HAL open science

A causal hierarchical Markov framework for the classification of multiresolution and multisensor remote sensing images

Alessandro Montaldo, Luca Fronda, Ihsen Hedhli, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia

To cite this version:

Alessandro Montaldo, Luca Fronda, Ihsen Hedhli, Gabriele Moser, Sebastiano B Serpico, et al.. A causal hierarchical Markov framework for the classification of multiresolution and multisensor remote sensing images. ISPRS 2020 - XXIV International Society of Photogrammetry and Remote Sensing Congress, Aug 2020, Nice / Virtual, France. pp.269-277, 10.5194/isprs-annals-V-3-2020-269-2020 . hal-02982420

HAL Id: hal-02982420

https://inria.hal.science/hal-02982420v1

Submitted on 28 Oct 2020

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.

A CAUSAL HIERARCHICAL MARKOV FRAMEWORK FOR THE CLASSIFICATION OF MULTIRESOLUTION AND MULTISENSOR REMOTE SENSING IMAGES

Alessandro Montaldo[1], Luca Fronda[1], Ihsen Hedhli[2], Gabriele Moser[1], Sebastiano B. Serpico[1], and Josiane Zerubia[3]

[1] University of Genoa, Italy, gabriele.moser@unige.it. [2] Université Laval, Quebec, Canada
[3] Inria, Université Côte d'Azur, Sophia Antipolis, France, josiane.zerubia@inria.fr

Commission III, WG III/6

KEY WORDS: Multiresolution and multisensor fusion, causality, hierarchical Markov random models, Markov chain, semantic image segmentation.

ABSTRACT:

In this paper, a multiscale Markov framework is proposed in order to address the problem of the classification of multiresolution and multisensor remotely sensed data. The proposed framework makes use of a quadtree to model the interactions across different spatial resolutions and a Markov model with respect to a generic total order relation to deal with contextual information at each scale in order to favor applicability to very high resolution imagery. The methodological properties of the proposed hierarchical framework are investigated. Firstly, we prove the causality of the overall proposed model, a particularly advantageous property in terms of computational cost of the inference. Secondly, we prove the expression of the marginal posterior mode criterion for inference on the proposed framework. Within this framework, a specific algorithm is formulated by defining, within each layer of the quadtree, a Markov chain model with respect to a pixel scan that combines both a zig-zag trajectory and a Hilbert space-filling curve. Data collected by distinct sensors at the same spatial resolution are fused through gradient boosted regression trees. The developed algorithm was experimentally validated with two very high resolution datasets including multispectral, panchromatic and radar satellite images. The experimental results confirm the effectiveness of the proposed algorithm as compared to previous techniques based on alternate approaches to multiresolution fusion.

1. INTRODUCTION

Thanks to the potential offered by current and forthcoming space missions, very high spatial resolution (VHR) optical (e.g., Pléiades, WorldView-3, SPOT-6/7) and synthetic aperture radar (SAR; e.g., COSMO-SkyMed Seconda Generazione, TerraSAR-X, RADARSAT-2) instruments are now available to serve many applications. From the viewpoint of image analysis, the joint exploitation of the resulting data requires novel methods that operate with images collected by multiple sensors on the same area at multiple spatial resolutions. This is highly promising because it allows to benefit from data associated with different physical natures, frequencies, polarizations, etc., and from the tradeoff between a synoptic view at coarser resolutions and the spatial detail of finer resolutions. In this paper, the semantic segmentation (Wang, 2016, Arnab et al., 2018) or dense supervised classification of images that are both multiresolution and multisensor is addressed. A major challenge in this joint multisensor-multiresolution scenario is the combination of heterogeneous statistics of the input images and of the need to characterize spatial information associated with different resolutions (Hedhli et al., 2017). Trivial well-known solutions mostly use resampling procedures and do not attempt to capture the multiresolution structure of the data explicitly.

Here, an approach to address multiresolution fusion is proposed based on hierarchical latent Markov modeling (Li, 2009). Within the family of structured output learning methods, Markov models postulated on planar or multilayer graphs are known as flexible and powerful stochastic models for spatial and possibly multimodal information (Li, 2009). A common shortcoming of generic Markov random fields (MRFs) is their non-causality that generally leads to possibly time-consuming inference algorithms. The resulting non-convex problems are addressed with stochastic optimizers such as simulated annealing (Li, 2009) or graph-theoretic algorithms (Boykov et al., 2001). In opposition to the case of classical MRFs, for causal probabilistic 1D models (e.g., Markov chains) efficient sequential techniques (e.g., the Baum-Welch algorithm (Baggenstoss, 2001)) are widely used for inference. Indeed, an effective approach to multiresolution and multisensor fusion addressed with Markov modeling is to postulate Markovianity on a hierarchical graph structure (e.g., systems of quadtrees (Hedhli et al., 2017)). However, despite hierarchical MRFs on quadtrees capture relations among sites located at different scales through the use of a Markov chain, a common limitation is that they do not explicitly characterize spatial dependences within the layer at each resolution (Laferté et al., 2000).

To join the benefits of both hierarchical and planar MRFs, the classification of multiresolution and multisensor remote sensing data is formulated in this paper through a causal hierarchical Markov framework. The contribution of the paper is twofold. First, we extend the approach developed in (Montaldo et al., 2019a, Montaldo et al., 2019b), in which hierarchical MRFs on quadtrees have been combined with planar Markov meshes and with various pixelwise probabilistic models, to a broader framework in which each layer of the quadtree is associated with a Markov model with respect to an arbitrary total order relation. The resulting framework is highly general and includes the Markov meshes in (Montaldo et al., 2019a, Montaldo et al., 2019b) as a special case. Secondly, within this generalized framework, we develop a novel algorithm that is based on a Markov chain model (Fjortoft et al., 2003) on each layer of the quadtree. In this model, Markov chains are formulated both across the scales of the tree and with respect to a 1D scan of the

pixel lattice of each layer. This joint strategy benefits from the spatial information within each layer and inherently supports multiresolution fusion. A case-specific scan is defined on each layer of the quadtree by integrating a simple zig-zag trajectory and a Hilbert space-filling curve (Abend et al., 1965). Methodologically, we prove that the proposed generalized framework ensures causality for the whole probabilistic graphical model, thus favoring efficient inference. Then, we derive the analytical formulation of the marginal posterior mode (MPM) inference algorithm for the proposed framework and we customize it to the special case of the aforementioned Markov chain. MPM is especially advantageous for classification and segmentation in a multiresolution scenario (Laferté et al., 2000) and its formulation for causal hierarchical Markov models explicitly relies on the causality of the models. In the proposed framework, the images coming from the various sensors are inserted in a quadtree based on their resolutions. On one hand, when distinct sensors provide data at different resolutions, multisensor fusion is accomplished naturally thanks to the adopted hierarchical topology. On the other hand, when data sources from distinct sensors are available at the same resolution, multisensor fusion is addressed by incorporating a nonparametric ensemble method within the proposed algorithm. Similar to (Montaldo et al., 2019a, Montaldo et al., 2019b), gradient boosting regression trees (GBRT) (Friedman, 2001) are employed in this role.

2. METHODOLOGY

2.1 Causal hierarchical Markov models

Among the main current trends of probabilistic graphical models in remote sensing, multiscale and multiresolution models are of primary importance. The rationale is that complementary information is appreciated at different spatial scales. While fine-scale observations provide lots of geometrical detail but are spatially heterogeneous and noise-sensitive, coarse-scale observations capture only large image regions and covers but with strong robustness to noise and outliers. Many multiscale or multiresolution graphical models have been introduced on the basis of hierarchical graphs (Willsky, 2002), such as quadtrees (Laferté et al., 2000). With this topology, hierarchical MRF models typically involve Markovianity along the scale axis. However, a limitation of a hierarchical MRF on a quadtree is that, while it captures relations among data at different scales through a Markov chain, it generally does not characterize spatial dependencies within the layer at each scale (Laferté et al., 2000). Here, a causal hierarchical framework that postulates Markovianity not only among pixels belonging to pixel lattices with different resolutions but also among pixels in the same lattice is defined. This framework describes both cross-layer (multiscale) and intra-layer (spatial) dependence relations.

2.1.1 Planar causal Markov models Let $R \subset \mathbb{Z}^2$ be a planar rectangular pixel lattice and let us define a total order relation $\trianglelefteq$ on R. This means that this relation fulfills the following properties ($r, s, t \in R$):

Antisymmetry:	$r \trianglelefteq s,$	$s \trianglelefteq r$	$\Longrightarrow$	$r = s$
Transitivity:	$r \trianglelefteq s,$	$s \trianglelefteq t$	$\Longrightarrow$	$r \trianglelefteq t$
Connexity:	at least one holds between $r \trianglelefteq s$ and $s \trianglelefteq r$.			

We shall also write $r \triangleleft s$ to indicate that $r \trianglelefteq s$ and $r \neq s$. Accordingly, the set $\{r \in R : r \triangleleft s\}$ formalizes the idea

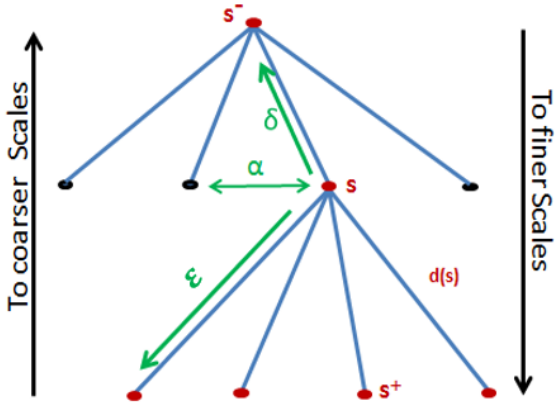

Figure 1. Quadtree structure and notations on the tree

of the "past" of site $s \in R$. Let us consider also a neighborhood relation to be defined in R consistently with the order relation $\triangleleft$, so that, intuitively, it makes sense to speak of the "past neighbors" or "recent past" of each site. We write $r \blacktriangleleft s$ to indicate that r is one of such past neighbors of s. Formally, $\{r \in R : r \blacktriangleleft s\} \subsetneq \{r \in R : r \triangleleft s\}$, i.e., the past neighbors of $s \in R$ are all included in the past of s but form a proper (strict) subset of its entire past. Normally, this subset is meant to be small to favor low computational burden in the inference.

Let a discrete random variable x_s be attached to each site $s \in R$ and let $\mathcal{X} = \{x_s\}_{s \in R}$ be the corresponding random field. A Markovianity property restricted to the past of each site holds for $\mathcal{X}$ if (Abend et al., 1965, Devijver, 1993, Willsky, 2002):

$$P(x_s | x_r, r \triangleleft s) = P(x_s | x_r, r \blacktriangleleft s) \qquad \forall s \in S. \tag{1}$$

This notion of causal Markovianity on a planar lattice extends the one discussed in (Montaldo et al., 2019a, Montaldo et al., 2019b), in which the order relation was referred to a 2D mesh. We also recall that it can be proved for several planar causal Markov models (including Markov meshes of second and third order and Markov chains) that the following factorization holds (Abend et al., 1965):

$$P(\mathcal{X}) = \prod_{s \in R} P(x_s | x_r, r \blacktriangleleft s), \tag{2}$$

up to properly defining the behavior of $\mathcal{X}$ at the image borders.

2.1.2 Hierarchical causal Markov random fields Let now $\{S^0, S^1, \ldots, S^L\}$ be a collection of pixel lattices organized as a quadtree, as shown in Fig. 1. This hierarchical structure is arranged so that the width and the height of $S^{\ell-1}$ are twice smaller than those of S^ℓ ($\ell = 1, 2, \ldots, L$). If $\ell \in \{1, 2, \ldots, L\}$, each site $s \in S^\ell$ in layer S^ℓ has a well-defined parent site $s^- \in S^{\ell-1}$. If $\ell \in \{0, 1, \ldots, L-1\}$, each site $s \in S^\ell$ has in $S^{\ell+1}$ four well-defined children sites, which we collect in the set $s^+ \subset S^{\ell+1}$. These definitions imply a hierarchy on the whole tree $S = S^0 \cup S^1 \cup \ldots \cup S^L$ from the root S^0 to the leaves S^L (Laferté et al., 2000). If again a discrete random variable x_s is associated with each $s \in S$, then $\mathcal{X} = \{x_s\}_{s \in S}$ is a hierarchical MRF on the quadtree if the following property holds for all

layers but the root ($\ell = 1, 2, \ldots, L$) (Laferté et al., 2000):

$$P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1},\mathcal{X}^{\ell-2},\ldots,\mathcal{X}^0) = \\ = P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}) = \prod_{s\in S^\ell} P(x_s|x_{s^-}), \quad (3)$$

where $\mathcal{X}^\ell = \{x_s\}_{s\in S^\ell}$ ($\ell = 0, 1, \ldots, L$). The first equality in (3) indicates that Markovianity is valid across the layers of the quadtree. The second equality expresses that the transition probability $P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1})$ from the $(\ell-1)$th to the ℓth layer factorizes on a pixelwise basis. Accordingly, a hierarchical MRF on a quadtree describes the dependence along the scale axis but not the spatial-contextual dependence associated with the random field $\mathcal{X}^\ell$ that is supported on each layer S^ℓ.

2.2 The proposed framework

2.2.1 Model assumptions In this paper, we formulate a general framework for the joint fusion of multisensor and multiresolution images in a supervised classification scenario. The key-idea of the proposed framework is to generalize the approaches in (Montaldo et al., 2019a, Montaldo et al., 2019b) by introducing a hierarchical Markov model with respect to a quadtree topology and to an arbitrary total order relation on each layer of the tree. Considering a set of well-registered images acquired by distinct VHR sensors at different spatial resolutions on the same area, each image is included in a corresponding layer of the quadtree. Hence, the resolutions of the input images should be in a power-of-2 relation. This is a restriction, but given the resolutions of current VHR sensors, it can be easily met up to minor resampling (Dikshit, Roy, 1996, Inglada et al., 2007) and appropriate antialiasing filtering (Alparone et al., 2008, Mallat, 2009). The hierarchical model naturally fits the requirements of multiresolution fusion. Furthermore, the relations $\lhd, \unlhd$, and $\blacktriangleleft$ are assumed to be defined within each planar layer of the quadtree. From a graphical viewpoint, this implies that every site $s \in S^\ell$ of every layer S^ℓ but the root ($\ell = 1, 2, \ldots, L$) is linked to one parent $s^- \in S^{\ell-1}$ located in the upper layer and to the past neighbors $\{r \in S^\ell : r \blacktriangleleft s\}$ located in its own layer. A pixel in the root S^0 has no parent and only the causal neighbors, if any.

With the same notations as above, x_s is now meant as the class label of site $s \in S$ and a finite set Ω of classes is assumed to be defined by a training map ($x_s \in \Omega, s \in S$). For each $s \in S$, let $\mathcal{C}_s$ be the vector collecting the labels of all nodes linked to s (the "context" of s), i.e., $\mathcal{C}_s = \{s^-\} \cup \{r \in S^\ell : r \blacktriangleleft s\}$ if s is not in the root ($s \in S \setminus S^0$), and $\mathcal{C}_s = \{r \in S^0 : r \blacktriangleleft s\}$ if $s \in S^0$ is in the root. As the input multisensor and multiresolution images are framed in the quadtree, each $s \in S$ is also associated with a feature vector y_s. Thus, a random field $\mathcal{Y} = \{y_s\}_{s\in S}$ of observations is defined. If a site s is not in the leaf layer ($s \in S \setminus S^L$), then we shall denote as $\mathcal{D}_s$ the vector collecting the feature vectors of all sites that descend from $s \in S$ in the quadtree. If s in the leaf layer ($s \in S^L$), we set $\mathcal{D}_s = y_s$. The proposed framework is defined by the following assumptions:

1. $\mathcal{X}$ is Markovian across the scales as in (3) and the following proportionality and factorization hold on all layers except the root ($\ell = 1, 2, \ldots, L$):

$$P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1},\mathcal{X}^{\ell-2},\ldots,\mathcal{X}^0) = P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1}) \propto \\ \propto \prod_{s\in S^\ell} P(x_s|x_r, r \blacktriangleleft s)P(x_s|x_{s^-}). \quad (4)$$

2. On the root, $\mathcal{X}^0$ satisfies the factorization in (2):

$$P(\mathcal{X}^0) = \prod_{s\in S^0} P(x_s|x_r, r \blacktriangleleft s). \quad (5)$$

3. Conditional independence holds for the observations given the class labels:

$$P(\mathcal{Y}|\mathcal{X}) = \prod_{s\in S} P(y_s|x_s) = \prod_{\ell=0}^{L}\prod_{s\in S^\ell} P(y_s|x_s). \quad (6)$$

4. The label of each site, given the labels of the sites linked to it, only depends on the observations of its descendants and not on those of the other sites:

$$P(x_s|\mathcal{C}_s, \mathcal{Y}) = P(x_s|\mathcal{C}_s, \mathcal{D}_s) \qquad \forall s \in S. \quad (7)$$

5. Given the observation field, the parent and neighboring labels of each site are conditionally independent:

$$P(\mathcal{C}_s|\mathcal{Y}) = \begin{cases} P(x_{s^-}|\mathcal{Y}) \prod_{r\blacktriangleleft s} P(x_r|\mathcal{Y}) & \forall s \in S \setminus S^0 \\ \prod_{r\blacktriangleleft s} P(x_r|\mathcal{Y}) & \forall s \in S^0 \end{cases} \quad (8)$$

6. The parent and neighboring labels of each site, when conditioned to the label of that site, are independent on the observations of its descendants and mutually independent:

$$P(\mathcal{C}_s|x_s, \mathcal{D}_s) = P(\mathcal{C}_s|x_s) = \\ = \begin{cases} P(x_{s^-}|x_s) \prod_{r\blacktriangleleft s} P(x_r|x_s) & \forall s \in S \setminus S^0 \\ \prod_{r\blacktriangleleft s} P(x_r|x_s) & \forall s \in S^0 \end{cases} \quad (9)$$

These assumptions play different roles within the proposed framework. On one hand, methodologically, the framework is defined by Assumptions 1, 2, and 3. Assumption 1 expresses the key idea of the proposed combination of a hierarchical MRF on a quadtree and a causal planar Markov model with respect to a generic order relation $\trianglelefteq$. It implies that Markovianity holds across the scales and that the corresponding transition probabilities factorize so that the labels in each layer convey both contextual and multiscale dependencies. Accordingly, the model can characterize both multiresolution and spatial information. From this perspective, Assumption 2 is necessary for completeness, as it ensures that the analogous spatial factorization holds on the root as well. The conditional independence in Assumption 3 is a commonly accepted statement in the application of MRFs to image classification and is used in the proposed framework to favor tractability. On the other hand, Assumptions 4, 5, and 6 are not meant to formalize specifically desired behaviors of the proposed model. As we shall argue in more detail in the next section, they are basically technical conditions that favor the analytical tractability of the inference. They are similar in nature to conditional independence hypotheses that are often stated for mathematical convenience when operating with either hierarchical (Laferté et al., 2000) or planar MRFs (Li, 2009).

Regarding inference within the proposed framework, we use the MPM criterion, i.e., $s \in S$ is assigned to the class $\omega \in \Omega$ that maximizes $P\{x_s = \omega|\mathcal{Y}\}$ (Li, 2009). MPM is especially appropriate for hierarchical MRFs because it penalizes errors according to the scale on which they are made, which is desirable to avoid error accumulation along the layers (Laferté et al.,

2000). We shall prove in the next section that the proposed framework is causal both spatially and across scales, which makes it possible to compute $P(x_s|\mathcal{Y})$ on each $s \in S$ recursively and efficiently. As data associated with different resolutions are associated with the observations in distinct layers of the quadtree, this recursive inference inherently accomplishes their fusion, thus jointly classifying multisensor data supported at different resolutions. If two or more considered input sensors correspond to the same resolution, then their data contribute to the feature vectors in the same layer (stacked vector). In this case, multisensor fusion is obtained by integrating the GBRT ensemble of decision trees (Friedman, 2001) in the MPM scheme (see Section 2.3). The flexibility of tree ensembles and their applicability to highly heterogeneous input feature vectors allows the proposed framework to address multisensor fusion in this case as well.

2.3 Methodological properties and MPM inference

Here, we prove the causality of the proposed hierarchical Markov framework and we derive a set of relations that allows computing the MPM functional $P(x_s|\mathcal{Y})$ of each site $s \in S$. Operatively, the causality is expressed in terms of the factorization of the joint distribution of all observations and labels in terms of causal transition probabilities.

Theorem 1. *Under Assumptions 1-3, the joint distribution $P(\mathcal{X}, \mathcal{Y})$ of all class labels and feature vectors in the quad-tree is entirely defined by the parent-child transition probabilities $\{P(x_s|x_{s^-})\}_{s \in S \setminus S^0}$, the past neighbor transition probabilities $\{P(x_s|x_r, r \blacktriangleleft s)\}_{s \in S}$, and the pixelwise data conditional likelihoods $\{P(y_s|x_s)\}_{s \in S}$.*

Proof. According to (3), Assumption 1 implies that:

$$P(\mathcal{X}, \mathcal{Y}) = P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}) = P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L, \mathcal{X}^{L-1}, \ldots, \mathcal{X}^0)$$
$$= P(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L|\mathcal{X}^{L-1}) \ldots P(\mathcal{X}^1|\mathcal{X}^0)P(\mathcal{X}^0). \quad (10)$$

Plugging the factorizations (6) for $P(\mathcal{Y}|\mathcal{X})$, (5) for $P(\mathcal{X}^0)$, and (4) for $P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1})$ ($\ell = 1, 2, \ldots, L$) into (10) yields:

$$P(\mathcal{X}, \mathcal{Y}) \propto \prod_{\ell=1}^{L} \prod_{s \in S^\ell} P(x_s|x_r, r \blacktriangleleft s)P(x_s|x_{s^-})P(y_s|x_s) \cdot$$
$$\cdot \prod_{s \in S^0} P(x_s|x_r, r \blacktriangleleft s)P(y_s|x_s). \quad (11)$$

Therefore, $P(\mathcal{X}, \mathcal{Y})$ is entirely defined by the parent-child transition probabilities $P(x_s|x_{s^-})$, the past neighbor transition probabilities $P(x_s|x_r, r \blacktriangleleft s)$ and the data conditional likelihoods $P(y_s|x_s)$. □

Theorem 2. *Under Assumptions 1-6, for each site $s \in S \setminus S^0$ not in the root* [1]*:*

$$P(x_s) = \sum_{x_{s^-}} P(x_s|x_{s^-})P(x_{s^-}), \quad (12)$$

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_s|x_{s^-})P(x_{s^-})P(x_s)^{-|\mathcal{C}_s|} \cdot$$
$$\cdot \prod_{r \blacktriangleleft s} P(x_s|x_r)P(x_r), \quad (13)$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{D}_s)P(x_{s^-}|\mathcal{Y}) \prod_{r \blacktriangleleft s} P(x_r|\mathcal{Y}). \quad (14)$$

[1] We denote as $|A|$ the cardinality (i.e., the number of elements) of a finite set A.

For each root site $s \in S^0$:

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto \frac{P(x_s|\mathcal{D}_s)}{P(x_s)^{|\mathcal{C}_s|}} \prod_{r \blacktriangleleft s} P(x_s|x_r)P(x_r), \quad (15)$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{D}_s) \prod_{r \blacktriangleleft s} P(x_r|\mathcal{Y}). \quad (16)$$

For each site $s \in S \setminus S^L$ not in the leaf layer:

$$P(x_s|\mathcal{D}_s) \propto P(x_s|y_s) \prod_{t \in s^+} \sum_{x_t} \frac{P(x_t|\mathcal{D}_s)P(x_t|x_s)}{P(x_t)}. \quad (17)$$

Proof. First, (12) is a straightforward application of the total probability theorem, and the proof of (17) is the same as that reported in (Laferté et al., 2000) for a hierarchical MRF. Then, the total probability theorem implies that:

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{Y})P(\mathcal{C}_s|\mathcal{Y}), \quad (18)$$

and (14) and (16) follow from plugging (7) and (8) into this equation. Moreover, thanks to Bayes theorem:

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(\mathcal{C}_s|x_s, \mathcal{D}_s)P(x_s|\mathcal{D}_s), \quad (19)$$

where the proportionality constant does not depend on x_s. If $s \in S \setminus S^0$ and if we plug (9) into this equality, we obtain:

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_{s^-}|x_s) \prod_{r \blacktriangleleft s} P(x_r|x_s), \quad (20)$$

from which (13) follows due to Bayes theorem. The case of (15) is analogous. □

Theorem 1 indicates that $(\mathcal{X}, \mathcal{Y})$ is a Markov chain with respect to the causal structure described above, which determines the causality of the proposed hierarchical Markov model. Hence, the proposed hierarchical Markov framework is causal both spatially and across scales, which allows an efficient recursive algorithm to be formulated for the MPM criterion. Based on Theorem 2, this criterion can be formulated in terms of three recursive steps. First, the prior probabilities $P(x_s)$ are initialized on the root and (12) is used to calculate them for each class on all sites of the quadtree through a top-down pass from the root to the leaves. Second, (17), (13), and (15) are used to compute $P(x_s|\mathcal{C}_s, \mathcal{D}_s)$ through a bottom-up pass from the leaves to the root and within the root. Third, (14) and (16) are used to derive $P(x_s|\mathcal{Y})$ through a second top-down pass within the root and from the root to the leaves.

2.4 Proposed Markov chain-based method

Different choices for the definition of the total order relation $\trianglelefteq$, meaning different choices of the set of pixels which represent the past of a site $s \in S$, lead to different algorithmic formulations. In (Montaldo et al., 2019a, Montaldo et al., 2019b), a Markov mesh model was considered, which, in the generalized formulation of the present paper, is equivalent to saying that $r \triangleleft s$ if pixel r is a neighbor of s and is reached before s as the image is raster-scanned. Here, a Markov chain formulation is used, which implies that the past of a pixel is determined on a 1D basis according to a certain way of scanning all the pixels in a lattice.

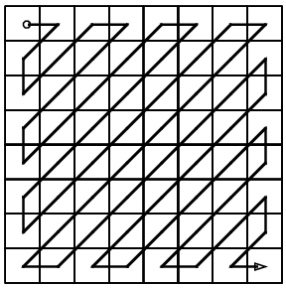

Figure 2. Zig-zag scan

On each layer S^ℓ of the quadtree ($\ell = 0, 1, \ldots, L$), let us consider a scan trajectory, i.e., a sequence that visits every pixel once and moves from one pixel to one of its neighbors. Given a pixel $s \in S^\ell$, let us indicate as $s^\star \in S^\ell$ the previous pixel in the scan. We focus here on the special case in which $r \blacktriangleleft s$ if and only if $r = s^\star$. Accordingly, (4) and (5) characterize the spatial behavior of $\mathcal{X}$ as a first-order Markov chain. In particular, (13) and (14) simplify as follows ($s \in S \setminus S^0$):

$$P(x_s|x_{s^-}, x_{s^\star}, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_s)^{-2} \cdot \\ \cdot P(x_s|x_{s^-})P(x_{s^-})P(x_s|x_{s^\star})P(x_{s^\star}), \tag{21}$$

$$P(x_s|\mathcal{Y}) = \sum_{x_{s^-}, x_{s^\star}} P(x_s|x_{s^-}, x_{s^\star}, \mathcal{D}_s)P(x_{s^-}|\mathcal{Y})P(x_{s^\star}|\mathcal{Y}),$$

and (15) and (16) simplify analogously.

Two well-known ways to perform such a pixel scan include a zig-zag trajectory and the Hilbert space-filling curve. In the former case, a zig-zag scan over the lattice S^ℓ is performed as shown in Fig. 2. With the exception of the pixels at the border of the lattice, $s^\star$ is the pixel adjacent to s and placed diagonally with respect to s in the direction of a zig-zag. Along the image borders, this trajectory aligns with the borders themselves. The latter scan option is based on the Hilbert space-filling curve over a power-two sized lattice S^ℓ (see Fig. 3). The rationale for the introduction of this order relation is twofold. First, the Hilbert curve is useful because it gives a mapping between 1D and 2D spaces that preserves locality, a precious property for contextual classification. This means that two points which are close to each other in the 1D scan are also close to each other after folding. The converse may not always be true. Second, the Hilbert curve is particularly interesting in the case of a multiresolution data set as the recursive construction of the curve matches the power-of-2 relation of the layer sizes of the quadtree.

A typical drawback of causal Markov approaches is the possible presence of anisotropic corner-dependent artifacts. To prevent them, symmetrization procedures that ensure that, within the overall execution of the method, pixels are visited in a balanced and isotropic manner, are often necessary. In this paper, this drawback is addressed by defining the scan according to a symmetrized combination of the aforementioned zig-zag and Hilbert curves. The proposed scan is shown in Fig. 4 and results as a combination of four zig-zag scans and two Hilbert scans. It is accomplished by moving onto two different directions of the Hilbert curve and onto the zig-zag curves along the two diagonals, in both possible directions. Through this scan, each site is visited multiple times in a symmetric manner, which prevents the risk of geometrical artifacts that may occur while integrating contextual information in a 1D scan on a 2D pixel lattice. Indeed, if asymmetric scanning was used, directional artifacts

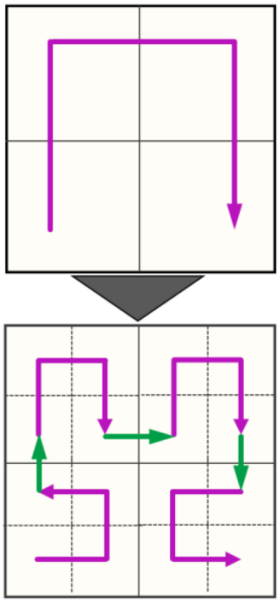

Figure 3. Hilbert space-filling curve

in the classification map may generally occur in the shapes of regions and edges (Yousefi, Kehtarnavaz, 2011).

The same approach used in (Montaldo et al., 2019a, Montaldo et al., 2019b) is used to define the transition probability $P(x_s|x_{s^-})$ across consecutive scales. It consists of a parametric stationary model in which $P\{x_s = \omega|x_{s^-} = \omega\} = \theta$ for all $\omega \in \Omega$, where θ is a hyperparameter of the method, and $P\{x_s = \omega|x_{s^-} = \omega'\}$ is constant over all $\omega \neq \omega'$ ($\omega, \omega' \in \Omega$). Following the results obtained in (Hedhli et al., 2016) in the case of a multitemporal hierarchical MRF model, the impact of θ on the results of the method are expected to be minor. An analogous model is used for the transition probability $P(x_s|x_{s^\star})$ along the scan trajectory.

GBRT is used to estimate the pixelwise posteriors $P(y_s|x_s)$ from the training samples of the classes. GBRT is an ensemble of decision trees that uses boosting (Schapire, 1990) to iteratively and adaptively manipulate the training set in order to generate multiple decision trees. A weight is assigned to each training sample, and at each iteration the weighted error on the training set is minimized by a decision tree. Then the weights are updated so that larger weight is given to wrongly classified samples and smaller weight is given to correctly classified samples. The final predictor is constructed by a weighted vote of the individual decision trees where the weights are given by the accuracy of the corresponding decision tree.

3. EXPERIMENTAL RESULTS

3.1 Data sets and experimental setup

The proposed method was experimentally validated with two VHR remote sensing datasets. The first one was collected over Port-au-Prince, Haiti, shortly after the 2010 earthquake. It consisted of an RGB GeoEye-1 image at 1.25 m resolution (1024×1024 pixels), an RGB-NIR (near infrared) QuickBird image at 2.5 m resolution, and a SAR COSMO-SkyMed stripmap image at 5 m resolution. For the COSMO-SkyMed stripmap modality, 5 m is the spatial resolution corresponding

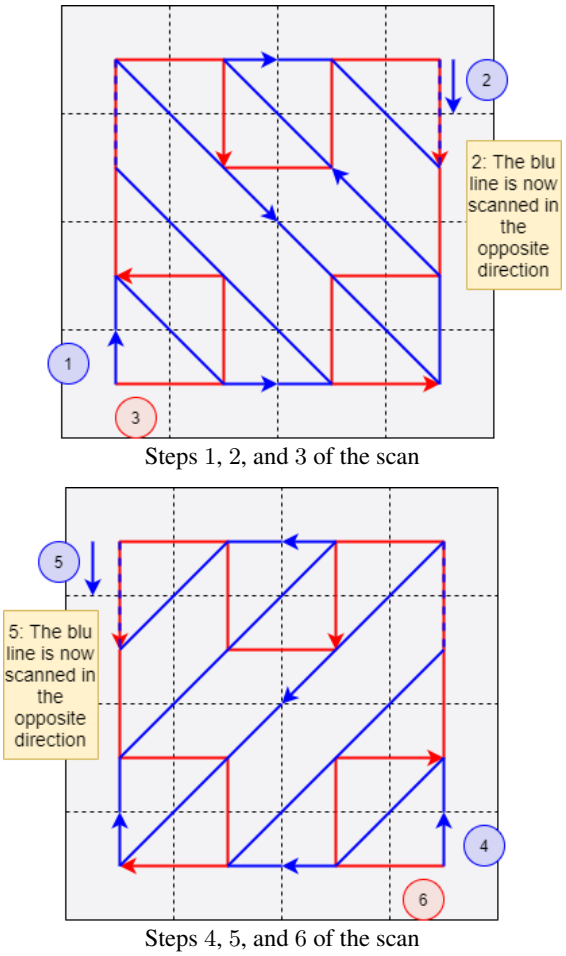

Figure 4. Symmetrized scan introduced. It consists of a combination of four zig-zag scans and two Hilbert scans (6 steps illustrated in the figure).

to 4-look data. The resolution of the multispectral channels of QuickBird is 2.4 m, so minor resampling was applied. In the case of GeoEye-1, the resolutions of the multispectral and panchromatic channels are 1.64 and 0.41 m, respectively. The 1.25 m image was obtained by resampling after pansharpening. Pansharpened data were used as this was the format in which GeoEye-1 imagery was provided by the Google Crisis Response initiative[2]. In all resampling processes, proper antialiasing filtering was applied (Alparone et al., 2008, Mallat, 2009). The thematic classes in the data set included "buildings," "asphalt," "containers," "vegetation," and "water." In the desired output map, these classes were meant to be discriminated at the finest observed resolution, i.e., 1.25 m. Yet, the proposed method also mapped them at the other observed scales, i.e., 2.5 and 5 m.

The second data set was acquired by the IKONOS optical multiresolution sensor in 2004 over the area of Alessandria, Italy. It consisted of a single-channel panchromatic image at 1 m resolution (1024×1024 pixels) and a 4-channel RGB-NIR multispectral image at 4 m resolution. The thematic classes in this data set were "urban," "agricultural," "rangeland," "forest," "bare soil," "wet soil," and "water." As in the case of the first data set, they were meant to be discriminated on the 1 m lattice, although the method also provided a classification map at 4 m as a by-

[2] http://www.google.com/relief/haitiearthquake/geoeye.html

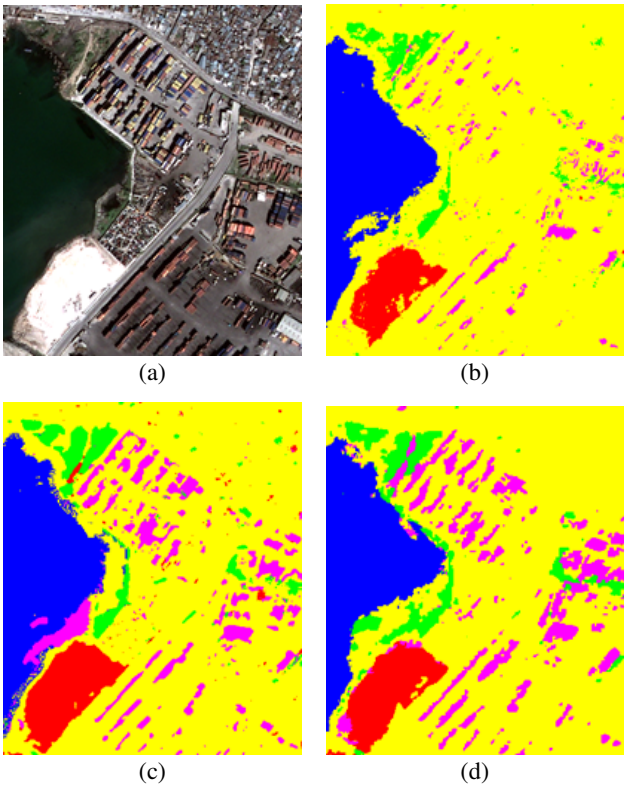

Figure 5. "Haiti" data set. Details of (a) the optical image at 1.25-m resolution and of the classification maps obtained using (b) a single-resolution MRF after resampling, (c) an adaptation of the method in (Moser et al., 2016) to the case of input multisensor optical-SAR imagery, and (d) the proposed algorithm. Class legend: containers, vegetation, asphalt, buildings, water.

product. The urban area was meant to be discriminated as a whole, without separating possible subclasses associated with ground materials.

For both data sets, non-overlapping training and test sets were manually annotated in homogeneous areas by a specialist. No borders were taken into account in the training and test sets to avoid mixed pixels in the ground truth.

For comparison purposes, three previous approaches to multiresolution and multisensor fusion were considered. The first one consists in a planar MRF classifier in which unary potentials were based on the pixelwise predictions obtained by random forest (Breiman, 2001) after resampling all the images to the finest resolution. The pairwise potential was the Potts model, and the energy minimization was through sequential tree-reweighted message passing (TRW-S) (Kolmogorov, 2006). The second one is the method in (Moser et al., 2016) for the classification of multiresolution optical images through MRF, graph cuts, and a linear mixture model. This method makes use of a Gaussian class-conditional model. In the application to SAR data, it was adapted by using a log-normal class-conditional model and by extending the technique accordingly. The third benchmark technique is the recently developed approach in (Montaldo et al., 2019a, Montaldo et al., 2019b) in which decision tree ensembles are combined with a causal hierarchical Markov model in which a Markov mesh is postulated at each resolution scale. The framework proposed in the present paper extends this previous technique. The algorithm

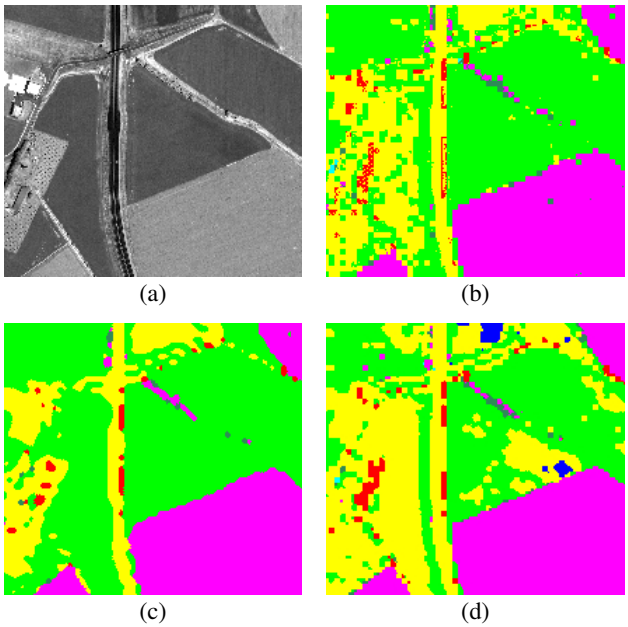

Figure 6. "Alessandria" data set. Details of (a) the IKONOS panchromatic image at 1-m resolution and of the classification maps obtained using (b) a single-resolution MRF after resampling, (c) the method in (Moser et al., 2016), and (d) the proposed algorithm. Class legend: urban, agricultural, rangeland, bare soil, forest, water, wet soil.

developed here within the proposed framework differs from that in (Montaldo et al., 2019a, Montaldo et al., 2019b) especially in the modeling of spatial-contextual information within each layer of the quadtree (through Markov chains in the developed method and Markov meshes in (Montaldo et al., 2019a, Montaldo et al., 2019b)).

3.2 Experimental results and comparisons

The overall and class-wise accuracies obtained on the test sets are presented in Tables 1 and 2. Spatial details of the classification maps are shown in Figs. 5 and 6. The proposed technique obtained high accuracies on the test sets of both data sets, a result that suggests the effectiveness of the proposed causal hierarchical Markov framework and of the Markov chain-based algorithm developed within this framework. Remarkably, the accuracies obtained by the technique in (Montaldo et al., 2019a, Montaldo et al., 2019b) were very similar (identical in some cases) to those of the proposed technique. This similarity is consistent with the fact that both algorithms are special cases of the general framework formalized in the present paper and share the same multiresolution hierarchical structure based on a quadtree. On one hand, this similarity suggests that the spatial-contextual models based on Markov meshes and Markov chains, when integrated in the proposed framework and at least in the case of the considered data sets, allows for analogous capabilities in terms of class discrimination. From this perspective, large accuracy differences between the two approaches are generally not expected. On the other hand, this similarity in accuracy is contrasted by significantly different computational burdens. In the proposed chain-based algorithm, a pixel s is linked to the preceding pixel in the 1D scanning curve. Hence, (13) and (14), in their special case (21), involve looping over the three labels x_s, x_{s^-}, and $x_{s^\star}$, which is computationally $O(|\Omega|^3)$. In the case of the previous method in (Montaldo et al., 2019a, Montaldo et al., 2019b), the mesh formulation involves additional neighboring pixels, which leads, for instance, to $O(|\Omega|^4)$ or $O(|\Omega|^5)$ in the case of meshes with second- or third-order, respectively. Accordingly, the algorithm developed here exhibits significant computational advantages over the previous formulation in (Montaldo et al., 2019a, Montaldo et al., 2019b).

The other two previous approaches also obtained high classification performances, yet with lower overall and/or class-wise accuracies (e.g., for "forest" in the "Alessandria" data set). This further suggests the effectiveness of the proposed framework based on a hierarchical causal probabilistic graphical model as compared not only to the use of a planar MRF together with a traditional resampling process but also to a previous recent Markov model in which multiresolution fusion is formalized through linear mixtures.

A visual analysis of the classification maps confirm these comments (Figs. 5 and 6). The map generated by the proposed algorithm was characterized by remarkable visual regularity without exhibiting oversmoothing issues. This is interpreted as a consequence of the contextual modeling obtained through Markovianity across both the scale and spatial domains.

4. CONCLUSIONS

A general hierarchical causal Markov framework has been proposed for the challenging problem of jointly classifying input image data that are both multiresolution and multisensor. The proposed framework generalizes previous formulations that made use of Markov meshes by combining a multiresolution quadtree structure with a spatial Markovianity condition formulated with respect to a generic total order relation. The causality of the overall framework and the formulation of MPM have been analytically proven, thus generalizing the previous methodological results we proved in (Montaldo et al., 2019a). Within this framework, a novel classification algorithm based on a Markov chain model within each layer of the quadtree has been developed. Experimental results obtained with VHR satellite images collected by panchromatic, multispectral, and SAR sensors pointed out the capability of the proposed algorithm to achieve high class-wise and overall accuracy. This behavior is shared by the previous technique in (Montaldo et al., 2019a, Montaldo et al., 2019b), which can be seen as a special case of the general framework developed here. On one hand, these results confirm the capability of this framework to address multiresolution and multisensor fusion through a causal hierarchical Markov approach. A further desirable property of the proposed framework is that it allows classification maps to be obtained at all spatial resolutions in the input data set. On the other hand, the specific algorithm developed in the present paper exhibits significantly lower computational complexity than the previous one in (Montaldo et al., 2019a, Montaldo et al., 2019b). Indeed, the approach of associating each layer of the quadtree with a 1D Markov chain along Hilbert and zig-zag curves demonstrated advantageous from a computational perspective thanks to its lower complexity and correspondingly shorter execution times. This reduction in the order of complexity allows to address more efficiently larger images and allows especially to consider a larger number of classes. Higher accuracies were also obtained by the proposed algorithm as compared both to a traditional resampling-based approach and to an advanced multiresolution model based on linear mixture

Table 1. "Haiti" data set. Overall accuracy (OA) [%], Cohen's κ, and class-by-class accuracies [%] on the test set for the proposed algorithm and the compared benchmark methods.

resolution 1.25 m	containers	vegetation	asphalt	buildings	sea	OA	κ
Single-res. MRF after resampling	63.97	74.81	98.64	99.30	76.24	94.58	0.8880
Adaptation of (Moser et al., 2016)	75.06	85.32	95.89	97.75	99.36	95.11	0.9010
Method in (Montaldo et al., 2019b)	87.08	33.27	95.17	99.04	97.18	96.36	0.9124
Proposed method	**86.97**	**34.57**	**97.82**	**99.36**	**100**	**96.90**	**0.9057**
resolution 2.5 m	containers	vegetation	asphalt	buildings	sea	OA	κ
Method in (Montaldo et al., 2019b)	86.31	32.71	94.87	100.00	96.04	96.63	0.9276
Proposed method	**86.92**	**33.08**	**97.29**	**100.00**	**100**	**97.00**	**0.9218**
resolution 5 m	containers	vegetation	asphalt	buildings	sea	OA	κ
Method in (Montaldo et al., 2019b)	87.98	25.39	96.55	100.00	88.88	96.01	0.9288
Proposed method	**87.98**	**25.39**	**96.55**	**100.00**	**88.88**	**96.01**	**0.9288**

Table 2. "Alessandria" data set. Overall accuracy (OA) [%], Cohen's κ, and class-by-class accuracies [%] on the test set for the proposed algorithm and the compared benchmark methods.

resolution 1 m	urban	agri	rangeland	forest	water	wet soil	bare soil	OA	κ
Single-res. MRF after resampling	98.58	99.12	92.23	36.98	100	98.30	96.82	96.03	0.9453
Method in (Moser et al., 2016)	99.70	99.21	97.34	64.92	100	100	99.66	98.60	0.9804
Method in (Montaldo et al., 2019b)	100	99.07	99.65	100	99.65	100	100	99.12	0.9976
Proposed method	**99.54**	**98.26**	**99.91**	**86.40**	**99.80**	**100**	**100**	**98.58**	**0.9853**
resolution 2 m	urban	agri	rangeland	forest	water	wet soil	bare soil	OA	κ
Method in (Montaldo et al., 2019b)	99.54	98.06	99.91	78.74	99.53	100	100	98.01	0.9825
Proposed method	**99.49**	**98.08**	**99.91**	**79.52**	**99.84**	**100**	**100**	**98.03**	**0.9829**
resolution 4 m	urban	agri	rangeland	forest	water	wet soil	bare soil	OA	κ
Method in (Montaldo et al., 2019b)	99.09	96.81	99.82	64.86	98.72	100	100	96.36	0.9697
Proposed method	**99.09**	**96.81**	**99.82**	**64.86**	**98.72**	**100**	**100**	**96.36**	**0.9697**

concepts (Moser et al., 2016). Furthermore, these previous approaches allow a classification map to be generated at the finest spatial resolution in the input data set but not at the other resolutions involved (unless case-specific resampling procedures are applied).

Future extensions of this work may involve integrating its multiresolution structure with convolutional neural networks, whose convolution and pooling layers intrinsically extract multiscale data representations (Goodfellow et al., 2016), and with adaptive multiresolution topologies based for instance on superpixel segmentations at different scales (Li, 2009).

ACKNOWLEDGMENTS

The authors would like to thank the Italian Space Agency (ASI), DigitalGlobe Inc., and Google Crisis Response for providing the COSMO-SkyMed, QuickBird, and GeoEye-1 images used for experiments, respectively.

REFERENCES

Abend, K., Harley, T. J., Kanal, L. N., 1965. Classification of Binary Random Patterns. *IEEE Transactions on Information Theory*, 11(4), 538-544.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., 2008. *Remote Sensing Image Fusion*. CRC Press.

Arnab, A., Zheng, S., Jayasumana, S., Romera-Paredes, B., Larsson, M., Kirillov, A., Savchynskyy, B., Rother, C., Kahl, F., Torr, P. H. S., 2018. Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction. *IEEE Signal Processing Magazine*, 35(1), 37-52.

Baggenstoss, P. M., 2001. A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. *IEEE Transactions on Speech and Audio Processing, vol. 9, no. 4, pp. 411-416*.

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

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), 5-32.

Devijver, P. A., 1993. Hidden Markov mesh random field models in image analysis. *J. Applied Stat.*, 20(5-6), 187-227.

Dikshit, O., Roy, D. P., 1996. An empirical investigation of image resampling effects upon the spectral and textural supervised classification of a high spatial resolution multispectral image. *Photogrammetric Engineering and Remote Sensing, 62(9):1085-1092*.

Fjortoft, R., Delignon, Y., Pieczynski, W., Sigelle, M., Tupin, F., 2003. Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Trans. Geosci. Rem. Sens.*, 41, 675-686.

Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 29(5), 1189-1232.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT Press.

Hedhli, I., Moser, G., Serpico, S. B., Zerubia, J., 2017. Classification of Multisensor and Multiresolution Remote Sensing Images Through Hierarchical Markov Random Fields. *IEEE Geosci. Rem. Sens. Lett.*, 14(12), 2448-2452.

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 Trans. Geosci. Rem. Sens.*, 54(11), 6333-6348.

Inglada, J., Muron, V., Pichard, D., Feuvrier, T., 2007. Analysis of artifacts in subpixel remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing, 45(1):254-264*.

Kolmogorov, V., 2006. Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1568-1583*.

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

Li, S., 2009. *Markov Random Field Modeling in Image Analysis*. Springer.

Mallat, S., 2009. *A wavelet tour of signal processing – The sparse way*. Academic Press.

Montaldo, A., Fronda, L., Hedhli, I., Moser, G., Zerubia, J., Serpico, S. B., 2019a. Causal markov mesh hierarchical modeling for the contextual classification of multiresolution satellite images. *Proc. of IEEE ICIP 2019*.

Montaldo, A., Fronda, L., Hedhli, I., Moser, G., Zerubia, J., Serpico, S. B., 2019b. Joint classification of multiresolution and multisensor data using a multiscale markov mesh model. *Proc. of IEEE IGARSS 2019*.

Moser, G., De Giorgi, A., Serpico, S. B., 2016. Multiresolution Supervised Classification of Panchromatic and Multispectral Images by Markov Random Fields and Graph Cuts. *IEEE Trans. Geosci. Rem. Sens.*, 54(9), 5054-5070.

Schapire, R., 1990. The strength of weak learnability. *Machine Learning*, 197227.

Wang, X., 2016. Deep learning in object recognition, detection, and segmentation. *Foundations and Trends in Signal Processing*, 8(4), 217-382.

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

Yousefi, S., Kehtarnavaz, N., 2011. Generating symmetric causal Markov random fields. *Electronics letters, vol. 47(22):1224-1225*.