BAYESIAN INFERENCE AND OPTIMIZATION STRATEGIES FOR SOME DETECTION AND CLASSIFICATION PROBLEMS IN SONAR IMAGERY

M.Mignotte ‡ *C.Collet* ‡ *P.Pérez* • *P.Bouthemy* •

‡ Groupe de Traitement du Signal, Ecole Navale, Lanvéoc-Poulmic, 29240 Brest-Naval France.
e-mail: name@poseidon.ecole-navale.fr

• IRISA/INRIA, Campus Universitaire de Beaulieu, 35042 Rennes cedex, France.
e-mail: name@irisa.fr

ABSTRACT

In this paper, we investigate the use of the Bayesian inference for some detection and classification problems of great importance in sonar imagery. More precisely this paper is concerned with the segmentation of sonar image, the classification of object lying on the sea-bottom and the classification of sea-floor. These aforementioned classification tasks are based on the identification of the detected cast shadows which correspond to a lack of acoustic reverberation behind the different natural or man-made objects lying on the sea-floor. The adopted Bayesian approach allows to model efficiently all the available prior information, for each detection and classification task under concern yielding a cost minimization problem. To this end, we associate to each Bayesian statistical modeling, a specific optimization strategy well suited to the global energy function to be minimized. These segmentation and classification schemes can be used separately for a specific application or can lead to an original Bayesian processing chain for the automatic classification of objects lying on the sea-floors. The efficiency and robustness of this unsupervised processing chain has been tested and demonstrated on a great number of real and synthetic sonar images.

keywords: *Bayesian inference, optimization strategy, detection, segmentation, classification, sonar imagery.*

1. INTRODUCTION

High-resolution sidescan sonar plays an important role in underwater sensing, due to its capability of providing high-quality acoustic images of the sea-bed far beyond the one supplied by optical means [1]. However, because of the large data amount to process, the exploitation of the collected data has now to be achieved automatically.

Amongst the sonar image processing applications of great interest, one can cite the detection of all objects lying on the sea-bed and their classification in order either to separate natural objects from man-made objects (low level classification step) or to identify different types of manufactured objects based on the shape of their shadows (higher level classification task). For a high resolution sonar, detection and classification of objects of interest are generally based on the extraction and the identification of their associated cast shadows. These shadows are due to a lack of acoustic reverberation behind each object. Thus, classification tasks require a preliminary segmentation of the sonar image in order to extract the different shadow regions. This segmentation map can also be interesting in other applications related to the sonar image processing like three dimensional reconstruction of underwater objects [2] or for the symbolic matching processes of sonar image. Another application, investigated in this paper, is the sea-floor classification into homogeneous geoacoustic regions such as sand, pebbles, ridges, etc... Sea-bed classification is essential in a wide range of applications, such as geological survey (cartography of sea-floors, geophysical exploration, ...), ocean engineering, military field (route surveying, ...). It may also improve the aforementioned detection and object classification tasks [3].

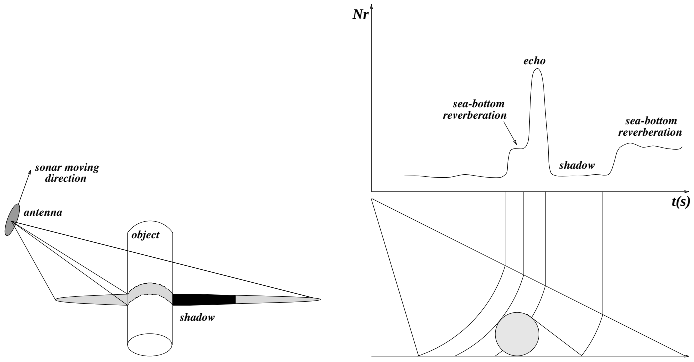
For all these detection and classification issues, some prior knowledge on the information to be extracted from the original image is available. Bayesian statistical theory is a convenient tool to take this *a priori* knowledge into consideration. This paradigm has become quite popular in image analysis and has been primarily applied, with success to low level image analysis tasks. We can cite, for example, image segmentation [4], image restoration [5], edge detection [6], surface reconstruction [7], visual motion analysis [8], etc. In these applications, the proper use of the available prior information is expressed by local prior models. However, the use of Bayesian theory makes it possible to apply more general models. Indeed global prior models can be used. In these cases, Bayesian inference has recently proved to be also successful for high level image analysis tasks such as scene interpretation [9], shape matching with deformable template-based methods [10] or classification problems [11]. The Bayesian methodology allows to combine a prior model that expresses the available prior information on the hidden variables, or the solution to be estimated, along with a statistical model describing the interactions between hidden and observed variables. For the aforementioned detection and classification issues, we have to deal, different types of *a priori* informations have to be considered:

[1] **Acknowledgments:** The authors thank GESMA (Groupe d'Étude Sous Marine de l'Atlantique), Brest, France, for having provided numerous real sonar pictures, and DGA, (Direction Générale de l'Armement, French Ministry of Defense) for partial financial support of this work (student grant).

• For the sonar image segmentation problem, the hidden variables are labels constituting the segmentation map (also called the label field). A simple prior may express the fact that nearby pixels are fairly likely to belong to the same class. In a probabilistic framework, such regularities are well captured by Markov Random Fields (MRF) [4]. Thanks to the Hammersley-Clifford theorem, this prior model can be described by a Gibbs distribution based on local interactions. Contextual knowledge is thus simply captured through the specification of spatially local interactions, called clique potentials. In order to extract a reliable and accurate segmentation map, contextual information at a higher level of representation, particularly at region level, is also interesting to be taken into account. In our application, there is a spatial dependency between the echo regions, due to the reflection of the acoustic wave on the object, and the shadow regions, corresponding to a lack of acoustic reverberation behind each object lying on the sea-bed. Such an *a priori* information relative to the interaction between two types of areas can be efficiently exploited in the Bayesian framework.
• For the sea-floor classification problem, we can exploit our knowledge about the type of shadow patterns corresponding to each sea-bottom type. Besides, we can also specify homogeneity properties of the desired sea-bottom segmentation map in order to favor homogeneous region of identical sea-floor.
• Finally, for the classification of objects lying on the sea-floor, prior can capture the fact that, contrary to the cast shadow of a natural object, the one of a manufactured object exhibits a regular and/or shape easily identifiable. This prior knowledge can be specified using a prototype template along with a set of admissible linear transformations accounting for the shape variability. This constitutes a global prior model, as opposed to the MRF-type local prior. In this case, a global model is better suited than a local one in order to obtain the geometric information that is inherent to this problem.
Using the available *a priori* knowledge, detection and classification issues can be stated in the Bayesian machinery which translates them into a cost function minimization problem. To cope with these minimizations, many deterministic or stochastic relaxation methods have been proposed in the literature. Nevertheless, some of them are inefficient when no good initial guess is available and other are computationally too expensive. In fact, depending on the type of prior model and likelihood model, energy functions may be very different and an optimization strategy may be well suited only for a specific global energy function type to be minimized.
This paper is organized as follows. In Section 2, we present the sonar imaging process, followed by the Bayesian inference principle in Section 3. Different Bayesian modelings and associated optimization strategies are presented and discussed in Section 4, 5 and 6 for the sonar image segmentation, object classification, and sea-floor classification respectively. Characteristics and advantages of each optimization strategy over the others are discussed. Section 7 describes a complete chain for the automatic classification of objects lying on the sea-bed. Finally, section 8 contains concluding remarks.

2. SONAR IMAGERY

As opposed to electromagnetic waves, acoustic waves allow the transmission of the signal in the sea environment with low attenuation. The sonar is an emission and reception system using this property. In the emission stage, the sidescan sonar generates a short electrical impulse whose frequency determines the sonar characteristics. This impulse is transmitted to the emission antenna that converts it into an highly directive acoustic wave in the orthogonal direction of the sonar displacement (see Fig. 1.a). For each impulse, a sampling of the reverberated signal in the reception stage allows to obtain the acoustic wave amplitude value and thus the pixel grey level of the corresponding sonar image line. Figure 1.b shows the collected signal evolution in function of time, or in function of the distance in the acoustic wave emission direction, for a cylindrical object lying on the sea-floor.



(a) (b)

Figure 1: Sonar image formation physical process. (**a**) Cylinder and cast shadow shape in the region acoustically obscured by this object. (**b**) Evolution of the reverberated signal as a function of time, or as a function of the distance in the observation direction.

Figure 1 shows that three kinds of regions can easily be identified in sonar imagery: the first one, called "sea bottom reverberation" area, is due to the sonar signal reverberation on the sea-bed. The second one, the "echo" information,

corresponds to the reflection of the acoustic wave on the object and the third region, the "shadow" zone, arises from the lack of acoustic reverberation behind each object or behind each morphological element (such as ridges, dunes of sand, pebbles, rocks, ...) located on the sea-floor.
On the pictures supplied by a sonar, the echo features are generally less discriminant than the shadow shapes of each object lying on the sea-bed. Besides, for each morphological elements located on the sea-bottom, the echo does not exist and the cast shadow is the only information that can be efficiently exploited. That is why classification of these objects or the identification of sea-floor type are based on the extraction and the identification of their detected cast shadows. In our application, the cast shadow constitute the useful two dimensional signal that will be detected during the segmentation step and then exploited during the classification step. Unfortunately, as all types of coherent imaging systems, such as Synthetic Aperture Radar (SAR), ultrasound medical or laser illuminated imagery, the input sonar image is strongly corrupted by speckle noise [1]. This partially correlated noise is due, both to the random interferences of the acoustic waves scattered by the microscopic fluctuations of the object or the sea-bottom surface within one resolution cell, and to the minor lobes of the acoustic antenna bringing back-scattered signal. Speckle noise is a granular noise that reduces noticeably the resolution of the sonar image and sharpness of the cast shadow contours. Thus, before any classification step, one must statistically model this speckle noise and estimate the associated parameters in order to perform an unsupervised preliminary segmentation step. This will be made in Section 4.

3. BAYESIAN INFERENCE

Let X represents the set of hidden variables to be recovered, and Y, the field of observed data or the observation variables. The general Bayesian inference defined by Besag [4] consists of four successive stages outlined below:

1) construct of a prior probability distribution or a prior model $P_X(x)$ which represents our initial prior knowledge on the solution.
2) combine the observed image with the underlying prior model through a conditional Probability Density Function (PDF) $P_{Y/X}(y/x)$, also called the data likelihood model.
3) compute the posterior density $P_{X/Y}(x/y)$ from the prior and the conditional probability by Bayes theorem:

$$P_{X/Y}(x/y) \propto P_X(x)\; P_{Y/X}(y/x) \tag{1}$$

4) finally, choose a criterion based on the posterior probability (1), in order to define the "best" x given y.

In Bayesian analysis, all kinds of inference are made from $P_{X/Y}(x/y)$. Finding the Maximum A Posteriori (MAP) estimate i.e., the occurence of x that is the most likely, is one of the most frequently used choice of inference. For $P_{X/Y}(x/y) \propto \exp(-U(x,y))$, for some "energy" function U, this MAP estimation problem is equivalent to an energy minimization problem:

$$\hat{x}_{\text{MAP}} = \arg\max_x \; P_{X/Y}(x/y) = \arg\min_x \; U(x,y) \tag{2}$$

The energy function $U(x,y)$ involves two components. One arising from $P_{Y/X}$ of which expresses the interaction between the hidden variables and the observed data (usually called the data likelihood term). The other one, stemming from $P_X(x)$, encode constraints on the desired solution. It is the *a priori* term. Depending on the type of prior (local or global), on the number of hidden variables to be estimated, and on the nature of the observation field, considered energy landscapes may be very different (i.e., nearly concave sometimes but usually very complex with many local extrema over the parameter space). A global search is usually impossible due to the size of the configuration space. Nevertheless, several other minimization methods exist. Deterministic relaxation techniques can be efficiently used when a good initial guess of the solution is available, otherwise these methods converge to configurations corresponding to local minima of the global energy function. Stochastic methods have the capability of avoiding local minima and no initial guess is required to initialize the search procedure. However, one of their major drawbacks is their high computational load. In our applications, we will have to find the more appropriate (i.e., the faster and the more efficient) minimization technique for each of these different energy functions we shall introduce.

4. SEGMENTATION

In order to efficiently model all the available *a priori* information about the desired segmentation map that will be used in the classification step, we decompose the segmentation task as a two-step process:
• In a first step, the sea-bottom reverberation and the echo classes are not distinguished. The "reverberation class" is thus considered as capturing both types of regions. We use a segmentation model specifically well suited to face the speckle noise effect and to extract the shadow regions from the reverberation ones.
• In a second step, we perform a Markovian segmentation into two classes of the region previously labelled as "reverberation area". This aims at distinguishing echo areas from sea-bottom reverberation ones. To this end, we use a spatial MRF prior model integrating *a priori* information at different levels of representation, pixels and regions.

4.1. Two-class segmentation strategy

The segmentation of the sonar images is made difficult by the presence of speckle noise [1]. This granular and correlated noise makes ineffective simple schemes, such as Maximum Likelihood (ML) segmentation. In order to extract a reliable and accurate segmentation map, contextual information is important to be taken into account. This can be done using MRF models [4] which provide a powerful tool to incorporate an *a priori* knowledge about the spatial statistics of the desired label field. More precisely, MRFs enable to model the local characteristics of the image content thanks to *a priori* specification of spatial or hierarchical dependencies between "neighboring" variables. The knowledge about the scene is incorporated into a Gibbs distribution whose energy function consists of the sum of appropriate clique potentials.
Nevertheless, purely spatial MRF models have a limited ability to describe properties on large scale, and may be not sufficient to ensure the regularization of the set of labels when the sonar image contains strong speckle noise. For that reason, a simple spatial Ising-type prior model with a 2^{nd} order neighborhood system cannot be efficiently used for the sonar image segmentation issue [12, 13]. In order to face speckle noise, an other way is to use a hierarchical MRF prior model. A number of such models have been proposed [14]. Bouman *et al.* has proposed a multiscale MRF model where each level is causally dependent on the preceding coarser one [15]. On one hand, this scheme is not iterative thanks to the causal structure and exhibits a good robustness against the speckle noise effect [12, 13]. On the other hand, the neighborhood structure induced by the quadtree is not shift invariant and thus does not completely capture some aspects of local image properties because two spatially adjacent pixels can be actually "far" apart in the graph structure.
Herein, the proposed segmentation model consists of a label field pyramid associated to a single observation level. This model, introduced in [12, 13] combines specificities of a spatial and a causal model. More precisely, it involves local connections between spatially adjacent sites as well as parent-child connexion between sites belonging to consecutive levels. We will refer to it as Scale Causal Multigrid (SCM). From the inference point of view, a site s interacts with spatially adjacent sites at a given scale and with its "father" at the upper coarser level (i.e., a spatial and causal neighborhood in scale). Figure 2a depicts the pyramidal structure and the used interactions for the two-class segmentation issue.

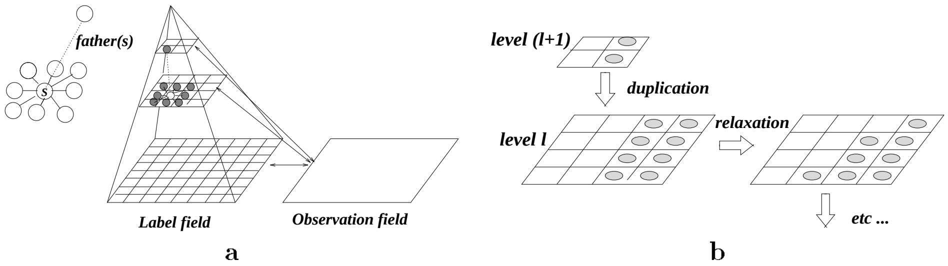


Figure 2: *a : Multiscale labeling structure with spatial and causal interactions involved in the two-class segmentation issue. b : "coarse-to-fine" minimization strategy for the two-class segmentation issue.*

In this model, the pyramidal structure is composed of a hierarchy $(x^L, \ldots, x^0)$ of label fields, where x^l is defined on grid S^l defined subsampling S by a factor 2^l in each direction ($S^0 \equiv S$ and $x^0 \equiv x$). $Y = \{Y_s, s \in S\}$ represents the field of observations located on the largest lattice S of N sites s (associated to N pixels of the sonar image), and $X = \{X_s, s \in S\}$ designates the label field at the finest resolution ($X^0 \equiv X$). Each of the Y_s takes its value in $\Lambda_{obs} = \{0, \ldots, 255\}$ and each X_s in $\{e_0 =$ shadow, $e_1 =$ reverberation$\}$. The segmentation of sonar images in two classes (shadow/reverberation) is stated as a causal statistical labeling problem according to a global Bayesian formulation in which we search for x^l, the labeling at the resolution level l, such as [13]:

$$\hat{x}^L = \arg\max_{x^L} \Big\{ P_{X^L/Y}(x^L/y) \Big\} \quad (3)$$

$$l = (L-1), \ldots, 0 \qquad \hat{x}^l = \arg\max_{x^l} \Big\{ P_{X^l/Y, X^{l+1}}(x^l/y, \hat{x}^{l+1}) \Big\} \quad (4)$$

$$= \arg\max_{x^l} \Big\{ \exp -U^l(x^l, y, \hat{x}^{l+1}) \Big\} \quad (5)$$

$$= \arg\min_{x^l} U^l(x^l, y, \hat{x}^{l+1}) \quad (6)$$

The posterior distribution at level 0 involved in equation (4) can be written as $P_{X/Y,X^1}(x/y, \hat{x}^1) \propto P_{X/X^1}(x/x^1) P_{Y/X}(y/x)$. In this expression $P_{X/X^1}(x/\hat{x}^1)$ represents the prior model and $P_{Y/X}(y/x)$ is assumed to factor as $\Pi_s P_{Y_s/X_s}(y_s/x_s)$ with $P_{Y_s/X_s}(y_s/x_s)$ the noise distribution associated to each region (i.e., shadow and reverberation) of the input sonar image. At the finest resolution, the global energy function to be minimized can be written as follows [13]:

$$U(x, y, \hat{x}^1) = \underbrace{-\sum_{s \in S} \ln P_{Y_s/X_s}(y_s/x_s)}_{U_1(x,y)} + \underbrace{\sum_{<s,t> \subset S} \beta_{st}\Big(1 - \delta\big(x_s, x_t\big)\Big)}_{U_2(x)} + \underbrace{\sum_{s \in S} \beta_5 \Big(1 - \delta\big(x_s, \hat{x}^1_{\text{father}(s)}\big)\Big)}_{U_3(x,\hat{x}^1)}$$

where $\delta(.)$ is the Kronecker delta function and $\beta_{st} = \beta_1, \beta_2, \beta_3$ or β_4 only depends on the "orientation" of the spatial binary clique and β_5 designates the parameter associated to the inter-level parameter. U_1 expresses the adequacy between observations and labels, U_2 the energy of the spatial *a priori* model and U_3 the energy term expressing the scale causal relation with the segmentation map at the coarser resolution level. Model at level l is now defined in a multigrid way [8], as:

$$U^l(x^l, y, x^{l+1}) \triangleq U\big(\Psi_0^l(x^l), y, \Psi_1^{l+1}(x^{l+1})\big)$$

where Ψ_k^l is the duplication operator from S^l into S^k (for $k < l$). Finally, we let:

$$U^L(x^L, y) = U_1\big(\Psi_0^L(x^L), y\big) + U_2\big(\Psi^L(x^L)\big)$$

Energy of the SCM prior model (i.e., U_2 and U_3) depends on a parameter vector $\Phi_x = (\beta_1, \ldots, \beta_5)$ while noise model term U_1 depends on an other parameter vector Φ_y. In [16], we have shown that the Weibull PDF is an appropriate distribution to describe the luminance y within the reverberation and the shadow regions. In order to make this two-classes segmentation completely unsupervised, we can use the Iterative Conditional Estimation, introduced in [16], to estimate simultaneously the noise model and the SCM prior model parameters. We use the deterministic ICM algorithm [4] to cope with each of the minimization in (6). The final estimate $\hat{x}^{l+1}$ obtained at a given level is then interpolated as $\Psi_l^{l+1}(\hat{x}^{l+1})$ to be used as an initialization for the relaxation process at the next finer level as schown in Figure 2b.
This multiscale energy minimization approach has appeared to be fast, far less sensitive to local minima than other standard relaxation algorithms and particularly well suited for the type of energy function related to the two-class segmentation problem of very noisy sonar images. Figure 3 display examples of unsupervised two-class segmentation, exploiting parameters estimated with the ICE procedure for a Weibull distribution-based model. Experiments and comparisons, given in [16], with other prior models combined with other energy minimization methods, have proved that the proposed scheme exhibits a better robustness against the speckle noise effect (false alarms have been eliminated), and allows to extract a more accurate segmentation map.

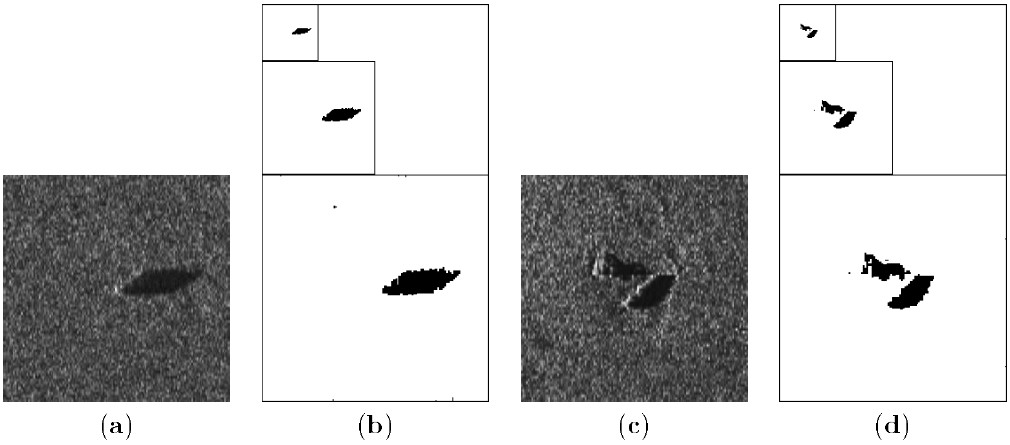


Figure 3: (**a and c**) Real sonar image of a sandy sea floor with the shadow of a man-made object or a rock. (**b and d**) SCM two-class segmentation results.

4.2. Three-class segmentation strategy

In order to ensure a proper detection of the echo information when the picture contains strong speckle noise, a solution consists in taking into account *a priori* information about the way echo physically appears. In sonar imagery, objects lying on the sea floor create a cast shadow and also an echo. This spatial dependency between these two areas can be incorporated in a MRF model.
Let $\hat{x}^{[1]}$ be the label field obtained after the two-class segmentation stage based on hierarchical Markovian modeling previously described. Label $\hat{x}_s^{[1]}$ belongs to $\{e_0, e_1\}$. Based on $x^{[1]}$, we now consider pixel subset S' ($S' \subset S$) such as $S' = \big\{s \in S : \hat{x}_s^{[1]} = e_1 = \text{reverberation}\big\}$. This set has now to be segmented into two classes in order to extract the echo signal from the sea-bottom reverberation one. Let $X^{[2]}$ the corresponding random binary process : $\forall s \in S', X_s^{[2]}$ takes values in $\{e_1 = \text{sea-bottom reverberation}, e_2 = \text{echo}\}$. The segmentation will make use of restricted data $y^{[1]} = \{y_s, s \in S'\}$. We define the distribution of $(X^{[2]}/Y^{[1]} = y^{[1]}, X^{[1]} = \hat{x}^{[1]})$. This last one is defined, firstly, by $P_{X^{[2]}/X^{[1]}}(x^{[2]}/\hat{x}^{[1]})$, the distribution of $X^{[2]}$ supposed to be stationary and Markovian, and secondly, by the site-wise data likelihoods $P_{Y_s^{[1]}/X_s^{[2]}}(y_s^{[1]}/x_s^{[2]})$, depending on the class label $x_s^{[2]}$. We now search for $x^{[2]}$ such as [16]:

$$\hat{x}^{[2]} = \arg\max_{x^{[2]}} P_{X^{[2]}/X^{[1]},Y^{[1]}}(x^{[2]}/\hat{x}^{[1]}, y^{[1]}) \quad (7)$$

$$= \arg\max_{x^{[2]}} P_{Y^{[1]}/X^{[1]},X^{[2]}}(y^{[1]}/\hat{x}^{[1]}, x^{[2]}) P_{X^{[2]}/X^{[1]}}(x^{[2]}/\hat{x}^{[1]}) \tag{8}$$

$$= \arg\min_{x^{[2]}} \Big\{ U_1(y^{[1]}, x^{[2]}, \hat{x}^{[1]}) + U_2(x^{[2]}, \hat{x}^{[1]}) \Big\} \tag{9}$$

where U_1 expresses the adequacy between observations and labels $\{e_1, e_2\}$ and U_2, related to the prior distribution $P_{X^{[2]}/X^{[1]}}(x^{[2]}/\hat{x}^{[1]})$, is the energy term corresponding to the *a priori* model. Parameters of the likelihood energy term U_1 can be estimated with the procedure given in [16]. U_2 has to express constraints on the desired solution. In our application, we adopt a 8-connexity spatial neighborhood in which β_1, β_2, β_3, β_4 represent the *a priori* parameters associated to the horizontal, vertical, right and left diagonal binary cliques respectively, and β_0 stands for the unary clique parameter.

• First, in order to favor homogeneous regions, we use a Potts model that associates to binary clique $<s,t>$, the potential $\beta\big(1 - \delta(x_s^{[2]}, x_t^{[2]})\big)$, where $\beta = \beta_1 = \beta_2 = \beta_3 = \beta_4$.

• Secondly, in order to express the dependency between the echo and shadows regions, we decide to defavour the choice of the echo label for a site that is too far away from a shadow region through the potential of the singleton clique : $-\beta_0 \ln \Psi_{\hat{x}^{[1]}}(s)\delta(x_s^{[2]}, e_2)$. In this expression, $\Psi_{\hat{x}^{[1]}}(s)$ is a potential field defined as follows [16]:

$$\Psi_{\hat{x}^{[1]}}(t) = \inf\Big\{ \sum_{s\in S:\hat{x}_s^{[1]}=e_0} \underbrace{\frac{1}{d(s,t)} \exp\big(-\frac{d(s,t)}{\sigma}\big)}_{\phi_s\left(d(s,t)\right)}, 1 \Big\} \tag{10}$$

where $d(s,t)$ is the distance between pixels s and t and σ is a standard deviation parameter controlling the interaction distance between echo and shadow regions. Finally, the global energy function to be minimized is defined as follows:

$$\begin{aligned} U(x^{[2]}, \hat{x}^{[1]}, y^{[1]}) = & \underbrace{- \sum_{s\in S'} \ln P_{Y_s^{[1]}/X_s^{[2]}}(y_s^{[1]}/x_s^{[2]})}_{U_1(y^{[1]}, x^{[2]}, \hat{x}^{[1]})} + \underbrace{\sum_{<s,t>\subset S'} \beta\Big(1 - \delta(x_s^{[2]}, x_t^{[2]})\Big)}_{U_{21}(x^{[2]}, \hat{x}^{[1]})} \\ & + \underbrace{\sum_{s\in S'} -\beta_0 \ln \Psi_{\hat{x}^{[1]}}(s)\cdot\delta(x_s^{[2]}, e_2)}_{U_{22}(x^{[1]}, x^{[2]}, \hat{x}^{[1]})} \end{aligned}$$

We use the deterministic relaxation algorithm ICM [4] to minimize this global energy function. For the initialization of this algorithm, we exploit the segmentation map obtained by a ML segmentation.
Due to the very small number of pixels associated to the echo region and the good initial guess given by the ML segmentation for the initialization process, the single-scale ICM minimization approach allows to find efficiently the echo area associated to each shadow region. Figures 4 and 5 display examples of three-class segmentation results on synthetic and real sonar images. Other results can be found in [16] that show the aptness of the *a priori* term expressing the spatial dependency between echo and shadow areas. This term allows to remove efficiently undesired echos induced by the speckle noise effects.

5. OBJECT CLASSIFICATION

In sonar imagery, the first goal of the classification step is to separate automatically natural objects from man-made objects. This low-level classification step in two classes can be based on the following a priori information: contrary to natural objects, a manufactured object is mainly composed of elements with simple and regular geometrical shape. For this reason, their cast shadow present geometric properties and exhibit straight lines as contours in the case of wrecks, pipe-lines, etc. In particular, the cast shadow of a cylindrical or of a cubic object is a perfect parallelogram. This global prior knowledge can be captured by a prototype template which is, in this case, a parallelogram characterized by its four vertices [3]. For a different application however, we have to detect spherical objects lying on the seabed. In this case, the associated cast shadow has a typical shape whose representation can easily be defined by a set of n points manually selected or equally sampled which approximate its outline. A cubic B-spline shape representation involving these n control points corresponding to "landmarks" is then defined. The outline of the spherical object cast shadow, used to build the prototype, can either be obtained from a real scene or synthetized with a ray tracing procedure (see Fig. 4.d).
Let γ_0 denote the prototype template for a class of objects. In order to take into account the variability of the considered object class, we introduce a set of admissible linear transformations on γ_0. Let γ_θ be a deformed version of the original prototype γ_0 according to a affine transformation with parameter vector θ. In the case of our first parametric template, used to detect manufactured objects, these deformations involve translation, scaling, rotation, stretching and skewing of the template as shown in Figure 6.a. Due to spherical symmetry, the cast shadow of a spherical object is symmetric relative to the sonar beam direction. Therefore, the only considered transformations are translation, scaling and stretching for this template (see Fig. 6.b). Such a scheme captures the global structure of a shape and can be exploited, in the same way, to detect manufactured object with another typical cast shadow shape.

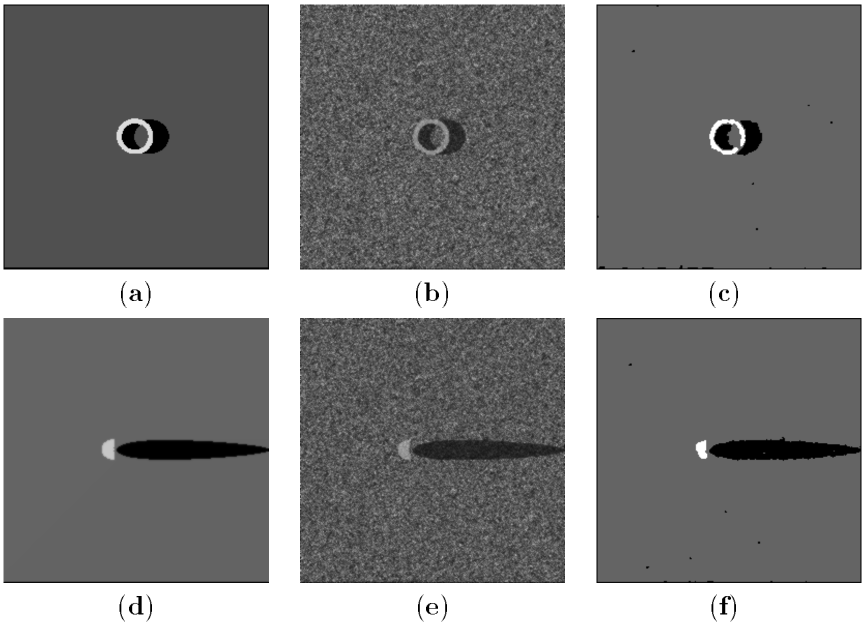


Figure 4: Synthetic shadow shape of a metallic core (**a**) and of a sphere (**d**) lying on a sea-bed obtained with a ray tracing procedure. (**b**) Synthetic sonar images, with synthetic speckle noise, of a metallic core (**b**) and a sphere (**e**) lying on a sandy sea-bed. (**c**)(**f**) Three-class segmentation results obtained with the proposed scheme.

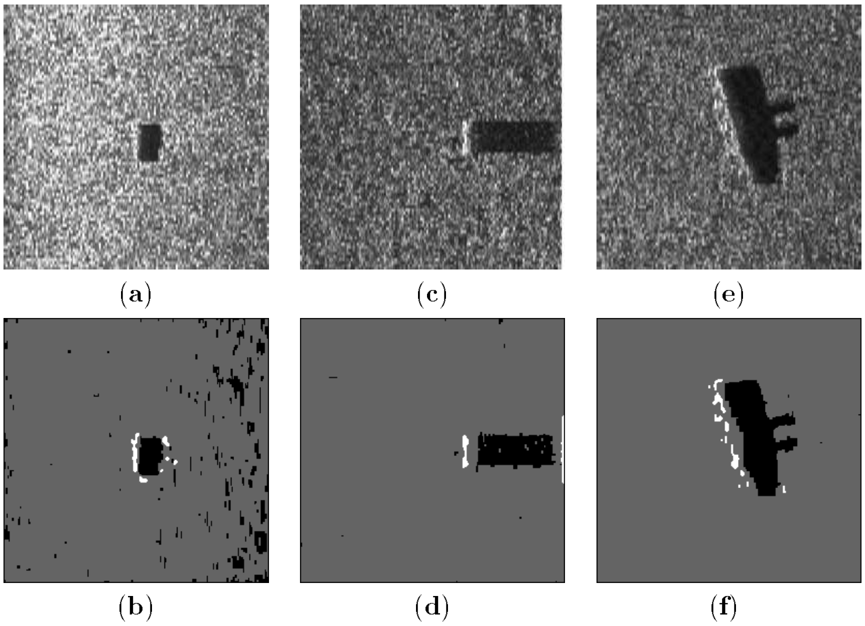


Figure 5: (**a**) (**c**) (**e**) Real sonar images involving a sandy sea floor and a man-made object (two cylinders and a trolley). (**b**) (**d**) (**f**) Three-class segmentation results obtained with the proposed scheme.

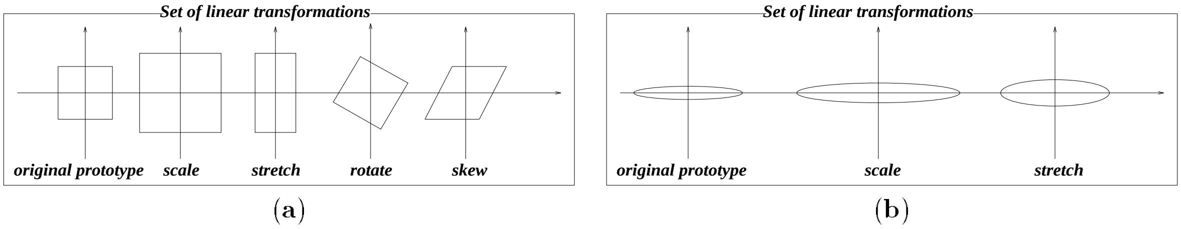


Figure 6: Considered linear transformations for the original prototype associated to (**a**) the cast shadow of a regular man-made object (**b**) the cast shadow of a spherical object.

The detection is based on an objective function ϵ measuring how well a given instance of deformed template γ_θ fits the content of the two-class segmented image x. From a probabilistic point of view, $\epsilon(\theta, x)$ defines the joint model through the Gibbs distribution [3, 17]:

$$P_{\Theta,X}(\theta, x) = \frac{1}{Z} \exp\big\{-\epsilon(\theta, x)\big\} \tag{11}$$

where Θ is the random vector of parameters, and Z a normalizing constant. The energy function $\epsilon(\theta, x)$ is composed of two terms :

$$\epsilon(\theta, x) = \underbrace{-\ln\Big\{\frac{1}{N_{\gamma_\theta}} \sum_{s\in\gamma_\theta} \Phi_{S'}(s)\Big\}}_{\epsilon_c(\theta,x)} \underbrace{-\ln\Big\{\frac{1}{N_{\gamma_\theta^\bullet}} \sum_{s\in\gamma_\theta^\bullet} \delta(x_s, e_0)\Big\}}_{\epsilon_r(\theta,x)} \tag{12}$$

where ϵ_c is the edge energy term that constraints the deformable template to be attracted and aligned to the salient contours of each object. To this end, $\Phi_{S'}$ is an edge potential field whose form is given in (10) and in which the summation is now over all the set of edges associated to detected cast shadows in the two-class segmented image x. This potential field induces a smooth version of the edge image in which a site close to an edge will get a potential value close to 1. Finally, the summation is over all the N_{γ_θ} pixels on the deformed template γ_θ. ϵ_r is the region-based energy term that aims at placing the inside of the deformed template in a region classified as shadow by the segmentation procedure. $\gamma_\theta^\bullet$ and $N_{\gamma_\theta^\bullet}$ represent respectively the set of pixels and the number of pixels inside the region delimited by γ_θ. We formulate the detection problem as the search of the *Maximum A Posteriori* (MAP) estimation of θ:

$$\hat{\theta}_{\text{MAP}} = \arg\max_\theta \big\{P_{\Theta/X}(\theta/x)\big\} = \arg\max_\theta \Big\{\frac{1}{Z_x}\exp\big(-\epsilon(\theta, x)\big)\Big\} = \arg\min_\theta \epsilon(\theta, x) \tag{13}$$

where Z_x is a normalizing constant depending on x only. ϵ is the global energy function to be minimized. This function is minimal when the deformed template exactly coincides with the edges of x and contains only pixels labeled as shadow. For the classification step, the resulting value of energy $\epsilon(\hat{\theta}_{\text{MAP}}, x)$ is used to measure the degree of fitness of the template with the region of interest, and then, to determine whether the extracted "object" actually belongs to the class of interest. If $\epsilon(\hat{\theta}_{\text{MAP}}, x)$ is lower than a given threshold, then the desired object is assumed to be present and the final configuration of the deformed template reveals the shape and the location of the detected object; otherwise, we decide that the desired object is not present. Due to the used detection and classification strategy, deterministic relaxation techniques such as gradient-based method cannot be used to minimize this energy function: this method requires a good initialization of the template, near the true location of the cast shadow shape to be detected. Otherwise, we obtain a sub-optimal solution (local minima) and the value of the resulting objective function ϵ cannot be exploited to affirm the presence of the desired object [16].

Stochastic methods based on Simulated Annealing (SA) [18] have the capability of avoiding local minima and no human interaction is required to initialize the model. However, one of the major drawbacks of this procedure is its high computational load [16, 19]. An alternate approach, adopted within the context of this application, consists in using a genetic exploration of the parameter space [16]. This global optimization procedure which mimics the evolution of natural systems [20] seems appropriate to minimize this L-dimensional ($L = 6$ or $L = 4$ depending on the considered deformable template) continuous function : a global minimum is found 25 times faster than with simulated annealing [16]. Figures 7, 8, 9 show a few examples of classification results from our database by setting the threshold value (empirically chosen after a set of experiments) at 0.2. Geometric shape, i.e., manufactured objects are well detected such as the geometric part of a wreck (Fig. 7.e), a section of a pipe-line (Fig. 7.d), a trolley (Fig 7.f) and different cylindrical objects (Fig. 7.a 7.b 7.c). In the same way, spherical objects (Fig. 8) are well detected even if the cast shadow is partially occluded. Figure 9 shows several natural objects and the associated values of the objective function ϵ.

6. SEA-FLOOR CLASSIFICATION

Another application of the two-classes segmented image x is the automatic segmentation and classification of the sea-bottom. Segmentation of sea-floor aims at partitioning the acoustic image into homogeneous regions with respect to

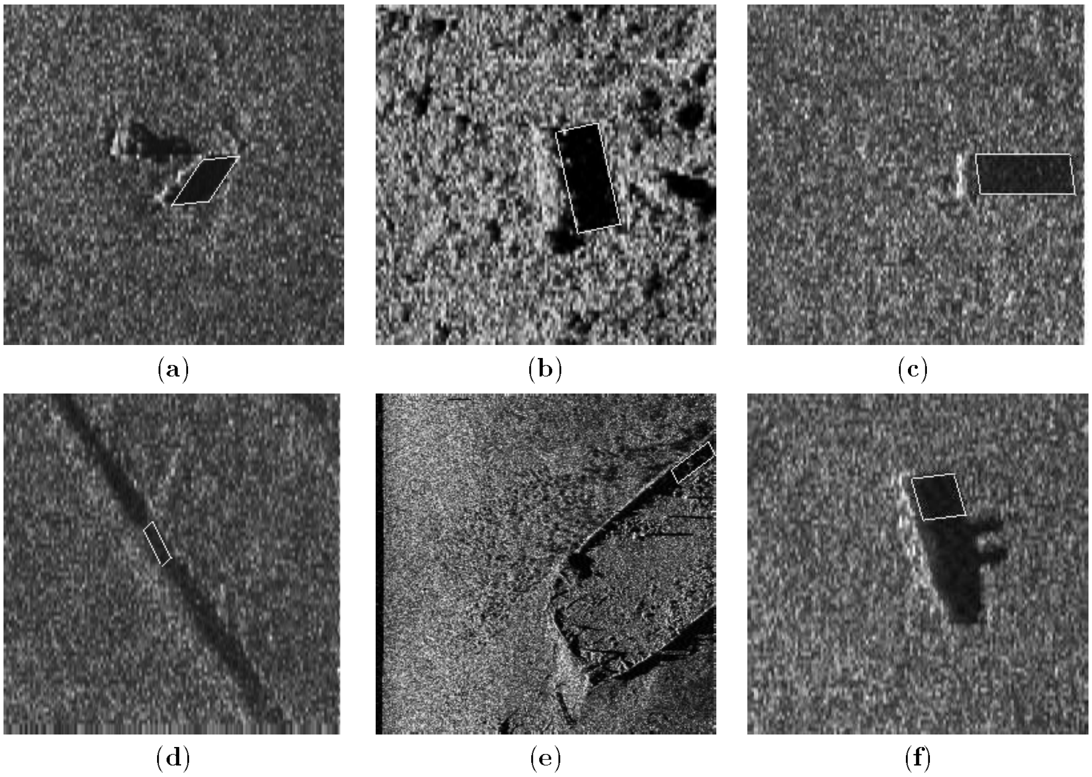


Figure 7: The parallelogram template and the low value obtained (with the genetic search) for ϵ ($\epsilon < 0.2$) allow us to classify these shadows as corresponding to manufactured objects. (**a-b-c**) Different cylindrical objects. ϵ=0.17, ϵ=0.15, and ϵ=0.14 respectively. (**d**) pipe-line: ϵ=0.12. (**e**) wreck: ϵ=0.15. (**f**) trolley: ϵ=0.2.

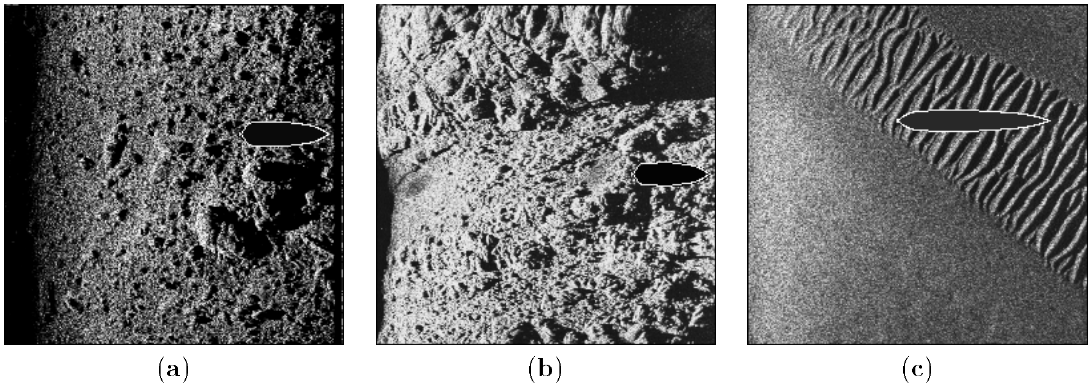


Figure 8: The cubic B-spline template and the low value obtained (with the genetic search) for ϵ ($\epsilon < 0.2$) allow us to classify these shadows as corresponding to spherical objects. (**a**) $\epsilon = 0.04$. (**b**) $\epsilon = 0.15$. (**c**) $\epsilon = 0.19$.

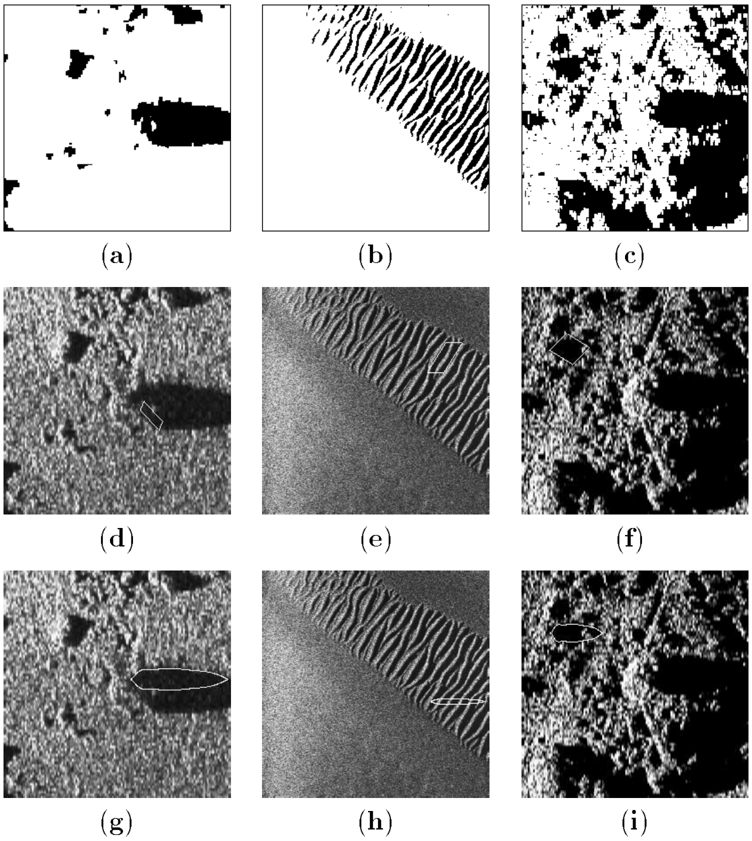


Figure 9: The values of ϵ (>0.2) (obtained with the genetic search) for both templates are sufficiently large so that we can reject the hypothesis of a manufactured or spherical object present in these images. (**a-c**) Markovian segmentation of the sonar images. (**d**) ϵ=0.40. (**e**) ϵ=0.41. (**f**) ϵ=0.24. (**g**) ϵ=0.55. (**h**) ϵ=0.57. (**i**) ϵ=0.38.

certain geological characteristics. The goal of the classification task is to assign these different geoacoustic regions to sea-floor classes such as $\{w_0 = \text{sand}, w_1 = \text{ridges}, w_2 = \text{dunes}, w_3 = \text{pebbles}, w_4 = \text{rocks}\}(= \Lambda)$.

Herein, the proposed method [21] adopts a pattern recognition approach, based on the identification of the cast shadow patterns for each sea-bottom type. To this end, x is high-pass filtered in order to extract the contour of each detected cast shadow. Then, the resulting edge image is partitioned in small sub-images from which feature vectors are extracted. The aim of this feature extraction process is to simplify the representation of the shadow patterns created by the different sea-floor types, and to cope with a relevant parameter vector giving maximal information about the extracted shadow profiles. For each sub-window, we compute the mean compactness, the directivity of the detected cast shadow, the maximal elongation cast shadow and the maximal size cast shadow outline contained in this sub-image [21]. Once these four parameters have been computed, they make up the feature vector v that will be exploited within a fuzzy classification method modeling our *a priori* knowledge about the cast shadow patterns created by each sea-bottom type [21]. This *a priori* knowledge model the fact that the cast shadows of ridges or dunes are geometrically elongated and parallel, the ones of pebbles and stones corresponds to non geometrical elements with random orientation, etc. [21]. Fuzzy formalism gives us, for each feature vector v_k computed on sub-window k, a degree of membership $\mu_{w_i}(v_k)$ $(0 \leq \mu_{w_i}(.) \leq 1)$ for each class w_i. This membership function allows to express the interaction between the sea-floor class and the observed data.

In order to obtain an accurate segmentation-and-classification map, contextual informations, i.e., relations between feature measures computed on adjacent sub-windows, are taken into account and we resort, once more, to MRF models. This allows us both to take into account the fuzzy classification results and to express constraints on the desired solution such as homogeneity properties of the expected segmentation map. We consider a couple of random fields (V, W), with $V = \{V_s, s \in \mathcal{S}\}$, the field of observations (v_s the feature vector computed on the sub-window s is considered as a realization of V_s), and $W = \{W_s, s \in \mathcal{S}\}$ the label field, located on a lattice $\mathcal{S}$ of N sub-windows. Each of the W_s takes its value in Λ. The distribution of W, $P_W(w)$, is supposed to be stationary and Markovian. The determination of segmentation map w is stated as a statistical contextual labeling problem, according to a global Bayesian formulation based on the MAP criterion:

$$\hat{w} = \arg\max_{w} P_{W/V}(w/v) = \arg\max_{w}\big\{P_W(w)\ P_{V/W}(v/w)\big\} = \arg\min_{w} U_1(w,v) + U_2(w) \qquad (14)$$

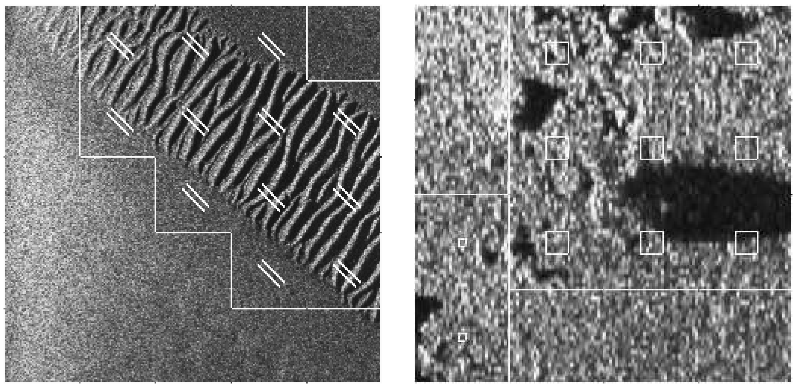
where $U_1(w,v)$ is the "feature-driven" term that expresses the adequacy between observations and labels, and $U_2(w)$ is the regularization energy term corresponding to the *a priori* model.
• First, in order to favor homogeneous regions, we use the same neighborhood and potentials as in Section 4.2.
• Secondly, in order to express the adequacy between our *a priori* knowledge about the cast shadows for each sea-bottom type and labels, we exploit the result obtained by the fuzzy pre-classification step and the energy term $U_1(w,v)$ is defined as follows:

$$U_1(w,v) = -\sum_{s\in S} \gamma(w_s, v_s) \tag{15}$$

where $\gamma(.)$ is a function that returns $\mu_{w_s}(v_s)$ if $w_s = \arg\max_{w_j}\big(\mu_{w_j}(.)\big)$ $(0 \le j \le 4)$ and 0 otherwise. Finally, the global energy function to be minimized is defined as follows:

$$U(w,v) \;=\; \underbrace{-\sum_{s\in S} \gamma(w_s, v_s)}_{U_1(w,v)} \;+\; \underbrace{\sum_{<s,t>} \beta_{st}\big(1-\delta(w_s,w_t)\big)}_{U_2(w)}$$

We use the deterministic relaxation algorithm ICM [4] to minimize this function. For the initialization of this algorithm, we employ the classification map obtained by the blind fuzzy classification scheme. Due to the initial guess given by the blind classification scheme, the monoscale ICM approach allows to find a global minimum of this energy function and consequently to obtain a reliable and accurate segmentation map.
Figure 10 represents original sea-floor images and classification results obtained with our scheme with the following display convention superimposed on the original sonar image. An empty square for the class "sand", and respectively a square with a small square inside it for the class "pebbles", a bigger square for the class "rocks", a straight line for the class "ridges" and two parallel lines for the class "dunes". Other results can be found in [21].



(a) (b)

Figure 10: (a) Original sidescan sonar image of an area composed by sand (with for the display convention, an empty square) and dune of sand (with for the display convention a square with two parallel lines) (b) Original sidescan sonar image of an area composed by sand, pebbles (display convention: a square with a small square inside it), rocks (display convention: a square with a bigger square inside it). Classification results showing detection of these three classes.

7. BAYESIAN PROCESSING CHAIN FOR AUTOMATIC OBJECT CLASSIFICATION

These segmentation and classification schemes can be used separately for a specific application. Nevertheless, in the context of the automatic classification of objects lying on the sea-floors, these different Bayesian methods can be exploited together in order to construct a Bayesian processing chain. Sea-bed classification information can be exploited to know if the cast shadows to be detected could be partially occulted in the case of objects lying on a complex backgrounds, i.e., sea-floor with pebbles, rocks, ridges or dunes of sand. In these cases, occlusions of the cast shadows to be detected as cast shadow of morphological elements (such as ridges or pebbles) make necessary to set the threshold value at a higher value. Moreover, echo information can be exploited in order to confirm or infirm the presence of a manufactured object.

8. CONCLUSION

In this paper, we have adressed different detection and classification problems in sonar imagery within the Bayesian framework. In this framework, the available prior information is expressed either by local or global prior models. In case

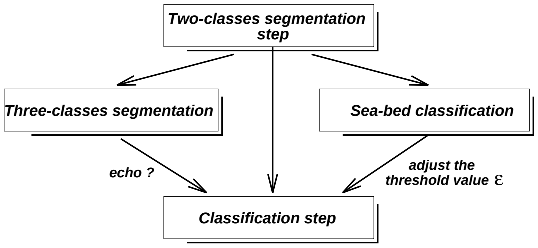


Figure 11: Bayesian processing chain for the automatic classification of object lying on the sea-bed.

of a local prior model, *a priori* information can be efficiently modeled at different levels of knowledge, at pixel level or at a higher level of representation, particularly at region level. The data models exploit either the grey level image *via* a noise model mixture estimation, the segmented image or some geometrical features extracted from it. Finally, the detection and classification tasks are turned into cost minimization problems. Different optimization strategies have been designed, depending on the type of energy function to be minimized. They are either deterministic (monoscale ICM, multiscale) or stochastic (GA). The different segmentation and classification methods can be combined in an original Bayesian processing chain for the automatic classification of objects lying on the sea-floors. This processing chain has been validated on a number of real sonar images demonstrating its efficiency and robustness and has proved useful for automatic processing of massive amounts of data.

References

[1] J.W. Goodman. Some fundamental properties of speckle. *Journal of Optical Society of America*, 66(11):1145–1150, Nov. 1976.

[2] B. Zerr and B. Stage. Three-dimensional reconstruction of underwater objects from a sequence of sonar images. In *Proc. 3rd International Conference on Image Processing*, pages 927–930, Lausanne, Sept. 1996.

[3] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Statistical model and genetic optimization : application to pattern detection in sonar images. In *Proc. International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2471–2475, Seattle, May 1998.

[4] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48:259–302, 1986.

[5] C. Regazzoni, F. Arduini, and G. Vernazza. A multilevel GMRF-based approach to image segmentation and restoration. *Signal Processing*, 34:43–67, 1993.

[6] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-12(7):609–628, july 1990.

[7] P. B. Chou and C. M. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

[8] F. Heitz, P. Pérez, and P. Bouthemy. Multiscale minimisation of global energy functions in some visual recovery problems. *CVGIP : Image Understanding*, 59(1):125–134, Jan. 1994.

[9] J.W Modestino and J. Zhang. A Markov random field Model-based approach to image interpretation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-14(6):601–615, 1992.

[10] A.K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(3):267–278, March 1996.

[11] M.P. Dubuisson Jolly, S. Lakshmanan, and A.K Jain. Vehicle segmentation and classification using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(3):293–308, March 1996.

[12] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Unsupervised hierarchical Markovian segmentation of sonar images. In *Proc. ICIP*, Santa Barbara, Californie, USA, Oct. 1997.

[13] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Trans. on Image Processing (accepted for publication (September 1998), to appear).*

[14] C. Collet, P. Thourel, P. Pérez, and P. Bouthemy. Hierarchical MRF modeling for sonar picture segmentation. In *International Conference on Image Processing - ICIP'96*, volume 3, pages 979–982, Lausanne, September 1996.

[15] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2):162–177, March 1994.

[16] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Three-class Markovian segmentation of high resolution sonar images. *Computer Vision and Image Understanding (submitted).*

[17] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence (submitted).*

[18] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, Nov. 1984.

[19] C. Kervrann and F. Heitz. Statistical model-based segmentation of deformable motion. In *Proc. International Conference on Image Processing*, pages 937–940, Lausanne, 1996.

[20] D.E Goldberg. *Genetic Algorithm.* Addison Wesley, 1989.

[21] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Seafloor classification using Markov random field and fuzzy logic. volume Mathematical Modeling, Bayesian Estimation and Inverse Problem (SD88), page submitted, SPIE's International Conference, Denver, Colorado, USA, July 1999.