HAL open science

Multisensor and Multiresolution Remote Sensing Image Classification through a Causal Hierarchical Markov Framework and Decision Tree Ensembles

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

To cite this version:

Martina Pastorino, Alessandro Montaldo, Luca Fronda, Ihsen Hedhli, Gabriele Moser, et al.. Multisensor and Multiresolution Remote Sensing Image Classification through a Causal Hierarchical Markov Framework and Decision Tree Ensembles. Remote Sensing, 2021, 13 (5), pp.25. 10.3390/rs13050849 . hal-03155379v2

HAL Id: hal-03155379

https://inria.hal.science/hal-03155379v2

Submitted on 4 Mar 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.

remote sensing

MDPI

Article

Multisensor and Multiresolution Remote Sensing Image Classification through a Causal Hierarchical Markov Framework and Decision Tree Ensembles

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

[1] DITEN Deptartment, University of Genoa, 16145 Genoa, Italy; martina.pastorino@edu.unige.it (M.P.); 4039367@studenti.unige.it (A.M.); 4047957@studenti.unige.it (L.F.); sebastiano.serpico@unige.it (S.B.S.)
[2] Inria, Université Côte d'Azur, 06902 Sophia-Antipolis, France; josiane.zerubia@inria.fr
[3] Institute on Intelligence and Data, Laval University, Quebec City, QC G1V0A6, Canada; ihsen.hedhli.1@ulaval.ca
* Correspondence: gabriele.moser@unige.it

Abstract: In this paper, a hierarchical probabilistic graphical model is proposed to tackle joint classification of multiresolution and multisensor remote sensing images of the same scene. This problem is crucial in the study of satellite imagery and jointly involves multiresolution and multisensor image fusion. The proposed framework consists of a hierarchical Markov model with a quadtree structure to model information contained in different spatial scales, a planar Markov model to account for contextual spatial information at each resolution, and decision tree ensembles for pixelwise modeling. This probabilistic graphical model and its topology are especially fit for application to very high resolution (VHR) image data. The theoretical properties of the proposed model are analyzed: the causality of the whole framework is mathematically proved, granting the use of time-efficient inference algorithms such as the marginal posterior mode criterion, which is non-iterative when applied to quadtree structures. This is mostly advantageous for classification methods linked to multiresolution tasks formulated on hierarchical Markov models. Within the proposed framework, two multimodal classification algorithms are developed, that incorporate Markov mesh and spatial Markov chain concepts. The results obtained in the experimental validation conducted with two datasets containing VHR multispectral, panchromatic, and radar satellite images, verify the effectiveness of the proposed framework. The proposed approach is also compared to previous methods that are based on alternate strategies for multimodal fusion.

Keywords: multimodal data fusion; multiresolution and multisensor fusion; causal Markov model; hierarchical Markov random field; Markov mesh random field; Markov chain; decision tree ensemble; semantic image segmentation; remote sensing

check for updates

Citation: Pastorino, M.; Montaldo, A.; Fronda, L.; Hedhli, I.; Moser, G.; Serpico, S.B.; Zerubia, J. Multisensor and Multiresolution Remote Sensing Image Classification through a Causal Hierarchical Markov Framework and Decision Tree Ensembles. *Remote Sens.* **2021**, *13*, 849. https://doi.org/10.3390/rs13050849

Academic Editor: Junshi Xia

Received: 8 January 2021
Accepted: 19 February 2021
Published: 25 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Due to the heterogeneity of high or very high resolution (HR/VHR) remote sensing imagery, including optical (e.g., Pléiades, WorldView-3, SPOT-6/7) and synthetic aperture radar (SAR; e.g., COSMO-SkyMed Second Generation, TerraSAR-X, RADARSAT Constellation) data, a challenging problem is the development of image classifiers that can exploit information from multimodal input observations. This multimodal framework allows to take advantage of data associated with different physical characteristics, frequencies, polarizations, and to exploit the advantages of information at different scales, which can be either HR or VHR images or even a combination of the two categories: fine-scale observations capture geometrical details but are noise-sensitive, while coarse-scale observations are characterized by large image regions with a strong robustness to outliers and noise [1–3].

A major issue in this framework is the heterogeneity of the data coming from different sensors and the characterization of spatial information associated with different resolutions [4].

Trivial well-known solutions mostly use resampling procedures, which are not computationally demanding, in general, but can produce artifacts (e.g., because of aliasing) [3], and do not manage to explicitly capture the multiresolution structure of the image data.

In this paper, an approach to address the semantic segmentation [5,6], or dense supervised classification, of multiresolution and multisensor images is proposed, based on hierarchical probabilistic graphical models (PGMs), in particular hierarchical latent Markov models [7]. On one hand, Markov models formulated on planar graphs or on multilayer graphs such as quadtrees are powerful and flexible stochastic models in the field of structured output learning [7]. However, a common drawback of many Markov random fields (MRFs) postulated on 2D grids of pixels is that they are generally non-causal. The resulting non-convex problems are addressed with time-consuming iterative inference algorithms such as stochastic optimizers (e.g., simulated annealing) [7] or graph-theoretic techniques [8]. On the other hand, an effective way to model multiresolution information is to adopt a hierarchical MRF on a quadtree, to take into account relations among sites at different scales with a Markov chain. These models are causal and it is possible to use efficient sequential techniques for inference (e.g., the Baum–Welch algorithm [9]), but they do not model explicitly contextual-spatial information at each resolution [10].

In the proposed approach, Markovianity is postulated within each scale according to a causal neighborhood system defined on the pixel lattice by an order relation. This choice makes it possible to jointly address spatial-contextual and multiresolution fusion within a unique joint causal framework. Operatively, within the pixel lattice at each scale, this is addressed either by a one-dimensional scan, combining both a zig-zag trajectory and a Hilbert space-filling curve [11,12], or by a Markov mesh random field, where a neighborhood system is defined on a planar lattice, allowing to formulate causality [11,13–16].

Multiresolution fusion is intrinsically supported by the topology of the proposed framework, while multisensor (optical and radar) fusion is addressed by the integration of non-parametric ensemble modeling, e.g., decision tree ensembles [17], into the proposed hierarchical Markov model. From this perspective, the developed framework generalizes and completes the preliminary formulations that were presented in the conference papers [18–21].

The marginal posterior mode (MPM) criterion for inference is analytically formulated and proved for the proposed model. This criterion, when applied to quadtree structures, is based on recursive equations leading to the direct computation of the target marginal posterior distribution. As compared to the aforementioned iterative stochastic or graph-theoretic minimization algorithms for non-causal planar Markov random fields, it allows reducing the computational burden and increasing the efficiency of the inference. It is especially favorable for classification tasks linked to multiresolution models [10].

The main contributions of this paper are twofold. Firstly, a novel framework is proposed to tackle multiresolution fusion and spatial-contextual classification of remote sensing images, based on the aforementioned PGM formulation with respect to a total order relation on each planar pixel lattice involved. Input multiresolution datasets including HR images, VHR images, or both are supported in the proposed approach. In particular, the causality of the whole hierarchical Markov framework is proved in the general case associated with an arbitrary total order relation. Furthermore, the analytical formulation of the MPM criterion for the proposed framework is also proved, and the related conditional independence assumptions are discussed. Secondly, two specific classification algorithms are developed within the proposed framework for input multiresolution and multisensor imagery. These algorithms are based on the definition, on each layer of the quadtree, of either a Markov mesh or a Markov chain with respect to a 1D scan of the pixel lattice. In both methods, the fusion of multisensor imagery associated with different spatial resolutions is naturally accomplished thanks to the quadtree topology. The fusion of multisensor imagery collected at the same spatial resolution is obtained by incorporating decision tree ensembles [22] in both methods. These ensemble approaches, such as random forest [23], rotation forest [24], ExtraTrees [25], and gradient boosted regression trees (GBRT) [22], are characterized by a strong flexibility and a non-parametric formulation, allowing them to

be applied to highly heterogeneous input data, such as the ones deriving from multisensor applications [17,26]. In this manner, multisensor and multiresolution fusion are jointly addressed in a unique probabilistic graphical framework. The experimental validation was carried out with two VHR satellite datasets of urban and agricultural areas associated with land cover mapping and, in one case, to flood risk prevention [27]. We recall that, in the previous conference paper in [21], the experimentation on the hierarchical PGM with the Markov chain was conducted with the sole GBRT.

This paper is organized as follows: after a review of the state of the art in the field of multiresolution and multisensor classification in Section 2, the developed methodology is presented in Section 3. First, Section 3.1 presents the hierarchical framework proposed, with the required model assumptions in Section 3.1.1 and the theoretical properties in Section 3.1.2. Then, the two multimodal classification methods developed within the proposed framework are presented in Section 3.2. Sections 4 and 5 report the experimental results obtained with the two VHR datasets considered and a discussion of the performances of the proposed methods. Conclusions are drawn in Section 6.

2. Previous Work

A major challenge regarding the efficient processing of data in remote sensing is given by the recent increase in availability of image data acquired by different sensors, with their corresponding heterogeneous natures. The aim of multisensor image fusion is to integrate complementary information from several images related to the same scene and collected by distinct sensors (e.g., optical and synthetic aperture radar—SAR, or hyperspectral and LiDAR). In the case of multiresolution fusion, multiple images associated with distinct spatial resolutions are jointly used to benefit from their complementary perspectives (more synoptic view vs. more precise spatial detail). Here, the focus is on multisensor and multiresolution fusion aimed at generating supervised image classification or semantic segmentation results from input multimodal imagery. In general terms, the literature on either multisensor or multiresolution fusion in remote sensing is vast and dates back a few decades [2,28], whereas the literature on joint multisensor and multiresolution classification is scarcer. In this framework, serious issues are both the heterogeneity of the data coming from different sensors and the need for a characterization of the spatial information associated with different resolutions [4]. Well-known solutions mostly use resampling procedures, in general computationally cheap but that may cause artifacts [3] and that do not explicitly aim to benefit from the information carried by multiresolution data. Here, we shall review the state of the art associated with multisensor and multiresolution fusion—jointly—and then we shall focus especially on such fusion approaches developed for classification purposes.

Popular approaches involve wavelet-based algorithms [29]. For example, in [30] a hybrid multiresolution method combining the stationary wavelet transform (SWT) with the nonsubsampled contourlet transform (NSCT) to perform image fusion on a visible satellite image and an infrared image is presented. Wavelet-based methods are usually employed to perform the fusion of images captured by different sensors with different resolutions. In general, optical images with fine spectral resolution tend to be coarser in spatial resolution as compared to images with fewer spectral bands, for reasons associated with the signal-to-noise ratio of the optical instrument [31]. In [32] a wavelet-based multiresolution pyramid applied to multitemporal or multisensor satellite data is combined with a stochastic gradient based on two similarity measures, correlation and mutual information. In [33], an image fusion based on a preregistration process is implemented using the scale-invariant feature transform (SIFT) with a reliable outlier removal procedure, a fine-tuning process, and the maximization of mutual information using a modified Marquardt—Levenberg search strategy, is then tested on various remote sensing optical and SAR images. In [34] several wavelet pyramids aimed at performing invariant feature extraction and accelerating image fusion through multiple spatial resolutions are evaluated. In these methods, in order to achieve computational efficiency, fusion is performed iteratively from the coarsest level

of the image pyramid, where the number of pixels is lower and the computational cost is limited, to the finest level of the image pyramid [32–34].

In [35] the optimal number of decomposition levels required in the wavelet transform in wavelet-based fusion is investigated. Based on the results obtained it was possible to deduce that too few decomposition levels result in fused images with poor spatial quality, while too many levels reduce the spectral similarity between the original multispectral and the pan-sharpened images.

Pan-sharpening is a very common family of multiresolution fusion techniques. Component substitution (CS)-based pan-sharpening methods spectrally transform the multispectral data into another feature space and separate spatial and spectral information into different components [36]. Among these methods, well-known techniques involve intensity-hue-saturation (IHS) [37], principal component analysis (PCA) [38], and Gram-Schmidt [39] transformations. On the other hand, multiresolution analysis (MRA)-based algorithms consist in the injection of spatial details (or high-frequency components) from a panchromatic image to a multispectral image [36,39]. Popular MRA approaches are based on box filtering [40], Gaussian filtering [41], wavelet transform [42,43], such as decimated wavelet transform (DWT) [29], and curvelet transform [44]. Indeed, pansharpening addresses multiresolution fusion from a signal-processing perspective and not from a supervised learning perspective. Accordingly, the resulting fusion product is not designed to be optimal from the viewpoint of supervised labeling with respect to a set of thematic classes described by a training set.

Furthermore, with the advent of unmanned aerial vehicles (UAVs), images with extremely high spatial resolution have become available at a relatively low cost. UAVs are often equipped with simple, lightweight sensors, such as RGB cameras [45] that capture small portions of land—a scenario that may raise the need for multimodal fusion to benefit from the available lightweight instruments. In [45] classification of a high spatial resolution color image and a lower spatial resolution hyperspectral image of the same scene is addressed. Contextual information is obtained from the color image through color attribute profiles (CAPs) and spectral information is extracted from the hyperspectral image; a composite decision fusion (CDF) strategy exploiting kernel-based techniques is proposed.

Joint optical-SAR data classification has been addressed for a couple of decades [46] with approaches stemming from several methodological areas, such as statistical pattern recognition [47], neural networks [48–50], consensus theory and decision fusion [51], kernel-based learning [52], and MRFs [53]. From a signal-processing perspective, multisensor image fusion algorithms to integrate panchromatic and SAR features into multispectral images were proposed in [54,55], extracting SAR texture by ratioing the despeckled SAR image to its low-pass approximation with a wavelet decomposition, and using the modified Brovey transform (MBT) conjointly with a wavelet decomposition, respectively. In [56] an image fusion algorithm to merge SAR and optical images based on the discrete multiwavelet transform (DMWT) performed at pixel level and an area-based selective fusion of SAR information to inject in the optical image is presented.

Following signal-level fusion through pan-sharpening, to perform image classification, either single-resolution classifiers [57] or multiresolution methods, based, for example, on Bayesian modeling [58] or hierarchical MRFs [7] can be applied. So far, the framework of supervised classification of multisensor and multiresolution remote sensing data has been scarcely investigated. Two examples of joint multisensor and multiresolution image classification are presented in [4,59]. In [4], a method based on multiple hierarchical Markov random fields on distinct quadtrees associated with multisensor optical and SAR data is introduced. This framework is integrated with finite mixture models and the MPM criterion. In [59], a copula-theoretic multivariate model allows to perform multisensor fusion and a multiscale Markovian model preserves details and avoids the smoothing effects generated by the single-scale MRF-based method. Recent approaches based on deep learning and convolutional neural networks (CNNs) in the field of multimodal—specifically multitemporal

and multisensor—classification of remote sensing images are [60,61], where multitemporal samples are collected from Sentinel-2A/B and Landsat-8 satellites and the focus is on input satellite image time series.

3. Methodology

In general terms, a semantic segmentation (or image classification) task can be considered as the problem of estimating latent information represented by class labels x from feature vectors y associated with one or more sites (i.e., pixels) s of an image data set. From a statistical perspective, x and y are usually regarded as occurrences of random vectors. The inference of x given y may be computationally expensive since, in many cases—and specifically in the case of non-causal PGMs on planar graphs associated with image lattices—iterative algorithms are necessary. On the contrary, probabilistic graphical models that involve the use of hierarchical MRFs on quadtrees use the concept of causality in order to develop efficient non-iterative inference algorithms. From a modeling viewpoint, this is accomplished through the factorization of the latent field distribution (i.e., the prior distribution of x) in terms of causal transition probabilities.

Here, this approach is adopted to address semantic segmentation with input multiresolution and multisensor images. The idea is, first, to make use of a quadtree (shown in Figure 1) to model the interactions across multiresolution images. Input image data are inserted in the layers of the quadtree as a function of their spatial resolutions. Then, we propose a hierarchical Markov model that takes into account both the relations between different scales and the spatial information at each scale, in order to favor applicability to HR and VHR imagery. The causality of the proposed model makes it possible to derive an efficient formulation of the MPM criterion that is used to predict the label of each pixel in each layer of the quadtree. The outcome is a multiresolution dense classification result generated from the input multiresolution images. For this purpose, decision tree ensembles are separately trained on all layers of the quadtree to incorporate image data coming from different sensors that operate at the same resolution. Details can be found in the next sections.

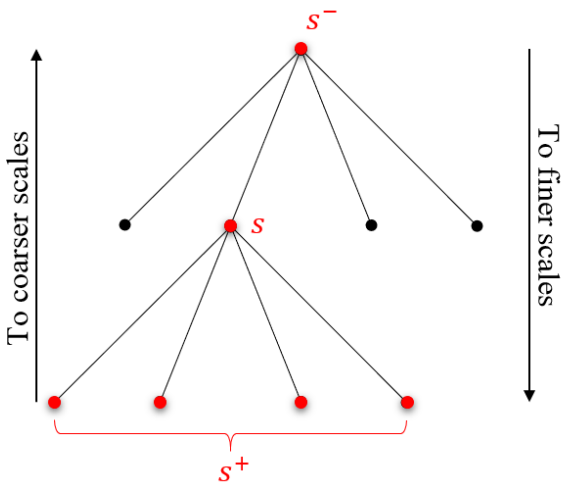


Figure 1. Structure of the quadtree and the corresponding parent–child relations.

3.1. The Proposed Hierarchical Causal Markov Framework

Multiresolution models are fundamental among the current trends of probabilistic graphical models. From an image processing perspective, the idea is to develop analysis methods that can take advantage of all available resolutions and model the trade-off between the coarser ones, more robust to noise and outliers, and the geometrical details of the finer ones. Probabilistic graphical models on either planar or hierarchical multilayer graphs [62], such as quadrees [10], have been popular for long as stochastic models for multiscale and possibly multisensor information. Hierarchical MRFs based on this topology typically capture

multiresolution relations across the different layers of the quadtree through a Markov chain. However, in general, they do not model the spatial dependencies within each layer [10]. In this paper, a causal hierarchical framework postulating Markovianity both inter-scale (multiscale information) and intra-scale (contextual spatial information) is defined to address multiresolution fusion while also capturing the spatial information at each scale—a desired property when HR and VHR imagery is considered.

In the following, we shall use $P(\cdot)$ and $p(\cdot)$ to indicate the probability mass function (PMF) of a discrete random variable or vector and the probability density function (PDF) of a continuous random variable or vector, respectively.

3.1.1. Model Assumptions

Consider $\{S^0, S^1, \ldots, S^L\}$, an ensemble of pixel lattices structured as a quadtree, as illustrated in Figure 1. The quadtree is built with the width and the height of $S^{\ell-1}$ being half those of S^ℓ ($\ell = 1, 2, \ldots, L$). Accordingly, S^0 is the root of the quadtree and corresponds to the coarsest spatial resolution, while S^L is the layer of the leaves of the tree and corresponds to the finest resolution. Following the usual terminology of hierarchical Markov models, the pixels in the various layers of the quadtree will be named sites. Each site $s \in S^\ell$ in layer S^ℓ except the root (i.e., with $\ell \in \{1, 2, \ldots, L\}$) has a parent site $s^- \in S^{\ell-1}$. Each site $s \in S^\ell$ in layer S^ℓ except the leaves (i.e., with $\ell \in \{0, 1, \ldots, L-1\}$) has four children sites in $S^{\ell+1}$, which are collected in the set $s^+ \subset S^{\ell+1}$. These definitions determine a hierarchy on the quadtree $S = S^0 \cup S^1 \cup \ldots \cup S^L$ starting from the root S^0 to the leaves S^L [10].

Let us also consider a discrete random variable x_s, representing a class label, associated with each site $s \in S$, and let Ω be the finite set of classes, such that $x_s \in \Omega$ for each $s \in S$. The random field $\mathcal{X} = \{x_s\}_{s \in S}$ of all class labels in the quadtree is a hierarchical MRF if and only if the following property holds for all layers, with the exception of the root layer ($\ell = 1, 2, \ldots, L$) [10]

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

where $\mathcal{X}^\ell = \{x_s\}_{s \in S^\ell}$ ($\ell = 0, 1, \ldots, L$). According to the first equality, the Markovian property holds across the layers of the quadtree, while the second equality indicates that the pixelwise transition probability $P(\mathcal{X}^\ell | \mathcal{X}^{\ell-1})$ from layer $(\ell - 1)$ to layer ℓ is supposed to factor out on a pixelwise basis. The former equality and the Markov chain structure it yields across the layers of the quadtree imply a well-known property of hierarchical MRFs on quadtrees, i.e., causality [10]. The latter factorization is also a customary condition in hierarchical MRFs for image classification [10] but has implications in terms of modeling properties. Indeed, as mentioned before, a hierarchical MRF on a quadtree accounts for the dependencies between pixels in different scales, but does not model the spatial-contextual dependencies associated with the random field $\mathcal{X}^\ell$ on each layer S^ℓ (see Figure 2a). In the proposed framework, we overcome this limitation by also incorporating a Markov model for the spatial context on the pixel grid at each scale. Simultaneously, it is desired that this spatial model preserves causality, because of the aforementioned implications in terms of efficient inference.

For this purpose, let us denote as $R \subset \mathbb{Z}^2$ a planar rectangular pixel lattice where a total order relation $\preceq$ is defined. We recall that a total order relation satisfies the following properties ($r, s, t \in R$) [63]:

$$\begin{array}{lll} \text{Antisymmetry:} & r \preceq s, \quad s \preceq r & \Longrightarrow \quad r = s \\ \text{Transitivity:} & r \preceq s, \quad s \preceq t & \Longrightarrow \quad r \preceq t \\ \text{Connexity:} & \text{at least one holds between } r \preceq s \text{ and } s \preceq r. & \end{array}$$

The relation $r \prec s$ indicates that $r \preceq s$ and $r \neq s$. The set $\{r \in R : r \prec s\}$ defines the "past" of a site $s \in R$. A neighborhood relation is also defined in order to have the "past neighbours" of each site, consistently with the order relation $\prec$. The relation $r \precsim s$ denotes that r is a past neighbor of s. Formally, $\{r \in R : r \precsim s\} \subsetneq \{r \in R : r \prec s\}$, which means

that the past neighbors of $s \in R$ are included in the past of s but constitute a strict subset of its entire past. If we consider again a discrete random variable x_s, representing a class label attached to each site $s \in R$, and let $\mathcal{X} = \{x_s\}_{s \in R}$ be the corresponding random field, then the Markovianity property constrained to the past of each site holds for $\mathcal{X}$ if and only if [11,13,62]:

$$P(x_s | x_r, r \prec s) = P(x_s | x_r, r \precsim s) \qquad \forall s \in R. \tag{2}$$

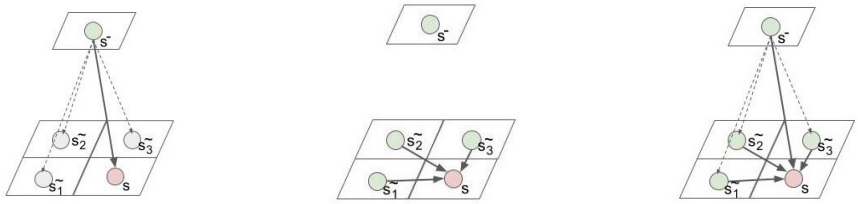


Figure 2. Causal models of (**a**) a hierarchical Markov random field (MRF), (**b**) a Markov mesh, and (**c**) the proposed symmetric Markov mesh random field (SMMRF)-based algorithm.

Again, this Markov model associated with an order relation on a planar lattice is known to be causal [11,13]. In particular, the following factorization holds for various causal Markov models postulated on planar lattices, such as Markov chains and second and third order Markov meshes [11]:

$$P(\mathcal{X}) = \prod_{s \in R} P(x_s | x_r, r \precsim s), \tag{3}$$

up to appropriately defining the behavior of $\mathcal{X}$ near the borders of the image.

Indeed, the proposed framework merges the key ideas of the two aforementioned models into a unique hierarchical Markov model combining a quadtree topology and an arbitrary total order relation on each layer of the quadtree for the joint fusion and supervised classification of multisensor and multiresolution images.

Specifically, we assume the same notations introduced above for the quadtree topology. To insert a set of images acquired by different HR or VHR sensors at distinct spatial resolutions on the same area in a quadtree, the requirement is that the pixel lattices of the input images should have a power-of-2 relation one with the other. This constraint can be easily satisfied with a minor resampling [64,65] and antialiasing filtering [3,29]. The fusion between images at different resolutions is naturally matched by hierarchical models, and the relations $\prec$,$\preceq$, and $\precsim$ are considered to be defined within each lattice in the quadtree separately. This suggests that every site $s \in S^\ell$ of every layer S^ℓ ($\ell = 1, 2, \ldots, L$) is connected to its parent $s^- \in S^{\ell-1}$ in the upper layer and to its past neighbors $\{r \in S^\ell : r \precsim s\}$ in the same layer, except for the root layer. A pixel in the root S^0 has no parent but only its past neighbors, if present. The multiresolution and multisensor input images are inserted in the quadtree according to their spatial resolutions, and each $s \in S$ is linked to a continuous feature vector y_s collecting the corresponding image data. Input sensors that output data with the same resolution contribute to the feature vector in the same layer. $\mathcal{Y} = \{y_s\}_{s \in S}$ is the random field of the observations. $\mathcal{X} = \{x_s\}_{s \in S}$ indicates again the random field of the class labels $x_s \in \Omega$ associated with all sites $s \in S$ in the quadtree. The following assumptions define the proposed framework.

Assumption 1. *$\mathcal{X}$ satisfies the Markovian property across the scales as in Equation (1). Furthermore, on every layer except the root ($\ell = 1, 2, \ldots, L$), the following proportionality and factorization hold valid:*

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

Assumption 2. *On the root layer, $\mathcal{X}^0$ satisfies Equation (3):*

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

Assumption 3. *The observations are conditionally independent given the class labels:*

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

Assumption 1 defines the rationale of the proposed framework involving a hierarchical MRF on a quadtree and a causal planar Markov model built on a total order relation $\preceq$, implying that the Markovian property holds across the scales and that the transition probabilities can, in turn, be factorized to model both spatial-contextual and multiscale dependencies. Assumption 2 ensures that the spatial factorization is also valid on the root layer. The conditional independence shown in Assumption 3 is normally accepted when using MRFs for image classification and it is used to favor tractability [7,10,14,66].

3.1.2. Methodological Properties

The proposed model is methodologically defined by Assumptions 1–3. Here, we analytically prove two properties of this model. First, we note that the hierarchical MRF in Equation (1) and the MRF on a planar lattice with respect to a total order relation in Equation (2) are well-known to be causal, but this property is not *a-priori* guaranteed for the proposed framework. Indeed, causality holds for the developed framework as well, as indicated by the following theorem.

Theorem 1. *According to Assumptions 1–3, the joint probability distribution of the class labels and the feature vectors in the whole quadtree, i.e., the joint distribution of the random vector $(\mathcal{X}, \mathcal{Y})$, is completely determined by the parent–child transition probabilities $\{P(x_s|x_{s^-})\}_{s \in S \setminus S^0}$, the past neighbor transition probabilities $\{P(x_s|x_r, r \precsim s)\}_{s \in S}$, and the pixelwise data conditional likelihoods $\{p(y_s|x_s)\}_{s \in S}$.*

The causality is expressed through the factorization of the joint distribution of the feature vectors and the labels as causal transition probabilities. Thanks to Theorem 1, it is possible to conclude that $(\mathcal{X}, \mathcal{Y})$ is a Markov chain with respect to the causal topology introduced in this paper and associated with the quadtree and with the total order relation defined on each one of its layers. The proof of Theorem 1 can be found in Appendix A.1.

As anticipated in Section 1, causality allows for an efficient recursive implementation of the MPM criterion. To formulate this criterion, three further assumptions are introduced. For each $s \in S$, denote as $\mathcal{C}_s$ the vector of the labels of all nodes linked to s (the "context" of s), i.e., $\mathcal{C}_s$ collects the labels of: (i) the sites in $\{s^-\} \cup \{r \in S^\ell : r \precsim s\}$ if s is not in the root ($s \in S \setminus S^0$), and (ii) the sites in $\{r \in S^0 : r \precsim s\}$ if $s \in S^0$ is in the root. Moreover, if a site s is not in the leaf layer ($s \in S \setminus S^L$), then it is possible to build the vector $\mathcal{D}_s$ of the feature vectors of all the sites that descend from s along all the lower layers of the quadtree. If s in the leaf layer ($s \in S^L$), then we set $\mathcal{D}_s = y_s$.

Assumption 4. *The label of each site, when conditioned to those of its connected sites, depends only on the feature vectors of its descendants and not on the feature vectors of the other sites in the quadtree:*

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

Assumption 5. *The parent and neighbor labels of each site, when conditioned to the field of the feature vectors, are conditionally independent:*

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

Assumption 6. *The parent and neighbor labels of each site, when conditioned to its own label, are mutually independent and independent on the observations of its descendants:*

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

Assumptions 4–6 are conditional independence hypotheses and, similar to numerous analogous assumptions that are often made when operating with either hierarchical [10] or planar MRFs [7], they are accepted here to make the inference based on the MPM criterion analytically tractable. This criterion is advantageous for hierarchical MRFs since it considers the errors based on the scale where they occur, avoiding the accumulation of errors along the layers [10]. According to MPM, a pixel $s \in S$ is labelled as belonging to the class $\omega \in \Omega$ that yields the maximum value of $P\{x_s = \omega|\mathcal{Y}\}$ [7].

Theorem 2. *According to Assumptions 1–6, for each site $s \in S \setminus S^0$, i.e., not in the root layer:*

$$P(x_s) = \sum_{x_{s^-} \in \Omega} P(x_s|x_{s^-}) P(x_{s^-}), \tag{10}$$

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s) P(x_s|x_{s^-}) P(x_{s^-}) P(x_s)^{-|\mathcal{C}_s|} \cdot \prod_{r \precsim s} P(x_s|x_r) P(x_r), \tag{11}$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{D}_s) P(x_{s^-}|\mathcal{Y}) \prod_{r \precsim s} P(x_r|\mathcal{Y}), \tag{12}$$

where in Equation (10), the summation is over all possible class labels assigned to site s^-; in Equation (11), $|B|$ indicates the cardinality of a finite set B, and the labels x_{s^-} and $x_r, r \precsim s$, are fixed to the values they take in the context $\mathcal{C}_s$; and in Equation (12), the summation is over all possible label configurations in the context $\mathcal{C}_s$ of site s, i.e., over all possible labels of its parent and past neighbors.

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

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s) P(x_s)^{-|\mathcal{C}_s|} \prod_{r \precsim s} P(x_s|x_r) P(x_r), \tag{13}$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{D}_s) \prod_{r \precsim s} P(x_r|\mathcal{Y}). \tag{14}$$

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

$$P(x_s|\mathcal{D}_s) \propto P(x_s|y_s) \prod_{t \in s^+} \sum_{x_t \in \Omega} \frac{P(x_t|\mathcal{D}_s) P(x_t|x_s)}{P(x_t)}. \tag{15}$$

Proof of Theorem 2 can be found in Appendix A.2. According to Theorem 2, the MPM criterion can be formulated with three recursive steps. First of all, after the initialization of the prior probabilities $P(x_s)$ on the root layer, Equation (10) is used to obtain the prior probabilities of all classes on all the other sites of the quadtree using a top-down pass from the root down to the leaves. Various initialization strategies are possible; here the prior probabilities $P(x_s)$ are assumed independent with respect to the site s considered and are estimated at the root level ($s \in S^0$) using the relative frequencies of the classes in the training set considered. Secondly, Equations (15), (11), and (13) are used to compute $P(x_s|\mathcal{C}_s, \mathcal{D}_s)$

using a bottom-up pass from the leaves up to the root and within the root layer. Lastly, Equations (12) and (14) are used to obtain $P(x_s|\mathcal{Y})$ through a second top-down pass.

The feature vectors enter the recursions through the pixelwise posterior probabilities $P(x_s|y_s)$ in Equation (15). As mentioned above, the fusion of multisensor observations coming at different spatial resolutions is ensured by the quadtree topology intrinsically. In the proposed framework, the pixelwise posteriors $P(x_s|y_s)$ are estimated using decision tree ensembles, such as random forest [23], rotation forest [24], ExtraTrees [25], and GBRT [22]. Data in different layers of the quadtree are associated with the observations at the corresponding resolution: the information collected by two or more sensors at the same resolution contributes to the feature vector in the same layer. Accordingly, this stacked vector may be highly heterogeneous. Decision tree ensembles, with their fully non-parametric formulation, are exploited to perform the estimation of the pixelwise posteriors, because they can flexibly characterize data with arbitrary statistics and are applicable to modeling the class-wise behavior of heterogeneous features. The integration of this ensemble approach for the $P(x_s|y_s)$ terms of the recursive inference of the proposed framework guarantees the fusion of multisensor data associated with the same resolution as well (further details on the role of the decision tree ensembles in the proposed approach can be found in Section 3.2.3).

3.2. The Proposed Multimodal Classification Methods

Two classification algorithms are developed within the proposed causal hierarchical Markov framework. They differ in the choice of the spatial model associated with the individual layers of the quadtree, which corresponds, in turn, to distinct definitions of the total order relation $\preceq$ used to formalize spatial Markovianity. This analytical difference in the PGM structure leads to different algorithmic formulations.

3.2.1. Markov Mesh-Based Classification Algorithm

Markov mesh random fields (MMRFs) are an important class of causal MRFs built upon an order relation $\preceq$ on a rectangular lattice $R \subset \mathbb{Z}^2$ to account for the spatial relationship between pixels in the lattice. In the case of an MMRF, $r \prec s$ if r precedes s while the image is raster scanned, i.e., the past of a site s is the set of all pixels crossed by a raster scan before getting to s (see Figure 2b). In particular, in the case of a second-order Markov mesh, $r \precsim s$ if and only if r is adjacent to s in its previous row and column in the scan. Denoting as i and j the pixel coordinates in R, this means that $r \precsim s = (i, j)$ if and only if either $r = (i-1, j)$ or $r = (i, j-1)$. In the case of a third-order Markov mesh, the three pixels adjacent to s and located in its previous row and column are its past neighbors, i.e., $r \precsim s = (i, j)$ if and only if $r \in \{(i-1, j), (i, j-1), (i-1, j-1)\}$. These definitions of the past neighborhood are obviously adapted to the image borders. The factorization in Equation (3) holds for Markov meshes of both orders [11].

In the first proposed algorithm, a Markov mesh model is assumed for the spatial information in each level (or layer) of the quadtree. Accordingly, the MPM equations are applied using the aforementioned definition of the $\precsim$ relation. In this respect, the intrinsic directionality of this definition may generally favor anisotropic artifacts in the output classification map—a common drawback of causal Markov model applications [13]. To prevent these artifacts, the symmetric MMRF (SMMRF) formulation defined in [15] is used in the first proposed algorithm (see Figure 3). As far as a planar lattice is concerned, it is based on a pixel visiting trajectory that ensures corner independence [15]. In the scanning trajectory presented in Figure 3, four rows are scanned at each step, two from the bottom and two from the top of the image, until all the rows are scanned. Consequently, the dependencies of pixels on all the neighboring sites are taken into consideration, guaranteeing an isotropic pixel visiting scheme [15]. In the proposed method, this scan is used within each layer of the quadtree separately. Accordingly, the transition probabilities of the Markov chain in Equation (4) take into account the Markov mesh dependencies, so that the components of $\mathcal{X}^\ell$ convey both the parent–child and the spatial dependencies, as shown in Figure 2c.

Operatively, applying the MPM formulation proved for the proposed framework in the first developed algorithm requires specifying both the pixelwise posteriors $P(x_s|y_s)$ and the transition probabilities $P(x_s|x_{s^-})$ and $P(x_s|x_r), r \precsim s$, between the label of a site and the labels of its parent and past neighbors. A parametric modeling approach is used for these transition probabilities, based on the formulation introduced by Bouman and Shapiro in [67]. This means that the transition probability across consecutive scales $P(x_s|x_{s^-})$ is modeled so that: (i) the probability that a site and its parent share the same label is the same for all classes and all site locations, i.e., $P\{x_s = \omega|x_{s^-} = \omega\} = \theta$ for all $\omega \in \Omega$ and all $s \in S \setminus S^0$, where $\theta \in [0,1]$ is a parameter of the method; and (ii) the probability that a site and its parent have different labels does not depend on their class memberships and on their location, i.e., $P\{x_s = \omega|x_{s^-} = \omega'\}$ is constant over all $\omega \neq \omega'$ $(\omega, \omega' \in \Omega)$ and all $s \in S \setminus S^0$. An analogous model is adopted for the causal neighborhood transition probability $P(x_s|x_r), r \precsim s$, on the same pixel grid, using a second parameter $\phi \in [0,1]$ that represents the common value of all $P\{x_s = \omega|x_r = \omega\}$ for $\omega \in \Omega$ and $r \precsim s$. More details can be found in [67,68]. Regarding the parametric models used for the transition probabilities, we recall that, in the context of hierarchical MRF modeling for multitemporal image classification, as shown by the results obtained in [68], the corresponding parameters did not influence significantly the results.

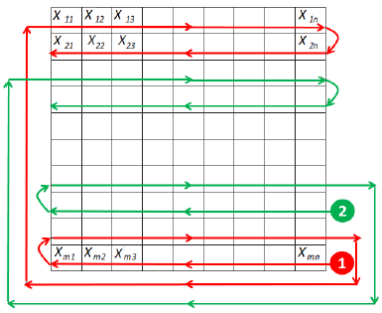


Figure 3. Visiting scheme of the SMMRF adopted in each layer of the first proposed method [18].

3.2.2. Markov Chain-Based Classification Algorithm

In the second developed algorithm, the spatial-contextual model associated with each layer S^ℓ of the quadtree ($\ell = 0, 1, \ldots, L$) is a Markov chain (MC). Consider a scan trajectory on S^ℓ, i.e., a sequence moving to one pixel from one of its neighbors, visiting every pixel once. Given a pixel $s \in S^\ell$, the previous pixel in the scan is indicated as $s^\star \in S^\ell$. In the second developed algorithm, $r \precsim s$ if and only if $r = s^\star$. Therefore, the spatial Markovianity condition in Equation (4) implies that $\mathcal{X}^\ell$ is a first-order Markov chain for the aforementioned 1D scan trajectory. Accordingly, Equations (4) and (5) define the spatial behavior of the whole random field $\mathcal{X}$ as a Markov chain of first-order. Specifically, Equations (11) and (12) result in the following special cases ($s \in S \setminus S^0$):

$$P(x_s|x_{s^-}, x_{s^\star}, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_s)^{-2} \cdot P(x_s|x_{s^-})P(x_{s^-})P(x_s|x_{s^\star})P(x_{s^\star}), \tag{16}$$

$$P(x_s|\mathcal{Y}) = \sum_{x_{s^-}, x_{s^\star} \in \Omega} P(x_s|x_{s^-}, x_{s^\star}, \mathcal{D}_s)P(x_{s^-}|\mathcal{Y})P(x_{s^\star}|\mathcal{Y}), \tag{17}$$

and Equations (13) and (14) simplify analogously.

The formulation of the second developed algorithm requires the definition of the scan trajectory to be used within each layer of the quadtree. Two popular pixel visiting schemes are the zig-zag path and the Hilbert space-filling curve [11]. A zig-zag scan over the lattice S^ℓ can be performed as illustrated in Figure 4b. Apart from the sites at the border of the pixel grid, $s^\star$ is the pixel adjacent to s and located diagonally from s in the direction of a zig-zag. The visiting scheme follows the image borders on the perimeter. The other scan makes use of the Hilbert space-filling curve on a power-of-2 sized grid S^ℓ (see Figure 4a).

The Hilbert curve is especially interesting because, while it maps a 1D to a 2D spaces, it preserves locality, an important property for spatial-contextual labeling: two points that are near each other in the 1D path remain near each other after folding. The opposite may generally be violated. Furthermore, the power-of-2 relation among the layer sizes in the quadtree is matched by the recursive construction of the Hilbert curve.

In the second developed method, to avoid the presence of anisotropic "corner-dependent" artifacts, a symmetric scan strategy is employed again. In fact, when using asymmetric scanning, directional artifacts in the shapes and edges of the regions in the classification maps generally occur [69]. Specifically, the scan used in the proposed method combines two Hilbert and four zig-zag paths, as shown in Figure 4c. Each pixel is reached multiple times in a symmetric way through a visiting scheme that involves two different directions of the Hilbert curve and zig-zag path along the two diagonals, in both directions, thus preventing any geometrical artifacts. The scanning trajectory is divided in six steps: in steps 1 and 2 the pixels are visited along a zig-zag curve on the main diagonal from the two opposite ends to the other, subsequently. Step 3 consists of a clockwise scan along the Hilbert curve. Steps 4 and 5 are symmetrical to steps 1 and 2, following the direction of the antidiagonal. Finally, step 6 involves moving anticlockwise onto the Hilbert curve.

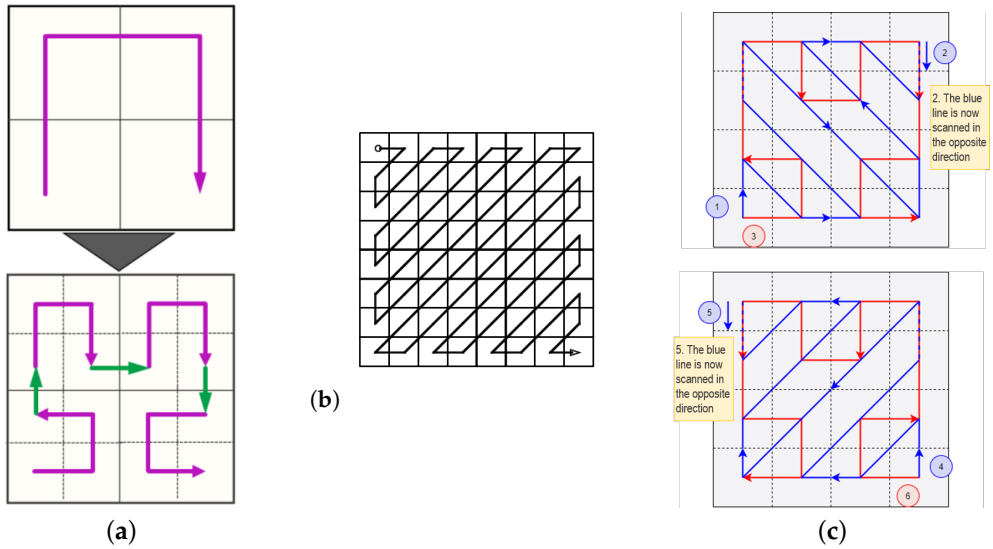


Figure 4. (**a**) Hilbert space-filling curve, (**b**) zig-zag scan, (**c**) symmetrized scan introduced for the second proposed algorithm through the combination of two Hilbert curve scans and four zig-zag paths.

Regarding the modeling of the transition probabilities $P(x_s|x_{s^-})$ and $P(x_s|x_{s^\star})$, the same approach as in the case of the first developed method is used, i.e., parametric Bouman–Shapiro models are adopted. In the case of the second method as well, these models involve two parameters θ and ϕ that represent the common values of all same-label transition probabilities $P\{x_s = \omega|x_{s^-} = \omega\} = \theta$ and $P\{x_s = \omega|x_{s^\star} = \omega\} = \phi$ ($\omega \in \Omega$; $\theta, \phi \in [0,1]$).

3.2.3. Role of Decision Tree Ensembles

The fusion of multiresolution data in the MPM scheme is performed by ensembles of decision trees [22]. In general, any classifier predicting pixelwise posterior probabilities $P(x_s|y_s)$ can be used in conjunction with the proposed model. Here, random forest (RanF) [23], rotation forest (RotF) [24], ExtraTrees (ET) [25], and gradient boosted regression trees (GBRT) [22] are used for the estimation of these pixelwise posteriors, since their non-parametric formulation is particularly advantageous for the integration of multisensor data into the proposed hierarchical Markov model [17]. Furthermore, the formulation in terms of pixelwise posteriors does not require models for the class-conditional PDF of the feature vectors to be defined.

Decision tree ensembles are a collection of decision trees that make independent predictions for test samples, while their outputs are aggregated to estimate the pixelwise posterior distribution. GBRT uses boosting [70] to adaptively and iteratively manipulate the training samples for building decision trees [22]. Each training sample is assigned a weight, minimized at each iteration while a decision tree is constructed. The final predictor is determined by a weighted vote of the single trees, depending on their accuracy. For the random forest, instead, each individual tree is trained on a bootstrap sample drawn from the training set (resampling the input data with replacement), resulting in different trees [23]. This process is known as bagging, and the same training sample may be drawn multiple times, while other training samples may never be selected. Random forest combines bagging and random feature-subset selection to favor independence among the trees in the ensemble. An unknown sample is classified by all the trees of the random forest and each tree casts a vote for a class membership. The class with the highest score will be the one finally selected. The ExtraTrees classifier differs from random forest in the selection of the split points in the various nodes of the trees and in the use of the training samples [25]: while in random forest, the optimum split is chosen in a random subset of features, in ExtraTrees this split is selected randomly; while random forest is trained with bootstrap replicas, ExtraTrees is trained with the whole training set available. Rotation forest is based on the idea of encouraging individual accuracy and diversity within the decision trees of the ensemble, by creating N random subsets of the training data and applying principal component analysis (PCA) to each one of them [24]. All the principal components are retained in order to preserve the variance of the data and N axis rotations take place to form the new features for a base classifier. All these four ensemble methods also provide estimates of the pixelwise posterior probabilities. Further algorithm details can be found in [22–25].

Previous work showed that the use of the bagging and boosting techniques generally achieves higher accuracy and/or enhanced robustness to overfitting than using single decision tree classifiers [71,72], as well as more stability and improved robustness to noise in the training set [73].

4. Experimental Results

4.1. Dataset and Experimental Setup

The experimental validation was carried out with two VHR datasets. The first one consisted of images taken over Port-au-Prince, Haiti, right after the 2010 earthquake (see Figure 5). This dataset is made up of a GeoEye-1 image with the resolution of 1.25 m and RGB channels, a QuickBird image with the resolution of 2.5 m and RGB-near infrared (NIR) channels, and a COSMO-SkyMed stripmap image (X-band SAR) at the resolution of 5 m. For the COSMO-SkyMed stripmap modality, the original resolution is 5 m, while for the multispectral channels of QuickBird and for the multispectral and panchromatic bands of GeoEye-1, whose resolutions are 2.4 m, 1.84 m, and 0.46 m, respectively, a minor resampling was applied. Proper antialiasing filtering was used for this purpose. For the GeoEye-1 image, the resampling was applied after the pan-sharpening to obtain the 1.25 m resolution image with a pixel grid of 1040 × 1360 pixels.

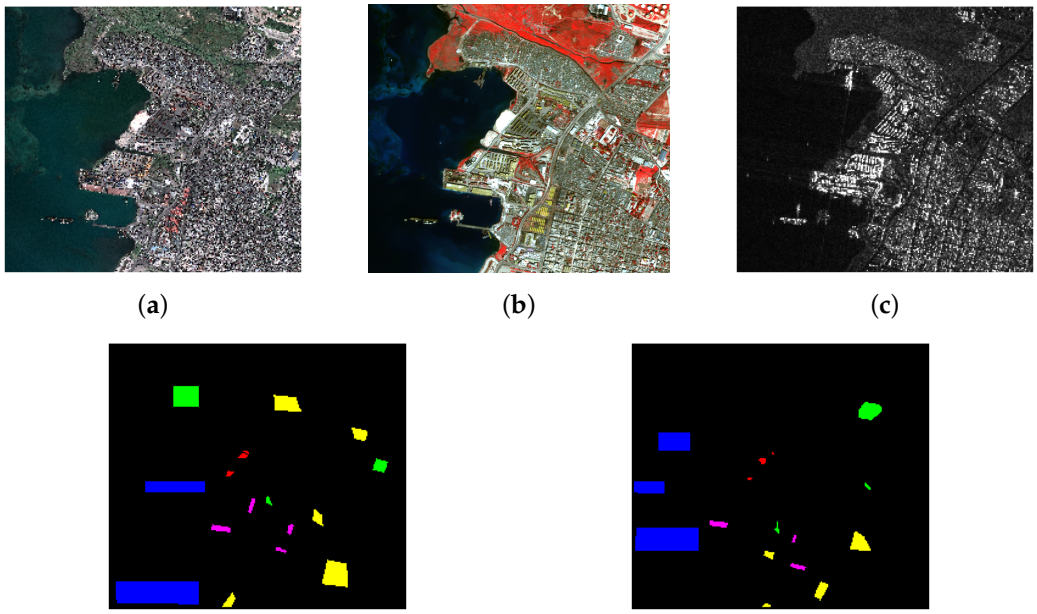


Figure 5. "Haiti" dataset: (**a**) 1.25 m resolution GeoEye-1 image (RGB true color composition), (**b**) 2.5 m resolution QuickBird image (RGB false color composition), (**c**) 5 m resolution synthetic aperture radar (SAR) COSMO-SkyMed stripmap image, (**d**) training map, and (**e**) test map. Color legend for the classes: containers, vegetation, asphalt, buildings, water.

The second dataset consisted of images collected by the IKONOS multiresolution optical sensor over the city of Alessandria, Italy, in 2004 (see Figure 6). It comprises a panchromatic image with the resolution of 1 m (1260 × 1400 pixels, single-channel) and a multispectral image at the resolution of 4 m (315 × 350 pixels, RGB-NIR channels). This dataset was collected in the framework of a case study of flood vulnerability assessment [27].

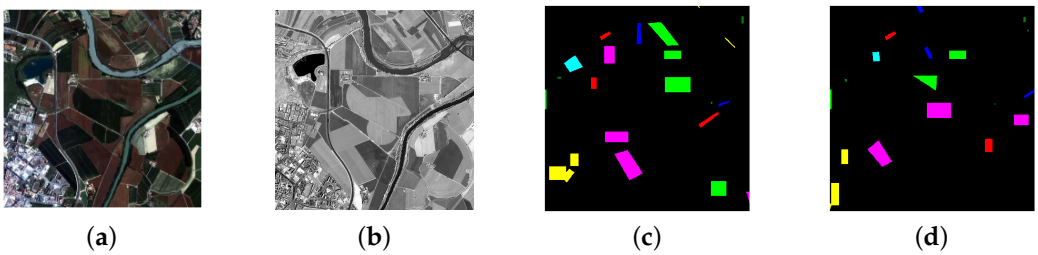


Figure 6. "Alessandria dataset": (**a**) RGB true-color composition of the IKONOS multispectral channels, (**b**) panchromatic channel, (**c**) training map, and (**d**) test map. Color legend for the classes: urban, agricultural, rangeland, bare soil, forest, water, wet soil.

A quadtree with three layers was employed for both datasets. For the IKONOS dataset, it contained the panchromatic image on the leaf layer and the multispectral image on the root layer (the resolutions of these two images were natively in a power-of-2 relation). The intermediate layer was filled with the result of pansharpening given by the Gram–Schmidt technique [39] and resampling (again with antialiasing) to obtain a 2 m resolution image. For the Haiti dataset, the GeoEye-1, QuickBird, and COSMO-SkyMed images were inserted in the leaf, intermediate, and root layers, respectively, with the aforementioned minor resampling for the first two images.

Both datasets were manually annotated in homogeneous areas at the finest resolution of the quadtree by a photo-interpretation specialist. The resulting ground truth at the finest resolution was divided into two non-overlapping sets, the training set and the test set. As

usual in land cover mapping through remote sensing image classification, no class borders were taken into account in this ground truth to avoid mixed pixels in the training and test sets. Training and test maps corresponding to coarser resolutions in the quadtree were obtained by downsampling the training and test maps associated with the finest resolution. On each layer of the quadtree, the training set was used to train the decision tree ensembles and estimate the pixelwise posteriors accordingly, and the test set was used to perform quantitative accuracy assessment.

The proposed SMMRF-based classification algorithm was applied using a second-order mesh. The results of both proposed algorithms were compared to previous approaches to multiresolution and multisensor fusion. First of all, a planar MRF classifier was used. In this method, the unary potentials were computed according to the pixelwise predictions given by random forest after a resampling of the dataset to the finest resolution, the pairwise potentials were expressed by the well-known Potts model [66], and the minimization of the MRF energy was accomplished using sequential tree-reweighted message passing (TRW-S) [74]. TRW-S is a well-known approach from the broader family of belief propagation methods and its sequential formulation exhibits regular convergence properties [74]. This first comparison is aimed at evaluating the performances of the developed methods as compared to those of a spatial-contextual but single-resolution method applied after resampling.

The second approach considered for comparison was the one proposed in [75] for multiresolution optical image classification, based on noncausal MRFs, graph cuts, and Gaussian random fields. In this case, multiresolution fusion is accomplished at the feature level through a linear mixture model rather than through the probabilistic structure of the classifier as in the proposed algorithms. When applied to the COSMO-SkyMed image of the Haiti dataset, it was adapted with a log-normal model for the class-conditional PDF, which is often accepted for the statistics of SAR data. We note that both benchmark techniques produce classification maps at the finest observed resolution. On the contrary, thanks to the MPM formulation on the quadtree, the proposed algorithms generate classification results at all observed resolutions, which is generally a desirable property when input multiresolution imagery is concerned.

4.2. Results and Performances

A quantitative and qualitative (visual) analysis of the results is possible based on the reported figures and tables. For each dataset, the accuracy parameters refer to the corresponding test set.

The results for the Haiti dataset are shown in Figure 7, including the maps generated by the algorithms developed here within the proposed framework (see Figure 7b,c for the SMMRF- and MC-based algorithms, respectively) and those obtained by the aforementioned benchmark techniques (see Figure 7d,e for the single-resolution MRF and the adaptation of the method in [75], respectively). Five land cover classes were present in the dataset: asphalt (yellow), containers (red), vegetation (green), buildings (magenta), and water (blue). The quantitative results of the experimentation conducted on this dataset are presented in two separate tables. Table 1 shows the general behavior of the two proposed algorithms in terms of overall accuracy (OA) and Cohen's κ, when applied in conjunction with the different decision tree ensembles. Following these results, which highlight that the highest performances were attained using GBRT, Table 2 focuses in more detail on the class-wise accuracies obtained on the test set by the two proposed methods when paired with this decision tree ensemble. The class-wise accuracies in the cases of the other ensembles were analogous and are not shown for brevity.

Table 1. "Haiti" dataset: OA and Cohen's κ with respect to the test set, for the proposed methods employing SMMRF and Markov chain (MC), when combined with random forest (RanF), rotation forest (RotF), ExtraTrees (ET), and gradient boosted regression trees (GBRT).

Resolution 1.25 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	96.06	94.48	95.07	**96.36**
	Cohen's κ coeff	0.9047	0.8613	0.8786	**0.9124**
Proposed-MC	OA %	96.80	96.37	**97.04**	96.90
	Cohen's κ coeff	0.9026	0.8868	**0.9098**	0.9057
Resolution 2.5 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	96.23	94.47	96.13	**96.63**
	Cohen's κ coeff	0.9174	0.8691	0.9148	**0.9276**
Proposed-MC	OA %	96.77	95.85	96.81	**97.00**
	Cohen's κ coeff	0.9139	0.8800	0.9150	**0.9218**
Resolution 5 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	90.00	85.81	87.58	**96.01**
	Cohen's κ coeff	0.7680	0.6561	0.7035	**0.9288**
Proposed-MC	OA %	90.97	87.92	88.78	**96.01**
	Cohen's κ coeff	0.7347	0.6284	0.6588	**0.9288**

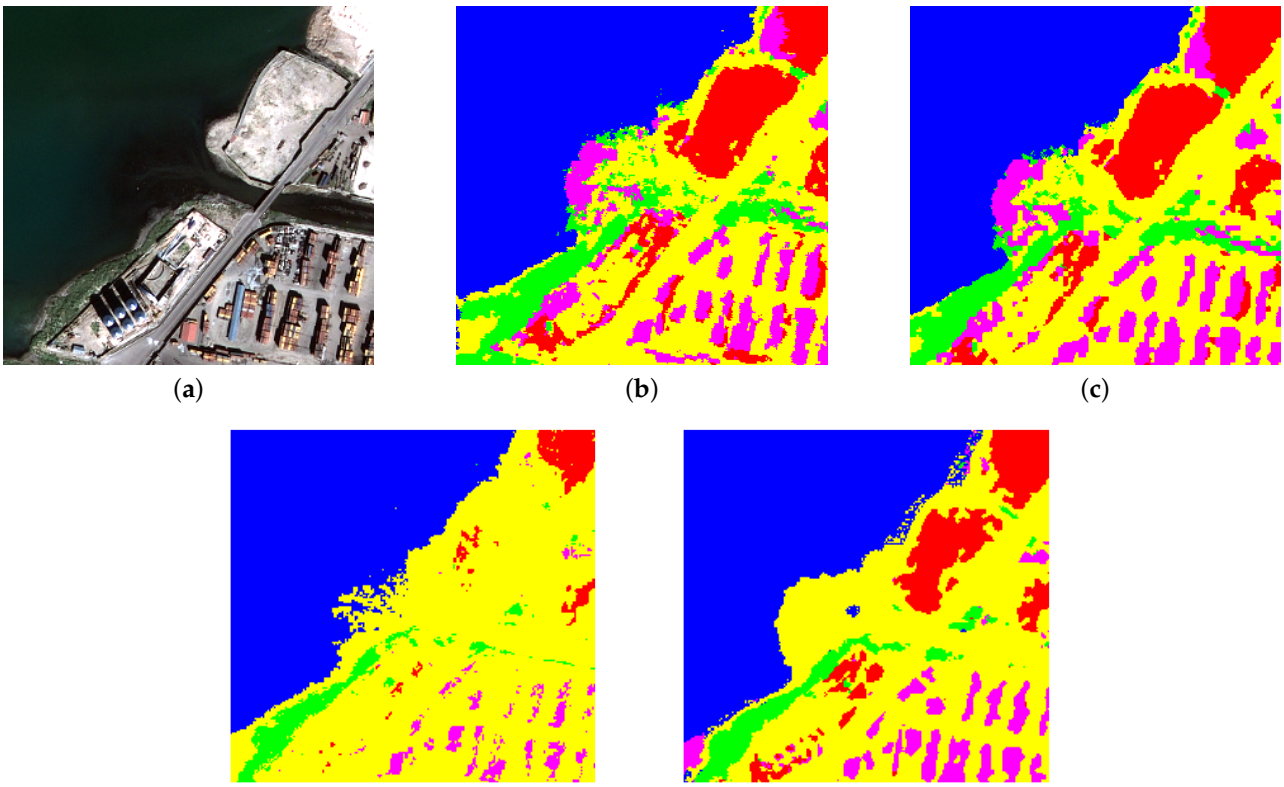
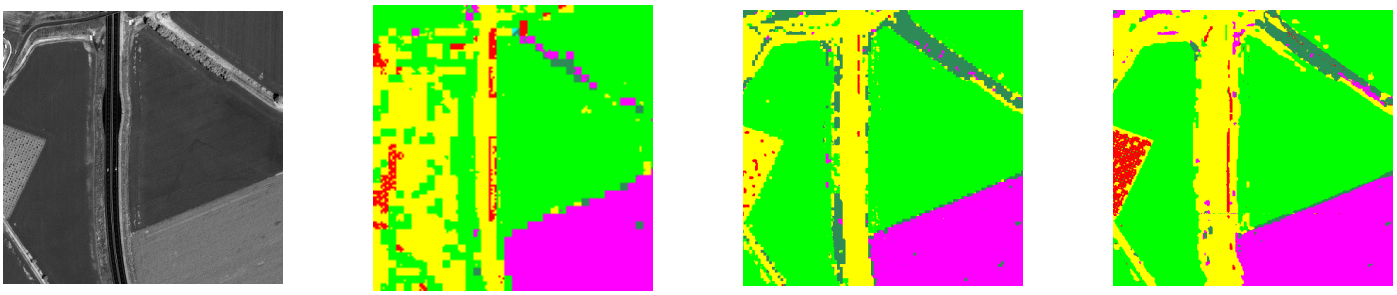


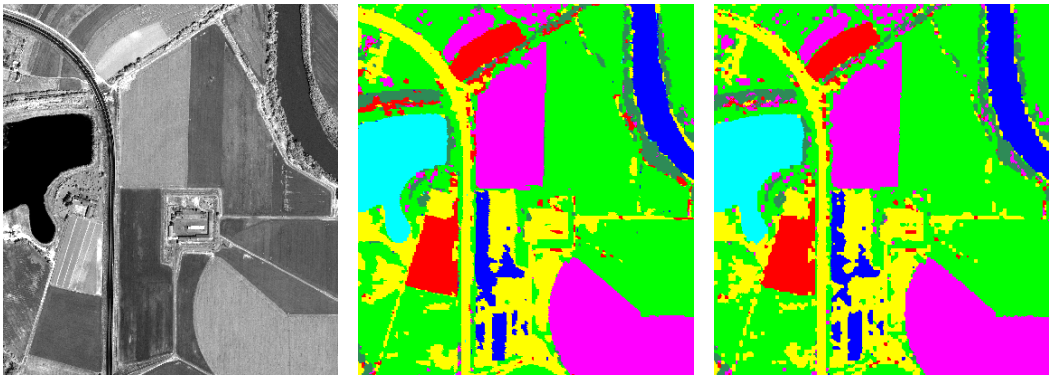
Figure 7. "Haiti" dataset: details of (**a**) the 1.25 m resolution image and of the classification maps generated by (**b**) the proposed SMMRF-based algorithm with GBRT, (**c**) the proposed MC-based algorithm with the zig-zag and Hilbert curve scan and GBRT, (**d**) the single-resolution MRF-based classifier after resampling, and (**e**) the adaptation of the technique in [75] to the case of optical-SAR data. Color legend for the classes: containers, vegetation, asphalt, buildings, water.

Regarding the classification maps obtained from the IKONOS dataset, crops on spatial details are shown in Figures 8 and 9. In particular, in Figure 9 it is possible to see the comparison between the results of the two proposed algorithms in the case of a spatially detailed test image. In this dataset there are seven classes: urban (yellow), bare soil (red), rangeland (green), forest (dark green), agricultural (magenta), water (blue), and wet soil (cyan). The classification accuracies, on the other hand, are reported in Table 3, where the overall behavior of the two algorithms is presented as a function of the tree ensembles involved, and in Table 4, where class-wise accuracies and comparisons with benchmark methods are shown while focusing on the use of a single ensemble method (again, the results obtained using the other ensemble methods are similar and omitted for brevity).



(**a**) (**b**) (**c**) (**d**)

Figure 8. "Alessandria" dataset: details of (**a**) the 1 m resolution panchromatic image and of the classification maps generated by (**b**) the single-resolution MRF-based classifier after resampling, (**c**) the algorithm in [75], and (**d**) the proposed SMMRF-based method with GBRT. Class legend: urban, agricultural, rangeland, bare soil, forest, water, wet soil. "Water" is not present in this detail.



(**a**) (**b**) (**c**)

Figure 9. "Alessandria" dataset: details of (**a**) the 1 m resolution panchromatic image and the classification maps generated by (**b**) the proposed SMMRF-based algorithm with GBRT, and (**c**) the proposed MC-based algorithm with the zig-zag and Hilbert curve scan and GBRT. Class legend: urban, agricultural, rangeland, bare soil, forest, water, wet soil.

Table 2. "Haiti" dataset: class-wise accuracies of the proposed algorithms, combined with GBRT and compared with benchmark methods, with respect to the test set.

Resolution 1.25 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	**87.08**	33.27	95.17	99.04	97.18	96.36	**0.9124**
Proposed-MC-GBRT	86.97	34.57	97.82	**99.36**	**100**	**96.90**	0.9057
Single-res. MRF after resampling	63.97	74.81	**98.64**	99.30	76.24	94.58	0.888
Adaptation of [75]	75.06	**85.32**	95.89	97.75	99.36	95.11	0.901
Resolution 2.5 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	86.31	32.71	94.87	**100**	96.04	96.63	**0.9276**
Proposed-MC-GBRT	**86.92**	**33.08**	**97.29**	**100**	**100**	**97.00**	0.9218
Resolution 5 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	**87.98**	**25.39**	**96.55**	**100**	**88.88**	**96.01**	**0.9288**
Proposed-MC-GBRT	**87.98**	**25.39**	**96.55**	**100**	**88.88**	**96.01**	**0.9288**

The proposed approach attained high accuracies on both datasets, suggesting the effectiveness of the developed methods. The two proposed techniques differ in the definition of spatial-contextual information found in each layer of the quadtree. However, it is noticeable from the quantitative results that the accuracies obtained with the two algorithms are similar, in some cases even identical, consistently with the fact that both techniques are special cases of the causal hierarchical PGM framework formulated in this paper. This is further demonstrated by Table 5, which focuses on the "Haiti" dataset and on the class-wise performances of the SMMRF-based algorithm, when paired with all of the considered decision tree ensembles. For the sake of brevity and to avoid redundancy, we focus in this table on the SMMRF-based method and on "Haiti" because the behavior in the case of the MC-based algorithm and of "Alessandria" is analogous.

A discussion of the experimental results and of their comparison with those of the benchmark techniques can be found in Section 5.

Table 3. "Alessandria" dataset: OA and Cohen's κ with respect to the test set for the proposed methods employing SMMRF and MC, when combined with random forest (RanF), rotation forest (RotF), ExtraTrees (ET), and GBRT.

Resolution 1 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	97.84	98.34	98.77	**99.12**
	Cohen's κ coeff	0.9749	0.9818	0.9879	**0.9976**
Proposed-MC	OA %	97.52	97.10	98.44	**98.58**
	Cohen's κ coeff	0.9704	0.9647	0.9833	**0.9853**
Resolution 2 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	97.45	96.65	**98.31**	98.01
	Cohen's κ coeff	0.9745	0.9635	**0.9867**	0.9825
Proposed-MC	OA %	97.23	96.61	**98.09**	98.03
	Cohen's κ coeff	0.9715	0.9629	**0.9836**	0.9829
Resolution 4 m		**RanF**	**RotF**	**ET**	**GBRT**
Proposed-SMMRF	OA %	96.25	94.79	**96.95**	96.36
	Cohen's κ coeff	0.9682	0.9477	**0.9781**	0.9697
Proposed-MC	OA %	96.25	94.75	**96.95**	96.36
	Cohen's κ coeff	0.9682	0.9477	**0.9782**	0.9697

Table 4. "Alessandria" dataset: class-wise accuracies of the proposed algorithms, applied with GBRT and compared with benchmark methods, with respect to the test set.

Resolution 1 m	**Urban %**	**Agriculture %**	**Rangeland %**	**Forest %**	**Water %**	**Wet Soil %**	**Bare Soil %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	**100**	99.07	99.65	**100**	99.65	**100**	**100**	**99.12**	**0.9976**
Proposed-MC-GBRT	99.54	98.26	**99.91**	86.40	99.80	**100**	**100**	98.58	0.9853
Single-res. MRF after resampling	98.58	99.12	92.23	36.98	**100**	98.30	96.82	96.03	0.9453
Method in [75]	99.70	**99.21**	97.34	64.92	**100**	**100**	99.66	98.60	0.9804
Resolution 2 m	**Urban %**	**Agriculture %**	**Rangeland %**	**Forest %**	**Water %**	**Wet Soil %**	**Bare Soil %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	**99.54**	98.06	**99.91**	78.74	99.53	**100**	**100**	98.01	0.9825
Proposed-MC-GBRT	99.49	**98.08**	**99.91**	**79.52**	**99.84**	**100**	**100**	**98.03**	**0.9829**
Resolution 4 m	**Urban %**	**Agriculture %**	**Rangeland %**	**Forest %**	**Water %**	**Wet Soil %**	**Bare Soil %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-GBRT	**99.09**	**96.81**	**99.82**	**64.86**	**98.72**	**100**	**100**	**96.36**	**0.9697**
Proposed-MC-GBRT	**99.09**	**96.81**	**99.82**	**64.86**	**98.72**	**100**	**100**	**96.36**	**0.9697**

Table 5. "Haiti" dataset: test-set class-wise accuracies of the proposed SMMRF-based method as a function of the adopted decision tree ensemble.

Resolution 1.25 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-RanF	85.76	34.29	95.03	98.86	95.35	96.06	0.9047
Proposed-SMMRF-RotF	83.37	**38.75**	81.51	**99.52**	67.73	94.48	0.8613
Proposed-SMMRF-ET	82.29	37.63	88.75	98.77	91.89	95.07	0.8786
Proposed-SMMRF-GBRT	**87.08**	33.27	**95.17**	99.04	**97.18**	**96.36**	**0.9124**
Resolution 2.5 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-RanF	85.95	27.51	95.32	99.58	95.85	96.23	0.9174
Proposed-SMMRF-RotF	76.85	**32.71**	81.18	**100**	92.29	94.47	0.8691
Proposed-SMMRF-ET	84.76	**32.71**	**95.77**	99.44	95.45	96.13	0.9148
Proposed-SMMRF-GBRT	**86.31**	**32.71**	94.87	**100**	**96.04**	**96.63**	**0.9276**
Resolution 5 m	**Container %**	**Vegetation %**	**Asphalt %**	**Buildings %**	**Water %**	**OA %**	**Cohen's κ Coeff**
Proposed-SMMRF-RanF	73.32	15.87	84.40	97.28	32.54	90.00	0.7680
Proposed-SMMRF-RotF	45.91	14.28	65.60	97.26	28.57	85.81	0.6561
Proposed-SMMRF-ET	57.93	20.63	72.24	97.23	31.74	87.58	0.7035
Proposed-SMMRF-GBRT	**87.98**	**25.39**	**96.55**	**100**	**88.88**	**96.01**	**0.9288**

5. Discussion

A visual analysis of the classification maps generated by applying the developed methods to the considered datasets highlights a remarkable visual regularity, obtained thanks to both the related multiresolution graph and the modeling of the contextual spatial information through planar Markov models at each layer (see Figures 7–9).

Regardless of the strong spectral overlapping present among several classes (e.g., the vegetated land covers), in particular for the IKONOS dataset, the proposed model attained high test-set accuracies for all the levels of the quadtree. Tables 1 and 3 show the results with the images taken over Port-au-Prince, Haiti, and Alessandria, Italy, respectively, for the two proposed algorithms. These results suggest the flexibility of the proposed framework in integrating pixelwise predictions from distinct tree ensembles. In terms of class-wise accuracy, both proposed algorithms obtained high values for all classes in the Alessandria dataset and all classes except "vegetation" in the Haiti dataset. To improve the discrimination of "vegetation" in the latter dataset, additional training samples could be added to better represent this land cover as it appears in different areas of the imaged scene.

The highest accuracies were achieved using the proposed algorithms paired with GBRT (and, in some cases, with ExtraTrees or random forest). In the case of both datasets, rather similar results were obtained at the finest and the intermediate resolutions (1.25 m and 2.5 m for "Haiti", and 1 m and 2 m for "Alessandria"). A remarkable improvement ($\geq$6%) was attained using GBRT as compared to the other ensemble methods at resolution equal to 5 m in the case of the Haiti dataset. These results suggest that, while the framework is flexible to incorporating the predictions of arbitrary ensemble methods and accurate maps were obtained in the case of all four ensembles considered, GBRT happens to be an especially appropriate choice within the applications associated with the considered datasets.

The proposed method surpasses in accuracy the performances of the single-resolution MRF and the adaptation of the method in [75], as can be seen in Tables 2 and 4. Both these benchmark approaches discriminated most classes in quite a satisfactory manner, however with remarkably lower class-wise accuracies as compared to the proposed framework, for example, for "forest" in the "Alessandria" dataset. Furthermore, while the previous techniques perform the classification exclusively at the finest scale, the developed methods map at all the resolutions considered, i.e., the three layers of the quadtree in the considered case studies. These results indicate the effectiveness of the developed approach, based on a causal hierarchical PGM on a quadtree and on ensemble modeling, to address multiresolution and multisensor fusion, as compared to a classical resampling strategy and to a multiresolution model using a linear mixture rationale, such as in [75].

The two algorithms developed within the proposed framework attained similar results, in some cases even identical, as mentioned in Section 4. This matches the fact that both techniques, that make use either of a Markov chain or of an SMMRF within each layer, are special cases of a general model defined in this paper, characterized by a multiresolution causal hierarchical structure postulated on a quadtree and an arbitrary total order relation. This is shown in Tables 2 and 4 and in Figures 7–9. Consistently with this perspective, in general, large difference in accuracy between these two spatial-contextual models based on either Markov meshes or Markov chains incorporated in the proposed hierarchical framework on a quadtree, are not expected. At least in the case of the considered datasets, these two proposed approaches exhibited similar capabilities from the viewpoint of class discrimination, as confirmed by the results obtained.

However, the two approaches differ in terms of computational burden. The MC-based method involves looping over the three labels x_s, x_{s^-}, and $x_{s^\star}$, according to Equations (11) and (12) and to their special case in Equation (16), because each site s is connected to its parent site and to the preceding pixel in the 1D scan. Therefore, the computational burden is $O(|\Omega|^3)$. With regard to the proposed method using the SMMRF, the formulation of the mesh includes more neighboring pixels, which can lead, for example, to $O(|\Omega|^4)$ and $O(|\Omega|^5)$ in the case of second- and third-order meshes, respectively. Clearly, the smaller the subset of neighbors, the lower the computational cost of the inference.

6. Conclusions

A general framework based on causal hierarchical Markov modeling has been proposed for the complex problem of the joint classification of multimodal data including multisensor and multiresolution input components. The developed framework extends approaches formalized previously, by introducing Markovianity at each resolution through an arbitrary total order relation on a pixel lattice. Two classification algorithms have been developed within the proposed framework, by making use of different total orders and of correspondingly different spatial Markov models (a Markov mesh and a Markov chain along a specific scan path). The two methods were experimentally validated using two VHR datasets involving images collected by multispectral, panchromatic, and SAR sensors.

Both the causality of the whole framework and the expression of the MPM criterion have been proven analytically. The proposed hierarchical approach allows to obtain classification maps at all the spatial resolutions considered (a frequently advantageous property when working with multiresolution datasets). On the contrary, previous tech-

niques used for comparison perform classification at one resolution. Multiresolution fusion is addressed intrinsically by the quadtree topology, while multisensor fusion is obtained thanks to the flexibility of tree ensemble modeling. Quite accurate discrimination was attained on the test set in all the layers of the quadtree, in spite of the high spectral overlapping of several classes, corresponding to vegetated land covers sensed with only a few multispectral bands. The results indicated that the proposed approach was flexible to the incorporation of pixelwise predictions from different tree ensembles. Yet, the highest performances were attained using GBRT, making it a particularly appropriate choice within the proposed experimentation.

The classification maps obtained by the proposed methods showed visual smoothness, thus indicating the effectiveness of a framework that exploits the spatial information within each layer of the quadtree. While the previous algorithms considered for comparisons also generated accurate maps, the presented framework increased the overall accuracy, thus confirming the potential of the integration of hierarchical PGMs with tree ensembles in the field of multiresolution and multisensor fusion.

The two proposed algorithms show different computational complexities, lower for the MC-based method than for the SMMRF-based technique. As a matter of fact, the method integrating a 1D Markov chain along zig-zag and Hilbert curves for each layer of the quadtree is computationally advantageous, owing to its lower complexity that leads to lower execution times. This reduced computational complexity makes it possible to tackle more efficiently both a larger image size and more classes.

Future generalizations of the present work may consider the integration of this multiscale framework with convolutional neural networks, which intrinsically match the multiresolution data representation through their convolution and pooling layers [76], and with adaptive multiscale topologies associated with superpixel segmentation results at several scales [77,78].

Author Contributions: Conceptualization, I.H., G.M., S.B.S. and J.Z.; methodology, All; software, L.F., I.H. and A.M.; validation, All; formal analysis, I.H., A.M., G.M., M.P., S.B.S. and J.Z.; investigation, L.F. and A.M.; resources, G.M., S.B.S. and J.Z.; data curation, G.M., S.B.S. and J.Z.; writing—original draft preparation, M.P.; writing—review and editing, All; visualization, M.P.; supervision, G.M., S.B.S. and J.Z.; project administration, G.M., S.B.S. and J.Z.; funding acquisition, G.M., S.B.S. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received partial funding from the French Space Agency (CNES) within project CNES/INRIA no. 131024/00. The support is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Acknowledgments: The authors wish to thank CNES for partial financial funding, DigitalGlobe Inc. for the QuickBird data, Google Crisis Response for the GeoEye-1 data, and the Italian Space Agency (ASI) for the COSMO-SkyMed data. They would also like to thank Gabriele Barabino for his assistance in conducting the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proof of Theorem 1

We consider Assumptions 1–3 presented in Section 3.1.1 to hold valid.

Theorem A1. *According to Assumptions 1–3, the joint probability distribution of the class labels and the feature vectors in the whole quadtree, i.e., the joint distribution of the random vector* $(\mathcal{X}, \mathcal{Y})$*, is completely determined by the parent–child transition probabilities* $\{P(x_s|x_{s^-})\}_{s \in S \setminus S^0}$*, the past neighbor transition probabilities* $\{P(x_s|x_r, r \precsim s)\}_{s \in S}$*, and the pixelwise data conditional likelihoods* $\{p(y_s|x_s)\}_{s \in S}$*.*

Proof. According to Equation (4), Assumption 1 implies that:

$$\begin{aligned} P(\mathcal{X},\mathcal{Y}) &= p(\mathcal{Y}|\mathcal{X})P(\mathcal{X}) = p(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L,\mathcal{X}^{L-1},\ldots,\mathcal{X}^0) \\ &= p(\mathcal{Y}|\mathcal{X})P(\mathcal{X}^L|\mathcal{X}^{L-1})\ldots P(\mathcal{X}^1|\mathcal{X}^0)P(\mathcal{X}^0). \end{aligned} \tag{A1}$$

Incidentally, we note that $P(\mathcal{X},\mathcal{Y})$, which formalizes the joint distribution of $(\mathcal{X},\mathcal{Y})$, is neither a PMF nor a PDF because the random vector $(\mathcal{X},\mathcal{Y})$ includes both discrete ($\mathcal{X}$) and continuous components ($\mathcal{Y}$). Yet, $P(\mathcal{X},\mathcal{Y})$ does represent a proper "density function," meant in the sense of a Radon–Nikodym derivative with respect to an appropriate product measure space [63].

By using the factorizations in Equation (4) for $P(\mathcal{X}^\ell|\mathcal{X}^{\ell-1})$ ($\ell = 1,2,\ldots,L$), Equation (5) for $P(\mathcal{X}^0)$, and (6) for $p(\mathcal{Y}|\mathcal{X})$, into Equation (A1) we obtain:

$$P(\mathcal{X},\mathcal{Y}) \propto \prod_{\ell=1}^{L}\prod_{s\in S^\ell} P(x_s|x_r, r \precsim s)P(x_s|x_{s^-})p(y_s|x_s) \cdot \prod_{s\in S^0} P(x_s|x_r, r \precsim s)p(y_s|x_s). \tag{A2}$$

This implies that the joint probability distribution of the feature vectors and the labels $P(\mathcal{X},\mathcal{Y})$ is completely defined by the class-conditional likelihoods $p(y_s|x_s)$, the past neighbor transition probabilities $P(x_s|x_r, r \precsim s)$, and the parent–child transition probabilities $P(x_s|x_{s^-})$. □

Appendix A.2. Proof of Theorem 2

The Assumptions 1–6 introduced in Sections 3.1.1 and 3.1.2 are considered to hold valid.

Theorem A2. *According to Assumptions 1–6, for each site $s \in S \setminus S^0$, i.e., not in the root layer:*

$$P(x_s) = \sum_{x_{s^-}\in\Omega} P(x_s|x_{s^-})P(x_{s^-}), \tag{10}$$

$$P(x_s|\mathcal{C}_s,\mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_s|x_{s^-})P(x_{s^-})P(x_s)^{-|\mathcal{C}_s|} \cdot \prod_{r\precsim s} P(x_s|x_r)P(x_r), \tag{11}$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s,\mathcal{D}_s)P(x_{s^-}|\mathcal{Y})\prod_{r\precsim s} P(x_r|\mathcal{Y}), \tag{12}$$

where in Equation (10), the summation is over all possible class labels assigned to site s^-; in Equation (11), $|B|$ indicates the cardinality of a finite set B, and the labels x_{s^-} and $x_r, r \precsim s$, are fixed to the values they take in the context $\mathcal{C}_s$; and in Equation (12), the summation is over all possible label configurations in the context $\mathcal{C}_s$ of site s, i.e., over all possible labels of its parent and past neighbors.

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

$$P(x_s|\mathcal{C}_s,\mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_s)^{-|\mathcal{C}_s|}\prod_{r\precsim s} P(x_s|x_r)P(x_r), \tag{13}$$

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s,\mathcal{D}_s)\prod_{r\precsim s} P(x_r|\mathcal{Y}). \tag{14}$$

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

$$P(x_s|\mathcal{D}_s) \propto P(x_s|y_s)\prod_{t\in s^+}\sum_{x_t\in\Omega}\frac{P(x_t|\mathcal{D}_s)P(x_t|x_s)}{P(x_t)}. \tag{15}$$

Proof. Equation (10) directly derives from the total probability theorem. The proof of Equation (15) for a hierarchical MRF is reported in [10]. According to the total probability theorem:

$$P(x_s|\mathcal{Y}) = \sum_{\mathcal{C}_s} P(x_s|\mathcal{C}_s, \mathcal{Y})P(\mathcal{C}_s|\mathcal{Y}). \tag{A3}$$

Equations (12) and (14) follow by inserting Equations (7) and (8) into (A3). Considering Bayes theorem:

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(\mathcal{C}_s|x_s, \mathcal{D}_s)P(x_s|\mathcal{D}_s). \tag{A4}$$

The constant of proportionality does not include x_s. If $s \in S \setminus S^0$ and if Equation (9) is inserted into Equation (A4), we derive:

$$P(x_s|\mathcal{C}_s, \mathcal{D}_s) \propto P(x_s|\mathcal{D}_s)P(x_{s^-}|x_s)\prod_{r \precsim s} P(x_r|x_s). \tag{A5}$$

Equation (11) follows from this proportionality and from Bayes theorem. Equation (13) is proved analogously. □

References

1. Moser, G.; Serpico, S.B.; Benediktsson, J.A. Land-Cover Mapping by Markov Modeling of Spatial–Contextual Information in Very-High-Resolution Remote Sensing Images. *Proc. IEEE* **2013**, *101*, 631–651. [CrossRef]
2. Richards, J.A. *Remote Sensing Digital Image Analysis: An Introduction*, 5th ed.; Springer: Berlin, Germany, 2013.
3. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, G. *Remote Sensing Image Fusion*; CRC Press: Boca Raton, FL, USA, 2015.
4. Hedhli, I.; Moser, G.; Serpico, S.B.; Zerubia, J. Classification of multisensor and multiresolution remote sensing images through hierarchical Markov random fields. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2448–2452. [CrossRef]
5. Wang, X. Deep learning in object recognition, detection, and segmentation. *Found. Trends Signal Process.* **2016**, *8*, 217–382. [CrossRef]
6. Arnab, A.; Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Larsson, M.; Kirillov, A.; Savchynskyy, B.; Rother, C.; Kahl, F.; Torr, P.H.S. Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction. *IEEE Signal Process. Mag.* **2018**, *35*, 37–52. [CrossRef]
7. Li, S.Z. *Markov Random Field Modeling in Image Analysis*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
8. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]
9. Baggenstoss, P.M. A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 411–416. [CrossRef]
10. Laferté, J.M.; Pérez, P.; Heitz, F. Discrete Markov image modeling and inference on the quadtree. *IEEE Trans. Image Process.* **2000**, *9*, 390–404. [CrossRef]
11. Abend, K.; Harley, T.J.; Kanal, L.N. Classification of binary random patterns. *IEEE Trans. Inf. Theory* **1965**, *11*, 538–544. [CrossRef]
12. Fjortoft, R.; Delignon, Y.; Pieczynski, W.; Sigelle, M.; Tupin, F. Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 675–686. [CrossRef]
13. Devijver, P.A. Hidden Markov mesh random field models in image analysis. *J. Appl. Stat.* **1993**, *20*, 187–227. [CrossRef]
14. Dunmur, A.P.; Titterington, D.M. Computational Bayesian analysis of hidden Markov mesh models. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 1296–1300. [CrossRef]
15. Yousefi, S.; Kehtarnavaz, N.; Cao, Y. Computationally tractable stochastic image modeling based on symmetric Markov mesh random fields. *IEEE Trans. Image Process.* **2013**, *22*, 2192–2206. [CrossRef] [PubMed]
16. Razlighi, Q.R.; Kehtarnavaz, N.; Nosratinia, A. Computation of image spatial entropy using quadrilateral Markov random field. *IEEE Trans. Image Process.* **2009**, *18*, 2629–2639. [CrossRef] [PubMed]
17. Criminisi, A.; Shotton, J. *Decision Forests for Computer Vision and Medical Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2013.
18. Hedhli, I.; Moser, G.; Serpico, S.B.; Zerubia, J. Multi-resolution classification of urban areas using hierarchical symmetric Markov mesh models. In Proceedings of the JURSE 2017, Dubai, United Arab Emirates, 6–8 March 2017; pp. 1–4.
19. Montaldo, A.; Fronda, L.; Hedhli, I.; Moser, G.; Serpico, S.B.; Zerubia, J. Causal Markov mesh hierarchical modeling for the contextual classification of multiresolution satellite images. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 2716–2720.
20. Montaldo, A.; Fronda, L.; Hedhli, I.; Moser, G.; Zerubia, J.; Serpico, S.B. Joint classification of multiresolution and multisensor data using a multiscale Markov mesh model. In Proceedings of the 2019 IEEE Geoscience and Remote Sensing Society (IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 2810–2813.

21. Montaldo, A.; Fronda, L.; Hedhli, I.; Moser, G.; Serpico, S.; Zerubia, J. A causal hierarchical Markov framework for the classification of multiresolution and multisensor remote sensing images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-3-2020*, 269–277. [CrossRef]
22. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
23. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
24. Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation forest: A New classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [CrossRef]
25. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
26. Merentitis, A.; Debes, C. Many hands make light work—On ensemble learning techniques for data fusion in remote sensing. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 86–99. [CrossRef]
27. Serpico, S.B.; Dellepiane, S.; Boni, G.; Moser, G.; Angiati, E.; Rudari, R. Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proc. IEEE* **2012**, *100*, 2946–2970. [CrossRef]
28. Pohl, C.; Van Genderen, J. *Remote Sensing Image Fusion: A Practical Guide*, 1st ed.; CRC Press: Baton Rouge, FL, USA, 2016.
29. Mallat, S. *A Wavelet Tour of Signal Processing—The Sparse Way*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2009.
30. Li, S.; Yang, B. Hybrid multiresolution method for multisensor multimodal image fusion. *IEEE Sensors J.* **2010**, *10*, 1519–1526.
31. Pajares, G.; Manuel de la Cruz, J. A wavelet-based image fusion tutorial. *Pattern Recognit.* **2004**, *37*, 1855–1872. [CrossRef]
32. Cole-Rhodes, A.; Johnson, K.; LeMoigne, J.; Zavorin, I. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE Trans. Image Process.* **2003**, *12*, 1495–1511. [CrossRef]
33. Gong, M.; Zhao, S.; Jiao, L.; Tian, D.; Wang, S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4328–4338. [CrossRef]
34. Zavorin, I.; Moigne, J. Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery. *IEEE Trans. Image Process.* **2005**, *14*, 770–782. [CrossRef] [PubMed]
35. Pradhan, P.S.; King, R.L.; Younan, N.H.; Holcomb, D.W. Estimation of the mumber of decomposition levels for a wavelet-based multiresolution multisensor image fusion. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3674–3686. [CrossRef]
36. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Höfle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; Atkinson, P.; Benediktsson, J. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [CrossRef]
37. Carper, W.; Lillesand, T.; Kiefer, P. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
38. Chavez, P.; Kwarteng, A. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
39. Baronti, S.; Aiazzi, B.; Selva, M.; Garzelli, A.; Alparone, L. A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 446–453. [CrossRef]
40. Liu, J.G. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *Int. J. Remote Sens.* **2000**, *21*, 3461–3472. [CrossRef]
41. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-tailored multiscale fusion of high-resolution MS and pan imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596. [CrossRef]
42. Nunez, J.; Otazu, X.; Fors, O.; Prades, A.; Pala, V.; Arbiol, R. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1204–1211. [CrossRef]
43. Otazu, X.; Gonzalez-Audicana, M.; Fors, O.; Nunez, J. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385. [CrossRef]
44. Nencini, F.; Garzelli, A.; Baronti, S.; Alparone, L. Remote sensing image fusion using the curvelet transform. *Inf. Fusion* **2007**, *8*, 143–156. [CrossRef]
45. Thoonen, G.; Mahmood, Z.; Peeters, S.; Scheunders, P. Multisource classification of color and hyperspectral images using color attribute profiles and composite decision fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 510–521. [CrossRef]
46. Gómez-Chova, L.; Tuia, D.; Moser, G.; Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **2015**, *103*, 1560–1584. [CrossRef]
47. Datcu, M.; Melgani, F.; Piardi, A.; Serpico, S.B. Multisource data classification with dependence trees. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 609–617. [CrossRef]
48. Benediktsson, J.A.; Swain, P.H.; Ersoy, O.K. Neural network approaches versus statistical methods in classification of multisource remote sensing data. In Proceedings of the 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 10–14 July 1989; Volume 2, pp. 489–492.
49. Serpico, S.B.; Roli, F. Classification of multi-sensor remote-sensing images by structured neural networks. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 562–578. [CrossRef]
50. Simone, G.; Farina, A.; Morabito, F.; Serpico, S.B.; Bruzzone, L. Image fusion techniques for remote sensing applications. *Int. J. Inf. Fusion* **2002**, *3*, 3–15. [CrossRef]
51. Waske, B.; Benediktsson, J.A. Fusion of support vector machines for classification of multisensor data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3858–3866. [CrossRef]

52. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Rojo-Alvarez, J.L.; Martinez-Ramon, M. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1822–1835. [CrossRef]
53. Solberg, A.H.S.; Taxt, T.; Jain, A.K. A Markov random field model for classification of multisource satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 100–113. [CrossRef]
54. Alparone, L.; Baronti, S.; Garzelli, A.; Nencini, F. Landsat ETM+ and SAR image fusion based on generalized intensity Modulation. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2832–2839. [CrossRef]
55. Chibani, Y. Integration of panchromatic and SAR features into multispectral SPOT images using the 'à trous' wavelet decomposition. *Int. J. Remote Sens.* **2007**, *28*, 2295–2307. [CrossRef]
56. Byun, Y.; Choi, J.; Han, Y. An area-based image fusion scheme for the integration of SAR and optical satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2212–2220. [CrossRef]
57. Colditz, R.R.; Wehrmann, T.; Bachmann, M.; Steinnocher, K.; Schmidt, M.; Strunz, G.; Dech, S. Influence of image fusion approaches on classification accuracy: A case study. *Int. J. Remote Sens.* **2006**, *27*, 3311–3335. [CrossRef]
58. Storvik, G.; Fjortoft, R.; Solberg, A.H.S. A Bayesian approach to classification of multiresolution remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 539–547. [CrossRef]
59. Voisin, A.; Krylov, V.A.; Moser, G.; Serpico, S.B.; Zerubia, J. Supervised classification of multisensor and multiresolution remote sensing images with a hierarchical copula-based approach. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3346–3358. [CrossRef]
60. Li, Z.; Chen, G.; Zhang, T. Temporal attention networks for multitemporal multisensor crop classification. *IEEE Access* **2019**, *7*, 134677–134690. [CrossRef]
61. Li, Z.; Chen, G.; Zhang, T. A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 847–858. [CrossRef]
62. Willsky, A.S. Multiresolution Markov models for signal and image processing. *Proc. IEEE* **2002**, *90*, 1396–1458. [CrossRef]
63. Rudin, W. *Real and Complex Analysis*, 3rd ed.; McGraw-Hill, Inc.: New York, NY, USA USA, 1987.
64. Dikshit, O.; Roy, D.P. An empirical investigation of image resampling effects upon the spectral and textural supervised classification of a high spatial resolution multispectral image. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1085–1092.
65. Inglada, J.; Muron, V.; Pichard, D.; Feuvrier, T. Analysis of artifacts in subpixel remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 254–264. [CrossRef]
66. Kato, Z.; Zerubia, J. Markov Random Fields in Image Segmentation. *Found. Trends Signal Process.* **2012**, *5*, 1–155. [CrossRef]
67. Bouman, C.A.; Shapiro, M. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. Image Process.* **1994**, *3*, 162–177. [CrossRef]
68. Hedhli, I.; Moser, G.; Zerubia, J.; Serpico, S.B. A new cascade model for the hierarchical joint classification of multitemporal and multiresolution remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6333–6348. [CrossRef]
69. Yousefi, S.; Kehtarnavaz, N. Generating symmetric causal Markov random fields. *Electron. Lett.* **2011**, *47*, 1224–1225. [CrossRef]
70. Schapire, R. The strength of weak learnability. *Mach. Learn.* **1990**, *5*, 197–227. [CrossRef]
71. Briem, G.J.; Benediktsson, J.A.; Sveinsson, J.R. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2291–2299. [CrossRef]
72. Miao, X.; Heaton, J.; Zheng, S.; Charlet, D.; Liu, H. Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *Int. J. Remote Sens.* **2012**, *33*, 1823–1849. [CrossRef]
73. DeFries, R.S.; Chan, J.C.W. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sens. Environ.* **2000**, *74*, 503–515. [CrossRef]
74. Kolmogorov, V. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1568–1583. [CrossRef] [PubMed]
75. Moser, G.; De Giorgi, A.; Serpico, S.B. Multiresolution supervised classification of panchromatic and multispectral images by Markov random fields and graph cuts. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5054–5070. [CrossRef]
76. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT Press: Boston, MA, USA, 2016.
77. Zhou, Y.; Li, J.; Feng, L.; Zhang, X.; Hu, X. Adaptive scale selection for multiscale segmentation of satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3641–3651. [CrossRef]
78. Akbas, E.; Ahuja, N. Low-level multiscale image segmentation and a benchmark for its evaluation. *Comput. Vis. Image Underst.* **2020**, *199*, 103026. [CrossRef]