A Voting-Based Approach for Fast Object Recognition in Underwater Acoustic Images

Gian Luca Foresti, *Member, IEEE,* Vittorio Murino, *Member, IEEE,*
Carlo S. Regazzoni, *Member, IEEE,* and Andrea Trucco

***Abstract*— This paper describes a voting-based approach for the fast automatic recognition of man-made objects and related attitude estimation in underwater acoustic images generated by forward-looking sonars or acoustic cameras. In general, the continuous analysis of sequences of images is a very heavy task for human operators and this is due to the poor quality of acoustic images. Hence, algorithms able to recognize an object on the basis of *a priori* knowledge of the model and to estimate its attitude with reference to a global coordinate system are very useful to facilitate underwater operations like object manipulation or vehicle navigation. The proposed method is capable of recognizing objects and estimating their two-dimensional attitude by using information coming from boundary segments and their angular relations. It is based on a simple voting approach directly applied to the edge discontinuities of underwater acoustic images, whose quality is usually affected by some undesired effects such as object blurring, speckle noise, and geometrical distortions degrading the edge detection. The voting approach is robust, with respect to these effects, so that good results are obtained even with images of very poor quality. The sequences of simulated and real acoustic images are presented in order to test the validity of the proposed method in terms of average estimation error and computational load.**

***Index Terms*— Acoustic imaging, image line pattern recognition, object attitude estimation, object recognition, underwater vehicles, voting methods.**

I. Introduction

IN MODERN underwater missions, the use of automatic or autonomous devices is constantly increasing in order to avoid the direct involvement of human operators in dangerous tasks and to support the operator in making decisions about manipulation and navigation tasks. For instance, the analysis of the acoustic images acquired with an underwater vehicle in the mission site is a very heavy task for humans as the quality of images is poor and the operator relaxes his attention owing to the long duration of the mission. As a consequence, algorithms that are able to recognize an object on the basis of *a priori* knowledge of the model and to estimate at the same time its attitude are surely very useful. Two related examples can be the underwater object manipulation by a robotic vehicle provided with a suitable acoustic sensor (e.g., an acoustic camera) and the AUV (autonomous underwater vehicle) navigation by using a sequence of landmarks fixed to the sea bottom and observed by an imaging sonar.

Modern acoustic cameras, only recently commercially available, have a planar array of sensors and provide for real-time three-dimensional (3-D) maps [1]–[3] of scenes that are some meters away from the array. These 3-D maps are frequently displayed in a projective two-dimensional (2-D) version, like orthoscopic images or section images. Another kind of sonar system that has been improved very much in the last few years is the multibeam forward-looking sonar [4]. This sonar is generally steered along the motion direction and provides for real-time 2-D images of the region of the sea bottom placed in front of the vehicle. Such acoustic devices find their applications on-board of both remotely operated vehicles (ROV's) and AUV's for a wide number of different tasks, and this is due to the medium range achievable (i.e., few tens of meters). Far from an exhaustive list of these tasks, we can mention the inspection, the small-scale survey, the manipulation and the positioning in offshore structures or mineral mining sites, along pipelines or communication cables, the relocation of lost objects, the attitude estimation and the removal of toxic wastes, the mine counter measures, and so on.

In this paper, we address the problem of the fast recognition of man-made objects with an estimation of the related orientation in 2-D acoustic images acquired with a forward-looking sonar or an acoustic camera. Man-made objects are generally characterized by a regular and well-defined structure, where *regular* indicates that the object can be approximated using geometric surfaces bounded by straight lines. If an optical camera is used, the projection on the image plane of such an object produces a geometric regular shape. Actually, well-defined geometrical relations exist between a configuration of 3-D lines and their projections onto an image plane. Many works have been done in this sense trying to exploit the use of such relations (see [5]). Unfortunately, the same problem considering acoustical images, where the resolution is lower and the speckle noise is quite degrading, has not been addressed often. It has been argued that extraction of straight edges is not useful as a basis for performing 3-D geometrical inferences on objects detected in classical sidescan imagery. However, in this paper, we show that this conjecture does not necessarily hold for classes of acoustic imaging systems that are different from sidescan which essentially operates at higher frequency and shorter range. In other words, the proposed method is not well suited to operate on images containing objects with completely deformed shape

Manuscript revised November 12, 1996.

G. L. Foresti and V. Murino are with the Department of Mathematics and Computer Science (DIMI), University of Udine, 33100 Udine, Italy.

C. S. Regazzoni and A. Trucco are with the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, 16145 Genova, Italy.

Publisher Item Identifier S 0364-9059(97)01265-X.

as sometimes occurs in classical sidescan imagery, due to the observation of large sea-bottom areas, but it can be suitable for short/medium range and the orthoscopic display of acquired data.

In general, not much work has been devised for pattern recognition upon acoustic imaging [6], and very little effort was mainly directed to the analysis of the acoustic images built by means of holographic methods. Regarding the sidescan imagery, in [7] a structured method was proposed to detect toxic waste deposits. It was based on image segmentation by statistic features leading to the discrimination between the object areas and the shadow areas, however, despite the good performances, the implicit complexity of the method results in a difficult fast implementation. In [8], a system for the automatic interpretation of sector scan sonar images was proposed. The classification of the patches that were present in the sonar image was essentially model-based, i.e., some qualitative descriptions of the objects were stored and the extracted features were matched with such descriptions, with particular reference to the invariant characteristics and the robustness to noise. In [9], sidescan images were analyzed considering the texture features: fractal-based and spatial point processes were applied to segmentation and object detection purposes. Concerning forward-looking sonar, in [10], a real-time obstacle avoidance method was directly presented working on the signals in output from the sonar processor. An acoustic camera was designed and presented in [11] together with some simple techniques for the object detection, but they only detect objects without recognition and orientation estimation.

Acoustic image segmentation beyond simple thresholding was also addressed by more intelligent techniques that are able to identify more precisely the parts of the object. In [12], a fuzzy clustering technique followed by an adaptive filter was used to extract image regions. In [13], a multiscale Markov random-fields method was used to segment a sonar image into regions. Both previous techniques apply *ad hoc* strategies to reduce the speckle noise. In [14], a voting-based method is used to enforce backscattered acoustic information, improving the image quality so that an easier edge extraction can be performed with classical techniques (e.g., classical edge detection [15]).

Typically, voting-based approaches have been employed in optical images. The related works referred to the Hough transform [16] and to some kinds of matching/recognition processes [17]. In these works, in general, the matching was considered a coherent set of associations between data features performed in a certain space whose degree of reliability was assessed on the basis of the number of "votes" that each association had received according with the available evidence. In [18], a voting method was proposed for perceptual organization. To group a set of features to be considered as a single object, a criterion of compatibility was defined as capable of associating pairs of image tokens (e.g., points, edges, lines, etc.), if they had voted in the same location in the parameter space.

In underwater environments, a few works address the problem of the recognition and determination of the position of objects considering optical images. A shape-from-shading method [19] was used to determine the size and the position of cylindrical objects, while in [20], a Hough-based method was employed in the vehicle position-keeping operation once the structure has been detected and its model identified.

All these works are in some way related to that proposed in this paper concerning either the application, the sensors utilized, or the methodology, but, unfortunately, no other method addresses the recognition and attitude estimation of objects by using such kind of acoustic images. Therefore, the proposed approach cannot be directly compared with the related methods at a quantitative level, but only a general comparison can be attempted. However, a quantitative analysis can be performed with respect to other classical "aerial" methods for image analysis (e.g., template matching and moments of inertia), showing the goodness of the proposed method in terms of accuracy and robustness.

In this paper, we present a technique that is able to recognize an object and to estimate at the same time its 2-D attitude. This method is based on a simple voting approach [21] that is directly applied to the edge discontinuities of the underwater acoustic image. Two different processing levels of increasing complexity are used to detect significative image features (e.g., straight lines) with differnt accuracy. A database of object models is generated off-line taking into account several 2-D attitudes depending on different object projections. Therefore, the first level aims at detecting the presence of an object and its rough attitude in a very fast way. The histogram obtained by summing up the votes received from the edge points with similar gradient orientation is inspected to find significative peaks, i.e., orientations common to several edge points in the image. To improve the robustness of the approach and to fasten the procedure, the model database provides information about angular relations between the adjacent object model sides that must be considered during the histogram analysis to avoid false peaks generated by noise, background points, or other boundary points belonging to different objects. The second level is computationally more complex as it allows one to obtain both a more accurate attitude estimation and a recognition of the object. The angles between the boundary segments of the 2-D object model are used to focus the attention on significative parts of the histogram containing votes coming from the segments of the object; then, the boundary segments of the 2-D object model are matched with 2-D straight segments extracted from the original image to provide a complete recognition [17].

The proposed method is intrinsically robust to noise and invariant to translation and scale allowing the discrimination between different sets of man-made objects and the detection of their orientations. The robustness is due to the voting approach leading to an efficient and fast process of noisy edge information. In fact, the low sensitivity to noise is not due to special methods for edge detection, it is actually performed by an adaptive thresholding (as typical in acoustic image processing) followed by morphology-based techniques [22], but, it is due to the voting technique that intrinsically allows one to disregard noisy information. In the authors' opinion, this is the main novelty and contribution of the proposed approach.

The paper is organized as follows. In Section II, the voting approach is detailed, stressing the inserted geometrical constraints. Section III shows some results obtained from both real and simulated images and proposes several comparisons with the results obtained by applying other recognition methods. Finally, in Section IV, conclusions are drawn.

II. The Voting Approach

A. General Considerations

Up to now, voting methods have generally been applied to visual images for feature detection and object recognition showing a high robustness to noise and to object occlusions [16], [21], [23], [24].[1]

The voting method described here is composed of two different processing levels (see Fig. 1) and has been developed specifically for underwater acoustic images which are often affected by undesired effects such as object blurring (i.e., the presence of halos around objects due to a poor sonar directivity), speckle generating spurious discontinuities (due to coherent interference among waves backscattered from the scene), and distortion (due, above all, to inhomogeneities of the medium). The first level is very fast and aims at detecting the presence of specifc landmarks or man-made objects in a scene. A confidence factor (CF) is associated with each detected object. It is also able to estimate a rough attitude of some classes of objects whose 2-D projection on the image plane produces a regular shape characterized by specific angular relations between its sides: triangles, rectangles, squares, etc. In particular, the orientation α_i of the object boundaries is directly searched for in the feature space [see Fig. 2(a)]. Then, the angle realtions β_{ij} among these straight lines i and j are computed and matched with those of some classes of 2-D object models.[2] The second level consumes more time and is focused on the object recognition and on a more accurate attitude estimation of the object. Information about very simple 2-D object models (based on the position of straight object boundaries and the angle relation between them) have been utilized.

B. Fast Object Detection

The first-level voting process is directly performed from the edge image E [obtained by applying a morphological operator [22] to the original acoustic image $I(x, y)$] into a 1-D parameter space $\Pi = \{\alpha\}$, composed by all possible orientations α of the straight boundaries of the object. α is the angle between the normal to the object boundary and the x axis of the image reference system [see Fig. 2(a)]. Let Θ be the upper limit of the parameter space (e.g., Θ =180°) and let $\Delta\theta$ be the quantization step, thus the parameter space dimensions are given by $\Theta \times \Delta\theta$ [24], [25].[3] The first-level voting equation G_1 is 1-D and favors the grouping of the edge points that have similar orientation values of the gradient $\gamma(\boldsymbol{x})$. Two edge points, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, belonging to straight lines having similar orientations vote for the same parameter space point α in the space Π [26]:

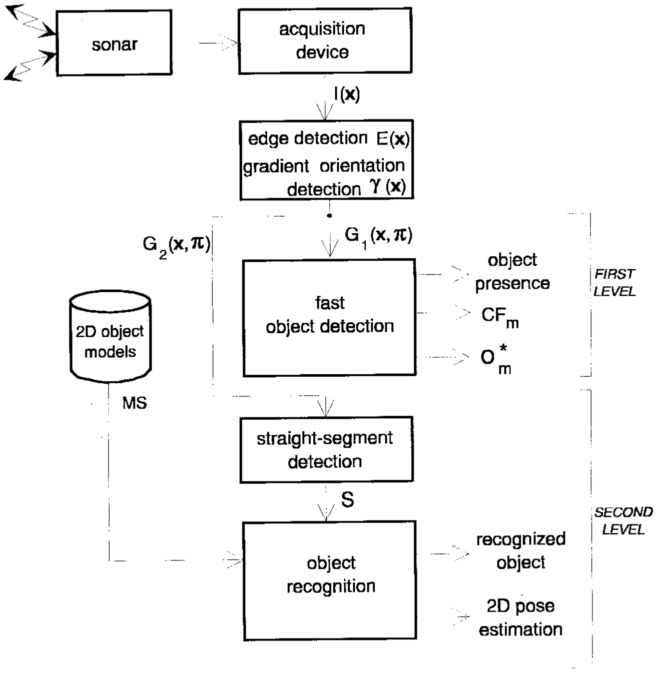


Fig. 1. General flow chart of the two-level voting method.

$$G_1(\boldsymbol{x}, \alpha) = \begin{cases} 1, & \forall\, \alpha \in \Delta\gamma(\boldsymbol{x}) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\Delta\gamma(\boldsymbol{x}) = [\gamma(\boldsymbol{x}) - \gamma_{th}, \gamma(\boldsymbol{x}) + \gamma_{th}]$ and γ_{th} is a local threshold [21]. For each edge pixel $\boldsymbol{x}$ in the image domain which satisfies the relation $G_1(\boldsymbol{x}, \alpha)$, a 1-D accumulator array $C(\alpha)$ (i.e., a histogram) is increased by one count. At the end of the voting phase, $C(\alpha)$ reports, for each α value, the global number of votes received by each cell.[4]

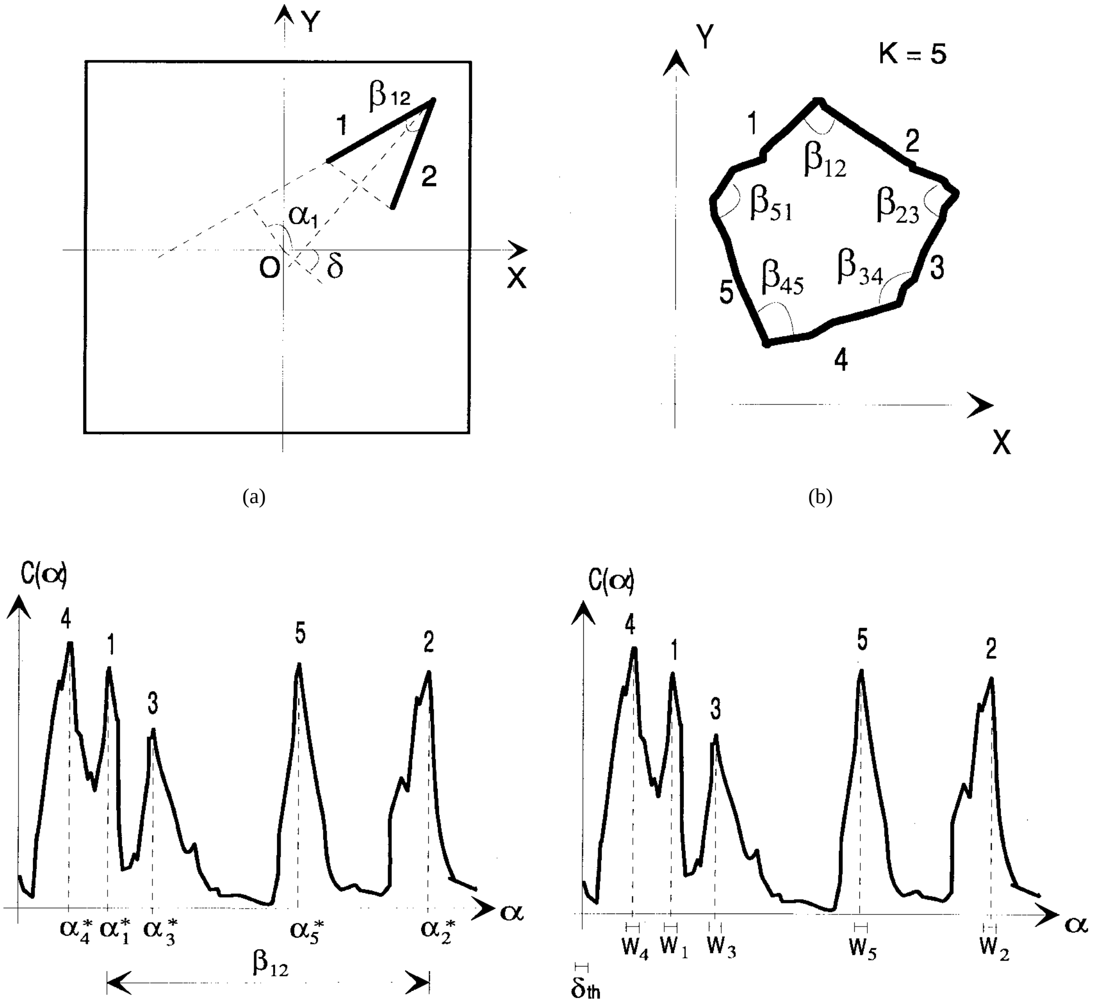
Fig. 2(c) reports the parameter space $C(\alpha)$ obtained by applying the voting method to an image containing a five-sided object shown in Fig. 2(b). It is important to note that other significative peaks appear in the parameter space, especially near the five main peaks. These peaks (i.e., spurious peaks) are generated by vote spreading effects due to both image quantization and noisy gradient effects [24], [25] and can create false object boundary detection when multiple objects are present in the same scene. In order to increase the robustness of the proposed approach and also to extend the search for images containing multiple objects, the information about the angle relations between pairs of 2-D closed boundaries

[1] These methods map each feature space point $\boldsymbol{e} \in E$ into a set of parameter space points $\boldsymbol{\pi} \in \Pi$ by means of a generating equation $G(\boldsymbol{e}, \boldsymbol{\pi}) = 0$, which depends on each class of features (e.g., straight-lines, circles, ellipses, etc.) to be detected.

[2] For example, the method can quickly detect the presence of an arrow landmark characterized by two long convergent straight segments with a given convergence angle β_{12} [see Fig. 2(a)] in an image acquired with a forward lookng sonar. Moreover, it can also estimate the orientation δ of the main axis of the arrow.

[3] To obtain an orientation accuracy equal to 1°, it is necessary to choose $\Delta\theta$ =1.

[4] Orientations α^*, common to several edge points into the image domain (e.g., object borders), generate multiple isolated peaks in the parameter space. The greater the number of points of a given edge, the taller the related peak.



(a) (b)

(c) (d)

Fig. 2. (a) Representation of an arrow landmark composed of two long straight segments with a given convergence angle β_{12}. (b) Input image containing a five-side object. (c) The related parameter space $C(\alpha)$ showing five significative peaks. (d) The five windows W_i centered about each peak.

(obtained from 2-D model objects) must be considered during the histogram analysis.[5] A set of angle relations corresponding to pairs of model object boundaries can be easily computed by

$$\beta_{ij} = |\alpha_i - \alpha_j| \quad \forall i, j \in [1, K], \qquad i \neq j \tag{2}$$

where α_i represents the orientation of the ith model object boundary and K is the global number of object model boundaries[6] [see Fig. 2(b)]. During the process, a set of mobile windows $W = \{W_k, k = 1, \cdots, K\}$ of width δ_{th} and interdistance $\beta_{ij} = |\omega_i - \omega_j|$, where ω_i is the medium point of W_i, have been shifted on the histogram to detect candidate peaks, whose position satisfies the angle relation characterizing the 2-D object model boundaries [see Fig. 2(d)]. The candidate peaks α_k^*, satisfying the angle relations that are related to the 2-D boundaries of the mth object model are grouped into the same set $O_m = \{\alpha_k^*, k = 1, \cdots, K\}$, $m = 1, \cdots, M$, where M is the number of the candidate detected objects. In this way, the probability of detecting false maxima is reduced by making the detection step more robust to false peaks and noisy data. Objects are detected on the basis of the groups O_m^* whose accumulators $C(\alpha_k^*)$ receive a number of votes higher than a predefined threshold C_{th} [21]:

$$O_m^* = \{\alpha_k^*\colon \; C(\alpha_k^*) > C_{\text{th}}, \; k = 1, \cdots, K\}. \tag{3}$$

A rough 2-D object attitude is also estimated on the basis of the detected α_k^*, and a confidence factor CF_m is computed and associated with each detected object:

$$CF_m = \frac{\frac{1}{K}\sum_{k=1}^{K} C(\alpha_k^*)}{\frac{1}{\Theta\times\Delta\theta}\sum_{\alpha=1}^{\Theta\times\Delta\theta} C(\alpha)}. \tag{4}$$

C. Object Recognition and Attitude Estimation

Once the object has been detected with a certain degree of reliability, the actual recognition phase occurs. This process is more complex. At first, it is able to extract from the edge image a set of N straight segments $S = \{s_n, \; n =$

[5] The method is able to search only for objects with different attitudes in the same image [11].

[6] We only consider the angle relations between adjacent boundaries, then the total number of angle relations β_{ij} will be equal to the total number K of model boundaries.

$1, \cdots, N\}$ and then to match these segments with multiple object model segments contained into an object database. However, the computational complexity of both a complete search for image segments and an exhaustive search for data-model correspondences is too high. In order to obtain an execution time of the process suitable to the application, a focus of attention phase that is able to reduce the input edge points to a set of points with a greater probability to belong to the object boundaries is performed [27]. During the voting phase applied to the first system level, a label $l(\alpha)$ has been used to mark the edge points that have contributed to improving the peaks of the histogram. The angles between the boundary segments of the 2-D model object have been used to focus the attention on the histogram peaks α_k^*, $k = 1, \cdots, K$, containing votes coming from object segments. Then, a second voting step, that is restricted to the edge points characterized by a label $l(\alpha_k^*)$, is performed to detect significant object segments.

Each straight line can be parametrically described in the reference system of the image by polar coordinates (ρ, α) [16], [23]:

$$G_2(\boldsymbol{x}, \boldsymbol{\pi}) = G_2(x, y, \rho, \alpha) = \rho - x \cdot \cos\alpha - y \cdot \sin\alpha = 0 \quad (5)$$

where ρ is the distance of the line from the origin of the reference system and $\alpha \in [-\pi/2, \pi/2]$ is the angle between the line and the x axis. Let $s(\boldsymbol{\pi})$ be a straight line on the image plane and $s_j(\boldsymbol{\pi})$ a limited straight segment belonging to $s(\boldsymbol{\pi})$.[7] The 2-D parameter space Π is quantized into cells of size $\Delta\rho$ and $\Delta\alpha$, chosen according to the rules indicated in [24] and [25] in order to reduce the spreading vote effects. Classical voting approaches [16], [23] allow the edge points $\boldsymbol{x}$, belonging to the same straight line $s(\boldsymbol{\pi})$ in the image space, to vote for the same parameter-space point $\boldsymbol{\pi}$ [24], [25]. The voting method proposed in this paper constitutes an improvement over the classical voting representation, as it allows one to associate edge points with different (suitably chosen) straight segments $s_j(\boldsymbol{\pi})$, depending on the edge-point positions along the line $s(\boldsymbol{\pi})$.

For each pixel in the image domain, which satisfies the relation $G_2(x, y, \rho, \alpha)$, the corresponding cell $C(\rho, \alpha)$ is incremented, in the 2-D parameter space, by one count and the corresponding segment endpoints are accordingly updated. At the end of the voting step, the parameter space Π is examined to select higher peaks $\boldsymbol{\pi}^* = (\rho^*, \alpha^*)$ in order to recover the segments $s_j(\boldsymbol{\pi}^*)$ representing the object boundaries in the image domain. A set of N segments, $S = \{s_n, \; n = 1, \cdots, N\}$, extracted from the input image, is created at this level. Given a set of model segments $MS = \{ms_j, \; j = 1, \cdots, J\}$, representing the 2-D projection on the image plane of the searched object, the next step is to verify if there is a group of image segments that matches that set. The problem reduces to a complete exploration of the solution space (i.e., the space of all possible matches between data and model segments) to search for feasible interpretations. It is easy to show how this search computationally becomes very heavy as soon as N and M assume reasonable values [17]. However, the number N of detected straight segments is normally very low, allowing one to perform a matching model/data that is not at all computationally heavy, despite the complexity of the matching problem.

The proposed method in this paper follows some basic ideas developed in [17] to implement the matching process and makes use of geometric constraints to reduce the search. The constraints imposed on segments may be intrinsic (e.g., length) or relational (e.g., segment-to-segment distance, angle, symmetry, etc.) and are stored in a database generated off-line. Therefore, the matching process result is the recognition of an object, i.e., an association between data and model segments. The recognition can also be performed in the case of incomplete data due to partial occlusion or to a high noise level if the detected pairs of data and model segments are greater than a fixed threshold, depending on the type of object to be searched.

After the recognition step, the presented approach allows for a more accurate object attitude estimation based on the position of the detected object segments. For example, if an isosceles triangle (e.g., an arrow) has been detected in the image, the object attitude (in particular, the orientation of the arrow, useful to drive an AUV) is given by the axis between the longer object sides. In general, if one or more segments are partially missing, but the object has been recognized thanks to the model information, it is possible to estimate equally the object orientation by substituting the lost segments with the corresponding model object segments [17].

III. Results

The results aim to verify and to compare the validity of the proposed method by testing it on images acquired with a simulated forward looking sonar and on real images acquired with an acoustic camera.

A. Forward-Looking Sonar Images

A set of images acquired with a forward-looking sonar installed on an underwater vehicle surveying the sea bottom at a depth of 5 m has been generated as follows. The data received by the array of sensors were simulated by properly modeling the characteristics of the scene objects and the sea bottom, then the images were formed by a dynamic focused beamforming. The simulation is a flexible way to generate test images with different combinations of the sea-bottom reverberation and the object speckle. The array was composed of 50 1.5λ-equispaced elements (λ being the wavelength), the carrier frequency was of 450 kHz, the angular aperture of the field of view was of ±18°, and the along-track extension of the field of view ranged between 30° and 60° of elevation angle (0° being the elevation angle of the vertical). Fig. 3 shows a set of six images in which an isosceles triangle is present in the field of view with different orientations and distances that are ranged from 6 to 9 m, as it may occur in a potential application in which an AUV is routed by

[7] Each collinear segment $s_j(\boldsymbol{\pi})$ is defined as a set of edge points $\boldsymbol{x}$ satisfying the generating equation and the two constrained endpoints, $s_j(\boldsymbol{\pi}) = \{x\colon \; G_2(\boldsymbol{x}, \boldsymbol{\pi}) = 0, \boldsymbol{x} \in s(\pi), \boldsymbol{x} \in [\boldsymbol{x}_j^{\min}(\pi), \boldsymbol{x}_j^{\max}(\pi)]\}$. A 1-D reference system, whose origin is placed at the foot point of the normal, is defined along each straight line $s(\boldsymbol{\pi})$ [23].

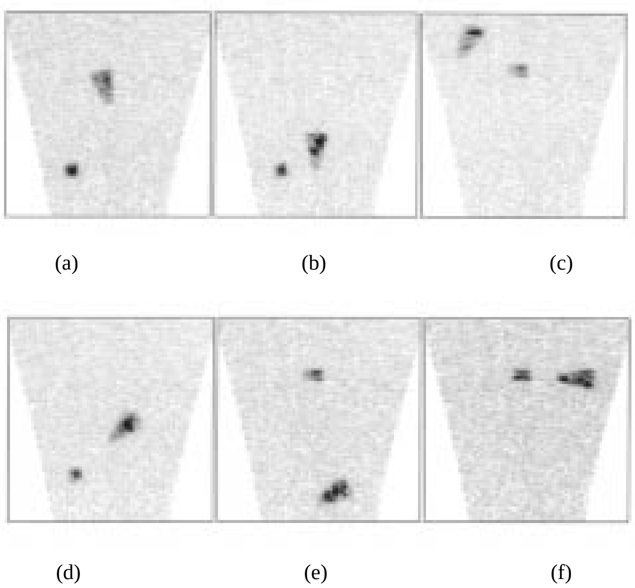


Fig. 3. Set of six simulated forward-looking sonar images in which an isosceles triangle is present together with another spurious object. From panel (a) to panel (f), the exact triangle orientation is respectively $-14°$, 0°, 25°, 45°, 60°, and 90°. The SNR is equal to 15 dB.

means of a sequence of known landmarks fixed to the sea bottom. The triangle, whose dimensions were 0.5 m of base and 1 m of height, was simulated like a slightly rough surface by using a densely packed distribution of random scatterers [28], [29]. In this manner, a realistic speckle noise affects the object image and it is possible to test the algorithm efficiency with respect to different occurrences of the speckle by modifying the relative position between the sonar antenna and the object and by repeating the simulation. Different from the object response, the sea-bottom scattering was simulated, as usual, by supposing the echo amplitude to have a Rayleigh distribution [30], [31]. In this way, it is possile to test the algorithm efficiency with respect to the level of the sea-bottom reverberation, by tuning the variance of the Rayleigh distribution. Since, for a given position of the object and the antenna, the echo of the object does not change, the tuning of the variance is about equivalent to acquire images with a different signal-to-noise ratio (SNR). The SNR can be defined as the ratio between the mean power of the object echo and the mean power of the bottom echo. The beam signals obtained by the above simulation were mapped on images of 256×256 pixels in a flat bottom hypothesis by displaying strong echo amplitudes by dark-grey-level pixels. For a given level of sea-bottom reverberation, six different images were generated (see Fig. 3); the results are six different realizations of the speckle noise. This generation of six images was repeated three times by assigning three different levels of sea-bottom reverberation; then, the total number of simulated images is 18. For the six images of Fig. 3, the SNR is equal to 15 dB; the other two sequences are differently characterized by 5 and 10 dB of SNR.

The current orientation of the triangle has been measured by the anticlockwise rotation angle that is necessary to place the triangle in the panel (b) of Fig. 3, i.e., the parallel axis to the motion track and the arrow pointed toward the vehicle. By this convention, the exact orientation of the triangle, from panel (a) to panel (f) of Fig. 3, is, respectively, $-14°$, 0°, 25°, 45°, 60°, and 90°. Every image represents a region of 6.5 m × 6.5 m^2 and also contains another spurious object [a 25 cm × 25 cm^2 square in panels (a), (b), and (d), and an 18 cm × 35 cm^2 rectangle in panels (c), (e), and (f)]. The algorithm was applied once for each of the 18 simulated images and it achieved the results that are presented in Table I. The algorithm recognized a triangle in all six images, for both SNR $=$ 15 dB and SNR $=$ 10 dB. The measured orientations were reported in Table I together with the mean value of the absolute error, equal to 1.4° at SNR $=$ 15 dB and equal to 3° at SNR $=$ 10 dB. When the SNR is lowered to 5 dB, the triangle is recognized only in three of the six images and in these cases the absolute error is between 3.5° and 4.5°.

TABLE I
MEASURED ANGLES OF THE TRIANGLE ORIENTATION AT 15, 10, AND 5 dB OF SNR, COMPARED TO THE EXACT ORIENTATION

Image	a	b	c	d	e	f	$E\{\|\varepsilon\|\}$
Exact orientation	-14.0°	0.0°	25.0°	45.0°	60.0°	90.0°	-
Measured, SNR = 15 dB	-10.5°	0.0°	25.5°	46.5°	61.5°	91.5°	1.4°
Measured, SNR = 10 dB	-16.5°	3.0°	28.5°	46.5°	55.5°	93.0°	3.0°
Measured, SNR = 5 dB	N.R.	-4.5°	21°	N.R.	N.R.	93.5°	-

The letter of image identification is referred to those shown in Fig. 3. N.R. means that the triangle was not recognized and $E\{|\epsilon|\}$ is the average value of the absolute error.

Moreover, in the image that is shown in Fig. 3(b) (SNR $=$ 15 dB), the little square was also recognized, while the rectangle was recognized in all the three images in which it was present [i.e., in Fig. 3(c), (e), and (f)]. When the SNR is equal to 10 dB, the square was recognized only in the image (d), while the rectangle was recognized only in the image (e). No recognition of the objects different from the triangle occurred when SNR $=$ 5 dB.

One can notice that a good accuracy has been achieved when the SNR is 15 dB while a little bit poorer accuracy has been obtained at SNR $=$ 10 dB. When the reverberation level of the sea bottom further increases (i.e., SNR $=$ 5 dB), the recognition of the triangle is more difficult as the initial segmentation of the image of the thresholding operation does not always result in a reliable edge image. Thus, in three propitious cases, the edges are sufficient to recognize the object with an error of some degrees, but in the other three cases the triangle is not recognized at all. To overcome this fact, it is necessary to improve the segmentation step by using more specific techniques. It is importat to stress that these results were obtained by keeping fixed all the parameters affecting the algorithm, i.e., no regulations based on the data were necessary. Fig. 4 presents some intermediate results: panels (a) to (c) show the images obtained by the edge detection performed as it has been explained in the previous sections on the basis of the original images, respectively shown in the panels (a) to (c) of Fig. 3, whereas, panels (d) to (f) of Fig. 4 show a reconstruction of the detected segments in the above-mentioned edge images.

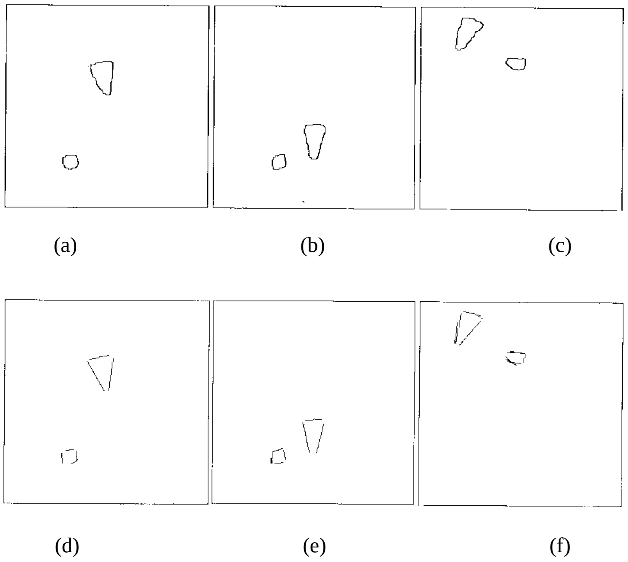


Fig. 4. Intermediate results. (a)–(c) The images obtained by the edge detection of the images shown in Fig. 3(a)–(c). (d)–(f) The reconstructions of the segments detected in the edge images.

Considering a 256×256-pixels image, the computation times of about 0.2 and 0.8 s (on a SUN Sparc 10 workstation) are respectively required by the 1-D and 2-D voting functions. The matching step requires 0.2 s to obtain a complete object recognition and 0.3 s for the object reconstruction.

B. Comparison with Other Methods

In order to demonstrate the superiority of the proposed method, we compare the obtained results on the simulated images with those of the other two classical methods applied in Computer Vision for object detection and attitude estimation: the template matching (TM) method [17] and the moment of inertia (MI) method [17]. The TM method is based on the definition of an ideal representation (template) of the object model to be detected in a scene. The template is shifted on the whole image and at each step a correlation measure is computed. The MI method is based on the estimation of the principal axes of inertia of the image. It can be estimated with respect to the coordinate system that finds its origin in the object centroid, and it gives the orientation of the main body of the object. Tables II and III show the obtained results for the TM and the MI methods, respectively. It is important to note that for all noise percentages that have been considered, the proposed method makes out better results: the object attitude is estimated more accurately than by the other approaches. In particular, the TM method shows comparable estimation results (absolute error equal to 1.7) for SNR = 15 dB, but, for lower SNR's, it is not able to detect the object. The MI method is able to detect the object also in the presence of high noise levels, but it makes out higher absolute errors.

TABLE II
MEASURED ANGLES OF THE TRIANGLE ORIENTATION AT 15, 10, AND 5 dB OF SNR, OBTAINED BY THE TEMPLATE MATCHING METHOD

Image	a	b	c	d	e	f	$E\{\lvert\varepsilon\rvert\}$
Exact orientation	-14.0°	0.0°	25.0°	45.0°	60.0°	90.0°	-
Measured, SNR = 15 dB	-12.5°	2.0°	23.5°	43.5°	58.5°	90.5°	1.7°
Measured, SNR = 10 dB	N.R.	6.5°	N.R.	39.5°	N.R.	85.5°	-
Measured, SNR = 5 dB	N.R.	N.R.	N.R.	N.R.	N.R.	N.R.	-

N.R. means that the triangle was not recognized and $E\{|\epsilon|\}$ is the average value of the absolute error.

TABLE III
MEASURED ANGLES OF THE TRIANGLE ORIENTATION AT 15, 10, AND 5 dB OF SNR, OBTAINED THE MOMENTS OF INERTIA METHOD

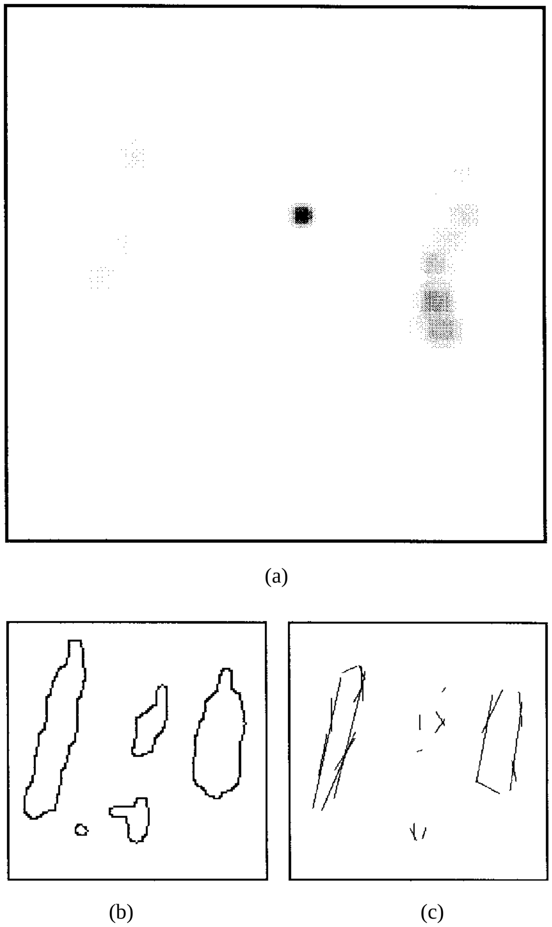
Image	a	b	c	d	e	f	$E\{\lvert\varepsilon\rvert\}$
Exact orientation	-14.0°	0.0°	25.0°	45.0°	60.0°	90.0°	-
Measured, SNR = 15 dB	-8.5°	2.5°	27.5°	49.5°	64.5°	87.5°	3.5°
Measured, SNR = 10 dB	-3.5°	5.0°	21.5°	53.5°	57.5°	81.0°	6.5°
Measured, SNR = 5 dB	N.R.	-10.5°	NR	N.R.	N.R.	80.5°	-

N.R. means that the triangle was not recognized and $E\{|\epsilon|\}$ is the average value of the absolute error.

C. Acoustic Camera Images

Acoustic cameras that are able to form a 3-D map of a scene and to project it on a plane, to thus obtain a 2-D image [2]–[3], have recently been put on the market. Among all the possible projections, the orthoscopic image (i.e., the projection on a plane parallel to the array) is something of common interest for its similarity to the optical images. Thanks to 3-D capabilities, each point in such an image has a known distance from the camera, but this fact does not influence our algorithm. Fig. 5 shows an orthoscopic image obtained by an Echoscope 1600, a real-tme acoustic 3-D camera that has been designed and produced by a Norwegian company. In this case also, the higher the amplitude of an echo, the darker the related pixels. The imaged scene is composed by two anchor chains and a small round object (8 cm in diameter), about 5 m away from the 40×40 array elements of the camera. The image in Fig. 5, acquired with a 600-kHz carrier and by displaying a region of 3.15 m × 3.15 m^2, shows two anchor chains like two very narrow rectangles that are closed to the verticality; more precisely, the chain on the left side has an inclination of 14° with respect to the verticality, whereas the other chain has an inclination of 8°. The navigation of underwater vehicles near man-made structures and the robotic manipualtion are two possible applications of the 3-D camera and, in this context, the capabilities of the proposed method represent useful tools.

We can notice that for the described method it is impossible to distinguish between an anchor chains and another object having a similar geometrical projection. In particular, in the ideal case, an anchor chain has been modeled as a pair of parallel straight segments characterized by a similar length and by a distance much shorter than the length. On the basis of this model, the discussed algorithm succeeded in the detection of two objects classified as anchor chains and their orientations were estimated equal to 15° for the left chain and 9° for the right chain (the actual orientation were, respectively, 14° and 8°). As for the simulated images, the algorithm also performed well enough in this case and orientation errors appeared small with respect to the mediocre quality of the acoustic images and the related understanding difficulties.



(a)

(b) (c)

Fig. 5. (a) Real image acquired with an acoustc camera, showing two anchor chains and a small round object in between. (b) Result of the edge-detection process. (c) Result of the voting-based (straight-line extraction) approach.

IV. Conclusion

In this paper, a voting approach that aimed at processing short-range acoustic images of submerse scenes for object recognition and orientation estimation has been presented. The method works on the projections of objects on the image plane and is able to recognize man-made objects on the basis of *a priori* known models. Thanks to the simple voting nature that is suitable to the addressed application, a considerable robustness to noises and typical defects of acoustic imaging and a valuable computational efficiency has been achieved. An assessment of the method performances has been performed on both simulated forward-looking sonar and real acoustic camera images, providing encouraging results.

The future improvement of the proposed method will lie in the capability to work with 3-D information, when are available (e.g., in the acoustic camera case). This should be possible, without a great increase of the computational complexity, by analyzing several orthoscopic projections of the same object from different viewpoints. An adequate model for each projection that is employed should be stored in an opportune database. In this way, different objects having the same shape in the orthoscopic projection may be correctly recognized also. The object recognition can be improved by searching not only for single segments in the image but searching for more specific segment groups as well. Collinear, parallel, and convergent segments could be directly detected in the parameter space [27]. A model-driven focus of attention mechanisms could be used to search for segment groups relating to different 3-D views of the object.

Acknowledgment

The authors would like to thank the anonymous reviewers for their useful comments and counsels, Dr. R. K. Hansen and Dr. R. Jakobsen of Omnitech S/A (Norway) who kindly provided real data acquired with their real-time 3-D acoustic camera, and S. Papini for the important support in implementing the proposed method and for his precious suggestions.

References

[1] R. K. Hansen and P. A. Andersen, “3-D acoustic camera for underwater imaging,” *Acoust. Imaging,* vol. 20, pp. 723–727, 1993.

[2] N. Cesbron, F. Ollivier, P. Alais, and P. Challande, “A 3-D acoustical imaging system,” in *Proc. Eur. Conf. Underwater Acoustics,* Luxembourg, 1992, pp. 756–759.

[3] C. J. M. Van Ruiten and J. C. Bu, “Development of an acoustic camera for ROV’s,” in *Proc. Eur. Conf. Underwater Acoustics,* Luxembourg, 1992, pp. 733–736.

[4] Manual of the SeaBat 6012 Multibeam Forward Looking Sonar System, Goleta, CA: Reson, 1995.

[5] D. G. Lowe, *Perceptual Organization and Visual Recognition.* Norwell, MA: Kluwer Academic, 1985.

[6] P. K. Keating, T. Sawatari, and G. Zilinskas, “Signal processing in acoustic imaging,” in *Proc. IEEE,* vol. 67, pp. 496–509, Apr. 1979.

[7] S. G. Johnson and M. A. Deaett, “The application of automated recognition techniques to side scan sonar imagery,” *IEEE J. Oceanic Eng.,* vol. 19, pp. 138–141, Jan. 1994.

[8] D. M. Lane and J. P. Stoner, “Automatic interpretation of sonar imagery using qualitative feature matching,” *IEEE J. Oceanic Eng.,* vol. 19, pp. 391–405, July 1994.

[9] L. M. Linnett and D. R. Carmichael, “The analysis of sidescan sonar images for seabed types and objects,” in *Proc. 2nd Conf. Underwater Acoustics,* Copenhagen, Denmark, July 1994, pp. 733–738.

[10] L. Henriksen, “Real-time underwater object detection based on an electrically scanned high-resolution sonar,” in *Proc. IEEE 1994 Symp. Autonomous Underwater Vehicle Technology,* Cambridge, MA, July 1994.

[11] M. Ashraf and J. Lucas, “Underwater object recognition techniques using ultrasonics,” in *Proc. IEEE Int. Conf. Oceans 94 Osates,* Brest, vol. I, September 1994, pp. 170–175.

[12] S. Guillaudeux, S. Daniel, and E. Maillard, “Optimization of a sonar image processing chain: A fuzzy rules based expert system approach,” in *Proc. Int. Conf. Oceans 96 MTS/IEEE,* Ft. Lauderdale, FL, Sept. 1996, pp. 1319–1323.

[13] C. Collet, P. Thourel, P. Pérez, and P. Bouthemy, “Hierarchical MRF modeling for sonar picture segmentation,” *IEEE Int. Conf. Image Processing ICIP-96,* Lausanne, Switzerland, Sept. 1996, vol. III, pp. 979–982.

[14] V. Murino and A. Trucco, “Ultrasonic imaging improvement by a cooperative approach,” in *Proc. Inst. Acoust., Sonar Signal Processing,* vol. 17, pt. 8, pp. 59–68, 1995.

[15] J. F. Canny, “Finding edges and lines in images,” in *Artificial Intelligence Laboratory,* Tech. Rep. TR-720, MIT, Cambridge, MA, 1983.

[16] H. Illingworth and J. Kittler, “A survey on the Hough transform,” *Computer Vision, Graphics and Image Processing,* vol. 44, pp. 87–116, 1988.

[17] W. E. L. Grimson, *Object Recognition by Computer, The Role of Geometric Constraints.* Cambridge, MA: MIT Press, 1990.

[18] S. Sarkar and K. Boyer, “A computational structure for preattentive perceptual organization: Graphical enumeration and voting methods,” *IEEE Trans. Syst., Man, Cybern.,* vol. 24, pp. 246–266, Feb. 1994.

[19] S. Negahdaripour and S. Zhang, “Determining the size and position of cylindrical objects from optical images in underwater environments,” in *9th Int. Symp. Unmanned Unthetered Submersible Tech.,* Durham, NH, Sept. 1995, pp. 232–242.

[20] B. A. A. P. Balasuriya, T. Fujii, and T. Ura, "Underwater pattern observation for positioning and communication of AUVs," in *9th Int. Symp. Unmanned Unthetered Submersible Tech.,* Durham NH, Sept. 1995, pp. 193–201.

[21] G. Foresti, V. Murino, C. S. Regazzoni, and G. Vernazza, "Grouping of straight segment by the labelled hough transform," *Computer Vision, Graphics and Image Processing: Image Understanding,* vol. 58, pp. 22–42, 1994.

[22] J. Serra, *Image Analysis and Mathematical Morphology.* London, U.K.: Academic, 1982.

[23] R. O. Duda and P. E. Hart, "Use of the HT to detect lines and curves in pictures," *Commun. ACM,* vol. 15, no. 1, pp. 11–15, 1972.

[24] T. M. Van Veen and F. C. A. Groen, "Discretization errors in the Hough transform," *Pattern Recogn.,* vol. 14, pp. 137–145, 1981.

[25] W. Niblack and D. Petkovick, "On improving the accuracy of the Hough transform," *Machine Vision Applicat.,* vol. 3, pp. 87–106, 1990.

[26] C. M. Brown, "Inherent bias and noise in the Hough transform," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-5, pp. 493–505, 1981.

[27] G. L. Foresti, V. Murino, and C. S. Regazzoni, "Grouping as a searching process for minimum-energy configurations of labelled random fields," *Computer Vision and Image Understanding,* vol. 64, no. 1, pp. 157–174, July 1996.

[28] O. George and R. Bahl, "Simulation of backscattering of high frequency sound from complex objects and sand sea-bottom," *IEEE J. Oceanic Eng.,* vol. 20, pp. 119–130, Apr. 1995.

[29] E. Jakeman and R. J. Tough, "Generalized K distribution: A statistical model for weak scattering," *J. Opt. Soc. Amer.,* vol. 4, pp. 1764–1772, Sept. 1987.

[30] T. L. Henderson and S. G. Lacker, "Seafloor profiling by a sideband sonar: Simulation, frequency-response, optimization, and results of a brief sea test," *IEEE J. Oceanic Eng.,* vol. 14, pp. 94–107, Jan. 1989.

[31] H. Boeheme, N. P. Chotiros, T. G. Goldsberry, S. P. Pitt, R. A. Lamb, A. L. Garcia, and R. A. Altenburg, "Acoustic backscattering at low grazing angles from the ocean bottom. Part II. Statistical characteristics of bottom backscatter at a shallow water site," *J. Acoust. Soc. Amer.,* vo. 77, pp. 975–982, 1985.

Gian Luca Foresti (S'93–M'95) was born in Savona, Italy, in 1965. He received the Laurea degree in electronic engineering in 1990 and the Ph.D. degree in computer science in 1994 from the University of Genoa, Italy.

Immediately after receiving the Laurea degree, he worked with the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa, Genoa, Italy, in the area of computer vision, signal processing, and image understandng. He was Visiting Professor at Trento University in 1994. Currently, he is an Assistant Professor at the Department of Mathematics and Computer Science (DIMI), at the University of Udine, Udine, Italy. His main interests involve distributed data fusion in multisensor systems, probabilistic and symbolic techniques in image processing, and artificial neural networks. The proposed techniques found their applications in the following fields: automatic systems for surveillance of outdoor environments (e.g., underground stations, railway lines, etc.), vision systems for autonomous vehicle driving and/or road traffic control, 3-D scene interpretation and reconstruction. He is author or co-author of more than 50 papers published in international journals and conferences, and a reviewer for several international scientific journals and EC research programs (MAST III, Long Term Research, CRAFT). He is a member of the Technical Committee of some International Conferences on image processing and vehicle automation,

Dr. Foresti is a member of IAPR and AEI.

Vittorio Murino (S'92–M'93) was born in Lavagna, Genova, Italy, in 1964. He received the Laurea degree in electronic engineering in 1989 and the Ph.D. degree in 1993 from the University of Genoa, Genoa, Italy.

He was a Post-Doctoral Fellow at the Univerity of Genoa, working in the Signal Processing and Understanding Group of the Department of Biophysical and Electronic Engineering, as supervisor of research activities concerning with signal and image processing in underwater environment. He worked at several national and European projects funded by the European Commission, especially in the context of the MAST (MArine Science and Technology) program concerning the investigation of underwater scenes by visual and acoustcal sensors. He is also an evaluator for the European Commission of project proposals for the MAST III, Industrial and Material Technology (BRITE-EURAM III) and Long Term Research (LTR) programs. At present, he is Assistant Professor at the Department of Mathematics and Computer Science of the University of Udine, Udine, Italy, and is working at a project funded by the European Commission concerning multisensorial underwater vision for object recognition. His main research interests include acoustic and optical underwater imaging, probabilistic techniques for signal/image reconstruction, data fusion, and pattern recognition.

Dr. Murino is a member of IAPR.

Carlo S. Regazzoni (S'90–M'92) received the Laurea degree in electronic engineering and the Ph.D. degree in telecommunications and signal processing from the University of Genoa, Genoa, Italy, in 1987 and 1992, respectively.

He joined the Signal Processing and Understanding Group (SPUG) in the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, in 1987, where he has been responsible for the Industrial Signal and Image Processing (ISIP) area since 1990. He was Visiting Scientist for different periods at the University of Toronto, Toronto, Canada, in 1993–1995. He has been a member of the Transport Research Centre of the University of Genoa since 1995. He has been Assistant Professor in Telecommunications at the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa since 1995. His main research interests are probabilistic nonlinear techniques for signal and image processing, nonconventional signal detection and estimation techniques, and distributed data fusion in multisensor systems. Moreover, he is strongly involved in the application of such methodologies within the support systems for the transport field (e.g., distributed surveillance systems, positioning systems, etc.). In this context, the ISIP area has participated to several EU research and development projects (ESPRIT (P7809 DIMUS, P8433 PASSWORDS, P6068 ATHENA)). He is a referee for several international journals and he has been a reviewer for EU projects (ESPRIT BRA, 1993, LTR 1994–1995). He has been chairman and a member of the technical committee at several conferences (EUROPTO '94, EUSIPCO '96).

Dr. Regazzoni is a member of IAPR and AIIA.

Andrea Trucco was born in Genoa, Italy, in 1970. He received the Laurea (M.Sc.) degree in electronic engineering from the University of Genoa, Genoa, Italy, in June 1994. He is currently working towards the Ph.D. degree in the Department of Biophysical and Electronic Engineering (DIBE) of the same University and cooperates with the Marine Research Area of the Signal Processing and Understanding Group at DIBE.

His main research interests are array synthesis, coherent and noncoherent algorithms for acoustic imaging, acoustic image improvement, interferometric sonar and radar, simulation methodologies.

Mr. Trucco is a member of IAPR. From 1987 to 1990, he was three times a finalist for the Philips Awards for European Young Scientists, and ranked third twice. In 1995, he won the Student Paper Competition organized by the 9th International Symposium on Unmanned Untethered Submersible Technology.