Unsupervised Segmentation of Multilook Polarimetric Synthetic Aperture Radar Images

Nizar Bouhlel, Stéphane Méric

Unsupervised Segmentation of Multilook Polarimetric Synthetic Aperture Radar Images

Nizar Bouhlel, *Member, IEEE*, and Stéphane Méric, *Member, IEEE*

***Abstract*— This paper proposes a new unsupervised image segmentation method for multilook polarimetric synthetic aperture radar (PolSAR) data. The statistical model for the PolSAR data is considered as a finite mixture of non-Gaussian compound distributions considered as the product of two statistically independent random variates, speckle, and texture. With different texture distributions, the product model leads to various expressions of the compound distribution. The method uses a Markov random field (MRF) model for pixel class labels. The expectation–maximization/maximization of the posterior marginals (EM/MPM) algorithm is used for the simultaneous estimation of texture and speckle parameters and for the segmentation of multilook PolSAR images. Simulated and real PolSAR data are shown to demonstrate the method.**

***Index Terms*— Expectation–maximization (EM) algorithm, Markov random field (MRF), maximum-likelihood (ML), maximization of the posterior marginals (MPM), polarimetric synthetic aperture radar (PolSAR).**

I. Introduction

CLASSIFICATION plays an important role in polarimetric synthetic aperture radar (PolSAR) image interpretation. The unsupervised classification approaches cluster the PolSAR data into classes according to their statistical characteristics. The Wishart distribution is considered as the most used in the classification and segmentation of low-resolution PolSAR images [1] due to its ease of implementation and low computational cost. For multilook PolSAR data, it is the scaled Wishart distribution that is used as the simplest model for classification. When resolution increases, the homogeneous hypothesis of the PolSAR is not valid anymore and non-Gaussian statistics are observed. Consequently, many heterogeneous models have been proposed in the literature based on the product model. The multivariate product model, introduced by Yueh *et al.* [2], considers the observed signal as the product of two independent random variates: a complex Gaussian speckle and a positive scalar texture. It was shown that a multilook PolSAR speckle followed a matrix-variate-scaled complex Wishart distribution [3]. With varied texture distributions, the product model leads to different expressions of the compound distribution. The following distributions are usually used in the literature to model the texture random variable: gamma (γ), inverse gamma (γ^{-1}), generalized inverse Gaussian (GIG) ($\mathcal{N}^{-1}$), and Fisher ($\mathcal{F}$) with the corresponding compound distributions $\mathcal{K}_d$, $\mathcal{G}_d^0$, $\mathcal{G}_d$, and Kummer-$\mathcal{U}_d$, respectively [4]–[9]. The last two non-Gaussian distributions are more flexible (two texture shape parameters instead of one parameter) and can fit an extremely heterogeneous clutter compared to the first two distributions with only one shape parameter.

As a consequence, polarimetric classifications are performed with non-Gaussian clustering algorithms. Doulgeris *et al.* [10] proposed a non-Gaussian $\mathcal{K}$-Wishart clustering algorithm. It was based upon the iterative expectation–maximization (EM) algorithm [11] for finite mixture modeling using the $\mathcal{K}_d$ probability density function (pdf) for the probabilities and the method of matrix log-cumulants (MoMLC) [12] for class parameter updates [10]. The method was extended in [13] to incorporate spatial contextual information for multilook PolSAR data. Their method was based on a Markov random field (MRF) model that integrated a $\mathcal{K}_d$ distribution for the PolSAR data statistics conditioned to each image cluster and a Potts model for the spatial context. The latter approach was recently proposed with the Kummer-$\mathcal{U}_d$ distribution [14]. Another unsupervised clustering algorithm was presented by Bombrun *et al.* [15] and based on the hierarchical segmentation algorithm adapted to the Kummer-$\mathcal{U}_d$ distributed target scattering vector. Khan and Doulgeris *et al.* [16] used the flexible $\mathcal{G}_d$ distribution to model the statistics of each image cluster and the method of multivariate fractional moment (MoMFM) to estimate the model parameters. The unsupervised clustering was achieved using the EM algorithm.

Parameter estimation is very important in the classification of PolSAR images. The underestimation of the parameters affects the quality of the classification, and then the analysis of the PolSAR images. For this, efficient parameter estimation of the compound distributions for each cluster requires careful consideration. Three parameters need to be estimated: 1) the equivalent number of looks (ENL); 2) the covariance matrix; and 3) the texture parameters. In the literature, there are several manners to estimate the parameters of the compound distribution. The most popular method is the

Manuscript received December 7, 2018; revised February 12, 2019; accepted March 7, 2019. *(Corresponding author: Nizar Bouhlel.)*

N. Bouhlel was with the SAR and Hyperspectral Multi-Modal Imaging and Signal Processing, Electromagnetic Modeling Team (SHINE), Institute of Electronics and Telecommunications of Rennes (IETR), National Institute for the Applied Sciences (INSA), 35043 Rennes, France. He is now with the Lab-STICC, École Nationale Supérieure de Techniques Avancées Bretagne (ENSTA), 29806 Brest, France (e-mail: nizar.bouhlel@ensta-bretagne.fr).

S. Méric is with Image and Remote Sensing, Institut d'Electronique et des Télécommunications de Rennes, Université de Rennes 1, UMR CNRS 6164, 35708 Rennes, France.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2904401

MoMLC for multilook polarimetric parameter estimation [12]. With the MoMLC, the ENL and the texture parameters are estimated from the first- and second-order MLC equations when the distribution has only one texture parameter, as $\mathcal{K}_d$ and $\mathcal{G}_d^0$ distributions [9], [12] and from the first-, second-, and third-order MLC equations when the distribution has two texture parameters, as $\mathcal{G}_d$ and Kummer-$\mathcal{U}_d$ distributions. The last two distributions involve higher order derivatives of special functions. Khan and Guida *et al.* [17] showed that the MoMFM had a lower bias and variance than the MoMLC-based estimators. Nevertheless, these estimators, MLC and MFM, were not maximum-likelihood (ML) parameter estimators. The latter benefits from very desirable mathematical properties (asymptotic efficiency). Usually, the ML-based parameter estimators are not available with closed-form solutions. Moreover, a developed method based on the EM algorithm was presented in [18] and proposed to compute the ML estimates of the unknown parameters of different compound distributions. The method showed a good performance in terms of bias and mean square error compared to the performance of other known estimators such as the MLC and MFM estimators.

The classification method of interest in this paper is a novel formulation of the unsupervised EM algorithm for finite mixture modeling using non-Gaussian compound distributions and ML parameter estimators for class parameter updates [18]. In particular, we estimate: 1) the ENL; 2) the covariance matrix of the speckle component; and 3) the texture distribution parameters. The EM algorithm is developed to iteratively solve the problem of finding the ML estimator when an exact algebraic solution is not possible. The EM algorithm interprets the multilook covariance matrix as incomplete data with a missing class label and a missing texture. The following $\mathcal{K}_d$, $\mathcal{G}_d^0$, and $\mathcal{G}_d$ pdfs are used as compound distributions describing the statistics of multilook polarimetric SAR data. Since the MRF have been successfully used in image segmentation to integrate spatial information, it is adopted in this paper to improve the accuracy and quality of the segmentation. The class posterior probabilities are estimated using the maximization of the posterior marginals (MPM) algorithm. In summary, our approach is based on EM/MPM where all parameters are estimated by EM and the class labels are updated by MPM.

The rest of this paper is organized as follows. Section II introduces the statistical product model for multilook PolSAR data. The application to parameter estimation and class label update for multilook PolSAR data using the EM algorithm is presented in Section III. The content of Section IV is related to the concepts of MRFs and the chosen MPM algorithm for class posterior probability estimates. Section V demonstrates the performance of the approach with the simulated and real multilook PolSAR data. Finally, Section VI concludes this paper.

II. Product Model for Multilook Polarimetric Radar Data

The polarimetric SAR system is characterized by the following scattering vector $\mathbf{s} = [s_{ij}]^T \in \mathbb{C}^d$, where $i = \{h, v\}$ is the transmit "horizontal" or "vertical" polarization and $j = \{h, v\}$ is the receive "horizontal" or "vertical" polarization [19], $[.]^T$ means transposition, and $d = \dim(\mathbf{s})$ is the vector dimension. In order to reduce the speckle noise, the polarimetric multilooking operation is applied and given by

$$\mathbf{C} = \frac{1}{L}\sum_{\ell=1}^{L}\mathbf{s}_\ell\mathbf{s}_\ell^H, \quad L \geq d \tag{1}$$

where L is the number of looks, $(.)^H$ denotes the Hermitian operator, and $\mathbf{C} \in \mathbf{\Omega}_+ \subset \mathbb{C}^{d\times d}$ is the multilook polarimetric covariance matrix defined on the cone $\mathbf{\Omega}_+$ of the positive definite complex Hermitian matrices. According to the multilook polarimetric product model, the covariance matrix $\mathbf{C}$ is the product of texture τ and speckle $\mathbf{X}$, and is given as follows:

$$\mathbf{C} = \tau\mathbf{X}. \tag{2}$$

The texture τ is a positive scalar random variable with a pdf $f_\tau(\tau)$. The speckle $\mathbf{X}$ is a random matrix following a scaled complex Wishart distribution $s\mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$ with a pdf given by the following:

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{L^{Ld}|\mathbf{X}|^{L-d}}{\Gamma_d(L)|\mathbf{\Sigma}|^L}\text{etr}(-L\mathbf{\Sigma}^{-1}\mathbf{X}) \tag{3}$$

where $\mathbf{\Sigma}$ is the covariance matrix of the speckle, $\text{etr}(.) = \exp[\text{tr}(.)]$ is the exponential trace operator, $|.|$ is the determinant operator, and $\Gamma_d(L)$ is the multivariate gamma function of the complex kind defined as follows:

$$\Gamma_d(L) = \pi^{d(d-1)/2}\prod_{i=0}^{d-1}\Gamma(L-i). \tag{4}$$

The function $\Gamma(L)$ is the standard Euler gamma function. The pdf of $\mathbf{C}$ using Bayes' theorem becomes

$$f_{\mathbf{C}}(\mathbf{C}) = \int_0^{+\infty} f_{\mathbf{C}|\tau}(\mathbf{C}|\tau)\, f_\tau(\tau)d\tau \tag{5}$$

where $f_{\mathbf{C}|\tau}(\mathbf{C}|\tau)$ is the pdf of $\mathbf{C}$ with a specific value of τ and given by

$$f_{\mathbf{C}|\tau}(\mathbf{C}|\tau) = \frac{L^{Ld}|\mathbf{C}|^{L-d}}{\Gamma_d(L)|\mathbf{\Sigma}|^L}\frac{1}{\tau^{dL}}\text{etr}\left(-\frac{L}{\tau}\mathbf{\Sigma}^{-1}\mathbf{C}\right) \tag{6}$$

Therefore, the pdf of $\mathbf{C}$ is obtained by substituting (6) in (5)

$$f_{\mathbf{C}}(\mathbf{C}) = \frac{L^{Ld}|\mathbf{C}|^{L-d}}{\Gamma_d(L)|\mathbf{\Sigma}|^L}\int_0^{+\infty}\frac{1}{\tau^{dL}}\text{etr}\left(-\frac{L}{\tau}\mathbf{\Sigma}^{-1}\mathbf{C}\right) f_\tau(\tau)d\tau \tag{7}$$

when the texture distribution is one of the following: a normalized gamma to a unit mean ($\bar{\gamma}$), a normalized inverse gamma to a unit mean ($\bar{\gamma}^{-1}$) and GIG ($\mathcal{N}^{-1}$), and the covariance matrix pdf is respectively, the matrix-variate $\mathcal{K}_d$, $\mathcal{G}_d^0$, and $\mathcal{G}_d$ distributions. Table I lists these distributions as possible choices of $f_\tau(\tau)$. It is worth noticing that the matrix variate $\mathcal{G}_d$ distribution is normalized by dividing the covariance matrix of the speckle $\mathbf{\Sigma}$ by $\text{tr}(\mathbf{\Sigma})/d$ [20].

TABLE I
TEXTURE AND COVARIANCE MATRIX DISTRIBUTION UNDER MULTILOOK POLARIMETRIC PRODUCT MODEL

Texture pdf		Compound pdf		Ref.
Symbol	$f_\tau(\tau)$	Symbol	$f_{\mathbf{C}}(\mathbf{C})$	
$\bar{\gamma}(\alpha)$	$\frac{\alpha^\alpha \tau^{\alpha-1}}{\Gamma(\alpha)} \exp(-\alpha\tau)$ $\tau, \alpha \in \mathbb{R}_+$	$\mathcal{K}_d(\mathbf{\Sigma}, L, \alpha)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^L} \frac{2\alpha^{\frac{\alpha+Ld}{2}}}{\Gamma(\alpha)} (L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}))^{\frac{\alpha-Ld}{2}} \times$ $K_{\alpha-Ld}\left(2\sqrt{L\alpha\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\right)$	[12]
$\bar{\gamma}^{-1}(\lambda)$	$\frac{(\lambda-1)^\lambda}{\Gamma(\lambda)} \frac{1}{\tau^{1+\lambda}} \exp(-\frac{\lambda-1}{\tau})$ $\tau \in \mathbb{R}_+^*, \lambda > 1$	$\mathcal{G}_d^0(\mathbf{\Sigma}, L, \lambda)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^L} \frac{(\lambda-1)^\lambda \Gamma(dL+\lambda)}{\Gamma(\lambda)} \left(L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\right)^{-(dL+\lambda)}$	[6]
$\mathcal{N}^{-1}(\alpha, w, \eta)$	$\frac{1}{\eta^\alpha 2K_\alpha(w)} \tau^{\alpha-1} \exp(-\frac{w}{2}(\frac{\eta}{\tau} + \frac{\tau}{\eta}))$ $\eta, \tau, w \in \mathbb{R}^+, \alpha \in \mathbb{R}$	$\mathcal{G}_d(\mathbf{\Sigma}, L, \alpha, w, \eta)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\eta^\alpha \Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^L} \frac{1}{K_\alpha(w)} \left(\frac{2L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})+w\eta}{w/\eta}\right)^{\frac{\alpha-Ld}{2}} \times$ $K_{\alpha-Ld}\left(\sqrt{\frac{w}{\eta}(2L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + w\eta)}\right)$	[9]

III. EM FOR PARAMETER ESTIMATION AND CLASS LABEL UPDATE

We consider $\mathbf{C} = \{\mathbf{C}_i, 1 \le i \le N\}$ a set of independent and identically distributed (i.i.d.) random matrices. We assume that the pdf of the random matrix $\mathbf{C}_i$ is described by the following finite mixture of M distributions:

$$f_{\mathbf{C}_i}(\mathbf{C}_i|\Theta) = \sum_{m=1}^{M} \pi_m f_{\mathbf{C}_i|X_i}(\mathbf{C}_i|\boldsymbol{\theta}_m) \tag{8}$$

where X represents a set of hidden random labels $\{X_i, 1 \le i \le N\}$, each achievement x_i of X_i takes one of M class labels, i.e., $x_i \in \{1, 2, \ldots, M\}$, π_m corresponds to the mixture weight of each component and represents the prior probabilities $P_{X_i}(x_i = m) = \pi_m$ and verifies $\pi_m \ge 0$ and $\sum_{m=1}^{M} \pi_m = 1$, $f_{\mathbf{C}_i|X_i}(\mathbf{C}_i|\boldsymbol{\theta}_m)$ is a compound distribution describing component m in the form given by (7) with $\boldsymbol{\theta}_m = (L, \mathbf{\Sigma}_m, \theta_{\tau,m})$ where $\theta_{\tau,m}$ are the parameters of the texture distribution of partition m, and Θ is a vector of the whole set of parameters of a given M-mixture to be estimated and is defined as $\Theta = \{\pi_1, \pi_2, \ldots, \pi_M, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$. The number of looks is considered the same for all clusters. We suppose that each covariance matrix $\mathbf{C}_i$ has a corresponding unobserved and hiding texture τ_i. We assume that the sequence $\{\tau_i, 1 \le i \le N\}$ is also i.i.d. random variables. The joint distribution of i.i.d. random matrices is given by

$$\begin{aligned} &f_{\mathbf{C}}(\mathbf{C}|\Theta) \\ &= \prod_{i=1}^{N} \sum_{m=1}^{M} \pi_m f_{\mathbf{C}_i|X_i}(\mathbf{C}_i|\boldsymbol{\theta}_m) \\ &= \prod_{i=1}^{N} \sum_{m=1}^{M} \pi_m \int_0^{+\infty} f_{\mathbf{C}_i|\tau_i,X_i}(\mathbf{C}_i|\tau_i, \boldsymbol{\theta}_m) f_{\tau_i|X_i}(\tau_i|\boldsymbol{\theta}_m) d\tau_i. \end{aligned} \tag{9}$$

To estimate the vectors of the parameters Θ, the maximum of the log-likelihood estimator $\hat{\Theta}$ is the following general choice:

$$\hat{\Theta} = \arg\max_{\Theta} \ln f_{\mathbf{C}}(\mathbf{C}|\Theta) \tag{10}$$

$$= \arg\max_{\Theta} \sum_{i=1}^{N} \ln\left(\sum_{m=1}^{M} \pi_m f_{\mathbf{C}_i|X_i}(\mathbf{C}_i|\boldsymbol{\theta}_m)\right). \tag{11}$$

It is well known that the ML estimate cannot be found analytically. The usual choice is the EM algorithm, which interprets $\mathbf{C}$ as the incomplete data with the missing class label X and the missing texture τ. With the complete data set $(\mathbf{C}, X, \tau)$, we can define the expected value of the complete-data log-likelihood $\ln f_{\mathbf{C},X,\tau}(\mathbf{C}, X, \tau|\Theta)$ with respect to the missing data X and τ given the observed covariance matrix $\mathbf{C}$ and the current parameter estimates Θ'. The EM can provide the estimation $\hat{\Theta}$ given a current estimate Θ'. The maximization of the conditional expectation of the complete log-likelihood according to parameters Θ is given by

$$\hat{\Theta} = \arg\max_{\Theta} E_{X,\tau|\mathbf{C}}\{\ln f_{\mathbf{C},X,\tau}(\mathbf{C}, X, \tau|\Theta)|\mathbf{C}, \Theta'\} \tag{12}$$

$$\begin{aligned} &= \arg\max_{\Theta} E_{X,\tau|\mathbf{C}}\left\{\sum_{i=1}^{N} \ln f_{\mathbf{C}_i,X_i,\tau_i}(\mathbf{C}_i, X_i, \tau_i|\Theta)|\mathbf{C}_i, \Theta'\right\} \\ &= \arg\max_{\Theta} \sum_{i=1}^{N} E_{X_i,\tau_i|\mathbf{C}_i}\{\ln f_{\mathbf{C}_i,\tau_i|X_i}(\mathbf{C}_i, \tau_i|X_i, \Theta)|\mathbf{C}_i, \Theta'\} \\ &\quad + \sum_{i=1}^{N} E_{X_i,\tau_i|\mathbf{C}_i}\{\ln P_{X_i}(X_i|\Theta)|\mathbf{C}_i, \Theta'\}. \end{aligned} \tag{13}$$

The conditional expectation of the complete log-likelihood given by (13) contains two independent terms, one depending on π_m and the other on $\boldsymbol{\theta}_m$, which can, therefore, be maximized separately. The first term of (13) is written as follows:

$$\begin{aligned} &E_{X_i,\tau_i|\mathbf{C}_i}\{\ln f_{\mathbf{C}_i,\tau_i|X_i}(\mathbf{C}_i, \tau_i|X_i, \Theta)|\mathbf{C}_i, \Theta'\} \\ &\quad = \sum_{m=1}^{M} \int_0^{+\infty} \ln f_{\mathbf{C}_i,\tau_i|X_i}(\mathbf{C}_i, \tau_i|m, \Theta) \\ &\qquad \times f_{X_i,\tau_i|\mathbf{C}_i}(x_i = m, \tau_i|\mathbf{C}_i, \Theta') d\tau_i \\ &\quad = \sum_{m=1}^{M} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') \\ &\qquad \times E_{\tau_i|X_i,\mathbf{C}_i}\{\ln f_{\mathbf{C}_i,\tau_i|X_i}(\mathbf{C}_i, \tau_i|m, \Theta)|m, \mathbf{C}_i, \Theta'\}. \end{aligned} \tag{14}$$

We used the propriety that $f_{X_i,\tau_i|\mathbf{C}_i}(x_i = m, \tau_i|\mathbf{C}_i, \Theta') = f_{\tau_i|X_i,\mathbf{C}_i}(\tau_i|m, \mathbf{C}_i, \Theta') P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')$ with $P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')$ being the posterior probability of X_i belonging to the mth component given $\mathbf{C}_i$ and written as

$$P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') = \frac{\pi_m f_{\mathbf{C}_i|X_i}\left(\mathbf{C}_i|m, \boldsymbol{\theta}'_m\right)}{\sum_{k=1}^{M} \pi_k f_{\mathbf{C}_i|X_i}\left(\mathbf{C}_i|k, \boldsymbol{\theta}'_k\right)}. \tag{15}$$

Let $\boldsymbol{\theta}_{\mathbf{X}} = (L, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M)$, $\theta_\tau = (\theta_{\tau,1}, \theta_{\tau,2}, \ldots, \theta_{\tau,M})$ and using the fact that $f_{\mathbf{C}_i,\tau_i|X_i}(\mathbf{C}_i, \tau_i|m, \Theta) = f_{\mathbf{C}_i|\tau_i,X_i}(\mathbf{C}_i|\tau_i, m, \Theta) f_{\tau_i|X_i}(\tau_i|m, \Theta)$, (14) can then be maximized separately according to the speckle parameter $\hat{\boldsymbol{\theta}}_{\mathbf{X}}$ and the texture parameter $\hat{\theta}_\tau$ as follows:

$$\hat{\boldsymbol{\theta}}_{\mathbf{X}} = \arg\max_{\boldsymbol{\theta}_{\mathbf{X}}} \sum_{i=1}^{N}\sum_{m=1}^{M} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') \times E_{\tau_i|X_i,\mathbf{C}_i}\{\ln f_{\mathbf{C}_i|\tau_i,X_i}(\mathbf{C}_i|\tau_i, m, \boldsymbol{\theta}_{\mathbf{X}})|m, \mathbf{C}_i, \Theta'\} \quad (16)$$

$$\hat{\theta}_\tau = \arg\max_{\theta_\tau} \sum_{i=1}^{N}\sum_{m=1}^{M} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') \times E_{\tau_i|X_i,\mathbf{C}_i}\{\ln f_{\tau_i|X_i}(\tau_i|m, \theta_\tau)|m, \mathbf{C}_i, \Theta'\}. \quad (17)$$

A. Estimation of Speckle Parameters

Substituting (6) in (16), differentiating with respect to L, and $\boldsymbol{\Sigma}_m$ by taking into account the following properties [21]: $\partial \ln|\boldsymbol{\Sigma}|/\partial\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}^{-1})^T$ and $\partial \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{W})/\partial\boldsymbol{\Sigma} = -(\boldsymbol{\Sigma}^{-1}\mathbf{W}\boldsymbol{\Sigma}^{-1})^T$, and setting the result to zero yields

$$\hat{\boldsymbol{\Sigma}}_m = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') g_i^{(m)} \mathbf{C}_i}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')} \quad (18)$$

$$d\ln\hat{L} + d - \psi_d(\hat{L}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') \times \left(g_i^{(m)} tr\left(\hat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{C}_i\right) - \ln\left|\hat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{C}_i\right| + d a_i^{(m)}\right) \quad (19)$$

where $a_i^{(m)}$, $g_i^{(m)}$, and $h_i^{(m)}$ are three posterior expectations defined by

$$a_i^{(m)} = E_{\tau_i|\mathbf{C}_i,X_i}\{\ln\tau_i|\mathbf{C}_i, X_i = m, \Theta'\} \quad (20)$$

$$g_i^{(m)} = E_{\tau_i|\mathbf{C}_i,X_i}\left\{\frac{1}{\tau_i}|\mathbf{C}_i, X_i = m, \Theta'\right\} \quad (21)$$

$$h_i^{(m)} = E_{\tau_i|\mathbf{C}_i,X_i}\{\tau_i|\mathbf{C}_i, X_i = m, \Theta'\}. \quad (22)$$

The function $\psi_d(L)$ is the multivariate digamma function defined as $\psi_d(L) = \partial \ln\Gamma_d(L)/\partial L = \sum_{j=0}^{d-1}\psi(L-j)$, where $\psi(.)$ is the digamma function. Equation (18) is the ML estimator of the covariance matrix of the speckle for the partition m and (19) is the equation needed to solve for the ENL. An explicit solution for L is not possible from (19) and needs to be solved numerically by using the trust region method [22]. Specifically, the *fsolve* function in MATLAB is utilized to obtain the parameter estimate.

B. Estimation of Prior Probabilities

The second term of (13) is written as

$$E_{X_i,\tau_i|\mathbf{C}_i}\{\ln P_{X_i}(X_i|\Theta)\Big|\mathbf{C}_i, \Theta'\} = \sum_{m=1}^{M}\int_0^{+\infty}\ln(\pi_m) f_{X_i,\tau_i|\mathbf{C}_i}(m, \tau_i|\mathbf{C}_i, \Theta')d\tau_i = \sum_{m=1}^{M}\ln(\pi_m) P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta'). \quad (23)$$

One can introduce the Lagrange multiplier Λ with the constraint $\sum_{m=1}^{M}\pi_m = 1$ and find the estimate of π_m as follows:

$$\hat{\theta}_\pi = \arg\max_{\theta_\pi} \sum_{i=1}^{N}\sum_{m=1}^{M}\ln(\pi_m) P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') + \Lambda\left(\sum_{m=1}^{M}\pi_m - 1\right) \quad (24)$$

where $\theta_\pi = \{\pi_1, \pi_2, \ldots, \pi_M\}$. Using the fact that $\sum_{m=1}^{M} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta') = 1$, we get $\Lambda = -N$ resulting in

$$\hat{\pi}_m = \frac{1}{N}\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta'). \quad (25)$$

C. Estimation of Texture Parameters

The texture distribution parameters $\hat{\theta}_\tau$ are estimated using (17), which need $f_{\tau_i|X_i}(\tau_i|m, \theta_\tau)$ and $f_{\tau_i|X_i,\mathbf{C}_i}(\tau_i|\mathbf{C}_i, m, \Theta')$. In the following, three texture pdfs are used.

1) Gamma Distribution: Substituting $\bar{\gamma}(\alpha_m)$ in (17), differentiating with respect to α_m and setting the result to zero yields

$$\ln(\hat{\alpha}_m) - \psi(\hat{\alpha}_m) + 1 = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')\left(h_i^{(m)} - a_i^{(m)}\right)}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')}. \quad (26)$$

An explicit solution for α_m is not possible from the ML equation. The solution is found using the trust region method. As stated before, the ML equation depends on the posterior pdf of the texture $f_{\tau_i|\mathbf{C}_i,X_i}(\tau_i|\mathbf{C}_i, m)$ through $a_i^{(m)}$ and $h_i^{(m)}$. Using Bayes' rule, the posterior pdf of τ_i given $X_i = m$ and $\mathbf{C}_i$ is given as follows:

$$f_{\tau_i|\mathbf{C}_i,X_i}(\tau_i|\mathbf{C}_i, m) = \frac{f_{\mathbf{C}_i|\tau_i,X_i}(\mathbf{C}_i|\tau_i, m) f_{\tau_i|X_i}(\tau_i|m)}{f_{\mathbf{C}_i|X_i}(\mathbf{C}_i|m)}. \quad (27)$$

The expression of $f_{\tau_i|\mathbf{C}_i,X_i}(\tau_i|\mathbf{C}_i, m)$ corresponds to the GIG distribution (see Appendix B), with parameters ($\alpha_1 = \alpha_m - dL$, $w_1 = 2\sqrt{L\alpha_m \mathrm{tr}(\boldsymbol{\Sigma}_m^{-1}\mathbf{C}_i)}$, $\eta_1 = \sqrt{\frac{L}{\alpha_m}\mathrm{tr}(\boldsymbol{\Sigma}_m^{-1}\mathbf{C}_i)}$). The posterior expectation expressions in terms of these parameters are easily calculated using the formulas given in Appendix B.

2) Inverse Gamma Distribution: Similarly, substituting $\bar{\gamma}^{-1}(\lambda_m)$ in (17), differentiating with respect to λ_m, and setting the result to zero yields

$$\ln(\hat{\lambda}_m - 1) - \psi(\hat{\lambda}_m) + \frac{\hat{\lambda}_m}{\hat{\lambda}_m - 1} = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')\left(g_i^{(m)} + a_i^{(m)}\right)}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')}. \quad (28)$$

The posterior pdf of τ_i given $\mathbf{C}_i$ and $X_i = m$ corresponds to an inverse gamma distribution with parameters ($\alpha_1 = dL + \lambda_m$, $\beta_1 = L\mathrm{tr}(\boldsymbol{\Sigma}_m^{-1}\mathbf{C}_i) + \lambda_m - 1$). The posterior expectation expression in terms of these parameters are calculated by the formulas given in Appendix C.

3) Generalized Inverse Gaussian Distribution: Substituting $\mathcal{N}^{-1}(\alpha_m, \omega_m, \eta_m)$ in (17), differentiating with respect to α_m, ω_m and η_m, and setting the result to zero yields

$$\frac{K_{\hat{\alpha}_m-1}(\hat{\omega}_m)+K_{\hat{\alpha}_m+1}(\hat{\omega}_m)}{K_{\hat{\alpha}_m}(\hat{\omega}_m)} = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')\left(\hat{\eta}_m g_i^{(m)}+\frac{h_i^{(m)}}{\hat{\eta}_m}\right)}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')} \tag{29}$$

$$\ln\hat{\eta}_m + \frac{1}{K_{\hat{\alpha}_m}(\hat{\omega}_m)}\frac{\partial K_{\hat{\alpha}_m}(\hat{\omega}_m)}{\partial\alpha} = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')a_i^{(m)}}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')} \tag{30}$$

$$2\frac{\hat{\alpha}_m}{\hat{\omega}_m} = \frac{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')\left(\frac{h_i^{(m)}}{\hat{\eta}_m}-\hat{\eta}_m g_i^{(m)}\right)}{\sum_{i=1}^{N} P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')}. \tag{31}$$

The posterior pdf of τ_i given $\mathbf{C}_i$ and $X_i = m$ is a GIG distribution with parameters $(\alpha_1 = \alpha_m - dL, w_1 = \sqrt{\frac{w_m}{\eta_m}(2L\mathrm{tr}(\mathbf{\Sigma}_m^{-1}\mathbf{C}_i)+w_m\eta_m)}, \eta_1 = \sqrt{\frac{2L\mathrm{tr}(\mathbf{\Sigma}_m^{-1}\mathbf{C}_i)+w_m\eta_m}{w_m/\eta_m}})$. The posterior expectation expressions in terms of these parameters are easily calculated by the formulas in Appendix B.

IV. Markov Random Field Model

The MRF has been used successfully for various image processing applications. In image segmentation, where the problem is reduced to assigning labels to pixels, label dependencies are modeled by MRF and an optimal labeling is determined by Bayesian estimation. The main advantage of MRF models is that prior information can be imposed locally through clique potentials [23].

The MRF models are defined via parameterized energy functions that characterize local interactions between neighboring pixels. The energy of a Gibbs distribution is expressed as the sum of several potential functions over all possible cliques $\mathcal{C}$, each ascribed to the cliques of a certain size. Here, only the cliques of size up to two are considered. Readers are referred to [24], [25] for details of MRF models. Let $X = \{X_s, \forall s \in S\}$ be the hidden label field. For a regular rectangular lattice S, X is considered as a family of random variables defined on the set S where each random variable X_s at a spatial location s takes a value x_s in the configuration set $E_s = \{1, 2, \ldots, M\}$. The notation $(X_s = x_s)$ is used to denote the event that X_s takes the value x_s and the notation $(X = x)$ to denote the joint event. Let the unobserved class label X be an MRF defined on S with respect to a neighborhood system $\mathcal{N} = \{\mathcal{N}_s, \forall s \in S\}$ where $\mathcal{N}_s$ is the set of sites neighboring s. The joint probability of X has the following form:

$$P_X(x) = \frac{1}{Z}\exp\left(-\sum_{\{s,r\}\in\mathcal{C}} V_c(x_s, x_r)\right) \tag{32}$$

where $\mathcal{C}$ is the set of all cliques. In this paper, it will include all pairs of spatially horizontal or vertical adjacent pixels. The normalizing constant Z is the sum of all possible realizations of the random field X and has the form of

$$Z = \sum_x \exp\left(-\sum_{\{s,r\}\in\mathcal{C}} V_c(x_s, x_r)\right) \tag{33}$$

with $V_c(.)$ being the clique potential associated with a clique c. The potential function is assumed to have the form

$$V_c(x_s, x_r) = \beta_c t(x_s, x_r), \quad t(x_s, x_r) = \begin{cases} 0 & \text{if } x_s = x_r \\ 1 & \text{if } x_s \neq x_r \end{cases} \tag{34}$$

where $\beta_c > 0$ is the spatial interaction parameter. The conditional probability $P_{X_s}(x_s|X_r = x_r, r \in \mathcal{N}_s)$ at site s is expressed explicitly with respect to clique potentials in the neighborhood $\mathcal{N}_s$ of s as follows:

$$P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c) = \frac{\exp\left(-\sum_{c\in\mathcal{C}/s\in c} V_c(x_s)\right)}{\sum_{x_s\in E_s}\exp\left(-\sum_{c\in\mathcal{C}/s\in c} V_c(x_s)\right)}. \tag{35}$$

We assume that $\forall s \in S$ the couples of random variables $(\mathbf{C}_s, \tau_s)$ are conditionally independent, given the label field X. Moreover, the conditional pdf of $(\mathbf{C}_s, \tau_s)$ given X depends only on the value of X at location s. Then, the conditional pdf of $(\mathbf{C}, \tau)$ given X can be written as

$$f_{\mathbf{C},\tau|X}(\mathbf{C},\tau|x,\boldsymbol{\theta}) = \prod_{s\in S} f_{\mathbf{C}_s,\tau_s|X}(\mathbf{C}_s,\tau_s|x,\boldsymbol{\theta}) = \prod_{s\in S} f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|x_s,\boldsymbol{\theta}). \tag{36}$$

The assumption (36) will be used in Section IV-B. We also assume that the random matrices $\mathbf{C}_s$ are conditionally independent, given the label field X, and in addition, it only depends on the value of X at location s

$$f_{\mathbf{C}|X}(\mathbf{C}|x,\boldsymbol{\theta}) = \prod_{s\in S} f_{\mathbf{C}_s|X}(\mathbf{C}_s|x,\boldsymbol{\theta}) = \prod_{s\in S} f_{\mathbf{C}_s|X_s}(\mathbf{C}_s|x_s,\boldsymbol{\theta}) \tag{37}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, \beta_c\}$ is a vector whose elements are the unknown parameters of the conditional pdf of $\mathbf{C}$ given X. A direct consequence of assumptions (36) and (37) is $f_{\tau|\mathbf{C},X}(\tau|\mathbf{C},x,\boldsymbol{\theta}) = \prod_{s\in S} f_{\tau_s|\mathbf{C}_s,X_s}(\tau_s|\mathbf{C}_s,x_s,\boldsymbol{\theta})$ justified as follows:

$$\begin{aligned} &f_{\mathbf{C},\tau|X}(\mathbf{C},\tau|x,\boldsymbol{\theta}) \\ &\quad= \prod_{s\in S} f_{\tau_s|\mathbf{C}_s,X_s}(\tau_s|\mathbf{C}_s,x_s,\boldsymbol{\theta}) \prod_{s\in S} f_{\mathbf{C}_s|X_s}(\mathbf{C}_s|x_s,\boldsymbol{\theta}) \\ &\quad= f_{\tau|\mathbf{C},X}(\tau|\mathbf{C},x,\boldsymbol{\theta}) f_{\mathbf{C}|X}(\mathbf{C}|x,\boldsymbol{\theta}). \end{aligned} \tag{38}$$

By using Bayes' rule and (37), the conditional probability mass function of X given $\mathbf{C}$ to segment the image is given by

$$\begin{aligned} P_{X|\mathbf{C}}(x|\mathbf{C},\boldsymbol{\theta}) &\propto f_{\mathbf{C}|X}(\mathbf{C}|x,\boldsymbol{\theta}) P_X(x) \\ &= \left(\prod_{s\in S} f_{\mathbf{C}_s|X_s}(\mathbf{C}_s|x_s,\boldsymbol{\theta})\right)\frac{1}{Z}\exp\left(-\sum_{\{s,r\}\in\mathcal{C}} V_c(x_s,x_r)\right) \\ &= \frac{1}{Z}\exp\left(\sum_{s\in S}\ln f_{\mathbf{C}_s|X_s}(\mathbf{C}_s|x_s,\boldsymbol{\theta}) - \sum_{\{s,r\}\in\mathcal{C}} V_c(x_s,x_r)\right). \end{aligned} \tag{39}$$

The conditional probability $P_{X|\mathbf{C}}(x|\mathbf{C},\boldsymbol{\theta})$ is a Gibbs distribution.

A. MPM Segmentation Algorithm

In [26], it was shown that the maximizer of the posterior marginal (MPM) estimate was more appropriate for image segmentation than the maximum *a posteriori* (MAP) estimate of the label field, given the observed image. The segmentation problem for the MPM algorithm is formulated as an optimization problem. It consists in minimizing the expected value of the number of misclassified pixels which is equivalent to maximizing the marginal posterior probability of the class label $P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta})$ over all $m \in \{1, 2, , \ldots M\}$ of each $s \in S$ as follows:

$$\hat{x}_{\text{MPM}} = \arg\max_{x_s} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}), \quad s \in S. \tag{40}$$

The main difficulty in calculating the MPM criterion is the computation of the marginal posterior probabilities. Marroquin *et al.* [26] presented an algorithm for approximating the MPM estimate by using a Markov chain Monte Carlo (MCMC) sampling process. It consists in using the Gibbs sampler [27] to generate a discrete-time Markov chain $X(t)$ that converges in distribution to a random field with a probability mass function given by (39). The marginal conditional probability mass functions that are to be maximized are then approximated as a fraction of time the Markov chain spends in state m at pixel s, for each m and s [26], [28]. Then, the approximation is given by

$$P_{X_s|\mathbf{C}}(m|\mathbf{C},\boldsymbol{\theta}) \approx \frac{1}{T_s}\sum_{t=1}^{T_s} u_{m,s}(t), \quad \forall m, s \tag{41}$$

with

$$u_{m,s}(t) = \begin{cases} 1, & \text{if } X_s(t) = m \\ 0, & \text{if } X_s(t) \neq m \end{cases} \tag{42}$$

where T_s is the number of visits to pixel s made by the Gibbs sampler.

B. Stochastic EM/MPM Algorithm for Parameter Estimation and Segmentation

The maximization of the conditional expectation of the complete log-likelihood according the parameters $\boldsymbol{\theta}$ by considering the assumption (36) is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} E_{X,\tau|\mathbf{C}}\{\ln f_{X,\tau,\mathbf{C}}(X,\mathbf{C},\tau|\boldsymbol{\theta})\Big|\mathbf{C},\boldsymbol{\theta}'\} \qquad (43)\\ &= \arg\max_{\boldsymbol{\theta}} E_{X,\tau|\mathbf{C}}\{\ln f_{\mathbf{C},\tau|X}(\mathbf{C},\tau|X,\boldsymbol{\theta}) + \ln P_X(X)|\mathbf{C},\boldsymbol{\theta}'\}\\ &= \arg\max_{\boldsymbol{\theta}} E_{X,\tau|\mathbf{C}}\left\{\sum_{s\in S}\ln f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|X_s,\boldsymbol{\theta})|\mathbf{C},\boldsymbol{\theta}'\right\}\\ &\quad + E_{X,\tau|\mathbf{C}}\{\ln P_X(X)|\mathbf{C},\boldsymbol{\theta}'\}. \qquad (44)\end{aligned}$$

Since the probability mass function $P_X(X)$ of the class label X depends only on β_c, (44) becomes as follows:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \sum_{s\in S} E_{X_s,\tau_s|\mathbf{C}}\left\{\ln f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|X_s,\boldsymbol{\theta})\Big|\mathbf{C},\boldsymbol{\theta}'\right\}\\ &= \arg\max_{\boldsymbol{\theta}} \sum_{s\in S}\sum_{x_s=1}^{M}\int_0^{+\infty}\ln f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|x_s,\boldsymbol{\theta})\\ &\quad\times f_{\tau_s,X_s|\mathbf{C}}(\tau_s,x_s|\mathbf{C},\boldsymbol{\theta}')d\tau_s \qquad (45)\\ &= \arg\max_{\boldsymbol{\theta}} \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\\ &\quad\times\int_0^{+\infty}\ln f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|x_s,\boldsymbol{\theta}) f_{\tau_s|X_s,\mathbf{C}}(\tau_s|x_s,\mathbf{C},\boldsymbol{\theta}')d\tau_s. \qquad (46)\end{aligned}$$

However, $f_{\tau_s|X_s,\mathbf{C}}(\tau_s|x_s,\mathbf{C},\boldsymbol{\theta}') = f_{\tau_s|X_s,\mathbf{C}_s}(\tau_s|x_s,\mathbf{C}_s,\boldsymbol{\theta}')$. Accordingly, we have

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\\ &\quad\times E_{\tau_s|X_s,\mathbf{C}_s}\{\ln f_{\mathbf{C}_s,\tau_s|X_s}(\mathbf{C}_s,\tau_s|X_s,\boldsymbol{\theta})|x_s,\mathbf{C}_s,\boldsymbol{\theta}'\} \qquad (47)\end{aligned}$$

Equation (47) can be maximized separately according to the speckle parameter $\hat{\boldsymbol{\theta}}_{\mathbf{X}}$ and the texture parameter $\hat{\theta}_\tau$ as follows:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathbf{X}} &= \arg\max_{\boldsymbol{\theta}_{\mathbf{X}}} \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\\ &\quad\times E_{\tau_s|X_s,\mathbf{C}_s}\{\ln f_{\mathbf{C}_s|\tau_s,X_s}(\mathbf{C}_s|\tau_s,X_s,\boldsymbol{\theta}_{\mathbf{X}})|x_s,\mathbf{C}_s,\boldsymbol{\theta}'\} \qquad (48)\\ \hat{\theta}_\tau &= \arg\max_{\theta_\tau} \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\\ &\quad\times E_{\tau_s|X_s,\mathbf{C}_s}\{\ln f_{\tau_s|X_s}(\tau_s|X_s,\theta_\tau)|x_s,\mathbf{C}_s,\boldsymbol{\theta}'\}. \qquad (49)\end{aligned}$$

As it can be seen, the difference between (16) and (17), and (48) and (49) is the expression of the posterior probability. In fact, the first is the marginal conditional probability of X_i belonging to the mth component, given $\mathbf{C}_i$, $P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')$, and the second is the marginal conditional probability of X_s, given all data set $\mathbf{C}$, $P_{X_s|\mathbf{C}}(m|\mathbf{C},\boldsymbol{\theta}')$. Using the same calculation as in Sections III-A and III-C, the equations resulting in the estimation of the speckle and texture parameters are the same with the difference that $P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i,\Theta')$ is replaced by $P_{X_s|\mathbf{C}}(m|\mathbf{C},\boldsymbol{\theta}')$.

C. MRF Parameter Estimation

The interaction parameter β_c can be estimated using the maximum-pseudolikelihood (MPL) [29]. It is the product of the conditional likelihood terms and is given as follows:

$$P_X(x) = \prod_{s\in S} P_{X_s}(x_s|x_{\mathcal{N}_s};\beta_c). \tag{50}$$

The MPL estimation method is easy to implement and retains good asymptotic properties in situations where the ML is more difficult to implement. From (44), the interaction parameter β_c is estimated as follows:

$$\begin{aligned}\hat{\beta}_c &= \arg\max_{\beta_c} E_{X,\tau|\mathbf{C}}\{\ln P_X(X)|\mathbf{C},\boldsymbol{\theta}'\} \qquad (51)\\ &= \arg\max_{\beta_c} \sum_{s\in S} E_{X_s,\tau_s|\mathbf{C}}\{\ln P_{X_s}(X_s|X_{\mathcal{N}_s},\beta_c)|\mathbf{C},\boldsymbol{\theta}'\} \qquad (52)\\ &= \arg\max_{\beta_c} \sum_{s\in S}\sum_{x_s=1}^{M}\int_0^{+\infty}\ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c)\\ &\quad\times f_{X_s,\tau_s|\mathbf{C}}(x_s,\tau_s|\mathbf{C},\boldsymbol{\theta}')\mathrm{d}\tau_s \qquad (53)\end{aligned}$$

Algorithm 1

1: **Input**: $N, \mathbf{C}_i, d, \epsilon, M, \beta_c, T_s$
2: **Output**: $\hat{\boldsymbol{\theta}} = (\hat{L}, \hat{\boldsymbol{\Sigma}}_1, \ldots, \hat{\boldsymbol{\Sigma}}_M, \hat{\theta}_{\tau,1}, \ldots, \hat{\theta}_{\tau,M}, \hat{\beta}_c)$
3: **Initialization**:
4: Set initial parameters $\boldsymbol{\theta'} = (\boldsymbol{\theta'_X}, \theta'_\tau, \beta'_c)$ by the initialization procedure given in the next section.
5: **Main Loop**:
6: **Repeat**
7: Use the Gibbs sampler to generate the Markov chain $X(t)$ and sample from the distribution (39) using the estimate of $\boldsymbol{\theta}$ obtained in the previous iteration
8: Approximate the class label probabilities $P_{X_s|\mathbf{C}}(x_s|\mathbf{C}, \boldsymbol{\theta})$ using (41)
9: For cluster m calculate $E_{\tau_s|\mathbf{C}_s,X_s}\{.|\mathbf{C}_s, X_s = m, \boldsymbol{\theta'}\}$ according to the chosen compound pdf
10: Estimate $\hat{\boldsymbol{\Sigma}}_m$ using (18).[1] For $\mathcal{G}_d$ use $\hat{\boldsymbol{\Sigma}}_m \leftarrow d\frac{\hat{\boldsymbol{\Sigma}}_m}{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_m)}$
11: Estimate $\hat{L}$ using $\hat{\boldsymbol{\Sigma}}_m$ and (19)[1]
12: Estimate $\hat{\theta}_{\tau,m}$ using (26)[1] for $\mathcal{K}_d$, or (28)[1] for $\mathcal{G}_d^0$, or (29, 30, 31)[1] for $\mathcal{G}_d$
13: Estimate $\hat{\beta}_c$ using (57)
14: Calculate stop criterion: $D \leftarrow \|\hat{\theta}_\tau - \theta'_\tau\| + \|\hat{\boldsymbol{\theta}}_{\boldsymbol{X}} - \boldsymbol{\theta'_X}\|$
15: Set inputs for next iteration: $L', \boldsymbol{\Sigma}'_m, \theta'_{\tau,m}, \beta'_c \leftarrow \hat{L}, \hat{\boldsymbol{\Sigma}}_m, \hat{\theta}_{\tau,m}, \hat{\beta}_c$
16: **Until** $D < \epsilon$
17: **Return** $(\hat{L}, \hat{\boldsymbol{\Sigma}}_1, \ldots, \hat{\boldsymbol{\Sigma}}_M, \hat{\theta}_{\tau,1}, \ldots, \hat{\theta}_{\tau,M}, \hat{\beta}_c)$
18: **Final segmentation**: The marginal conditional probability mass functions is maximized according to (40)

$$= \arg\max_{\beta_c} \sum_{s\in S} \sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C}, \boldsymbol{\theta}') \int_0^{+\infty} \ln P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c) \times f_{\tau_s|X_s,\mathbf{C}}(\tau_s|x_s, \mathbf{C}, \boldsymbol{\theta}')\mathrm{d}\tau_s \quad (54)$$

$$= \arg\max_{\beta_c} \sum_{s\in S} \sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C}, \boldsymbol{\theta}') \ln P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c). \quad (55)$$

Finding a solution typically requires taking the derivatives of the previous expression (55) with respect to β_c as follows:

$$\sum_{s\in S} \sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C}, \boldsymbol{\theta}') \frac{\partial}{\partial \beta_c} \ln P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c) = 0. \quad (56)$$

Substituting the derivative of $\ln P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c)$ in the previous expression, we have the following equation:

$$\sum_{s\in S} \sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C}, \boldsymbol{\theta}') \sum_{r\in\mathcal{N}_s} t(x_s, x_r) = \sum_{s\in S} \sum_{x_s=1}^{M} P_{X_s}(x_s|x_{\mathcal{N}_s}, \beta_c) \sum_{r\in\mathcal{N}_s} t(x_s, x_r). \quad (57)$$

This equation is solved numerically to find the solution $\hat{\beta}_c$. For more details of derivation, see Appendix A.

[1]Replace $P_{X_i|\mathbf{C}_i}(m|\mathbf{C}_i, \Theta')$ by $P_{X_s|\mathbf{C}}(m|\mathbf{C}, \boldsymbol{\theta}')$

D. Implementation of EM/MPM Technique

Equations (18), (19), (26), and (57); (18), (19), (28), and (57); and (18), (19), (29), (30), (31), and (57) represent the iterative EM algorithm for the estimation of the matrix-variate $\mathcal{K}_d$, $\mathcal{G}_d^0$ and $\mathcal{G}_d$ distribution parameters and the interaction parameter of the MRF model, respectively. The algorithm is iterative and can be initialized with any suitable values of $(\boldsymbol{\theta'_X}, \theta'_\tau, \beta'_c)$. The algorithm is stopped when the convergence criteria is satisfied. The $(\partial/\partial\alpha)K_\alpha(.)$ function is approximated using the following relation:

$$\frac{\partial}{\partial\alpha} K_\alpha(x) \approx \frac{K_{\alpha+h}(x) - K_{\alpha-h}(x)}{2h} \quad (58)$$

with $h = 10^{-4}$. The unsupervised segmentation and the parameter estimation approach are summarized in Algorithm 1. In our case, the Frobenius norm is used in step 14.

E. Number of Classes

One of the necessary parameters leading to unsupervised image segmentation is the number of classes used for clustering. Usually, this number is assumed to be known, but recently, many methods have been dedicated to automatically determine the appropriate number of classes. The most used method is the goodness-of-fit (GoF) testing performed using the matrix log cumulants [10], [13], [14], [30]. Subsequently, the GoF test is incorporated within the EM algorithm and allows to split or merge the classes. When classes do not fit well, they are split into two clusters and the EM is reapplied again. The appropriate number of classes is obtained when all classes are considered good-fits to the data histograms. In this paper, the number of classes is determined by the analyst and then assumed to be known.

F. Initialization Procedure

The initialization procedure is an important step in the unsupervised segmentation process. With an adequate initialization, we accelerate the estimation of hidden class labels and associated parameters, and, thus, the convergence of the algorithm. The idea is to estimate the local values of ENL, the local estimates of the texture parameters, and the local estimates of the covariance matrix of the speckle by using a sliding window of size $k \times k$ covering the whole image. The goal is to make a repartition map of the values of these parameters. A simple clustering algorithm will use these data to generate a first clustering where the estimated parameters of each class will be utilized as an initialization of the unsupervised classification. The EM-based estimator [18] is used to estimate the ENL, the texture parameters, and the covariance matrix of the speckle, assuming the existence of only one class inside the sliding window, which means that $M = 1$ and $P_{X_i|\mathbf{C}_i}(x_i|\mathbf{C}_i, \Theta') = 1$. Then, the initialization procedure takes place as follows.

1) Compute the local estimation of the covariance matrix of the speckle $\hat{\boldsymbol{\Sigma}}$ and its determinant $|\hat{\boldsymbol{\Sigma}}|$ by EM-based estimator using a sliding window of size $k \times k$ where $k \in \{7, 9, 11, 13, 15\}$, covering the whole image.

The estimated covariance matrix of the speckle is given by

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} g_i^{(1)}\mathbf{C}_i. \tag{59}$$

2) In a similar manner, compute the local ENL estimates using the EM-based estimator. The equation to use is as follows:

$$d\ln\hat{L} + d - \psi_d(\hat{L}) = \frac{1}{N}\sum_{i=1}^{N} g_i^{(1)} tr(\hat{\mathbf{\Sigma}}^{-1}\mathbf{C}_i) - \ln|\hat{\mathbf{\Sigma}}^{-1}\mathbf{C}_i| + da_i^{(1)}. \tag{60}$$

A kernel density estimator (KDE) implemented with the Epanechnikov kernel function and a kernel bandwidth of $h' = 0.1$ is used to estimate the distribution of the EM estimator of the ENL. Since the number of correlated looks L is replaced with an equivalent number of uncorrelated looks ENL, the mode value of the distribution of ENL estimates is used as an estimate of the global ENL.

3) Compute the local estimation of the texture parameters using the EM estimator utilizing a sliding window of size $k \times k$. For example, if a gamma distribution is used as a texture model, the equation to solve is as follows:

$$\ln(\hat{\alpha}) - \psi(\hat{\alpha}) + 1 = \frac{1}{N}\sum_{i=1}^{N} h_i^{(1)} - a_i^{(1)}. \tag{61}$$

4) Any clustering algorithm, like Fuzzy C-means (FCM) or K-means, is applied to a vector image consisting of the local estimation of the texture parameters and the local determinant of the estimated covariance matrix of the speckle.

V. Experiment, Results, and Discussion

In this section, we try to find out the performance of the proposed method on both simulated and real SAR data. Classification with simulated data permits evaluating the accuracy of the method, whereas the real data allow a visual comparison.

A. Simulated Multilook PolSAR Images

1) First Example: The simulated image consists of an image of 200 × 200 pixels with six distinct classes of $\mathcal{K}_d$ distribution generated with ten looks and in full PolSAR. The areas represent homogeneous, heterogeneous, and extremely heterogeneous texture. To test the robustness of the segmentation, only a few parameters change between two neighboring regions. Then, the parameters of the $\mathcal{K}_d$ pdf are different in some regions and similar in others. Fig. 4(a) and (b) shows the true classes and the corresponding simulated data with six classes and ten-look covariance, respectively. The elements of the covariance matrix of the speckle and the texture parameters are given in Tables II and III for each area, respectively. Regions 1 and 2 have the same covariance matrix of the speckle, and regions 2 and 3 have the same texture parameter α. Moreover, region 1 represents extremely heterogeneous texture characteristics of an urban area. Regions 2, 3, and 4 represent a forest area with a moderate texture. Regions 5 and 6 represent low heterogeneous regions and homogeneous regions, respectively. In addition, region 6 has a texture parameter α that tends to infinity leading to a realization of a scaled Wishart distribution. Fig. 1(a) shows the pdf when the texture is gamma distributed with different values of the texture parameter α.

TABLE II
Covariance Matrices of Speckle per Area

Area	Σ_{11}	Σ_{22}	Σ_{33}	Σ_{12}	Σ_{13}	Σ_{23}
1	0.8	1	0.5	0.3j	0.2j	0.1
2	0.8	1	0.5	0.3j	0.2j	0.1
3	3	1	0.5	0.3j	0.2j	0.1
4	2	1	0.5	0.3j	0.5j	0.1
5	3.5	0.8	0.42	0.5+0.3j	0.2j	0.1-0.3j
6	1	1	0.5	0.2-0.3j	0.1+0.5j	0.1-0.01j

TABLE III
Texture Parameters per Area

Texture		Area					
Symbol	Para.	1	2	3	4	5	6
$\mathcal{K}_d$	α	1.5	3	3	7	12	$+\infty$
	$\hat{\alpha}$	1.52	2.96	2.97	7.05	12.21	7577.5
$\mathcal{G}_d$	α	1	2	2	4	6	$+\infty$
	w	1	3	3	1	2	$+\infty$
	η	1	4	4	3	2	1
	$\hat{\alpha}$	0.96	1.22	1.50	3.79	6.28	161.10
	$\hat{w}$	0.97	3.22	3.23	1.37	1.61	46.23
	$\hat{\eta}$	1.00	4.96	4.67	4.26	1.57	0.14

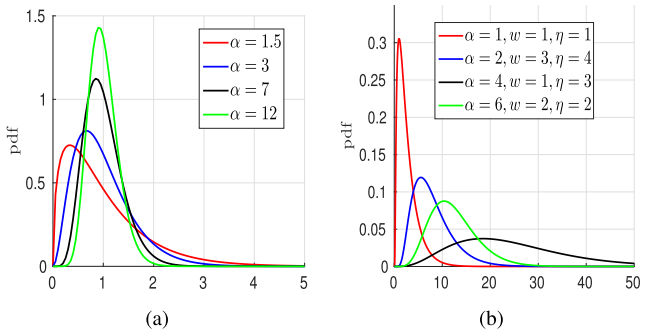

Fig. 1. (a) Gamma and (b) GIG texture pdfs used for simulation. Lower values of α represent extreme heterogeneous areas and higher values represent homogeneous areas.

Our aim is to show how accurately the algorithm performs the parameter estimation. The results of the initialization are presented in Figs. 2 and 3. Fig. 2(a) depicts the local ENL estimation images computed for the EM-based estimator using a sliding window of size $k = 9$ covering the whole image. Fig. 2(b) shows the distribution of the local estimates. The distributions are computed with KDE with an Epanechnikov kernel as it is mentioned before in Section IV-F. The distribution is centered around the true number of looks $L = 10$ since the random generated data are not correlated. The mode of the distribution estimates for the EM estimator is considered as an estimate of the global ENL. Fig. 2(c) illustrates the results of the local estimation image of the texture parameters using the EM-based estimator. The matrix-variate $\mathcal{K}_d$ distribution is used to model the multilook polarimetric covariance matrix. For a

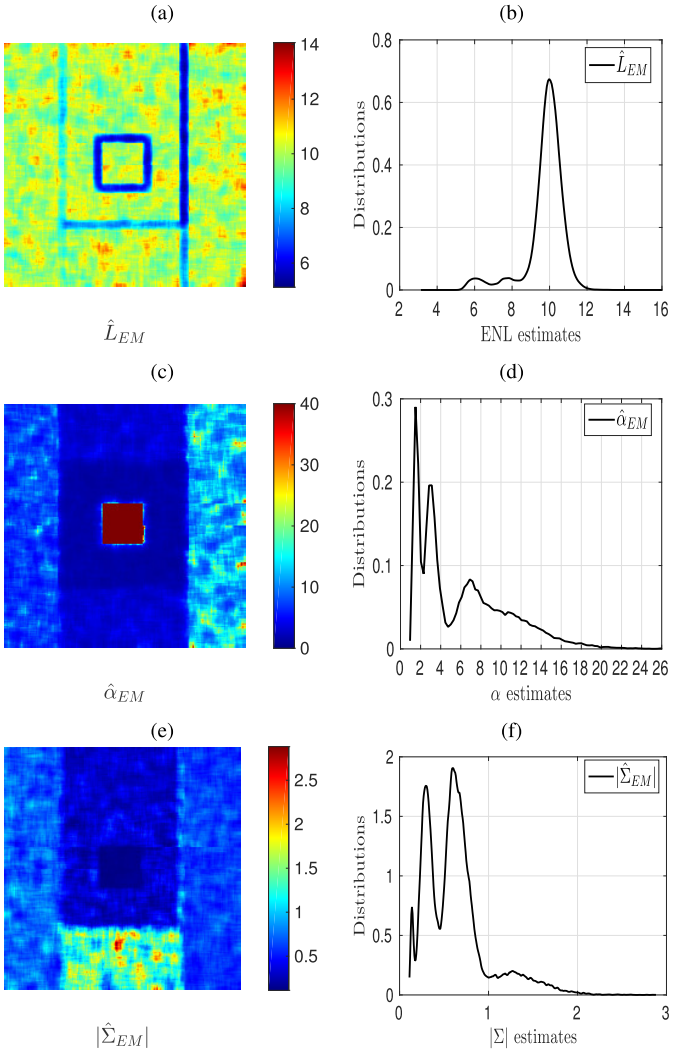

Fig. 2. Parameter estimates as part of initialization procedure. (a) Local ENL estimates calculated by using EM estimator. (b) Distribution estimates calculated when slide window is $k = 9$. (c) Local $\hat{\alpha}_{EM}$ estimates for EM estimator. (d) Distribution of $\hat{\alpha}$ for latter estimator. (e) Local $|\hat{\mathbf{\Sigma}}|$ estimates for EM estimator. (f) Distribution of $|\hat{\mathbf{\Sigma}}|$ for latter estimator.

better display, we have limited the scale between 0 and 40 for $\hat{\alpha}_{EM}$. As a consequence, the homogeneous region 6 is marked with dark red color. Fig. 2(d) shows the distribution of the estimates for the previous estimator. Fig. 2(e) highlights the local determinant of the covariance matrix of the speckle $|\hat{\mathbf{\Sigma}}|$ estimated on the same sliding window. Fig. 2(f) depicts the distribution of $|\hat{\mathbf{\Sigma}}|$ estimates. As we can see, several modes can be detected indicating the existence of several regions. The final initialization step is shown in Fig. 3, where the result of a clustering using FCM applied to the local estimation of the texture parameters is shown in Fig. 3(a). Five different classes can clearly be recognized corresponding to the different values of α. When FCM is only applied to the local determinant of the estimated covariance matrix of the speckle, the clustering is illustrated in Fig. 3(b). Fig. 3(c) corresponds to the result of clustering applied to a vector image composed of both the local estimation of the texture parameters and the local determinant of the estimated covariance matrix of the speckle. Six classes can be clearly seen. Fig. 3(c) is used as initialization to our

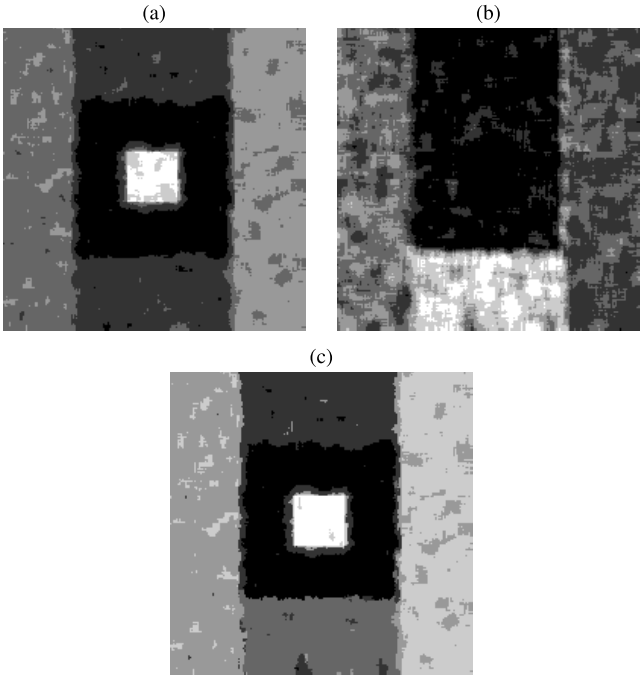

Fig. 3. (a) Clustering results using FCM applied to local estimation of texture parameters. (b) Local determinant of estimated covariance matrix of speckle. (c) Vector image composed of local estimation of texture parameters and local determinant of estimated covariance matrix of speckle.

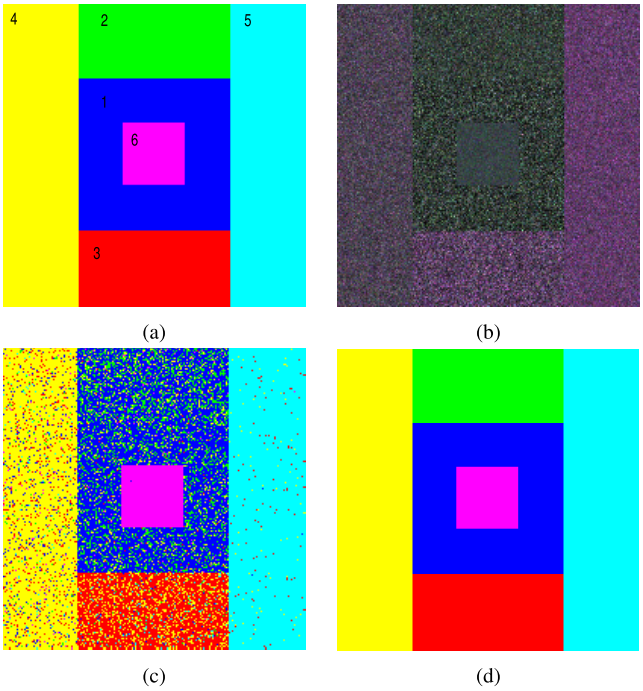

Fig. 4. Simulated data classification, six classes, ten-look covariance. (a) True classes. (b) Colored composition of simulated data in Pauli basis. (c) Unsupervised segmentation using matrix variate $\mathcal{K}_d$. (d) Unsupervised segmentation using matrix-variate $\mathcal{K}_d$-MRF.

unsupervised segmentation algorithm. The initial parameters given by step 4 of Algorithm 1 are estimated on each class of the image initialization.

The main loop of Algorithm 1 is undergoing (step 5). The result of the segmentation using the compound distribution

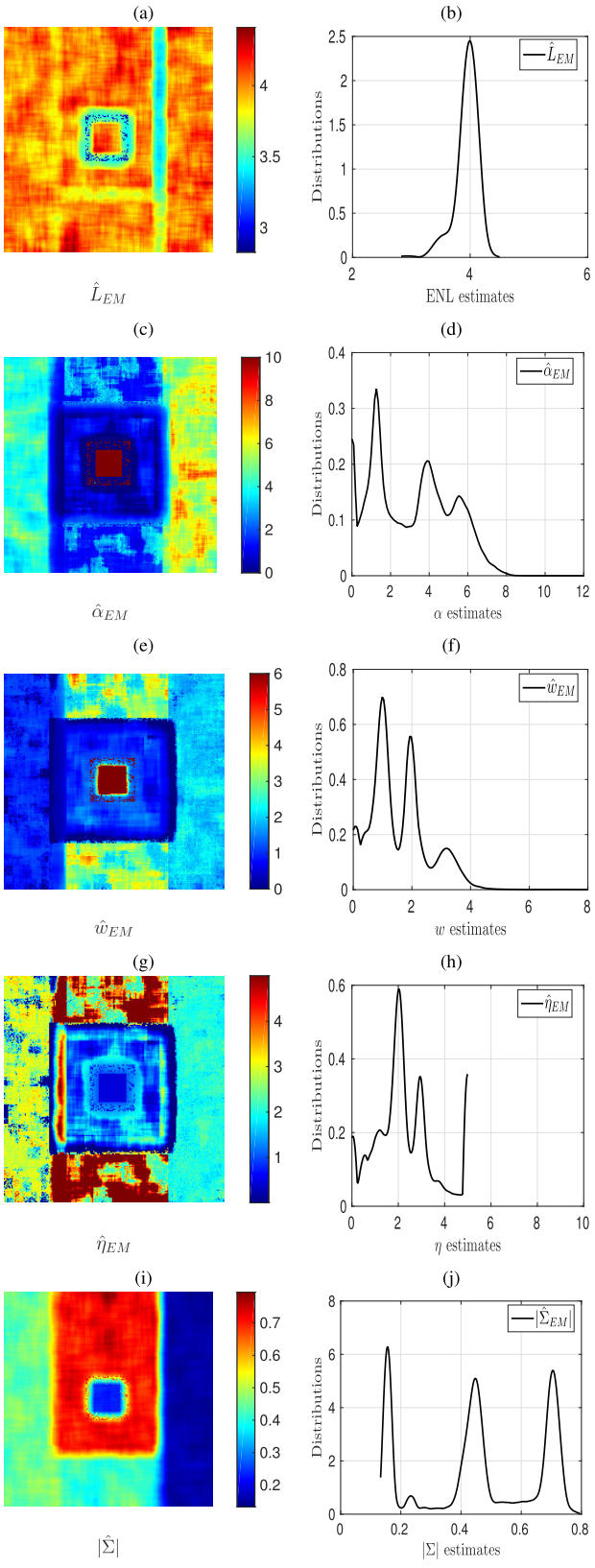

Fig. 5. Local $\hat{\alpha}, \hat{w}, \hat{\eta}$ estimates by EM estimator. (a) and (b) $\hat{\alpha}_{EM}$ and their distributions. (c) and (d) $\hat{w}_{EM}$ and their distributions. (e) and (f) $\hat{\eta}_{EM}$ and their distributions. The slide window is $k = 17$.

without an MRF is shown in Fig. 4(c) and with an MRF in Fig. 4(d). To quantify the quality of classification, the confusion matrix is used. Table IV shows the accuracy of both

TABLE IV
CONFUSION MATRIX

Area	1	2	3	4	5	6
$\mathcal{K}_d$	73.3%	24.5%	73%	86%	98%	99.9%
$\mathcal{K}_d$-MRF	100%	100%	100%	100%	100%	100%
$\mathcal{G}_d$	74%	67.5%	68.9%	70.1%	81.8%	99.2%
$\mathcal{G}_d$-MRF	100%	100%	100%	100%	100%	100%

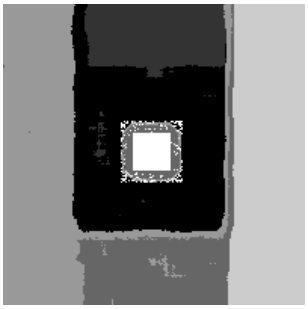

Fig. 6. Clustering results using FCM applied to vector image composed of local estimation of texture parameters (α, ω) and local determinant $|\mathbf{\Sigma}|$ of estimated covariance matrix of speckle.

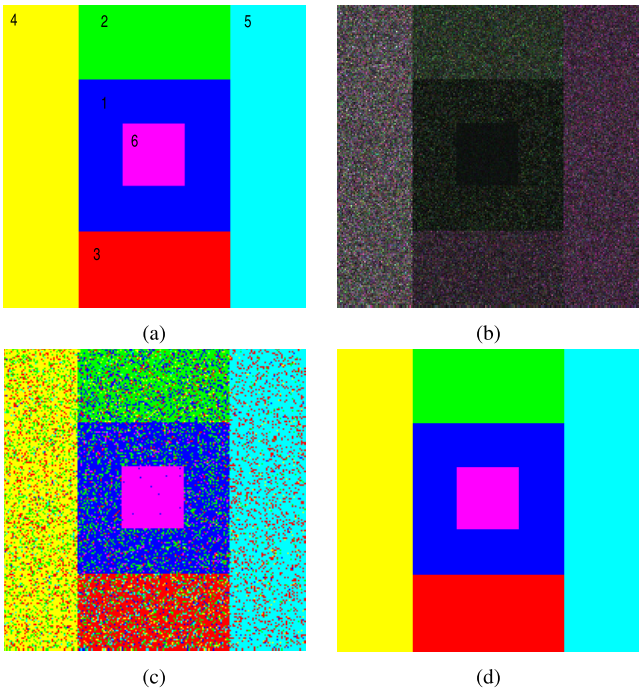

Fig. 7. Simulated data classification, six classes, four-look covariance. (a) True classes. (b) Colored composition of simulated data in Pauli basis. (c) Final unsupervised segmentation using matrix variate $\mathcal{G}_d$. (d) Final unsupervised segmentation using matrix-variate $\mathcal{G}_d$-MRF.

classifiers. A value of 100% indicates a perfect classification and underneath 100% indicates a nearly perfect classification performance. With the final classification by $\mathcal{K}_d$-MRF, the estimation of the texture parameters $\hat{\alpha}$ are given in Table III. As we can conclude, both classifiers already provide good classification performances, with accuracies higher than 80% for classes 4, 5, and 6 characterized by low texture. Low performances are observed for the highly non-Gaussian classes (class 1 and 3) that have classification accuracies of 73.3%

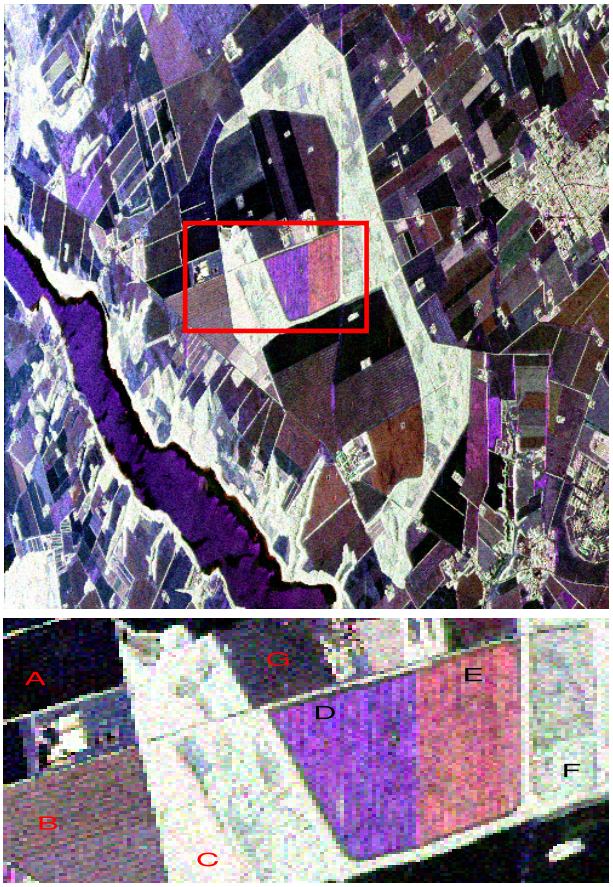

Fig. 8. (Top) EMISAR L-band quad-pol acquisition over Foulum, Denmark, multilooked with six looks shown as Pauli RGB image. (Bottom) Zoomed-in view of selected test region surrounded in main image by red rectangle.

and 73% for $\mathcal{K}_d$ distribution. The lowest performance is registered for class 2 with classification accuracy of 24.5% for the $\mathcal{K}_d$. Since class 2 and class 1 have the same covariance matrix of the speckle and very close texture parameter values, $\mathcal{K}_d$ has difficulty in distinguishing between classes 2 and 1. The classification with $\mathcal{K}_d$-MRF corrects this weakness and improves the results, yielding a perfect segmentation.

2) Second Example: The simulated quad-pol data contains four-look PolSAR images of 200 × 200 pixels and have three polarimetric channels. The simulation contains six areas where five of them contain independent realizations of the matrix variate $\mathcal{G}_d$ distribution, and the last one is a realization of the scaled Wishart distribution which is a special case of $\mathcal{G}_d$ when the two shape parameters α and w tend toward infinity. Fig. 7(a) shows the true classes chosen the same as example 1. The corresponding simulated data with six classes and four-look covariance are illustrated in Fig. 7(b). The covariance matrix of the speckle and the texture parameters are given, respectively, in Tables II and III for each region. The same values of the covariance matrix used in the first example are utilized here. It is recalled that the covariance matrix of the speckle is normalized, meaning that the covariance matrices verify the normalization condition $\text{tr}(\mathbf{\Sigma}) = d = 3$. Regions 2 and 3 have the same texture parameter α and w values. Moreover, region 1 represents extremely heterogeneous texture characteristics of an urban area. Regions 2, 3, and 4 represent a forest area with a moderate texture. Regions 5 and 6 represent homogeneous regions and have the lowest texture. Fig. 1(b) shows the pdf of the GIG distribution as a texture model with different values of shape parameters α and w.

Initialization results are shown in Figs. 5 and 6. Fig. 5(a) depicts the local ENL estimation images computed for the EM estimator, using a sliding window of size $k = 17$ covering the whole image. Fig. 5(b) depicts the distribution estimates. They produce distributions that are centered around the true number of looks $L = 4$. Fig. 5(c), (e), and (g) shows, respectively, the results of the local estimations image of the texture parameters (α, w) and scale parameter (η) using the EM estimator utilizing a sliding window of size $k = 17$. We recall that the matrix-variate $\mathcal{G}_d$ distribution is used in this example as a compound distribution for the product model. The estimated α and w values for region 6 are significantly higher than the other regions. This is due to the lack of texture. For a reliable display, we have limited the scale between 0 and 10 for α [Fig. 5(c)] and between 0 and 6 for w [Fig. 5(e)]. Then, for region 6, $\hat{\alpha}$ and $\hat{w}$ values greater than 10 and 6, respectively, are, therefore, marked by the dark red color. We can easily see the different regions corresponding to the different $\hat{\alpha}$, $\hat{w}$, and $\hat{\eta}$ values. Fig. 5(d), (f), and (h) shows the distributions of these estimates. Fig. 5(i) illustrates the local determinant of the covariance matrix of the speckle $|\hat{\mathbf{\Sigma}}|$ estimated on a sliding window of size $k = 17$. Fig. 5(j) depicts the distribution of $|\hat{\mathbf{\Sigma}}|$ estimates. As we can see, several modes can be detected. The result of clustering applied to a vector image composed of both the local estimation of the texture parameters $(\hat{\alpha},\hat{w})$ and the local determinant $(|\hat{\mathbf{\Sigma}}|)$ of the estimated covariance matrix of the speckle is shown in Fig. 6. There are six clearly seen classes. Fig. 6 is used as an initialization to our unsupervised segmentation algorithm. The initial parameters given by step 4 of Algorithm 1 are estimated on each class of the image initialization.

After applying the Algorithm 1 from step 5 to step 18, the result of the unsupervised segmentation using the compound distribution without an MRF is shown in Fig. 7(c) and with an MRF shown in Fig. 7(d). This example clearly demonstrates that the $\mathcal{G}_d$-MRF clustering gives good segmentation results. With the final classification, the estimation of the texture parameters are given in Table III for the matrix-variate $\mathcal{G}_d$ distribution. The results of classification show that both classifiers $\mathcal{G}_d$ and $\mathcal{G}_d$-MRF already provide good classification performances, with accuracies higher than 80% for classes 5 and 6, which are homogeneous regions. Low performances are observed for the highly non-Gaussian classes (class 1, 2, 3, and 4). The lowest performance is registered for class 2 and class 3 with classification accuracies of 67.5% and 68.9%, respectively, for the $\mathcal{G}_d$ classifier. Despite the fact that these two classes have the same texture parameter values, classifier $\mathcal{G}_d$ can distinguish between them in contrast to example 1. The classification with $\mathcal{G}_d$-MRF improves the results drastically, yielding a perfect segmentation.

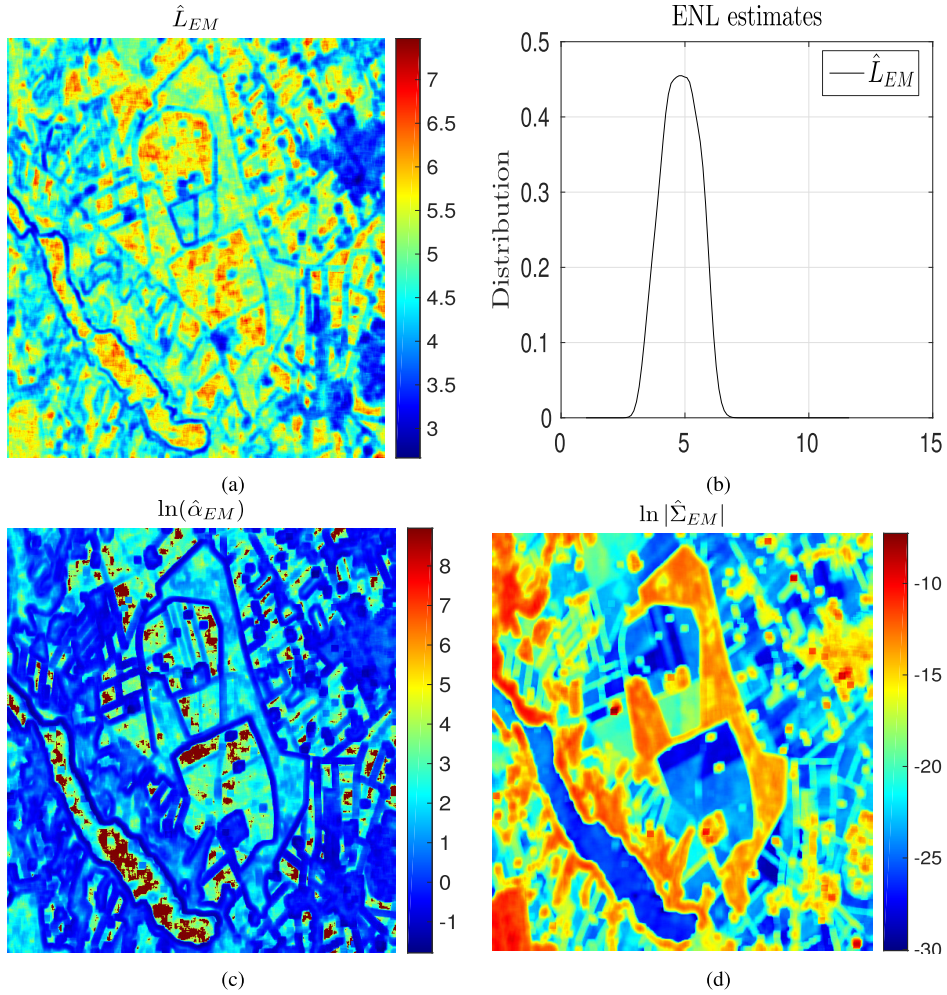

Fig. 9. Local estimation of speckle and texture parameters of the whole Foulum image. (a) Local ENL estimation. (b) Distribution estimates of ENL. (c) Local texture parameter $\ln\hat{\alpha}$ estimate. (d) Local log determinant of the covariance matrix of speckle $\ln|\hat{\mathbf{\Sigma}}|$.

B. Real SAR Images

The unsupervised segmentation is demonstrated on real Pol-SAR data. The test image is a small section of the image of an agricultural area from an airborne Electromagnetics Institute Synthetic Aperture Radar (EMISAR) L-band quad-pol SAR acquisition over Foulum, Denmark, in 1998. A Pauli RGB composite image is shown in Fig. 8(top) and the selected test region is shown in Fig. 8(bottom). The data are multilooked to six looks.

The local estimation of the covariance matrix of the speckle and texture parameters for the whole image are illustrated in Fig. 9. The local ENL estimation image computed for the EM-based estimator using a sliding window of size $k = 9$ covering the whole image is depicted in Fig. 9(a). The distribution estimates are shown in Fig. 9(b). As we can see, the estimated ENL is 4.8. Fig. 9(c) presents the local texture parameter $\ln\hat{\alpha}$ estimate using the EM-based estimator. The logarithmic scale is used to reduce the dynamics of the image. The extremely heterogeneous texture regions (low α value) are marked by dark blue color and the homogeneous regions are marked by dark red color. Fig. 9(d) shows the local log determinant of the covariance matrix of the speckle $\ln|\hat{\mathbf{\Sigma}}|$ estimated on a sliding window of size $k = 9$. The same behavior is observed but, this time, in the opposite direction. Thus, the heterogeneous texture regions are marked by red color and the homogeneous regions are marked by blue color.

The number of classes for the selected test image is fixed to 14. As illustrated in Fig. 10(d), the $\mathcal{K}_d$-MRF classifier splits fields A and G into two classes (red and magenta), whereas the $\mathcal{K}_d$ classifier [Fig. 10(c)] groups them as a single class (red). This happens also for field C (forest area), where the $\mathcal{K}_d$-MRF classifier splits field C into four classes: class 12 (dark green), class 13 (dark magenta), class 14 (dark red), and class 4 (yellow), whereas the $\mathcal{K}_d$ classifier groups them into two classes: class 12 (dark green) and class 9 in low proportion. For forest areas (fields C and F), the $\mathcal{K}_d$-MRF classifier can distinguish significant differences in forest density correctly.

Despite the low number of looks and the noisy aspect of the image, $\mathcal{K}_d$ and $\mathcal{K}_d$-MRF classifiers are able to generate spatially homogeneous segmentation results. In fields B, D, and E, it is observed that the $\mathcal{K}_d$ classifier already does a good job in smoothing the fields into quite homogeneous areas, as shown in Fig. 10(c). The fields are further smoothed by the $\mathcal{K}_d$-MRF classifier but without removing small distinct targets. The conclusion of this experiment is that the combination

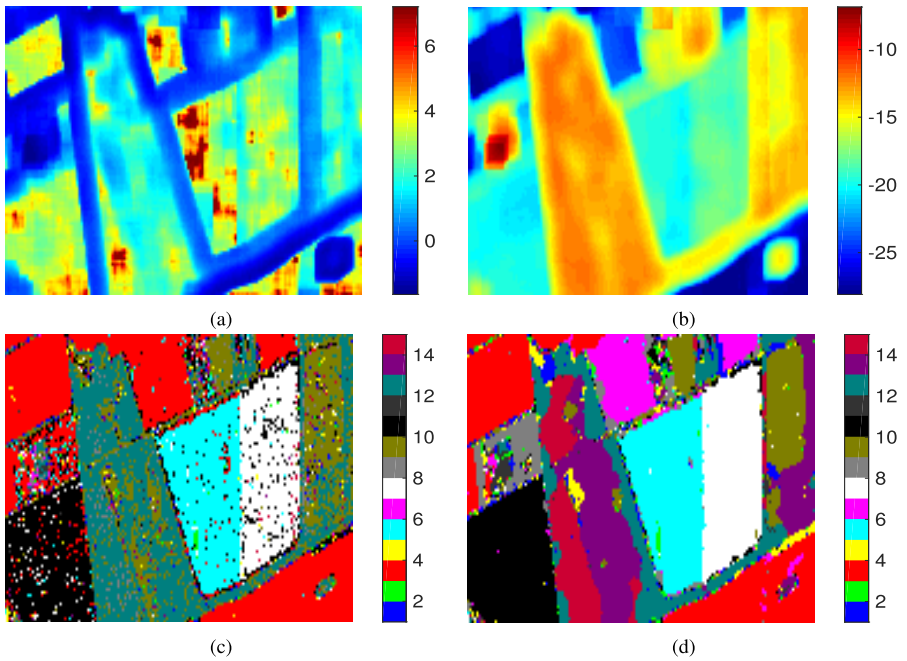

Fig. 10. Unsupervised segmentation of selected test region. (a) Local estimation of texture parameters ($\ln \hat{\alpha}$) on selected test region. (b) Local estimation of speckle ($\ln |\hat{\mathbf{\Sigma}}|$) on selected test region. Fourteen class segmentation results with (c) $\mathcal{K}_d$ and (d) $\mathcal{K}_d$-MRF classifiers.

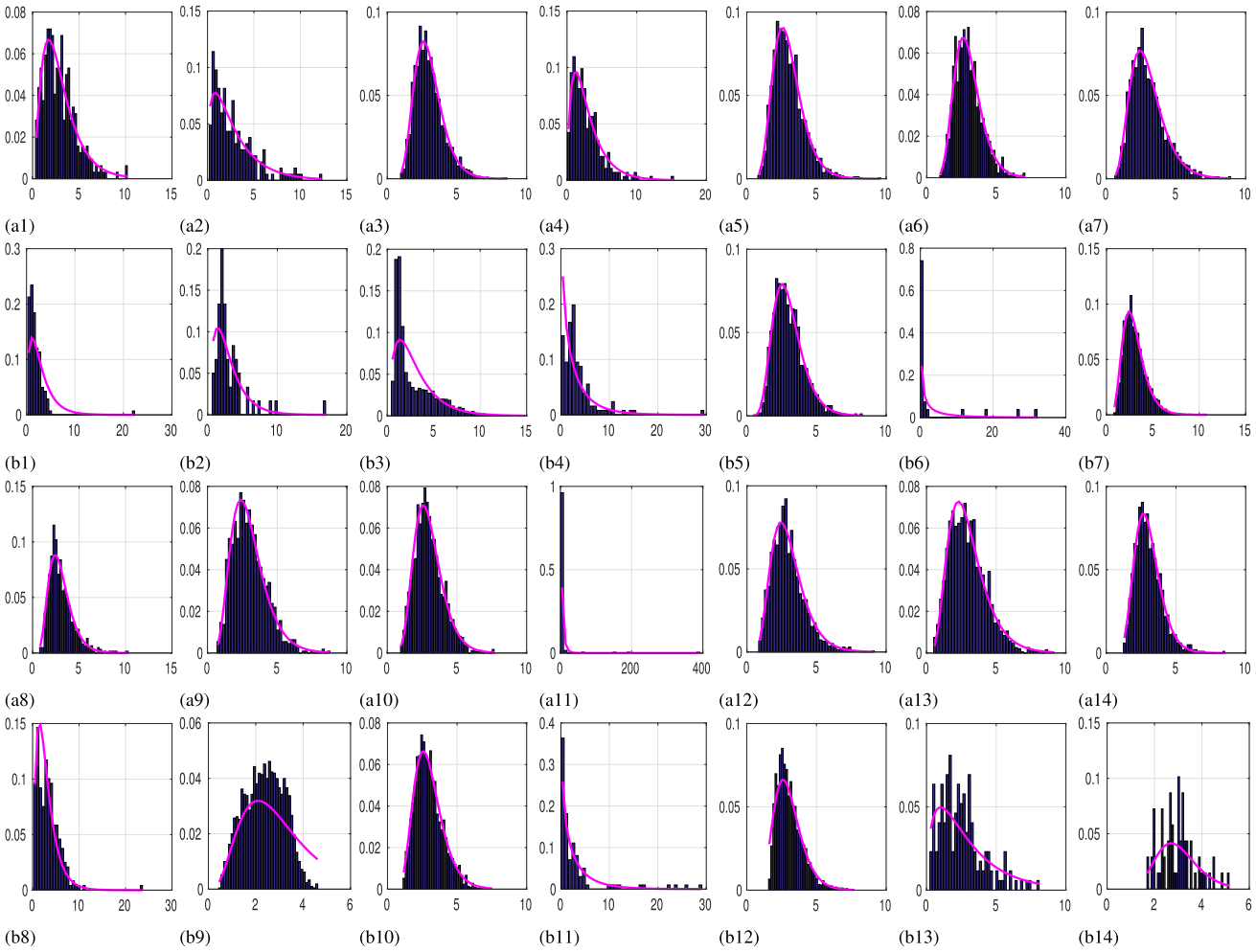

Fig. 11. Comparison between estimated pdf of $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})$ and class histograms of each cluster. (ai) and (bi) show respectively, for class 'i' the fitting of the estimated entity $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})$ pdf produced by the $\mathcal{K}_d$-MRF and $\mathcal{K}_d$ classifier to the class histograms of real data. The figure shows a good visual fit of the $\mathcal{K}_d$-MRF classifier to all class histograms.

TABLE V
PDF OF $Y = \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})$ FOR DIFFERENT TEXTURE DISTRIBUTIONS

τ	$f_Y(y)$
$\bar{\gamma}(\alpha)$	$\frac{2(L\alpha)^{\frac{\alpha+dL}{2}}}{\Gamma(\alpha)\Gamma(dL)} y^{\frac{\alpha+dL}{2}-1} K_{\alpha-dL}(2\sqrt{L\alpha y})$
$\bar{\gamma}^{-1}(\lambda)$	$\frac{(\lambda-1)^{\lambda} L^{dL}}{\Gamma(\lambda)} \frac{\Gamma(dL+\lambda)}{\Gamma(dL)} y^{dL-1}(Ly+\lambda-1)^{-(dL+\lambda)}$
$\mathcal{N}^{-1}(\alpha, w, \eta)$	$\frac{L^{dL} y^{dL-1}}{\eta^{\alpha}\Gamma(dL)} \frac{1}{K_{\alpha}(w)} \left(\frac{2Ly+w\eta}{w/\eta}\right)^{\frac{\alpha-dL}{2}} \times K_{\alpha-dL}\left(\sqrt{\frac{w}{\eta}(2Ly+w\eta)}\right)$

of compound $\mathcal{K}_d$ distribution with an MRF model improves the segmentation results. The $\mathcal{K}_d$-MRF classifier produces the most reliable clustering, as shown in Fig. 10(d).

To fit the histogram of each class, the multilook polarimetric covariance matrix is transformed into scalar by the variable $Y = \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})$. The statistical model for $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) = \tau \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{X})$ is derived using the distribution of the texture τ and the distribution of $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{X})$ [18]. The pdf of the scalar variable Y is given in Table V for each texture pdf. Fig. 11 shows the fitting of the estimated entity $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})$ pdf produced by the $\mathcal{K}_d$ and $\mathcal{K}_d$-MRF classifiers to the class histograms of real data. Fig. 11(ai) and (bi) illustrates the fitting for class i produced by the $\mathcal{K}_d$-MRF and $\mathcal{K}_d$ classifier, respectively. It is clear to see that the $\mathcal{K}_d$-MRF model fits well the class histograms.

VI. CONCLUSION

In this paper, a new unsupervised segmentation of multilook PolSAR images has been presented. The method combines non-Gaussian compound distribution as an observed model for the PolSAR data statistics conditioned to each class with spatial information as an MRF used as prior model. Three distributions, $\mathcal{K}_d$, $\mathcal{G}_d^0$, and $\mathcal{G}_d$, have been used in this paper as compound distributions. The ML method has been proposed to estimate the parameters of these distributions for each class. Parameter estimation by ML is a very difficult task due to the complexity of the likelihood function. Thus, the EM algorithm has been developed to derive the ML estimates for the product model. The MRF has been adopted in this paper to integrate spatial information and then to improve the accuracy and quality of segmentation. The class posterior probabilities have been estimated using the MPM algorithm. In summary, the unsupervised segmentation method is based on the combination of EM and MPM where all parameters are estimated by EM and the class labels are updated by MPM. Three examples have been presented and tested with the method. The segmentation results are perfect. A comparison with a segmentation without MRF leads to improvement brought by the MRF.

APPENDIX A

Finding a solution typically requires taking the derivatives of the previous expression (55) with respect to β_c as follows:

$$\sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}') \frac{\partial}{\partial \beta_c} \ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c) = 0. \quad (62)$$

According to (35), $\ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c)$ is given

$$\ln P_{X_s}(x_s|X_{\mathcal{N}_s},\beta_c) = -\beta_c \sum_{r\in\mathcal{N}_s} t(x_s,x_r) - \ln \sum_{x_s=1}^{M} \exp\left(-\beta_c \sum_{r\in\mathcal{N}_s} t(x_s,x_r)\right). \quad (63)$$

The derivative of $\ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c)$ with respect to β_c is given as follows:

$$\frac{\partial}{\partial \beta_c} \ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c) \quad (64)$$

$$= -\sum_{r\in\mathcal{N}_s} t(x_s,x_r) - \frac{\frac{\partial}{\partial\beta_c}\sum_{x_s=1}^{M}\exp\left(-\beta_c\sum_{r\in\mathcal{N}_s}t(x_s,x_r)\right)}{\sum_{x_s=1}^{M}\exp\left(-\beta_c\sum_{r\in\mathcal{N}_s}t(x_s,x_r)\right)} \quad (65)$$

$$= -\sum_{r\in\mathcal{N}_s} t(x_s,x_r) + \frac{\sum_{x_s=1}^{M}\sum_{r\in\mathcal{N}_s}t(x_s,x_r)\exp\left(-\beta_c\sum_{r\in\mathcal{N}_s}t(x_s,x_r)\right)}{\sum_{x_s=1}^{M}\exp\left(-\beta_c\sum_{r\in\mathcal{N}_s}t(x_s,x_r)\right)}$$

$$= -\sum_{r\in\mathcal{N}_s} t(x_s,x_r) + \sum_{x_s=1}^{M} P_{X_s}(x_s|x_r, r\in\mathcal{N}_s)\sum_{r\in\mathcal{N}_s} t(x_s,x_r). \quad (66)$$

Substituting the derivative of $\ln P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c)$ in (62), we have the following equation:

$$\sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\sum_{r\in\mathcal{N}_s} t(x_s,x_r) = \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}') \times \left(\sum_{x_s=1}^{M} P_{X_s}(x_s|x_{\mathcal{N}_s},\beta_c)\sum_{r\in\mathcal{N}_s} t(x_s,x_r)\right). \quad (67)$$

Knowing that $\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}') = 1$, then the expression becomes as follows:

$$\sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s|\mathbf{C}}(x_s|\mathbf{C},\boldsymbol{\theta}')\sum_{r\in\mathcal{N}_s} t(x_s,x_r) = \sum_{s\in S}\sum_{x_s=1}^{M} P_{X_s}(x_s|x_r, r\in\mathcal{N}_s,\beta_c)\sum_{r\in\mathcal{N}_s} t(x_s,x_r). \quad (68)$$

APPENDIX B
GIG DISTRIBUTION

The pdf of the GIG distribution is given by

$$f(z;\alpha_1,w_1,\eta_1) = \frac{z^{\alpha_1-1}}{\eta_1^{\alpha_1} 2K_{\alpha_1}(w_1)} \exp -\frac{w_1}{2}\left(\frac{\eta_1}{z}+\frac{z}{\eta_1}\right). \quad (69)$$

The following moments exist and are finite as follows:

$$E\{Z^k\} = \eta_1^k \frac{K_{\alpha_1+k}(w_1)}{K_{\alpha_1}(w_1)}, \quad k\in\mathbb{Z} \quad (70)$$

$$E\{\ln Z\} = \ln\eta_1 + \frac{\frac{\partial}{\partial a}K_{\alpha_1+a}(w_1)|_{a=0}}{K_{\alpha_1}(w_1)}. \quad (71)$$

APPENDIX C
INVERSE GAMMA DISTRIBUTION

The pdf of the inverse gamma distribution is given by

$$f(z;\alpha_1,\beta_1)=\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{1}{z^{1+\alpha_1}}\exp\left(\frac{-\beta_1}{z}\right). \tag{72}$$

The following moments exist and are finite:

$$E\{Z\}=\frac{\beta_1}{\alpha_1-1}\ \text{for}\ \alpha_1>1,\ \ E\{Z^{-1}\}=\frac{\alpha_1}{\beta_1} \tag{73}$$

$$E\{\ln Z\}=-\psi(\alpha_1)+\ln\beta_1. \tag{74}$$

REFERENCES

[1] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.

[2] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Appl.*, vol. 3, no. 8, pp. 747–768, Jan. 1989.

[3] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution: An introduction," *Ann. Math. Statist.*, vol. 34, no. 81, pp. 152–177, Mar. 1963.

[4] G. Gao, "Statistical modeling of SAR images: A survey," *Sensors*, vol. 10, no. 1, pp. 775–795, Jan. 2010.

[5] J. S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, "K-distribution for multi-look processed polarimetric SAR imagery," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4, Aug. 1994, pp. 2179–2181.

[6] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric $\mathcal{G}$ distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.

[7] L. Bombrun and J. M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.

[8] L. Bombrun, S. N. Anfinsen, and O. Harant, "A complete coverage of log-cumulant space in terms of distributions for polarimetric SAR data," in *Proc. 5th Int. Workshop Sci. Appl. SAR Polarimetry Polarimetric Interferometry*, Frascati, Italy, Jan. 2011, pp. 1–8.

[9] S. Khan and R. Guida, "Application of Mellin-kind statistics to polarimetric $\mathcal{G}$ distribution for SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3513–3528, Jun. 2014.

[10] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3665–3676, Oct. 2011.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, no. 1, pp. 1–22, Sep. 1977.

[12] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.

[13] V. Akbari, A. P. Doulgeris, G. Moser, T. Eltoft, S. N. Anfinsen, and S. B. Serpico, "A textural–contextual model for unsupervised segmentation of multipolarization synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2442–2453, Apr. 2013.

[14] A. P. Doulgeris, "An automatic U-distribution and Markov random field segmentation algorithm for PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1819–1827, Apr. 2015.

[15] L. Bombrun, G. Vasile, M. Gay, and F. Totir, "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 726–737, Feb. 2011.

[16] S. Khan and A. P. Doulgeris, "Unsupervised clustering of PolSAR data using polarimetric G distribution and Markov random fields," in *Proc. 10th Eur. Conf. Synth. Aperture Radar*, Jun. 2014, pp. 1–4.

[17] S. Khan and R. Guida, "On fractional moments of multilook polarimetric whitening filter for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3502–3512, Jun. 2014.

[18] N. Bouhlel and S. Méric, "Maximum-likelihood parameter estimation of the product model for multilook polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1596–1611, Mar. 2019.

[19] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2009.

[20] F. Gini and M. S. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[21] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[22] Y. X. Yuan, "A review of trust region algorithms for optimization," in *Proc. 4th Int. Congr. Ind. Appl. Math. (ICIAM)*, Edinburgh, U.K., vol. 99, 2000, no. 1, pp. 271–282.

[23] Z. Kato and J. Zerubia, "Markov random fields in image segmentation," *Found. Trends Signal Process.*, vol. 5, nos. 1–2, pp. 1–155, Oct. 2012.

[24] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer-Verlag, 2009.

[25] C. Gaetan and X. Guyon, *Spatial Statistics and Modeling* (Springer Series in Statistics). New York, NY, USA: Springer-Verlag, 2010.

[26] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, no. 397, pp. 76–89, Mar. 1987.

[27] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[28] M. L. Comer and E. J. Delp, "The EM/MPM algorithm for segmentation of textured images: Analysis and further experimental results," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1731–1744, Oct. 2000.

[29] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. B*, vol. 36, no. 2, pp. 192–225, Jan. 1974.

[30] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Goodness-of-fit tests for multilook polarimetric radar data based on the Mellin transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2764–2781, Jul. 2011.

Nizar Bouhlel (M'13) received the B.E. degree in electrical engineering and the master's degree in communication systems from the National Engineering School of Tunis (ENIT), Tunis, Tunisia, in 1997 and 1999, respectively, and the Ph.D. degree in applied mathematics from the Applied Mathematics at Paris 5 (MAP5), University of Paris Descartes, Paris, France, in 2006.

From 2007 to 2011, he was a Researcher with the Communication Department, Faculty of Sciences of Monastir, Monastir, Tunisia. From 2012 to 2016, he was a Post-Doctoral Researcher in several French laboratories: the Biomedical Imaging Laboratory (LIB), University of Pierre-et-Marie-Curie (UPMC), Paris, France, and the Laboratory of Digital Sciences of Nantes (LS2N), School of Centrale Nantes (ECN), Nantes, France. He is currently a Senior Researcher with the with the SHINE Team, Institute of Electronics and Telecommunications of Rennes (IETR), National Institute for the Applied Sciences (INSA) Rennes, France. His research interests include stochastic image processing, radar signal processing, Bayesian inference, parameter estimation, Markov random fields, and classification.

Stéphane Méric (M'08) received the Diploma degree in electrical engineering from the National Institute for the Applied Sciences (INSA), Rennes, France, in 1991, the M.S. degree in signal processing and telecommunications from the University of Rennes 1, Rennes, in 1991, and the Ph.D. degree in electronics from INSA in 1996.

Since 2000, he has been an Assistant Professor with INSA. In 2005, he joined the SAPHIR Team, Institute of Electronics and Telecommunications of Rennes (IETR)-CNRS UMR 6164, Rennes. He is currently involved in radar system [continuous wave (CW), frequency-modulated continuous-wave (FMCW)] dedicated to specific SAR applications [radar imaging in motorway context, remote sensing, multi-in multi-out (MIMO) configuration, and passive radar imaging]. His education activities are about analog electronics, signal processing, radar and radar imaging, and electromagnetic diffraction. He has been supervising five Ph.D. students and currently three Ph.D. students. He has been designed as a Reviewer for the National Research Programs and was with the Educational Team, Teaching Department, Engineering School, Oujda, Morrocco. He has been supervising three Ph.D. students and currently three Ph.D. students. He has authored or co-authored 40 conference papers, 11 journal papers, and 2 book chapters. He holds one patent. His research interests include SAR data in radargrammetric applications.