A Hierarchical Markovian Model for Multiscale Region-Based Classification of Vector-Valued Images

Antonis Katartzis, Iris Vanhamel and Hichem Sahli

Abstract

We propose a new classification method for vector-valued images, based on (i) a causal Markovian model, defined on the hierarchy of a multiscale region adjacency tree (MRAT), and (ii) a set of non-parametric dissimilarity measures that express the data likelihoods. The image classification is treated as a hierarchical labeling of the MRAT, using a finite set of interpretation labels (e.g. land cover classes). This is accomplished via a non-iterative estimation of the modes of posterior marginals (MPM), inspired from existing approaches for Bayesian inference on the quadtree. The paper describes the main principles of our method and illustrates classification results on a set of artificial and remote sensing images, together with qualitative and quantitative comparisons with a variety of pixel-based techniques that follow the Bayesian-Markovian framework either on hierarchical structures or the original image lattice.

Index Terms

Region-based image segmentation, classification, scale-space, hierarchical Markovian models, distribution dissimilarity measures, remote sensing

I. Introduction

Vector-valued images with a wide range of band frequencies, has been extensively used in the field of remote sensing, where several multi- or hyper-spectral sensors are now effectively used for earth observation [1]. The classification of such images is generally based on the discrimination between spectral signatures, using pixel-wise classifiers, such as Bayesian classifiers [2], neural networks [3] or fuzzy clustering [4]. One drawback of these

approaches is that they do not take into account contextual information during the labeling process. The issue of combining both spectral and spatial information has been addressed by several authors, in the framework of Markov random field (MRF) theory. Non-causal Markovian (or energy-based) models, defined on the image lattice, have been widely used in the area of computer vision [5], as well as the application domain of remote sensing [6]. Unfortunately, the non-causal nature of these models generally leads to iterative inference algorithms that are computationally demanding.

In order to cope with the computational burden of large images (like typical remote sensing data), several hierarchical, multiresolution Markovian models have been proposed for pixel-based image classification/segmentation. In their majority, they are associated with a set of parametric models that describe the data likelihoods. Stochastic models defined on hierarchical graph structures, such as quadtrees or pyramids [7][8][9], yield in-scale causality properties that allow fast and efficient non-iterative inference procedures, similar to those used for discrete Markov chain models. Unfortunately, such hierarchical structures can exhibit shift-, rotation- or scale-variance, something that it is generally acknowledged and proven in [10]. Particularly, quadtree-based approaches have the inherent disadvantage of inducing block artifacts in the final estimates [8]. The problem of block artifacts can be avoided, in some extent, by using a more inter-leaved hierarchy, such as the one proposed by Bouman and Shapiro [7]. This is accomplished by increasing the number of coarse level neighbors at the original quadtree structure. On the other hand, more complex hierarchies that incorporate also intra-level spatial interactions, like multigrid approaches [11], methods based on truncated trees [12] or three dimensional adjacency graphs [13], although being more accurate, they lack the practical advantages of trees or pyramids, leading to iterative or semi-iterative labeling processes.

In this paper, we present a new classification method for vector-valued images, based on a Markovian model, defined on a multiscale region adjacency graph, in the form of a truncated tree structure, denoted as Multiscale Region Adjacency Tree (MRAT) [14]. Although the concept of region-based classification is not new in the area of remote sensing, the majority of the existing approaches consider the classification as a post-processing module, applied on the result of an initial segmentation [15]. In our scheme, the classification is applied, in a stochastic way, on an hierarchy of image partitions, and not on a predetermined segmentation. This hierarchy captures the multiscale nature of each image, since it is obtained from the *deep image structure* [16]. The latter implies that the original image is extended to a family of derived images for which the scale increases [17].

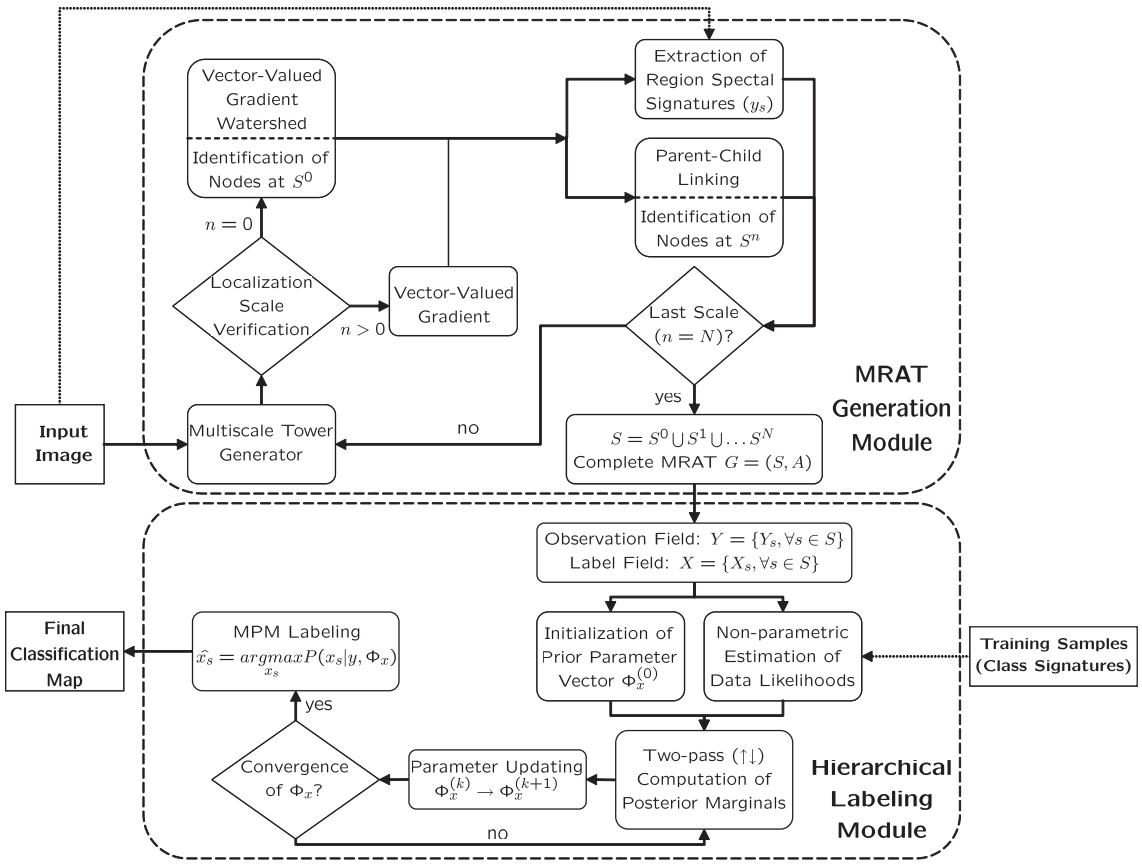


Fig. 1. Schematic diagram of the proposed hierarchical, region-based classification scheme.

The proposed classification method consists of two basic modules, namely the *MRAT generation* and the *hierarchical labeling*, as depicted in Fig. 1. The MRAT is generated as follows [18]: Initially, a hierarchical representation of the image is obtained in the form of a multiscale tower, generated by anisotropic diffusion filtering. The particular diffusion scheme is based on inherent image properties and leads to noise reduction and good localization of the boundaries of the image structures. At the finest scale, the watershed transformation is performed on the generalized gradient of the vector-valued image [19], in order to identify the position of all the contours in the image. At higher scales, the duality between the regional minima of the generalized gradient and the catchment basins of the watershed is exploited to create a robust region-based parent-child linking list along the scale-space stack. This linking scheme provides the nodes and branches of the MRAT.

The image classification is then treated as a hierarchical labeling problem, applied to the previously defined MRAT, using a finite set of interpretation labels (e.g. land cover classes). We associate to the nodes of the hierarchy

a causal label field in the form of a coarse-to-fine Markov chain and an observation field, which represents the spectral signature of each region in the original image. The data likelihoods of the field of observations can be expressed in terms of a series of dissimilarity measures, and have the advantage to be model-free, in the sense that the underlying probability distributions are not assumed to belong to a parametric model class. The nature of the proposed MRAT allows us to incorporate efficient, non-iterative schemes for Bayesian inference, which have been inspired from the state-of-the-art developments of hierarchical Markovian models defined on quadtrees. In particular, an optimal labeling is achieved via a non-iterative procedure [8] that provides the exact estimates of the *modes of posterior marginals (MPM)* and facilitates the estimation of the involved parameters.

Despite the similarities of our hierarchical model with the ones proposed by [7] or [8], there are some fundamental differences that promote region-based multiscale hierarchies over pixel-based hierarchical graphs (such as pyramids or quadtrees):

1) In the proposed scheme, the sites under labeling are not individual pixels but correspond to perceptually uniform regions, which implicitly encompass contextual information and offer a better description of the statistical properties of each site.
2) Pixel-based hierarchies always follow a predefined structure, whereas our statistical model is defined on an information-driven, non-regular, hierarchical partition of the image, which captures the hierarchical nature of the image content through scale-space evolution.
3) In the case of pixel-based hierarchical models defined on a mono-resolution (single resolution) image, the observations are expressed only at the finest resolution level, from which the data likelihoods at coarser levels are inferred. In our approach there is an *explicit* assignment of attributes (observations) at each node of the tree, which allows to estimate the data likelihoods at each scale. This improves the statistical inference all over the hierarchical structure, not only at the finest resolution level.

The paper is organized as follows. Section II is dedicated to the generation of the scale-space tower and the creation of the multiscale region adjacency tree. Section III describes the core structure of the proposed Bayesian hierarchical labeling scheme. In section IV, we illustrate classification results on a set of artificial and remote sensing images, and perform a qualitative and quantitative comparison of our method with a variety of pixel-based classification approaches that follow the Bayesian inference. Finally, conclusions and directions for future research

work are given in section V.

II. Multiscale Region Adjacency Tree (MRAT)

The creation of the multiscale region adjacency tree is based upon scale-space theory [16] [17], thus it consists of two main parts: (i) the scale-space filtering and (ii) the deep image structure. Scale-space filtering concerns the mechanism that embeds the image into a family of derived images (referred as multiscale tower), for which the image content is causally simplified. The deep image structure refers to the methodology of relating image features along the multiscale tower.

A. Generation of the Multiscale Tower: Scale-space filtering

In order to avoid blurring and delocalization of the image features, an image adaptive scale-space filter is used. In this work, we opted for a method that guides the filtering process in such a way that intra-region smoothing is preferred over inter-region smoothing and edges are gradually enhanced. The employed filter belongs to the class of nonlinear anisotropic diffusion filters. It is a combination of the Catté et al. [20] regularized Perona and Malik filter [21] and the one proposed by You et al. [22].

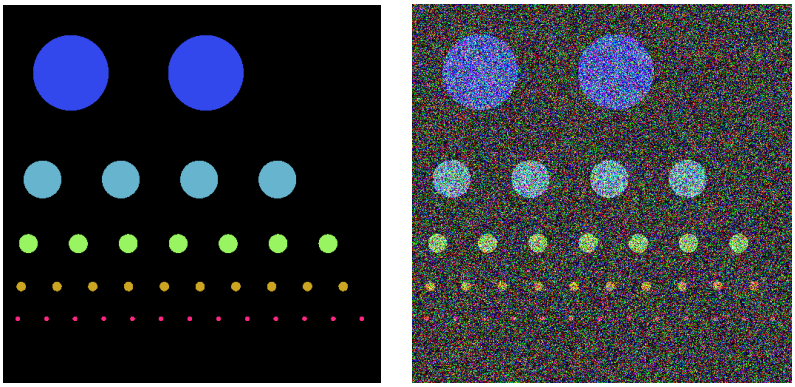
Let $I = \{I^{(1)}, I^{(2)}, \ldots, I^{(\mathcal{R})}\}$ be a vector-valued image defined on a finite domain Ω. The multiscale tower (u) of I is governed by the following system of coupled parabolic partial differential equations (PDEs):

$$\begin{cases} \partial_t u^{(r)} & = \text{div}\left[g\left(|\nabla u_\sigma|\right)\frac{\nabla u^{(r)}}{|\nabla u^{(r)}|}\right] \quad \forall r = 1, 2, \ldots, \mathcal{R} \\ u\left(t=0\right) & = I \\ \partial_{\mathbf{n}} u & = 0 \qquad\qquad\qquad\qquad \text{on } \delta\Omega \end{cases} \tag{1}$$

where $u^{(r)}$ represents the r^{th} image band, t is the continuous scale parameter, $\delta\Omega$ is the image boundary (with $\mathbf{n}$ denoting the normal direction to it), and σ is a regularization parameter, which ensures the well-posedness of the above system; g is the Lorentzian edge stopping function [23], which is formulated as:

$$g(|\nabla u_\sigma|) = \frac{1}{1 + \frac{|\nabla u_\sigma|^2}{k^2}} \tag{2}$$

The so-called contrast parameter k in (2) separates backward from forward diffusion and is estimated using the cumulative histogram of the reguralized gradient magnitude ($|\nabla u_\sigma|$) [18]. A discrete version of the multiscale tower u, denoted as $U = \{u_0, \ldots, u_n, \ldots, u_N\}$, is obtained by applying the natural scale-space sampling method [16].



(a) (b)

Fig. 2. (a) An example of a 512x512 synthetic color image; (b) Noisy counterpart, corrupted with Gaussian white noise.

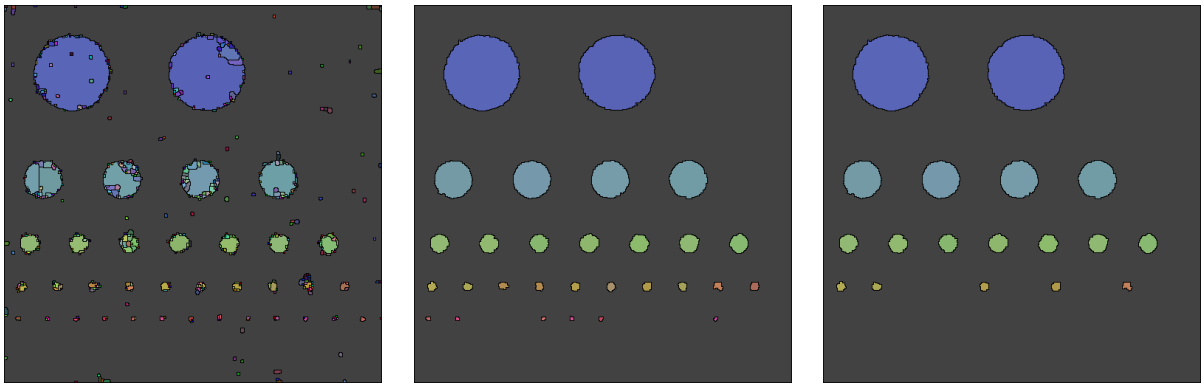
(a) (b) (c)

Fig. 3. Multiscale tower U. (a) Localization scale u_0; (b) Scale u_3; (c) Scale u_5.

The finest scale u_0, hereafter referred as the *localization scale*, is determined empirically and corresponds to the scale that obtains a maximum noise reduction, while retaining all important image features.

Figure 3 illustrates the generated multiscale tower U of a synthetic image, shown in Fig.2(b). The latter is a noisy version of the 512x512 synthetic color image of Fig.2(a), which is composed of disks with various radii in front of a homogeneous background. Each color channel is corrupted with Gaussian white noise of zero mean and a standard deviation equal to 0.5 (considering image intensities ranging from 0 to 1). The resulted signal-to-noise ratio [1] for each of the three color channels is: $SNR_R = -1.2$db, $SNR_G = -0.85$db and $SNR_B = -0.49$db.

[1]The definition used for the estimation of the SNR is the following: $SNR \triangleq \log_{10} \frac{\text{Variance}\{image\}}{\text{Variance}\{noise\}}$



(a) (b) (c)

Fig. 4. Segmentation of the multiscale tower U. (a) Segmentation at the localization scale u_0 with 628 detected regions; (c)-(d) Segmentations at scale u_3 and scale u_5 with 30 and 19 regions, respectively.

B. Construction of the MRAT: The Deep Image Structure

At the localization scale, we perform a gradient watershed transformation to detect a set of regions with well-localized contours. By using the duality between the regional minima of the gradient and the catchment basins of the watershed, the regions at u_0 are tracked across the scales [24]. This is achieved by spatially projecting the regional minima at each scale, into the coarser one. In particular, each regional minimum residing in a given scale is linked with a regional minimum in the coarser scale, if and only if its projection falls in the catchment basin of that regional minimum. This process renders a robust parent-child linking. Fig. 4 depicts a segmentation of the multiscale tower, obtained by merging the detected regions at u_0 that have the same parent at three different scales.

The produced hierarchy allows the construction of a multiscale graph of regions $G = (S, A)$. G has the form of a truncated tree (MRAT), with each node (apart from the ones at the highest scale) having a unique predecessor (its parent). The set of nodes S are partitioned into the different scales (indexed by $n = 0, \ldots, N$), so that $S = S^0 \bigcup S^1 \bigcup \ldots S^N$. We use the following notations: $\overline{s}$ is the parent of a node s (if s does not belong to the coarser scale); $\underline{s}$ is the set of its children (if s does not belong to the localization scale); finally, $\underline{\underline{s}}$ is the set of all descendants of s (including s). An illustration of the MRAT, between three successive scales, is presented in Fig. 5.

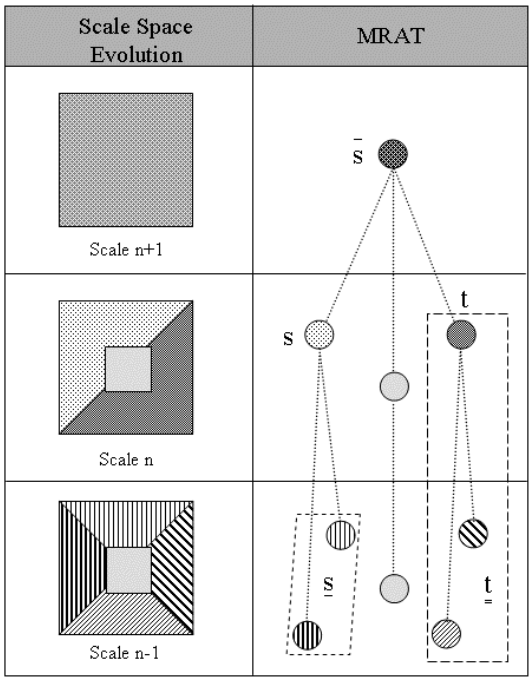


Fig. 5. An example of the Multiscale Region Adjacency Tree in three successive scales.

III. HIERARCHICAL LABELING

Image classification falls in the broad category of inverse problems, where one attempts to estimate the optimal realization of some hidden variables (image classes), given a set of observations. In our approach, the image classification is treated as a hierarchical labeling problem, applied to the nodes S of the previously defined MRAT (G). The labeling is performed using a finite set $L = \{1, \ldots, M\}$ of interpretation labels. We consider a couple of random fields (X, Y) on G, with $X = \{X_s, \forall s \in S\}$ being the label field, and $Y = \{Y_s, \forall s \in S\}$ a field of observations that represent spectral region properties. At each scale n, we denote as $Y^n \triangleq \{Y_s,\, s \in S^n\}$ and $X^n \triangleq \{X_s,\, s \in S^n\}$ the restriction of both fields to the subset of nodes S^n, with similar notations holding for their realizations $y^n \triangleq \{y_s,\, s \in S^n\}$ and $x^n \triangleq \{x_s,\, s \in S^n,\, x_s \in L\}$.

We associate a causal Markovian model to the label field X and identify the optimal configuration of labels $\hat{x_s}$ based on a non-iterative procedure that provides the modes of posterior marginals (MPM):

$$\forall s \in S\,,\ \hat{x_s} = \underset{x_s \in L}{argmax} P(X_s = x_s | Y = y) \triangleq \underset{x_s \in L}{argmax} P(x_s | y) \tag{3}$$

The MPM estimation process, hereafter denoted as *MRAT-MPM*, is inspired from the state-of-the-art developments

of Bayesian inference on quadtrees [8] [9]. For its realization, we have put forward a series of statistical assumptions that characterize the prior and data likelihood distributions ($P(X = x) \triangleq P(x)$ and $P(Y = y|X = x) \triangleq P(y|x)$, respectively). The explicit form of both distributions is described in details in the following two sections.

A. Prior Markovian model

As the labeling process is applied on the independence graph of the MRAT, we can put forward the following three assumptions regarding the *a priori* distribution $P(x)$:

- X can be considered as a first order coarse-to-fine Markov chain:

$$P(x^n|x^{n'}, n' > n) = P(x^n|x^{n+1}), \forall n < N \tag{4}$$

- The labels of X^n are conditionally spatially independent, given the labels of X^{n+1}. Additionally, for each node in S^n, the conditioning in X^{n+1} reduces to a dependence with respect to its parent only:

$$P(x^n|x^{n+1}) = \prod_{s \in S^n} P(x_s|x_{\overline{s}}), \forall n < N \tag{5}$$

- $P(x^N)$ is uniformly distributed (no intra-scale spatial interactions are taken into account), so that:

$$P(x^N) = \prod_{s \in S^N} P(x_s) \tag{6}$$

The combination of (4), (5) and (6) gives the following form for the *a priori* distribution:

$$\begin{array}{ll} P(x) = P(x^N) \prod_{n<N} P(x^n|x^{n+1}) & = P(x^N) \prod_{s \in S \backslash S^N} P(x_s|x_{\overline{s}}) \\ & = \prod_{s \in S^N} P(x_s) \prod_{s \in S \backslash S^N} P(x_s|x_{\overline{s}}) \end{array} \tag{7}$$

where $S \backslash S^N$ denotes all nodes of S excluding S^N. For the transition probabilities $P(x_s|x_{\overline{s}})$, we have adopted the Potts-like function used by Bouman and Shapiro [7]:

$$\forall s \in S \backslash S^N,\ P(x_s|x_{\overline{s}}) = \begin{cases} \theta_n & \text{if } x_s = x_{\overline{s}} \\ \frac{1-\theta_n}{M-1} & \text{otherwise} \end{cases} \tag{8}$$

The parameter $\theta_n > \frac{1}{M}$ is the probability that the labeling will remain the same from scale $n+1$ to n. If a class change does occur, it is equally likely to be any of the remaining class types. Finally, by denoting as $\pi_s(i) \triangleq P(X_s = x_s = i), \forall s \in S^N$, the prior Markovian model $P(x)$ is fully characterized by the parameter vector $\Phi_x = [\pi_s(i), \theta_n]$.

B. Observation Model

For each region $s \in S$, we consider as observation a *unary* region property Y_s that represents the spectral signature of the area that region s encompasses at the original image (an image with $\mathcal{R}$ spectral bands). We assume that the components of the observation field Y are all mutually independent given X and that, for each of them, the conditioning w.r.t the corresponding hidden variable is independent from the other hidden variables. This leads to the following expression for the conditional likelihood function:

$$P(y|x) = \prod_{s \in S} P(y_s | X_s = x_s) \tag{9}$$

By denoting as H_i the spectral signature of label i, the site-wise likelihood $P(y_s|X_s = i)$ can be expressed as a function of a *dissimilarity measure* $D(y_s, H_i)$ between the region (y_s) and label (H_i) distributions:

$$P(y_s|X_s = i) \propto exp(-\lambda\, D(y_s, H_i)) \ , \forall s \in S \, , \forall i \in L \tag{10}$$

where λ is a weighting factor, which was empirically fixed to the same value in all experiments reported in this paper. The likelihood function in (10) has been frequently used in computer vision applications that involve color similarity [25], and allows a free selection of the functional form of the dissimilarities measures D. For our application, we have investigated a set of non-parametric dissimilarity measures that do not conform to a given parametric distribution model.

Several non-parametric dissimilarity measures have been reported in the literature [26] and have been successfully applied in the field of image classification, retrieval and unsupervised segmentation. In our case, the multivariate histograms of both samples y_s and H_i serve as a non-parametric estimator of their empirical distributions. As multivariate histograms with regular binning often result in poor performance (curse of dimensionality), we have chosen to use instead the marginal histograms of the $\mathcal{R}$ bands of each sample. For a given data sample I, its r^{th} marginal histogram at each bin b will be denoted as $f^{(r)}(b\,; I) \, , \forall\, r = 1, \ldots, \mathcal{R}$. For each band r, we have chosen the following two representative distance measures:

- χ^2*-statistic* [27]:

$$D^{(r)}(y_s, H_i) = \sum_{b}^{\#bins} \frac{(f^{(r)}(b\,; y_s) - \hat{f}^{(r)}(b))^2}{\hat{f}^{(r)}(b)} \tag{11}$$

  where $\hat{f}^{(r)}(b) = \left\{ f^{(r)}(b\,; y_s) + f^{(r)}(b\,; H_i) \right\} /2$.

- *Kolmogorov-Smirnov distance* [28]:

$$D^{(r)}(y_s, H_i) = \max_b |F^{(r)}(b\,; y_s) - F^{(r)}(b\,; H_i)| \tag{12}$$

where $F^{(r)}(b\,; I)$ denotes the corresponding cumulative histogram of $f^{(r)}(b\,; I)$.

For the combination of the $\mathcal{R}$ independently evaluated comparisons in both (11) and (12), we use the L_∞ norm, so that:

$$D(y_s, H_i) = \underset{(r)}{argmax}\ D^{(r)}(y_s, H_i)$$

whereas the involved histograms were discretized using 20 bins.

The comparison of the two dissimilarity measures in (11) and (12), in terms of classification performance, is presented in section IV.

C. Joint Distribution - MPM Estimation

Using the assumptions of conditional independence in (7) and (9), the joint probability $P(x, y)$ can be expressed as:

$$P(x, y) = \prod_{s \in S^N} P(x_s) \prod_{s \in S \setminus S^N} P(x_s | x_{\bar{s}}) \prod_{s \in S} P(y_s | x_s) \tag{13}$$

Given the data likelihoods of (10), $P(x, y)$ depends only on the prior parameter vector Φ_x. For a given instance of Φ_x, it is possible to compute exactly the posterior marginals $P(x_s|y)$ and the optimal labels $\hat{x_s}$ (through equation (3)) at each node $s \in S$, within two passes on the MRAT. In the following description, we will assume that the prior parameters Φ_x are known, while its exact estimation will be described in details in section III-D.

Based on the Bayes rule, we can determine the following downward recursion for the estimation of the posterior marginals:

$$P(x_s|y) = \sum_{x_{\bar{s}}} P(x_s | x_{\bar{s}}, y) P(x_{\bar{s}}|y)\,, \forall s \in S \setminus S^N \tag{14}$$

Using the properties of conditional independence of the MRAT, $P(x_s|x_{\bar{s}}, y) = P(x_s|x_{\bar{s}}, y_{\underline{\underline{s}}})$, the recursion in (14) becomes:

$$P(x_s|y) = \sum_{x_{\bar{s}}} P(x_s | x_{\bar{s}}, y_{\underline{\underline{s}}}) P(x_{\bar{s}}|y)\,, \forall s \in S \setminus S^N \tag{15}$$

The recursive form of (15) is initialized by estimating the posterior marginals at the highest scale $P(x_s|y)$, $\forall s \in S^N$, and requires the estimation of probabilities $P(x_s|x_{\overline{s}}, y_{\underline{\underline{s}}})$, $\forall s \in S \backslash S^N$. According to Laferte et al. [8]:

$$P(x_s|x_{\overline{s}}, y_{\underline{\underline{s}}}) = \frac{P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}})}{P(x_{\overline{s}}|y_{\underline{\underline{s}}})} = \frac{P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}})}{\sum_{x_s} P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}})}$$

$$\text{with } P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}}) = P(x_{\overline{s}}|x_s)P(x_s|y_{\underline{\underline{s}}}) = \frac{P(x_s|x_{\overline{s}})P(x_{\overline{s}})}{P(x_s)} P(x_s|y_{\underline{\underline{s}}})$$

where the prior marginals $P(x_s)$ are computed using an initial top-down recursion:

$$\begin{cases} \forall s \in S^N, \forall i \in L: \; P(X_s = i) = \pi_s(i) \\ \forall s \in S \backslash S^N: \; P(x_s) = \sum_{x_{\overline{s}}} P(x_s|x_{\overline{s}}) P(x_{\overline{s}}) \end{cases}$$

Finally, the calculation of $P(x_s|y_{\underline{\underline{s}}})$ is carried out through the following bottom-up recursion:

$$\begin{aligned} P(x_s|y_{\underline{\underline{s}}}) \propto P(x_s, y_{\underline{\underline{s}}}) &= \sum_{x_{\underline{s}}} P(y_{\underline{\underline{s}}}, x_{\underline{s}}|x_s) P(x_s) \\ &= P(x_s) P(y_s|x_s) \prod_{t \in \underline{s}} P(y_{\underline{\underline{t}}}|x_s) \\ &= P(x_s) P(y_s|x_s) \prod_{t \in \underline{s}} \sum_{x_t} \underbrace{P(y_{\underline{\underline{t}}}|x_t)}_{\propto \frac{P(x_t|y_{\underline{\underline{t}}})}{P(x_t)}} P(x_t|x_s) \end{aligned}$$

or, equivalently

$$P(x_s|y_{\underline{\underline{s}}}) = \frac{1}{Z} P(x_s) P(y_s|x_s) \prod_{t \in \underline{\underline{s}}} \sum_{x_t} P(x_t|y_{\underline{\underline{t}}}) \frac{P(x_t|x_s)}{P(x_t)}$$

where Z is a normalization factor, satisfying the condition $\sum_{x_s} P(x_s|y_{\underline{\underline{s}}}) = 1$.

Based on these formulations, the MPM estimates that provide the optimal labels $\hat{x_s}$ can be identified through two passes on the MRAT as indicated in table I.

At this point we have to note that the non-regular nature of the MRAT may induce a problem of *underflow*: during the upwards recursion, the partial posterior marginal $P(x_s|y_{\underline{\underline{s}}})$ in (18) may become so small that the usual precision of computers is not sufficient. This can occur in the case of an abrupt reduction of nodes between two successive scales, i.e. from an over-segmented scale to an under-segmented one, inducing the assignment of a large number of children to a given parent. In order to avoid this problem, a good selection of the localization scale should be made.

D. Parameter Estimation

As stated in the previous section, the joint probability in (13) and the two-pass computation of the posterior marginals in table I require the estimation of the parameter vector $\Phi_x = [\pi_s(i), \theta_n]$. The estimation of these

□ **Initial top-down recursion**

- Computation of the prior marginals:

$$\begin{aligned} &\forall s \in S^N, \forall i \in L:\ P(X_s = i) = \pi_s(i) \\ &\forall s \in S \backslash S^N:\ P(x_s) = \textstyle\sum_{x_{\overline{s}}} P(x_s|x_{\overline{s}}) P(x_{\overline{s}}) \end{aligned} \tag{16}$$

△ **Bottom-up pass**

- Leaves, $s \in S^0$:

$$\begin{aligned} P(x_s|y_s) &= \tfrac{P(y_s|x_s)P(x_s)}{\sum_{x_s} P(y_s|x_s)P(x_s)} \\ P(x_s, x_{\overline{s}}|y_s) &= \tfrac{P(x_s|x_{\overline{s}})P(x_{\overline{s}})}{P(x_s)} P(x_s|y_s) \end{aligned} \tag{17}$$

- Recursion, for $n = 1, \ldots, N-1$, $s \in S^n$:

$$\begin{aligned} P(x_s|y_{\underline{\underline{s}}}) &= \tfrac{1}{Z} P(x_s) P(y_s|x_s) \textstyle\prod_{t \in \underline{s}} \sum_{x_t} P(x_t|y_{\underline{\underline{t}}}) \tfrac{P(x_t|x_s)}{P(x_t)} \\ P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}}) &= \tfrac{P(x_s|x_{\overline{s}})P(x_{\overline{s}})}{P(x_s)} P(x_s|y_{\underline{\underline{s}}}) \end{aligned} \tag{18}$$

◇ **Highest scale**

- $\forall s \in S^N$:

$$\begin{aligned} &P(x_s|y) = P(x_s|y_{\underline{\underline{s}}}) = \tfrac{1}{Z} P(x_s) P(y_s|x_s) \textstyle\prod_{t \in \underline{s}} \sum_{x_t} P(x_t|y_{\underline{\underline{t}}}) \tfrac{P(x_t|x_s)}{P(x_t)} \\ &\hat{x_s} = \underset{x_s}{argmax} P(x_s|y) \end{aligned} \tag{19}$$

▽ **Top-down pass**

- Recursion, for $n = N-1, \ldots, 0$, $s \in S^n$:

$$\begin{aligned} P(x_s|y) &= \textstyle\sum_{x_{\overline{s}}} \tfrac{P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}})}{\sum_{x_s} P(x_s, x_{\overline{s}}|y_{\underline{\underline{s}}})} P(x_{\overline{s}}|y) \\ \hat{x_s} &= \underset{x_s}{argmax} P(x_s|y) \end{aligned} \tag{20}$$

TABLE I

TWO-PASS COMPUTATION OF THE MPM ESTIMATES ON THE MRAT

parameters is an *incomplete data* problem, as only y is observed, while x remains hidden. Using only the incomplete (observed) data y, the maximum likelihood estimate of Φ_x,

$$\hat{\Phi}_x \triangleq \underset{\Phi_x}{argmax} \log P(y|\Phi_x)$$

becomes intractable. In order to cope with this problem, we use the *Expectation-Maximization (EM)* algorithm [29], by considering the expectation of the complete data (x, y). In particular, the EM algorithm iteratively computes the

maximum likelihood estimate of Φ_x by repeating, until convergence, the following steps:

- **E-step** (expectation) Computation of $Q(\Phi_x|\Phi_x^{(k)}) = \mathbb{E}[logP(x,y|\Phi_x)|y,\Phi_x^{(k)}]$.
- **M-step** (maximization) Update $\Phi_x^{(k+1)}$ such that: $\Phi_x^{(k+1)} = \underset{\Phi_x}{argmax}\, Q(\Phi_x|\Phi_x^{(k)})$

It is possible to implement the exact EM algorithm, in which both expectation and maximization can be conducted without any approximation. In particular, the expectation $Q(\Phi_x|\Phi_x^{(k)})$ is expressed as:

$$\begin{aligned} Q(\Phi_x|\Phi_x^{(k)}) &= \mathbb{E}\left[\log P(x,y|\Phi_x)|y,\Phi_x^{(k)}\right] \\ &= \textstyle\sum_{s\in S^N}\sum_{i\in L} P(X_s=i|y,\Phi_x^{(k)})\log P(X_s=i|y,\Phi_x) \\ &+ \textstyle\sum_{s\in S\setminus S^N}\sum_{(i,j)\in L^2} P(X_s=j,X_{\overline{s}}=i|y,\Phi_x^{(k)})\log P(X_s=j|X_{\overline{s}}=i,\Phi_x) \\ &+ \textstyle\sum_{s\in S}\sum_{i\in L} P(X_s=i|y,\Phi_x^{(k)})\log P(y_s|x_s) \end{aligned} \tag{21}$$

Let $\zeta_s^{(k)}(i) \triangleq P(X_s=i|y,\Phi_x^{(k)})$ and $\xi_s^{(k)}(i,j) \triangleq P(X_s=j,X_{\overline{s}}=i|y,\Phi_x^{(k)})$. The latter can be estimated during the downward pass on the MRAT, as:

$$P(X_s=j,X_{\overline{s}}=i|y) = \frac{P(X_s=j,X_{\overline{s}}=i|y_{\underline{\underline{s}}})}{\sum_{j\in L}P(X_s=j,X_{\overline{s}}=i|y_{\underline{\underline{s}}})}P(X_{\overline{s}}=i|y) \tag{22}$$

Using these notations and incorporating the unknown parameters $\pi_s(i)$ and θ_n in (21), we obtain:

$$\begin{aligned} Q(\Phi_x|\Phi_x^{(k)}) &= \textstyle\sum_{s\in S^N}\sum_{i\in L}\zeta_s^{(k)}(i)\log\pi_s(i) \\ &+ \textstyle\sum_{n=0}^{N-1}\sum_{s\in S^n}\left[\log\theta_n\sum_{i\in L}\xi_s^{(k)}(i,i) + \log\frac{1-\theta_n}{M-1}(1-\xi_s^{(k)}(i,i))\right] \\ &+ \textstyle\sum_{s\in S}\sum_{i\in L}\zeta_s^{(k)}(i)\log P(y_s|x_s) \end{aligned} \tag{23}$$

The maximization of $Q(\Phi_x|\Phi_x^{(k)})$ should satisfy the following two constraints:

$$\sum_{i\in L}\pi_s(i) = 1\,,\ \forall i\in L,\ \forall s\in S^N\ ;\ \sum_{j\in L}P(X_s=j|X_{\overline{s}}=i) = 1\,,\ \forall i\in L,\ \forall s\in S\setminus S^N \tag{24}$$

Using Lagrange multipliers one gets the following updates for both parameters $\pi_s(i)$ and θ_n:

$$\forall i\in L\,,\ \forall n<N\ \begin{cases} \pi_s^{(k+1)}(i) = \zeta_{s\in S^N}^{(k)}(i) \\ \theta_n^{(k+1)} = \frac{\sum_{s\in S^n}\sum_{i\in L}\xi_s^{(k)}(i,i)}{|S^n|} \end{cases} \tag{26}$$

Having all information in hand, the outline of the overall MRAT-MPM algorithm, from initialization to final classification is summarized in table II.

1) *Parameter initialization* $k = 0$:

$$\forall i \in L\,,\, \forall n < N \begin{cases} \pi_s^{(0)}(i) = \frac{1}{M} \\ \theta_n^{(0)} = 0.5 \end{cases} \tag{25}$$

2) *Estimation procedure*: two-pass computation (table I) of posterior marginals $P(X_s = i|y, \Phi_x^{(k)})\,,\, \forall i \in L$ and $P(X_s = j, X_{\overline{s}} = i|y_{\underline{\underline{s}}}, \Phi_x^{(k)})\,,\, \forall (i,j) \in L \times L$ (eq. (22)).

3) *Parameter updating*: The updates of the parameter vector $\Phi_x^{(k+1)}$ are estimated using (26).

4) *Repeat step 2 until* Φ_x *converges.*

5) *Classification step*: The modes of posterior marginal (MPM) provide the final classification map.

TABLE II

THE MRAT-MPM ALGORITHM

IV. EXPERIMENTAL RESULTS

A validation process is conducted in two ways: a) using the synthetic image of Fig.2(b), we assess the influence of the non-parametric dissimilarity measures (section III-B) to the overall performance of our method and perform an additional comparison with a parametric distance measure, which assumes Gaussian distributions for the image samples; b) we present a comparison of the MRAT-MPM approach with supervised versions of four pixel-based techniques that follow the Bayesian-Markovian framework either on the hierarchical graphs or the original image lattice, and present both qualitative (visual) and quantitative results, for the image of Fig.2(b) and an example of a multispectral remote sensing image, shown in Fig. 6(a). The latter corresponds to a 800x800 real color airborne image of an area near Saalbach, Austria.

The quantitative evaluation of the different classification maps is carried out, in the pixel level, via the concept of *confusion matrix* (a_{ij}) [30]. Given a validation set of $\mathcal{N}$ pixels with known ground-truth labels, a_{ij} stands for the number of pixels that are attributed to class i by the classifier and to class j by the ground-truth. Based on a $M \times M$ confusion matrix, the rates of correct classification were expressed in terms of two global indicators, namely the *overall accuracy* (τ) and the so-called *Kappa indicator* (κ):

$$\tau = \frac{1}{\mathcal{N}} \sum_i \mathrm{a}_{ii} \tag{27}$$

$$\kappa = \frac{\mathcal{N}\sum_i \mathsf{a}_{ii} - \sum_i \mathsf{a}_{i+}\mathsf{a}_{+i}}{\mathcal{N}^2 - \sum_i \mathsf{a}_{i+}\mathsf{a}_{+i}} \tag{28}$$

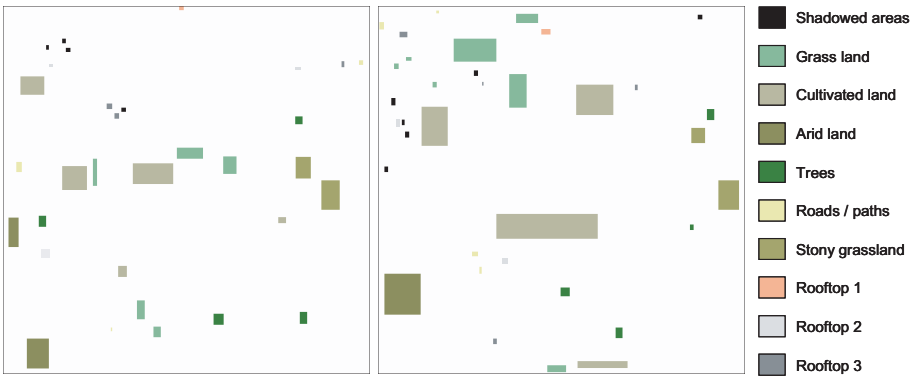
where $\mathsf{a}_{i+} = \sum_j \mathsf{a}_{ij}$ and $\mathsf{a}_{+i} = \sum_j \mathsf{a}_{ji}$.

In our experiments, a set of samples with known ground truth labels is collected and divided in two parts: a learning set, through which we extract the spectral signatures of each class, and a validation set, which is used to estimate the aforementioned evaluation measures. The total number of evaluation samples ($\mathcal{P}$) was set to $\mathcal{P} = 43259$ for the synthetic image of Fig.2(b) with six distinct classes ($M = 6$), and $\mathcal{P} = 72945$ for the airborne image of Fig. 6(a) with ten land cover classes ($M = 10$) . The exact location of these samples in the case of the image of Fig. 6(a) is shown in Fig. 6(b).

A. Comparison of dissimilarity measures

1) Initially, we evaluate the performance of our method using the non-parametric dissimilarity measures in (11) and (12), without any model assumptions about the data distributions. A very useful property of these measures is that they are both bounded in the interval $[0, 1]$, something that reduces the risk of underflow during the bottom-up pass of the MPM estimation (see section III-C). In all our experiments the value of λ in (10), was empirically set to $\lambda = 10$. This value satisfies a good trade-off between the influence of data likelihoods and the prior parent-child transitions. The classification performance on the image of Fig.2(b), using both dissimilarity measures, is presented in table III. The results of this analysis are in accordance with the findings of [26], showing that the χ^2-statistic has better discriminant properties than the Kolmogorov-Smirnov distance.
2) Considering again the synthetic example of Fig.2(b), we can now assume that the image samples are normally distributed and use the Mahalanobis distance as a parametric measure for the description of the data likelihoods. For the sake of direct comparison and avoiding possible problems of underflow during the MPM estimation, we have normalized the values of the Mahalanobis distance in the interval $[0, 1]$. The classification results, using this parametric distance measure are also presented in table III. Comparing the classification rates of all three measures, we can see that the χ^2-statistic outperforms its parametric counterpart, even in the case of strong Gaussianity assumptions. A visualization of the classification maps obtained using both χ^2-statistic and Mahalanobis distance is shown in Figures 7(e)-(f).

(a)



(b)

Fig. 6. (a) Real color airborne image of an area near Saalbach, Austria; (b) Ground-truth: training (left) and test (right) set.

B. Comparison with pixel-based approaches

1) *Hierarchical Markovian schemes*: As representatives of the stochastic, hierarchical schemes, we chose the ”Hierarchical MPM” (H-MPM) method of Laferté et al. [8] and the ”Sequential MAP” (SMAP) approach of Bouman and Shapiro [7]. Both schemes are based on hierarchical models defined as a coarse-to-fine Markov chain of levels, where the estimates (MPM and SMAP, respectively) are derived through a non-iterative, two-sweep inference procedure on the quadtree. For each class, the data likelihoods are described with a

Dissimilarity measure	**Fig.2(b)** τ / κ
χ^2-statistic	$98.4\%/97.8\%$
Kolmogorov-Smirnov distance	$97.1\%/95.9\%$
Mahalanobis distance	$97.8\%/96.9\%$

TABLE III

CLASSIFICATION PERFORMANCE

simple Gaussian model, defined by a mean vector and a covariance matrix, whereas the prior parameters of the hierarchical Markovian model are, in both cases, automatically estimated.

2) *Non-hierarchical MRF schemes*: In the case of single-resolution data, we derive approximated MAP estimates of a standard lattice-based classification model. This lattice-based model is defined with the same Gaussian likelihoods as in the SMAP and H-MPM approaches, and is based on a Potts prior on a first-order neighborhood. This non-hierarchical MAP estimate was obtained iteratively by two methods: (a) simulated annealing [5] (we denote the resulting estimate by NH-MAP); (b) a deterministic ICM algorithm [31] whose final classification will be referred as NH-ICM. These two non-hierarchical iterative algorithms are stopped when the number of actual updates, after a complete sweep of the image, falls below a given threshold (one per one thousand of the total number of pixels). The cooling schedule in the simulated annealing procedure is defined as $T_k = T_0/(1.01)^k$ where T_0 is the initial temperature, set to 100 and k stands for the current number of image sweeps.

Figures 7 and 8 illustrate the results of the five classification methods for the images of figures 2(b) and 6(a), respectively. For the MRAT-MPM approach we used the χ^2-statistic as distribution dissimilarity measure, whereas for the prior parameters Φ_x we obtained convergence after a maximum number of five iterations. The performance of each classification approach is summarized in table IV. As it can be noticed, in terms of accuracy, the MRAT-MPM scheme outperforms the pixel-based methods. This is also the case even when we use the assumption of Gaussian distribution for the data likelihoods (see Fig. 7 and last row of table III). From a visual inspection of figures 7

and 8, it is also evident that the quadtree-based H-MPM method exhibits block artifacts that are evident in both examples. These effects are alleviated by the SMAP, due to the none regular form of the hierarchy at the coarser levels. However, the non-contextual nature of both SMAP and H-MPM produces a less homogeneous classification map than the MRAT-MPM approach. The latter manages to implicitly encompass contextual information in the form of perceptually uniform regions while preserving smoothed and well localized object boundaries. Finally, both non-hierarchical schemes (NH-MAP and NH-ICM) although resulting in a compact classification map, they lack of accuracy (especially the deterministic NH-ICM scheme).

As far as the computational load is concerned, our method is the slowest compared to its pixel-based counterparts (SMAP and H-MPM), mainly because of the computational demands for the generation of the multiscale tower (described in section II-A). As the involved anisotropic diffusion process of equation (1) is steered by the image content, its convergence depends on the noise level and image complexity. For the severely noise-contaminated example of Fig. 2(b), the entire MRAT generation module, producing 6 discrete scales and a graph of order $|S| = 851$, lasted approximately 300 seconds on a Pentium III at 1 GHz. For the larger airborne image of Fig. 6(a), a MRAT of 10 scales and $|S| = 120491$ was constructed in approximately the same time. However, the complexity of the subsequent hierarchical labeling step, as being proportional to the order of the considered hierarchical graph, is in our case systematically lower than in the pixel-based hierarchical schemes. An indicative example concerns the image of Fig. 2(b), where the order of the quadtree is equal to 87381, almost 100 times larger than in the produced MRAT. Finally, regarding the complexity of the iterative non-hierarchical methods, the NH-MAP, as expected, exhibits the slowest convergence, whereas its deterministic counterpart (NH-ICM) is comparable with the hierarchical schemes.

V. Conclusions

Recent developments in the field of computer vision, such as scale space theory and Markovian modeling, can provide new horizons to the development of region-based classification algorithms in the area of remote sensing. The contribution of the proposed method to the state of the art of classification schemes is manifold. First of all, the sites under labeling are not individual pixels but correspond to perceptually uniform regions, which describe both contextual and spectral information. Although the generation of the multiscale tower is computationally demanding,

Method	**Fig.2(b)** τ / κ	**Fig.6(a)** τ / κ
NH-ICM	$94.2\%/91.9\%$	$90.9\%/87.3\%$
NH-MAP	$97.4\%/96.4\%$	$92.8\%/89.9\%$
H-MPM	$94.9\%/92.9\%$	$90.1\%/86.0\%$
SMAP	$97.3\%/96.3\%$	$95.4\%/93.6\%$
MRAT-MPM (χ^2)	$98.4\%/97.8\%$	$95.6\%/93.8\%$

TABLE IV

CLASSIFICATION PERFORMANCE

the use of a multiscale image structure, based on a non-linear diffusion scheme characterized by noise removal and a good localization of the boundaries of image structures, overcomes the geometric distortions of most pyramidal approaches. Moreover, the proposed hierarchical inference scheme results in high classification accuracies as shown in the previous section. Furthermore, the use of nonparametric dissimilarity measures offers an additional flexibility in the estimation of the data likelihoods, as the underlying probability distributions are not restricted to a given parametric model class. Note that the proposed classification method is presented in a supervised way. However, it could also be extended to an unsupervised classification scheme by using the described EM formulation of section III-D and updating both the prior parameters and the signature of the classes.

Future work will involve two main aspects of the proposed approach, namely an optimal MRAT generation and an extension of the field of observations. For the MRAT generation the following issues have to be considered:

- An automated localization scale selection [32] in order to alleviate the problems of underflow, mentioned in section III-C.
- The application of our approach to a variety of region hierarchies, like the morphological partitions of [33] or the region growing and clustering approach of [34].

Finally, the field of observations could be extended by including both spectral and texture information [35].

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of this work by the EC-IST-INFSO, ”Airborne minefield area reduction” project IST-2000-25300, and the Fund for Scientific Research - Flanders (FWO, Belgium) ”Multiscale Stochastic Models for Image Denoising and Segmentation” project FWOAL264.

REFERENCES

[1] R. A. Schowengerdt, *Remote sensing: Models and methods for image analysis*. Academic Press, 1996.

[2] P. Mantero, G. Moser, and S. Serpico, “Partially supervised classification of remote sensing images using SVM-based probability density estimation,” in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data (WARSD 2003)*, NASA Goddard Visitor Center, Greenbelt Maryland, USA, 2003.

[3] F. Roli, S. Serpico, and G. Vernazza, “Neural networks for classification of remotely sensed images,” in *Fuzzy Logic and Neural Network Handbook (Part 2, Chapter 15)*. McGraw-Hill Pub., 1996.

[4] F. Melgani and S. T. B.A.R. Al Hashemy, “An explicit fuzzy supervised classification method for multispectral remote sensing images,” *IEEE Trans. On Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 287–295, 2000.

[5] S. Geman and D. Geman, “Stohastic relaxation, Gibbs distributions, and Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.

[6] Q. Jackson and D. Landgrebe, “Adaptive Bayesian contextural classification based on Markov random fields,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40 (11), pp. 2454–2463, 2002.

[7] C. A. Bouman and M. Shapiro, “A multiscale random field model for Bayesian image segmentation,” *IEEE Trans. on Image Processing*, vol. 3 (2), pp. 162–177, 1994.

[8] J. M. Laferté, P. Perez, and F. Heitz, “Discrete markov modeling and inference on the quad-tree,” *IEEE Trans. Image Processing*, vol. 9, pp. 390–404, 2000.

[9] J.-N. Provost, C. Collet, P. Rostaing, P. Perez, and P. Bouthemy, “Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps,” *Computer Vision and Image Understanding*, vol. 93, pp. 155–174, 2004.

[10] M. Bister, J. Cornelis, and A. Rosenfeld, “A critical view of pyramid segmentation algorithms,” *Pattern Recognition Letters*, vol. 11, pp. 605–617, 1990.

[11] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, “Sonar image segmentation using an unsupervised hierarchical MRF model,” *IEEE Trans. on Image Processing*, vol. 9 (7), pp. 1216–1231, 2000.

[12] A. Chardin and P. Perez, “Semi-iterative inference with hierarchical models,” in *IEEE Int. Conf. on Image Processing (ICIP 1998)*, Chicago, USA, 1998, pp. 630–634.

[13] Z. Kato, M. Berthod, and J. Zerubia, “A hierarchical Markov random field model and multitemperature annealing for parallel image classification,” *Graphical Models and Image Processing*, vol. 58 (1), pp. 18–37, 1996.

[14] A. Katartzis, I. Vanhamel, and H. Sahli, "A hierarchical Markovian model for multiscale region-based classification of multispectral images," in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data (WARSD 2003)*, NASA Goddard Visitor Center, Greenbelt Maryland, USA, 2003.

[15] A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majumder, and D. K. Pal, "A MRF model-based segmentation approach to classification for multispectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, pp. 1102– 1113, 2002.

[16] J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

[17] A. Witkin, "Scale-space filtering," in *International Joint Conference on Artificial Intelligence*, vol. 2, Kalsruhe, Germany, 1983, pp. 1019–1022.

[18] I. Vanhamel, I. Pratikakis, and H. Sahli, "Multiscale gradient watersheds of color images," *IEEE Trans. on Image Processing*, vol. 12 (6), pp. 617–626, 2003.

[19] S. D. Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics and Image Processing*, vol. 33, pp. 116–125, 1986.

[20] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM Journal on Numerical Analysis*, vol. 29 (1), pp. 182–193, 1992.

[21] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12(7), pp. 629–639, 1990.

[22] Y.-L. You, M. Kaveh, W.-Y. Xu, and A. Tannenbaum, "Analysis and design of anisotropic diffusion for image processing," in *IEEE International Conference Image Processing (ICIP-94)*, vol. 3, Austin, TX USA, 1994, pp. 497–501.

[23] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. on Image Processing*, vol. 7 (3), pp. 421–432, 1998.

[24] I. Pratikakis, "Watershed-driven image segmentation," PhD Dissertation, Vrije Universiteit Brussel, 1998.

[25] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *7th European Conference on Computer Vision (ECCV 2002)*, Copenhagen, Denmark, 2002, pp. 661–675.

[26] J. Puzicha, J. M. Buhmann, Y. Rubner, and C. Tomasi, "Empirical evaluation of dissimilarity measures for color and texture," in *IEEE Int. Conf. on Computer Vision (ICCV 1999)*, Corfu, Greece, 1999, pp. 1165–1173.

[27] T. Hofmann, J. Puzicha, and J. Buhmann, "Unsupervised texture segmentation in a deterministic annealing framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20 (8), pp. 803–818, 1998.

[28] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12 (7), pp. 609–628, 1990.

[29] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal Statistical Society*, pp. 1–38, 1976.

[30] X. Zhuang, B. A. Engel, X. Xiong, and C. J. Johannsen, "Analysis of classification results of remotely sensed data and evaluation of classification algorithms," *Photogrammetric Engineering and Remote Sensing*, vol. 61, pp. 427–433, 1995.

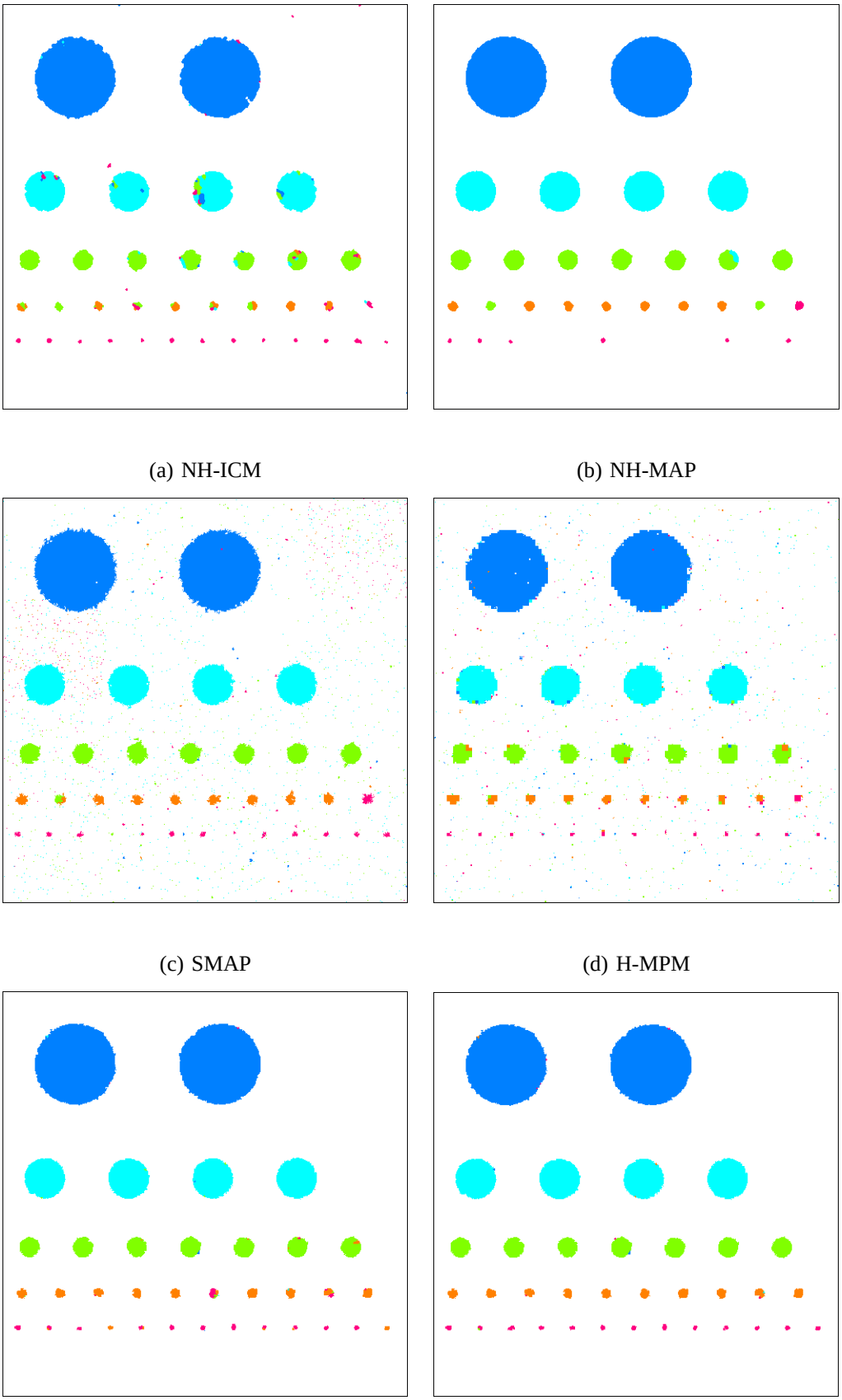
[31] J. Besag, "On the statistical analysis of dirty images," *J. Roy. Statist. Soc*, vol. B 48, p. 259, 1986.

[32] O. Olsen, "Multi-scale segmentation of grey-scale images," MsC Thesis, University of Copenhagen, 1996.

[33] F. Meyer, "Hierarchies of partitions and morphological segmentations," in *Scale-Space and Morphology in Computer Vision*, M. Kerckhove, Ed. Springer, 2001, pp. 161–182.

[34] J. Tilton, “Analysis of hierarchically related image segmentations,” in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data (WARSD 2003)*, NASA Goddard Visitor Center, Greenbelt Maryland, USA, 2003.

[35] I. Vanhamel, A. Katartzis, and H. Sahli, “Hierarchical segmentation via a diffusion scheme in color/texture feature space,” in *IEEE Int. Conf. on Image Processing (ICIP 2003)*, Barcelona, Spain, 2003.



(a) NH-ICM (b) NH-MAP

(c) SMAP (d) H-MPM

(e) MRAT-MPM (Mahalanobis) (f) MRAT-MPM (χ^2)

Fig. 7. Classification results on image of Fig. 2(b)

(a) NH-ICM

(b) NH-MAP

(c) SMAP

(d) H-MPM

(e) MRAT-MPM (χ^2)

Fig. 8. Classification results on image of Fig. 6(a)