HAL
open science

Hierarchical Markov random fields for high resolution land cover classification of multisensor and multiresolution image time series

Ihsen Hedhli, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia

To cite this version:

Ihsen Hedhli, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia. Hierarchical Markov random fields for high resolution land cover classification of multisensor and multiresolution image time series. Abdourrahmane Atto; Francesca Bovolo; Lorenzo Bruzzone. Change Detection and Image Time Series Analysis, 2, ISTE-Wiley, pp.1-27, 2021, Supervised Methods, 9781789450576. hal-03466612

HAL Id: hal-03466612

https://inria.hal.science/hal-03466612v1

Submitted on 6 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter **1**

Hierarchical Markov random fields for high resolution land cover classification of multisensor and multiresolution image time series

Ihsen HEDHLI[1], Gabriele MOSER[2], Sebastiano B. SERPICO[2], Josiane ZERUBIA[3]

[1] Institute Intelligence and Data (IID), Université Laval, Canada; [2] University of Genoa, Italy; and [3] Inria, Université Cote d'Azur, France

1.1. Introduction

1.1.1. *The role of multisensor data in time series classification*

Accurate and time-efficient classification methods for multitemporal imagery and satellite image time series are important tools required to support rapid and reliable extraction of information on a monitored region, especially when an extensive area is considered. Given the substantial amount and variety of data available currently from last-generation very-high spatial resolution satellite missions, the main difficulty is to develop a classifier that utilizes the benefits of input time series that are possibly composed of multimission, multisensor, multiresolution, and multifrequency imagery (Gómez-Chova *et al*. 2015). From an application-oriented viewpoint, the goal is to take advantage of this variety of input sources to maximize accuracy and effectiveness of the resulting thematic mapping products. From a methodological viewpoint, this goal claims for the development of novel data fusion techniques that are flexible enough to support the joint classification of time series of images collected on the same area by different sensors at different times and associated with multiple spatial resolutions and wavelength ranges.

In this chapter, this joint fusion problem is addressed. First, an overview of the major concepts and of the recent literature in the area of remote sensing data fusion is presented (see Section 1.1.3). Then, two advanced methods for the joint supervised classification of multimission image time series including multisensor optical and Synthetic Aperture Radar (SAR) components acquired at multiple spatial resolutions are described (see Section 1.2). The two techniques address different problems of supervised classification of satellite image time series and share a common methodological formulation based on hierarchical Markov random field (MRF) models. Examples of experimental results obtained by the proposed approaches in the application to very high resolution time series are also presented and discussed (see Section 1.3).

On one hand, the use of multiresolution and multiband imagery has been previously shown to optimize the classification results in terms of accuracy and computation time. On the other hand, the integration of the temporal dimension into a classification scheme can both enhance the results in terms of reliability and capture the evolution in time of the monitored area. However, the joint problem of the fusion of several distinct data modalities (e.g., multitemporal, multiresolution, and multisensor) has been much more scarcely addressed in the remote sensing literature so far.

1.1.2. *Multisensor and multiresolution classification*

The availability of different kinds of sensors is very advantageous for land cover mapping applications. It allows us to capture a wide variety of properties of the objects contained in a scene as measured by each sensor at each acquisition time. These properties can be exploited to extract richer information about the imaged area. In particular, the opportunity of joint availability of SAR and optical images within a time series can possibly offer high resolution, all-weather, day/night, short revisit time data with polarimetric, multifrequency, and multispectral acquisition capabilities. This potential is especially emphasized by current satellite missions for Earth Observation (EO), e.g., Sentinel-1 and -2, Pléiades, TerraSAR-X, COSMO-SkyMed and COSMO-SkyMed Second Generation, RADARSAT-2 and RADARSAT Constellation, GeoEye-1, WorldView-1, -2, -3, and WorldView Legion, or PRISMA, which convey a huge potential for multisensor optical and SAR observations. They allow a spatially distributed and temporally repetitive view of the monitored area at multiple spatial scales. However, the use of multisource image analysis for land cover classification purposes has been mostly addressed so far by focusing on single-resolution multisensor optical-SAR imagery whereas the joint use of multisensor and multiresolution capabilities within a time series of images of the same scene has been more scarcely investigated. This approach bears the obvious advantage of simplicity but is, in general, suboptimal. From a methodological viewpoint, when multisensor (optical and SAR) or multiresolution images of a given

scene are available, using them separately discards part of the correlations among these multiple data sources and, most importantly, their complementarity. As

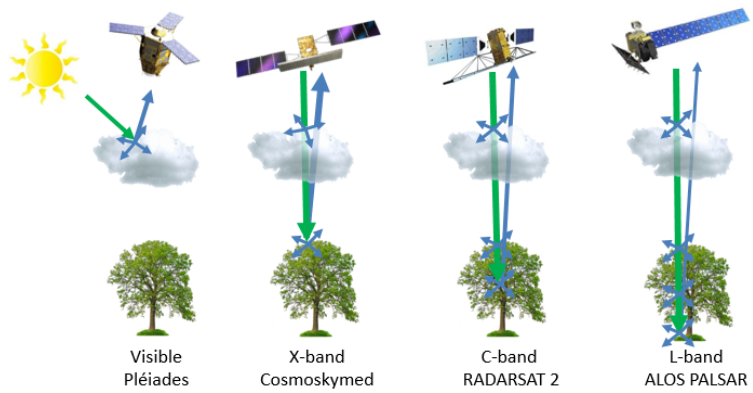

Figure 1.1. *Sensitivity to cloud cover and object size using different wavelength ranges*

illustrated in Figure 1.1, SAR and multispectral images exhibit complementary properties in terms of wavelength range (active microwave vs. passive visible and infrared), noisy behavior (often strong in SAR due to speckle, usually less critical in optical imagery), feasibility of photo-interpretation (usually easier with optical than with SAR data), impact of atmospheric conditions and cloud cover (strong for optical acquisitions and almost negligible for SAR) and sensitivity to Sun-illumination (strong for optical imagery and negligible for SAR) (Landgrebe 2005, Ulaby and Long 2015). This makes the joint use of high-resolution optical and SAR imagery particularly interesting for many applications related to environmental monitoring and risk management (Serpico *et al*. 2012).

Within this framework, there is a definite need for classification methods that automatically correlate different sets of images taken at different times on the same area from different sensors and at different resolutions. One way to address this problem is to resort to an explicit statistical modeling by finding a joint probability distribution given the class-conditional marginal probability density function (pdf) of the data collected by each sensor (see Figure 1.2). The joint statistics can be designed by resorting to meta-Gaussian distributions (Storvik *et al*. 2009), multivariate statistics such as multivariate copulas (Voisin *et al*. 2014), or non-parametric density estimators (Fukunaga 2013). However, employing heterogeneous data (SAR-optical in our case) makes the task of finding an appropriate multivariate statistical model complex, time demanding, and possibly prone to overfitting.

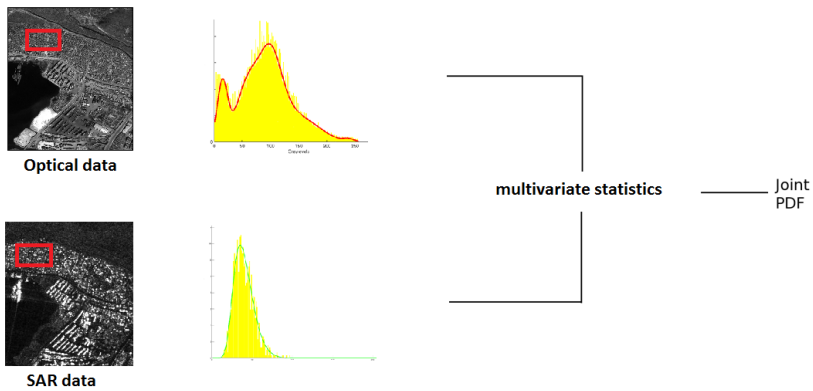

Figure 1.2. *Multivariate statistical modeling for optical-SAR data fusion*

In this context, the rationale of both approaches described in Section 1.2 is to take benefit of the data fusion capabilities of hierarchical MRFs and to avoid the computation of joint statistics. An approach based on multiple quad-trees in cascade and applied to multisensor and multiresolution fusion is described. In the first proposed method, for each sensor, the input images of the series are associated with separate quad-tree structures according to their resolutions. The goal is to generate a classification map based on a series of SAR and optical images acquired over the same area. The proposed approach formalizes, within this multiple quad-tree topology, a supervised Bayesian classifier that combines a class-conditional statistical model for pixel-wise information and a hierarchical MRF for multisensor and multiresolution contextual information. The second proposed method regards the case of the multimission fusion of multifrequency SAR data collected by the COSMO-SkyMed and RADARSAT-2 sensors together with optical Pléiades data. A multiple quad-tree structure is used again but optical and SAR images are both included in all cascaded quad-trees to take into account the specifics of the spatial resolutions of the considered satellite instruments. Compared to the first method, which considers the fusion of data from generally arbitrary SAR and optical sensors, this second method focuses on a specific combination of spaceborne SAR and optical sensors to investigate the synergy among the multifrequency and multiresolution information they provide.

1.1.3. *Previous work*

The literature in remote sensing data fusion is extensive, indicating the intense interest in this topic, as highlighted by the recent sharp increase in the number of papers published in the major remote sensing journals, and the increasing number of

related sessions in international conferences. Indeed, data fusion have given rise to a continuing tradition in remote sensing, since Earth observation is by definition dynamic (thus implying the multitemporal capability of remote sensing instruments), multiresolution (multiple spatial and spectral resolutions), and related to different physical quantities (thus requiring multiview/multisensor capability) (Waltz *et al*. 1990).

Data fusion is defined differently depending on the final goal of the user. Indeed, (Pohl and van Genderen 2014) and (Li *et al*. 1995) considered data fusion in remote sensing as the combination of two or more algorithms. This may include but is not restricted to *multiresolution fusion and pansharpening techniques* whose aim is to obtain multispectral images of increased spatial resolution (Vivone *et al*. 2015), *resolution blending* that consists in providing time series of data at their maximum spatial and spectral resolutions (referred to as parallel pansharpening in the multitemporal domain) (Huang and Song 2012), and *data fusion for missing information reconstruction* by using complementary data (Wang and Liang 2014).

An alternative perspective is to define data fusion in remote sensing as a decision fusion process that combines the information that is obtained from different data sets and provides sufficient generalization capability (Wald 1999). According to this definition, any type of image processing that combines two or more data sets, for example for land cover classification, atmospheric correction, or application of vegetation indices, could be considered as data fusion.

Within the former definition, various families of data fusion techniques have been proposed in the literature. On one hand, these methods may generally differ in their *application requirements*, such as the availability of ground reference data, the collected prior information, and/or some ancillary data that can be used in the development of the system according to a multisource processing architecture. On the other hand, it is important to properly understand the *user needs* with respect to economic costs and processing time and performance. Figure 1.3 summarizes the general architecture of a data fusion technique. As exposed before, the availability of remote sensing imagery at varying resolution has increased. Merging images of different spatial resolutions has become a significant operation in the field of digital remote sensing. A variety of different multiscale fusion approaches have been developed since the late Eighties. In the following, we give an overview on the most common approaches found in the literature. We can broadly divide them into two groups, (i) transformation techniques and (ii) modeling techniques.

Methods in (i) consist in replacing the entire set of multiscale images by a single composite representation that incorporates all relevant data. The multiscale transformations usually employ pyramid transforms (Burt 1984), the discrete wavelet transform (Piella 2003, Forster *et al*. 2004, Zhang and Hong 2005), the undecimated wavelet transform (Rockinger 1996, Chibani and Houacine 2003), the dual-tree

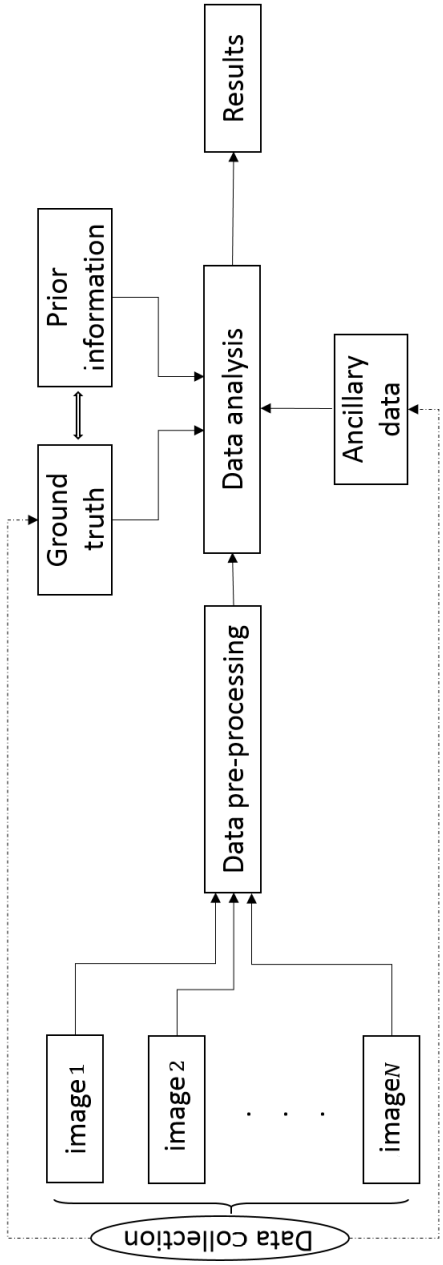

Figure 1.3. *General data fusion architecture. Images 1 through N have generally been acquired by distinct sensors, at different spatial resolutions, and/or with different radar frequencies and spectral bands,*

complex wavelet transform (Demirel and Anbarjafari 2010, Iqbal *et al*. 2013, Nelson *et al*. 2018, Zhang and Kingsbury 2015), the curvelet transform (Nencini *et al*. 2007, Choi *et al*. 2005), the contourlet transform (ALEjaily *et al*. 2008, Shah *et al*. 2008) and the nonsubsampled contourlet transform (Yang *et al*. 2007).

Techniques in (ii) include multiscale approaches with a focus on the use of the coarser resolutions in the data set in order to obtain fast computational algorithms. In the seminal papers (Basseville *et al*. 1992*a*,*b*), the basis for multiscale autoregressive modeling in dyadic trees was introduced. Since then, straightforward approaches were performed to deal with multiresolution images using trees (Pérez 1993, Kato and Zerubia 2012, Laferté *et al*. 2000, Chardin 2000, Voisin 2012, Hedhli *et al*. 2014). A detailed review of some of these methods can be found in (Graffigne *et al*. 1995) and (Willsky 2002).

In broader terms, multisensor analysis encompasses all processes dealing with data and information from multiple sensors to achieve refined/improved information as compared to the result that could be obtained by using data from only one individual source (Hall and Llinas 2001, Waltz *et al*. 1990, Pohl 1998). The accuracy of the classification of remote sensing images, for instance, is generally improved when multiple source image data are introduced in the processing chain in a suitable manner (e.g., (Nguyen *et al*. 2011, Gamba *et al*. 2011, Dousset and Gourmelon 2003, Hedhli *et al*. 2015)). As mentioned above, images from microwave and optical sensors provide complementary information that helps in discriminating the different classes. Several procedures have been introduced in the literature including, on one hand, post-classification techniques in which, first, the two data sets are separately segmented, and then the joint classification is produced by using, for example, random forest (e.g., (Waske and van der Linden 2008)), support vector machines with *ad-hoc* kernels (Muñoz-Marí *et al*. 2010), and artificial neural networks (Mas and Flores 2008). On the other hand, other methods directly classify the combined multisensor data by using for instance, statistical mixture models (e.g., (Dousset and Gourmelon 2003, Voisin *et al*. 2012, Prendes 2015)), entropy based techniques (e.g., (Roberts *et al*. 2008)), and fuzzy analysis (e.g., (Benz 1999, Stroppiana *et al*. 2015)). Furthermore, for complex data, especially when dealing with urban areas, radar images can contribute to differentiate between different land covers owing to the differences in surface roughness, shape, and moisture content of the observed ground surface (e.g., (Brunner *et al*. 2010)). The use of multisensor data in image classification becomes more and more popular with the increased availability of sophisticated software and hardware facilities to handle the increasing volumes of data. The decision on which of these techniques is the most suitable is very much driven by the applications and the typology of input remote sensing data.

Recently with the exposure of neural networks, several multisensor data fusion techniques have been proposed based on feed-forward multilayer perceptron and Convolutional Neural Network (CNN) architectures. Indeed, the huge amount of data

makes the use of Deep Neural Network (DNN) models possible. Many effective multi-task approaches have been recently developed to train DNN models on some large-scale remote-sensing benchmarks (e.g., (Cheng *et al*. 2020, Carvalho *et al*. 2019, Chen *et al*. 2017)). The aims of these multi-task methods is to learn an embedding space from different sensors (i.e., task). This could be done by first learning the embedding of each modality separately and then combining all the learned features as a joint representation. Then, this representation is utilized as an input of the last layers for different high level visual applications, e.g., remote sensing classification, monitoring, or change detection. Alternatively, DNN models could be used as an heterogeneous data fusion framework, learning the related parameters from all the input sources (e.g., (Benedetti *et al*. 2018, Minh *et al*. 2018, Ghamisi *et al*. 2016)). Despite the regularization techniques used to mitigate the high computational complexity of DNN methods (Pan *et al*. 2015), the training of these techniques is still greedy and hard to converge especially with remote sensing data sets.

In the next section, we will describe two advanced methods for the supervised classification of multisource satellite image time series. These methods have the advantage of being applicable to series of two or more images taken by single or multiple sensors operating at the same or at different spatial resolutions and with the same or with different radar frequencies and spectral bands. In general, the available images in the series are temporally and spatially correlated. Indeed, temporal and spatial contextual constraints are unavoidable in multitemporal data interpretation. Within this framework, Markov models provide a convenient and consistent way of modeling context-dependent spatio-temporal entities originated from multiple information sources, such as images in a multitemporal, multisensor, multiresolution, and multimission context.

1.2. Methodology

1.2.1. *Overview of the proposed approaches*

Let us consider a time series $(\mathcal{I}_k(p,q))_{k,p,q}$ composed of K images, acquired over the same area on K acquisition dates by up to K optical and SAR different sensors. Each image in the series is generally composed of multiple features (i.e., it is vector-valued), possibly corresponding to distinct spectral bands or radar polarizations. Specifically, $\mathcal{I}_k(p,q)$ indicates the feature vector of pixel (p,q) in the k-th image in the series ($k = 1, 2, \ldots, K$). In general, each sensor may operate at a distinct spatial resolution, hence a multisensor and multiresolution time series is being considered.

The acquisition times of the images in the series are assumed to be close enough so that no significant changes occur in the land cover of the observed area. In particular,

we assume that no abrupt changes (e.g., due to natural disasters such as floods or earthquakes) occur within the overall time span of the series. This assumption makes it possible to use the whole time series to classify the land cover in the scene, by taking benefit from the complementary properties of the images acquired by different sensors and at different spatial resolutions. Furthermore, this assumption may be especially relevant when the temporal dynamic of the ground scene *per se* is an indicator of land cover membership, such as in the case of forested (e.g., deciduous vs evergreen) or agricultural areas. We denote as $\omega_1, \omega_2, \ldots, \omega_M$ the land cover classes in the scene and as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ their set. We operate in a supervised framework, hence we assume that training samples are available for all of these classes.

The overall formulation introduced in (Hedhli *et al*. 2016) to address multitemporal fusion in the case of single-sensor imagery and based on multiple quad-trees in cascade is generalized here to take benefit from the images acquired by different sensors and from their mutual synergy. The multiscale topology of the quad-trees and of hierarchical MRFs defined on quad-trees intrinsically allow multiresolution and multisensor data to be naturally fused in the land cover mapping process.

In this framework, two specific algorithms are defined. In the first one, the k-th image in the series is assigned to a separate quad-tree based on its own spatial resolution. A hierarchical MRF is defined on this quad-tree topology and inference on the resulting probabilistic graphical model is addressed using the Bayesian marginal posterior mode (MPM) criterion (Kato and Zerubia 2012). In the second proposed algorithm, the focus is on a specific case of multimission, multifrequency, and multiresolution time series: multi-frequency X-band COSMO-SkyMed and C-band RADARSAT-2 SAR images are used alongside with optical visible and near-infrared (VNIR) Pléiades data. This scenario is of special current interest both because of the potential of exploiting the synergy among these missions and especially in view of the recent COSMO-SkyMed Second Generation and RADARSAT Constellation programs. In the case of the second method, different quad-trees are also used, but optical and SAR data are both associated with each quad-tree to benefit from the finest resolution available from the considered sensors. Both approaches exploit the potential of hierarchical probabilistic graphical models (Kato and Zerubia 2012) to address challenging problems of multimodal classification of an image time series.

1.2.2. *Hierarchical model associated with the first proposed method*

Let us first define the multiple quad-tree structure associated with the first proposed method. The K images $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_K$ in the series are included in the finest-scale layers (i.e., the leaves) of K distinct quad-trees. Coarser-scale layers of each quad-tree are filled in by applying wavelet transforms to the image on the finest-scale layer (Mallat 2008). The roots of the K quad-trees are assumed to correspond to the

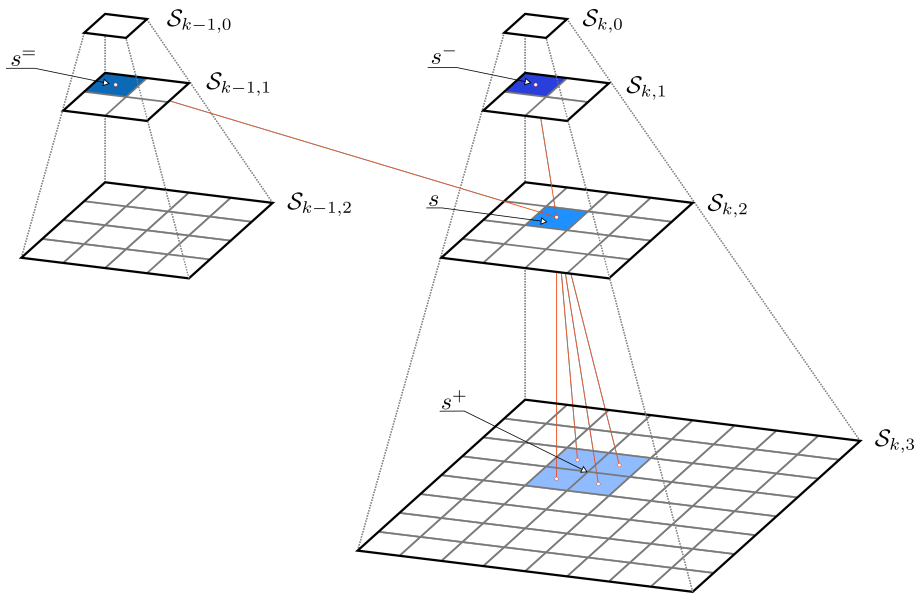

Figure 1.4. *Quad-trees associated with the input multisensor and multiresolution time series and related notations.*

same spatial resolution. The rationale of this hierarchical structure is that each image in the input series originates a separate multiscale quad-tree, each with generally a different number of layers and the input image on the leaves, and that the roots of these quad-trees share a common spatial resolution (see Fig. 1.4). This graph topology implicitly means that the spatial resolutions of the input images in the series are in a power-of-2 mutual relation. In general terms, this is a restriction but, with concern to current high-resolution satellite missions, this condition is easily met up to possible minor resampling.

Let $\boldsymbol{\mathcal{I}}_{k\ell}$ be the image associated with the ℓ-th layer of the k-th quadtree in the series. We shall index the common root with $\ell = 0$ and the leaves of the k-th quad-tree with $\ell = L_k$, i.e., $\boldsymbol{\mathcal{I}}_{k,L_k}$ coincides with the original input image $\boldsymbol{\mathcal{I}}_k$ ($k = 1, 2, \ldots, K$). The images $\boldsymbol{\mathcal{I}}_{k\ell}$ in the other layers ($\ell = 0, 1, \ldots, L_k - 1$) have been obtained through wavelets from $\boldsymbol{\mathcal{I}}_{k,L_k} = \boldsymbol{\mathcal{I}}_k$. The whole time series of multiscale images, either acquired by the considered sensors or obtained through wavelets, will be denoted as $\boldsymbol{\mathcal{I}} = (\boldsymbol{\mathcal{I}}_{k\ell}(p, q))_{k,\ell,p,q}$.

We will also indicate as $\mathcal{S}_{k\ell}$ the pixel lattice of the ℓ-th layer of the k-th quadtree ($\ell = 0, 1, \ldots, L_k; k = 1, 2, \ldots, K$). We shall denote as $s = (p, q)$ the coordinate pair of a generic pixel in one of these layers ($s \in \mathcal{S}_{k\ell}$). Following the literature of

hierarchical MRFs, s will be named site in the following. Sites in the described quad-tree structure are linked by parent-child relations – within each quad-tree and across consecutive quad-trees – as a function of their spatial scale. Specifically, if $s \in \mathcal{S}_{k\ell}$ is a site in the ℓ-th layer of the k-th quadtree and $\ell \in \{1, 2, \ldots, L_k\}$, i.e., s is not on the root layer, then $s^- \in \mathcal{S}_{k,\ell-1}$ indicates its parent node in the same quad-tree ($k = 1, 2, \ldots, K$). Similarly, if $s \in \mathcal{S}_{k\ell}$ with $\ell \in \{0, 1, \ldots, L_k - 1\}$, i.e., s is not on leaves layer, then $s^+ \subset \mathcal{S}_{k,\ell+1}$ denotes the set of its four children nodes in the same quad-tree. Finally, if $s \in \mathcal{S}_{k\ell}$ with $\ell \in \{1, 2, \ldots, L_k\}$ and $k \in \{2, 3, \ldots, K\}$, i.e., if s is neither in the first quad-tree of the series nor in the root of the other quad-trees, then $s^= \in \mathcal{S}_{k-1,\ell-1}$ indicates its parent node in the $(k-1)$-th quad-tree, i.e., in the quad-tree associated with the previous image of the series (see Fig. 1.4). From a graph-theoretic perspective, if the sites in the quad-trees are meant as nodes in a graph, then the pairs (s, s^-), $(s, s^=)$, and (s, r) with $r \in s^+$ define the corresponding edges.

Given this multiple quad-tree topology, a probabilistic graphical model based on a hierarchical MRF is defined. It is made of a series of random fields associated with the various scales and connected by transition relations associated with the links $s \mapsto s^-$, $s \mapsto s^=$, and $s \mapsto s^+$ among the sites. In particular, the quad-trees are meant to be in cascade, consistently with the input time series. Let $\mathcal{M}_{k\ell}(s) \in \Omega$ be the class label of site $s \in \mathcal{S}_{k\ell}$ ($\ell = 0, 1, \ldots, L_k; k = 1, 2, \ldots, K$), and let $\mathcal{M} = (\mathcal{M}_{k\ell}(s))_{k,\ell,s}$ be the corresponding time series of random fields associated with all multiscale layers. Each realization of $\mathcal{M}$ corresponds to a set of classification maps for all images in the series and all scales in the corresponding quad-trees.

The key assumption in the hierarchical MRF model is that the random fields $\mathcal{M}$ are Markovian both across scales and time ($\ell = 1, 2, \ldots, L_k; k = 2, 3, \ldots, K$) (Kato and Zerubia 2012):

$$P(\mathcal{M}_{k\ell} | \mathcal{M}_{hm}, h \leq k, m \leq \ell, (h, m) \neq (k, \ell)) = \\ = P(\mathcal{M}_{k\ell} | \mathcal{M}_{k,\ell-1}, \mathcal{M}_{k-1,\ell-1}), \quad [1.1]$$

where $P(\cdot)$ indicates the probability mass function (pmf) of discrete random variables and fields. Eq. [1.1] implies that the distribution of the labels in each layer of each quad-tree, conditioned on the labels in all above layers of the same quad-tree and of the previous quad-trees in the series, can be restricted to the distribution conditioned only on the labels of the upper layers in the same and previous quad-trees. Furthermore, this distribution factorizes in a conditionally independent fashion – a common assumption in the area of latent Markov models ($\ell = 1, 2, \ldots, L_k; k = 2, 3, \ldots, K$) (Li 2009, Kato and Zerubia 2012):

$$P(\mathcal{M}_{k\ell} \mid \mathcal{M}_{k,\ell-1}, \mathcal{M}_{k-1,\ell-1}) = \\ = \prod_{s \in \mathcal{S}_{k\ell}} P(\mathcal{M}_{k\ell}(s) | \mathcal{M}_{k,\ell-1}(s^-), \mathcal{M}_{k-1,\ell-1}(s^=)). \quad [1.2]$$

In the case of the first quad-tree in the series, these Markovianity and conditional independence assumptions are naturally adapted as follows ($\ell = 1, 2, \ldots, L_0$):

$$P(\mathcal{M}_{1,\ell}|\mathcal{M}_{1,m}, m < \ell) = P(\mathcal{M}_{1,\ell}|\mathcal{M}_{1,\ell-1}) = \\ = \prod_{s \in \mathcal{S}_{1,\ell}} P(\mathcal{M}_{1,\ell}(s)|\mathcal{M}_{1,\ell-1}(s^-)). \quad [1.3]$$

Finally, the feature vectors in the image time series $\boldsymbol{\mathcal{I}}$ are also assumed conditionally independent on the labels in $\mathcal{M}$:

$$p(\boldsymbol{\mathcal{I}}|\mathcal{M}) = \prod_{k=1}^{K} \prod_{\ell=0}^{L_k} p(\boldsymbol{\mathcal{I}}_{k\ell}|\mathcal{M}_{k\ell}) = \prod_{k=1}^{K} \prod_{\ell=0}^{L_k} \prod_{s \in \mathcal{S}_{k\ell}} p(\boldsymbol{\mathcal{I}}_{k\ell}(s)|\mathcal{M}_{k\ell}(s)), \quad [1.4]$$

where $p(\cdot)$ denotes the pdf of continuous random variables and fields. Again, this assumption is widely accepted in the literature of latent MRF models (Li 2009, Kato and Zerubia 2012).

To ease the notations, in the following, we shall write the feature vector $\boldsymbol{\mathcal{I}}_{k\ell}(s)$ and the class label $\mathcal{M}_{k\ell}(s)$ of site $s \in \mathcal{S}_{k\ell}$ simply as $\boldsymbol{x}_s$ and c_s, respectively, dropping the explicit dependence on k and ℓ for the sake of clarity. For the same reason, we shall explain the formulation of the first proposed method in the case of a series composed of $K = 2$ images $\boldsymbol{\mathcal{I}}_1$ and $\boldsymbol{\mathcal{I}}_2$ acquired by two different sensors and at two different resolutions on the considered area. In this case, two quad-trees in cascade are used. The extension to the case $K > 2$ is straightforward.

The formulation of MPM defined in (Hedhli *et al*. 2016) with regard to the case of multitemporal classification of single-sensor multiresolution imagery is generalized here to the case of multisensor data. The MPM decision rule assigns site $s \in \mathcal{S}_{k\ell}$ ($\ell = 0, 1, \ldots, L_k; k = 1, 2, \ldots, K$) the class label that maximizes the posterior marginal probability $P(c_s|\boldsymbol{\mathcal{I}})$, i.e., the distribution of its own individual label, given all feature vectors in the image series (Li 2009, Kato and Zerubia 2012). This decision rule is especially advantageous in the case of hierarchical graphs because it penalizes classification errors as a function of the scale at which they occur. Intuitively, an error on a site in the leaves layer directly affects only the corresponding pixel, whereas an error in a single pixel in the root layer may propagate into many erroneously labeled pixels on the leaves layer. MPM correctly penalizes the latter scenario more strongly than the former (Laferté *et al*. 2000).

As proven in (Laferté *et al*. 2000) and (Hedhli *et al*. 2016), under suitable conditional independence assumptions, the posterior marginal $P(c_s|\boldsymbol{\mathcal{I}})$ can be recursively expressed as a function of the posterior marginal $P(c_{s^-}|\boldsymbol{\mathcal{I}})$ of the parent

node s^- in the same quad-tree and, in the case $k = 2$, also of the posterior marginal $P(c_{s^=}|\boldsymbol{\mathcal{I}})$ of the parent node $s^=$ in the previous quad-tree:

$$P(c_s|\boldsymbol{\mathcal{I}}) = \begin{cases} \sum\limits_{x_{s^-}, x_{s^=}} P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s) P(c_{s^-}|\boldsymbol{\mathcal{I}}) P(x_{s^=}|\boldsymbol{\mathcal{I}}), & s \in \mathcal{S}_2^* \\ \sum\limits_{x_{s^-}} P(c_s|c_{s^-}, \boldsymbol{X}_s) P(c_{s^-}|\boldsymbol{\mathcal{I}}), & s \in \mathcal{S}_1^*, \end{cases} \quad [1.5]$$

where $\mathcal{S}_k^* = \bigcup_{k=1}^{L_k} \mathcal{S}_{k\ell}$ collects all sites except the root in the k-th quad-tree ($k = 1, 2$) and $\boldsymbol{X}_s$ is a vector collecting the features of all the descendants of site s. Through this formulation, MPM takes into consideration the information conveyed by the input multisensor data within the labeling of each site $s \in \mathcal{S}_2^*$.

1.2.3. *Hierarchical model associated with the second proposed method*

In the second proposed method, the case-specific problem of jointly classifying a multimission, multifrequency (radar X band, radar C band, and optical VNIR), and multiresolution series of images acquired by COSMO-SkyMed, RADARSAT-2, and Pléiades is addressed. All these sensors support multiple spatial resolutions, up to 0.5 m for Pléiades, approximately 1 m for COSMO-SkyMed, and 1 m $\times$ 2 m for RADARSAT-2.

Let a time series of three images, collected by COSMO-SkyMed, RADARSAT-2, and Pléiades on the same area at times close enough to assume land cover stability, be available. A case-specific hierarchical model, using again two quad-trees in cascade, is introduced to generate a supervised classification map at the finest among the spatial resolutions of the input images in the series. Unlike in the case of the hierarchical model of the first proposed approach, in which distinct quad-trees are related to different sensors, in this second approach, for each of the two considered radar sensors, the input image is included in a separate quad-tree according to its spatial resolution (see Fig. 1.5). These resolutions are generally expected to be coarser than the finest resolution that is observed through Pléiades. Hence, the input Pléiades data are inserted in the finest-resolution layers of both aforementioned quad-trees (see Fig. 1.5). The rationale is to both benefit from all input multisensor imagery and to map land cover at the finest resolution available. Empty layers of each of the two quad-trees are filled in with wavelet transforms computed from the Pléiades image included in the leaves layer (see Fig. 1.5). The Markovian formulation defined in the previous section, the recursive equations [1.5], and the remarks on the power-of-2 relation amongst the resolutions of the input images in the series are valid in this case as well.

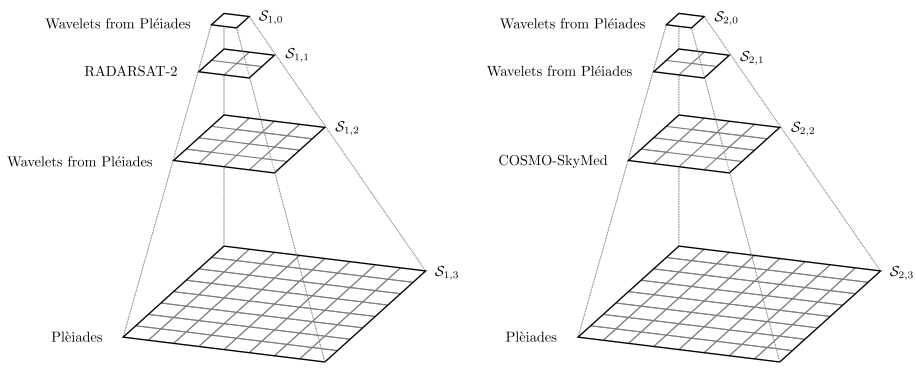

Figure 1.5. *Cascaded quad-trees associated with the second proposed method. In this example, we assume that the input time series includes Pléiades data at 0.5 m resolution, COSMO-SkyMed spotlight data at 1 m resolution, and RADARSAT-2 data sampled on a 2 m pixel lattice. Accordingly, the Pléiades image is included in the leaves layer of both quad-trees. The COSMO-SkyMed and RADARSAT-2 images are inserted in intermediate layers of separate quad-trees according to their resolutions. The empty layers of both quad-trees are filled in with wavelet transforms of the Pléiades imagery.*

1.2.4. *Multisensor hierarchical MPM inference*

Equation [1.5] allows recursively calculating the posterior marginal $P(c_s|\boldsymbol{\mathcal{I}})$ at each site $s \in \mathcal{S}_2^*$ of the second quad-tree (except the root) as long as the probabilities $P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s)$ become known. In particular, the focus on the second quad-tree is consistent with the cascade approach used within the proposed methods: given the input time series, the output classification map is obtained on the leaves of this second quad-tree (or on the leaves of the last quad-tree in the case of a longer time series).

Specifically, under appropriate assumptions we proved in (Hedhli *et al*. 2016) that, for all $s \in \mathcal{S}_2^*$:

$$P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s) \propto \frac{P(c_s|c_{s^-}, c_{s^=})P(c_{s^-}|c_{s^=})P(c_{s^=})P(c_s|\boldsymbol{X}_s)}{P(c_s)}, \qquad [1.6]$$

where $P(c_s|c_{s^-}, c_{s^=})$ is the parent-child transition probability across the two quad-trees, $P(c_s)$ is the prior probability, $P(c_{s^-}|c_{s^=})$ is the transition probability between two sites at the same scale, and $P(c_s|\boldsymbol{X}_s)$ is the partial posterior marginal probability. Given the obvious constraint $\sum_{c_s} P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s) = 1$, Eq. [1.6] determines $P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s)$ uniquely on all sites $s \in \mathcal{S}_2^*$. To determine these probabilities, the

approach developed in (Hedhli *et al*. 2016) for the multitemporal single-sensor case is generalized to the multisensor tasks considered here. This process is formalized within three recursive passes across the second quad-tree – one bottom-up and two top-down passes – which are described in the next subsections.

1.2.4.1. *Initializing on the first quad-tree*

First, to initialize the process, classification is performed using only the data included in the first quad-tree through a classical MPM on a single quad-tree. For details we refer the reader to (Laferté *et al*. 2000). Here, we only briefly recall that the algorithm is initialized by choosing the prior distribution on the root of this first quad-tree. In order to favor spatial regularity, this prior is selected here according to a Potts MRF model (Li 2009, Kato and Zerubia 2012). Details can be found in (Hedhli *et al*. 2016). From the perspective of the classification of the input series of multisensor images, the key point is that, after this initialization stage, the posterior marginal $P(c_s|\boldsymbol{\mathcal{I}})$ is known for each site $s \in \mathcal{S}_{1,\ell}$ ($\ell = 0, 1, \ldots L_1$) of the first quad-tree. Furthermore, $P(c_s|\boldsymbol{X}_s)$ is also obtained as an intermediate by-product on the same sites.

1.2.4.2. *First top-down pass*

In the first top-down pass, the second quad-tree is swept downward from the root to the leaves to calculate the prior $P(c_s)$ recursively. This prior is initialized in each site $s \in \mathcal{S}_{2,0}$ of the root of the second quad-tree as $P(c_s) = P(c_r|\boldsymbol{X}_r)$, where $r \in \mathcal{S}_{1,0}$ is the site of the root of the first quad-tree with the same spatial location as s. The partial posterior marginal $P(c_r|\boldsymbol{X}_r)$ has been derived in the aforementioned initialization. Then, the top-down pass travels along the other layers until reaching the leaves ($s \in \mathcal{S}_2^*$):

$$P(c_s) = \sum_{c_{s^-}} P(c_s|c_{s^-})P(c_{s^-}). \qquad [1.7]$$

This formulation encourages that parent and children sites share the same class label, although it does not enforce deterministically this condition. It also implies a model for the statistical relations between labels in consecutive layers. Here, the parent-child transition probability $P(c_s|c_{s^-})$ is expressed using the parametric model in (Bouman 1991):

$$P(c_s|c_{s^-}) = \begin{cases} \theta & \text{if } c_s = c_{s^-} \\ \dfrac{1-\theta}{M-1} & \text{otherwise,} \end{cases} \qquad [1.8]$$

where $\theta \in [0, 1]$ is a hyperparameter of the method. Experiments conducted in (Hedhli *et al*. 2016) indicated limited sensitivity of the result of MPM on multiple cascaded

quad-trees to the value of this hyperparameter. We also note that Eq. [1.8] implicitly yields a stationary model for the considered transitions, i.e., the probability $P\{c_s = \omega_m | c_{s^-} = \omega_n\}$ depends on the pair (ω_m, ω_n) of classes but not on the specific site location s ($s \in \mathcal{S}_2^*; m, n = 1, 2, \ldots, M$).

After the first top-down pass, the prior $P(c_s)$ is known on every site s of the second quad-tree.

1.2.4.3. *Bottom-up pass*

To calculate $P(c_s | c_{s^-}, c_{s^=}, \boldsymbol{X}_s)$ ($s \in \mathcal{S}_2^*$), a bottom-up step, traveling through the second quad-tree from the leaves to the root is used. It is based on Eq. [1.6], in which, besides the priors $P(c_s)$, which are known from the first top-down pass, three further probability distributions are necessary: (i) the transition probabilities at the same scale $P(c_{s^-} | c_{s^=})$; (ii) the parent-child transition probabilities $P(c_s | c_{s^-}, c_{s^=})$; and (iii) the partial posterior marginals $P(c_s | \boldsymbol{X}_s)$.

Concerning (i), the algorithm in (Bruzzone *et al*. 1999) is applied to estimate the multitemporal joint probability matrix, i.e., the $M \times M$ matrix J whose (m, n)-th entry is $J_{mn} = P\{c_{s^-} = \omega_m, c_{s^=} = \omega_n\}$ ($m, n = 1, 2, \ldots, M$). This technique is based on the expectation-maximization (EM) algorithm and addresses the problem of learning these joint probabilities as a parametric estimation task. Details can be found in (Hedhli *et al*. 2016). Once J has been estimated, $P(c_{s^-} | c_{s^=})$ is derived as an obvious by-product.

With regard to (ii), the parametric model in Eq. [1.8] is extended as follows:

$$P(c_s | c_{s^-}, c_{s^=}) = \begin{cases} \theta & \text{if } c_s = c_{s^-} = c_{s^=} \\ \phi & \text{if } (c_s = c_{s^-} \text{ or } c_s = c_{s^=}) \text{ and } c_{s^-} \neq c_{s^=} \\ \dfrac{1-\theta}{M-1} & \text{if } c_s \neq c_{s^-} \text{ and } c_s \neq c_{s^=} \text{ and } c_{s^-} = c_{s^=} \\ \dfrac{1-2\phi}{M-2} & \text{otherwise,} \end{cases} \quad [1.9]$$

where θ has the same meaning as in Eq. [1.8] and $\phi \in [0, 1]$ is a second hyperparameter.

Concerning (iii), it has been proved that, on all layers except the leaves (Laferté *et al*. 2000):

$$P(c_s | \boldsymbol{X}_s) \propto p(\boldsymbol{x}_s | c_s) P(c_s) \prod_{r \in s^+} \sum_{c_r} \frac{P(c_r | \boldsymbol{X}_r) P(c_r | c_s)}{P(c_r)} \quad [1.10]$$

for all sites $s \in \mathcal{S}_{2,\ell}, \ell = 0, 1, \ldots, L_2 - 1$. First, $P(c_s | \boldsymbol{X}_s)$ is initialized on the leaves of the second quad-tree by setting $P(c_s | \boldsymbol{X}_s) = P(c_s | \boldsymbol{x}_s) \propto p(\boldsymbol{x}_s | c_s) P(c_s)$

for $s \in \mathcal{S}_{2,L_2}$. Then, $P(c_s|\boldsymbol{X}_s)$ is calculated by using Eq. [1.10] while sweeping the second quad-tree upward until reaching the root. This recursive process makes use of the pixelwise class-conditional pdf $p(\boldsymbol{x}_s|c_s)$, whose modeling is discussed in Section 1.2.5. After the bottom-up pass, $P(c_s|c_{s^-}, c_{s^=}, \boldsymbol{X}_s)$ is known on every site s of the second quad-tree.

1.2.4.4. *Second top-down pass*

Finally, based on Eq. [1.5], the posterior marginal is initialized at the root of the second quad-tree as $P(c_s|\boldsymbol{\mathcal{I}}) = P(c_s|\boldsymbol{X}_s)$ for $s \in \mathcal{S}_{2,0}$. Then, given the probabilities that have been determined or modeled within the previous stages, $P(c_s|\boldsymbol{\mathcal{I}})$ is obtained on all sites $s \in \mathcal{S}_2^*$ of all other layers through Eq. [1.5] by sweeping the second quad-tree downward in a second top-down pass.

1.2.4.5. *Generation of the output map*

The aforementioned stages lead to the computation of the posterior marginal $P(c_s|\boldsymbol{\mathcal{I}})$ on every site s of the second quad-tree. In principle, site s could directly be given the label $\arg\max_{c_s} P(c_s|\boldsymbol{\mathcal{I}})$, i.e., the label that maximizes $P(c_s|\boldsymbol{\mathcal{I}})$ over the set Ω of classes. However, this strategy is often avoided in the literature of hierarchical MRFs because of its computational burden and of the risk of blocky artifacts (Laferté *et al*. 2000, Voisin *et al*. 2014). As an alternative, the case-specific formulation of the modified Metropolis dynamics (MMD) algorithm (Berthod *et al*. 1996) that was combined in (Hedhli *et al*. 2016) with MPM for the case of multitemporal single-sensor classification is generalized here to the multi-sensor case. For details we refer the reader to (Hedhli *et al*. 2016). In the case of both proposed methods, after this integrated MPM-MMD labeling, the classification result on the leaves of the second quad-tree provides the output classification map at the finest of the observed resolutions.

1.2.5. *Probability density estimation through finite mixtures*

For each class, each layer, and each quad-tree, a finite-mixture model (FMM) is used for the corresponding pixelwise class-conditional pdf. This means that the function $f_{k\ell m} : \boldsymbol{x}_s \mapsto p(\boldsymbol{x}_s|c_s = \omega_m)$ for $s \in \mathcal{S}_{k\ell}$ ($\ell = 0, 1, \ldots, L_k; k = 1, 2; m = 1, 2, \ldots, M$) is supposed to belong to the following class of pdfs:

$$\mathcal{F}(g) = \left\{ \sum_{n=1}^{N} \pi_n g(\cdot|\boldsymbol{\psi}_n) : N \in \mathbb{N},\ \pi_n \in [0,1],\ \sum_{n=1}^{N} \pi_n = 1,\ \boldsymbol{\psi}_n \in \Psi \right\}, \quad [1.11]$$

where $g(\cdot|\boldsymbol{\psi})$ is a pdf model depending on a vector $\boldsymbol{\psi} \in \Psi$ of parameters that takes values in a parameter set Ψ and every function $f(\boldsymbol{x}) = \sum_{n=1}^{N} \pi_n g(\boldsymbol{x}|\boldsymbol{\theta}_n)$ is a convex

linear combination of N such pdfs, parameterized by the parameter vectors $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_N \in \Psi$ and weighted by the proportions $\pi_1, \pi_2, \ldots, \pi_N$.

This modeling choice is motivated by the remarkable flexibility that FMMs offer in the characterization of data with heterogeneous statistics – a highly desirable property in the application to high spatial resolution remote sensing imagery (Hedhli *et al*. 2016). In the proposed methods, for each layer of each quad-tree, if the corresponding data are multispectral, then for all class-conditional pdfs, g is chosen to be a multivariate Gaussian, i.e., a Gaussian mixture model is used. In this case, the parameter vector $\boldsymbol{\psi}_n$ of each component obviously includes the related vector mean and covariance matrix ($n = 1, 2, \ldots, N$) (Landgrebe 2005). This model is also extended to the layers populated by wavelet transforms of optical data, consistently with the linearity of the wavelet operators.

On the contrary, for each layer that is populated by SAR data, all class-conditional pdfs are modeled using FMMs in which g is a generalized Gamma distribution, i.e., generalized Gamma mixtures are used. In this case, the parameter vector $\boldsymbol{\theta}_n$ of each n-th component includes a scale parameter and two shape parameters ($n = 1, 2, \ldots, N$). The choice of the generalized Gamma mixture is explained by its accuracy in the application to high spatial resolution SAR imagery (Li *et al*. 2011, Krylov *et al*. 2013). Here, we also generalize it – albeit empirically – to the layers populated with wavelet transforms of SAR imagery.

In all these cases, the FMM parameters are estimated through the stochastic expectation maximization (SEM) algorithm. SEM is an iterative stochastic parameter estimation technique that has been introduced for problems characterized by data incompleteness and that approaches maximum likelihood estimates under suitable assumptions (Celeux *et al*. 1996). It is separately applied to the training set of each class ω_m in each ℓ-th layer of each k-th quad-tree to model the corresponding class-conditional pdf $f_{k\ell m}$ ($\ell = 0, 1, \ldots, L_k; k = 1, 2; m = 1, 2, \ldots, M$). In the case of the generalized Gamma mixtures for the SAR layers, it is also integrated with the method of log-cumulants (Krylov *et al*. 2013). Details on this combination can be found in (Moser and Serpico 2009). We recall that SEM also automatically determines the number N of mixture components, for which only an upper bound has to be provided by the operator. This upper bound was set to 10 in all our experiments.

1.3. Examples of experimental results

1.3.1. *Results of the first method*

To experimentally validate the first method, a time series of two high-resolution images acquired in 2010 over Port-au-Prince, Haiti, has been used. The series was made of an HH-polarized single-look COSMO-SkyMed stripmap image with 2.5 m

Table 1.1. *First proposed method: classification accuracies and computation times of the proposed technique and of the previous algorithms in (Storvik* et al*. 2009) and (Voisin* et al*. 2014) on the test set of the time series composed of COSMO-SkyMed and GeoEye-1 images. Computation times refer to an Intel i7 quad-core, 2.40 GHz, 8-GB RAM, 64-bit Linux system.*

	Class-wise accuracies					overall	time
	water	urban	vegetation	bare soil	containers	accuracy	[s]
Proposed method	100%	78.12%	89.46%	98.78%	47.12%	82.69%	254
(Storvik *et al*. 2009)	99.95%	97.32%	90.81%	96.22%	37.25%	79.44%	298
(Voisin *et al*. 2014)	100%	75.24%	87.16%	98.89%	49.31%	82.12%	668

pixel spacing (325×400 pixels; see Fig. 1.6(a)) and of a GeoEye-1 image with 2.5 m spatial resolution (see Fig. 1.6(b)) and three channels in the visible wavelength range. The time lag between the two acquisitions was a few days. Five main land cover classes were present in the scene: urban, water, vegetation, soil, and containers. These classes were defined by an expert photointerpreter who also annotated their training and test samples.

In the approach taken by the first proposed method, the quad-trees are ordered, and consistently with the cascade approach, the output classification map is the result obtained on the leaves of the second quad-tree. Here, the GeoEye-1 and COSMO-SkyMed images were associated with the first and second quad-trees, respectively. The rationale of this choice is to initialize the land cover mapping result using the optical data and to finalize it through the fusion with SAR imagery. In the case of both quad-trees, to fill in the empty levels of the quad-tree, 2D Daubechies wavelets of order 10 were applied (Mallat 2008).

The classification result obtained by the proposed technique (see Fig. 1.6(g)) were compared to those generated by several previous approaches to multisensor and/or multiresolution classification. First, to compare with the result of a multiscale but single-sensor approach, the hierarchical MRF on a single quad-tree of (Laferté *et al*. 2000) was applied to classify the image collected by each sensor. In this case as well, the MPM criterion was used and the class-conditional pdfs were estimated using SEM together with multivariate Gaussian or generalized Gamma mixtures in the case of optical and SAR data, respectively (see Figs. 1.6(c) and (d)). Then, to compare with a multisensor multiscale approach, the algorithm in (Voisin *et al*. 2014) was used. It makes use of a hierarchical MRF on a single quad-tree whose layers are filled in with both optical and SAR data in a stacked vector fashion. Multisensor fusion is accomplished using multivariate copula functions (see Fig. 1.6(e)). Finally, to compare with a multisensor but single-scale approach, the technique in (Storvik *et al*. 2009)

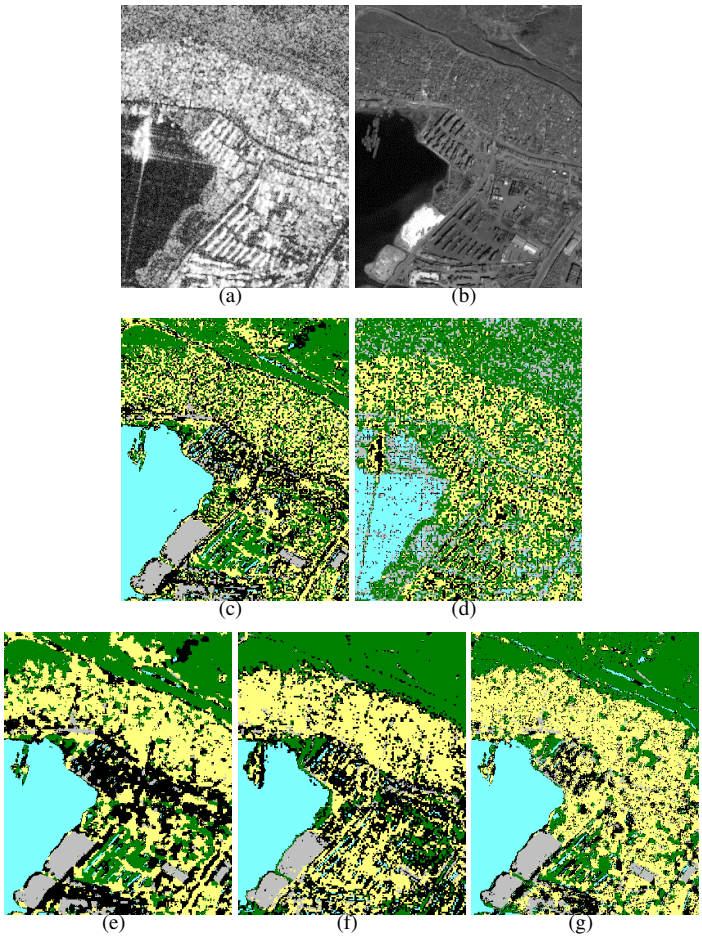

Figure 1.6. *First proposed method. (a) COSMO-SkyMed (©ASI 2010) and (b) GeoEye-1 (©GeoEye 2010) images of the input series. The former is shown after histogram equalization. The R-band of the latter is displayed. Classification maps obtained by separately classifying (c) the GeoEye-1 and (d) the COSMO-SkyMed images through a hierarchical MRF on a single quad-tree. Classification maps generated by (e) the multisensor multiscale method in (Voisin* et al. *2014), (f) the multisensor single-scale technique in (Storvik* et al. *2009), and (g) the proposed algorithm. Color legend:*
water ☐ *urban* ☐ *vegetation* ■ *bare soil* ■ *containers* ■

was considered after upsampling all the data to the pixel lattice at the finest resolution. In (Storvik *et al*. 2009), the joint class-conditional distributions of multisensor data are estimated using meta-Gaussian density functions (essentially equivalent to Gaussian copulas) and the maximum likelihood decision rule is applied to generate the output classification map (see Fig. 1.6(f)).

First, a visual qualitative inspection of the classification maps generated by the proposed and benchmark techniques suggest that the first proposed algorithm yielded quite accurate results and obtained improvements as compared to the previous methods, especially to the separate multiscale classification of the individual images coming from COSMO-Skymed and GeoEye-1. Specifically, in the results achieved only through the use of the SAR image, roads were discriminated quite accurately but most other classes were not. In the results obtained through the use only of the optical image, classes that were spatially homogeneous were discriminated more effectively. The proposed technique is able to benefit from both satellite data sources in order to produce a classification output in which most classes in the high-resolution data set are visually well detected. Furthermore, as compared to the case of multisensor but single-scale classification through the algorithm described in (Storvik *et al*. 2009), the proposed method improved in terms of spatial regularity of the classification map. This result is interpreted as a consequence of the contextual modeling components that are integrated in the proposed approach and are due to MRF modeling over a quad-tree and wavelet transformation.

The effectiveness of the results of the proposed technique and the improvement as compared to those achieved by the algorithm in (Storvik *et al*. 2009) are confirmed quantitatively by the classification accuracies on the test samples of the aforementioned classes (see Table 1.1). All considered methods obtained a poor discrimination of the "containers" class, due to its significant overlapping with the "urban" class in the multisensor or multispectral feature space. As a further evolution of the present technique, improvements in the discrimination of this class could be obtained by using texture features. On one hand, accurate performance on the test samples, quite similar to those achieved by the proposed method, were also obtained by the multisensor multiscale technique in (Voisin *et al*. 2014). On the other hand, the proposed algorithm granted improvements over this benchmark approach with regard to the spatial detail in the resulting maps (see Figs. 1.6 (e) and (g)) and to the overall computational burden (see Table 1.1). The latter advantage is due to the fact that addressing multisensor fusion using multiple quad-trees in cascade can leverage on the time efficiency of the sequential recursive formulation of MPM, without requiring the challenging and possibly time-expensive problem of the modeling of the joint pdf of optical and SAR data. On the contrary, this problem is involved in the approaches in (Storvik *et al*. 2009) and (Voisin *et al*. 2014), in which copulas and meta-Gaussian densities are used for this purpose.

Table 1.2. *Second proposed method: classification accuracies on the test set of the time series composed of Pléiades, COSMO-SkyMed (CS), and RADARSAT-2 (RS) images.*

	Class-wise accuracies					overall
	water	urban	vegetation	bare soil	containers	accuracy
Pléiades only	100%	61.66%	81.69%	82.82%	56.72%	76.57%
Pléiades & CS	100%	44.32%	83.54%	74.75%	49.12%	70.34%
Pléiades & RS	92.56%	44.85%	79.85%	78.62%	42.15%	67.60%
Proposed method	90.79%	91.45%	82.59%	81.02%	54.85%	80.14%

1.3.2. *Results of the second method*

Two very-high resolution time series, acquired again over Port-au-Prince, Haiti, were used for experiments with the second proposed method. They both consist of Pléiades pansharpened data at the spatial resolution of 0.5 m (see Figs. 1.7(a) and 1.8(a)), of HH-polarized X-band COSMO-SkyMed spotlight data at the resolution of 1 m (see Fig. 1.7(b) and 1.8(b)), and of HH-polarized C-band RADARSAT-2 ultrafine data with a pixel spacing of 1.56 m (see Fig. 1.7(c) and 1.8(c)). The acquisition dates of the three images in the series were a few days apart from one another. They correspond to two different sites of the Port-au-Prince area, which are shown in Figs. 1.7 and 1.8 and are related to 1000×1000 and 2400×600 pixel grids at the finest resolution, respectively. The main classes in the two scenes are the same as in the previous section. Training and test samples associated with the two sites and annotated by an expert photointerpreter were used to train the second proposed method and to quantitatively measure its performance. The pixel grid at the resolution of 0.5 m of the Pléiades image was used as the reference finest resolution and the RADARSAT-2 image was slightly resampled to a pixel spacing of $4 \cdot 0.5 = 2$ m in order to match the power-of-2 structure associated with the quad-tree (see also Fig. 1.5). Antialiasing filtering was applied within this minor downsampling from 1.56 to 2 m, which is expected to have a negligible impact on the classification output, since the resolution ratio between the original and resampled images is close to 1.

In principle, the second proposed method can be applied in two distinct ways that differ in the ordering of the two SAR data sources in the two quad-trees, i.e., the COSMO-SkyMed image in the first quad-tree and the RADARSAT-2 image in the second one or vice versa. Preliminary experiments, which we omit for brevity, indicated that this choice of order did not have relevant impact on the output classification map.

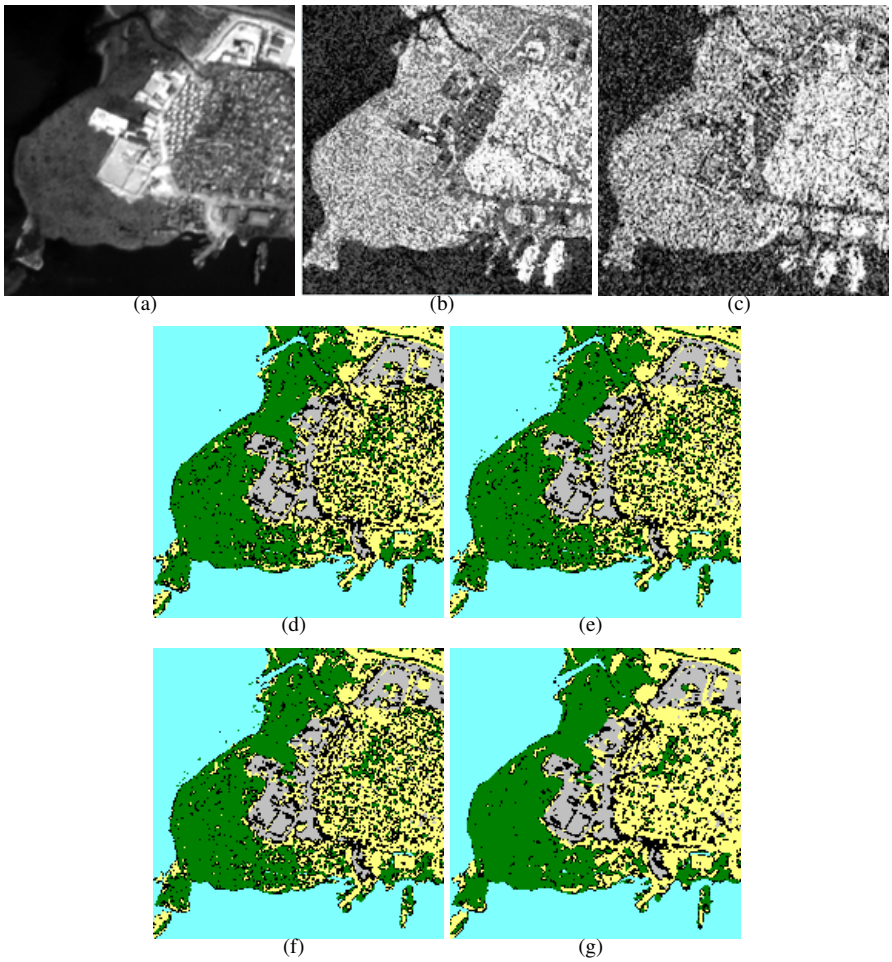

Figure 1.7. *Second proposed method - First test site. (a) Pléiades (©CNES distribution Airbus DS), (b) COSMO-SkyMed (©ASI 2011), and (c) RADARSAT-2 (©CSA 2011) images of the input series. The SAR images are shown after histogram equalization. The R-band of the optical image is displayed. Classification maps obtained by operating with (d) only Pléiades data, (e) Pléiades and COSMO-SkyMed data, (f) Pléiades and RADARSAT-2 data, and (g) the proposed technique with all data in the series. Color legend:*
water ☐ *urban* ☐ *vegetation* ■ *bare soil* ■ *containers* ■ .

Quite accurate performance was obtained on the test samples by the proposed method in the case of the multimission, multifrequency, and multiresolution fusion task addressed in the present experiment (see Table 1.2). The maps obtained from the classification of the compound COSMO-SkyMed / RADARSAT-2 / Pléiades time series of the two sites also exhibited remarkable spatial regularity (see Figs. 1.7(g) and 1.8(g)). In this experiment as well, rather low accuracy was achieved in the case of the "containers" class, again because of its overlapping with the "urban" class in the multisensor feature space.

In order to explore more in detail the capability of the technique to benefit from the synergy of the VNIR optical, X-band radar, and C-band radar imagery in the input series, the aforementioned results were compared to those achieved when (i) only the Pléiades data were used for classification or (ii) the Pléiades data were used in conjunction only with one of the two SAR images (see Table 1.2 and Figs. 1.7 and 1.8(d), (e), and (f)). In all such cases, the same classification scheme based on quad-trees, MPM, and FMM as in the proposed method was applied. We omit the results obtained when only the two SAR images in the series were used because they corresponded to low accuracy values – an expected result in the case of the classification of the aforementioned classes solely with a short series of two SAR images.

The results in Table 1.2 confirm that jointly exploiting all three satellite data sources made it possible to achieve remarkably higher accuracies than when only a subset of these sources was used. When only the Pléiades image was employed, the "water," "vegetation," and "bare soil" classes were discriminated quite accurately but the "urban" class was not. When the second proposed method was applied to these VNIR data together with both the COSMO-SkyMed and the RADARSAT-2 data, the enhancement in the discrimination of the "urban" class was approximately $+30\%$. Furthermore, in this case, the results were more accurate than those generated by jointly using the Pléiades image along with only one of the two SAR images in the series. This scenario suggests the capability of the second proposed method to take benefit from a multimission time series including multifrequency and multiresolution imagery from current satellite instruments at very high spatial resolution. A drawback in the results of the proposed algorithm was the lower accuracy for the "water" class as compared to when only the Pléiades data or Pléiades and COSMO-SkyMed imagery were employed. The "water" class exhibits a significant texture in the RADARSAT-2 image (see Fig. 1.7(c)), and the proposed algorithm does not involve any texture analysis component. The impact of this issue is limited, as the "water" class is discriminated by the proposed algorithm with an accuracy of around 91%. However, extending the method by integrating texture extraction appear as a promising possible generalization.

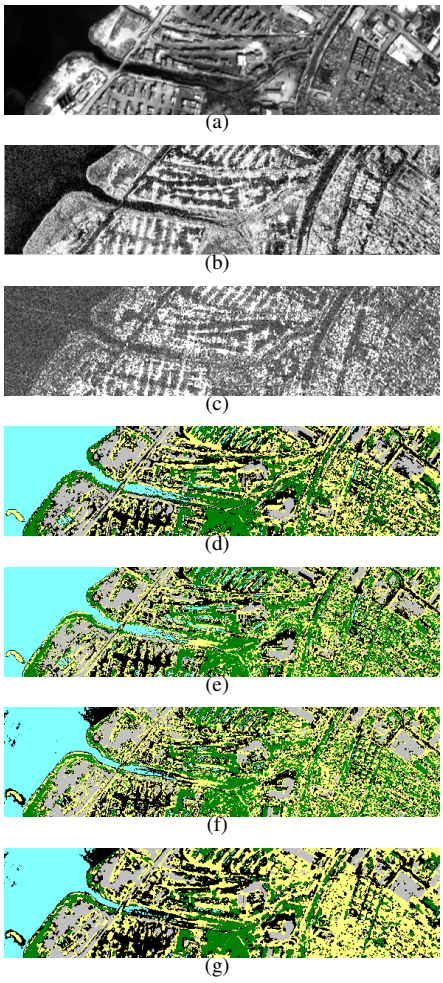

Figure 1.8. *Second proposed method - Second test site. (a) Pléiades (©CNES distribution Airbus DS), (b) COSMO-SkyMed (©ASI 2011), and (c) RADARSAT-2 (©CSA 2011) images of the input series. The SAR images are shown after histogram equalization. The R-band of the optical image is displayed. Classification maps obtained by operating with (d) only Pléiades data, (e) Pléiades and COSMO-SkyMed data, (f) Pléiades and RADARSAT-2 data, and (g) the proposed technique with all data in the series. Color legend:*
water ☐ *urban* ☐ *vegetation* ■ *bare soil* ■ *containers* ■

1.4. Conclusion

In this chapter, the problem of the generation of a classification map from an input time series composed of multisensor multiresolution remote sensing images has been discussed. First, the related literature in the area of remote sensing data fusion has been reviewed. Then, an advanced approach based on multiple quad-trees in cascade has been described. It derives from the multisensor generalization of a previous technique focused on time series of single-sensor data, and addresses the challenging problem of multisensor, multifrequency, and multiresolution fusion for classification purposes.

In the framework of this approach, two algorithms have been developed for two different multimodal fusion objectives. In the first one, the general task of jointly classifying a time series of multisensor multiresolution imagery is considered. In the second one, the focus is on the special case of the fusion of multimission data acquired by COSMO-SkyMed, RADARSAT-2, and Pléiades. In the case of both techniques, the fusion task is formalized in the framework of hierarchical probabilistic graphical models – most remarkably hierarchical Markov random fields on cascaded quad-trees. Inference and parameter estimation are addressed through the maximum posterior mode criterion and finite mixture models, respectively.

Examples of experimental results provided by the two proposed algorithms have been shown with regard to high or very high resolution time series associated with case studies in Port-au-Prince, Haiti. The results have suggested that the described algorithms successfully benefit from the data sources in the input multisensor time series, improving the classification result as compared to those obtained using single-mission, single-scale, or previous methods in terms of classification accuracy, computation time, or spatial regularity.

A major property of the proposed hierarchical Markovian framework is its flexibility. The graphical architecture associated with multiple quad-trees in cascade allows incorporating input image sources associated with different sensors, acquisition times, and spatial resolutions – jointly. Relevant extensions of this framework may involve the combination with spatial-contextual models within each layer of the quad-trees or with the intrinsically multiscale structure of convolutional neural networks (Goodfellow *et al*. 2016).

1.5. Acknowledgments

The authors would like to thank the Italian Space Agency (ASI), the French Space Agency (CNES), and the Canadian Space Agency (CSA) for providing COSMO-SkyMed, Pléiades, and RADARSAT-2 data, respectively. The COSMO-SkyMed and RADARSAT-2 images were procured in the context of the SOAR-ASI 5245 project framed within the joint ASI-CSA announcement of opportunity.

1.6. Bibliography

ALEjaily, A., El Rube, I., Mangoud, M. (2008), Fusion of remote sensing images using contourlet transform, *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering*, pp. 213–218.

Basseville, M., Benveniste, A., Willsky, A. (1992*a*), Multiscale autoregressive processes. I. Schur-Levinson parametrizations, *IEEE Transactions on Signal Processing*, 40(8), 1915–1934.

Basseville, M., Benveniste, A., Willsky, A. (1992*b*), Multiscale autoregressive processes. II. lattice structures for whitening and modeling, *IEEE Transactions on Signal Processing*, 40(8), 1935–1954.

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G., Dupuy, S. (2018), M3fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12), 4939–4949.

Benz, U. (1999), Supervised fuzzy analysis of single-and multichannel SAR data, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 1023–1037.

Berthod, M., Kato, Z., Yu, S., Zerubia, J. (1996), Bayesian image classification using Markov random fields, *Image and Vision Computing*, 14(4), 285–295.

Bouman, C. (1991), *A multiscale image model for Bayesian image segmentation*, Purdue University, School of Electrical Engineering.

Brunner, D., Lemoine, G., Bruzzone, L. (2010), Earthquake damage assessment of buildings using VHR optical and SAR imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2403–2420.

Bruzzone, L., Prieto, D. F., Serpico, S. (1999), A neural-statistical approach to multitemporal and multisource remote-sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), 1350–1359.

Burt, P. (1984), *The pyramid as a structure for efficient computation*, Springer.

Carvalho, M., Le Saux, B., Trouvé-Peloux, P., Champagnat, F., Almansa, A. (2019), Multitask learning of height and semantics from aerial images, *IEEE Geoscience and Remote Sensing Letters*, .

Celeux, G., Chauveau, D., Diebolt, J. (1996), Stochastic versions of the EM algorithm: an experimental study in the mixture case, *Journal of Statistical Computation and Simulation*, 55(4), 287–314.

Chardin, A. (2000), Modeles energetiques hierarchiques pour la resolution des problemes inverses en analyse d'images, application a la teledetection, PhD thesis, University of Rennes 1, France.

Chen, Y., Li, C., Ghamisi, P., Jia, X., Gu, Y. (2017), Deep fusion of remote sensing data for accurate classification, *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1253–1257.

Cheng, X., Zheng, Y., Zhang, J., Yang, Z. (2020), Multi-task multi-source deep correlation filter for remote sensing data fusion, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, .

Chibani, Y., Houacine, A. (2003), Redundant versus orthogonal wavelet decomposition for multisensor image fusion, *Pattern Recognition*, 36(4), 879–887.

Choi, M., Kim, R., Nam, M.-R., Kim, H. (2005), Fusion of multispectral and panchromatic satellite images using the curvelet transform, *IEEE Geoscience and Remote Sensing Letters*, 2(2), 136–140.

Demirel, H., Anbarjafari, G. (2010), Satellite image resolution enhancement using complex wavelet transform, *IEEE Geoscience and Remote Sensing Letters*, 7(1), 123–126.

Dousset, B., Gourmelon, F. (2003), Satellite multi-sensor data analysis of urban surface temperatures and landcover, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1), 43–54.

Forster, B., Van De Ville, D., Berent, J., Sage, D., Unser, M. (2004), Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images, *Microscopy Research and Technique*, 65(1-2), 33–42.

Fukunaga, K. (2013), *Introduction to statistical pattern recognition*, Academic press.

Gamba, P., Lisini, G., Tomás, L., Almeida, C., Fonseca, L. (2011), Joint VHR-LIDAR classification framework in urban areas using a priori knowledge and post processing shape optimization, *in* IEEE Urban Remote Sensing Event (JURSE), pp. 93–96.

Ghamisi, P., Höfle, B., Zhu, X. X. (2016), Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6), 3011–3024.

Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G. (2015), Multimodal classification of remote sensing images: a review and future directions, *Proceedings of the IEEE*, 103(9), 1560–1584.

Goodfellow, I., Bengio, Y., Courville, A. (2016), *Deep Learning*, MIT Press.

Graffigne, C., Heitz, F., Perez, P., Preteux, F., Sigelle, M., Zerubia, J. (1995), Hierarchical Markov random field models applied to image analysis: a review, *in* International Symposium on Optical Science, Engineering, and Instrumentation (SPIE), International Society for Optics and Photonics, pp. 2–17.

Hall, D., Llinas, J. (2001), *Multisensor data fusion*, CRC press.

Hedhli, I., Moser, G., Serpico, S., Zerubia, J. (2015), New hierarchical joint classification method of SAR-optical multiresolution remote sensing data, *in* IEEE European Signal Processing Conference.

Hedhli, I., Moser, G., Zerubia, J., Serpico, S. (2014), New cascade model for hierarchical joint classification of multitemporal, multiresolution and multisensor remote sensing data, *in* IEEE International Conference on Image Processing (ICIP),

2014.

Hedhli, I., Moser, G., Zerubia, J., Serpico, S. B. (2016), A New Cascade Model for the Hierarchical Joint Classification of Multitemporal and Multiresolution Remote Sensing Data , *IEEE Transactions on Geoscience and Remote Sensing*, 54(11), 6333–6348.

Huang, B., Song, H. (2012), Spatiotemporal reflectance fusion via sparse representation, *IEEE Transactions on Geoscience and Remote Sensing*, 50(10), 3707–3716.

Iqbal, M., Ghafoor, A., Siddiqui, A. (2013), Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means, *IEEE Geoscience and Remote Sensing Letters*, 10(3), 451–455.

Kato, Z., Zerubia, J. (2012), *Markov random fields in image segmentation*, Now Pub.

Krylov, V., Moser, G., Serpico, S., Zerubia, J. (2013), On the method of logarithmic cumulants for parametric probability density function estimation, *IEEE Transactions on Image Processing*, 22(10), 3791–3806.

Laferté, J.-M., Pérez, P., Heitz, F. (2000), Discrete Markov image modeling and inference on the quadtree, *IEEE Transactions on Image Processing*, 9(3), 390–404.

Landgrebe, D. (2005), *Signal theory methods in multispectral remote sensing*, vol. 29, John Wiley & Sons.

Li, H.-C., Hong, W., Wu, Y.-R., Fan, P.-Z. (2011), On the empirical-statistical modeling of SAR images with generalized Gamma distribution, *IEEE Journal of Selected Topics in Signal Processing*, 5(3), 386–397.

Li, H., Manjunath, B., Mitra, S. (1995), Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing*, 57(3), 235–245.

Li, S. (2009), *Markov random field modeling in image analysis*, Springer Science & Business Media.

Mallat, S. (2008), *A wavelet tour of signal processing*, 3rd edition, Academic press.

Mas, J., Flores, J. (2008), The application of artificial neural networks to the analysis of remotely sensed data, *International Journal of Remote Sensing*, 29(3), 617–663.

Minh, D. H. T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P. (2018), Deep recurrent neural networks for winter vegetation quality mapping via multitemporal sar sentinel-1, *IEEE Geoscience and Remote Sensing Letters*, 15(3), 464–468.

Moser, G., Serpico, S. (2009), Unsupervised change detection from multichannel SAR data by Markovian data fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2114–2128.

Muñoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., Camp-Valls, G. (2010), Semisupervised one-class support vector machines for classification of remote

sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 48(8), 3188–3197.

Nelson, J. D. B., Gibberd, A. J., Nafornita, C., Kingsbury, N. (2018), The locally stationary dual-tree complex wavelet model, *Statistics and Computing*, 28(6), 1139–1154.

Nencini, F., Garzelli, A., Baronti, S., Alparone, L. (2007), Remote sensing image fusion using the curvelet transform, *Information Fusion*, 8(2), 143–156.

Nguyen, N., Nasrabadi, N., Tran, T. (2011), Robust multi-sensor classification via joint sparse representation, *in* International Conference on Information Fusion.

Pan, S., Wu, J., Zhu, X., Zhang, C., Philip, S. Y. (2015), Joint structure feature exploration and regularization for multi-task graph classification, *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 715–728.

Pérez, P. (1993), Champs markoviens et analyse multirésolution de l'image: application à l'analyse du mouvement, PhD thesis, University of Rennes 1, France.

Piella, G. (2003), Adaptive wavelets and their applications to image fusion and compression, PhD thesis, University of Amsterdam.

Pohl, C., van Genderen, J. (2014), Remote sensing image fusion: an update in the context of digital earth, *International Journal of Digital Earth*, 7(2), 158–172.

Pohl, C.and Van Genderen, J. (1998), Review article – Multisensor image fusion in remote sensing: concepts, methods and applications, *International Journal of Remote Sensing*, 19(5), 823–854.

Prendes, J. (2015), New statistical modeling of multi-sensor images with application to change detection, PhD thesis, Université Paris-Sud, France.

Roberts, J., Van Aardt, J., Ahmed, F. (2008), Assessment of image fusion procedures using entropy, image quality, and multispectral classification, *Journal of Applied Remote Sensing*, 2(1).

Rockinger, O. (1996), Pixel-level fusion of image sequences using wavelet frames, *in* Proceedings of the 16th Leeds applied shape research workshop, Leeds University Press, Citeseer.

Serpico, S., Dellepiane, S., Boni, G., Moser, G., Angiati, E., Rudari, R. (2012), Information extraction from remote sensing images for flood monitoring and damage evaluation, *Proceedings of the IEEE*, 100(10), 2946–2970.

Shah, V., Younan, N., King, R. (2008), An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets, *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1323–1335.

Storvik, B., Storvik, G., Fjortoft, R. (2009), On the combination of multisensor data using meta-Gaussian distributions, *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2372–2379.

Stroppiana, D., Azar, R., Calo, F., Pepe, A., Imperatore, P., Boschetti, M., Silva, J., Brivio, P., Lanari, R. (2015), Remote sensing of burned area: A fuzzy-based framework for joint processing of optical and microwave data, *in* IEEE, Geoscience and Remote Sensing Symposium (IGARSS), pp. 1409–1412.

Ulaby, F., Long, D. G. (2015), *Microwave radar and radiometric remote sensing*, Artech House.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G., Restaino, R., Wald, L. (2015), A critical comparison among pansharpening algorithms, *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2565–2586.

Voisin, A. (2012), Classification supervisée d'images d'observation de la Terre à haute résolution par utilisation de méthodes markoviennes, PhD thesis, University of Nice Sophia Antipolis, France.

Voisin, A., Krylov, V., Moser, G., Serpico, S., Zerubia, J. (2012), Multichannel hierarchical image classification using multivariate copulas, *in* IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics.

Voisin, A., Krylov, V., Moser, G., Serpico, S., Zerubia, J. (2014), Supervised Classification of Multi-sensor and Multi-resolution Remote Sensing Images with a Hierarchical Copula-based Approach, *IEEE Transactions on Geoscience and Remote Sensing*, 52(6), 3346–3358.

Wald, L. (1999), Some terms of reference in data fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), 1190–1193.

Waltz, E., Llinas, J. *et al*. (1990), *Multisensor data fusion*, vol. 685, Artech house Boston.

Wang, D., Liang, S. (2014), Improving LAI mapping by integrating MODIS and CYCLOPES LAI products using optimal interpolation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(2), 445–457.

Waske, B., van der Linden, S. (2008), Classifying multilevel imagery from SAR and optical sensors by decision fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1457–1466.

Willsky, A. (2002), Multiresolution Markov models for signal and image processing, *Proceedings of the IEEE*, 90(8), 1396–1458.

Yang, B., Li, S., Sun, F. (2007), Image fusion using nonsubsampled contourlet transform, *in* IEEE International Conference on Image and Graphics, pp. 719–724.

Zhang, G., Kingsbury, N. (2015), Variational bayesian image restoration with group-sparse modeling of wavelet coefficients, *Digital Signal Processing: A Review Journal*, 47, 157–168.

Zhang, Y., Hong, G. (2005), An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images, *Information Fusion*, 6(3), 225–234.