Image into Planar Region," *Computer vision, graphics, and image processing*, vol. 45 pp. 42-60, 1989.
[21] P.J. Besl and R.C. Jain "Invariant Surface Characteristic for 3D Object Recognition in Range Images," *Computer vision, graphics, and image processing*, vol. 33 pp. 33-80, 1986.
[22] S.Z. Li, "Toward 3D Vision Range Images: An Optimization Framework and Parallel Networks," *CVGIP: Image Understanding*, vol. 55, n. 3, pp. 231-260, 1992.
[23] E. Trucco and R.B. Fisher "Experiments in Curvature - Based Segmentation of Range Data," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 17, n. 2, pp. 177-182, 1995.
[24] N. Yokoya and M.D. Levine "Range Image Segmentation of Differential Geometry: A Hybrid Approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 11, n. 6, pp. 643-649, June 1989.
[25] P.J. Besl and R.C. Jain, "Segmentation Through Variable-Order Surface Fitting," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 10, n. 2, pp. 167-192, March 1988.
[26] S.M. Bhandarkar and A. Siebert, "Integrating Edge and Surface Information for Range Image Segmentation," *Pattern Recognition*, vol. 25, n. 9, pp. 947-962, 1992.
[27] D.P. Mital, E.K. Teoh, and A.W.T. Lim, "A Hybrid Method towards the Segmentation of Range Images for 3-D Object Recognition," *Int. Journal on Pattern Recognition and Artificial Intelligence*, vol. 8, pp. 869-995, 1994.
[28] R.S. Beattie and S.C. Elder, "Side Scan Sonar Image Restoration using Simulated Annealing and Iterative Conditional Modes," *Proc. Int. Conf. on Sonar Signal Processing*, Loughborough (UK), pp. 161-167, December 1995.
[29] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, 1987.
[30] S. Show and J. Arnold, "Automated Error Detection in Multibeam Bathymetry Data," *Proc. IEEE OCEANS '93*, Victoria (Canada), vol. II, pp. 89-94, 1993.
[31] R. Szelinsky, *Bayesian Modeling of Uncertainty in Low-Level Vision*, Kluver Academic Publ., Boston, 1989.
[32] S. Dugelay, J.M. Augustin, and C. Graffigne, "Segmentation of Multibeam Acoustic Imagery in the Exploration of the Deep Sea Bottom," *13th Int. Conf. on Pattern Recognition*, Vienna, pp. 437-445, August 1996.
[33] L.V. Subramaniam and R. Bahl, "Segmentation and Surface Fitting of Sonar Images for 3D Visualization," *Proc. 8th Int. Symp. on Unmanned Untethered Submersible*

Technology, Durham (NH, USA), pp. 290-298, September 1995.

[34] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Unsupervised segmentation applied on sonar images," in *Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science 1223*, pp. 491-506, 1997.

[35] B.R. Calder, L.M. Linnett, and S.J. Clarke, "Spatial Interaction Models for Sonar Image Data," *Proc. Int. Conf. on Sonar Signal Processing*, Loughborough (UK), pp. 81-90, December 1995.

[36] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society*, vol. B-48, pp. 259-302, 1986.

[37] D.R. Carmichael, L.M. Linnett, and S.J. Clarke, "A Multiresolution Directional Operator for Sidescan Sonar Image Analysis," *Proc. Int. Conf. on Sonar Signal Processing*, Loughborough (UK), pp. 49-57, December 1995.

[38] L.M. Linnett, S.J. Clarke, and D.R. Carmichael, "The Analysis of Sidescan Sonar Images for Seabed Types and Objects," *Proc. 2nd Conf. on Underwater Acoustics*, pp. 733-738, Copenhagen, July 1994.

[39] B. Zerr and B. Stage, "Three-dimensional reconstruction of underwater objects from a sequence of sonar images," *IEEE Int. Conf. on Image Processing*, vol. 3, pp. 927-930, September 1996.

[40] P.G. Auran and K.E. Malvig, "Realtime extraction of connected components in 3d sonar range images," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 580-585, San Francisco, CA, USA, June 1996.

[41] A. Elfes, "Sonar-Based Real-World Mapping and Navigation," *IEEE Journal on Robotics and Automation*, vol. 3, n. 3, pp. 249-265, 1987.

[42] V. Murino, "Acoustic Image Reconstruction by Markov Random Fields," *Electronics Letters*, vol. 32, n. 7, pp. 697-698, 1996.

[43] V. Murino, A. Trucco, and C.S. Regazzoni, "A Probabilistic Approach to the Coupled Reconstruction and Restoration of Underwater Acoustic Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, n. 1, pp. 9-22, 1998.

[44] V. Murino and A. Trucco, "Markov-based Methodology for the Restoration of Underwater Acoustic Images," *International Journal of Imaging Systems and Technology*, vol. 8, n. 4, pp. 386-395, 1997.

[45] V. Murino, "Reconstruction and Segmentation of Underwater Acoustic Images combining Confidence Information in MRF Models," *Pattern Recognition*, vol. 34, n. 5, pp. 981-997, May 2001.

[46] V. Murino, A. Fusiello, N. Iuretigh, and E. Puppo, "3D Mosaicing for Environment Reconstruction," *15th Int. Conf. on Pattern Recognition*, Barcelona, Spain, pp. 362-366, September 2000.

[47] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, n. 2, pp. 239-256, February 1992.

[48] T. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing*, vol. 10, n. 3, pp. 145-155, 1992.

[49] Z. Zhang, "Iterative point matching of free-form curves and surfaces," *Int. Journal of Computer Vision*, vol. 13, n. 2, pp. 119-152, 1994.

[50] G. Blais and M.D. Levine, "Registering Multiview Range Data to Create 3D Computer Objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, n. 8, pp. 540-547, 1995.

[51] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Towards a general multiview registration technique," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, n. 5, pp. 540-547, May 1996.

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

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

[54] A.E. Johnson, M. Hebert, "Surface Registration by Matching Oriented Points," Int. *Conf. on 3D Digital Imaging and Modeling*, Ottawa, Canada, pp. 121-128, 1997.

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

[56] R.K. Hansen and P.A. Andersen, "A 3D Underwater Acoustic Camera - Properties and Applications," in *Acoustical Imaging*, P. Tortoli, L. Masotti Eds., Plenum Press, pp. 607-611, New York, 1996.

[57] 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.

[58] V. Murino, A. Grion, and S. Bianchini, "A Geometric Approach to the Segmentation and Reconstruction of Acoustic Three-Dimensional Data," *OCEANS '98 MTS/IEEE*, pp. 582-586, Nice, France, September 1998.

[59] J. Serra, *Image analisys and mathematical morphology*, Academic Press, London, 1982.

[60] V. Murino and R. Giannitrapani, "Three-Dimensional Skeleton Extraction by Point Set Contraction," *IEEE Int. Conf. on Image Processing ICIP '99*, pp. 565-569, Kobe, Japan, October 1999.

[61] R. Giannitrapani, V. Murino, and A. Trucco, "Segmentation of Underwater 3D Acoustical Images for Augmented and Virtual Reality Applications," *Int. Conf. Oceans '99 MTS/IEEE*, pp. 459-465, Seattle, WA (USA), September 1999.