Sonar Image Segmentation Using an Unsupervised Hierarchical MRF Model

Max Mignotte, Christophe Collet, Patrick Pérez, and Patrick Bouthemy

***Abstract*—This paper is concerned with hierarchical Markov random field (MRF) models and their application to sonar image segmentation. We present an original hierarchical segmentation procedure devoted to images given by a high-resolution sonar. The sonar image is segmented into two kinds of regions: shadow (corresponding to a lack of acoustic reverberation behind each object lying on the sea-bed) and sea-bottom reverberation. The proposed unsupervised scheme takes into account the variety of the laws in the distribution mixture of a sonar image, and it estimates both the parameters of noise distributions and the parameters of the Markovian prior. For the estimation step, we use an iterative technique which combines a maximum likelihood approach (for noise model parameters) with a least-squares method (for MRF-based prior). In order to model more precisely the local and global characteristics of image content at different scales, we introduce a hierarchical model involving a pyramidal label field. It combines coarse-to-fine causal interactions with a spatial neighborhood structure. This new method of segmentation, called scale causal multigrid (SCM) algorithm, has been successfully applied to real sonar images and seems to be well suited to the segmentation of very noisy images. The experiments reported in this paper demonstrate that the discussed method performs better than other hierarchical schemes for sonar image segmentation.**

***Index Terms*—Hierarchical MRF, parameter estimation, sonar imagery, unsupervised segmentation.**

I. Introduction

In high-resolution sonar imagery, three kinds of regions can be visualized: *echo*, *shadow*, and *sea-bottom reverberation*. The echo is caused by the reflection of the acoustic wave on an object while the shadow zone corresponds to a lack of acoustic reverberation behind this object. The remaining information constitutes so-called sea-bottom reverberation. On the pictures given by a classification sonar [1], the echo features are generally less discriminant than the shadow shapes for the classification of objects lying on the sea-bed. For this reason, the detection of each object located on the sea-bottom and its classification (as wreck, rock, man-made object, etc.) are generally based on the extraction and the identification of its associated cast shadow [2]. Thus, before any classification step, one must segment the sonar image in terms of shadow areas and sea-bottom reverberation areas.

Manuscript received November 18, 1997; revised August 20, 1999. This work was supported in part by Direction des Recherches (DRET), Etudes et Technique, French Ministry of Defense. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jeffrey J. Rodriguez.

M. Mignotte and C. Collet are with the Groupe de Traitement du Signal, Ecole Navale, Lanvéoc-Poulmic, France.

P. Pérez, and P. Bouthemy are with IRISA/INRIA, Rennes, France.

Publisher Item Identifier S 1057-7149(00)05669-4.

Few studies describe complete approaches allowing to perform automatically such a segmentation of sonar images, with results that can be efficiently used afterward for object identification. Some of them are based on simple, and often ad-hoc, clustering techniques (such as fuzzy k-means) working on luminance mean and variance within small windows [3], [4]. In that case, only a coarse grain classification is obtained, and no modeling of the luminance within the different types of regions is introduced. This latter aspect results in a lack of robustness for this type of methods. Similar techniques can however include more advanced elements of image formation modeling [5].

Unfortunately, this is not sufficient in general because the speckle noise present in sonar images affects any simple segmentation scheme. To cope with these problems, contextual information is important to be taken into account. This can be done *a posteriori*, using either morphological filters [3], [4], or MRF-based models [5]. As in many other low-level vision issues (especially classification and segmentation issues), it is known that the introduction of contextual dependencies is more proper and efficient if considered *a priori* within the modeling step. Such an approach is particularly well formulated within MRF-based modeling framework.

The advantages of this consistent statistical framework are already exploited in the context of sea-bottom reconstruction from range sonar data [6], [7]. For this task, which has received a larger attention than segmentation problem from sonar community, one copes with the fusion of high resolution side-scan sonar images with independent low resolution bathymetric data, to recover a three-dimensional (3-D) model of observed sea-floor.

As concerns the segmentation of sonar images, a MRF-based model is used in [5], but, as already said, this is done *a posteriori*, in order to "clean" the binary classification obtained beforehand by a simple clustering technique.[1] By contrast, we think that MRF framework should be used to capture in a joint and versatile way both a precise modeling of sonar luminance in the different regions and a regularizing prior on the regions to be recovered [8]. This is the type of approach that is investigated in [2]and [9], and that we further study in this paper, with a special emphasis on the additional problem of estimating the parameters in an automatic way.

The generic problem of *unsupervised* Markovian segmentation is quite complex, and remains a very active domain of research in the low-level vision community. The main difficulty

[1]Moreover, the classification method proposed in [5] is applied to images provided by a low frequency sonar, where the pixel resolution is around 400×400 m. We suspect that the proposed model is not well-suited to the high resolution sonar images we wish to segment for object classification purposes.

1057–7149/00$10.00 © 2000 IEEE

is that the estimation of parameters is required for the segmentation, while one or several segmentations are usually required for parameter estimation. To circumvent this difficulty, a scheme is proposed in [10] where the estimation and the segmentation are conducted alternatively. Although the method proved to converge in the case of independent Gaussian models, it is not clear that it can be extended to spatial MRF models. Besides, the method requires very heavy computations. The alternative approach we choose to solve the unsupervised MRF-based segmentation problem consists in having a two-step process. First, a parameter estimation step is conducted to infer both the noise model parameters and the MRF model parameters. Then, a second step is devoted to the segmentation itself based on the values of estimated parameters.

For the parameter estimation step, we adopt an iterative method called iterative conditional estimation (ICE) [11]–[13]. Similar to stochastic expectation-maximization (SEM) methods [14], we shall see that it offers however more flexibility in the choice of the actual techniques used for estimating each type of parameters. In particular, it allows to accomodate easily the "complete data" based estimators that seem the better suited to the different parts of the model.

The data model we introduced in [15] has proved successful in capturing the variety of the noise laws present in the distribution mixture of sonar images. The parameters involved by the different laws that compose this model will be estimated in the proposed scheme. As for the prior, we have introduced in [16] a novel hierarchical model based on both hierarchical and spatial Ising-type interactions. This prior has proven suitable for the purpose of segmenting sonar images with strong speckle noise. We show here that its parameters can be estimated within the ICE approach, and that it can used in conjonction with abovementionned data model in order to yield unsupervised sonar image segmentations of good quality.

This paper is organized as follows. In Section II, we detail the estimation step for the chosen data model combined, at this stage, with a standard spatial (anisotropic) Ising prior. The application of the estimation procedure is illustrated on both synthetic pictures and real sonar images. Section III presents the segmentation step and the proposed hierarchical prior model that is actually used. Experimental results obtained by the method on real scenes are reported in Section IV, along with comparisons to those obtained by other classical MRF-based approaches. Then, we conclude with some perspectives.

II. Parameter Estimation

A. Introduction

First, let us consider the estimation of the noise model parameters. To perform this task, a number of methods use the image histogram. Most of them (Fourier, polynomials, and cumulate histogram methods) are inefficient in case of important noise and cannot be used to estimate a mixture of not purely Gaussian laws [17].

Other techniques have been proposed to determine a maximum likelihood estimate of the noise model parameters from a given image. Expectation maximization (EM) or stochastic expectation maximization (SEM) algorithms can be used in the case of Gaussian distribution mixtures [14], [18]. However, in the specific case of sonar imagery, one has to deal with a mixture of laws that are not all Gaussian [1], and these techniques, like the former ones, are not well suited. By contrast, ICE technique is able to cope with the various non-Gaussian distributions that are relevant in this context, as experimentally demonstrated in [1].

Let us stress out that all the abovementionned approaches rely on the restrictive assumption that all the underlying "hidden" class labels are *independent*. When these hidden labels are considered as mutually interacting within a prior Gibbs distribution (or, equivalently, they are assumed to constitute a MRF), new parameters appear which are even more intricate to estimate. However, MRF modeling provides a powerful tool to incorporate an *a priori* knowledge about the spatial statistics of the label field, and it is thus worth the pain.

Using Markovian modeling, the prior knowledge about the "scene" is incorporated within an energy function which consists of a sum of appropriate local interaction potentials involving a few parameters. In many works that address segmentation and classification problems with this formalism, the parameters of the prior model are either assumed to be known, or determined in an ad-hoc fashion [6], [7], [19]–[22]. However, in our application, it is difficult to find manually appropriate values for the prior parameters since the real scenes may vary dramatically from one image to another (e.g., sea-floor with pebbles, dunes, ridges, sand, etc.). Thus, estimating these parameters in an automatic way is a crucial issue for the practical relevance of the labeling technique. One way to estimate these parameters from a given image is to use a simulated annealing-based method which alternates the estimation of parameters and the classification itself, for a number of iterations [23]. Unfortunaltely, this method is slow even with a first order isotropic model ruled by a single parameter. In [24], the algorithm is extended to Markovian prior. The resulting iterative scheme also requires a lot of computing time, and may get caught in local maxima without reaching a proper estimate. Also note that the EM formalism does not allow explicitely to estimate the parameters of standard (Ising or Potts) labeling priors. It only gives access to local conditional specifications which define imperfectly the prior. In [25], the authors propose to implement the estimation of the parameters associated with clique functions as a neural network whose weights are learned from examples by the error backpropagation algorithm. This method requires a learning step from a training data set, which can be problematic.

Considering the type of model we are going to deal with (non-Gaussian data likelihoods such as Weibull or Rayleigh laws, and Ising-type prior), we found more appropriate to use the ICE framework, which is more general and flexible [26], for the inference of parameters. In the following we first briefly review the ICE approach in case of generic modeling elements. We shall then specify the underlying estimators that we used in our context, and further describe how the proposed ICE procedure is initialized and led.

B. Iterative Conditional Estimation

We consider a couple of random fields $Z = (X, Y)$, with $Y = \{Y_s, s \in S\}$ the field of observations located on a lattice S

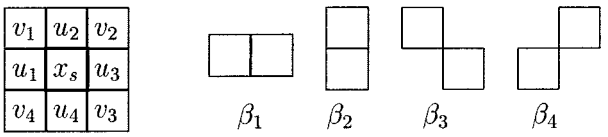

Fig. 1. Second-order neighborhood and associated labeling notations; the four types of two-site cliques and associated parameters in the anisotropic Ising model.

of N sites s, and $X = \{X_s, s \in S\}$ the label field. Each of the Y_s takes its value in $\Lambda_{obs} = \{0, \cdots, 255\}$ and each X_s in $\{e_0 = shadow, e_1 = sea - bottom\}$. The distribution of (X, Y) is defined, first, by $P_X(x)$, the distribution of X supposed to be stationary and Markovian, and secondly, by the site-wise conditional data likelihoods $P_{Y_s|X_s}(y_s|x_s)$, which depend on the concerned class label x_s. If the data are assumed to be independent *conditioned* on the labeling process X, one gets $P_{Y|X} = \Pi_{s\in S} P_{Y_s|X_s}$. We shall stick to this assumption throughout. The observable Y is called the *incomplete data* whereas Z constitutes the *complete data*. Prior distribution $P_X(x)$ depends on some parameter vector Φ_x, while data likelihood $P_{Y|X}(y|x)$ depends on another parameter vector Φ_y. Joint and posterior distributions $P_{X,Y}(x,y) = P_X(x)P_{Y|X}(y|x)$ and $P_{X|Y}(x|y) \propto P_X(x)P_{Y|X}(y|x)$ thus depend on $\Phi \triangleq (\Phi_x, \Phi_y)$. This dependence will be made explicit when necessary [e.g., denoting posterior distribution as $P_{X|Y,\Phi}(x|y,\Phi)$].

The single scale segmentation of sonar images in two classes can be viewed as a statistical labeling problem according to a global Bayesian formulation in which the posterior distribution $P_{X|Y}(x|y) \propto \exp\{-U(x,y)\}$ has to be maximized [27]. This is the maximum *a posteriori* (MAP) estimation. In the standard case of Ising-type prior [8], the corresponding posterior energy to be minimized is

$$U(x,y) = \sum_{s\in S} \Psi_s(x_s, y_s) + \sum_{\langle s,t\rangle} \beta_{st}[1 - \delta(x_s, x_t)] \quad (1)$$

where $\Psi_s(x_s, y_s) = -\ln\ P_{Y_s|X_s}(y_s|x_s)$, δ is the delta Kronecker function, and $\beta_{st} = \beta_1, \beta_2, \beta_3$ or β_4 depending on whether the pair of neighboring sites (relative to the second-order neighborhood system), or *clique*, $\langle s,t\rangle$ is horizontal, vertical, right diagonal, or left diagonal (see Fig. 1). In this energy setting, the first energy term expresses the adequacy between observations and labels, whereas the second one is related to the *a priori*.

To perform the unsupervised segmentation, we have to estimate the parameter vectors Φ_x and Φ_y. To this end, we resort, within a first step, to an iterative method of estimation called iterated conditional estimation (ICE) [11]. This method first requires to get two estimators $\hat{\Phi}_x(x)$ and $\hat{\Phi}_y(x,y)$ which provide respectively an estimate of Φ_x based on a configuration x, and an estimate of Φ_y based on a complete data configuration (x,y). Random field X being unobservable, the iterative ICE procedure defines parameter fits $\Phi_x^{[k+1]}$ and $\Phi_y^{[k+1]}$ at step $k+1$ as the conditional expectations of $\hat{\Phi}_x$ and $\hat{\Phi}_y$ given $Y = y$ and the current parameter fits $\Phi_x^{[k]}$ and $\Phi_y^{[k]}$. The fixed point of this iteration corresponds to the better approximations of Φ_x and Φ_y in terms of the mean squares error [13]. By denoting E_k the conditional expectation based on $\Phi^{[k]} = (\Phi_x^{[k]}, \Phi_y^{[k]})$, this iterative procedure is defined as follows:

- one takes an initial value $\Phi^{[0]} = (\Phi_x^{[0]}, \Phi_y^{[0]})$;
- $\Phi^{[k+1]}$ is computed from $\Phi^{[k]}$ and from y using

$$\Phi_x^{[k+1]} = E_k[\hat{\Phi}_x(X)|Y = y], \quad (2)$$
$$\Phi_y^{[k+1]} = E_k[\hat{\Phi}_y(X,Y)|Y = y]. \quad (3)$$

The computation of these expectations is impossible in practice, but we can approach them thanks to the law of large numbers

$$\Phi_x^{[k+1]} = \frac{1}{n}[\hat{\Phi}_x(x_{(1)}) + \cdots + \hat{\Phi}_x(x_{(n)})] \quad (4)$$
$$\Phi_y^{[k+1]} = \frac{1}{n}[\hat{\Phi}_y(x_{(1)}, y) + \cdots + \hat{\Phi}_y(x_{(n)}, y)] \quad (5)$$

where $x_{(i)}, i = 1, \cdots, n$ are realizations drawn from the posterior distribution $P_{X|Y,\Phi}(x|y,\Phi^{[k]})$. As it turns out, $n = 1$ is sometimes found sufficient (or even better) to get good estimates when convergence is reached, in case of stationary prior with low-dimensional parameter vector [11]. It is the case in our unsupervised classification model, and we actually chose $n = 1$ in our experiments.

By letting free the choice of estimators $\hat{\Phi}_x$ and $\hat{\Phi}_y$, the ICE procedure offers a great deal of flexibility which allows an efficient adaptation to the MRF model at hand. Contrary to EM-type approaches, the ICE allows to handle properly the estimation of Ising-type prior parameters, using the least-squares (LSQR) estimator [28] introduced by Derin *et al.* (see Section II-C2). As for $\hat{\Phi}_y(x,y)$, we used a maximum likelihood (ML) approach that proved well suited in our context where the speckle distribution in the sonar images is not exactly known, and may vary according to experimental conditions (see Section II-C1).

Finally, we need two other ingredients in order to use the ICE procedure, namely:

- a technique to get an initial value $\Phi^{[0]}$ which is not too far from the optimal parameters. To this end we use a clustering approach which is described in Section II-D.
- a way of simulating realizations from posterior distribution $P_{X|Y,\Phi}(x|y,\Phi)$. This is performed by using the Gibbs sampler [29].

C. Estimation of the Model Parameters for the Complete Data

1) Noise model parameters: The Gaussian law, $\mathcal{N}(\mu, \sigma^2)$, is a well-adapted degradation model to describe the luminance Y within shadow regions (where noise is essentially due to electronic aspects) [1]. We thus let

$$P_{Y_s|X_s,\Phi_y}(y_s|e_0,\Phi_y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_s-\mu)^2}{2\sigma^2}\right). \quad (6)$$

The corresponding ML estimator is defined by the empirical mean and the empirical variance. If $N_0 \triangleq \#\{s \in S\colon x_s = e_0\}$[2] pixels are located in the *shadow* areas, we have

$$\hat{\mu}(x,y) = \frac{1}{N_0} \sum_{s\in S: x_s = e_0} y_s,$$
$$\hat{\sigma}^2(x,y) = \frac{1}{N_0 - 1} \sum_{s\in S: x_s = e_0} (y_s - \hat{\mu})^2.$$

[2]Recall that e_0 stands for the "shadow" label while e_1 corresponds to "sea-bottom" class.

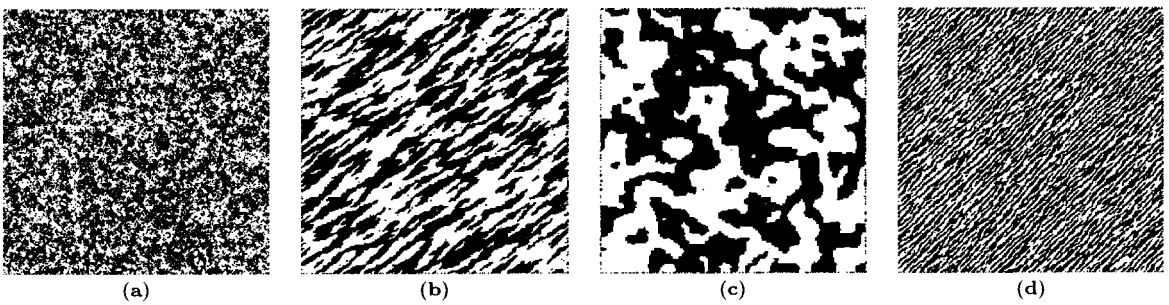

Fig. 2. Various realizations of the second-order anisotropic Ising model, corresponding to specified parameters reported in Table I.

In the reverberation areas, Weibull law provides an appropriate model for luminance to cope with speckle noise phenomenon [30]. This law implies a shape parameter that can be estimated from complete data. It turns out that estimated values are often close to 2. We thus restrict ourselves to Rayleigh law which exactly corresponds to 2 shape parameter [31]. More precisely, we model the conditional density function for the *sea-bottom* class by a shifted Rayleigh law $\mathcal{R}\,(min, \alpha^2)$

$$P_{Y_s|X_s,\Phi_y}(y_s|e_1,\Phi_y) = \frac{\min\{0, y_s - min\}}{\alpha^2} \cdot \exp\left(-\frac{(y_s - min)^2}{2\alpha^2}\right). \quad (7)$$

If $N_1 \triangleq \#\{s \in S\colon x_s = e_1\}$ is the number of pixels of the sea-bottom regions, we obtain the following expression of the ML estimator [1]:

$$\widehat{min}(x,y) \approx \min\{y_s\colon x_s = e_1\} - 1,$$
$$\hat{\alpha}^2(x,y) = \frac{1}{2N_1} \sum_{s\in S: x_s = e_1} (y_s - \widehat{min})^2.$$

If required [cf. Fig. 4(a)], proportions $\pi_0 \triangleq P_{X_s}(e_0)$ and $\pi_1 \triangleq P_{X_s}(e_1) = 1 - \pi_0$ of the mixture of laws (independent from s) can be approximated by the empirical frequencies, i.e., $\pi_i \approx N_i/|S|$.

2) A priori model parameters: At this stage, we consider the standard prior model introduced in (1). It is a second-order anisotropic[3] Ising model ruled by four parameters $\Phi_x = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ (see Fig. 1). The corresponding local conditional specification can be written down as follows. If s is some site and ν_s denotes its neighborhood, the probability that $X_s = x_s$ given some configuration $X_{\nu_s} = (u_1, u_2, u_3, u_4, v_1, v_2, v_3, v_4)^T \triangleq \eta$ on the neighborhood (refer to Fig. 1) is

$$P_{X_s|X_{\nu_s},\Phi_x}(x_s|\eta,\Phi_x) \propto \exp\{-\Theta(x_s,\eta)^T\Phi_x\} \quad (8)$$

where

$$\Theta(x_s,\eta) \triangleq [\mathcal{I}(x_s,u_1) + \mathcal{I}(x_s,u_3), \mathcal{I}(x_s,u_2) + \mathcal{I}(x_s,u_4), \mathcal{I}(x_s,v_1) + \mathcal{I}(x_s,v_3), \mathcal{I}(x_s,v_2) + \mathcal{I}(x_s,v_4)]^T$$

[3]The concept of anisotropy we refer to throughout the paper is related to the dependence of local interactions on the direction modulo 2π, that is only on the orientation.

with $\mathcal{I} = 1 - \delta$. Considering the following expression for the two possible values of X_s, $X_s = e_0$ (*shadow* label) and $X_s = e_1$ (*sea-bottom* label), for a same neighborhood configuration $X_{\nu_s} = \eta$, one gets

$$\ln \frac{P_{X_s|X_{\nu_s}}(e_1|\eta)}{P_{X_s|X_{\nu_s}}(e_0|\eta)} = \ln \frac{P_{X_s,X_{\nu_s}}(e_1,\eta)}{P_{X_s,X_{\nu_s}}(e_0,\eta)} = [\Theta(e_0,\eta) - \Theta(e_1,\eta)]^T\Phi_x \quad (9)$$

where Φ_x is the unknown parameter vector to be estimated. The model being stationary (shift-invariant), the second ratio can be approximated for each possible neighborhood configuration η, using simple histograming:

$$\frac{P_{X_s,X_{\nu_s}}(e_1,\eta)}{P_{X_s,X_{\nu_s}}(e_0,\eta)} \approx \frac{\#\{s \in S\colon x_s = e_1, x_{\nu_s} = \eta\}}{\#\{s \in S\colon x_s = e_0, x_{\nu_s} = \eta\}}.$$

By substituting these empirical ratio estimates in (9), we obtain $2^8 = 256$ linear equations in four unknowns. A given combination (e_i, η) may of course not occur at all in the label fields. In this case, we cannot obtain linear equation (9) because of the logarithm. Moreover, neighborhood configurations such that $\Theta(e_0,\eta) = \Theta(e_1,\eta)$ (e.g., $\Theta(e_0,\eta) - \Theta(e_1,\eta) = 0$ if $u_1 \neq u_3$, $u_2 \neq u_4$, $v_1 \neq v_3$, and $v_2 \neq v_4$) imply equations of type $0 =$ constant. Therefore we simply ignore these cases. The over-determined linear system of equations thus obtained is solved with the least-squares method.

This method for estimating Ising-type (and Potts-type) clique parameters from histograming has been proposed by Derin *et al.* [28]. It is not iterative and provides estimated parameters that are optimal in the least-squares sense.

We now present some results of this parameter estimation procedure, obtained on various realizations of the second-order Ising model, for different parameters Φ_x (see Fig. 2). It is a complete data problem, and the estimates provided by LSQR estimator $\hat{\Phi}_x$ are compared to real parameters in Table I. These few examples illustrate the accuracy that can be achieved by the method in case of observed data.

Synthetic textures presented in Fig. 2 also show that the Ising model with its four parameters offers an interesting variety of priors to capture different types of sea-floor: Fig. 2(a) shows inhomogeneous shapes reminiscent to segmentation of a pebbly sea-bottom; Fig. 2(b) exhibits homogeneous oriented shapes that are representative of segmented dunes; geometric shapes within Fig. 2(c) recall the shadows of manufactured objects; Fig. 2(d) looks like the segmentation of an image containing ridges of sand.

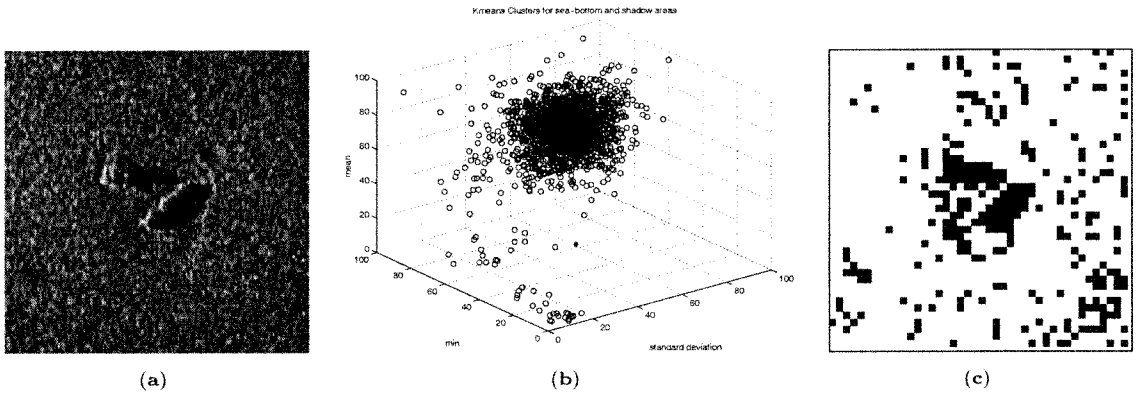

Fig. 3. (a) Real 256×256 sonar picture (object and rock shadows on a sandy sea-floor), (b) representation of the three-component samples computed from a partitioning of the original image into 1764 windows of size 6×6, and (c) block-wise segmentation of the original image corresponding to the clustering of sample by the k-means procedure.

TABLE I
SPECIFIED *a priori* PARAMETERS Φ_x FOR THE ANISOTROPIC SECOND-ORDER ISING MODEL, AND ESTIMATED PARAMETERS $\hat{\Phi}_x$ OBTAINED FROM REALIZATIONS

Fig.		β_1	β_2	β_3	β_4
2(a)	Φ_x	0.3	0.3	0.3	0.3
	$\hat{\Phi}_x$	0.3	0.29	0.31	0.31
2(b)	Φ_x	2	1	−1	1
	$\hat{\Phi}_x$	1.77	0.94	−0.86	0.98
2(c)	Φ_x	2	2	0	0
	$\hat{\Phi}_x$	1.93	1.94	−0.05	0.01
2(d)	Φ_x	1	1	−1	1
	$\hat{\Phi}_x$	0.95	0.93	−0.96	1.03

D. Initialization

The initial parameter values have a significant impact on the rapidity of the convergence of the ICE procedure and on the quality of the final estimates. In our application, we use the following method. The initial parameter values of the noise model $\Phi_y^{[0]}$ are determined by applying a small nonoverlapping sliding window over the image and by computing the empirical mean and variance of luminance, as well as the minimum grey level, for each location of the window. Each window location thus provides a three-component "sample" $\mathbf{x}_m$. The collected samples $\{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$ are then clustered into two classes $\{e_0, e_1\}$ using the k-means clustering procedure [32]. This algorithm uses a similarity measure that is the Euclidean distance between samples. For a given assignment of the M samples among the K clusters, an objective function J is then defined by

$$J = \sum_{i=1}^{K} \sum_{\mathbf{x}_m \in C_i} \|\mathbf{x}_m - c_i\|^2$$

where the second sum is over all samples in the ith cluster with center c_i. It is easy to show that for a given set of samples and of class assignments, J is minimized by choosing c_i to be the average sample of the ith cluster. For given centers c_i, J is minimized by assigning $\mathbf{x}_m$ to the cluster whose center is the closest. A number of other criteria are given in [32]. The complete algorithm is as follows.

1) Choose K initial cluster centers $c_1^{[1]}, \cdots, c_K^{[1]}$. They could be picked arbitrarily, but are usually defined by
$$c_i^{[1]} = \mathbf{x}_i \qquad 1 \le i \le K.$$
2) At the kth step, assign sample $\mathbf{x}_m$, $1 \le m \le M$, to cluster i if
$$\|\mathbf{x}_m - c_i^{[k]}\| < \|\mathbf{x}_m - c_j^{[k]}\| \quad \forall j \ne i.$$
Each sample is reassigned to the cluster with the nearest center. In the case of ties, the assignment is arbitrarily chosen among the competing clusters.
3) Let $C_i^{[k]}$ denote the ith cluster after Step 2. Determine new cluster centers by
$$c_i^{[k+1]} = \frac{1}{N_i} \sum_{\mathbf{x}_m \in C_i^{[k]}} \mathbf{x}_m$$
where N_i is the number of samples in $C_i^{[k]}$. Thus, the new cluster center is the mean of the samples in the cluster.
4) Repeat until convergence is achieved ($c_i^{[k+1]} = c_i^{[k]}, \forall i$).

Although it is possible to find pathological cases where convergence never occurs [32], the algorithm does converge in all tested examples. The rapidity of convergence depends on the number of clusters, the choice of initial cluster centers, and the order in which the samples are considered. In our application, $K = 2$.

One has to find a good compromise for the size of the window used in the constitution of samples. On the one hand, a small window increases the accuracy of the segmentation, hence providing a more precise estimation of the distribution mixture. On the other hand, decreasing the number of pixels on which the three attributes are computed may result in a higher misclassification error. In our application, satisfactory results are obtained with a 6×6 sliding window. Fig. 3 illustrates the k-means clustering procedure (and resulting segmentation) on a real sonar image.

Once the block-wise segmentation associated to final clustering is obtained, ML estimation is applied to it to get initial data parameter estimate $\Phi_y^{[0]}$. Based only on these parameters, a

first pixel-wise ML segmentation is readily obtained according to

$$\forall s \in S, \quad \hat{x}_s^{[0]} = \arg \max_{x_s} P_{Y_s|X_s,\Phi_y}(y_s|x_s, \Phi_y^{[0]}).$$

A first estimation of prior parameters can then be conducted on this segmentation, yielding $\Phi_x^{[0]} = \hat{\Phi}_x(\hat{x}^{[0]})$.

E. Parameter Estimation in the Incomplete Data Case

We can now summarize the complete parameter estimation procedure for the partially observed model defined as

$$P_{X|Y,\Phi}(\cdot|y,\Phi) \propto P_{X|\Phi_x}(\cdot|\Phi_x) \prod_{s\in S} P_{Y_s|X_s,\Phi_y}(y_s|\cdot,\Phi_y) \tag{10}$$

where $P_{X|\Phi_x}$ is the Gibbs distribution of an Ising-type MRF with parameters $\Phi_x = (\beta_1,\beta_2,\beta_3,\beta_4)^T$, $P_{Y_s|X_s,\Phi_y}(\cdot|c_0,\Phi_y)$ (within shadow area) is a Gaussian law with parameters (μ,σ^2), and $P_{Y_s|X_s,\Phi_y}(\cdot|c_1,\Phi_y)$ (within sea-bottom area) is a shifted Rayleigh law with parameters (min,α^2).

- **Parameter initialization.** As already explained, the initial parameter fit $\Phi^{[0]} = (\Phi_x^{[0]}, \Phi_y^{[0]})$ is derived from k-means clustering of image blocks.
- **ICE procedure.** $\Phi^{[k+1]}$ is computed from $\Phi^{[k]}$ in the following way.
  - Using the Gibbs sampler, n realizations $x_{(1)},\cdots,x_{(n)}$ are simulated according to the posterior distribution $P_{X|Y,\Phi}(\cdot|y,\Phi^{[k]})$.
  - For each $x_{(i)}$ $(i = 1\cdots n)$, the parameter vector Φ_x is estimated as $\hat{\Phi}_x(x_{(i)})$ by Derin *et al.* algorithm, and Φ_y is estimated as $\hat{\Phi}_x(x_{(i)},y)$ with the ML estimator of each class.
  - $\Phi^{[k+1]}$ is obtained by averaging $\{\hat{\Phi}_x(x_{(i)}), \hat{\Phi}_y(x_{(i)}, y)\}_{i=1\cdots n}$.

We designed a stopping criterion based on the variance of the estimators, which is empirically computed on the k_0 last parameter fits (where k_0 is a fixed parameter). When this indicator of the "stability" of the procedure falls below a given threshold, the sequence of $\Phi^{[k]}$ is assumed to have reached an equilibrium and the procedure is ended. One proceeds to the actual segmentation using the estimated parameters.

Note that, since shifted Rayleigh law (7) forbids luminances below $\min$, no pixel with luminance below the initial estimate of $\min$ would ever be classified as "sea-bottom" afterwards. As a consequence, any new estimate of $\min$ in the ICE procedure would be greater or equal to the initial estimate. In order to soften this constraint (i.e., to allow a possible decreasing of $\min$ estimate), we slightly modified $\mathcal{R}(\min,\alpha^2)$, such that it now associates a very low, but non-null, likelihood to pixels with a grey level lower than min [27].

We calibrate the weight of the "stochastic" aspect of the ICE by choosing n, the number of realizations drawn from the posterior distribution. When n increases, empirical averaging of estimator $\hat{\Phi}$ get closer to its posterior expectation, and each step of the ICE procedure becomes almost deterministic. The choice of a small value for n (e.g., $n = 1$) can increase the efficiency of the method by letting largely randomized each iteration [11]. This is what we observed in our experiments.

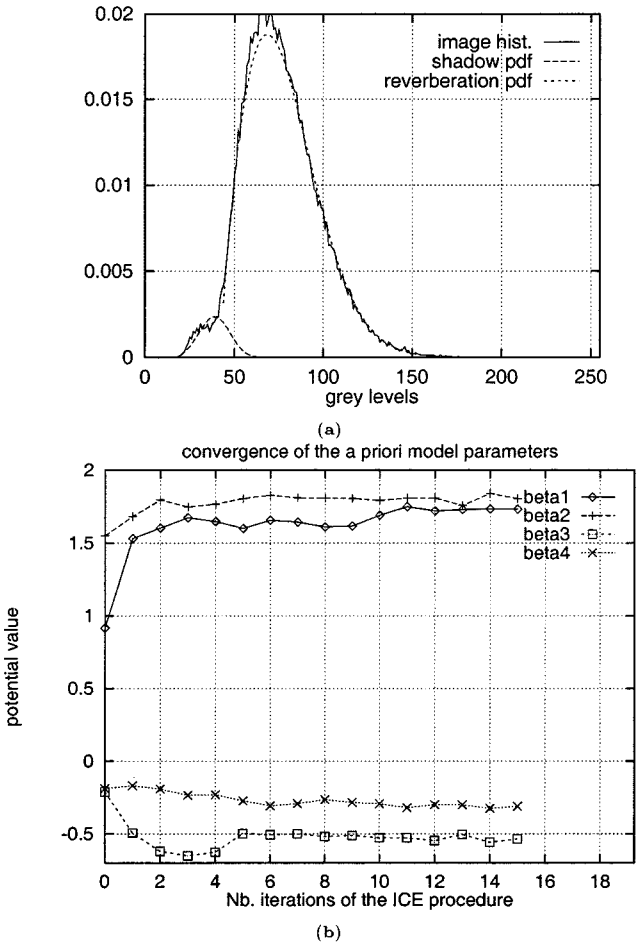

Fig. 4. (a) Image histogram of the sonar image from Fig. 3(a) and the two weighted mixture components that are estimated. (b) Evolution of the *a priori* parameter estimates $\hat{\Phi}_x^{[k]}$.

The quality of the parameter estimation is difficult to assess in absence of "true" values to compare with. As for the data model, we can roughly perform such an evaluation by comparing the grey level histogram of the image with the two components, $\hat{\pi}_0\mathcal{N}(\hat{\mu},\hat{\sigma}^2)$ and $\hat{\pi}_1\mathcal{R}(\widehat{min},\hat{\alpha}^2)$, of the estimated mixture weighted according to their estimated proportions. In Fig. 4(a) we propose such a visual assessment, by superimposing the two weighted densities estimated from a real sonar image, on the histogram of this image. One can appreciate the good fitting thus exhibited. This example illustrates the ability of the model to capture, thanks to the described ICE technique, the mixture of distributions involved in a real sonar image. The values of associated prior parameter estimates for that example are plotted against the number of iterations in Fig. 4(b). Although convergence is not theoretically established, we always observed it in practice. Stability is reached after around 15 iterations [requiring 60 s on a sonar picture of size 256 × 256 pels, with a 43P IBM Workstation (200 MHz)]. The original and final estimates for all parameters are given in Table II.

Now that we are able to estimate both the parameters of the mixture-based data modeling and those of the prior Ising modeling of the unknown labeling, we can turn our attention to the actual classification issue itself.

TABLE II
ESTIMATED PARAMETERS OF THE MIXTURE DATA MODEL AND OF THE SECOND-ORDER ANISOTROPIC PRIOR MODEL, FOR THE SONAR IMAGE IN FIG. 3(a)

	Initialization: k-means			
$\Phi^{[0]}_{y_{(shadow)}}$	$0.16_{(\pi_0)}$	$55_{(\mu)}$	$286_{(\sigma^2)}$	
$\Phi^{[0]}_{y_{(sea-bottom)}}$	$0.84_{(\pi_1)}$	$37_{(min)}$	$952_{(\alpha^2)}$	
$\Phi^{[0]}_{x}$	$0.9_{(\beta_1)}$	$1.6_{(\beta_2)}$	$-0.2_{(\beta_3)}$	$-0.2_{(\beta_4)}$
	ICE procedure (15 iterations)			
$\Phi^{*}_{y_{(shadow)}}$	$0.03_{(\pi_0)}$	$25_{(\mu)}$	$32_{(\sigma^2)}$	
$\Phi^{*}_{y_{(sea-bottom)}}$	$0.97_{(\pi_1)}$	$35_{(min)}$	$1424_{(\alpha^2)}$	
Φ^{*}_{x}	$1.7_{(\beta_1)}$	$1.8_{(\beta_2)}$	$-0.6_{(\beta_3)}$	$-0.3_{(\beta_4)}$

III. UNSUPERVISED HIERARCHICAL SEGMENTATION

A. Introduction

In order to capture more efficiently the larger scale characteristics of sonar image contents, the standard second-order Ising prior we have used so far, should be improved. This can be done by using a larger neighborhood structure, but this would increase dramatically the complexity of segmentation and parameter estimation procedures.

To circumvent this difficulty, different multiresolution schemes have been proposed. In [33], a tree-based hierarchical auto-regressive modeling is introduced to capture the multiscale structure of images that are similar, to some extent, to sonar images (namely, they deal with synthetic aperture radar (SAR) images). In this case, the hierarchical approach remains however at the level of data modeling, and the segmentation itself is then performed in the maximum likelihood (ML) sense, in-scale causality giving complete access to required likelihoods.

With MRF-based labeling priors we are looking at, a multiresolution approach can be classically introduced as follows: a hierarchical decomposition of the original image to be segmented is computed and "similar" labeling MRF's of proper sizes are attached to the different levels of resolution. Based on this structure, the segmentation is then performed in a coarse-to-fine way, using at each level the segmentation at previous level to get an initialization [9], [34]. The problem with such approaches lies in the issue of deriving the parameters of both prior model and data model at each level, given those of the finest level. As for data model parameters, a consistent treatment of this issue is introduced in [34] for Gaussian "textures." For other data and/or prior models, the issue remains widely open.

Another approach consists in devising a hierarchical labeling MRF which simply interacts with the original data (no multiresolution decomposition of the data is performed in this case). A number of such models have been proposed. Bouman *et al.* introduced a peculiar hierarchical MRF defined as a coarse-to-fine Markov chain of levels [35]. The associated interaction structure is a quadtree. It allows to devise a noniterative two-sweep segmentation procedure (in case of known parameters), but it induces undesirable spatial nonstationarities. As a fact, the model is not shift-invariant since two pixels that would be adjacent in terms of spatial lattice may be actually "far" apart in the graph structure. In [36], Kato *et al.* introduced a novel hierarchical model: they considered a pyramidal label field with a three-dimensional neighborhood system. Unfortunately, the resulting parameter estimation and segmentation procedures require a lot of computing time, even for the case (considered therein) where the spatial part of the prior corresponds to the first-order isotropic Ising model. Therefore, it did not appear to us as a good candidate in the complex context of sonar imagery where numerous sources of anisotropy and content variability are to be met (e.g. sea-floor with pebbles, dunes, ridges, sand, tires or various objects, etc.)

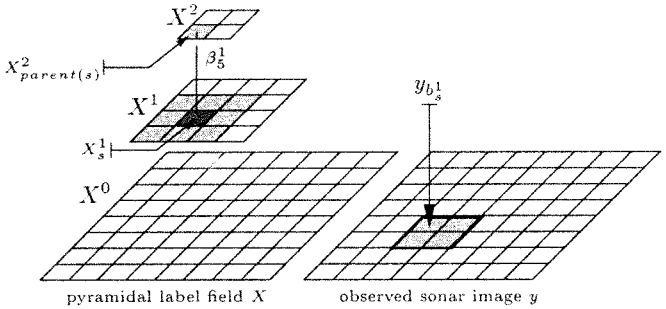

Fig. 5. Hierarchical labeling structure involved in the SCM method; sites in light gray hold variables that interact with X^1_s within the posterior conditional distribution $P_{X^1|X^2,Y}$.

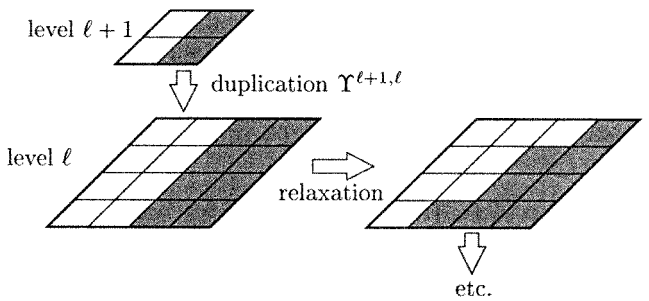

Fig. 6. "Coarse-to-fine" minimization strategy.

Herein, we propose a different hierarchical approach combining a scale-causal specification (modeling part) with a coarse-to-fine multigrid minimization technique (algorithmic part). As in [36], this model involves local interactions between spatially adjacent sites as well as parent-child interactions between sites belonging to consecutive levels. We will refer to this approach as scale causal multigrid (SCM). From an algorithmic point of view, this structure will allow to propagate efficiently contextual information. In terms of modeling, it offers the opportunity to capture some *a priori* characteristics of the underlying labeling process within a range of different scales.

The use of this original SCM prior for unsupervised segmentation requires an adaptation of the parameter estimation technique we used for nonhierarchical Ising prior. This shall be exposed, after the coming presentation of the SCM approach.

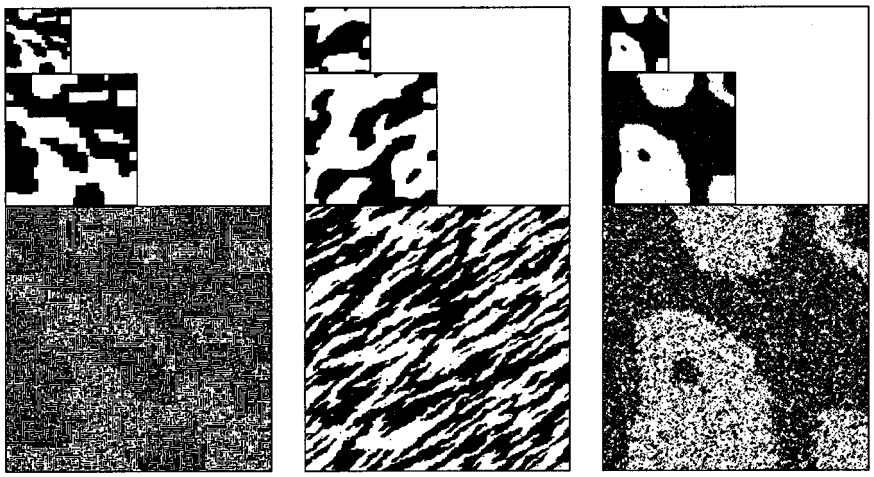

Fig. 7. Realizations of the hierarchical SCM model for the specified parameters reported in Table III.

B. SCM Hierarchical Modeling and Segmentation

The SCM model first consists in a hierarchy $(X^L \cdots X^0)$ of label fields which interact with the original image Y (see Fig. 5). Labeling X^ℓ is defined on grid S^ℓ which results from the reduction of $S^0 \equiv S$ by 2^ℓ in each direction. The segmentation of sonar images in two classes is now stated as a causal cascade of *conditional* MAP estimations

$$\hat{x}^L = \arg \max_{x^L} P_{X^L|Y}(x^L|y), \quad \text{and}$$
$$\hat{x}^\ell = \arg \max_{x^\ell} P_{X^\ell|X^{\ell+1},Y}(x^\ell|\hat{x}^{\ell+1}, y) \qquad \ell = L-1 \cdots 0. \tag{11}$$

The definition of the different inter-level transition distributions $P_{X^\ell|X^{\ell+1},Y}$ is based on the monoresolution model introduced in Section II-B. We first define $P_{X^0|X^1,Y}$ as the following Gibbs distribution:

$$P_{X^0|X^1,Y}(x^0|x^1,y) \propto \exp\{-U^0(x^0,x^1,y)\} \quad \text{with}$$
$$U^0(x^0,x^1,y) = U(x^0,y) + \sum_{s\in S^0} \beta_5 \mathcal{I}(x_s^0, x_{parent(s)}^1) \tag{12}$$

where

$$U(x^0,y) = \sum_{s\in S^0} \Psi_s(x_s^0, y_s) + \sum_{\langle s,t\rangle \subset S^0} \beta_{st} \mathcal{I}(x_s^0, x_t^0)$$

is the energy function defined in (1) applied to x^0, and a new parameter, β_5, is introduced in inter-level clique potentials. We now derive the other inter-level transition distributions extending the multigrid construction technique proposed in [37] to the two-level energy in (12).

Let $\Upsilon^{\ell,k}(k < \ell)$ be the "projection" from S^ℓ to S^k which associates, by duplication, a blockwise constant configuration on S^k to any configuration on S^ℓ. Using the multigrid approach from [37], we can define for each $\ell < L$

$$P_{X^\ell|X^{\ell+1},Y}(x^\ell|x^{\ell+1},y) \propto \exp\{-U^\ell(x^\ell,x^{\ell+1},y)\} \quad \text{with}$$
$$U^\ell(x^\ell,x^{\ell+1},y) \triangleq U^0[\Upsilon^{\ell,0}(x^\ell), \Upsilon^{\ell+1,1}(x^{\ell+1}), y]. \tag{13}$$

TABLE III
SPECIFIED *A PRIORI* PARAMETERS Φ_x FOR THE SCM MODEL, AND ESTIMATED PARAMETERS $\hat{\Phi}_x$ OBTAINED FROM REALIZATIONS

Fig.		β_1	β_2	β_3	β_4	β_5
7(a)	Φ_x	1	1	−1	−1	1
	$\hat{\Phi}_x$	0.91	0.91	−0.91	−0.90	0.94
7(b)	Φ_x	2	1	−1	1	0
	$\hat{\Phi}_x$	1.84	1.04	−0.94	0.84	−0.07
7(c)	Φ_x	0.25	0.25	0.25	0.25	0.25
	$\hat{\Phi}_x$	0.26	0.25	0.26	0.26	0.21

From (12) and (13), simple computations provide [37]

$$U^\ell(x^\ell,x^{\ell+1},y) = \sum_{s\in S^\ell} \Psi_s^\ell(x_s^\ell, y_{b_s^\ell}) + \sum_{\langle s,t\rangle \subset S^\ell} \alpha_{st}^\ell \beta_{st} \mathcal{I}(x_s^\ell, x_t^\ell) + \sum_{s\in S^\ell} \alpha_5^\ell \beta_5 \mathcal{I}(x_s^\ell, x_{parent(s)}^{\ell+1}) \tag{14}$$

with

$$\Psi_s^\ell(x_s^\ell, y_{b_s^\ell}) \triangleq \sum_{p\in b_s^\ell} \Psi_s(x_s^\ell, y_p) \tag{15}$$

where $b_s^\ell \subset S^0$ is the $2^\ell \times 2^\ell$ block of the "descendants" of s (see Fig. 5), and $y_{b_s^\ell} \triangleq \{y_p\colon p \in b_s^\ell\}$ stands for the data attached to that block. It is readily established that the intra-level clique parameters, $\beta_i^\ell \triangleq \alpha_i^\ell \beta_i$, $i = 1 \cdots 4$, associated to the four different possible orientations of clique $\langle s,t\rangle \subset S^\ell$, and the inter-level clique parameter $\beta_5^\ell \triangleq \alpha_5^\ell \beta_5$ are obtained by weighting those in (12) according to [37]

$$\alpha^\ell = \alpha_2^\ell = [2^\ell + 2(2^\ell - 1)],$$
$$\alpha_3^\ell = \alpha_4^\ell = 1, \quad \alpha_5^\ell = 4^\ell, \tag{16}$$

where α_i^ℓ, $i = 1 \cdots 5$, accounts for the number of site cliques straddling two $2^\ell \times 2^\ell$ blocks that constitute a block clique of type i, and α_5^ℓ is the number of site cliques contained in one $2^\ell \times 2^\ell$ block.

Conditional MAP estimation at level ℓ amounts to minimizing $U^\ell(x^\ell, \hat{x}^{\ell+1}, y)$ in x^ℓ. The interaction structure

1) Non-hierarchical Parameter Estimation

(a) Initialization:

– classification of 6×6 image blocks by k-means clustering technique

– ML fitting of a Gaussian law $\mathcal{N}(\mu, \sigma^2)$ in *shadow* area, and of a shifted Rayleigh law $\mathcal{R}(min, \alpha^2)$ in *sea-bottom* area $\Rightarrow \Phi_y^{[0]}$

– ML segmentation of y based on $\Phi_y^{[0]}$, and LSQR fitting of a second-order anisotropic Ising model on this segmentation $\Rightarrow \Phi_x^{[0]} = (\beta_1^{[0]}, \beta_2^{[0]}, \beta_3^{[0]}, \beta_4^{[0]})^T$

(b) ICE iteration:

– draw one sample x from $P_{X|Y,\Phi}(\cdot|y, \Phi^{[k]})$

– ML fitting of data mixture model based on $(x, y) \Rightarrow \Phi_y^{[k+1]}$

– LSQR fitting of second-order anisotropic Ising model on $x \Rightarrow \Phi_x^{[k+1]}$

– $k \leftarrow k + 1$

(c) Termination:

– if var$\{\Phi^{[k-k_0]} \dots \Phi^{[k]}\}$ >threshold return to 1.(b)

– else $\Phi^* \leftarrow \Phi^{[k]}$

– ICM segmentation based on $P_{X|Y,\Phi}(\cdot|y, \Phi^*) \Rightarrow$ "ICM segmentation"

2) Unsupervised Hierarchical Segmentation

(a) Initialization: $\Phi_y \leftarrow \Phi_y^*$, $\Phi_x \leftarrow (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \beta_5 = 0)^T$

(b) Multigrid segmentation: while $\ell \geq 0$, do

– initialization: if $\ell = L$, $x_s^L \leftarrow \arg\min_{x_s^L} \Psi_s^L(x_s^L, y_{b_s^L})$, $\forall s$; else $x^\ell \leftarrow \Upsilon^{l+1,l}(\hat{x}^{\ell+1})$

– starting from x^ℓ, ICM segmentation based on $P_{X^\ell|X^{\ell+1},Y,\Phi}(\cdot|\hat{x}^{\ell+1}, y, \Phi^*)$
$\Rightarrow$ "MG segmentation" at level ℓ, $\hat{x}^\ell$

– $\ell \leftarrow \ell - 1$

(c) Parameter re-estimation:

– ML fitting of data mixture model based on $(\hat{x}^0, y) \Rightarrow \Phi_y^{\text{final}}$

– LSQR fitting of SCM model on $(\hat{x}^L \cdots \hat{x}^0) \Rightarrow \Phi_x^{\text{final}}$

(d) Final segmentation: as in 2.(b) but using $\Phi^{\text{final}} \Rightarrow$ "SCM segmentation"

Fig. 8. Unsupervised SCM segmentation algorithm.

concerned by this minimization is the following: each variable x_s^ℓ interacts with the (unknown) second-order neighborhood labeling $x_{\nu_s}^\ell$, with the label $\hat{x}_{parent(s)}^{\ell+1}$ of its parent provided by minimization at level $\ell + 1$, and with the block of data $y_{b_s^\ell}$.

Finally, at coarsest level, we define

$$P_{X^L|Y}(x^L|y) \propto \exp\{-U^L(x^L, y)\}, \quad \text{with}$$
$$U^L(x^L, y) \triangleq U^0[\Upsilon^{L,0}(x^L), y]. \tag{17}$$

Energy U^L is the same as the one in (14) with $\ell = L$, apart from the last inter-level term which is obviously absent in this case.

We now have to deal with the coarse-to-fine recursive estimation (11). Each of the associated energy minimization problems is coped with the ICM algorithm [8]. The deterministic iterative minimization at level ℓ is initialized by the interpolation $\Upsilon^{l+1,l}(\hat{x}^{l+1})$ of the labeling previously obtained at coarser level $\ell + 1$ (see Fig. 6).

C. Estimation of Hierarchical Model Parameters

The introduction of the hierarchical modeling requires to re-address the issue of parameter estimation. This has been decomposed into two steps. We first let $\beta_5 = 0$ and use the estimates of parameters $\Phi_y, \beta_1, \beta_2, \beta_3$, and β_4 provided by the nonhierarchical ICE procedure described in Section II-B, to solve the SCM labeling problem (11).[4] The first multigrid segmentation $(x^L \cdots x^0)$ thus obtained is used in turn to re-estimate the whole set of parameters (including β_5) of the SCM

[4]Since $\beta_5 = 0$, we deal at this stage with the multigrid MAP estimation procedure introduced in [37], based on the ICE parameter estimates.

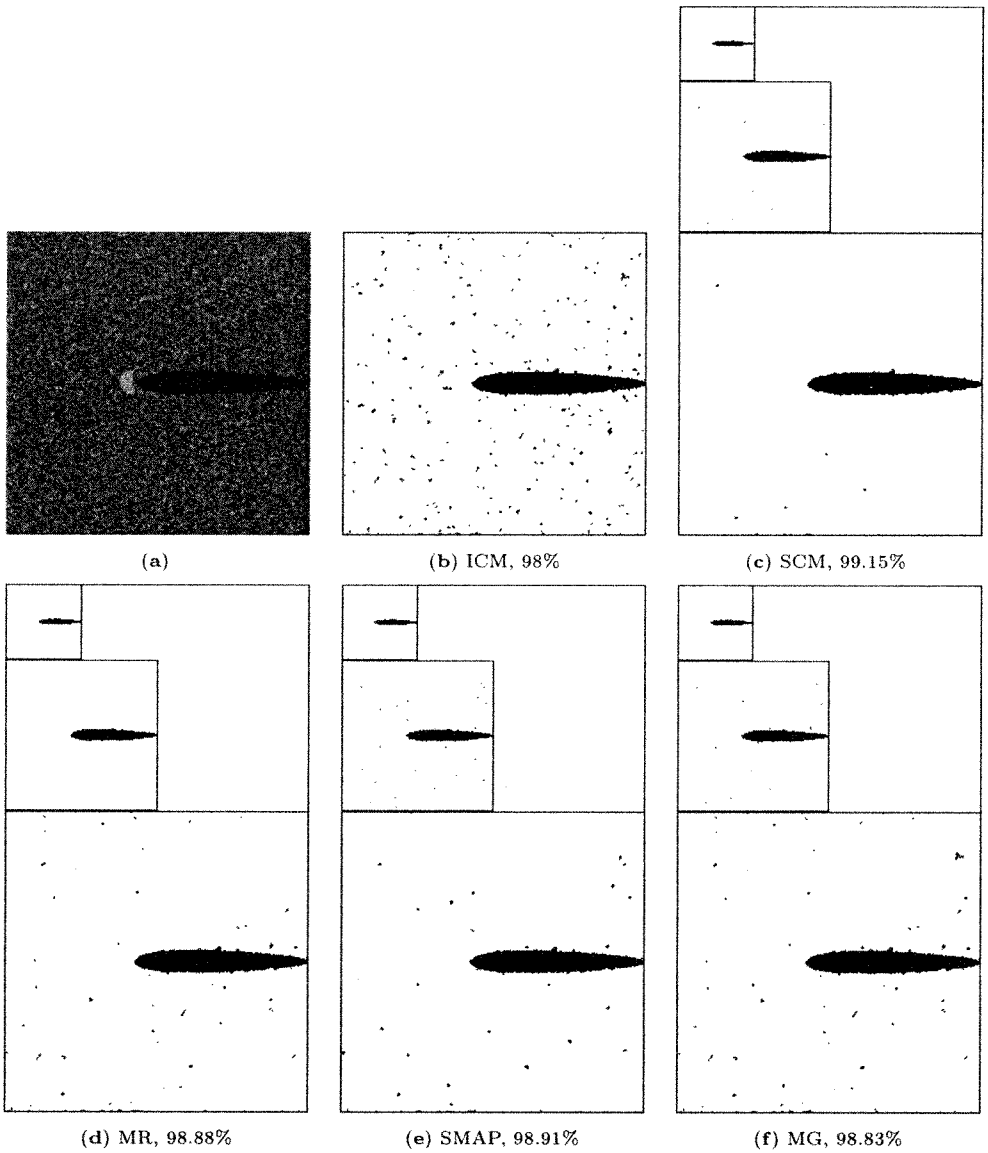

Fig. 9. (a) Synthetic sonar image of a spherical object lying on a sandy sea-floor and (b)–(f) two-class segmentation results and associated rates of correct classification.

model. Once this is done, a final SCM labeling is estimated in a coarse-to-fine way. We shall assess, with coming experimental comparisons, the impact of the SCM approach in terms of both final parameter estimates and final image segmentation, as compared to the nonhierarchical unsupervised approach introduced in Section II.

Given the hierarchical segmentation $(x^L \cdots x^1, x^0 = x)$ the ML estimation of data model parameters Φ_y remains unchanged. It is provided by $\hat{\Phi}_y(x^0, y)$. As concerns the estimation of parameters $\Phi_x = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$ which define the prior part of SCM model we need to extend the nonhierarchical method presented in Section II-C2. We derived this extension by expressing relation (9) at each level of the pyramidal structure, conditioned on the labeling of above level. Using empirical approximations computed on given $(x^L \cdots x^1, x^0 = x)$, one gets for each of the $2^9 = 512$ possible configurations η of $(X^\ell_{\nu_s}, X^{\ell+1}_{parent(s)})$ the following set of equations:

$$[\Theta(e_0,\eta) - \Theta(e_1,\eta)]^T \underline{M}^\ell \Phi_x = \ln \frac{\#\{s \in S^\ell \colon x^\ell_s = e_1, (x^\ell_{\nu_s}, x^{\ell+1}_{parent(s)}) = \eta\}}{\#\{s \in S^\ell \colon x^\ell_s = e_0, (x^\ell_{\nu_s}, x^{\ell+1}_{parent(s)}) = \eta\}}, \quad \ell = 0 \cdots L-1$$

with diagonal matrix $\underline{M}^\ell \triangleq \text{diag}\ (\alpha^\ell_1 \cdots \alpha^\ell_5)$. Vector $\underline{M}^\ell \Phi_x$ corresponds to the clique parameters of posterior conditional Gibbs distribution (13). We obtain $2^9 \times L$ linear equations in β_i's. We ignore equations associated with any combination (e_i, η) that does not occur in the label fields and equations for which $\Theta(e_0, \eta) = \Theta(e_1, \eta)$. The resulting over-determined linear system of equations is solved in the least-squares sense.

Like for nonhierarchical model, we present some examples of estimation of the SCM model, based on various realizations

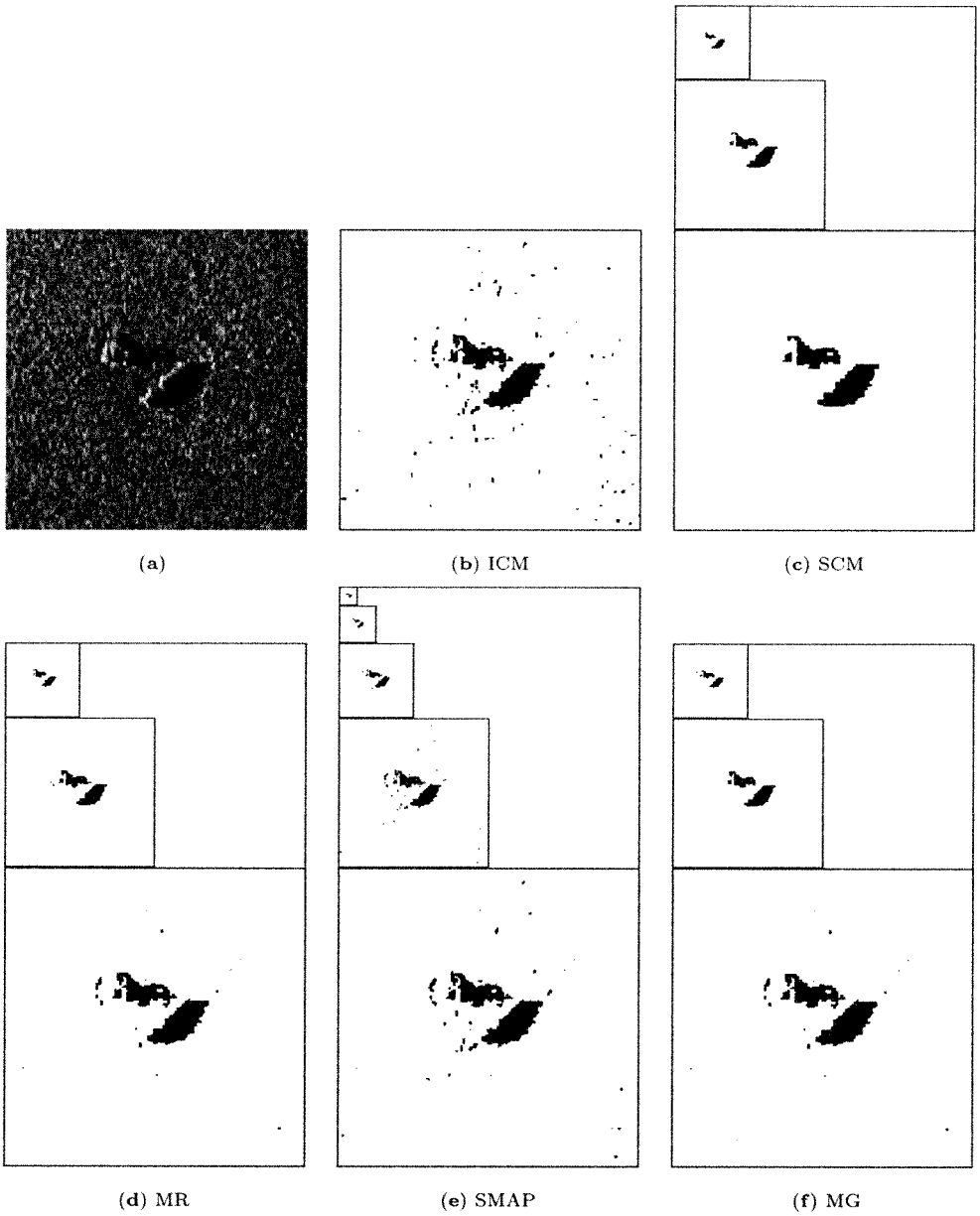

Fig. 10. (a) Real sonar image involving object and rock shadows and (b)–(f) two-class segmentation results.

of the model itself (see these samples in Fig. 7). The obtained estimates are compared to the specified values of the parameters in Table III. It appears that the SCM model offers a larger range of modeling possibilities than the one of the monoresolution model (with the expected ability of capturing in a joint way patterns of very different scales), and that good estimates of its parameters can be obtained from a given hierarchical labeling configuration.

Before we report experiments of unsupervised segmentation of sonar image with SCM approach, we summarize the whole procedure in Fig. 8.

IV. Experimental Results

The whole SCM algorithm actually provides three successive unsupervised segmentations (see Fig. 8): the first one is obtained by ICM from the nonhierarchical modeling (10), with estimated parameters $\Phi_y^* = (\mu^*, \sigma^*, min^*, \alpha^*)^T$ and $\Phi_x^* = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)^T$; a second hierarchical segmentation is obtained by coarse-to-fine minimization on the SCM model with same parameters as before and $\beta_5 = 0$; the last segmentation is obtained as the previous one, but with re-estimated parameters $\Phi_y^{\text{final}} = (\mu^{\text{final}}, \sigma^{\text{final}}, min^{\text{final}}, \alpha^{\text{final}})^T$ and $\Phi_x^{\text{final}} = (\beta_1^{\text{final}}, \beta_2^{\text{final}}, \beta_3^{\text{final}}, \beta_4^{\text{final}}, \beta_5^{\text{final}})^T$. In the following, we shall refer to these three outputs as the ICM, MG (for multigrid), and SCM segmentation, respectively.

We have compared on a number of images these three segmentations to those obtained by two other standard techniques: the noniterative SMAP algorithm on a quadtree proposed by Bouman *et al.* [35], and a classical multiresolution (MR) approach where a multiresolution pyramid of images is derived and a set of "similar" spatial models is considered on the dif-

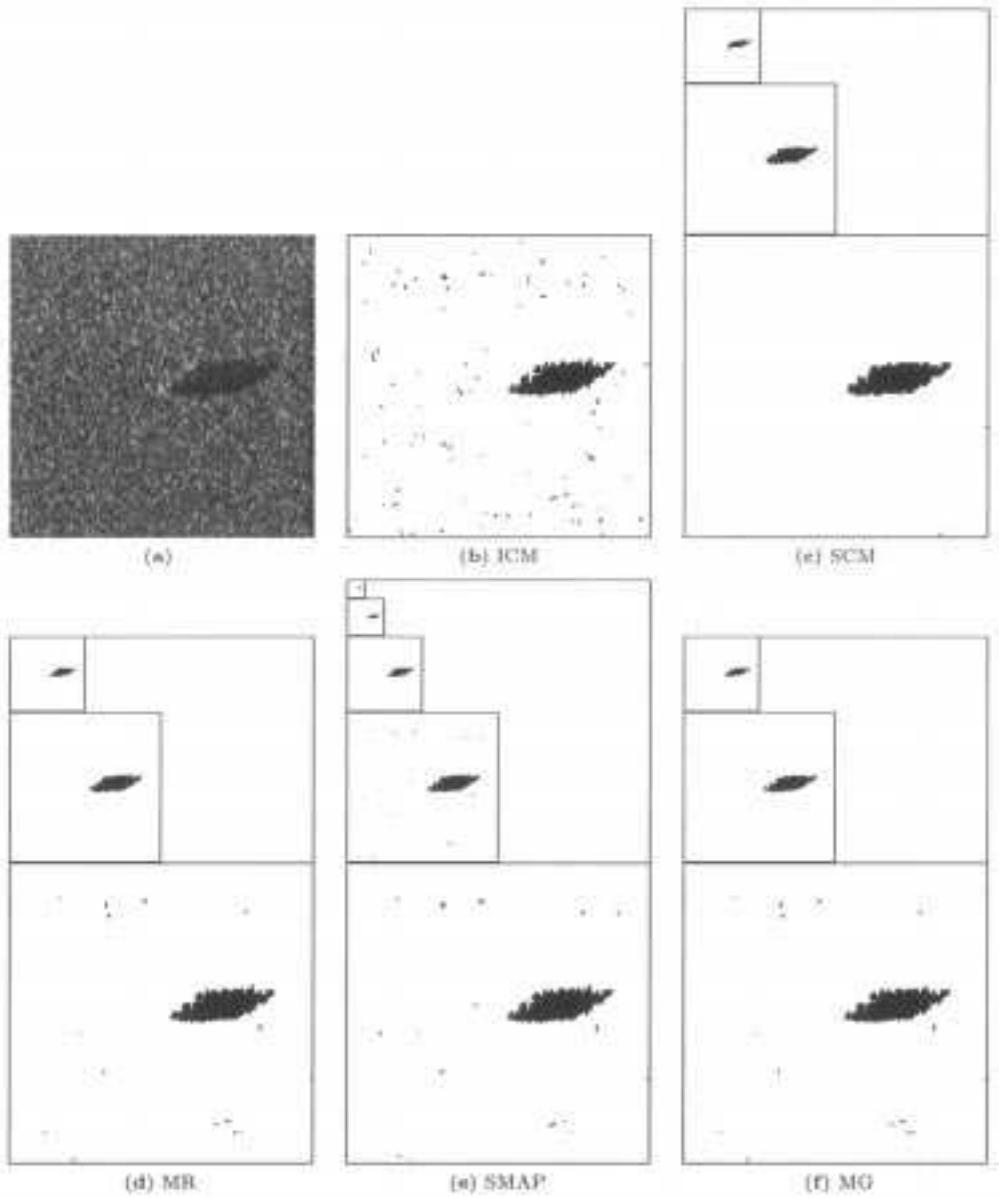

Fig. 11. (a) Real sonar image of sandy sea-floor with the shadow of a man-made object and (b)–(f) two-class segmentation results.

ferent resolution levels. In both cases, data modeling is the same as in SCM (mixture of Gaussian and Rayleigh laws) with parameters Φ_y^*. The only prior parameter involved by the tree-based model of Bouman *et al.*(the probability that a node of the tree exhibits the same label as the one of its parent) has been manually tuned. At each level of the model used for MR segmentation, the anisotropic second-order Ising MRF with parameters Φ_x^* is used as a prior. In all experiments, MG, SCM, and MG segmentations were obtained on three levels, whereas the SMAP estimation takes place on $0.5 \log_2 N$ levels, by definition.

We first report experimental comparisons obtained on a synthetic sonar image (Fig. 9) of a spherical object lying on a sand sea-bed. The image is simulated by using a ray tracing procedure. The different segmentations obtained on this example are presented in Fig. 9, along with their associated rates of correct classification. As visible from the different rates of success (all are at least 98%), this is an easy example, and one has to be careful when trying to extrapolate these results toward real sonar images. Nevertheless, we would like to note that the complete SCM approach slightly out-performs the others in terms of global classification performance, while providing a classification that is visually cleaner than the others. As concerns the estimated parameters for this synthetic example, one can observe that the β_5 parameter of SCM model remains null: this shows that the inter-level part of the model is not relevant on this particular synthetic image. It is not always the case however with real sonar images, as coming experiments will show.

Thorough experiments have been conducted on real sonar images. Figs. 10–12 show the ICM, MG, SMAP, MR, and SCM segmentation results obtained on images with various contents. The associated SCM parameter estimates Φ^{final} are given in Table IV. The size of these pictures is 256 × 256 pixels corresponding to a sea-floor surface of 25 m × 25 mm. (The DUBM41 sonar frequency is around 500 KHz.)

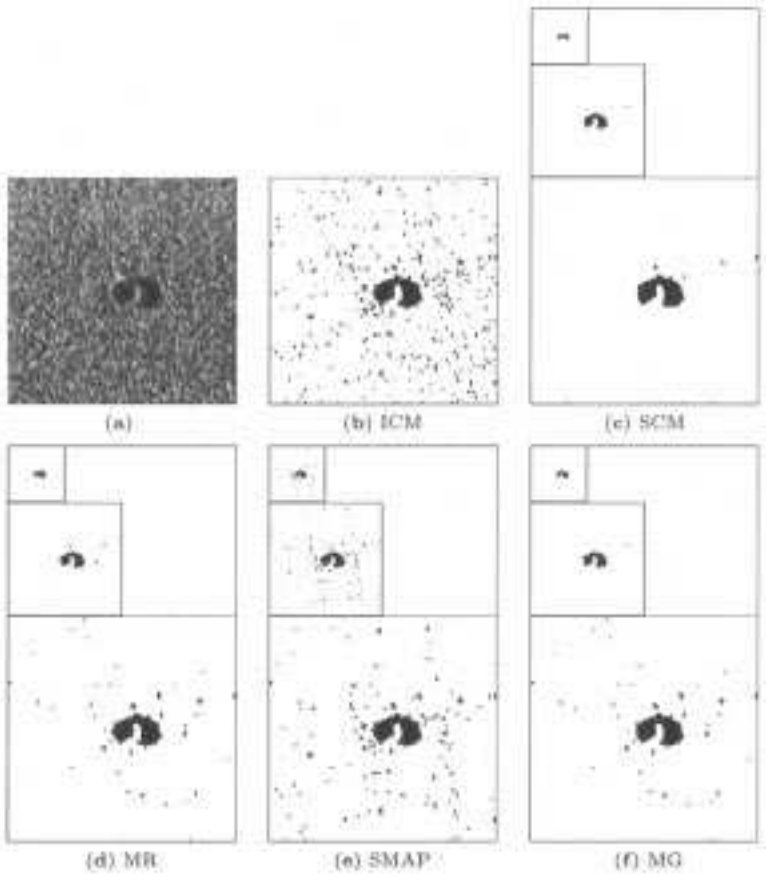

Fig. 12. (a) Real sonar image of sandy sea-floor with the cast shadow of a tire and (b)–(f) two-class segmentation results.

Let us first stress that the estimated prior parameters are quite different according to the type of objects involved in the input image (man-made or natural objects). In presence of small shadows, it also turns out that estimated values of parameters $\beta_1 \cdots \beta_4$ are significantly smaller (see Table IV). The resulting model is then able to extract shadows of only a few pixels large while avoiding false alarms. We can also notice that the estimated noise model parameters depend a lot on the nature of the sea-floor (e.g., sandy or pebbly sea-floor, etc.).

In some cases (e.g., on image from Fig. 12), the final estimate for β_5 remains null and the SCM approach amounts to the multigrid segmentation, but with *a multigrid refinement of all the other parameter estimates*. Even in these cases, final estimates of other parameters are different from those issued from the first nonhierarchical ICE step. The difference between the parameter values which MG and SCM segmentations respectively rely on can be assessed on the example from Fig. 3(a) [same image as in Fig. 10(a)] by looking at Φ^* in Table II and Φ^{final} reported for Fig. 10 in Table IV.

Table V presents the computational cost associated to the different segmentations we compared. It is expressed in terms of both cpu times (on a 200 MHz 43P IBM Workstation) and number of "equivalent iterations." One "equivalent iteration" corresponds to the update of N labels. One complete image sweep of ICM algorithm at level ℓ (for MR, MG, or SCM segmentation) thus amounts to $N/4^{\ell}$ equivalent iterations. As expected, the noniterative SMAP requires less computation than any other technique. However, its peculiar tree-based neighborhood structure does not seem able to capture properly the local properties of sonar images, and this model often produces blocky segmentations at the boundaries of the shadows. The other less expensive technique is nonhierarchical ICM which also provides segmentations of lower quality. Proposed SCM method is as expensive as MR technique for improved results, and requires twice as much time as the multigrid (MG) segmentation does. The latter method being an ingredient of the former one, this was expected. We do believe that reported experiments demonstrate that the extra cost is worth the pain.

It is seen from Figs. 10–12 that SMAP, ICM, MG, and MR segmentations are all plagued by a significant number of false alarms (wrong shadow detections) due to the speckle noise effect. In contrast, SCM performs better: it exhibits a good robustness against speckle noise (most false alarms are eliminated) while preserving the border of shadows. The versatility of SCM prior is further demonstrated by two other examples (Fig. 13). Indeed, it allows to get satisfactory segmentations even in case of numerous small rocks [Fig. 13(a) and (b)], as well as in presence of sand ridges with strong orientation [Fig. 13(c) and (d)].

These experiments show that the shadows of both manufactured objects and rocks are better segmented by SCM method than by other methods to which it is compared. In particular, the extracted cast shadow of the manufactured object (a cylinder) in Fig. 10 exhibits a regular geometric shape (contrary to the cast

TABLE IV
ESTIMATED PARAMETERS BY SCM APPROACH ON SYNTHETIC AND REAL SONAR IMAGES

Fig. 9	$\Phi^{final}_{y(shadow)}$		$0.04_{(\pi_0)}$	$11_{(\mu)}$	35_{σ^2}	
	$\Phi^{final}_{y(sea-bottom)}$		$0.96_{(\pi_1)}$	$41_{(min)}$	$50_{(\alpha^2)}$	
	Φ^{final}_{x}	$1.0_{(\beta_1)}$	$0.8_{(\beta_2)}$	$0.4_{(\beta_3)}$	$0.2_{(\beta_4)}$	$0.0_{(\beta_5)}$
Fig. 10	$\Phi^{final}_{y(shadow)}$		$0.03_{(\pi_0)}$	$24_{(\mu)}$	$28_{(\sigma^2)}$	
	$\Phi^{final}_{y(sea-bottom)}$		$0.97_{(\pi_1)}$	$32_{(min)}$	$1713_{(\alpha^2)}$	
	Φ^{final}_{x}	$2.9_{(\beta_1)}$	$2.6_{(\beta_2)}$	$-1_{(\beta_3)}$	$-0.8_{(\beta_4)}$	$0.1_{(\beta_5)}$
Fig. 11	$\Phi^{final}_{y(shadow)}$		$0.03_{(\pi_0)}$	$32_{(\mu)}$	$34_{(\sigma^2)}$	
	$\Phi^{final}_{y(sea-bottom)}$		$0.97_{(\pi_1)}$	$38_{(min)}$	$1230_{(\alpha^2)}$	
	Φ^{final}_{x}	$2.1_{(\beta_1)}$	$3_{(\beta_2)}$	$-1.2_{(\beta_3)}$	$-1_{(\beta_4)}$	$0.6_{(\beta_5)}$
Fig. 12	$\Phi^{final}_{y(shadow)}$		$0.02_{(\pi_0)}$	$35_{(\mu)}$	$104_{(\sigma^2)}$	
	$\Phi^{final}_{y(sea-bottom)}$		$0.98_{(\pi_1)}$	$43_{(min)}$	$2272_{(\alpha^2)}$	
	Φ^{final}_{x}	$2.6_{(\beta_1)}$	$2.9_{(\beta_2)}$	$-1.1_{(\beta_3)}$	$-1.1_{(\beta_4)}$	$0_{(\beta_5)}$
Fig. 13(a)	$\Phi^{final}_{y(shadow)}$		$0.1_{(\pi_0)}$	$33_{(\mu)}$	$66_{(\sigma^2)}$	
	$\Phi^{final}_{y(sea-bottom)}$		$0.9_{(\pi_1)}$	$44_{(min)}$	$3768_{(\alpha^2)}$	
	Φ^{final}_{x}	$1.1_{(\beta_1)}$	$0.4_{(\beta_2)}$	$-0.2_{(\beta_3)}$	$0.5_{(\beta_4)}$	$0.2_{(\beta_5)}$
Fig. 13(c)	$\Phi^{final}_{y(shadow)}$		$0.13_{(\pi_0)}$	$30_{(\mu)}$	$39_{(\sigma^2)}$	
	$\Phi^{final}_{y(sea-bottom)}$		$0.87_{(\pi_1)}$	$40_{(min)}$	$3815_{(\alpha^2)}$	
	Φ^{final}_{x}	$0.07_{(\beta_1)}$	$1_{(\beta_2)}$	$-0.01_{(\beta_3)}$	$1.3_{(\beta_4)}$	$0_{(\beta_5)}$

TABLE V
COMPUTATIONAL COST OF THE SEGMENTATION TECHNIQUES, EXPRESSED IN NUMBER OF "EQUIVALENT ITERATIONS" (*TOTAL NUMBER OF SITE UPDATES/N*), AND CPU TIMES, FOR 256×256 SONAR IMAGES

	ICM	MG	SCM	MR	SMAP
# iter.	5	7.75	15.25	12.5	3.5
cpu time	13s	21s	40s	32s	9s

shadow of the rock) which is in excellent agreement with the ground truth provided by an expert.

V. CONCLUSION

In this paper, we addressed the problem of segmenting high resolution sonar images into shadow areas and sea-bottom areas. Although this issue is important in many applications relying on this kind of images, it has only received few attention in the literature, where people mainly address the problems of sonar image restoration and surface reconstruction from 3-D sonar data.

For this specific unsupervised segmentation purpose, we propose a complete hierarchical Markovian approach which has been validated on a number of real high resolution sonar images. These experiments show that the proposed method allows to get accurate and robust results for a wide range of noise levels and image contents.

The modeling relies on the joint use of Ising-type priors and appropriate non-Gaussian conditional luminance distributions (a Gaussian likelihood in shadow areas, but a Rayleigh likelihood in sea-bottom areas). Based on these two ingredients, the approach is composed of two parts. The first part consists in a nonhierarchical estimation of all model parameters using the flexible ICE technique. Based on limited sampling, it allows to estimate the parameters of the two different data distributions with maximum likelihood techniques, and to recover the different parameters of the anisotropic prior with least-squares techniques. The versatility of ICE would allow to deal with other types of distributions if required. In particular, Weibull distribution (which Rayleigh law is a special instance of) can also be used to model the sea-bottom reverberation.

The initialization of the iterative parameter estimation scheme is provided by a simple clustering technique based on luminance distribution in small windows. Although convergence is not theoretically established, we observed it in all experiments (on more than thirty real 256×256 sonar images, and one 2000×6000 image).

The second part of the approach introduces a hierarchical component to perform the final segmentation. It is based on a multi-level prior model involving both scale-causal interactions and spatial interactions. The associated parameterization is derived in a consistent way from the one at the finest level, thanks to a multigrid technique. The ICE procedure previously em-

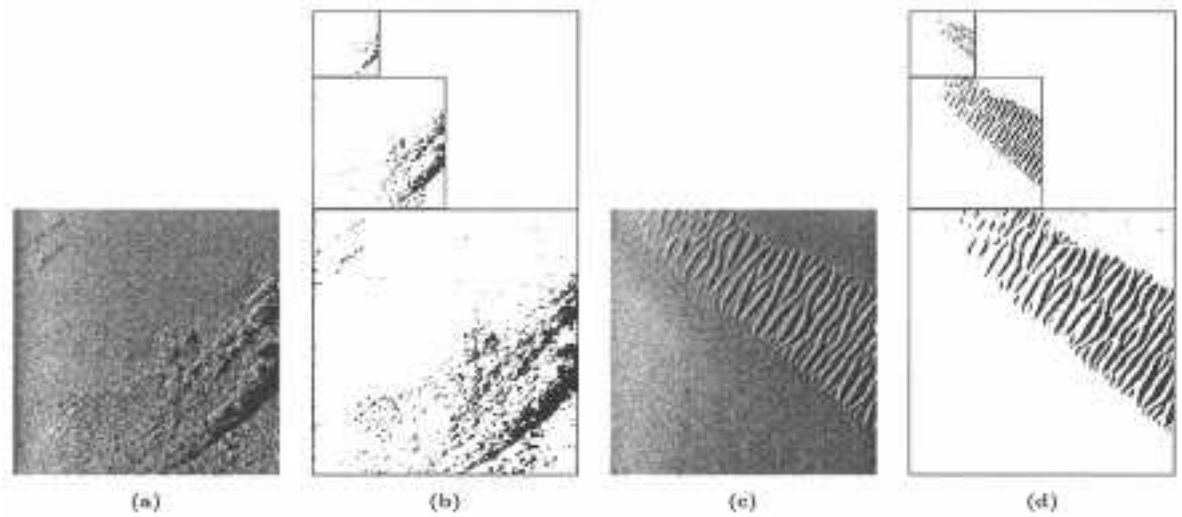

Fig. 13. (a)–(b) Real sonar image of pebbly sea-floor with no manufactured object, and associated SCM segmentation and (c)–(d) real sonar image of sandy sea floor and ridges of sand, and associated SCM segmentation.

ployed is then extended to deal simultaneously with the different levels of this new prior model. This allows to refine the parameter estimates obtained in the first part of the approach. The final segmentation is eventually obtained by solving a coarse-to-fine cascade of energy minimizations.

Like for other hierarchical approaches, the advantages of the method are twofold. From a modeling point of view, this hierarchical part of the approach offers an appealing ability to capture *a priori* characteristics of the underlying labeling process within a range of different scales. From an algorithmic point of view, contextual information is propagated in a more efficient way. But based on our experiments, the proposed SCM approach seems to perform better, for sonar image segmentation, than other standard hierarchical techniques. Whereas tree-based, multiresolution, and multigrid segmentation techniques remain plagued by a significant number of false alarms (wrong shadow detections), SCM technique exhibits a good robustness against speckle noise (most false alarms are eliminated) while preserving the border of shadows.

The complete twofold method thus allows an automatic and robust extraction of shadows from a large variety of high resolution sonar images with strong speckle noise. Provided segmentations can then be used for further treatments. We especially investigate a refined segmentation into three classes (*shadow*, *sea-bottom* reverberation, and *echo*), and the statistical identification of objects lying on sea-floor based on the shape of their extracted cast shadow [38].

ACKNOWLEDGMENT

The authors thank Groupe d'Étude Sous Marine de l'Atlantique (GESMA), Brest, for having provided numerous real sonar pictures.

REFERENCES

[1] F. Schmitt, M. Mignotte, C. Collet, and P. Thourel, "Estimation of noise parameters on sonar images," in *Proc. SPIE Signal Image Processing* Denver, CO, Aug. 1996, vol. 2823, pp. 1–12.

[2] C. Collet, P. Thourel, P. Pérez, and P. Bouthemy, "Hierarchical MRF modeling for sonar picture segmentation," in *Proc. 3rd IEEE Int. Conf. Image Processing*, vol. 3, Lausanne, Switzerland, Sept. 1996, pp. 979–982.

[3] S. Guillaudeux and S. Daniel, "Optimization of a sonar image processing chain: A fuzzy rules based expert system approach," in *Proc. OCEANS'96*, Fort Lauderdale, FL, 1996, pp. 1319–1323.

[4] S. Daniel and S. Guillaudeux, "Adaptation of a partial shape recognition approach," in *Proc. IEEE Conf. Systems, Man, Cybernetics*, Oct. 1997, pp. 2157–2162.

[5] S. Dugelay, C. Graffigne, and J. M. Augustin, "Deep seafloor characterization with multibeam echosounders by image segmentation using angular acoustic variations," in *Proc. SPIE'96 Int. Symp. Optical Science, Engineering, Instrumentation*, vol. 2847, 1996.

[6] V. Murino and A. Trucco, "Edge/region-based segmentation and reconstruction of acoustic underwater images by Markov random fields," in *Proc. Int. Conf. on Computer Vision Pattern Recognition*, Santa Barbara, CA, June 1999, pp. 408–413.

[7] V. Murino, A. Trucco, and C. Regazzoni, "A probabilistic approach to the coupled reconstruction and restoration of underwater acoustic images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 9–22, Jan. 1998.

[8] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, vol. B-48, pp. 259–302, 1986.

[9] P. Thourel, C. Collet, P. Bouthemy, and P. Pérez, "Multiresolution analysis and MRF modeling applied to the segmentation of shadows in sonar pictures," in *Proc. 2nd Asian Conf. Computer Vision*, vol. 2, Singapore, Dec. 1996, pp. 81–85.

[10] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, Aug. 1989.

[11] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. Vis.*, vol. 1, no. 1/2, pp. 261–268, 1992.

[12] B. Braathen, P. Masson, and W. Pieczynski, "Global and local methods of unsupervised Bayesian segmentation of images," *Mach. Graph. Vis.*, vol. 2, no. 1, pp. 39–52, 1993.

[13] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graph. Models Image Process.*, vol. 59, no. 4, pp. 205–220, 1997.

[14] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 618–633, May 1993.

[15] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Unsupervised Markovian segmentation of sonar images," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, München, Germany, May 1997, pp. 2781–2785.

[16] ——, "Unsupervised hierarchical Markovian segmentation of sonar images," in *Proc. Int. Conf. Image Processing*, Santa Barbara, CA, Oct. 1997.

[17] J. G. Postaire and P. A. Vasseur, "An approximate solution to normal mixture identification with application to unsupervised pattern classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 2, pp. 163–179, 1981.
[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, pp. 1–38, 1976.
[19] J. Zhang, J. Modestino, and D. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 404–419, Apr. 1994.
[20] S. A. Barker and P. J. W. Rayner, "Unsupervised image segmentation using markov random field models," in *Proc. Int. Workshop EMMCVPR'97: Energy Minimization Methods Computer Vision Pattern Recognition*, vol. 1223, Venice, Italy, May 1997, pp. 165–178.
[21] X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Maximum likelihood estimation of Markov random field parameters using markov chain Monte Carlo algorithms," in *Proc. Int. Workshop EMMCVPR'97: Energy Minimization Methods Computer Vision Pattern Recognition*, vol. 1223, Venice, Italy, May 1997, pp. 133–148.
[22] S. G. Nadabar and A. K. Jain, "Parameter estimation in Markov random field contextual models using geometric models of objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 326–329, Mar. 1996.
[23] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using a hierarchical Markovian model," in *Proc. Int. Conf. Computer Vision*, vol. 1, Cambridge, MA, June 20–23, 1995, pp. 169–174.
[24] B. Chalmond, "An iterative Gibbsian technique for reconstruction and m-ary images," *Pattern Recognit.*, vol. 22, no. 6, pp. 747–761, 1989.
[25] Y. Kim and H. S. Yang, "Efficient image labeling based on Markov random field and error backpropagation network," *Pattern Recognit.*, vol. 26, no. 11, pp. 1695–1707, 1993.
[26] J.-P. Delmas, "An equivalence of the EM and ICE algorithm for exponential family," *IEEE Trans. Signal Processing*, vol. 45, pp. 2613–2615, Oct. 1997.
[27] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Unsupervised segmentation applied on sonar images," in *Proc. Int. Workshop EMMCVPR'97: Energy Minimization Methods Computer Vision Pattern Recognition*, vol. LNCS 1223, Venice, Italy, May 1997, pp. 491–506.
[28] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 1, pp. 39–55, 1987.
[29] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721–741, 1984.
[30] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptative restoration of images with speckle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 373–383, Mar. 1987.
[31] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145–1150, 1976.
[32] S. Banks, *Signal Processing, Image Processing and Pattern Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[33] C. Fosgate, H. Krim, W. Irving, W. Karl, and A. Willsky, "Multiscale segmentation and anomaly enhancement of SAR imagery," *IEEE Trans. Image Processing*, vol. 6, pp. 7–20, Jan. 1997.
[34] S. Krishnamachari and R. Chellapa, "Multiresolution GMRF models for image segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 251–267, Feb. 1997.
[35] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 162–177, Feb. 1994.
[36] Z. Kato, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model and multitemperature annealing for parallel image classification," *Graph. Models Image Process.*, vol. 58, no. 1, pp. 18–37, 1996.
[37] F. Heitz, P. Pérez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," in *CVGIP: Image Understand.*, vol. 59, 1994, pp. 125–134.
[38] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Statistical model and genetic optimization: Application to pattern detection in sonar images," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Seattle, WA, May 1998, pp. 2471–2475.

Max Mignotte received the D.E.A degree in digital signal, image, and speech processing from the Institut National Polytechnique of Grenoble, France, in 1993, and the Ph.D. degree in digital signal and image processing from the University of Bretagne Occidentale, France, in 1998. His thesis work was done at the Laboratory Groupe de Traitement du Signal, French Naval Academy.

He is currently an INRIA Postdoctoral Fellow at Montreal University (DIRO), Montreal, P.Q., Canada. His current research interests include Bayesian inference, image segmentation (hierarchical markovian models or deformable models), parameters estimation, tracking, classification, deconvolution/restoration, and optimization problems in sonar or medical images.

Christophe Collet received the D.E.A. degree in signal processing from the University Paris-Sud, Orsay, France, in 1989 and the Ph.D. degree in image analysis from the University of Toulon, France, in 1992. His thesis research was carried out at the Ecole Nationale Superieure des Telecom. de Bretagne (ENST-Br) and at the French Naval Academy.

He has been Assistant Professor at the French Naval Academy and Chairman of the Laboratory Groupe de Traitement du Signal (GTS) since 1994. The line of research of the GTS laboratory concerns signal and image processing for underwater acoustic applications. His main research interests include image modeling, Markovian segmentation, Bayesian estimation, pattern recognition, and statistical optimization techniques for image analysis. The main applications deduced from this reasearch framework deal with motion segmentation of infrared image sequences, segmentation of seafloor for man-made pattern classification, and multispectral segmentation of SPOT images for bathymetric mapping.

Patrick Pérez was born in 1968. He graduated from École Centrale, Paris, France, in 1990. He received the Ph.D. degree in signal processing and telecommunications from the University of Rennes, France, in 1993.

He now holds a full-time research position at the INRIA Center, Rennes. His research interests include statistical and/or hierarchical models for large inverse problems in image analysis.

Patrick Bouthemy graduated from Ecole Nationale Superieure des Telecommunications, Paris, in 1980, and received the Ph.D degree in computer science from the University of Rennes, France, in 1982.

From December 1982 to February 1984, he was with the Department of Visual Communications, INRS-Telecommunications, Montreal, P.Q., Canada. Since April 1984, he has been with INRIA/IRISA, Rennes, France. He is currently Directeur de Recherche and Head of the Vista Project. His major research interests are concerned with image sequence analysis: 2-D motion analysis (rigid motion and deformable motion), statistical models for image sequence processing (MRF models, robust estimation, Bayesian estimation), dynamic scene interpretation, active vision, and video indexing.

Dr. Bouthemy is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.