

Unsupervised change detection from multichannel SAR data by Markov random fields

Gabriele Moser and Sebastiano B. Serpico *

April 12, 2007

Abstract

In the contexts of environmental monitoring and disaster management, multichannel synthetic aperture radar (SAR) data present a good potential, thanks both to their insensitivity to atmospheric and Sun-illumination conditions, and to the improved discrimination capability they may provide as compared to single-channel SAR. However, exploiting this potential requires accurate and possibly automatic techniques to generate change maps from multichannel SAR images acquired from the same geographic area at different times. In this paper an unsupervised contextual change-detection method is proposed for two-date multichannel SAR images, by adopting a data-fusion approach. Each SAR channel is modelled as a distinct information source and a Markovian approach to data fusion is adopted. A Markov random field model is introduced, that combines together the information conveyed by each SAR channel and the spatial contextual information related to the correlation among neighboring pixels by using “energy functions.” In order to address the task of the estimation of the model parameters, the expectation-maximization algorithm is combined with the recently proposed “method of log-cumulants.” The proposed technique is experimentally validated on SIR-C/XSAR data.

1 Introduction

In the contexts of environmental monitoring and disaster management, techniques allowing one to detect the changes that occurred in a given area between successive acquisition dates may be efficient data-analysis tools. Good discrimination between changed and unchanged areas can often be achieved by analyzing optical multispectral data, by operating in the related multidimensional feature space. However, such data are affected by atmospheric and Sun-illumination conditions. Synthetic aperture radar (SAR) [16] is insensitive to such issues but usually provides only one amplitude/intensity observation, thus possibly resulting in poor discrimination. Multichannel (i.e., multipolarization and/or multifrequency) SAR exhibits an improved potential: as compared with single-channel SAR, it is expected to provide an increased discrimination capability, while maintaining its insensitivity to atmospheric and illumination issues. This potential is also reinforced by the availability of multichannel SAR data with short revisit time which

*University of Genoa, Dept. of Biophysical and Electronic Eng. (DIBE), I-16145 Genoa, Italy, e-mail: gemini@dibe.unige.it

are granted by future missions (e.g., COSMO/SkyMed or TerraSAR), but stresses the need for accurate change-detection techniques for such data. Recently some efforts have been directed toward the automation of the change-detection process with single-channel SAR [1] [11]. However, change detection on multichannel SAR is still mostly addressed by using time-expensive “trial-and-error” procedures [3] [5] [18].

When working in a completely unsupervised context (i.e., assuming no ground-truth to be available at any acquisition date) a SAR change-detection problem is typically addressed by image ratioing, i.e., a multivariate ratio image is generated by computing pixel-by-pixel and channel-by-channel the ratio of the SAR amplitudes (or intensities) at the first date by the ones acquired at the second date [11] [16]. This is motivated by the multiplicative nature of the speckle (noise) phenomenon intrinsically affecting SAR data [16]. However, when dealing with multichannel SAR, the development of accurate parametric models for the related multivariate statistics of SAR amplitude ratios is a difficult task. In [12] this problem was overcome by iteratively identifying a suitable mapping of the multivariate ratio channels onto a one-dimensional space and by classifying only in the transformed space. However, a drawback of this iterative method was that, even though a good convergence behavior was experimentally noted, no theoretical proof of convergence was available.

In the present paper the problem of multichannel SAR change detection is formalized in terms of a multichannel data-fusion problem and is addressed by a Markov random field (MRF) approach [6] [23]. Each SAR channel is modelled as a separate “information source” and MRFs are used to define a decision rule based on multichannel fusion. MRFs are adopted thanks to their capability to integrate both distinct information sources [23] and the spatial contextual information into an image classification process, by defining suitable “energy functions” and by formulating the maximum-*a-posteriori* classification criterion as a minimum-energy decision rule [6] [23]. The use of contextual information also allows the speckle impact to be reduced. In order to address the task of the estimation of the model parameters, the expectation-maximization (EM) algorithm [19] and one of its variants (the EM algorithm with semilabeled samples developed by Landgrebe and Jackson, LJ-EM for short [9]) are combined with the recently proposed “method of log-cumulants” (MoLC) for parametric probability density function estimation [13] [14] [15]. The proposed technique is experimentally validated on SIR-C/XSAR data.

The paper is organized as follows. Sec. 2 defines the proposed change detection method and the related MRF model. Sec. 3 describes the experimental results obtained by applying the method to real multichannel SAR data. Finally conclusions are drawn in Sec. 4. Proofs and mathematical details are reported in the Appendix.

2 Methodology

2.1 The proposed MRF model

Let \mathcal{I}_0 and \mathcal{I}_1 be two coregistered SAR images, acquired over the same area at times t_0 and t_1 , respectively ($t_1 > t_0$), and consisting of n SAR amplitude channels each. The change-detection problem is formulated as a binary hypothesis testing problem [7], by denoting the “change” and “no-change” hypotheses as H_0 and H_1 , respectively. The image-ratioing approach, which generates a ratio image \mathcal{R} by dividing pixel-by-pixel and channel-by-channel the amplitudes in \mathcal{I}_0 by the amplitudes in \mathcal{I}_1 [5], is adopted; hence, \mathcal{R} is an n -channel image as well.

The key-idea of the proposed technique is to model each channel as a distinct information source and to accordingly formalize the change-detection problem as a multichannel data-fusion problem addressed by MRFs. Specifically, \mathcal{R} is modelled as a set $\{u_1, u_2, \dots, u_N\}$ of n -variate identically distributed random vectors (N being the number of pixels) and the hypothesis label of the k -th pixel is denoted by $\ell_k \in \{H_0, H_1\}$ ($k = 1, 2, \dots, N$). The label configuration $\{\ell_1, \ell_2, \dots, \ell_N\}$ (i.e., the two-dimensional binary random process of the pixel labels) is assumed to be an MRF, i.e., for each k -th pixel a neighborhood is defined and the following properties are assumed to hold [6] [9]:

- positivity: $P(\ell_1, \ell_2, \dots, \ell_N) > 0$ for all $\ell_1, \ell_2, \dots, \ell_N \in \{H_0, H_1\}$;
- Markovianity: $P\{\ell_k = H_i | \ell_h, h \neq k\} = P\{\ell_k = H_i | \mathcal{C}_k\}$, where \mathcal{C}_k is the set of the labels of the neighboring pixels (i.e., the context) of the k -th pixel ($k = 1, 2, \dots, N; i = 0, 1$).

Under mild assumptions, the following expression results for the (local) posterior distribution of each pixel label ℓ_k , given u_k and \mathcal{C}_k ($k = 1, 2, \dots, N; i = 0, 1$) [6]:

$$P\{\ell_k = H_i | u_k, \mathcal{C}_k\} = \frac{\exp[-U(H_i | u_k, \mathcal{C}_k)]}{\sum_{j=0}^1 \exp[-U(H_j | u_k, \mathcal{C}_k)]}, \quad (1)$$

where $U(H_i | u_k, \mathcal{C}_k)$ is an “energy function” characterizing the specific MRF model. According to the MRF-based approach to data fusion [23], the energy function of the proposed model is expressed as a linear combination of energy contributions, each related either to the spatial correlation among neighboring pixels in the image (“spatial energy contribution”) or to the information conveyed by each channel of the ratio image. The former is formulated as an isotropic Potts MRF model with respect to a 2nd-order neighborhood¹ [6] and the latter is directly related to the single-variate probability density function (PDF) of the corresponding SAR ratio channel conditioned to “change” and “no-change.” Hence, the proposed MRF model is defined by the following energy function² ($k = 1, 2, \dots, N; i = 0, 1$):

$$U(H_i | u_k, \mathcal{C}_k, \theta) = \sum_{r=1}^n \alpha_r [-\ln p_{ir}(u_{kr} | \xi_{ir})] - \beta m_{ik}, \quad (2)$$

where $p_{ir}(\cdot | \xi_{ir})$ is a parametric model for the PDF of the r -th amplitude ratio u_{kr} conditioned to H_i , ξ_{ir} is the vector of the parameters of this PDF, m_{ik} is the number of labels equal to H_i in the neighborhood of the k -th pixel, α_r ($r = 1, 2, \dots, n$) and β are positive parameters, and $\theta = (\beta, \alpha_1, \alpha_2, \dots, \alpha_n, \xi_{01}, \xi_{11}, \xi_{02}, \xi_{12}, \dots, \xi_{0n}, \xi_{1n})$ is a vector collecting all the model parameters to be estimated (in Eq. (2) the dependence of the energy on θ has been stressed by including θ in the corresponding notation).

The spatial contextual information is represented by m_{ik} thanks to the Potts model. According to the minimum-energy MRF decision rule, the spatial term ($-m_{ik}$) encourages the generation of connected regions including pixels assigned to the same hypothesis and thus minimizing the impact of speckle on the detection results (further details can be found in [6]). The model formalized by Eqs. (1) and (2) is intrinsically stationary. This condition is usually not fulfilled by real images, but is often accepted as a simplifying analytical assumption [8].

¹In a 2nd-order neighborhood the neighbors of each pixels are, by definition, the 8 surrounding pixels.

²Given $v \in \mathbb{R}^m$, we denote by v_j the j -th component of v ($j = 1, 2, \dots, m$).

Table 1: NR, WR, and LN parametric families for the H_i -conditional PDF of the r -th amplitude ratio u_{kr} and their MoLC equations (all the PDFs are defined for $u_{kr} > 0$ and are zero elsewhere; $k = 1, 2, \dots, N$; $r = 1, 2, \dots, n$; $i = 0, 1$). $\Gamma(\cdot)$ is the usual Gamma function [20] and $\Psi(1, \cdot)$ is the first-order polygamma function (i.e., the second logarithmic derivative of $\Gamma(\cdot)$ [22]).

Parametric family	PDF	Parameter vector	MoLC equation
LN	$p_{ir}(u_{kr} \xi_{ir}) = \frac{\exp\left[-\frac{(\ln u_{kr} - \mu_{ir})^2}{2\sigma_{ir}^2}\right]}{\sigma_{ir}u_{kr}\sqrt{2\pi}}$	$\xi_{ir} = (\mu_{ir}, \sigma_{ir})$ $\mu_{ir} \in \mathbb{R}, \sigma_{ir} > 0$	$\kappa_{1ir} = \mu_{ir}$ $\kappa_{2ir} = \sigma_{ir}^2$
NR	$p_{ir}(u_{kr} \xi_{ir}) = \frac{2\Gamma(2L_{ir})}{\Gamma^2(L_{ir})} \frac{\gamma_{ir}^{L_{ir}} u_{kr}^{2L_{ir}-1}}{(\gamma_{ir} + u_{kr}^2)^{2L_{ir}}}$	$\xi_{ir} = (L_{ir}, \gamma_{ir})$ $L_{ir}, \gamma_{ir} > 0$	$2\kappa_{1ir} = \ln \gamma_{ir}$ $2\kappa_{2ir} = \Psi(1, L_{ir})$
WR	$p_{ir}(u_{kr} \xi_{ir}) = \eta_{ir}\lambda_{ir}^{\eta_{ir}} \frac{u_{kr}^{\eta_{ir}-1}}{(\lambda_{ir}^{\eta_{ir}} + u_{kr}^{\eta_{ir}})^2}$	$\xi_{ir} = (\eta_{ir}, \lambda_{ir})$ $\eta_{ir}, \lambda_{ir} > 0$	$\kappa_{1ir} = \ln \lambda_{ir}$ $\kappa_{2ir} = 2\Psi(1, 1)\eta_{ir}^{-2}$

The parameters $\beta, \alpha_1, \alpha_2, \dots, \alpha_n$ tune the relative importance between the energy terms. In particular, α_r plays the role of a “reliability factor” assigned to the information source represented by the r -th channel ($r = 1, 2, \dots, n$) and is usually assumed to take values in $[0, 1]$. This constraint can be analytically expressed as follows³:

$$\|2\alpha - \mathbf{1}\|_\infty \leq 1, \quad (3)$$

where $\mathbf{1}$ is an n -dimensional vector with n unitary components and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$. As detailed in the Appendix, a differentiable constraint is desirable for analytical convenience. Hence, we replace the nondifferentiable constraint in Eq. (3) by the following one:

$$\|2\alpha - \mathbf{1}\|_q \leq 1, \quad (4)$$

where $q \geq 1$. For sufficiently large values of q (e.g., $q = 10$) Eq. (4) is basically a differentiable approximation of Eq. (3). Furthermore, it also allows undesired degenerate parameter values to be avoided (e.g., the undesired configuration $\alpha = 0$, which removes from the MRF model all the energy contributions related to the ratio channels, is permitted by Eq. (3) but forbidden by Eq. (4)).

2.2 Overview of the proposed change-detection method

The proposed SAR multichannel change-detection method postulates the MRF model of Eqs. (1) and (2) to hold and aims at generating a change map which minimizes the corresponding energy function. Hence, suitable parametric models for the involved PDFs and a parameter-estimation and classification strategy need to be defined.

As far as the former issue, several parametric families have been introduced in the literature to describe the marginal statistics of each amplitude ratio feature u_{kr} conditioned to H_i ($k =$

³Given $v \in \mathbb{R}^m$ and $q \geq 1$, we denote by $\|v\|_\infty$ and $\|v\|_q$ the Tchebitchev and the q -th order Minkowski norms (∞ -norm and q -norm for short) of v , i.e.: $\|v\|_\infty = \max\{|v_1|, |v_2|, \dots, |v_m|\}$ and $\|v\|_q = (|v_1|^q + |v_2|^q + \dots + |v_m|^q)^{1/q}$ [21].

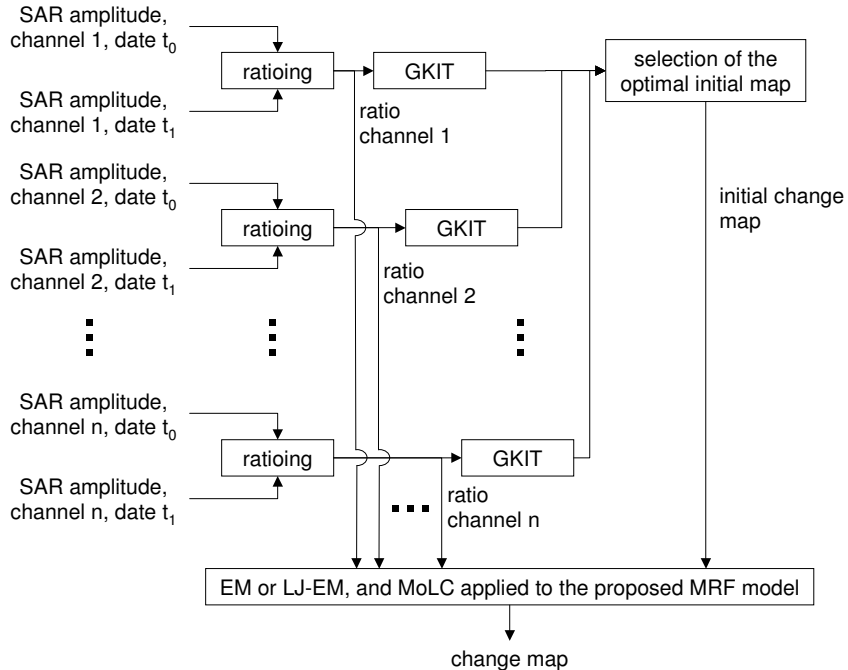


Figure 1: Block diagram of the proposed change-detection method.

$1, 2, \dots, N; r = 1, 2, \dots, n; i = 0, 1)$ [11] [16]. Here, the Nakagami-ratio (NR), Weibull-ratio (WR), and log-normal (LN) distributions are adopted, that are analytically derived (under mild assumptions) from the well-known Nakagami-Gamma, Weibull and log-normal distributions for the SAR amplitudes acquired at the two observation dates, respectively (see Table 1; further details can be found in [11]).

Concerning the parameter-estimation and classification tasks, case-specific variants of the EM and LJ-EM methods are developed, by integrating such approaches with MoLC (see Sec. 2.3). The resulting method is iterative; in order to initialize it, the change map generated by the (single-channel) “generalized Kittler and Illingworth thresholding” (GKIT) algorithm [11], applied to an automatically selected channel in \mathcal{R} , is used (see Sec. 2.4). A block diagram of the proposed method is shown in Fig. 1.

2.3 Parameter estimation and classification

EM is an iterative parameter estimation methodology, which deals with problems of data incompleteness and aims to compute maximum-likelihood estimates [19]. LJ-EM is a modified version of EM, specifically aimed at computing parameter estimates which allow a given set of hypotheses to be well discriminated [9]. EM has been proved to converge, under mild assumptions, to a local maximum of the log-likelihood function (even though convergence to a global maximum is not ensured, usually a good solution is obtained) [19]. However, the application of EM and LJ-EM to the proposed MRF model arises two difficulties.

First, the application of EM-type algorithms to contextual image models typically involves long computation times, due to the need to run a Gibbs sampler or to compute MonteCarlo approximations at each EM iteration [2]. To overcome this problem, the “mode-field” approach is adopted, which introduces an approximation in the log-likelihood function in order to reduce the computational burden, and which represents a good tradeoff between execution time

and classification accuracy (a general description of this approach can be found in [2]; similar techniques have been applied in [4] [8] [12]).

Secondly, the maximization problems involved by LJ-EM and EM have no close-form solution, when using the NR and WR models [11] [13]. In order to address this issue, we propose to combine EM and LJ-EM with MoLC, recently developed parameter estimation method, which exploits the Mellin transform theory [22] to state a set of equations relating the unknown parameters to logarithmic moments or logarithmic cumulants of the gray-level distribution, thus expressing the parameter-estimation problem as the solution of a set of (typically nonlinear) equations [15]. MoLC has been developed in the context of SAR image analysis and has been proven to be numerically feasible and fast for many SAR-specific models and to exhibit good theoretical estimation properties [14] [15]. It turns out to be feasible for all the LN, NR and, WR distributions [11]: denoting by $\kappa_{1ir} = E\{\ln u_{kr} | \ell_k = H_i\}$ and $\kappa_{2ir} = \text{Var}\{\ln u_{kr} | \ell_k = H_i\}$ the logarithmic mean and variance of u_{kr} , given H_i ($k = 1, 2, \dots, N; r = 1, 2, \dots, n; i = 0, 1$), MoLC allows the equations in Table 1 to be obtained [11].

As analytically proved in the Appendix, if q is an even integer number, the following operations are performed at the t -th iteration of the mode-field EM or LJ-EM algorithms, as integrated with MoLC and applied to the MRF model in Eqs. (1) and (2) (the current parameter estimates and pixel labels are marked with superscript “ t ”; $t = 0, 1, 2, \dots$):

1. compute for each k -th pixel the current estimate of $P\{\ell_k = H_i | u_k, \mathcal{C}_k^t, \theta^t\}$ ($k = 1, 2, \dots, N; i = 0, 1$) using Eqs. (1) and (2);
2. update the label of each k -th pixel according to the MRF minimum-energy rule [6], by setting ℓ_k^{t+1} as the label H_i that corresponds to the lowest value of $U(H_i | u_k, \mathcal{C}_k^t, \theta^t)$ ($k = 1, 2, \dots, N; i = 0, 1$);
3. update the estimates of the spatial parameter β and of the logarithmic means and variances as follows ($r = 1, 2, \dots, n; i = 0, 1$):

$$\begin{cases} \kappa_{1ir}^{t+1} = \frac{\sum_{k=1}^N w_{ik}^t \ln u_{kr}}{\sum_{k=1}^N w_{ik}^t} \\ \kappa_{2ir}^{t+1} = \frac{\sum_{k=1}^N w_{ik}^t (\ln u_{kr} - \kappa_{1ir}^{t+1})^2}{\sum_{k=1}^N w_{ik}^t} \\ \beta^{t+1} = \arg \max_{\beta > 0} \sum_{k=1}^N \left[\beta \sum_{i=0}^1 w_{ik}^t m_{ik}^t - \ln \sum_{i=0}^1 \exp(\beta m_{ik}^t) \right]; \end{cases} \quad (5)$$

where the weight w_{ik}^t is computed as follows ($k = 1, 2, \dots, N; i = 0, 1$):

- if EM is used, $w_{ik}^t = P\{\ell_k = H_i | u_k, \mathcal{C}_k^t, \theta^t\}$ as computed in step 1;
 - if LJ-EM is used, $w_{ik}^t = P\{\ell_k = H_i | u_k, \mathcal{C}_k^t, \theta^t\}$ if the k -th pixel has been assigned to H_i in step 2 and $w_{ik}^t = 0$ otherwise;
4. compute, for each hypothesis H_i and each r -th channel, the updated estimate ξ_{ir}^{t+1} of the parameter vector ξ_{ir} of the adopted PDF model (either NR, WR, or LN) by plugging the logarithmic mean and variance estimates κ_{1ir}^{t+1} and κ_{2ir}^{t+1} computed in step 3 in the corresponding MoLC equations (see Table 1) and by solving for the unknown parameters ($r = 1, 2, \dots, n; i = 0, 1$);

5. update the estimates of the reliability factor α_r of each r -th channel as follows ($r = 1, 2, \dots, n$):

$$\alpha_r^{t+1} = \frac{1}{2} + \frac{1}{2} \sqrt[q-1]{\frac{c_r^t}{\|c^t\|_{q'}}}, \quad (6)$$

where:

$$q' = \frac{q}{q-1}, \quad c_r^t = \sum_{k=1}^N \sum_{i=0}^1 w_{ik}^t \ln p_{ir}(u_{kr} | \xi_{ir}^{t+1}), \quad (7)$$

and $\|c^t\|_{q'}$ is the q' -norm of the vector $c^t = (c_1^t, c_2^t, \dots, c_n^t)$.

The maximization problem to compute β^{t+1} in step 3 is numerically solved by the Newton-Raphson method [17]. Step 4 requires solving the MoLC equations for the adopted PDF model (see Table 1). No actual solution process is required for LN, since the PDF parameters are exactly equal to the logarithmic mean and variance. The solution process for WR can be analytically carried out in close form. A numerical procedure is needed for NR, in order to solve the equation involving κ_{2ir} and L_{ir} . However, thanks to the strict monotonicity of the involved polygamma function, the simple bisection method [17] can be effectively applied. The update of the estimates of the reliability factors (step 5) is performed by a close-form calculation.

The difference between the EM and the LJ-EM versions of the method is that, at each iteration, the parameters of H_i are updated according to all image pixels in the case of EM, and only according to the samples assigned to H_i ($i = 0, 1$) in the case of LJ-EM. The latter estimation procedure, even though possibly biased, aims to reduce the overlapping between H_0 and H_1 and may be expected to generate parameter estimates which are more effective for classification. In both cases, since the iterative procedure performed by the proposed change detection method is actually derived from EM (see Appendix), a convergent behavior of the method can be theoretically expected [9] [19] [24].

2.4 Initialization by the GKIT algorithm

The initial change map is generated on a single-channel basis. The GKIT technique [11] is adopted here; it is still based on an NR, WR, or LN conditional model for SAR ratio data. Given one of these three models, GKIT automatically computes the optimal threshold to be applied to a single-channel ratio image in order to distinguish “change” and “no-change,” by minimizing a “criterion function” defined according to a Bayesian approach and related to the probability of error [11]. Let $J_r(\cdot)$ be the criterion function computed by GKIT when applied to the r -th channel in \mathcal{R} , $\tau_r^* = \arg \min_{\tau} J_r(\tau)$ the corresponding optimal threshold, and \mathcal{M}_r the resulting change map ($r = 1, 2, \dots, n$) [11]. The initial change map is defined as the map \mathcal{M}_r corresponding to the lowest optimal value $J_r(\tau_r^*)$ of the criterion function ($r = 1, 2, \dots, n$). According to the relationship between the criterion function and the probability of error [10] [11], this change map is expected to be, at least, a suboptimal choice among the single-channel maps.

For initialization purposes, this map is used to define the initial context \mathcal{C}_k^0 of each k -th pixel ($k = 1, 2, \dots, N$) and the initial estimates of the PDF parameters. In fact, for each r -th channel and each hypothesis H_i , a sample mean estimate κ_{1ir}^0 of κ_{1ir} and a sample-variance estimate κ_{2ir}^0 of κ_{2ir} are computed according to the set of samples assigned to H_i in the initial change map. Then, an estimate ξ_{ir}^0 of the related parameter vector is derived from the MoLC equations and used in the first iteration of the proposed method (see Table 1; $r = 1, 2, \dots, n; i = 0, 1$).

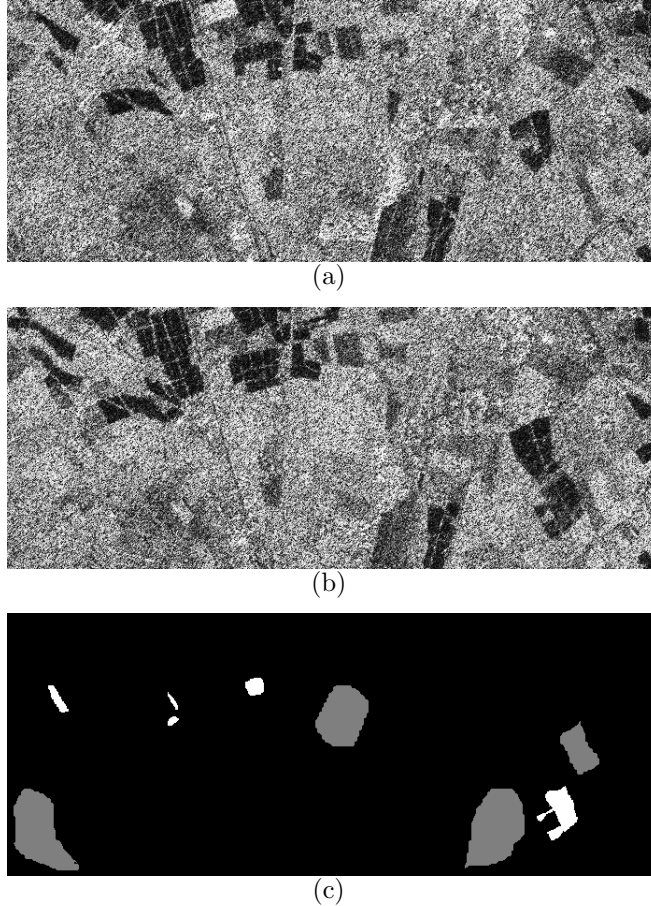


Figure 2: XSAR channels acquired on April 16th, 1994 (a) and on April 18th, 1994 (b) after histogram equalization, and test map (c). Legend for the test map: white = “change”, grey = “no-change.”

3 Experimental results

3.1 Data set for experiments

Two 700×280 pixel-sized coregistered multipolarization and multifrequency SAR images, acquired over an agricultural area near the city of Pavia on April 16th and 18th, 1994, were used for experiments. At each date, a 4-look XSAR image (X band, VV polarization) and three 4-look SIR-C channels (C band, HH, HV polarizations and TP (total power) channels) were available (Fig. 2). Ground changes occurred in the scene, as several rice fields were artificially flooded (for cultivation purpose). A test map, presenting 11287 “no-change” test pixels and 1870 “change” test pixels, was available (Fig. 2(c)).

3.2 Experimental results and comparisons

Six versions of the proposed method are defined according to the possible use of two parameter estimation strategies (EM and LJ-EM) and of three PDF models (LN, NR, and WR). In order to validate the method, the six versions were applied to the considered data set, by initializing the α_r ($r = 1, 2, 3, 4$) and β parameters to unitary values. The initialization GKIT procedure,

Table 2: Change-detection performances of the proposed method, of the initialization procedure, and of the previously proposed method developed in [12].

Norm index	Parametric model	Estimation method	False-alarm rate	Detection accuracy	Overall error rate
2	LN	EM	0.16%	94.27%	0.95%
		LJ-EM	0.16%	95.08%	0.84%
	WR	EM	0.72%	95.13%	1.31%
		LJ-EM	0.62%	97.05%	0.95%
	NR	EM	0.11%	89.78%	1.55%
		LJ-EM	0.11%	92.24%	1.20%
10	LN	EM	0%	93.79%	0.88%
		LJ-EM	0%	94.70%	0.75%
	WR	EM	0.72%	95.13%	1.31%
		LJ-EM	0.62%	97.05%	0.95%
	NR	EM	0%	89.73%	1.45%
		LJ-EM	0%	91.01%	1.27%
∞	LN	EM	2.01%	99.14%	1.84%
		LJ-EM	2.01%	99.14%	1.84%
	WR	EM	0.72%	95.13%	1.31%
		LJ-EM	0.62%	97.05%	0.95%
	NR	EM	0.12%	88.02%	2.78%
		LJ-EM	1.06%	88.02%	2.58%
method in [12]			0%	93.10%	0.98%
initialization			0.33%	48.09%	7.52%

applied with all the LN, NR, and WR models, selected the change map generated from the SIR-C-HV channel. As theoretically expected, the proposed method reached convergence in all the cases (fewer than 50 iterations were sufficient to reach convergence in all the experiments). Table 2 reports the detection accuracy (i.e., the percentage of correctly labelled “change” test pixels), the false-alarm rate (i.e., the percentage of erroneously labelled “no-change” test pixels), and the overall error rate (i.e., the percentage of erroneously labelled test pixels) of the resulting change maps computed according to the test map in Fig. 2(c).

Focusing first on the results obtained with $q = 2$ (i.e., Euclidean norm), accurate change maps were obtained by all the six versions of the method, with very low false-alarm rates (always below 1% and often below 0.2%) and high detection accuracies (above 90% in almost all the cases), even though the initialization map had a very low accuracy (see Table 2 and Fig. 3(a)). This suggests a good capability of the proposed method to jointly exploit both the information conveyed by the four ratio channels and the spatial information related to the correlations among neighboring pixels by Markovian data fusion.

The three PDF models yielded results quite similar to one another. Specifically, the lowest overall error rate (0.84%) was obtained by LJ-EM with LN, while the highest detection accuracy (97.05%) was achieved by LJ-EM with WR. A slightly worse result was obtained with NR. A further similarity can also be noted between the results of EM and the ones of LJ-EM, even though the latter always provided slightly higher detection accuracies and slightly lower error rates. This is an expected result, since LJ-EM explicitly aims at identifying parameter estimates

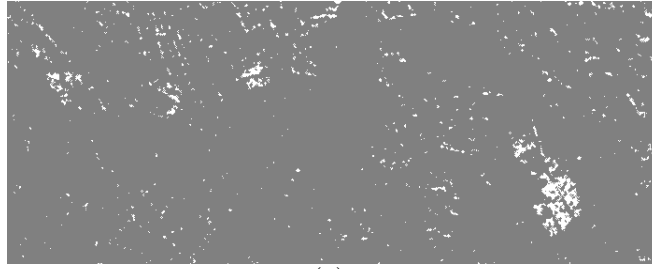
which allows the hypotheses to be better discriminated, as compared to EM.

In particular, even though no preliminary despeckling filter was applied, the change maps generated by the proposed technique were not significantly affected by speckle. This is interpreted as a consequence of the effectiveness of MRFs in exploiting the spatial contextual information. A visual analysis of the change maps (see, for instance, Fig. 3(b), which refers to LJ-EM and LN) further confirms the limited impact of speckle and the correct detection of the changed areas. On the contrary, the initialization map (Fig. 3(a)) generated by the noncontextual GKIT method, is strongly affected by speckle, which yields a large number of false and missed alarms. Several false alarms can be noted in Fig. 3(b), as well, especially at the borders of the agricultural fields which are present in the imaged scene. However, they are mainly due to small misregistration errors between the images acquired at the two dates.

A remarkably stable behavior of the results was noted as the Minkowski-norm index q was varied in the range $[2, 10]$ (the results obtained for $q = 10$ are reported in Table 2; the ones obtained for $q = 4, 6, 8$ are similar and are omitted for the sake of brevity). When using WR, the results of both EM and LJ-EM were actually unaffected by the variation of $q \in [2, 10]$; when either LN or NR were employed, only small variations in the numerical accuracies were noted (Table 2). A visual analysis of the related change maps (see, for example, Fig. 3(c), which still refers to LJ-EM and LN) points out that such variations are actually related to a stronger spatial regularization effect obtained by using the 10-norm, as compared with the 2-norm. Accordingly, most of the false alarms, which were obtained when $q = 2$ due to misregistration errors, were not present when $q = 10$ (Fig. 3(c)). On the other hand, a slightly stronger smoothing of the borders between changed and unchanged areas can be noted in the change map generated with the 10-norm than in the one obtained with the 2-norm. This can be interpreted as a consequence of the different convergence values reached by the estimated reliability factors. Focusing, for instance, on LJ-EM with LN, $\alpha = (0.224, 0.238, 0.281, 0.260)$ and $\alpha = (0.060, 0.062, 0.071, 0.066)$ were reached at convergence in the 2- and 10-norm cases, respectively, i.e., much lower values were obtained for $q = 10$ than for $q = 2$. On the other hand, the method provided almost unitary values for the estimate of β in both cases. Therefore, in the 10-norm case, the spatial energy contribution was assigned a comparatively higher relative weight than the energy terms related to the ratio channels, as compared to the 2-norm case, thus resulting in a stronger spatial regularization effect.

A specific experiment was performed to focus on the role of the constraints in Eqs. (3) and (4). A variant of the proposed method was run, such that, at each EM or LJ-EM iteration, the reliability factors were updated according to the ∞ -norm constraint. As discussed in the Appendix, this choice results in a linear programming problem, and yields binary-valued solutions (i.e., the updated value of each α_r parameter is either 0 or 1; $r = 1, 2, 3, 4$; see Appendix). As suggested by Table 2, the same results as those obtained with the 2- and 10-norms were achieved when using WR. On the other hand, significantly higher error rates were obtained with LN and NR with the ∞ -norm, as compared with the 2- and 10-norms. In particular, when using the LN model, EM and LJ-EM even converged to parameter configurations with zero values for all the α_r factors ($r = 1, 2, 3, 4$). In such degenerate situations, the minimum-energy decision rule is affected only by the spatial energy contribution and simply performs a smoothing of the initial change map, thus giving a very poor result (Fig. 3(d)). These critical configurations are forbidden by the q -norm constraint employed in the proposed method; this further confirms the usefulness of the adopted formulation.

Finally, a comparison was performed between the results of the proposed method and the ones given by the previously developed technique in [12], which combines MRFs with LJ-



(a)



(b)



(c)



(d)



(e)

Figure 3: Change maps generated by: (a) the initialization GKIT procedure; (b) the proposed method applied with LJ-EM, the LN model, and $q = 2$; (c) the proposed method applied with LJ-EM, the LN model, and $q = 10$; (d) the variant of the proposed method characterized by the ∞ -norm and applied with LJ-EM and LN; (e) the previously proposed method in [12] (the color legend is the same as in Fig. 2(c)).

EM and with an *ad hoc* version of the Fisher transform [7] in order to identify a suitable scalar mapping of the multichannel ratio channels and to classify in the transformed (one-dimensional) space. As shown in Fig. 3(e) and Table 2, accurate results were generated by this method, as well, even though globally higher performances were provided by the proposed technique. In particular, the detection accuracy of the proposed approach, applied with LJ-EM and WR, was almost 4% higher than the one provided by the algorithm in [12]. This suggests, at least on the adopted data set, a higher effectiveness for the data fusion approach than for the multichannel transformation approach, to the multichannel SAR change-detection problem.

4 Conclusions

In the present paper an unsupervised contextual change detection technique has been proposed for multichannel SAR images, by adopting a data-fusion approach and by integrating Markov random fields with EM-type and MoLC parameter-estimation algorithms. Two versions of the technique, based on the standard EM algorithm and on its LJ-EM variant, respectively, have been combined with three feasible models for the conditional distributions of the pixel intensities. The method was experimentally validated on real data and provided accurate results in all its versions. Slightly more accurate maps were generated by LJ-EM than by the standard EM algorithm, essentially because the former explicitly aims at computing parameter estimates that well separates the considered hypotheses. No preliminary despeckling was needed before applying the proposed method, thanks to the capability of MRF-based approaches to exploit the spatial contextual information in the classification process.

The proposed technique is iterative and a good convergent behavior in a small number of steps has been remarked in the experiments. The convergence of the method is a theoretically expected result, thanks to the well-known asymptotic properties of EM-like procedures. This also represents a theoretical advantage of this approach as compared with a recently proposed multi-channel SAR change detection method, based on the combination of MRFs, LJ-EM, and of a SAR-specific version of the Fisher transform [12]. The latter is an iterative technique, as well, but the convergence is not theoretically ensured. In addition, an experimental comparison between the proposed algorithm and the one in [12] also points out a remarkable improvement of the proposed approach, especially in the detection accuracy.

The three considered PDF models (i.e., log-normal, Weibull-ratio, and Nakagami-ratio) derive from corresponding well-known models for SAR amplitude statistics. They provided quite similar results to one another, with slightly lower overall error rates for the log-normal distribution. Slightly longer execution times were required by the Nakagami-ratio PDF, due to the need to numerically solve nonlinear equations at each iteration. Anyway, the computation time was small in all the cases and no more than a few minutes were needed to reach convergence in any run of the method.

The proposed technique presents an internal parameter whose value has to be preliminary chosen, i.e., the index q of the Minkowski norm involved in the proposed MRF model. A remarkable stability of the change-detection results as functions of q has been pointed out by the experiments. This suggests that preliminary choosing a proper value for q is not critical. In any case, the automation of this parameter-setting aspect could be an interesting future extension of this work. A further future generalization could be the integration of multispectral data and/or of other (e.g., ancillary) information sources, by suitably extending the adopted MRF-based data fusion approach.

5 Acknowledgment

The authors would like to thank Dr. Paolo Gamba from the University of Pavia (Italy) for providing the data employed for experiments and Mr. Andrea Sciarrone for his support to implementation.

Appendix

The mode-field EM algorithm expresses the parameter estimation problem as the iterative maximization of a modified log-likelihood function [2]. Adopting the notations in Secs. 2.1-2.3, the mode-field EM algorithm for the MRF model defined by Eqs. (1) and (2) computes the updated parameter vector θ^{t+1} at the t -th iteration by maximizing, with respect to θ , the following, pseudo-likelihood function [2]:

$$Q(\theta|\theta^t) = \sum_{k=1}^N \sum_{i=0}^1 w_{ik}^t \left[\ln P\{\ell_k = H_i | \mathcal{C}_k^t, \beta\} + \sum_{r=1}^n \alpha_r \ln p_{ir}(u_{kr} | \xi_{ir}) \right], \quad (8)$$

where $w_{ik}^t = P\{\ell_k = H_i | u_k, \mathcal{C}_k^t, \theta^t\}$. Thanks to the Markovianity of the label process, the local prior probability distribution $P\{\ell_k = H_i | \mathcal{C}_k^t, \beta\}$ of the k -th pixel label ℓ_k conditioned to its context is proved to be ($k = 1, 2, \dots, N$) [6]:

$$P\{\ell_k = H_i | \mathcal{C}_k^t, \beta\} = \frac{\exp(\beta m_{ik}^t)}{\sum_{j=0}^1 \exp(\beta m_{jk}^t)}, \quad (9)$$

Plugging Eq. (9) into Eq. (8), one may obtain after standard algebraic manipulations:

$$Q(\theta|\theta^t) = \Phi_0^t(\beta) + \sum_{r=1}^n \alpha_r \sum_{i=0}^1 \Phi_{ir}^t(\xi_{ir}) \quad (10)$$

where

$$\begin{cases} \Phi_0^t(\beta) = \sum_{k=1}^N \left[\beta \sum_{i=0}^1 w_{ik}^t m_{ik}^t - \ln \sum_{i=0}^1 \exp(\beta m_{ik}^t) \right] \\ \Phi_{ir}^t(\xi_{ir}) = \sum_{k=1}^N w_{ik}^t \ln p_{ir}(u_{kr} | \xi_{ir}). \end{cases} \quad (11)$$

The function in Eq. (10) is maximized with respect to θ , subject to the constraint in Eq. (4) on $\alpha_1, \alpha_2, \dots, \alpha_n$, to a positivity constraint on β and to possible further constraints on each parameter vector ξ_{ir} (e.g., in the case of NR, both components of $\xi_{ir} = (L_{ir}, \gamma_{ir})$ are constrained to be positive; see Table 1). Since only the first term $\Phi_0^t(\beta)$ depends on β , the expression in Eq. (5) for the updated spatial parameter β^{t+1} immediately follows.

Concerning the PDF parameters, since ξ_{ir} is included only in the term $\Phi_{ir}^t(\xi_{ir})$, which is weighted in Eq. (10) by the positive factor α_r , the mode-field EM algorithm updates ξ_{ir} as follows ($r = 1, 2, \dots, n; i = 0, 1$):

$$\xi_{ir}^{t+1} = \arg \max_{\xi_{ir}} \Phi_{ir}^t(\xi_{ir}). \quad (12)$$

If the LN model is used, then $\xi_{ir} = (\mu_{ir}, \sigma_{ir})$ and it is possible to prove that this maximization process yields ($r = 1, 2, \dots, n; i = 0, 1$) [2] [12]:

$$\begin{cases} \mu_{1ir}^{t+1} = \frac{\sum_{k=1}^N w_{ik}^t \ln u_{kr}}{\sum_{k=1}^N w_{ik}^t} \\ \sigma_{2ir}^{t+1} = \left[\frac{\sum_{k=1}^N w_{ik}^t (\ln u_{kr} - \mu_{1ir}^{t+1})^2}{\sum_{k=1}^N w_{ik}^t} \right]^{1/2} \end{cases} \quad (13)$$

Since the parameters of LN are exactly the logarithmic mean and variances of the distribution, this actually proves Eq. (5) in the LN case.

On the other hand, if either NR or WR are adopted, the maximization process in Eq. (12) does not yield a close-form solution [11]. On the contrary, the MoLC approach is feasible for such parametric families. Hence, we plug this approach in the EM iterative process. Specifically, Eq. (5), which has just been proved in the case of LN, is still applied to compute updated estimates of the logarithmic mean κ_{1ir} and variance κ_{2ir} also in the cases of NR and WR. Corresponding estimates of the parameters of the NR and WR distributions are then derived by solving the related MoLC equations (see Table 1). A similar hybrid MoLC-EM approach was applied in [13] for probability modelling purposes.

The vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ of the reliability factors is involved only in the double summation term in Eq. (10). Hence, plugging the updated PDF parameter estimates ξ_{ir}^{t+1} ($r = 1, 2, \dots, n; i = 0, 1$) into Eq. (10) and taking into account the definition of the c_r^t coefficients ($r = 1, 2, \dots, n$; see Eq. (7)), the vector α^{t+1} is obtained by maximizing the following function:

$$F^t(\alpha) = \sum_{r=1}^n \alpha_r \sum_{i=0}^1 \Phi_{ir}^t(\xi_{ir}^{t+1}) = \sum_{r=1}^n c_r^t \alpha_r, \quad (14)$$

subject to the following constraint:

$$G(\alpha) = \|2\alpha - \mathbf{1}\|_q^q - 1 = \sum_{r=1}^n (2\alpha_r - 1)^q - 1 \leq 0. \quad (15)$$

The domain $\Omega = \{\alpha \in \mathbb{R}^n : G(\alpha) \leq 0\}$ defined by the constraint is a compact subset of \mathbb{R}^n and F^t is continuous on Ω^t : hence, a global constrained maximum point of F^t exists, thanks to the Weierstrass theorem [20]. Both F^t and G are differentiable and F^t has a constant and (in general) nonzero gradient. Hence, the constrained maximum point will lie on the boundary of Ω , i.e., it will satisfy the equality condition $G(\alpha) = 0$. According to the Lagrange multiplier theorem, a necessary condition for α to be a solution of this constrained maximization problem is that:

$$\frac{\partial F^t(\alpha)}{\partial \alpha_r} + \lambda \frac{\partial G(\alpha)}{\partial \alpha_r} = 0, \quad r = 1, 2, \dots, n, \quad (16)$$

where λ is a Lagrange multiplier. Therefore:

$$c_r^t + \lambda \cdot 2q(2\alpha_r - 1)^{q-1} = 0, \quad r = 1, 2, \dots, n, \quad (17)$$

and, assuming q to be an even integer number (and so $(q-1)$ to be odd), we obtain:

$$2\alpha_r - 1 = -\sqrt[q-1]{\frac{c_r^t}{2\lambda q}}, \quad r = 1, 2, \dots, n. \quad (18)$$

Plugging this result into the equality constraint, we have:

$$G(\alpha) = \frac{1}{|2\lambda q|^{q/(q-1)}} \sum_{r=1}^n |c_r^t|^{q/(q-1)} - 1 = \left(\frac{\|c^t\|_{q'}}{|2\lambda q|} \right)^q - 1 = 0 \implies |2\lambda q| = \|c^t\|_{q'}. \quad (19)$$

This yields two solutions α^+ and α^- , i.e.:

$$\alpha_r^\pm = \frac{1}{2} \pm \frac{1}{2} {}^{q-1}\sqrt{\frac{c_r^t}{\|c^t\|_{q'}}}, \quad r = 1, 2, \dots, n. \quad (20)$$

They correspond to local extrema of F^t subject to the constraint defined by G . In order to choose the maximum point, we substitute into Eq. (14), which yields, after standard algebraic calculations:

$$F^t(\alpha^\pm) = \frac{1}{2} \left(\sum_{r=1}^n c_r^t \pm \|c^t\|_{q'} \right). \quad (21)$$

Hence, α^+ is the solution corresponding to the maximum constrained value of F^t and the reliability factors are updated as in Eq. (6).

If LJ-EM is used, the same algorithmic procedure as in the case of EM is applied, but the weights w_{ik}^t ($k = 1, 2, \dots, N; i = 0, 1$) are modified as discussed in Sec. 2.3 in order to reduce the overlapping between the two hypotheses.

We note that the above calculations were feasible thanks to the differentiability of both the objective function F^t and of the constrain G . The latter holds thanks to the replacement of the ∞ -norm by the q -norm. Furthermore, the ∞ -norm constraint in Eq. (3) means that $0 \leq \alpha_r \leq 1$ for $r = 1, 2, \dots, n$, i.e., α belongs to the n -dimensional unitary hypercube $[0, 1]^n$. Since F^t is linear, if the ∞ -norm constraint was used, a linear programming problem would result, and the corresponding solution would be, in general, one of the vertices of the hypercube. Hence, binary values for the reliability factors would be obtained. This is not formally forbidden but is intuitively quite undesired from a data-fusion viewpoint, because this solution does not actually fuse the information conveyed by the n ratio channels, but simply selects a subset of the channels, while discarding the remaining ones.

References

- [1] Y. Bazi, L. Bruzzone, and F. Melgani. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sensing*, 43(4):874–887, 2005.
- [2] G. Celeux, F. Forbes, and N. Peyrand. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recogn.*, 36:131–144, 2003.
- [3] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver. A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Trans. Geosci. Remote Sensing*, 41(1):4–19, 2003.
- [4] R. Cossu, S. Chaudhuri, and L. Bruzzone. A context-sensitive bayesian technique for the partially supervised classification of multitemporal images. *IEEE Geosci. Remote Sensing Lett.*, 2(2):352–356, 2005.

- [5] W. Dierking and H. Skriver. Change detection for thematic mapping by means of airborne multitemporal polarimetric SAR imagery. *IEEE Trans. Geosci. Remote Sensing*, 40(3):618–636, 2002.
- [6] R. C. Dubes and A. K. Jain. Random Field models in image analysis. *J. Appl. Stat.*, 16:131–163, 1989.
- [7] K. Fukunaga. *Introduction to statistical pattern recognition*. 2nd edition, Academic Press, 1990.
- [8] Q. Jackson and D. Landgrebe. Adaptive Bayesian contextual classification based on Markov Random Fields. *IEEE Trans. Geosci. Remote Sensing*, 40(11):2454–2463, 2002.
- [9] Q. Jackson and D. A. Landgrebe. An adaptive classifier design for high-dimensional data analysis with a limited training data set. *IEEE Trans. Geosci. Remote Sensing*, 39(12):2664–2679, 2001.
- [10] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recogn.*, 19:41–47, 1986.
- [11] G. Moser and S. B. Serpico. Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery. *IEEE Trans. Geosci. Remote Sensing*, 40(10):2972–2982, 2006.
- [12] G. Moser, S. B. Serpico, and G. Vernazza. Unsupervised change detection from multichannel SAR images. *IEEE Geosci. Remote Sensing Letters*, 4(2):278–282, 2007.
- [13] G. Moser, J. Zerubia, and S. B. Serpico. Dictionary-based Stochastic Expectation-Maximization for SAR amplitude probability density function estimation. *IEEE Trans. Geosci. Remote Sensing*, 44(1):188–200, 2005.
- [14] G. Moser, J. Zerubia, and S. B. Serpico. SAR amplitude probability density function estimation based on a generalized Gaussian model. *IEEE Trans. Image Process.*, 15(6):1429–1442, 2005.
- [15] J.-M. Nicolas. Introduction aux statistiques de deuxième espèce: applications des log-moments et des log-cumulants a l’analyse des lois d’images radar. *Traitement du Signal (in French)*, 19:139–167, 2002.
- [16] C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. Artech House, Norwood (MA), 1998.
- [17] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, Cambridge (UK), 2002.
- [18] M. Qong. Polarization state conformation and its application to change detection in polarimetric SAR data. *IEEE Geosci. Remote Sensing Lett.*, 1(4):304–308, 2004.
- [19] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

- [20] W. Rudin. *Principles of Mathematical Analysis*. 2nd edition, McGraw-Hill, New York, 1976.
- [21] W. Rudin. *Functional Analysis*. 2nd edition, McGraw-Hill, New York, 1990.
- [22] I. Sneddon. *The use of integral transforms*. McGraw-Hill, New York, 1972.
- [23] A. H. S. Solberg, T. Taxt, and A. K. Jain. A Markov Random Field model for classification of multisource satellite imagery. *IEEE Trans. Geosci. Remote Sensing*, 34(1):100–113, 1996.
- [24] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Ann. Stat.*, 11:95–103, 1983.