Bayesian Segmentation of Oceanic SAR Images: Application to Oil Spill Detection

Sónia Pelizzari and José M. Bioucas-Dias

Abstract— This paper introduces Bayesian supervised and unsupervised segmentation algorithms aimed at oceanic segmentation of SAR images. The data term, *i.e.*, the density of the observed backscattered signal given the region, is modeled by a finite mixture of Gamma densities with a given predefined number of components. To estimate the parameters of the class conditional densities, a new expectation maximization algorithm was developed. The prior is a multi-level logistic Markov random field enforcing local continuity in a statistical sense. The smoothness parameter controlling the degree of homogeneity imposed on the scene is automatically estimated, by computing the evidence with loopy belief propagation; the classical coding and least squares fit methods are also considered. The maximum a posteriori segmentation is computed efficiently by means of recent graph-cut techniques, namely the α-Expansion algorithm that extends the methodology to an optional number of classes. The effectiveness of the proposed approaches is illustrated with simulated images and real ERS and Envisat scenes containing oil spills.

Index Terms— Oceanic SAR images, Segmentation, Markov Random Fields, Energy minimization, Graph cuts, Mixture of Gammas, Oil spills.

I. INTRODUCTION

A wide number of oceanic phenomena become visible on SAR images as they have distinct scattering characteristics, namely the sea surface roughness and thus the normalized radar cross section (NRCS). Among these phenomena are gravity waves, convective cells, oceanic internal waves, current and coastal fronts, eddies, upwelling processes, ship wakes and oil pollution [1]. The automatic detection of these signatures is of outmost interest for a panoply of ocean monitoring systems, both for security, for commercial and for research applications. An example of an automated ocean feature detection scheme, able to detect fronts, ice edges and polar lows, is described in [2]. Another example is the Ocean Monitoring Workstation (OMW), developed by the company Satlantic. Usually, the first two steps of the processing chains of such systems are segmentation followed by classification. The Segmentation step computes a set of regions defining an image partition, where the features of each region, for example the gray levels, are *similar* in some sense. The classification focus then on each region attaching a label to it. For example, oil spill detection

based on SAR images has bee approached by many authors with the referred to classification-segmentation scheme(see for example [3]); the SAR image is first segmented and then the classification is focused on the regions with lower scattering; the classifier then computes a number of features from these regions including shape, moments, scale parameters, etc. [3] based on which a decision on whether the region corresponds to oil or look-alike is taken. We should refer, however,

An application where segmentation of oceanic SAR images is important is oil spill detection. In fact, many approaches to this issue have been proposed in recent years. Most of them follow the described segmentation-classification structure, although other methods exist that do not, like for example kernel-based anomaly detectors [4]. A review of SAR segmentation techniques for oil spill detection can be found in [3]. The present work, which is an elaboration of our previous works [5], [6], [7], describes algorithms for the segmentation of dark signatures in oceanic SAR images, following a Bayesian approach. The adopted technique uses as data model a finite Gamma mixture, with a given predefined number of components, for modeling each class density. In fact, this density is well suited to filtered intensity SAR images as shown, *e.g.*, in [8] and in [9]. By using a mixture, we aim at describing the continuous backscattering variability that may be observed in the SAR sea data. Moreover, a mixture is able to describe densities presenting more than one maximum, as it is the case of oceanic multi-look SAR images histograms. This cannot be achieved by none of the statistical models for SAR intensities proposed in the literature, like for example any of the Pearson distributions.

In this work we propose two supervised and one unsupervised algorithm. The supervised algorithms demand an interaction with the user that manually selects a region containing pixels from the dark signature of interest and a region containing water pixels. These regions should be representative and can be made up of different not connected parts. The unsupervised algorithm is an improvement of the supervised ones and is completely automatic.

To estimate the parameters of the class conditional densities, a new expectation maximization (EM) algorithm was developed. Details are described in appendix. When segmenting small sub-scenes of the image, a simplified data model with only one Gamma function per class can be used.

The prior used to impose local homogeneity is a multi-level logistic (MLL) model Markov random field (MRF) [10], with 2nd order neighborhood. To infer the prior smoothness parameter controling the degree of scene homogeneity, we develop an EM algorithm that uses loopy believe propagation (LBP) [11]. We have also exploited different classic estimation

methods, namely the least squares fit (LSF) and the coding method (CD) (see [10] for details) for comparison.

To infer the labels, we adopt the maximum a posterior (MAP) criterion, which we implement efficiently and exactly with graph-cut techniques [12].

Although the segmentation of oceanic dark patches is typically based on the assumption of two classes, we have generalized the problem to an optional number of classes. The underlying integer optimization problem is now attacked with the graph-cut based α-Expansion algorithm [13].

To evaluate the accuracy of the proposed algorithms, different simulations addressing both the referred Gamma model as well as intensity images corrupted with Gaussian noise have been carried out for error rate assessment.

The algorithms have also been applied to real SAR images containing well documented oil spills. For doing so, the scenes are divided in tiles and segmented individually.

A. Related Work

Related approaches to the problem of oil spill segmentation are built on off-the-shelf segmentation algorithms such as the *adaptive image thresholding* and the *hysteresis thresholding*. Entropy methods based on the maximum descriptive length (MDL) and wavelet based approaches have also been proposed. Another recently proposed segmentation methodology applies Hidden Markov Chains (HMC) to a multiscale representation of the original image. Hereby the wavelet coefficients are statistical characterized by the Pearson system and by the the generalized Gaussian family [14]. When other SAR products are available, for example polarimetric data, other methods have been described in the literature, like constant false alarm rate filters [15].

An example of an elaborated adaptive thresholding technique is provided by [16]. In this method, an image pyramid is created by averaging pixels in the original image. From the original image, the next level in the pyramid is created with half the pixel size of the original image. A threshold is then computed for each level based on local estimates of the roughness of the surrounding sea and on a look-up table containing experimental values obtained from a training data set.

Hysteresis thresholding has been used as the base for detecting oil slicks in [17]. The method includes two steps: applying a so-called directional hysteresis thresholding (DHT) and performing the fusion of the DHT responses using a Bayesian operator. The MDL technique, which basically consists in applying information theory in order to find the image description which has the lowest complexity, has been applied in [18] to segment speckled SAR images, namely those containing oil slicks. This segmentation method describes the image as a polygonal grid and determines the number of regions and the location of the nodes that delimit the regions. The two-dimensional wavelet transform, used as a bandpass filter to separate processes at different scales, has also been adopted to oil slick detection in the framework of an algorithm for automated detection and tracking of mesoscale features from satellite imagery. [2].

B. Contributions

We approach oil spill segmentation using a Bayesian framework and a multi-level Logistic (MLL) prior. Several methods in the same vein have been proposed since the seminal work of Geman and Geman [19], see *e.g.*, [20]. Applications of these ideas in the segmentation of SAR images can be found, *e.g.*, in [21], [22], [23].

The main contributions of this work to the state-of-art in oil spill segmentation are the following:

- the development of an EM algorithm to estimate the parameters of a mixture of a pre-defined number of Gamma distributions, in order to model the intensities in a SAR image
- the development of an EM algorithm using LBP to estimate the smoothness parameter in the MRF used as pior in our framework
- the application of recent graph-cut techniques for solving the energy minimization problem that arises from the followed Bayesian methodology.
- the design of a semisupervised algorithm for oil spill segmentation supported on the tools referred to above.

C. Paper Organization

The article is organized as follows: Section 1 introduces the problem, with references to related work and giving the main contributions of the present work; Section 2 overviews the Bayesian methodology that builds the base to the proposed algorithms. In addition, it briefly reviews the concept of the α-Expansion technique that has been implemented to generalize the methodology to an optional number of classes; Section 3 describes the used parameter estimation techniques; Section 4 describes the supervised segmentation algorithms; Section 5 describes the unsupervided algorithm and Section 6 presents results of segmenting simulated and real images applying the algorithms proposed in Section 4 and 5; Finally Section 7 contains the main conclusions and future work remarks. The article also includes an appendix where the referred to EM algorithm, developed to estimate the parameters of the class conditional densities of the Gamma mixture data model, is described.

II. PROBLEM FORMULATION

A. Bayesian Approach

let $\mathcal{L} := \{1, \ldots, c\}$ be a set of c classes and $\mathcal{P} := \{1, 2, \ldots, N\}$ be the set of N pixels (sites) where measurements $y := \{y_1, y_2, \ldots, y_N\}$, the SAR intensities, are available. A labeling $x := \{x_1, x_2, \ldots x_N\}$ is a mapping from $\mathcal{P}$ to $\mathcal{L}$, *i.e.*, it assigns to each pixel $p \in \mathcal{P}$ a label $x_p \in \mathcal{L}$. Any labeling x can be uniquely represented by a partition of image pixels $P = \{P_l | l \in \mathcal{L}\}$, where $P_l = \{p \in \mathcal{P} | x_p = l\}$ is the subset of pixels to which the label l have been assigned. Since there is an one-to-one correspondence between labelings x and partitions P, we use these notions interchangeably. By applying a segmentation algorithm to the image y, we get $\hat{x} := \{\hat{x}_1, \hat{x}_2 \ldots \hat{x}_N\}$, where $\hat{x}_i$ is the inferred label for pixel $i \in \mathcal{P}$.

B. Observation Model

In our problem formulation, we assume conditional independence of the measurements given the labels, *i.e.*,

$$p(y|x) = \prod_{i=1}^{N} p(y_i|x_i) = \prod_{l=1}^{c} \prod_{i \in \mathcal{P}_l} p(y_i|\phi^l), \quad (1)$$

where $p(\cdot|\phi^l)$ is the density corresponding to class l and ϕ_l the correspondent vector of parameters. The adopted density is a finite Gamma mixture given by

$$p(y_i|\phi^l) = \sum_{s=1}^{K} \alpha_s^l \, p(y_i|\theta_s^l), \quad (2)$$

where K is the number of Gamma modes in the mixture, i indexes the pixel, and, for the class l, θ_s^l is the vector of parameters of the Gamma mode s, and α_s^l is the a priori probability of mode s. We denote $\theta^l := (\theta_1^l, \dots, \theta_K^l)$, $\alpha^l := (\alpha_1^l, \dots, \alpha_K^l)$, and $\phi^l := (\alpha^l, \theta^l)$.

Given that $p(y_i|\theta_s^l)$ is Gamma distributed, we have then

$$p\left(y_i|\theta_s^l\right) = \frac{(\lambda_s^l)^{a_s^l}}{\Gamma\left(a_s^l\right)} y_i^{a_s^l - 1} \exp\left(-\lambda_s^l y_i\right), \quad y_i \geq 0, \quad (3)$$

where $\theta_s^l := \left(a_s^l, \lambda_s^l\right)$. The mean and variance of a random variable with the density (3) is, respectively, a_s^l/λ_s^l and $a_s^l/(\lambda_s^l)^2$.

The mixture parameters θ^l are estimated from the data by applying the EM Gamma mixture estimation algorithm described in Appendix. The procedure is the same both for the supervised and for the unsupervised algorithms.

We now make a brief comment on our choice of the Gamma mixture for modeling the observation densities. Under the assumption of fully developed speckle, the complex radar amplitude is zero-mean circular Gaussian distributed [24]. The average intensity computed over a number of independent random variables with the same density is, thus, Gamma distributed. It happens, however, that in sea SAR imaging one or more of the above assumptions may fail, rendering the Gamma density a poor model for SAR intensity [14]. A line of attack to obtain better models is to use more flexible parametric families, such as the Pearson System [25, Ch. 4.1] (see also [9]). However, in our problem each class density exhibits much more variability than that of accommodated by the the Pearson System. We have very often, for example, multi-modal densities. We should resort, therefore, to a mixture of Pearson System densities, what would lead to complex learning procedures. We have experimentally observed, however, that the Gamma mixture yields very good fittings for real SAR histograms, obtained with a moderately complex learning algorithm. For this reason, we have adopted the Gamma mixture model.

C. Prior

A second assumption we are making is of local smoothness of the labels in a statistical sense. It is more likely to have neighboring sites with the same label than the other way around. We model this local smoothness with a second order

MRF, $P(x)$; more specifically, by an MLL model (Ising model in the case of two classes). The Markov property assumes that

$$p\left(x_i| \; x_j, j \in \mathcal{P}\right) = p\left(x_i| \; x_j, j \in \mathcal{N}_i\right), \quad (4)$$

where $\mathcal{N}_i$ is the set of neighbors of site i. If $p\left(x_i| \; x_j, j \in \mathcal{N}_i\right) > 0$, then the Hammersley-Clifford Theorem states that $p(x)$ has the form of the Gibbs distribution

$$p\left(x\right) = \frac{1}{\mathcal{Z}} \exp^{-U(x)}, \quad (5)$$

where $\mathcal{Z}$ is the so called partition function and U is the energy function

$$U\left(x\right) = \sum_{cl \in C} V_{cl}\left(x\right), \quad (6)$$

where C is the set of cliques and $V_{cl}\left(x\right)$ is the clique potential defined over clique cl. In this work, we have pair-wise cliques defined on a second-order neighborhood (8 pixels). That is, $C = \{(i,j) : i \in \mathcal{N}_j, \; j \in \mathcal{N}_i, \; i > j\}$.

In these conditions, the MLL clique potentials, in the isotropic case, is given by

$$V_{cl}\left(x_r, x_s\right) = -\beta \delta\left(x_r - x_s\right), \quad (7)$$

where $r, s \in cl$, $\delta(x) := I_0(x)$ is the indicator function of set $\{0\}$, and parameter $\beta > 0$ controls the degree of scene homogeneity.

D. Maximum a Posteriori Estimate

The posterior of the labeling given the observed data is

$$p\left(x|y\right) \propto p\left(y|x\right) p\left(x\right). \quad (8)$$

In order to infer x, we adopt the MAP criterion. This amounts to maximize the posterior density of the labeling given the observed data:

$$\hat{x} = \arg\max_x p\left(x|y\right), \quad (9)$$

and is equivalent to minimize the negative logarithm of (8). In this sense, we may rewrite the problem in the following way:

$$\hat{x} = \arg\min_x E(x_1, \dots, x_n), \quad (10)$$

where

$$E\left(x_1, \dots, x_N\right) = -\log p(x|y) + c^{te}, \quad (11)$$

and c^{te} denotes an irrelevant constant. From equation (8), we have

$$E\left(x_1, \dots, x_N\right) = \sum_{p=1}^{N} E^i\left(x_i\right) + \sum_{i,j \in cl} E^{i,j}\left(x_i, x_j\right), \quad (12)$$

with

$$E^i\left(x_i\right) = -\log p\left(y_i|x_i\right) \quad (13)$$
$$E^{i,j}(x_i, x_j) = -\beta \delta(x_i - x_j). \quad (14)$$

E. Energy Minimization

As already stated, we are concerned with the minimization of $E(x_1, \ldots x_N)$ given by (12), which we term energy. For two classes, (*i.e.*, $c = 2$), the global minimum of $E(x_1, \ldots x_N)$ can be computed exactly by applying the graph-cut algorithm described in [12]. This is a consequence of energy being graph-representable, *i.e.*, $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0)$ (for details see [12]). For more than two classes, the solution of (9) can be approximately computed by the α-Expansion technique [13], also based on graph-cut concepts. This algorithm finds the local minimum of the energy within a known factor of the global minimum.

We now give a brief description of the α-Expansion algorithm. Given a label α, a move from a partition P (with correspondent labeling x) to a new partition P' (with correspondent labeling x') is called an α-Expansion if $P_\alpha \subset P'_\alpha$ and $P'_l \subset P_l$ for any label $l \neq \alpha$. In other words, an α-Expansion move allows any set of image pixels to change their label to α. The algorithm cycles through the labels in $\mathcal{P}$ in some fixed or random order and finds the lowest α-Expansion move from the current labeling. If the expansion move has lower energy than the current labeling, then it becomes the current labeling. The algorithm terminates when no α-Expansion move exist corresponding to a local minimum of the energy.

III. ESTIMATION OF PARAMETER β

In this work, we consider three different techniques to determine the smoothness parameter β: a new EM method hereafter introduced that uses loopy belief propagation (LBP) and the classical LSF and CD methods. Because LSF and CD assume the existence of labeled data, we have conceived an iterative labeling-estimation scheme, which alternates between a labeling step and an estimation step until the convergence of β is attained. On the contrary, the EM estimation algorithm, that we have called "Loopy-β-Estimation", is a one-shot technique. In the following sections, we briefly review the LSF and CD methods and provide a detailed description of the Loopy-β-Estimation method. In this section, the class parameters $\phi := (\phi^1, \phi^2, \ldots, \phi^c)$ are assumed known.

A. Least Squares Fit

This procedure for parameter estimation in MRF, described in detail in [10], is based in the following equation that holds for the MLL model, for every pixel p, with neighborhood $\mathcal{N}_p$, and for every label pair $x_p, x'_p \in \mathcal{L}$:

$$
\begin{aligned}
\beta \left[n(x_p) - n(x'_p) \right] &= \log \left(\frac{p(x_p | x_{\mathcal{N}_p}, y)}{p(x'_p | x_{\mathcal{N}_p}, y)} \right) \\
&\quad - \left(E^p(x'_p) - E^p(x_p) \right),
\end{aligned} \tag{15}
$$

where $n(x_p)$ is the number of pixels in the set $\mathcal{N}_p$ with the same label as x_p and $x_{\mathcal{N}_p} := \{x_i, i \in \mathcal{N}_p\}$. We use histograms to estimate the joint probabilities $p(x_p | x_{\mathcal{N}_p}, y)$ and $p(x'_p | x_{\mathcal{N}_p}, y)$ in (15): assuming that there are a total of M distinct 3x3 blocks in the image lattice with a given label configuration $x_{\mathcal{N}_p}$, then we take

$$
\widehat{p}(x_p | x_{\mathcal{N}_p}, y) = \frac{\mathcal{H}(x_p | x_{\mathcal{N}_p}, y)}{M}, \tag{16}
$$

where $\mathcal{H}(x_p | x_{\mathcal{N}_p}, y)$ is the number of times that a particular 3x3 configuration $(x_p | x_{\mathcal{N}_p}, y)$ occurs. The expression is then evaluated for a number of distinct combinations of x_p, x'_p and $x_{\mathcal{N}_p}$ in order to obtain an over-determined linear system of equations that is solved in order to β.

B. Coding Method

In this method the key idea (see [10]) is to partition the set $\mathcal{P}$ into sets $\mathcal{P}^{(k)}$, called codings, such that no two sites in one $\mathcal{P}^{(k)}$ are neighbors. In the present work, the neighborhood $\mathcal{N}_p$ is of 2nd order thus yielding four codings. As the pixels in $\mathcal{P}^{(k)}$ are not neighbors, the variables associated with these pixels, given the labels at all other pixels, are mutually independent under the Markovian assumption. The following simple product is thus obtained for the likelihood:

$$
p^{(k)}(x | \beta, y) = \prod_{p \in \mathcal{P}^{(k)}} p(x_p | x_{\mathcal{N}_p}, \beta, y), \tag{17}
$$

Maximizing (17) in order to β gives the coding estimate $\hat{\beta}^{(k)}$. Although it is not clear how to combine the results optimally, the arithmetic average, as suggested in [10], is an intuitive scheme that was adopted in this work.

C. Loopy-β-Estimation

We are seeking $\hat{\beta}_{ML} = \arg\max_\beta p(y|\beta)$, the ML estimate of the smoothness parameter β. Based on the fact that the marginal density $p(y|\beta)$, the so-called evidence, is a sum over the missing labels x, *i.e.*,

$$
\begin{aligned}
p(y|\beta) &\equiv \sum_x p(y, x|\beta) \\
&= \sum_x p(y|x) p(x|\beta),
\end{aligned} \tag{18}
$$

we develop an EM algorithm [26] for the ML estimation of the parameter β. The EM algorithm alternates between two steps: the E-step computes the conditional expectation of the logarithm of the complete a posteriori probability function, with respect to the missing variables, based on the actual parameter value; the M-step updates the value of the parameter, by maximizing the expression obtained in the E-step with respect to that parameter. We now derive the E-step and the M-step.

E-step:

$$
\begin{aligned}
Q(\beta; \beta_t) &= E[\log p(y, x|\beta)|y, \beta_t] \tag{19} \\
&= E[\log p(y|x)|y, \beta_t] \tag{20} \\
&\quad + E[\log p(x|\beta)|y, \beta_t]. \tag{21}
\end{aligned}
$$

Recalling that the MLL prior is given by

$$
p(x|\beta) = \frac{1}{\mathcal{Z}(\beta)} \exp \left[\beta \sum_{i,j \in cl} \delta(x_i - x_j) \right], \tag{22}
$$

with

$$
\mathcal{Z}(\beta) = \sum_x \exp \left[\beta \sum_{i,j \in cl} \delta(x_i - x_j) \right], \tag{23}
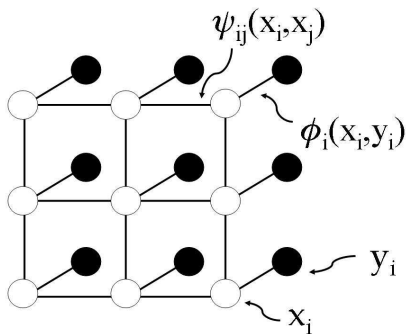$$

Fig. 1. Lattice representing the pairwise MRF.

we obtain, up to an irrelevant constant,

$$Q\left(\beta;\beta_t\right) = -\log \mathcal{Z}\left(\beta\right) + \beta \sum_{i,j \in cl} E\left[\delta\left(x_i - x_j\right) \mid y, \beta_t\right].$$

(24)

M-step:

$$\beta_{t+1} = \arg\max_{\beta} Q\left(\beta, \beta_t\right).$$

(25)

The stationary points of $\mathcal{Q}$ are the solution of

$$\frac{\partial Q}{\partial \beta} = -\frac{\partial \log\left[\mathcal{Z}\left(\beta\right)\right]}{\partial \beta}$$

(26)

$$+ \sum_{i,j \in cl} \sum_{x_i, x_j \in \mathcal{L}} \delta\left(x_i - x_j\right) p\left(x_i, x_j \mid y, \beta_t\right)) = 0.$$

By introducing expression (23) into (26) and if we consider that $\delta\left(x_i - x_j\right)$ takes non-zero values only for equal labels, we obtain

$$\frac{\partial Q}{\partial \beta} = \sum_{i,j \in cl} \sum_{k=1}^{c} p\left(x_i = k, x_j = k \mid y, \beta_t\right)$$

$$- p\left(x_i = k, x_j = k \mid \beta_t\right) = 0.$$

(27)

Since computing exact marginal distributions is infeasible in our case, we replace them by pseudo-marginals using BP. This approach has been successfully applied in problems of approximate parameter learning in discriminative fields [27]. BP is an efficient iterative algorithm in which local messages are passed in graphical models. For singly-connected (loop free) pairwise MRFs, the two-node beliefs will correspond to the exact two-node marginal probabilities. In our case, however, the graph that corresponds to the MRF contains cycles, preventing the basic BP algorithm to be applied. We resort to a slightly modified version called loopy belief propagation. In practice, this algorithm has often delivered good results [11]. We approximate the marginal probabilities with the two-node beliefs $b_{ij}\left(x_i, x_j\right)$ and $b_{ij}^y\left(x_i, x_j\right)$. These will provide approximations respectively for the marginals $p\left(x_i, x_j \mid \beta_t\right)$ and $p\left(x_i, x_j \mid y, \beta_t\right)$. By doing so, the M-Step of our EM algorithm is given by the sum of the differences between the two node beliefs that take the evidence y into consideration and those that do not make use of y:

$$\frac{\partial Q}{\partial \beta} = \sum_{i,j \in cl} \sum_{k=1}^{c} b_{ij}^y\left(k, k\right) - b_{ij}\left(k, k\right) = 0.$$

(28)

Fig. 1 depicts the graph that represents our pairwise MRF for computing the two-node beliefs. In the square lattice, $\psi_{ij}\left(x_i, x_j\right)$ stands for the interaction potential that penalizes every dissimilar pair of neighboring labels, and $\phi\left(x_i, y_i\right) = p\left(y_i \mid x_i\right)$ represents the statistical dependency between the labels x_i and the measurements y_i. For computing $b_{ij}\left(x_i, x_j\right)$, $\phi\left(x_i, y_i\right)$, is set to a constant value, making the result independent of the data values y.

We solve (28) by a line search type algorithm, ensuring that $\frac{\partial^2 Q}{\partial^2 \beta} \leq 0$, thus corresponding to a maximum of Q.

D. A few Remarks about the Vector of Parametres ϕ

In the previous section, we have considered that the class parameter vector ϕ is known and only the smoothness parameter β is to be inferred. However, still using the beliefs computed by LBP, the vector ϕ could have been inferred simultaneously with β by including in the Q function the additive term (20) corresponding to the class densities. We would have obtained

$$Q(\phi; \phi_t, \beta_t) = E[\log p(y|x)|y, \phi_t, \beta_t]$$

$$= \sum_{i=1}^{N} \sum_{l=1}^{c} \log p(y_i|\phi^l) p(x_i = l|y, \phi_t, \beta_t),$$

where the probabilities $p(x_i = l|y, \phi_t, \beta_t)$ are given by the LPB method. The **M-step** would consist then in two decoupled maximizations; one with respect to β and another with respect to ϕ. This approach is, however, beyond the scope of this paper. Nevertheless, we present in Section V an unsupervised algorithm, which is suboptimal but faster than the LBP based EM algorithm aimed at the inference of both the parameters β and ϕ.

An alternative to the proposed EM scheme based on LPB is using Monte Carlo techniques to cope with the difficulty in computing the **E-step** and the partition function Z [28], [29], [30]. Supported on the performance of the LBP, we believe, however, that the EM scheme based on LPB is, for the present problem, much more faster than the Monte Carlo based techniques.

IV. SUPERVISED SEGMENTATION

In this section, we introduce two supervised algorithms aimed at the segmentation of sea SAR images. In both algorithms, the first step is the estimation of the class parameters used in the data model. This is done by asking the user to define representative regions of interest (ROI) in the image. Once the regions are defined, different approaches may be followed: if a Gamma mixture is assumed for the observed data, as in the case of SAR images, the EM class parameters estimation algorithm described in Appendix is applied to infer the class conditional densities. If we are segmenting small sub-scenes and the radar range spreading loss has been compensated in the underlying SAR images, one single Gamma function per class often provides a good modeling for the SAR intensities. In this case, a common ML Gamma estimator is used instead of the EM procedure. After this step, the data model is considered to be known and is used thereafter.

The pseudo-code for two supervised algorithms is presented in Algorithm 1 and Algorithm 2. Algorithm 1 is of generalized likelihood type [31] implementing an iterative labeling scheme with two steps being performed alternately: the β-*Estimation* and the *segmentation*. Algorithm 2 is a one-shot procedure that performs β-Estimation using the Loopy β-Estimation method described in Section III-C.

Algorithm 1 Supervised Segmentation Using LSF/CD β-Estimation

---

Require: Initial parameter $\hat{\beta} = \beta_0$ and estimated class parameters $\hat{\phi}$

1: Compute $eLabel = E^p(x_p)$ for every pixel p using $\hat{\phi}$
2: **while** $\left|\Delta\hat{\beta}\right| \le \delta$ or nrIterations $<$ NrIterationsMax **do**
3:     Compute $\hat{x} = \alpha$-Expansion$(\hat{\beta}, eLabel)$
4:     Compute $\hat{\beta} = \beta$-Estimation$(\hat{x}, eLabel)$ {LSF or CD}.
5: **end while**
6: **return** $\left(\hat{x}, \hat{\beta}\right)$

Algorithm 2 Supervised Segmentation Using Loopy-β-Estimation

Require: estimated class parameters $\hat{\phi}$

1: Compute $eLabel = E^p(x_p)$ for every pixel p using $\hat{\phi}$
2: Compute $\hat{\beta} = $ Loopy-β-Estimation$(eLabel)$
3: Compute $\hat{x} = $ GraphCut-Segmentation$(\hat{\beta}, eLabel)$
4: **return** $\left(\hat{x}, \hat{\beta}\right)$

V. UNSUPERVISED SEGMENTATION

The unsupervised method is an improvement of the supervised algorithms described in the previous sections. The main difference is that the data model is not considered to be known but is also iteratively estimated along with the smoothness parameter and the segmentation. The scheme needs a rough initialization of the data model parameters. We computed this initialization by fitting an EM Gamma mixture to the complete data.

As our EM algorithm automatically eliminates unnecessary modes, we start with an overestimate of K, the number of modes in the mixtures. In our experiments with real data, the maximum number of modes we got was four.

Then, different strategies are possible. In the experiments reported in the next section, when considering $c = 2$ classes, we have assigned the mode with a lower mean value to one of the classes and the remaining modes to the other class.

In each iteration, we compute the ML estimate of the vector ϕ based on the previous segmentation, compute the new segmentation, and finally the new value of β. The algorithm may use any of the three $\beta-$parameter estimation methods and is applicable to an optional number of classes. The algorithm stops when both the parameter β and the class parameters ϕ converge to stable values.

Algorithm 3 Unsupervised Segmentation

Require: arbitrary parameter $\hat{\beta} = \beta_0$, initial class parameters $\hat{\phi} = \phi_0$ (EM Gamma Mixture Estimation)

1: Compute $eLabel_0 = E^p(x_p)$ for every pixel p, using ϕ_0
2: Compute initial labeling
    $\hat{x} = x_0 = \alpha$-Expansion$(\beta_0, eLabel_0)$
3: **for** stop criterium is not met **do**
4:     Compute $\hat{\phi} = $ ML-Estimation$(\hat{x})$
5:     Compute $eLabel = E^p(x_p)$ for every pixel p, using $\hat{\phi}$
6:     Compute $\hat{x} = \alpha$-Expansion$(\hat{\beta}, eLabel)$
7:     Compute $\hat{\beta} = \beta$-Estimation$(\hat{x}, eLabel)$ {use LSF, CD or Loopy}.
8: **end for**
9: **return** $\left(\hat{x}, \hat{\beta}, \hat{\phi}\right)$

VI. RESULTS

This section presents results of applying the proposed methodology to simulated and real SAR images, as well as some considerations regarding time complexity of the algorithms.

A. Simulations

We have performed simulations corresponding to Gamma data terms with two classes, and evaluated the overall accuracies of the obtained segmentations using Algorithms 1, 2 and 3. We also illustrate in detail the EM Gamma mixture estimation algorithm, by applying it to an example of simulated data.

1) Segmentation Results with Gamma Data Term: In order to compare the performance of the three proposed methods, we have tested Algorithms 1, 2 and 3 on simulated images generated by adding Gamma noise to ground-truth images containing two classes. We have used the ground-truth depicted in Fig. 2 (a) and added noise with Gamma distributions having mean values of five and nine. The parameters of the distributions were choosen in order to obtain increasing values of variance σ^2, corresponding to noisier images. We have then applied Algorithm 1, with LSF and CD methods, Algorithm 2, and Algorithm 3 with Loopy-β-Estimation. Furthermore, for comparison, the images were also segmented tuning the beta value manually and for the case that no prior is used ($\beta = 0$). Fig. 2 shows the segmentations obtained for $\sigma = 2.6$. As we can see from $\beta = 0$ this is a hard problem, on which LSF and CD fails the inner structures, but Loopy-β-Estimation provides better results, both in the supervised as in the unsupervised method. This behavior is further confirmed by the rank in Fig. 3, which gives the OA obtained for images with different σ values (corresponding to images with more noise) for the six different segmentation processes referred above.

We have also tested the proposed algorithms in images generated by adding Gamma noise to a ground-truth similar to an oil patch (see Fig. 4). Two images, Image A and Image B, with the histograms depicted in Fig. 4, corresponding to different segmentation difficulty levels are segmented. The best segmentation possible in this framework, achieved by tuning
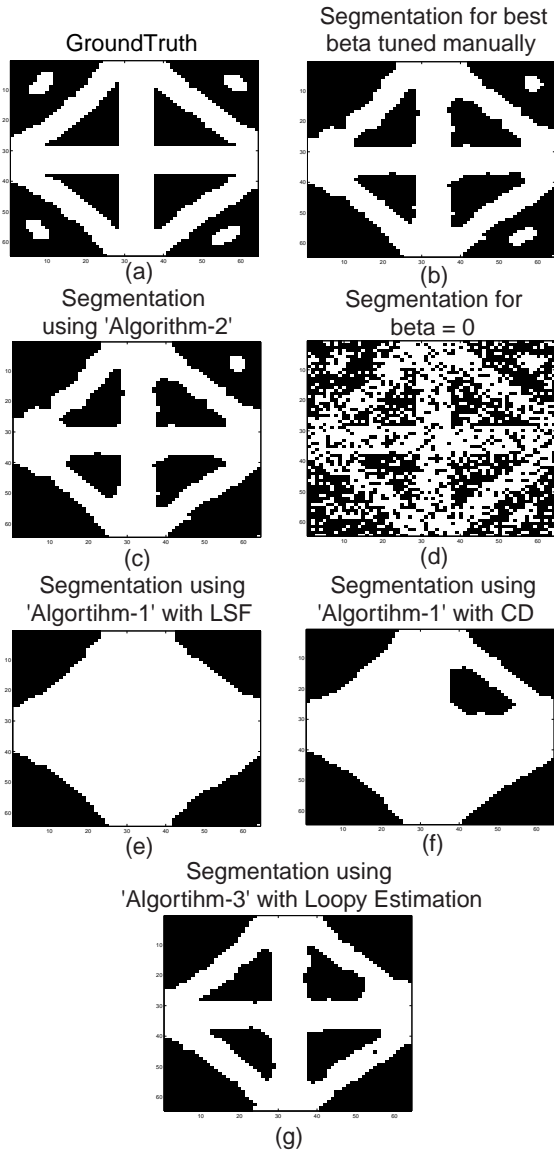
GroundTruth

(a)

Segmentation for best
beta tuned manually

(b)

Segmentation
using 'Algorithm-2'

(c)

Segmentation for
beta = 0

(d)

Segmentation using
'Algortihm-1' with LSF

(e)

Segmentation using
'Algortihm-1' with CD

(f)

Segmentation using
'Algortihm-3' with Loopy Estimation

(g)

Fig. 2. Ground-truth and results of different segmentation processes for an image with Gamma noise having mean values of five and nine and with a σ value = 2.6. Notice the good performance of Algorithm 2 and Algorithm 3, implementing the Loopy-β-Estimation.

the β value manually, and the segmentation obtained with no prior information (setting $\beta = 0$) are also displayed for comparison. Fig. 5 shows the results obtained for Image A and Fig. 6 the results obtained for Image B. For Image A, the segmented images using Algorithm 1, with CD and with LSF, are not shown, as they are almost equal to the image segmented with Algorithm 2. For Image B, Algorithm 3 has not provided good results and the segmentation is not displayed. The bad performance of Algorithm 3 in this case arises from the not so good initialisation of the class parameters, due to the complete overlapping from the oil and the water histogramms.

2) EM Algorithm for Gamma Mixture: In this subsection we illustrate the behavior of the EM algorithm designed to infer Gamma mixtures. For details on the theoretical issues, we refer to the Appendix. In Fig. 7, we can see the ground-
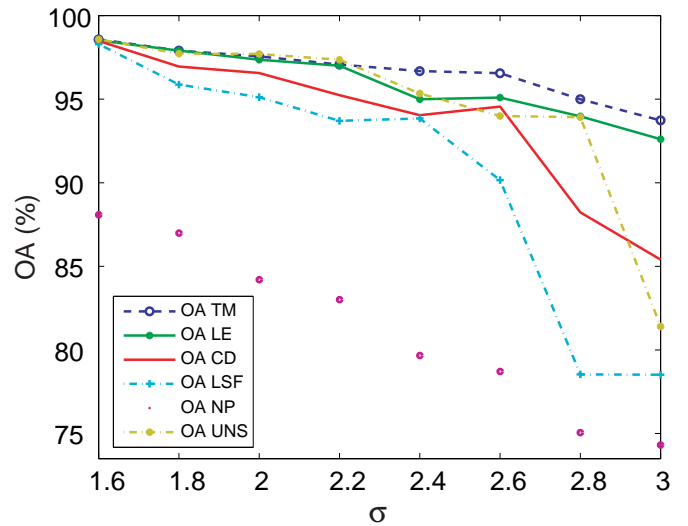
Fig. 3. OA for images corrupted with Gamma noise with increasing σ values: TM $\equiv$ for β tuned manually; LE $\equiv$ Algorithm 2, using Loopy Estimation; LSF $\equiv$ Algorithm 1 using LSF; CD $\equiv$ Algorithm 1 using CD, NP $\equiv$ no prior and UNS $\equiv$ Algorithm 3 using Loopy Estimation.

(a)

(b)

(c)

(d)

(e)
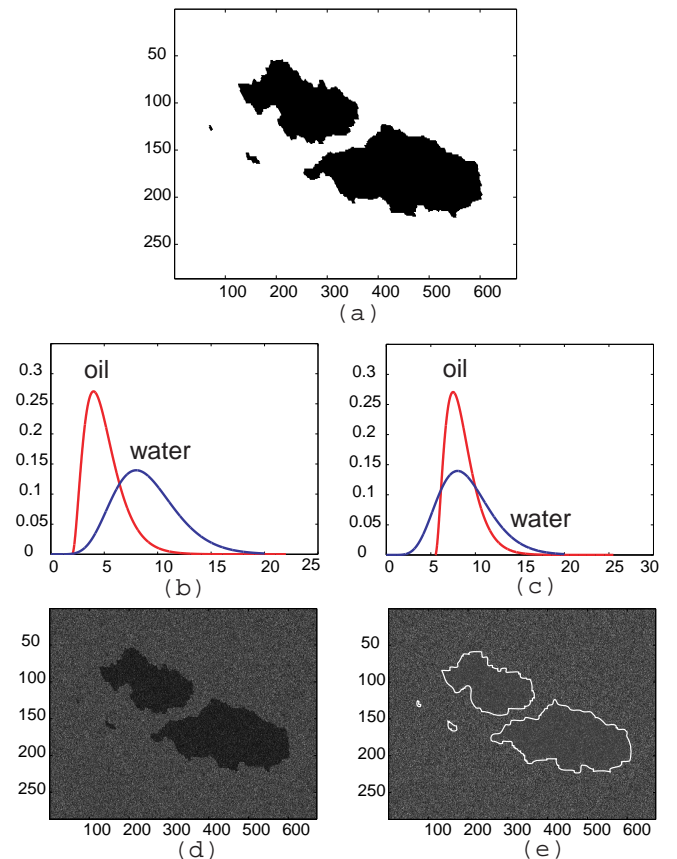
Fig. 4. Gamma Model: (a) Ground-truth, (b) Histogram of Image A, (c) Histogram of Image B, (d) Image A, (e) Image B with superimposed delimiting line, for better visualization
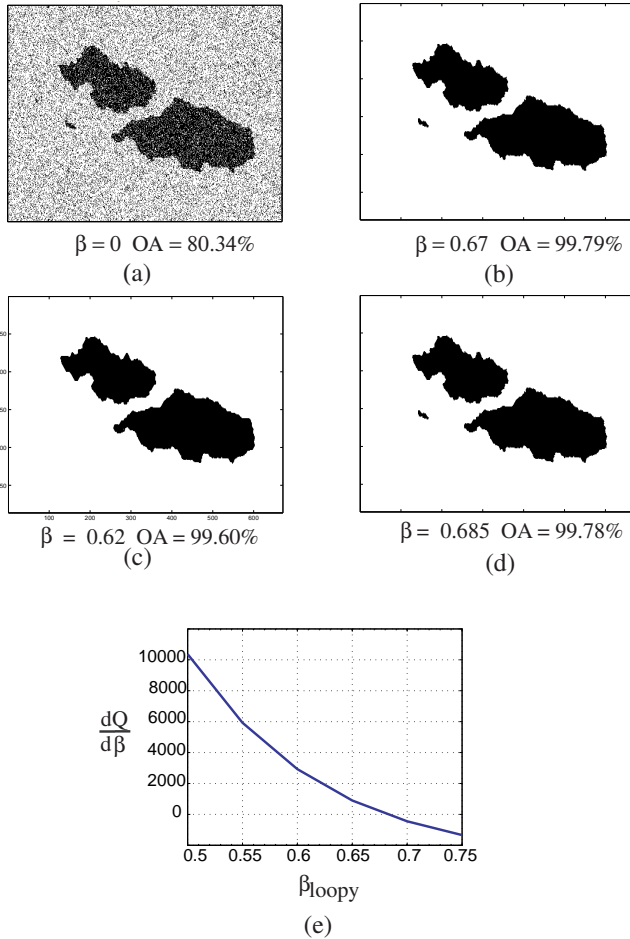
$\beta = 0$ OA = 80.34%
(a)

$\beta = 0.67$ OA = 99.79%
(b)

$\beta = 0.62$ OA = 99.60%
(c)

$\beta = 0.685$ OA = 99.78%
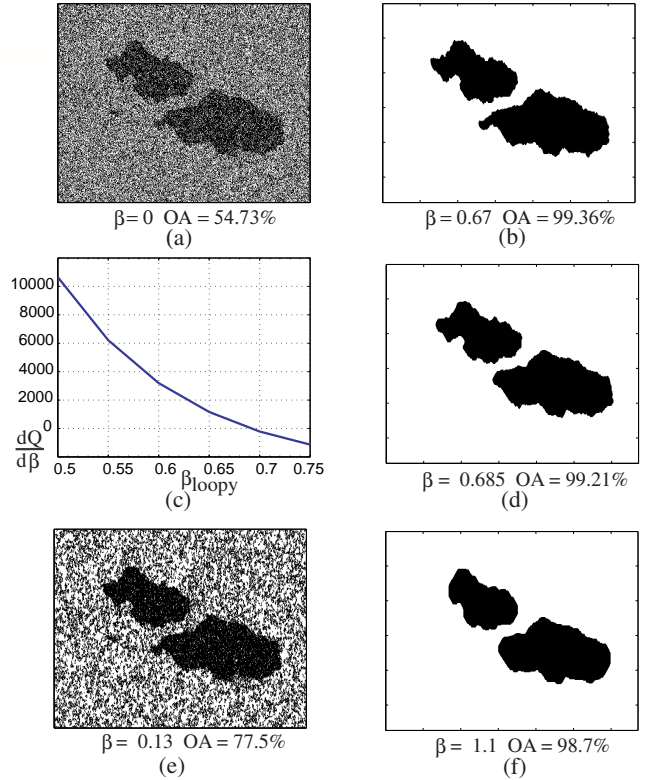(d)

$\dfrac{dQ}{d\beta}$

β_{loopy}

(e)

Fig. 5.   Segmentation results for image A: (a) No prior information, (b) β value tuned manually, (c) Segmentation using Algorithm 3 with Loopy-β-Estimation, (d) Segmentation using Algorithm 2 (Algorithm 1 provided the same results with only 0.1% difference in OA), (e) graph depicting the β estimation in Algorithm 2.

$\beta = 0$ OA = 54.73%
(a)

$\beta = 0.67$ OA = 99.36%
(b)

$\dfrac{dQ}{d\beta}$ β_{loopy}
(c)

$\beta = 0.685$ OA = 99.21%
(d)

$\beta = 0.13$ OA = 77.5%
(e)

$\beta = 1.1$ OA = 98.7%
(f)

Fig. 6.   Segmentation results for Image B: (a) No prior information, (b) β value tuned manually, (c) graph depicting β estimation using Algorithm 2, (d) Segmentation using Algorithm 2, (e) Segmentation using Algorithm 1 with LSF and (f) Segmentation using Algorithm 1 with CD.
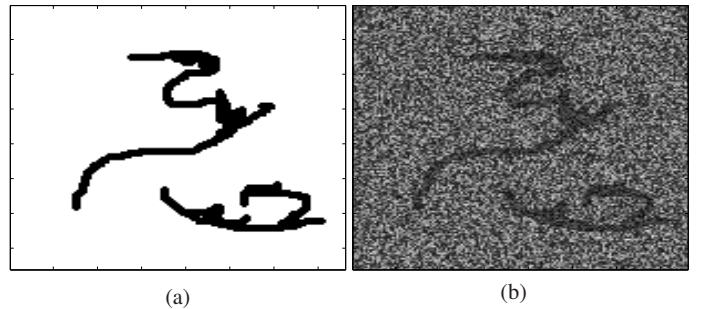

(a)   (b)

Fig. 7.   (a) Ground-truth used for simulating an oil spill. Black represents oil and white water; (b)Simulated SAR image

truth used for simulating the image and the generated image, according to the densities shown in Fig. 8. The two classes, oil and water have been modeled respectively by a mixture of two Gamma functions and a mixture of three Gamma functions. For this particular case, we have selected a ROI containing 178 pixels for representing the water and a ROI containing 142 pixels for representing the oil. After 20 iterations of the EM algorithm, we obtained the approximations for the probability distributions depicted in Fig. 9 and Fig. 10. The obtained results are quite reasonable for the small sample sizes used.

B. Real Images

We have tested the proposed methodology with three real SAR images for the special purpose of oil spill detection. For this type of application, an unsupervised algorithm is more indicated, and so we decided to apply Algorithm 3. Because Loopy Estimation has proven to be an effective method in the simulations, we have choosen Algorithm 3 with Loopy Estimation. Nevertheless, in order to have a comparision unsupervised approach versus supervised approach, we also applied Algorithm 2 to two of the images. In oil spill detection,

the number of classes is typically set to $c = 2$, although more classes can be considered if we are interested in distinguishing other phenomena occuring at the same time in the region of interest. The application of the algorithms is straightforward: the key idea, like in most state-of-the-art oil spill detection methods (see for example [16]), is to partition the image in tiles and run the algorithm separately for each part. This step is preceeded by the application of a landmask to the image, what can be done using external coastline information or by adopting some coastline self-extraction procedure. After the segmentation of each tile, a procedure for grouping patches detected on the tile borders should be carried on. Another possibility to increase segmentation coherency in the borders is to define overlapping tiles or to force continuity to some
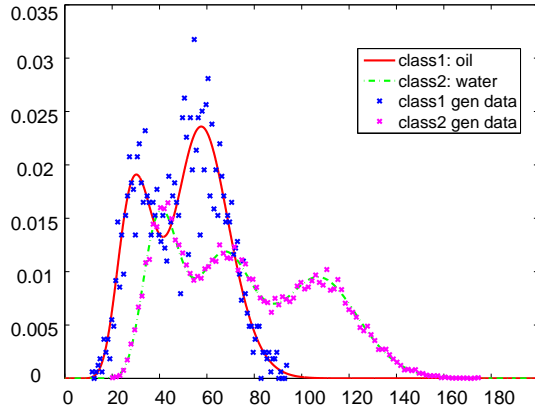
Fig. 8. Probability functions used to generate the simulated image with superimposed histogram of generated data set. A three modes function for water and two modes function for oil was used.
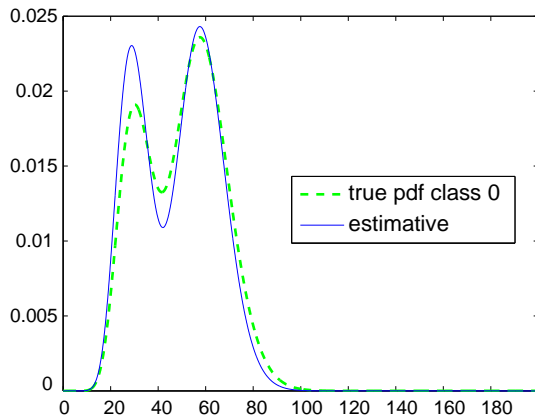

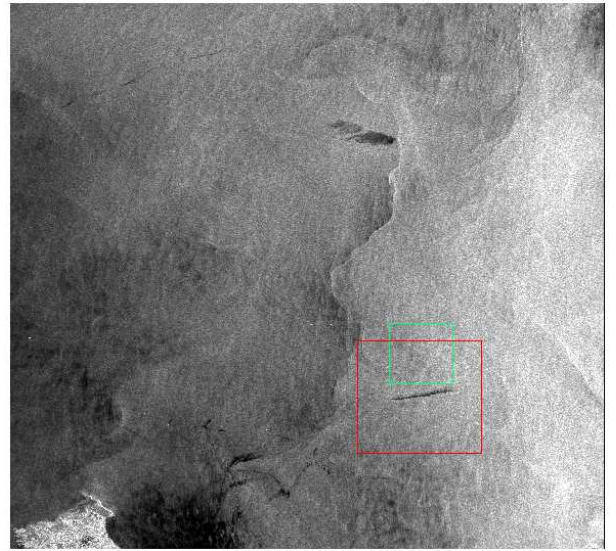Fig. 11. ERS-1 image from the Sicily Channel, Italy, acquired on 30th January of 1992. The smaller and larger squares are sub-scenes that have been used respectively to estimate the wind direction and to apply our algorithm.
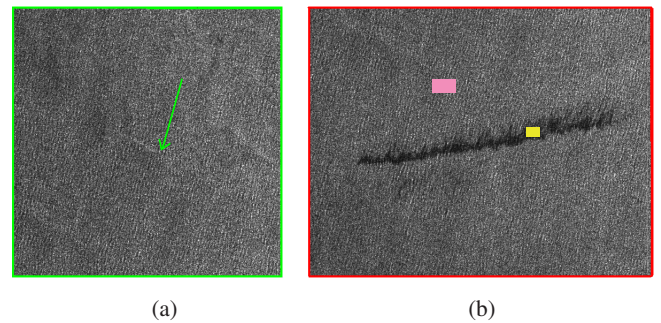

Fig. 9. True and estimated class densities for oil.

(a)   (b)

Fig. 12. Closer look to the sub-scenes of the ERS-1 image. Part (a) is used for wind direction estimation; the estimated direction has been overlapped as a green arrow. Part (b) contains an oil spill. The coloured regions correspond to the ROI's selected for class parameter estimation in the supervised algorithm

Fig. 10. True and estimated class densities for water.

degree on the estimated class and/or smoothness parameters from one tile to the next. Nevertheless, at the moment we are not doing this and these are considered future possible improvements to our methodology. We segmented oil spills contained in three different scenes, described in the following subsections.

1) Segmentation of an ERS-1 sub-scene: We have segmented part of an ERS-1 image from the Sicily Channel, Italy, that has been acquired on the 30th January of 1992. This image is referred in the ESA web pages regarding oil slicks http://earth.esa.int/ew/oil_slicks/ and contains three oil slicks, along with information regarding wind direction and intensity and existence of ships, ship wakes, natural oil films and currents.

Fig. 11 provides a quicklook of the scene with two squares superimposed: the larger representing the part to be segmented and the smaller representing a part used for wind estiamtion. Fig. 12 provides zooms of the referred squares. By applying the Radon Transform to the smaller square, the estimated direction has been calculated and is depicted in the image. The direction was consistent with the measured value reported in
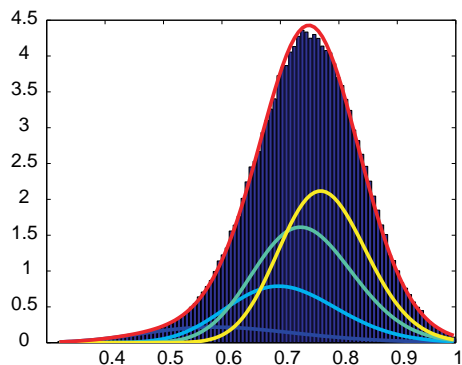
Fig. 13. Fitting of a mixture of Gammas to the data from the red square in the ERS-1 image.
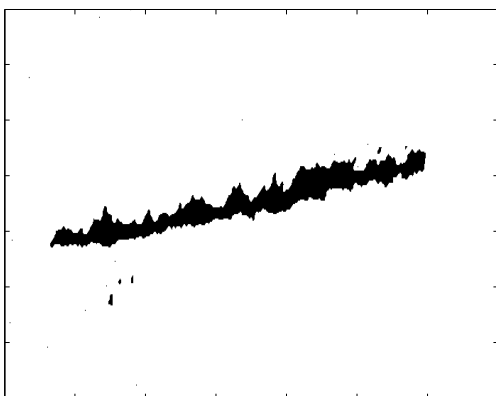

Fig. 15. Class Parameters Estimation for Algorithm-2.

Fig. 14. Segmentation of ERS-1 subscene containing oil with $\beta = 1.4$, estimated using Algorithm-3 with the Loopy-β-Estimation method.

Fig. 16. Segmentation of ERS-1 subscene containing oil with $\beta = 1.44$, estimated using Algorithm-2.

the url site and is at the origin of the well-known "feathering" effect that can be observed in this linear spill. We have computed the backscattering values of the image, by performing calibration using the ESA provided BEST software (`http://earth.esa.int/services/best/`) and then applied Algorithm-3 with Loopy-β-Estimation. The result of the Gamma mixture estimation, in the initialisation of the algorithm, is depicted in Fig. 13. After only three iterations, both the class parameters and the smoothness parameter have converged. A β value equal to 1.4 was estimated. The segmentation is displayed in Fig. 14. We also applied Algorithm-2: we selected two ROI's (160 pixels for water and 77 for oil) in the image (shown in Fig. 12) and computed ML Gamma Estimators for the two classes (see Fig. 15). We then applied the algorithm and obtained an estimated β value equal to 1.44. The segmentation result is given in Fig. 16 and is similar to the result obtained with Algorithm-3.

For comparison, we also provide the segmentation obtained with no prior (corresponding to $\beta = 0$) in Fig. 17. In practice, the estimated value has proved to deliver a good segmentation. When lower β values were used, the result was noisier and for higher β values, the details of the spill disappeared.

2) Segmentation of an Envisat ASAR IM sub-scene: We have also applied Algorithm 3 to a fragment of an
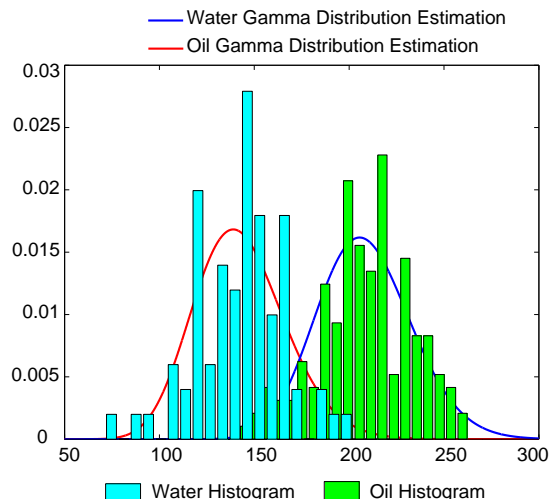

Fig. 17. Segmentation of ERS-1 subscene containing oil, with $\beta = 0$.
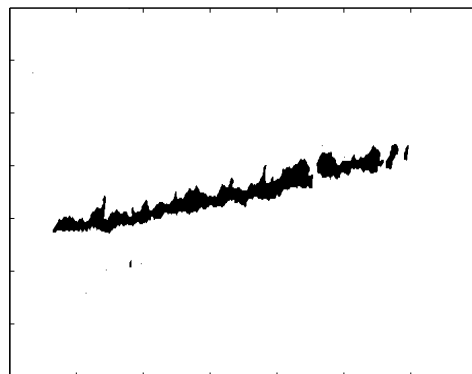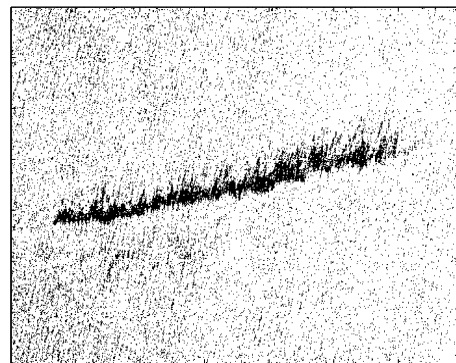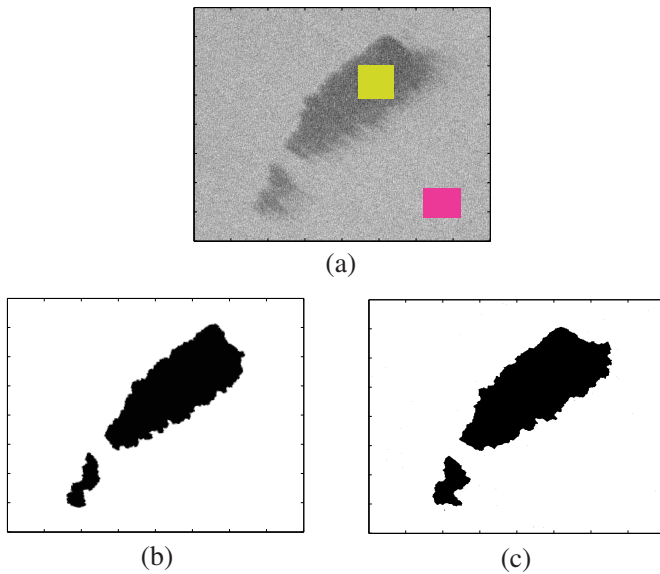
Fig. 18. (a) ASAR image fragment with coloured regions for class parameter estimation in Algorithm-2; (b) Segmentation result applying Algorithm-2; (c) Segmentation result applying Algorithm-3 with Loopy-β-Estimation.

Fig. 19. Display of the ASA WSM full resolution image of the Prestige accident occured in November 2002 in Galicia.

ASAR Image Mode image, acquired on 19 July 2004, in the ocean between Cyprus and Lebanon. The fragment contained an occurred oil spill of circa 10 km's (centered on $\approx$ 33°N, 33°E39') that was documented on the EC Oceanides project (a project in the framework of the Europe 's Global Monitoring for Environment and Security initiative) database. After six iterations, we achieved convergence of the β and θ values. The segmentation result is given in Fig. 18, corresponding to a estimated $\beta = 1.83$. The unsupervised segmentation result is compared with the one provided by Algorithm-2, where the user provided ROI's for water and for oil that were used to estimate the class parameters at the beginning of the process (see Fig. 18). In this case a $\beta = 1.75$ was obtained.

3) Segmentation of an Envisat ASAR WSM sub-scene: In order to demonstrate the viability of applying Algorithm 3 to a whole ASAR_WSM scene, we have run it over the very well known image of a confirmed oil spill, namely the Prestige case (see Fig. 19). This accident took place in November 2002, in Galicia (Spain), when a tanker carrying more than 20 million gallons (around 67,000 tons) of oil split in half off the northwest coast of Spain on 19 November 2002, threatening one of the worst environmental disasters in history. For segmenting the image, we have first partitioned it in tiles of 600x600 pixels each and then applied Algorithm-3 to each tile independently. No post-processing of the borders or "clean-up" operations (like for example morphological operations) have been carried out. Fig. 20 shows the segmented image, made up of the concatenation of the segmentation of the individual tiles. As we can see, the result can be considered in general very good, with only some discrepancies located on the tiles' borders. For initialising the algorithm, the EM Gamma mixture estimation procedure was applied with four modes ($K = 4$). In fact this number is enough to provide a good fitting of the
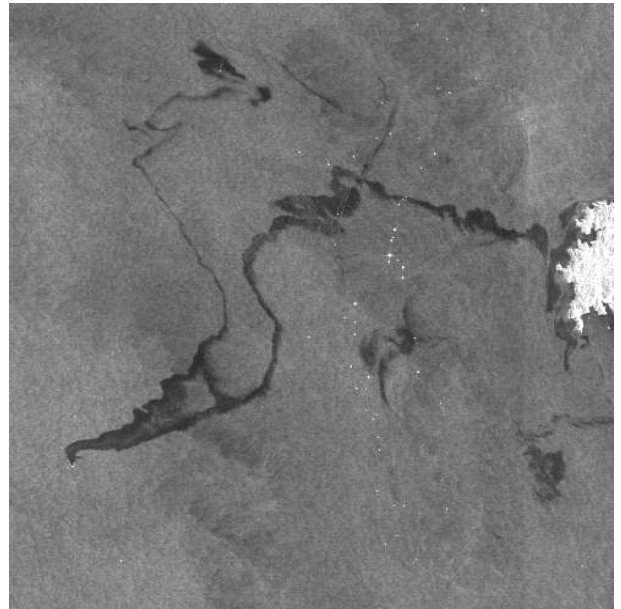

Fig. 20. Display of the segmentation results of applying Algorithm 3 with Loopy-β-Estimation to the image displayed in Fig. 19.

intensity data of the SAR image. On the other hand, our EM algorithm allows us to start with a higher mode number, and decreases this number automatically. We depict the results of this fitting on one of the tiles, shown in Figure 23, by showing the histogram of the tile with the superimposed estimated Gamma mixture (see Fig. 21 (a)). As we have explained, the mode corresponding to the lower mean value is assigned to the oil class and the others to the water class, providing an initialisation to the class parameters. The estimations of these parameters are actualised along the algorithm and, after nine iterations, we obtain the distributions shown in Fig. 21 (b).

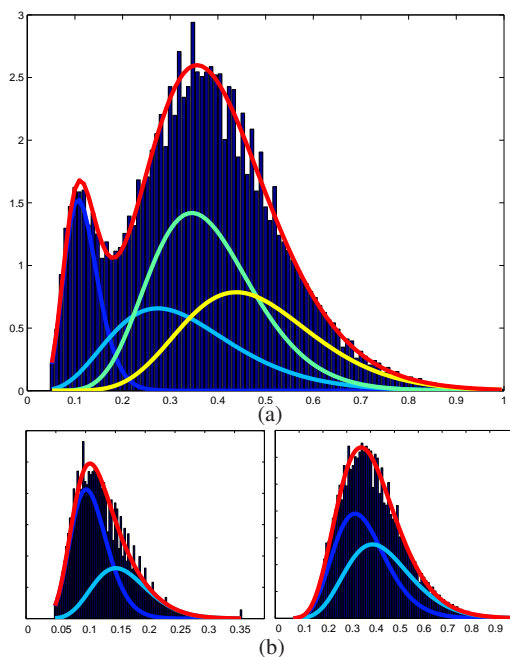To fully demonstrate the possibilities of our algorithm,

Fig. 21. Results of applying the EM Gamma mixture estimation procedure to the tile depicted in Figure 23: (a) Histogramm of the data values (normalised) with superimposed estimated Gamma mixture; (b) Distributions corresponding to the oil class (left) and water class (right) after nine iterations of Algorithm 3: Histogramm and superimposed fitting.
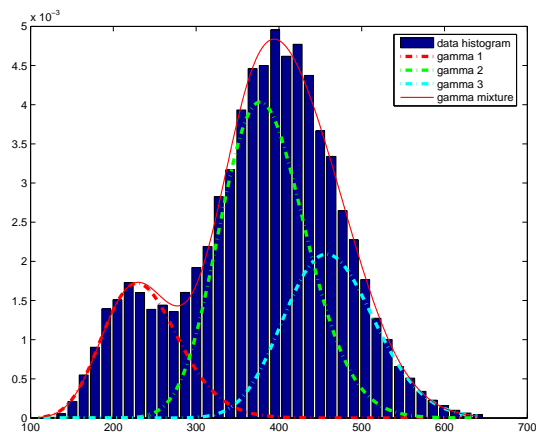

Fig. 22. Histogramm of the data values (normalised) with superimposed estimated Gamma mixture when the number of classes is set to $c = 3$

we again run it on the tile shown in Figure 23, but this time setting the number of classes to $c = 3$. By doing so, we hope to be able to segment a third ambiguous zone, corresponding to intermediate radiometry levels and probably due to atmospheric conditions originating a front. In this case we choose to fit a Gamma mixture of c modes to the data for initialising the class parameters, as depicted in Fig. 22. With this initialisation, the first obtained segmentation using Algorithm 3 is displayed in Fig. 24, after only nine iterations, all parameters have already converged to the segmentation given in Fig. 25. When comparing this segmentation with the one provided by a state-of-the-art algorithm, namely by multiscale HMC model in [14], we consider to have obtained a very good result.
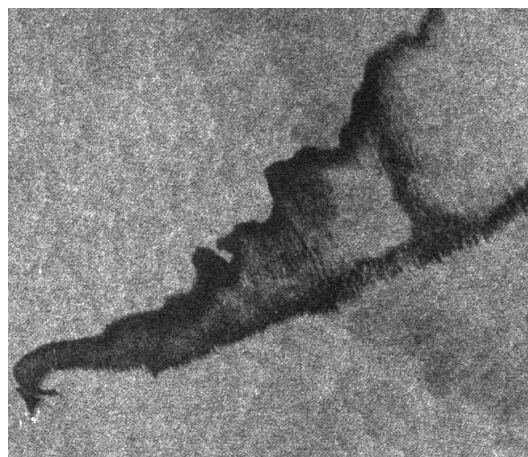

Fig. 23. Display of one of the individual tiles in which the image in Figure 19 has been divided for segmentation.
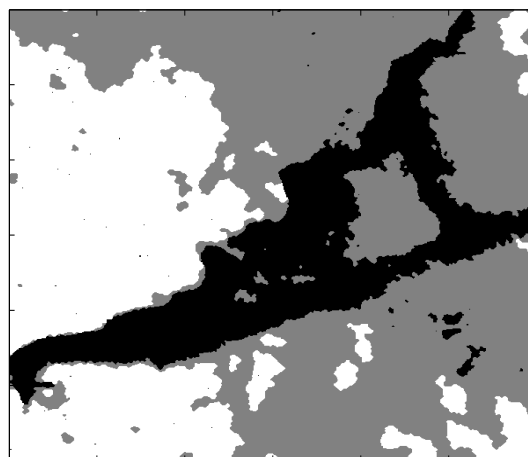

Fig. 24. Display of the initial segmentation result of applying Algorithm 3 with Loopy-β-Estimation and three classes to the tile depicted in Figure 23.
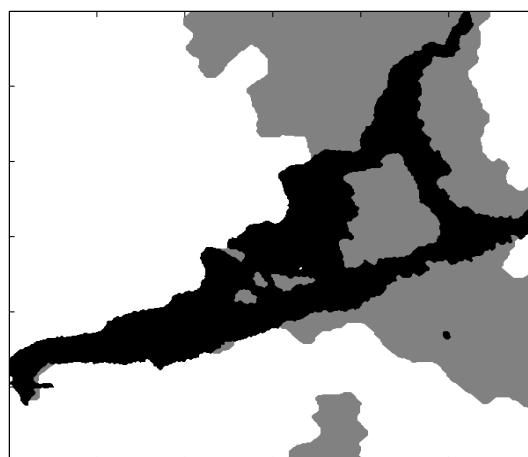

Fig. 25. Display of the final segmentation result, after nine iterations, of applying Algorithm 3 with Loopy-β-Estimation and three classes to the bottom left tile of Fig. 17.

4) Time considerations: The proposed algorithms display, in practical cases, a computational complexity of $O(N)$, with N being the number of pixels in the image. In fact, the most time-consuming steps in the algorithms are the α-Expansion and Graph-Cut segmentation routines. These use the max-flow code implementation referred in [32] that has complexity $O(n^{2.5})$ in the worst-case, however, in most situations is $O(N)$. Both Algorithm 1 and Algorithm 3 converge in few iterations and, as an example, for images of the dimension used in the first simulation (64x64) (see VI-A.1), only a few seconds are needed to run the algorithms. For a 600x600 pixel tile, like those in the Prestige example in VI-B.3, Algorithm 3 is taking ≈ 1 minute, but this time performance could be improved by a more efficient implementation of our procedure, namely by improving the implementation of the Loopy-β-Estimation and by migrating our matlab code fully to C-code.

VII. Conclusions

The results of applying the proposed methodology to simulated images with Gamma data models and to real SAR data are promising. The developed EM Gamma mixture estimation algorithm, when incorporated into the proposed segmentation algorithms, has proved to be an efficient tool for data modeling in SAR intensity images.

With the supervised algorithms, high OA accuracies have been achieved even for simulated images with high levels of noise. In general the Bayesian approach has resulted in an OA increase in the segmentation process between 10 and 15%, when compared to the segmentation using no prior information. Hereby, Algorithm 1 with LSF and Algorithm 1 with CD methods provided similar results, with a performance close to that obtained by setting the β value manually. On the other hand, the LSF estimation procedure seems to become less reliable for noisier images. When compared to Algorithm 2, using Loopy-β-Estimation, we see that this last method allways provides equal or better results, outcoming Algorithm 1. Another advantage of Algorithm 2 is that it is usually faster than Algorithm 1, being a one-shot process. Both algorithms, by introducing prior information into the segmentation process, increase the OA significantly. As a conclusion, Algorithm 2 should be preferred to Algorithm 1, when using a supervised method for oceanic sar images segmentation.

Algorithm 3, totally unsupervised, has been conceived as an improvement to Algorithm 1 and 2. When testing it on simulated data the obtained results were very good, although slightly worst than the supervised ones, as expected. In general, after a few iterations, the class and smoothness parameters converge to stable and meaningful values. By applying Algorithm 3 to real images containing documented oil spills, the segmentation has been considered successfull. We could segment both linear and patch type oil spills. Furthermore, the applicability of the method to segment whole scenes, as well as to segment more than two classes, has also been demonstrated in a well known image from the Prestige accident. As a conclusion, we believe the presented methods are suitable to be used for segmenting oceanic SAR images. In particular for oil spill detection, Algorithm 3 seems to be a suitable method.

Based on observations that oil spills in the ocean are often dragged by the wind and align more or less perpendicular to its direction, an interesting future issue is the incorporation of wind information into our segmentation algorithms. By adopting anisotropic MRF in the prior, we intend to reflect this directional dependency of the clique potentials. In the practice, the smoothness parameter β is no longer considered to be a constant value but is clique-dependent according to the wind direction and/or velocity. For estimating the wind profile from the SAR data, state-of-the-art algorithms are used. We have run promising simulations, are currently testing this extension on real data, and expect to obtain interesting results on a near future.

Appendix

Fitting a Mixture of Gamma Densities

As stated in the main text of this work, when segmenting real SAR images, the adopted data model for each class is a finite Gamma mixture. For completely defining the data model for each class (for lightness of the notation we have dropped the class index), we need to estimate $2K$ Gamma parameters, $\theta = (\theta_1, \ldots, \theta_K)$, with $\theta_s = (a_s, \lambda_s)$, and K a priori probabilities $\alpha = (\alpha_1, \ldots, \alpha_K)$). We infer $\phi = (\alpha, \theta)$ by computing its ML estimate from a training set. The ML estimate is computed via an EM algorithm [26]. Fitting a Gamma mixture is addressed in [33]. However, the authors consider N-look SAR images meaning that the underlying random variables are the average of N independent and identically distributed exponential random variables, thus, having a Gamma density but with just one parameter free; if the mean is μ, then the variance is given by μ^2/N. We estimate both the mean and the variance for each Gamma distribution in the mixture, rendering the algorithm more adaptable to real measurements.

The key point in the EM technique is the introduction of the so called missing data z, such that $p(y|\phi) = \int p(y, z|\phi)\,dz$ and $p(y, z|\phi)$ is easier to manipulate than $p(y|\phi)$. In the particular case of a mixture of densities, we will use as missing data a random variable, z_i, per site, with distribution $p(z_i = s) = \alpha_s$. It is interpretable as the probability of the $s-th$ Gamma mode is selected at pixel i. The EM algorithm alternates between two steps: the E-step computes the conditional expectation of the logarithm of the complete a posteriori probability function, with respect to the missing variables, based on the actual parameter value. The M-step updates the values of the parameters, by maximizing the expression obtained in the E-step with respect to each parameter on turn, *i.e.*,

$$\text{E-step} \quad : \quad Q\left(\phi; \phi^t\right) = E\left\{\log p\left(y, z|\phi\right)|y, \phi^t\right\} \quad (29)$$

$$\text{M-step} \quad : \quad \phi^{t+1} = \arg\max_{\phi} Q\left(\phi; \phi^t\right) \quad (30)$$

Denoting

$$w_{si}^t = P\left(z_i = s|y_i, \phi^t\right) \quad (31)$$

and taking into account that $\sum \alpha_s = 1$, then ϕ^{t+1} can be found among the stationary points of the Lagrangean

$$\mathcal{L}(\phi) = \sum_{i=1}^{N} \sum_{s=1}^{K} \left(L_{si}(\theta_s, \alpha_s) w_{si}^t \right) + \lambda \left(\sum_{s=1}^{K} \alpha_s - 1 \right), \quad (32)$$

with

$$
\begin{aligned}
L_{si}(\theta_s, \alpha_s) = & \log \left(\lambda_s^{a_s} \right) - \log \left[\Gamma \left(a_s \right) \right] \\
& + \log \left(y_i^{a_s - 1} \right) - \lambda_s y_i + \log \left(\alpha_s \right), \quad (33)
\end{aligned}
$$

where λ denotes a Lagrange multiplier. The expression for w_{si}^t (see [34]) is given by

$$w_{si}^t = \frac{\alpha_s^t p(y_i | \theta_s^t)}{\sum_{r=1}^{K} \alpha_r^t p(y_i | \theta_r^t)}. \quad (34)$$

In the M-step, after differentiating $\mathcal{L}$ in order to the unknown parameters and setting the derivatives to zero, we obtain a closed solution for the updating of the a priori probabilities α_i's, but numerical iteration is needed for determining parameters a_i's and λ_i's of the Gamma densities. Expression (35) gives the update expression for α_i's.

$$\alpha_s^{t+1} = \frac{1}{N} \sum_{i=1}^{N} w_{si}^t. \quad (35)$$

Equation (36) and (37) give the update expressions for the parameters λ_i's and a_i's.

$$\lambda_s^{t+1} = \frac{a_s^t \sum_{i=1}^{N} w_{si}^t}{\sum_{i=1}^{N} y_i w_{si}^t}, \quad (36)$$

$$a_s^{t+1} = \Psi^{-1} \left[\frac{\log \left(\lambda_s^t \right) \sum_{i=1}^{N} w_{si}^t + \sum_{i=1}^{N} \log \left(y_i \right) w_{si}^t}{\sum_{i=1}^{N} w_{si}^t} \right], \quad (37)$$

where

$$\Psi \left(a_s \right) = \frac{\Gamma' \left(a_s \right)}{\Gamma \left(a_s \right)}, \quad (38)$$

is the psi function. We refer to the Appendix B of [35] for a very fast Newton procedure to compute the inverse of the $\text{psi}(\cdot)$ function. Expressions (36) and (37) are iteratively recomputed until convergence is obtained, starting from initial values computed from the observed data y. The initial parameter values are calculated in such a way, that the initial probability function is a sum of equidistant Gammas that span the most representative data range. The EM scheme converges in a few tens of iterations.

ACKNOWLEDMENTS

REFERENCES

[1] C. R. Jackson and J. R. Apel, *Synthetic Aperture Radar Marine User's Manual*, Commerce Dept., NOAA, 2005.

[2] S. Y. Wu and A. K. Liu, "Towards an automated ocean feature detection, extraction and classification scheme for SAR imagery," *International Journal Remote Sensing*, vol. 24, no. 5, pp. 935–951, 2003.

[3] C. Brekke and A. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sensing of Environment*, vol. 95, pp. 1–13, 2005.

[4] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2839–2846, February 2006.

[5] S. Pelizzari and José M. B. Dias, "Bayesian adaptive oil spill segmentation of SAR images via graph cuts," in *Proceedings of the SeaSAR*, 2006.

[6] S. Pelizzari and José M. B. Dias, "Bayesian oil spill segmentation of SAR images via graph cuts," in *Proceedings of the IbPRIA*, 2007, vol. 2, pp. 637–644.

[7] S. Pelizzari and José M. B. Dias, "Oil spill segmentation of SAR images via graph cuts," in *Proceedings of the IGARSS*, 2007, pp. 1318 – 1321.

[8] T. Silva J. B. Dias and J. Leit ao, "Adaptive restoration of speckled SAR images using a compound random Markov field," in *Proceedings of ICIP*, 1998, vol. 2, pp. 79–83.

[9] R. Garello Y. Delignon and A. Hillion, "Statistical modelling of ocean SAR images," *IEEE Proc.-Radar, Sonar Navigation*, vol. 144, no. 6, pp. 348–354, December 1997.

[10] S.Z.Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, 1995.

[11] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding belief propagation and and its generalizations," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2001.

[12] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, February 2004.

[13] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[14] S. Derrode and G. Mercier, "Unsupervised multiscale oil slick segmentation from SAR images using a vector hmc model," *Pattern Recognition*, vol. 40, pp. 1135–1147, 2007.

[15] A. Gambardella M. Migliaccio and M. Tranfaglia, "SAR polarimetry to observe oil spills," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 2, pp. 506–511, February 2007.

[16] C. Brekke A. Solberg and P.Husøy, "Oil spill detection in Radarsat and envisat SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, 2007.

[17] G. Mercierb V.P. Onanac J. Mvogo Ngonoc P.L. Frisond J.P. Rudantd T. Kanaa, E. Tonyea and R. Garellob, "Detection of oil slick signatures in SAR images by fusion of hysteresis thresholding responses," in *Proceedings of IGARSS*, 2003, vol. 4, pp. 2750–2752.

[18] P. Réfrégier F. Galland and O. Germain, "Synthetic aperture radar oil spill segmentation by stochastic complexity minimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, 2004.

[19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Patt. Analysis and Mach. Intel.*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[20] S. YuE M. Berthod, Z. Kato and J. Zerubia, "Bayesian image classification using Markov random fields," *Image and Vision Computing*, vol. 14, pp. 285–295, 1996.

[21] H. Derin P. Kelly and K.D.Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1628–1641, 1988.

[22] G. Vézina H. Derin, P. Kelly and S.G.Labitt, "Modeling and segmentation of speckled images using complex data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 1, pp. 76–87, 1990.

[23] P.C.Smits and S.G.Dellepiane, "Synthetic aperture radar image segmentation by a detail preserving Markov random field approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 844–857, 1997.

[24] C. Jakowatz, D. Wahl, D. Eichel, D. Ghiglia, and P. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer, Norwell, MA, 1996.

[25] N. Jonhnson and S. Kotz, *Distribution in Statistics: Continuous Univarite Distributions*, vol. 1, Wiley, New York, 1996.

[26] N. Laird A. Dempster and D. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[27] S. Kumar, J. August, and M. Hebert, "Discontinuity preserving surface reconstruction through global optimization," in *EMMCVPR, LNCS*, 2005, vol. 3757, pp. 153–168.

[28] M. Ibáñez and A. Simó, "Parametric estimation in Markov random fiels image modeling with imperfect observations. a comparative study," vol. 24, pp. 2377–2389, 2003.

[29] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields and some comments on Chalmond's EM Gibbsian algorithm," *Probability Theory and Related Fields*, vol. 82, pp. 625–645, 1989.

[30] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields and some comments on Chalmond's EM Gibbsian algorithm," in *Proc. Stochastic Models, Statistical Methods and Algorithms in Image Analysis*, P. Barone and A. Frigessi, Eds., Berlin, Germany, 1991, vol. 47 of *Lecture Notes in Statistics*, pp. 240–258, Springer.

[31] S. Lakshamanan and H. Derin, "Simultaneous parameter estimation and segmentation of gibbd random fields using simulated annealing," *IEEE Trans. Patt. Analysis and Mach. Intel.*, vol. 11, pp. 793–813, 1989.

[32] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[33] S. Wang Q. Jiang A. E. Zaart, D. Ziou and G. B. Bénié, "SAR images segmentation using mixture of Gamma distributions," in *Proceedings of Vision Interface*, 1999.

[34] G. McLachlanand and D. Peel, *Finite Mixture Models*, John Wiley and Sons, New York, 2000.

[35] José M. Nascimento, *Unsupervised Hyperspectral Unmixing*, Ph.D. thesis, Instituto Superior Tecnico, 2006.