



Multichannel hierarchical image classification using multivariate copulas

Aurélie Voisin^{a#}, Vladimir A. Krylov^a, Gabriele Moser^b, Sebastiano B. Serpico^b
and Josiane Zerubia^a

^a EPI Ariana, CR INRIA Sophia Antipolis Méditerranée, 2004, Route des Lucioles, B.P.93,
F-06902 Sophia Antipolis (France);

^b Dept. of Biophysical and Electronic Engineering (DIBE), University of Genoa, Via Opera
Pia 11a, I-16145, Genoa (Italy);

ABSTRACT

This paper focuses on the classification of multichannel images. The proposed supervised Bayesian classification method applied to histological (medical) optical images and to remote sensing (optical and synthetic aperture radar) imagery consists of two steps. The first step introduces the joint statistical modeling of the coregistered input images. For each class and each input channel, the class-conditional marginal probability density functions are estimated by finite mixtures of well-chosen parametric families. For optical imagery, the normal distribution is a well-known model. For radar imagery, we have selected generalized gamma, log-normal, Nakagami and Weibull distributions. Next, the multivariate d -dimensional Clayton copula, where d can be interpreted as the number of input channels, is applied to estimate multivariate joint class-conditional statistics. As a second step, we plug the estimated joint probability density functions into a hierarchical Markovian model based on a quad-tree structure. Multiscale features are extracted by discrete wavelet transforms, or by using input multiresolution data. To obtain the classification map, we integrate an exact estimator of the marginal posterior mode.

Keywords: supervised classification, multichannel data, multivariate copulas, hierarchical Markov random fields, mixture models, marginal posterior mode (MPM), wavelet transform, synthetic aperture radar (SAR), histological images.

1. INTRODUCTION

A wide variety of images are available nowadays, and the high diversity of camera/sensor properties explains the difficulty of developing general-purpose image processing algorithms. In this paper we develop a general classifier that can be used on both monoband and multiband imagery as well as at different resolutions, with the underlying assumption that the acquired data are coregistered¹. Initially developed to deal with Synthetic Aperture Radar² (SAR) image classification, the proposed model is sufficiently flexible to deal with other types of data, hence giving the method an additional degree of freedom by allowing the use of multisensor data.

A common way to classify multisensor data is to combine multiple classifiers, by various methods such as boosting³, bagging⁴, majority voting⁵, decision fusion^{6,7}, support vector machines (SVM)⁸ or methods based on the Dempster-Shafer theory of evidence⁹. In a Bayesian context¹⁰, posterior probabilities are used for classification via statistical consensus theory¹¹ or neural networks¹². In a same Bayesian context, it is also possible to model the likelihood term by a joint probability¹³, given the marginal probabilities related to each input data. The classification is then obtained by applying standard classification methods, such as, for instance, Markov random field (MRF)-based methods¹⁴. The classification proposed in this paper is based on a Markovian context, and the likelihood term is estimated by combining the marginal probability density functions (PDFs) related to each input data by using multivariate copulas¹⁵. Moreover, the use of a hierarchical Markovian model^{16,17} allows to consider multiresolution acquisitions, without resorting to pixel-based fusion¹⁸. Such classifiers can be used in different application fields such as medical or biological image processing, as well

E-mail: aurelie.voisin@inria.fr

as in remote sensing, where the classification may help to determine land-use or land-cover maps or damaged areas after a natural disaster.

The proposed supervised Bayesian classification method consists of two steps. The first one is the statistical modeling of the coregistered input images. Such statistical modeling is flexible enough to allow multisource data to be dealt with. For each class and each channel in the chosen stacked-vector input dataset, the class-conditional marginal probability density functions (PDFs) are estimated by finite mixtures of well-chosen parametric families¹⁹. For optical imagery, the normal distribution is an often accepted model, and the parameters of the employed normal PDF mixtures are estimated in the proposed method by the stochastic expectation maximization (SEM)²⁰ algorithm. For SAR imagery, the generalized Gamma distribution²¹ is an adequate choice, both thanks to its accurate results with this data typology and because it extends earlier developed classical parametric SAR models. Nevertheless, log-normal, Nakagami and Weibull distributions may be other feasible options²². The mixture parameters in the case of a SAR channel are determined by a modified SEM algorithm that integrates the method of log-cumulants (MoLC)²³ instead of maximum likelihood estimates that are unfeasible for many SAR-specific parametric families. Next, the multivariate d -dimensional Clayton copula¹⁵ (d being the number of input channels) is applied to estimate multivariate joint class-conditional statistics, merging the marginal PDF estimates of the input channels. The copula parameters are estimated by using the relationship with the Kendall's tau correlation coefficient¹⁵. As a second step, we plug the estimated joint PDFs into a contextual model that uses a multiscale approach via a hierarchical Markovian model^{16,17} based on a quad-tree structure. Multiscale features are extracted by discrete wavelet transforms^{24,25} or by using input multiresolution data. The consideration of a quad-tree allows to integrate an exact estimator of the marginal posterior mode (MPM)²⁶ that aims to estimate the unknown class labels. The prior probability is iteratively updated at each level of the tree, leading to an algorithm more robust with respect to noise²⁷ (e.g., speckle in SAR acquisitions²) when compared to a non-updated prior²⁶.

For single-resolution data, it is also possible to use a spatial context via a hidden Markov random field (MRF) model²⁸ that employs a modified Metropolis dynamics (MMD) scheme²⁹ or graph-cuts³⁰⁻³² for energy minimization. Comparisons between the two contextual methods (spatial and hierarchical) will be studied in Sec. 5.

The rest of the paper is organized as follows. In Sec. 2, we introduce the statistical multivariate copula-based model that combines marginal PDF models of the input images. In Sec. 3, we briefly recall the main concepts and properties of the hierarchical approaches. In Sec. 4, we develop the MPM model based on the quad-tree that we use to classify our data. In Sec. 5, we present classification results obtained on histological (medical) and on remote sensing imagery.

2. JOINT PDF MODEL

Considering the problem of supervised classification, we need to model the joint statistics of the input images. These inputs may be, for instance, color bands of optical images and/or single or multi-polarized radar images. To proceed, we propose, first, to model independently the statistics of each input band for each class, and then to estimate a joint PDF for each class by using the statistical instrument of copulas.

2.1 Marginal PDF estimation

We want to model the distributions of each class ω_m considered for the classification, $m \in [1; M]$, given a training set, for each input image. For each class, the PDF $p_m(z|\omega_m)$ is modeled via finite mixtures³³ of independent greylevel distributions:

$$p_m(z|\omega_m) = \sum_{i=1}^K P_{mi} p_{mi}(z|\theta_{mi}), \quad (1)$$

where z is a greylevel, $z \in [0; Z - 1]$, and ω_m is the m^{th} class. P_{mi} are the mixing proportions such that for a given m , $\sum_{i=1}^K P_{mi} = 1$ with $0 \leq P_{mi} \leq 1$. θ_{mi} is the set of parameters of the i^{th} PDF mixture component of the m^{th} class. The use of finite mixtures instead of single PDFs offers the possibility to consider heterogeneous

PDFs, usually reflecting the contributions of the components present in each class (for instance, different kinds of crops for the vegetation class when considering high resolution remote sensing data). Moreover, the use of finite mixtures can be seen as a generalization of the determination of a single PDF, and allows to estimate both the best finite mixture model and/or the best single PDF model.

2.1.1 Optical case

When the input is an optical acquisition, we consider that the PDF $p_m(z|\omega_m)$ related to each class can be modeled by a finite mixture of Gaussian distributions, hence

$$p_{mi}(z|\theta_{mi}) = \frac{1}{\sqrt{2\pi\sigma_{mi}^2}} \exp\left[-\frac{(z - \mu_{mi})^2}{2\sigma_{mi}^2}\right], \quad \text{with } \theta_{mi} = \{\mu_{mi}, \sigma_{mi}^2\}, \quad (2)$$

where the mean μ_{mi} and the variance σ_{mi}^2 are estimated within a stochastic expectation maximization²⁰ algorithm.

2.1.2 Radar case

The radar acquisitions are known to be affected by speckle². For this reason, we use distributions more adapted to such images such as the generalized Gamma distribution³⁴. Each class conditional PDF is then modeled by a mixture of generalized Gamma distributions, hence

$$p_{mi}(z|\theta_{mi}) = \frac{\nu_{mi}}{\sigma_{mi}\Gamma(\kappa_{mi})} \left(\frac{z}{\sigma_{mi}}\right)^{\kappa_{mi}\nu_{mi}-1} \exp\left\{-\left(\frac{z}{\sigma_{mi}}\right)^{\nu_{mi}}\right\}, \quad \text{with } \theta_{mi} = \{\nu_{mi}, \sigma_{mi}, \kappa_{mi}\}, \quad (3)$$

where $\Gamma(\cdot)$ is the Gamma function³⁵.

When the generalized Gamma distribution is not applicable³⁶, the PDF $p_{mi}(z|\theta_{mi})$ is automatically chosen among the following distributions: log-normal, Weibull, Nakagami. These distributions are commonly used to model radar imagery³⁷. A modified SEM algorithm is then used to estimate the best-fitting mixture model for each considered class. It combines a density parameter estimation via the method of log-cumulants²³ and a stochastic expectation maximization (SEM) algorithm²⁰. For more details concerning the mixture estimation, see Refs. 19, 22 and 37.

2.2 Combination of marginal distributions via multivariate copulas

We now very briefly introduce some relevant properties and definitions for copulas. For a comprehensive introduction see Ref. 15.

The multivariate copula is a d -dimensional joint distribution defined on $[0, 1]^d$ such that marginal distributions are uniform on $[0, 1]$. The importance of copulas in statistics is explained by Sklar's theorem¹⁵, which states the existence of a copula C_m that models the joint distribution function H_m of arbitrary random variables $\{Z_1, \dots, Z_d\}$ with cumulative distribution functions (CDFs) $\{F_{1m}, \dots, F_{dm}\}$:

$$H_m(z_1, \dots, z_d) = C_m(F_{1m}(z_1), \dots, F_{dm}(z_d)), \quad (4)$$

for all z_1, \dots, z_d in \mathbb{R} .

Taking the derivative in Eq. (4) over the d continuous random variables $\{z_1, \dots, z_d\}$ with PDFs $\{f_{1m}, \dots, f_{dm}\}$, we obtain the joint PDF distribution:

$$h_m(z_1, \dots, z_d) = \prod_{j=1}^d f_{jm}(z_j) \times \frac{\partial^d C_m}{\partial z_1 \dots \partial z_d}(F_{1m}(z_1), \dots, F_{dm}(z_d)) = \prod_{j=1}^d f_{jm}(z_j) \times c_m(F_{1m}(z_1), \dots, F_{dm}(z_d)). \quad (5)$$

We wish to estimate, for each class m , the joint PDF h_m given the marginal distributions $\{f_{1m}, \dots, f_{dm}\}$ that correspond to the marginal PDFs of the different input images, i.e. $f_{jm} = p_m(z|\omega_m)$ estimated for the j^{th} input image and for the m^{th} class. The CDF F_{jm} is the integral on $]-\infty; z_j]$ of its corresponding PDF f_{jm} . Thus, the parameters of the cumulative distribution functions $\{F_{1m}, \dots, F_{dm}\}$ are the same as the parameters of the

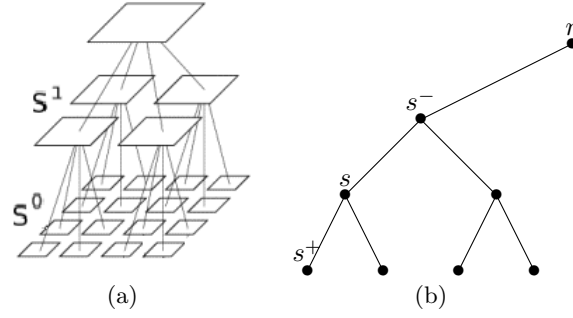


Figure 1. (a): Hierarchical model structure: quad-tree; (b): Quad-tree notations.

marginal estimates in Eq. (1). According to Eq. (5), to determine h_m , we have to determine the copula family C_m .

The bivariate copulas have been studied extensively in the literature, which is not necessarily the case for multivariate copulas. We only focus in this paper on the Clayton copula family, due to the fact that the densities of e.g., Frank or Marchal copulas, well-known in a bivariate case, are not straightforward for a d -dimensional case. The Clayton copula density can be written:

$$c_m(u_1, \dots, u_d) = \prod_{j=1}^d u_j^{-(\alpha+1)} \cdot \prod_{n=0}^{d-1} (1 + \alpha n) \cdot \left(\sum_{j=1}^d u_j^{-\alpha} - d + 1 \right)^{\left(\frac{-1-d\alpha}{\alpha} \right)} \quad \text{where } u_j = F_{jm}(z_j). \quad (6)$$

To estimate the unknown copula parameter α , we use the relationship between copulas and Kendall's τ which is a ranking correlation coefficient¹⁵. By definition, Kendall's τ is a concordance-discordance measure that can be empirically estimated by using the training sets. Once the estimate $\hat{\tau}$ is computed, we get the parameter estimate $\hat{\alpha} = \frac{2\hat{\tau}}{1-\hat{\tau}}$, according to Refs. 15,38. Such procedure is stressed for each considered class.

3. THE HIERARCHICAL MODEL

Among the various techniques that can be applied on hierarchical graphs¹⁶ (renormalization group,³⁹ constrained configuration subspaces,⁴⁰ etc.), we employ an explicit hierarchical graph-based model^{16,26} to address our classification problem. The specific graphs on which we base our study have a tree structure. The set of sites s ($s \in S$) is, therefore, hierarchically partitioned as $S = S^0 \cup S^1 \cup \dots \cup S^R$, where R corresponds to the coarsest resolution, the root, and 0 corresponds to the reference level (finest resolution). In this tree structure, there is a parent-child relationship: for each site s of any tree-level n a unique parent s^- and several children s^+ can be defined. $d(s)$ refers to the set including s and its descendants. For the specific case where s owns four children, the tree structure is called a quad-tree. Such a structure is depicted in Fig. 1.

Various good properties have led us to consider this type of hierarchical Markov random fields. First among them is the scale causality generated by the quad-tree structure, which allows a non-iterative algorithm to be used, thus implying a computational time decrease⁴¹. Second, the multigrid properties reduce the probability to find a local minimum, hence the applied algorithm is more likely to converge to a global solution. Moreover, such a model is able to take into account different kinds of statistics, and thus to use different kinds of images (different resolutions, different sensors, etc.)⁴².

The aim of classification is to estimate a set of hidden labels X given a set of observations Y attached to the sites. X and Y are considered to be random processes. The restriction of X (resp. Y) to the level n is $X^n = \{X_s, s \in S^n\}$ (resp. $Y^n = \{Y_s, s \in S^n\}$) where the realization x^n takes its values in Ω . Some extra hypotheses are needed to ensure that X is a Markov random field on the graph: $\forall s \in S, \forall x \in \Omega$,

- (i) $p(X = x) > 0$,

(ii) $p(X_s = x_s | X_t = x_t, t \in S - \{s\}) = p(X_s = x_s | X_t = x_t, t \in V_s)$, V_s describing the neighborhood of the site s .

One last hypothesis is the pointwise dependance of Y with respect to X , thus implying that each couple (X, Y) is Markovian on the quad-tree:

$$p(y|x) = \prod_{n=0}^R p(y^n|x^n) = \prod_{n=0}^R \prod_{s \in S^n} p(y_s|x_s).$$

4. HIERARCHICAL CLASSIFICATION APPROACH

A variety of algorithms were proposed to estimate the labels on hierarchical graphs¹⁷. Typically, a global energy minimization is done via iterative relaxation algorithms⁴³. The consideration of a quad-tree allows to benefit from its good properties (e.g., causality) and to apply non iterative algorithms. To avoid the underflow generated by the use of the maximum a posteriori (MAP) criterion, we take into account an exact estimator of the marginal posterior mode (MPM)^{26,44}. The cost function associated to this estimator offers the possibility to penalize the errors according to their number and the scale at which they occur: an error at the coarsest scale is more strongly penalized than an error at the finest scale, which is a desired property because a site located at the root corresponds to 4^R pixels at the finest scale.

To estimate the posterior probability, we need the following prior information: the likelihood, the prior probability and the transition probability at each site s of the quad-tree. In this section, we present how this information is estimated and then we focus on the maximization of the posterior probability. The employed likelihood model has already been developed in Sec. 2.

4.1 Transition probabilities

The fundamental hypothesis in the application of the described hierarchical MRF is that we consider the random process X Markovian with respect to scale, i.e. $p(x^n|x^k, k > n) = p(x^n|x^{n+1}) = \prod_{s \in S^n} p(x_s|x_{s-})$, where n and k are scales. These transition probabilities between the scales, $p(x_s|x_{s-})$, determine the hierarchical MRF since they represent the causality of the statistical interactions between the different levels of the tree. We use the transition probability in the form introduced by Bouman et al.⁴⁵: for all sites $s \in S$ and all scales $n \in [0; R - 1]$,

$$p(x_s = \omega_m | x_{s-} = \omega_k) = \begin{cases} \theta_n, & \text{if } \omega_m = \omega_k \\ \frac{1-\theta_n}{M-1}, & \text{otherwise} \end{cases}, \quad (7)$$

where $m, k \in [1; M]$. This model favors an identical parent-child labeling. Typically, we choose a fixed $\theta_n \approx 0.8$, which means that a site s at scale n has a probability of about 80% to belong to the same class as its ascendant s^- . These transition probabilities are used to estimate the prior probabilities at different scales (see Sec. 4.2).

4.2 Prior probability

The prior distribution at level n in $[0; R - 1]$ is given by:

$$p(x_s^n) = \sum_{x_{s-}^n} p(x_s^n | x_{s-}^n) p(x_{s-}^n). \quad (8)$$

Thus, the prior information at the coarsest level R allows to determine the prior information at the other levels since the transition probabilities are known (see Sec. 4.1).

The first step is, thus, to determine the prior information at the coarsest level R . The priors at other scales are estimated by using Eq. (8). We choose to model by considering the equiprobability between classes. We then apply an MPM estimation on a R -scale tree, and use the classification results as an updated prior. Then, we consider a smaller tree of scale $R - 1$ to which we apply the MPM algorithm to estimate a new prior. We proceed iteratively until scale 0 is reached. Such an update allows a better prior estimation, leading to a final classification map more accurate than without any prior update^{26,27}. This update is illustrated in Fig. 3.

To estimate the priors given a classification map, we use a Markovian model which takes into account the contextual information at each level, and therefore leads to a better prior estimation. By employing the Hammersley-Clifford theorem⁴⁶, we can define a local characteristic for each site:

$$p(x_s) = \frac{1}{Z} \exp(-\beta \sum_{s:\{s,t\} \in C} \delta_{x_s=x_t}) \quad \text{with } \delta_{x_s=x_t} = \begin{cases} 1, & \text{if } x_s = x_t \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

where Z is the normalization constant, s, t denote the sites in the same clique and x_s, x_t their labels. A clique is a non-empty subset c of neighboring sites of size equal to or higher than 1.

Instead of considering a widely used second-order neighborhood based on the 8 pixels surrounding a given pixel⁴⁷, we suggest to use an adaptive neighborhood, which means that we consider different kinds of neighbor sets, as in Ref. 48, and at each site we select the one that leads to the smallest energy^{48,49}. The adaptivity of the neighborhood aims to take into account the geometrical properties of the different areas in our original image, a spatial feature that plays a primary role especially in high-resolution imagery.

In Eq. (9), we notice the presence of an unknown positive parameter β to estimate. This parameter can be estimated by minimizing a pseudo-likelihood over a training set. We stress here that this method brings to adequate estimates only when using an exhaustive ground truth, or, at least, by taking into account a sufficient amount of class borders, which is rarely the case, in particular in remote sensing. For this reason, we determine this parameter by trial-and-error. The results of different experiments have led us to choose $\beta = 4.8$, which is quite a high value.

4.3 Posterior probabilities and their estimation using MPM

Since the quad-tree has, by definition, no cycles, the labels can be estimated exactly and non iteratively by MPM via a forward-backward algorithm, similar to the classical Baum algorithm for Markov chains⁵⁰. The aim is to maximize the posterior marginal at each site s :

$$\hat{x}_s = \underset{x_s}{\operatorname{argmax}} p(x_s|y). \quad (10)$$

A classical MPM algorithm²⁶ would need to estimate the posterior probability at level 0 given the observations at each level $y = \{y_s, \forall s \in S, \forall n \in [0; R]\}$. This estimation is generally done in 2 passes, referred to as bottom-up (“forward”) and top-down (“backward”) passes. In our case, we truncate the top-down pass, by using the highest tree-level label estimates to update the prior, and to run a novel MPM algorithm on a smaller quad-tree (see Figs. 2 and 3).

Bottom-up pass

This pass aims to estimate for each site $s \in S$ the partial posterior marginals $p(x_s|y_{d(s)})$ that are needed for the complete posterior probabilities $p(x_s|y)$ estimation (top-down pass). The probabilities $p(x_s|y_{d(s)})$ at a given level n are used to estimate $p(x_s|y_{d(s)})$ at level $n + 1$. In fact, Laferte et al.²⁶ showed that

$$p(x_s|y_{d(s)}) = \begin{cases} p(x_s|y_s) = \frac{1}{Z} p(y_s|x_s) p(x_s), & \text{at level 0} \\ \frac{1}{Z} p(y_s|x_s) p(x_s) \prod_{t \in s^+} \sum_{x_t} \left[\frac{p(x_t|y_{d(t)})}{p(x_t)} p(x_t|x_s) \right], & \text{otherwise} \end{cases}. \quad (11)$$

Thus, we proceed to a recursion, starting from the leaves and proceeding until the root is reached.

Top-down pass

As already mentioned, the modified MPM algorithm proposed in this paper combines the MPM-based algorithm of Laferte et al.²⁶ and the prior update introduced in Sec. 4.2. In that case, the estimate \hat{x}_s is only needed at the highest level of the currently considered quad-tree. At the coarsest level R , $p(x_s|y) = p(x_s|y_{d(s)})$. Hence, the classification map at this level is directly estimated via the relation $\hat{x}_s = \operatorname{argmax} p(x_s|y)$ for any $s \in S^R$. The $p(x_s|y)$ maximization is done by employing a modified Metropolis Dynamics algorithm²⁹ (MMD). To apply

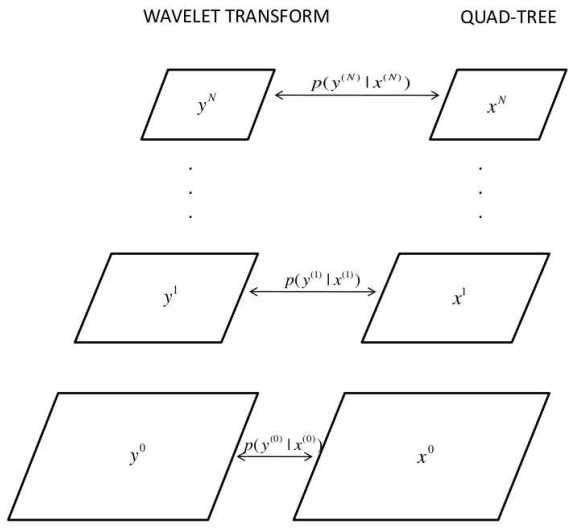


Figure 2. Generic hierarchical graph-based model of the quad-tree.

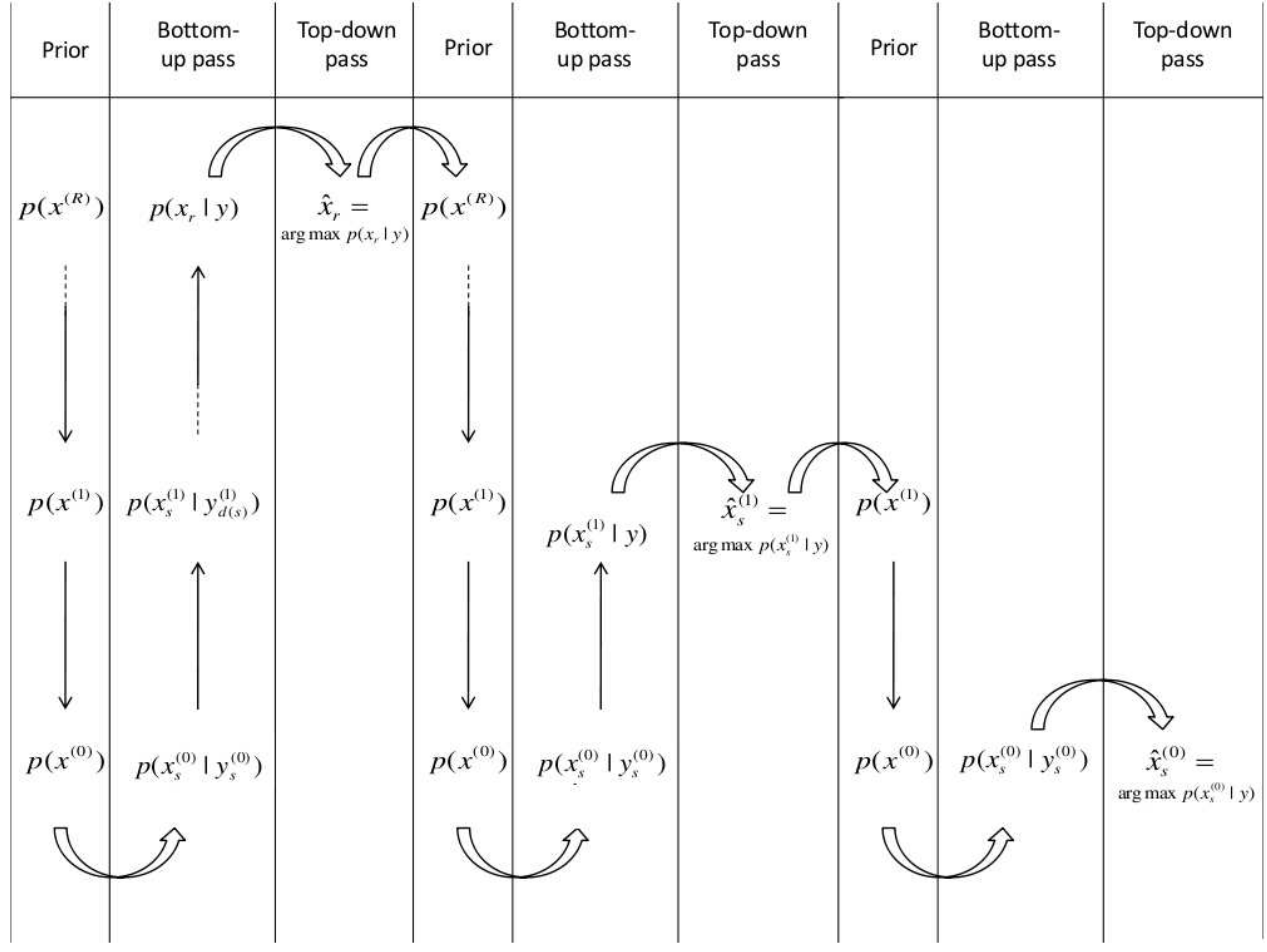


Figure 3. Proposed MPM estimation on the quad-tree represented in Fig. 2. In this representation, $R = 2$.

this algorithm, we do not maximize directly $p(x_s|y)$, but minimize the negative of its logarithm instead, which is possible since the logarithm is an increasing function. MMD algorithm has good properties for both its relative low computation time and the good precision of its results²⁹.

Finally, in order to improve the properties of the classification maps reported by our MPM-based method, we have applied a morphological majority voting procedure⁵¹. This smoothing part of the classification algorithm is optional and can be recommended when the final classification map demonstrates excessive salt-and-pepper type of noise due to the consideration of noisy input images (typically, SAR acquisitions).

5. EXPERIMENTAL RESULTS

The developed hierarchical classification approach is applied to three different data sets. Two of them are based on the classification of remote sensing images of cities, and the third one deals with the classification of an histological image. For each considered method and each data set, we give the final classification map and the corresponding classification accuracies, obtained on test sets that do not overlap with the training sets. Each training and test set is endowed with a manually annotated ground truth map and each represents between 7% and 10% of the whole image. In general, these manually-built ground truths are selected in homogeneous areas, meaning that the borders are not taken into account in order to guarantee the precision of the sets.

In the case of monoresolution acquisitions, we apply a 2-D discrete wavelet transform²⁴ to create a multiresolution input image that can be introduced in the hierarchical model. We then decompose our original image (mono- or multi-band) along different scales, corresponding to the tree levels. Empirically, we have concluded that good results are obtained with the hierarchical decomposition on $R = 2$ levels, R being the depth of decomposition. For each scale, we consider solely the approximation coefficients. The scale factor is always a power of 2, thus leading to the required quad-tree configuration. The approximation coefficients at scale 0 correspond to the original image. By filtering and decimating this image through the application of a low-pass filter to the rows and columns, we obtain the approximation coefficients for scale 1. Similarly, the approximation coefficients at scale j are decomposed by filtering and decimation so as to obtain the coefficients at scale $j + 1$, for $j < R$. A wide choice of wavelet functions exists, such as Daubechies, (bi-)orthogonal, and after comparison of the diverse classification performances, we finally chose Daubechies-10 wavelets²⁵ to decompose SAR images and Haar wavelets²⁴ to decompose optical images.

The β parameter of the Markov random field in the MPM-based algorithm (see Sec. 4.2) was empirically set to $\beta = 4.8$. As already mentioned in the introduction, when considering monoresolution acquisitions, we can simply combine the mixture-based likelihood modeling to a spatial MRF (already used in Ref. 52), instead of using a hierarchical model. In that case, the β parameter (see Sec. 4.2) is set to $\beta = 1.3$, and the optimization is done preferentially via a graph-cut scheme³⁰⁻³², faster than the MMD. We compared the two Markovian models on the remotely sensed and the histological images.

The use of graph-cuts for energy minimization has not been studied for the proposed MPM-based algorithm (Sec. 4.3). We did not apply the same graph-cuts implementation as in the MRF-based method because in this case, the Potts-based prior estimation is integrated in the graph-cut scheme. In the hierarchical model proposed in this paper, the prior is estimated preliminary. This main difference between the two algorithms also explains why the β parameter is set to different values.

5.1 Histological image

The considered image is a red, green, blue (RGB) histological image of the skin provided by Galderma of 550×1020 pixels. The R, G, and B bands are considered as the input features and the class conditional PDFs are modeled by using normal mixtures (see Sec. 2.1.1). This image is classified into 4 classes that were interpreted by a dermatological expert as the cytoplasm (in yellow in the classification maps), the nuclei (in blue) and the background (in red). The green class gathers the dermis matrix, the collagen and the stratum corneum keratin. Each of these classes is modeled by using our multivariate copula-based model (Sec. 2). Given the fact that this image is monoresolution, we applied, for comparison, the proposed hierarchical model and a single-scale MRF model. The multiresolution decompositions are obtained by Haar wavelet transform on $R = 2$ levels.

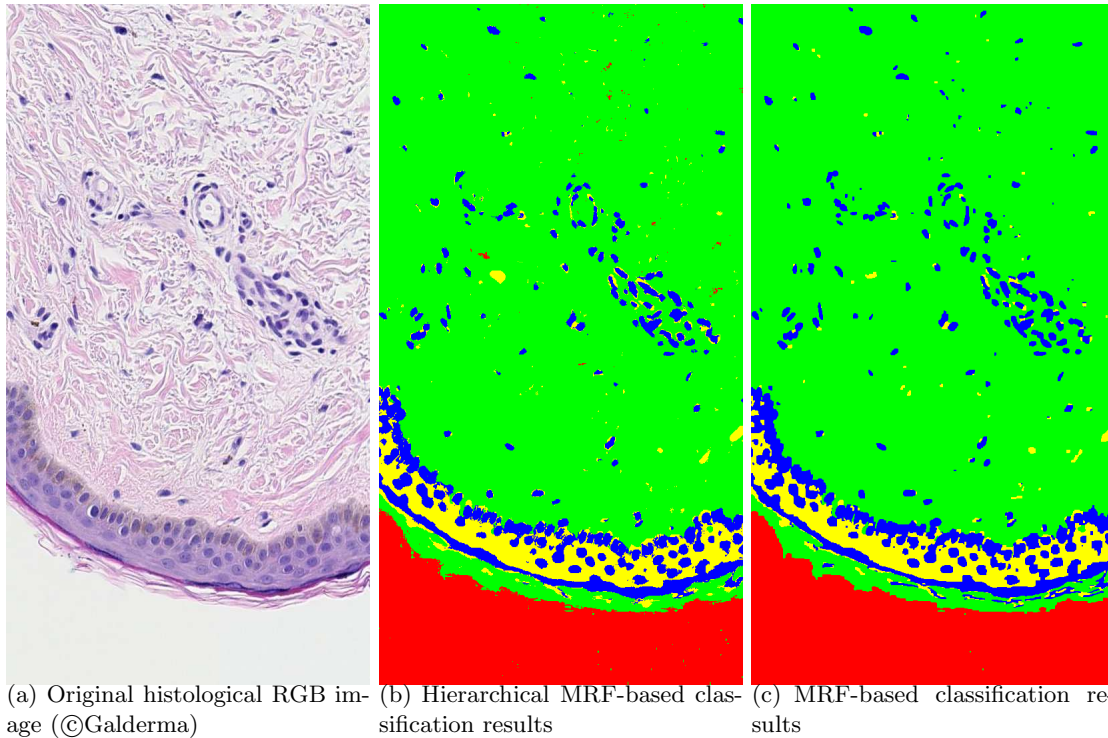


Figure 4. Original RGB histological image and classification results obtained with the two contextual methods.

Table 1. Accuracy for each of the 4 classes and overall results for the test areas of the histological image.

	Histological image				
	nuclei	dermis	background	cytoplasm	overall
Hierarchical MRF-based classif.	97.08%	99.87%	97.71%	97.13%	97.95%
MRF-based classif.	99.92%	99.97%	97.72%	96.65%	98.56%

For this specific case, by looking at both visual (Fig. 4) and numerical (Tab. 1) results, we can notice that the two Markovian-based methods lead to similar results. For this reason, we tend to favor the fastest method, that is the MRF-based method. In fact, we computed computation time for the various experiments that were conducted on Intel Xeon 2.40GHz, Linux system. The MRF-based algorithm runs in approximately 3 minutes, whereas our hierarchical method takes approximately 8 minutes to estimate the final classification map.

5.2 Amiens, France

The considered images are two single-pol COSMO-SkyMed SAR images of the city of Amiens (France) (©ASI, 2011):

- a StripMap acquisition (2.5 m pixel spacing), HH polarized, geocoded, single-look image. 510×1200 pixels, shown in Fig. 5(a).
- a PingPong acquisition (5 m pixel spacing), HH polarized, geocoded, 255×600 pixels, shown in Fig. 5(b).

In this case, we deal with four classes: urban (in red), water (in blue), vegetation (in green) and trees (in yellow). The hierarchical tree considered here has a scale $R = 1$.

To improve the urban area detection, we propose to extract a Greylevel co-occurrence matrix (GLCM)-based textural information from the original SAR images at each decomposition level, as already suggested in Ref. 52.

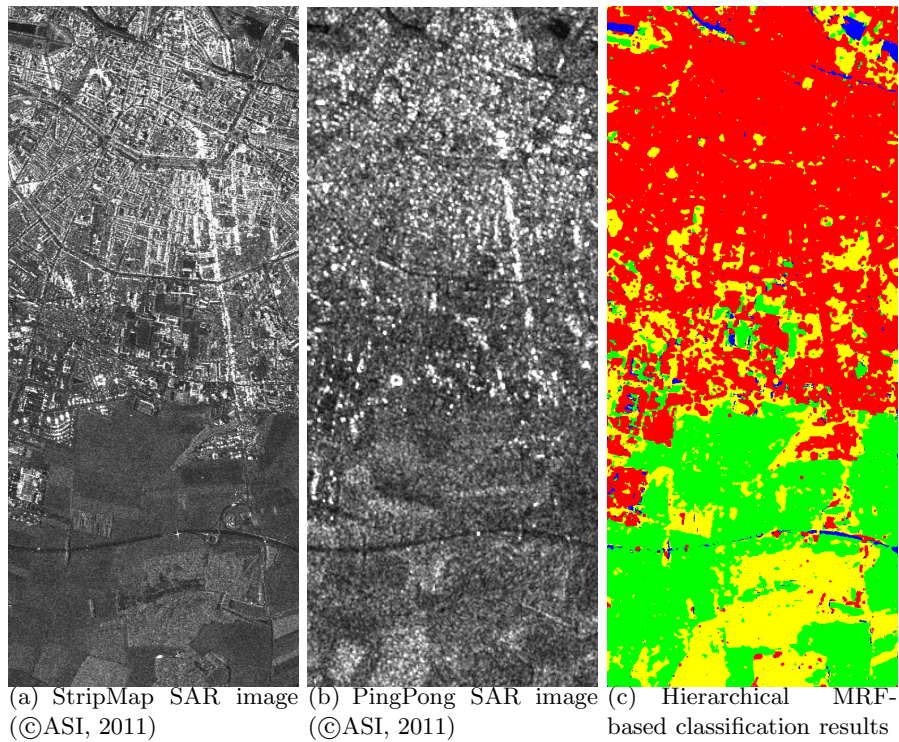


Figure 5. Original coregistered SAR images of Amiens in France acquired at different resolutions and the resulting classification map, obtained by using as input a combination of the SAR images and their GLCM textural features.

Table 2. Accuracy for each of the 4 classes and overall results for the test areas of Amiens.

	Amiens				
	water	urban	vegetation	tree	overall
Hierarchical MRF-based classif. with feature	98.05%	97.69%	85.90%	94.85%	94.12%
Hierarchical MRF-based classif. without feature	97.52%	93.80%	75.81%	93.16%	90.07%

The PDFs of both the SAR image and its textural feature are independently modeled by using generalized Gamma mixtures (see Sec. 2.1.2), and then combined via multivariate copulas.

The final classification map is given in Fig. 5(c). The final numerical results (Tab. 2) are more accurate when introducing a textural information that efficiently improves the urban area detection. The loss of accuracy in the vegetation class for all the methods (Tab. 2) is due to misclassifications of this area with trees at the bottom of the image. By visually analyzing the SAR image, we cannot really see a difference between these areas. Multiple polarization acquisition would be helpful in that case.

5.3 Port-au-Prince, Haiti

The considered input consists of two images of the quay of Port-au-Prince (Haiti):

- a single-polarized COSMO-SkyMed SAR image (©ASI, 2010), HH polarization, StripMap acquisition mode (2.5 m pixel spacing), geocoded, single-look, 920×820 pixels, shown in Fig. 6(a).
- a coregistered panchromatic GeoEye acquisition (©GeoEye, 2010), 920×820 pixels, shown in Fig. 6(b).

Five classes were chosen in that case: the water class (in blue in the classification maps), the urban areas (in red), the vegetation (in green), the sand (in yellow) and the containers (in pink).

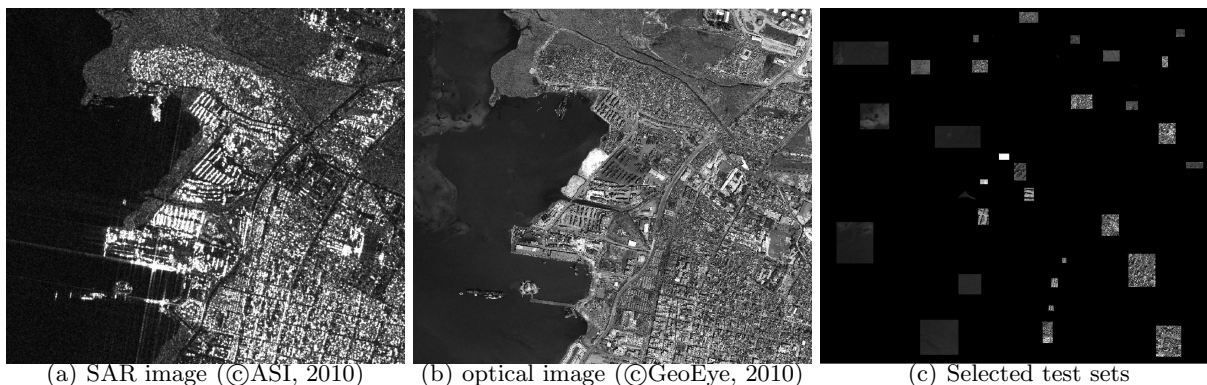
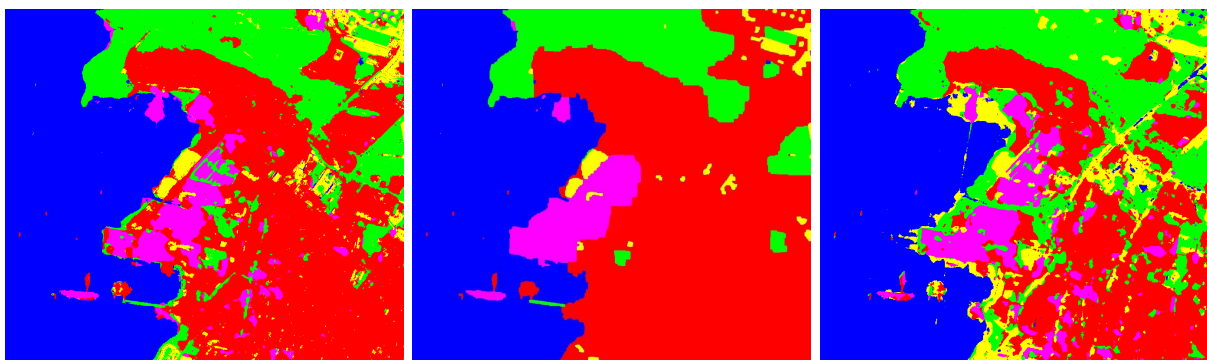


Figure 6. Original SAR image of Port-au-Prince in Haiti, coregistered panchromatic optical image and the selected ground truth used to test the results.

Table 3. Accuracy for each of the 5 classes and overall results for the test areas of Port-au-Prince.

	Port-au-Prince					
	water	urban	vegetation	sand	containers	overall
Hierarchical MRF-based classif.	99.60%	94.15%	98.48%	100%	79.28%	94.30%
MRF-based classif.	99.38%	100%	98.07%	100%	99.91%	99.47%
SAR only (Hierarchical MRF)	99.08%	87.80%	95.41%	52.38%	84.26%	83.79%
Panchro. only (Hierarchical MRF)	96.87%	89.48%	97.97%	100%	38.19%	84.50%



(a) Hierarchical MRF-based classification obtained for both optical and SAR images (b) MRF-based classification obtained for both optical and SAR images (c) Hierarchical MRF-based classification obtained for the SAR image

Figure 7. (a,b) Port-au-Prince classification maps obtained applying different Markovian contexts on optical/SAR data. (c) Classification map obtained with the hierarchical method applied to the SAR image.

As in the case of the histological image, we compare visual (Fig. 7) and numerical (Tab. 3) results obtained when using the hierarchical and the monoresolution MRF-based methods. For the hierarchical method, the multiresolution decompositions are obtained by Haar wavelet (for panchromatic) and Daubechies (for SAR) transforms on $R = 2$ levels. Numerical results indicate that the MRF-based method leads to a higher numerical accuracy. But we stress here that the map obtained by using the MRF-based method is severely oversmoothed, and this affects only marginally the numerical accuracies due to the localization of the test samples inside homogeneous areas (see Fig. 6(c)). However, when looking at the classification maps (Fig. 7), we can notice that the classification is more detailed when using a hierarchical decomposition, that makes us favor this method for the optical/SAR classification. Moreover, we also compared numerically (Tab. 3) and visually (Fig. 7) the results obtained with the proposed hierarchical method when considering respectively only SAR, only panchromatic and both images. We can obviously see the improvements related to the combination of the two images. In fact,

the optical image has a relevant effect in the sand discrimination, and the SAR acquisition is very helpful to detect the containers. Thus, the consideration of both of them allows to obtain good classification results for the container and the sand areas, and to improve the urban detection.

6. CONCLUSION

The method proposed in this paper allows to deal with multisensor, multiband, and/or multiresolution acquisitions. It combines a joint statistical modeling of considered input images (optical or radar imagery), with a hierarchical Markov random field, leading to a statistical supervised classification approach. We have proposed a copula-based multivariate statistical model that enables to fuse multisensor acquisitions, and we have developed a novel MPM-based hierarchical Markov random field model that iteratively updates the prior probabilities and, thus, leads to the improved robustness of the classifier. The hierarchical MRF considered here has two advantages: it is quite robust to speckle noise, and we can apply a non-iterative optimization algorithm (MPM estimation).

We analyzed the results obtained with the proposed method. Moreover, when dealing with monoresolution images, we compared the developed multiresolution hierarchical model to a spatial monoresolution MRF-based classification algorithm. The results were assessed both qualitatively (classification maps) and quantitatively (classification accuracies). The proposed method leads to good classification results for both optical and speckle-affected radar images. The MRF-based method is also efficient (faster and numerically better) and leads to good results. However, we could notice the smoothing effects of this method that may degrade the real accuracy of the final classification map by hiding some details.

The consideration of a quad-tree tends to limit the multiresolution approach by requiring a dyadic decomposition. A possible improvement would be then to find a new hierarchical Markovian-based algorithm that may take into account all resolution sizes, so as to make this method even more general.

ACKNOWLEDGMENTS

The authors would like to thank:

- The Direction Générale de l'Armement (DGA, France) and Institut National de Recherche en Informatique et Automatique (INRIA, France) for the partial financial support.
- The Italian Space Agency (ASI) for providing the COSMO-SkyMed images.
- GeoEye Inc. and Google crisis response for providing the GeoEye images available on the website <http://www.google.com/relief/haitiearthquake/geoeye.html>.
- Galderma R&D Early Development (Sophia Antipolis, France) for providing the histological images.
- Dr. Michaela De Martino from the University of Genoa (Italy) for her help with the ground truth maps of the remote sensing images.
- Pr. Zoltan Kato from the University of Szeged (Hungary) for his help with the graph-cuts implementation.

REFERENCES

- [1] Goshtasby, A. A., [*2-D and 3-D image registration: for medical, remote sensing, and industrial applications*], Wiley-Interscience (2005).
- [2] Oliver, C. and Quegan, S., [*Understanding Synthetic Aperture Radar images*], SciTech Publishing (2004).
- [3] Schapire, R. E., "A brief introduction to boosting," in [*Proceedings of the 16th International Joint Conference on Artificial Intelligence*], **16**(2), 1401–1406 (1999).
- [4] Breiman, L., "Bagging predictors," *Machine learning* **24**(2), 123–140 (1996).
- [5] Xu, L., Krzyzak, A., and Suen, C. Y., "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man, Cybern.* **22**(3), 418–435 (1992).

- [6] Benediktsson, J. A. and Kanellopoulos, I., “Classification of multisource and hyperspectral data based on decision fusion,” *IEEE Trans. Geosci. Remote Sens.* **37**(3), 1367–1377 (1999).
- [7] Waske, B. and van der Linden, S., “Classifying multilevel imagery from SAR and optical sensors by decision fusion,” *IEEE Trans. Geosci. Remote Sens.* **46**(5), 1457–1466 (2008).
- [8] Waske, B. and Benediktsson, J. A., “Fusion of support vector machines for classification of multisensor data,” *IEEE Trans. Geosci. Remote Sens.* **45**(12), 3858–3866 (2007).
- [9] Al-Ani, A. and Deriche, M., “A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence,” *Journal Of Artificial Intelligence Research* **17**, 333–361 (2011).
- [10] Barnard, G. A. and Bayes, T., “Studies in the history of probability and statistics: IX. Thomas Bayes’s essay towards solving a problem in the doctrine of chances,” *Biometrika* **45**(3/4), 293–315 (1958).
- [11] Benediktsson, J. A. and Swain, P. H., “Consensus theoretic classification methods,” *IEEE Trans. Syst., Man, Cybern.* **22**(4), 688–704 (1992).
- [12] Benediktsson, J. A., Sveinsson, J. R., and Swain, P. H., “Hybrid consensus theoretic classification,” *IEEE Trans. Geosci. Remote Sens.* **35**(4), 833–843 (1997).
- [13] Solberg, A. H. S., Taxt, T., and Jain, A. K., “A Markov random field model for classification of multisource satellite imagery,” *IEEE Trans. Geosci. Remote Sens.* **34**(1), 100–113 (1996).
- [14] D’Elia, C., Poggi, G., and Scarpa, G., “A tree-structured Markov random field model for Bayesian image segmentation,” *IEEE Trans. Image Process.* **12**(10), 1259–1273 (2003).
- [15] Nelsen, R. B., [*An introduction to copulas*], Springer, New York, 2nd ed. (2006).
- [16] Graffigne, C., Heitz, F., Perez, P., Preteux, F., Sigelle, M., and Zerubia, J., “Hierarchical Markov random field models applied to image analysis: a review,” in [*Proc. of the Conf. on Neural, Morphological and Stochastic Methods in Image Proc. in SPIE’s International Symposium on Optical Science, Engineering and Instrumentation*], (1995).
- [17] Fieguth, P., [*Statistical image processing and multidimensional modeling*], Springer (2011).
- [18] Wang, Z., Ziou, D., Armenakis, C., Li, D., and Li, Q., “A comparative analysis of image fusion methods,” *IEEE Trans. Geosci. Remote Sens.* **43**(6), 1391–1402 (2005).
- [19] Moser, G., Serpico, S. B., and Zerubia, J., “Dictionary-based Stochastic Expectation Maximization for SAR amplitude probability density function estimation,” *IEEE Trans. Geosci. Remote Sens.* **44**(1), 188–199 (2006).
- [20] Celeux, G., Cheveau, D., and Diebolt, J., “On stochastic versions of the EM algorithm,” Research report 2514, INRIA, France (1995).
- [21] Li, H.-C., Hong, W., Wu, Y.-R., and P.-Z.-Fan, “On the empirical-statistical modeling of SAR images with generalized gamma distribution,” *IEEE J. Sel. Top. Signal Process.* **5**(3), 386–397 (2011).
- [22] Krylov, V., Moser, G., Serpico, S. B., and Zerubia, J., “Enhanced dictionary-based SAR amplitude distribution estimation and its validation with very high-resolution data,” *IEEE Geosci. Remote Sens. Lett.* **8**(1), 148–152 (2011).
- [23] Tison, C., Nicolas, J.-M., Tupin, F., and Maitre, H., “A new statistical model for Markovian classification of urban areas in high-resolution SAR images,” *IEEE Trans. Geosci. Remote Sens.* **42**(10), 2046–2057 (2004).
- [24] Mallat, S. G., [*A wavelet tour of signal processing*], Academic Press, 3rd ed. (2008).
- [25] Daubechies, I., “Orthonormal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics* **41**(7), 909–996 (1988).
- [26] Laferte, J.-M., Perez, P., and Heitz, F., “Discrete Markov modeling and inference on the quad-tree,” *IEEE Trans. Image Process.* **9**(3), 390–404 (2000).
- [27] Voisin, A., Krylov, V., Moser, G., Serpico, S. B., and Zerubia, J., “Classification of very high resolution sar images of urban areas,” Research report 7758, INRIA, France (oct 2011).
- [28] Dubes, R. C. and Jain, A., “Random field models in image analysis,” *Journal of Applied Statistics* **16**(2), 131–164 (1989).
- [29] Berthod, M., Kato, Z., Yu, S., and Zerubia, J., “Bayesian image classification using Markov random fields,” *Image and Vision Computing* **14**(4), 285–295 (1996).
- [30] Boykov, Y., Veksler, O., and Zabih, R., “Efficient approximate energy minimization via graph cuts,” *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).

- [31] Kolmogorov, V. and Zabih, R., “What energy functions can be minimized via graph cuts?,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(2), 147–159 (2004).
- [32] Boykov, Y. and Kolmogorov, V., “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004).
- [33] Figueiredo, M. A. T. and Jain, A., “Unsupervised learning of finite mixture models,” *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 381–396 (2002).
- [34] Anastassopoulos, V., Lampropoulos, G. A., Drosopoulos, A., and Rey, M., “High resolution radar clutter statistics,” *IEEE Trans. Aerosp. Electron. Syst.* **35**(1), 43–60 (1999).
- [35] Sneddon, I., [*The use of integral transforms*], McGraw-Hill, New York (1972).
- [36] Krylov, V., Moser, G., Serpico, S. B., and Zerubia, J., “On the method of logarithmic cumulants for parametric probability density function estimation,” Research report 7666, INRIA, France (jul 2011).
- [37] Krylov, V., Moser, G., Serpico, S. B., and Zerubia, J., “Supervised high resolution dual polarization SAR image classification by finite mixtures and copulas,” *IEEE Journal of Selected Topics in Signal Proc.* **5**(3), 554–566 (2011).
- [38] Joe, H., “Multivariate concordance,” *Journal of multivariate analysis* **35**(1), 12–30 (1990).
- [39] Gidas, B., “A renormalization group approach to image processing problems,” *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(2), 164–180 (1989).
- [40] Heitz, F., Perez, P., and Bouthemy, P., “Multiscale minimization of global energy functions in some visual recovery problems,” *CVGIP: image understanding* **59**(1), 125–134 (1994).
- [41] Fabre, E., “New fast smoothers for multiscale systems,” *IEEE Trans. Signal Process.* **44**(8), 1893–1911 (1996).
- [42] Laferte, J.-M., Heitz, F., Perez, P., and Fabre, E., “Hierarchical statistical models for the fusion of multiresolution image data,” in [*Proceedings of the 5th International Conference on Computer Vision (ICCV’95)*], 908–913 (June 1995).
- [43] Kato, Z., Berthod, M., and Zerubia, J., “A hierarchical Markov random field model and multitemperature annealing for parallel image classification,” *Graphical models and image processing* **58**(1), 18–37 (1996).
- [44] Marroquin, J., Mitter, S., and Poggio, T., “Probabilistic solution of ill-posed problems in computational vision,” *Journal of the American Statistical Association* **82**(397), 76–89 (1987).
- [45] Bouman, C. and Shapiro, M., “A multiscale random field model for Bayesian image segmentation,” *IEEE Trans. Image Process.* **3**(2), 162–177 (1994).
- [46] Besag, J., “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society* **36**(2), 192–236 (1974).
- [47] Geman, S. and Geman, D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(6), 721–741 (1984).
- [48] Smits, P. C. and Dellepiane, S. G., “Synthetic aperture radar image segmentation by a detail preserving Markov random field approach,” *IEEE Trans. Geosci. Remote Sens.* **35**(4), 844–857 (1997).
- [49] Zhong, P., Liu, F., and Wang, R., “A new MRF framework with dual adaptive contexts for image segmentation,” in [*International Conference on Computational Intelligence and Security*], 351–355 (Dec. 2007).
- [50] Baum, L. E., Petrie, T., Soules, G., and Weiss, N., “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” *IEEE Ann. Math. Stats* **41**(1), 164–171 (1970).
- [51] Soille, P., [*Morphological Image Analysis - Principles and Applications*], Springer Verlag, Berlin, Germany, 2nd ed. (2003).
- [52] Voisin, A., Moser, G., Krylov, V., Serpico, S. B., and Zerubia, J., “Classification of very high resolution SAR images of urban areas by dictionary-based mixture models, copulas and Markov random fields using textural features,” in [*Proc. of SPIE Symposium on Remote Sensing 2010*], **7830**, 78300O (Sept. 2010).