

Unsupervised Bayesian wavelet domain segmentation using a Potts-Markov random field modeling

Patrice Brault [†] and Ali Mohammad-Djafari [‡]

[†] IEF, Institut d'Electronique Fondamentale, CNRS UMR-8622,
Université Orsay Paris-Sud, 91405 Orsay Cedex, FRANCE.

[‡]LSS, Laboratoire des Signaux et Systèmes, CNRS UMR-8506,
Supelec, Plateau du Moulon, 91192 Gif sur Yvette Cedex, FRANCE.
patrice.brault@lss.supelec.fr, djafari@lss.supelec.fr

Abstract

This paper describes a new fully unsupervised image segmentation method based on a Bayesian approach and a Potts-Markov Random Field (PMRF) model that are performed in the wavelet domain. A Bayesian segmentation model, based on a PMRF in the direct domain, has already been successfully developed and tested in [23, 12]. This model performs a fully unsupervised segmentation, on images composed of homogeneous regions, by introducing a hidden Markov Model (HMM) for the regions to be classified and Gaussian distributions for the noise and for the pixels pertaining to each regions. The computation of the posterior laws, deduced from these a priori distributions for the pixels, is done by a Markov Chain Monte Carlo (MCMC) approach and uses a Gibbs sampling algorithm. The use of a high number of iterations to reach convergence in a segmentation, where the number of segments, or “classes” labels, is important, makes the algorithm rather slow for the processing of a large quantity of data like image sequences [4, 5]. To overcome this problem we have taken advantage of the property of the wavelet coefficients, in an orthogonal decomposition, to be modeled by a mixture of two Gaussians. Thus, by projecting an observable noisy image in the wavelet domain, we are able to segment, in this same domain, the wavelet subbands in only two classes. After a decomposition up to a scale J , the main idea is to segment the coarse, and small, approximation subband with a high number of classes, and to segment all the detail (wavelet) subbands with only two classes. The segmented wavelet domain coefficients are then reconstructed to obtain a final segmented image in the direct domain. Our tests on synthetic and natural images show that the segmentation quality stays good, even with noisy images, and shows that the segmentation times can be significantly reduced.

Keywords.

Unsupervised segmentation, noisy images, wavelet coefficients domain, Orthogonal Wavelet Decomposition (OWT), unsupervised Bayesian classification, Potts-Markov Random Field (PMRF), Markov Chain Monte Carlo (MCMC) Gibbs sampling.

1 Introduction

Numerous methods for the segmentation of images have been developed and are today available. Statistical methods to which are closed to our approach can be classified into two main groups :

- Contour based methods. The extraction of contours in an image provides a first way to implement the segmentation of an image into regions. Markovian modeling as well as *Maximum a Posteriori* (MAP) deterministic optimization algorithms, like the Graduated Non-Convexity (GNC) [21] and the Mean Field Annealing (MFA) [27], are used in contour-based methods.
- Region based methods :

The *first type* is based on the *mono-dimensional* modeling of the histogram, with inter-modes thresholding and Multi-Gaussian modeling. Also when the image is considered like a realization of a discrete random process, local or global, stochastic attributes can be used. A second type is represented by *clustering* methods [1], hierarchical or not, like K-means, Fuzzy C-means or Cluster, which is also called “unsupervised learning” [17], pages xiii and 134, and works by organizing data following an underlying structure that groups individuals or makes a hierarchy of groups. A third type is finally the one to which our method pertains and englobes the Markovian methods, supervised and unsupervised. In this last category, we must cite one

of the most important works using the regularization and pseudo Maximum Likelihood (pML) in a supervised segmentation of textured images [13, 14]. More recently, numerous works have been made in the development on wavelet features based models in a Bayesian segmentation approach.

Our work pertains to the group of region-based segmentation methods and in particular of unsupervised segmentation methods. A brief presentation of unsupervised segmentation methods could start with multivariate Gaussian Markov Random Field (GMRF), models that have been extensively applied for the segmentation of still images [15].

Since several years, the multiscale Bayesian approaches have demonstrated good results in the domain of still image segmentation as well as recently in the domain of animated sequences segmentation. The supervised, partially-unsupervised and fully unsupervised segmentation of still images with multiscale Bayesian approaches has been widely studied and described in the literature. Under the Bayesian framework, multiscale approaches have proven to efficiently integrate image features (like wavelet coefficients) as well as contextual information (like labels in a HMM approach) for the classification. Image features can be represented by different statistical image models. Contextual information can be obtained by using multiscale contextual models like, for example, the interscale dependency of class labels between scales in a multiscale approach [2]. Also, the Joint-Multicontext and Multiscale Segmentation (JMCMs) was developed by using the fusion between intra-scale and inter-scale information [10]. In order to fusing the multiscale contextual information, several methods have been used : a Multiscale Random Field (MRF), combined with a *sequential maximum a posteriori estimator* (SMAP), was developed [2, 6].

More recently, the transfer of the observed model to a dual domain like the wavelet domain has enabled to fully take advantage of the very interesting property of wavelet coefficients to be modeled : intra scale, by an Independent Gaussians Mixture (IGM) and inter scales by considering the evolution of each wavelet coefficient. Models using the evolution of information between scales, like the wavelet domain Hidden Markov Tree (HMT) [9, 20], and an improved version using the dependencies across subbands (HTM-3S) [11], have been developed. Also in [7] a multiscale segmentation was developed based on the HMT and the inter-scale fusion of contextual information. A method has also been proposed recently to realize the unsupervised segmentation using HMM in the wavelet domain together with clustering methods [25, 24].

In our group, we have developed a fully unsupervised segmentation also based on a Bayesian approach in the direct domain [23, 12]. This method uses a HMM for the classification label assigned to the different regions of an image. The difference of this HMM compared with HMT models is that it is based on a Potts-Markov Random Field (PMRF) for the pixels. This enables to render the strength of dependency of neighbor pixels to insure a good homogeneity of the region being segmented. In this direct domain approach the PMRF uses a first order neighborhood. More recently, we have used the same model to segment $2D + T$ videos sequences, i.e. regularly sampled sequences of images [4]. We have thus demonstrated a significant improvement in the segmentation of a sequence of images. Turning this result to motion detection and quantification and also to sequence compression, is being under way and might turn to be an efficient method amongst other recent developments for

motion acquisition, quantification and compression.

Nevertheless, though very efficient for long term segmentation processes, we are also very much interested in reducing the segmentation time. This has motivated a new approach based on the transfer of the segmentation process in the wavelet transform domain to perform the segmentation. Due to their specific property of fast “local-decay”, the wavelet coefficients, obtained by decomposition on an orthogonal basis, behave like a mixture of zero-centered Gaussians [18, 9, 20]: a first, high variance, distribution representative of a few strong coefficients of major importance, and a second, “peaky” low variance distribution representative of a large number of low magnitude and low importance coefficients. Thus the segmentation of these band-pass “subband” coefficients can be made using only 2 classes labels. Furthermore, by zeroing coefficients pertaining to the class of low importance coefficients, we denoise the data and increase the performances of the segmentation, see recent works on wavelet denoising using Scale Mixture of Gaussians (SMG) in [19]. Based now on a decomposition in the wavelet domain we found also pertinent to improve the PMRF model used in the direct domain. This is why we adopted a first + second order, 8-connexity model, for the PMRF. This was motivated by the fact that wavelet subbands are split in vertical, diagonal and horizontal details. We thus tuned the Markov Field neighborhood to the wavelet subbands orientations through a specific α dependency parameter for the Potts model [26]. So we have naturally fixed three independent parameters: α_V , α_D and α_H as dependency parameters for our new three-orientations Potts model.

The rest of the paper is organized as follows: Section 2 describes our PMRF-Bayesian unsupervised segmentation method in the direct domain. Section 3 points out the important properties of the wavelet coefficients and what motivated the choice of this transform domain to perform the equivalent “dual” of our direct domain segmentation algorithm. Section 4 introduces our new 8-connexity PMRF for the Bayesian segmentation. Section 5 describes in detail the segmentation algorithm in the wavelet domain, its initialization with the segmentation of its approximation subband, the segmentation of its multiscale detail subbands in a coarse to fine scale scheme and the final inverse wavelet transform to get the final segmentation in the direct domain. Sections 6 and 7 make a comparison of the results of our new wavelet-domain Bayesian segmentation with the same method in the direct domain and with other techniques and recall the important features and advantages of this new unsupervised Bayesian segmentation in the wavelet domain.

2 Unsupervised segmentation in the direct domain

2.1 Potts-Markov field modeling and Bayesian approach

The description of still image segmentation and fusion by a Bayesian approach and a Hidden Markov Model (HMM) has been developed and well described by Feron et al. in [12]. We summarize in this section the main steps for the computation of the posterior laws necessary in the final segmentation-fusion algorithm.

We formulate the segmentation problem as an inverse problem where the objective is to find a classification $z(\mathbf{r})$ of an original

image $f(\mathbf{r})$. The observed image $g(\mathbf{r})$ is assumed to be a noisy version of the original image $f(\mathbf{r})$:

$$g(\mathbf{r}) = f(\mathbf{r}) + \epsilon(\mathbf{r}), \quad \mathbf{r} \in \mathcal{R} \quad (1)$$

where \mathcal{R} is the set of sites \mathbf{r} of the image.

If we assume the noise $\epsilon(\mathbf{r})$ to be centered, white and Gaussian. We have :

$$\epsilon(\mathbf{r}) \propto \mathcal{N}(\epsilon(\mathbf{r})|0, v_\epsilon) \quad \implies \quad p(g(\mathbf{r})|f(\mathbf{r})) = \mathcal{N}(g(\mathbf{r})|f(\mathbf{r}), v_\epsilon) \quad \forall \mathbf{r} \in \mathcal{R}$$

If we note by the vectors $\mathbf{g} = \{g(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$, $\mathbf{f} = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ and $\boldsymbol{\epsilon} = \{\epsilon(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ the discretized version of the images, we can write :

$$p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{f}, v_\epsilon \mathbf{I}) \quad (2)$$

where \mathbf{I} represents the identity matrix.

Because the goal is to have a reconstructed image \mathbf{f} segmented in a limited number of statistically homogeneous regions, a hidden variable z , which can take the discrete “label” values $k \in \{1, \dots, K\}$, is introduced to represent the image \mathbf{f} classified in K classes. Such a classification enables to segment \mathbf{f} in regions $\mathcal{R}_k = \{\mathbf{r} : z(\mathbf{r}) = k\}$. When regions pertain to the same class but are not contiguous, the number of segments increases, thus the number of segments is at least equal to the number of classes. At this point we also make the hypothesis that each region of the segmented image is modeled by a Gaussian distribution of mean m_k and variance v_k :

$$p(f(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(f(\mathbf{r})|m_k, v_k) \quad (3)$$

So we finally write a model for the distribution of the pixels of the image $f(\mathbf{r})$ as :

$$p(f(\mathbf{r})) = \sum_{k=1}^K \alpha_k \mathcal{N}(f(\mathbf{r})|m_k, v_k), \quad \text{with} \quad \alpha_k = p(z(\mathbf{r}) = k) \quad (4)$$

We can also write the dual expression for the observable $g(\mathbf{r})$:

$$p(g(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(g(\mathbf{r})|m_k, v_k + v_\epsilon) \quad (5)$$

and

$$p(g(\mathbf{r})) = \sum_{k=1}^K \alpha_k \mathcal{N}(g(\mathbf{r})|m_k, v_k + v_\epsilon), \quad \text{with} \quad \alpha_k = p(z(\mathbf{r}) = k) \quad (6)$$

In order to build homogeneous regions, the spatial dependency between the label of each pixel (the hidden Markov variable $z(\mathbf{r})$) and the label of its neighbors is modeled by a Potts-Markov Random Field (PMRF). The Markovian modeling assumes

that the value of $z(r)$ at a pixel position is related to the value of its neighbors (the four closest vertical and horizontal neighbors in a first order neighboring). The Potts model enables to control, by means of an attractive/repulsive α parameter, the mean value of the size of a region. Thus the homogeneity for each class is proportional to the strength of α .

$$p(z(\mathbf{r}), \mathbf{r} \in \mathcal{R}) = \frac{1}{T(\alpha)} \exp \left\{ \alpha \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{s} \in V(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{s})) \right\} \quad (7)$$

where $V(\mathbf{r})$ represents the neighborhood of \mathbf{r} . In the sequel of this section we consider $V(\mathbf{r})$ as the first order neighborhood (4-connexity) of the pixel \mathbf{r} .

More explicitly, if we consider a first order neighborhood, this one includes the four pixels of the horizontal and vertical neighborhoods. The Potts model can be written explicitly with the indices (i, j) of each pixel $r(i, j)$:

$$p(z(i, j), (i, j) \in \mathcal{R}) = \frac{1}{T(\alpha)} \times \exp \left\{ \alpha_V \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i - 1, j)) \right. \\ \left. + \alpha_H \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i, j - 1)) \right\} \quad (8)$$

where we assume $\alpha_V = \alpha_H = \alpha$.

Using the Bayes rule, the joint posterior law of \mathbf{f} and \mathbf{z} can then be expressed :

$$p(\mathbf{f}, \mathbf{z} | \mathbf{g}) \propto p(\mathbf{g} | \mathbf{f}, \mathbf{z}) p(\mathbf{f} | \mathbf{z}) p(\mathbf{z}) \quad (9)$$

The prior laws, defined formerly, need some of their parameters, which are called the model *hyperparameters*, to be defined.

These are v_ϵ , v_k and m_k . If we want to realize an unsupervised segmentation, the set $\boldsymbol{\theta}$ of so-called “hyperparameters” :

$$\boldsymbol{\theta} = \left\{ v_\epsilon, (m_k, v_k), \quad k \in \{1 \dots K\} \right\} \quad (10)$$

has also to be estimated. For this purpose we also assign *prior* laws to $\boldsymbol{\theta}$. These prior laws are taken as *conjugate priors* and depend themselves on *hyper-hyperparameters* which are α_0 , β_0 , m_0 and v_0 . We refer to [23] for the choice and the values of these final hyper-hyperparameters. The *priors* for the set $\boldsymbol{\theta}$ then takes the form :

$$\begin{cases} p(v_\epsilon) & \sim \mathcal{IG}(v_\epsilon | \alpha_0^\epsilon, \beta_0^\epsilon) \\ p(m_k) & \sim \mathcal{N}(m_k | m_0^k, v_0^k), \quad k = \{1 \dots K\} \\ p(v_k) & \sim \mathcal{IG}(v_k | \alpha_0^k, \beta_0^k), \quad k = \{1 \dots K\} \end{cases} \quad (11)$$

where \mathcal{IG} is the Inverse-Gamma.

The expression of the posterior law for an unsupervised segmentation becomes finally :

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}) \propto p(\mathbf{g} | \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}) p(\boldsymbol{\theta}) \quad (12)$$

2.2 Markov-Chain Monte-Carlo (MCMC) and Gibbs sampling algorithm

The Bayesian approach consists now in estimating the whole set of variables $(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ following the joint *a posteriori* distribution $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g})$ after Eq. 12. The MCMC method consists in generating samples from this *posterior* law from which we can estimate the mean, the median or any other statistics for each variable \mathbf{f} , \mathbf{z} or $\boldsymbol{\theta}$. For example the mean value of \mathbf{f} becomes :

$$\hat{\mathbf{f}} = \int \mathbf{f} \cdot p(\mathbf{f} | \mathbf{g}) d\mathbf{f} \simeq \frac{1}{N} \sum_{n=1}^N \mathbf{f}^{(n)} \quad (13)$$

where $\mathbf{f}^{(n)}$ are samples drawn from $p(\mathbf{f} | \mathbf{g})$.

To generate samples from the joint distribution in Eq. 12, we use the following Gibbs sampling algorithm :

$$\begin{cases} \mathbf{f}^n & \sim p(\mathbf{f} | \mathbf{g}, \mathbf{z}^{(n-1)}, \boldsymbol{\theta}^{(n-1)}) \\ \mathbf{z}^n & \sim p(\mathbf{z} | \mathbf{g}, \boldsymbol{\theta}^{(n-1)}, \mathbf{f}^{(n-1)}) \\ \boldsymbol{\theta}^n & \sim p(\boldsymbol{\theta} | \mathbf{g}, \mathbf{z}^{(n-1)}, \mathbf{f}^{(n-1)}) \end{cases} \quad (14)$$

where we need to write the expressions of those three posterior laws :

1) for \mathbf{f}

$$\begin{aligned} p(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta}) & \propto p(\mathbf{g} | \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}) \\ & \propto \prod_k \prod_{\mathbf{r} \in R_k} \mathcal{N}(g(\mathbf{r}) | f(\mathbf{r}), v_\epsilon) \mathcal{N}(f(\mathbf{r}) | m_k, v_k) \\ & \propto \prod_k \prod_{\mathbf{r} \in R_k} \mathcal{N}(f(\mathbf{r}) | \hat{m}_k, \hat{v}_k) \end{aligned} \quad (15)$$

with $\hat{m}_k = \hat{v}_k \left(\frac{m_k}{v_k} + \frac{\sum_{\mathbf{r} \in R_k} g(\mathbf{r})}{v_\epsilon} \right)$, $\hat{v}_k = \left(\frac{1}{v_\epsilon} + \frac{1}{v_k} \right)^{-1}$ and $R_k = \{\mathbf{r} : z(\mathbf{r}) = k\}$

2) for \mathbf{z}

$$\begin{aligned} p(\mathbf{z} | \mathbf{g}, \mathbf{f}, \boldsymbol{\theta}) & \propto p(\mathbf{g} | \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} | \boldsymbol{\theta}) \\ & \propto \left[\prod_k p(\mathbf{g}_k | \mathbf{f}_k, v_\epsilon) p(\mathbf{f}_k | m_k, v_k) \right] p(\mathbf{z}) \\ & \propto \left[\prod_k \prod_{\mathbf{r} \in R_k} \mathcal{N}(f(\mathbf{r}) | \hat{m}_k, \hat{v}_k) \right] p(\mathbf{z}) \end{aligned} \quad (16)$$

We can notice that this posterior law is also a PMRF where the prior probabilities are weighted by the posterior likelihood.

3) for $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta}|\mathbf{f}, \mathbf{g}, \mathbf{z}) \propto p(v_\epsilon|\mathbf{f}, \mathbf{g}) \prod_k p(m_k|v_k, \mathbf{f}, \mathbf{z}) \cdot p(v_k|\mathbf{f}, \mathbf{z}) \quad (17)$$

where :

- For the noise : $p(v_\epsilon|\mathbf{f}, \mathbf{g}) \propto \mathcal{IG}(v_\epsilon, |\alpha, \beta|)$, with $\alpha = n/2 + \alpha_0^\epsilon$ and $\beta = \frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} (g(\mathbf{r}) - f(\mathbf{r}))^2 + \beta_0^\epsilon$
and $n = \text{Card}(R)$, is the total number of pixels.

- For the mean in each region k : $p(m_k|\mathbf{f}, \mathbf{z}, v_k, m_0, v_0) \propto \mathcal{N}(m_k|\mu_k, \xi_k)$
with $\mu_k = \xi_k \left(\frac{m_0^k}{v_0^k} + \frac{\sum_{\mathbf{r} \in \mathcal{R}_k} f(\mathbf{r})}{v_k} \right)$, $\xi_k = \left(\frac{n_k}{v_k} + \frac{1}{v_0^k} \right)^{-1}$, and $n_k = \text{Card}(R_k)$

- For the variance in each region k : $p(v_k) \propto \mathcal{IG}(v_k|\alpha_k, \beta_k)$
with $\alpha_k = \alpha_0^k + \frac{n_k}{2}$, and $\beta_0^k = \beta_0 + \frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}_k} (f(\mathbf{r}) - m_k)^2$

This algorithm is iterated a “sufficient” number of times (*itermax*) in order to reach the convergence of the segmentation. We do not use any real “convergence criterion” and by convergence we mean that the segmented regions do not change significantly in the next iterations. After convergence, we take the max(histogram) for each pixel value and for all iterations. Also, our experience on the images analyzed has showed us that *itermax* depends essentially on the complexity of the image and the number of luminance levels, as well as on the number of classes taken for the segmentation. In general for a number of classes $K = 4$ we take *itermax* between 20 and 50. If we generate a number of samples *itermax* = N ,

$$(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})^{(1)}, (\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})^{(2)}, \dots, (\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})^{(L)}, \dots, (\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})^{(N)}$$

the algorithm starts providing homogeneous segments only after a “heating time” of L samples. The final value for each pixel is given by the median, the maximum of its histogram or the mean of the $(N - L)$ last values :

$$(\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}) \simeq \frac{1}{(N - L)} \sum_{n=L+1}^N (\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})^n \quad (18)$$

3 Projection in the wavelet domain

As mentioned in the introduction, our Bayesian segmentation method, based on a Potts-Markov random field modeling in the direct domain, gives good classification results. Nevertheless the main drawback of such an algorithm, using iterative sampling, is to exhibit very important computation times. For one of our concerns is to perform the segmentation on video sequences, we are interested in lowering these computation times. Using the projection of the observable onto a transform domain can provide interesting properties. In particular the wavelet domain coefficients exhibit properties that are very inter-

esting for our segmentation application :

- It gives a sparse representation of the observable.
- It presents the property of fast local decay of the wavelet coefficients that enables a modeling of these coefficients by a mixture of two Gaussians.
- It reflects, by analyzing the evolution of the wavelet coefficients between scales, the strength of singularities (Hölder coefficients) of each pixel of an image, and in the same way can model the dependency of pixels between scales.

More precisely it has been established [8, 9, 20] that the wavelet coefficients of “real-world” signals exhibit a local decay property which means that the coefficients of highest energy, and of utmost representativity for the signal, are very sparse and that the coefficients of low energy, and of low importance for the signal, are in large quantity. In the probability domain, we can then express the marginal density of the wavelet coefficients by one spread and heavy-tailed, i.e. with a large variance, Gaussian density, for the important coefficients, and by another very peaky, and low variance, Gaussian density for the coefficients of low importance. Such a property is well expressed by a mixture of Gaussians and the coefficients can be considered approximately decorrelated due to the Orthogonal Wavelet Transform (OWT) decomposition used (Fig. 2).

As a result of this we may notice that the model of mixture of two Gaussians can be very interesting when performing our Bayesian segmentation in the wavelet domain. It means that the segmentation in the wavelet subbands can be done by using only two classes. Then making a projection of the observable g , segmenting the coarse approximation subband with a high number of classes and finally segmenting the successive detail subbands with only two classes up to the image resolution significantly reduces the segmentation cost.

We will now introduce the wavelet decomposition, then will describe the complete segmentation algorithm in the wavelet domain.

In order to proceed with the same Bayesian segmentation approach in the wavelet domain, we first need to decompose the observable g on an orthogonal basis of scaling and wavelet functions. We use the classical multiresolution pyramidal decomposition of Mallat. This decomposition uses shifted and dilated versions of the scaling ϕ and wavelet ψ functions.

The observable g is written as a function of its decomposition coefficients, a_{J,b_1,b_2} and d_{j,b_1,b_2}^B , as :

$$g(x, y) = \sum_{(b_1, b_2) \in \mathbb{Z}} a_{J,b_1,b_2} \phi_{J,b_1,b_2}^{LL}(x, y) + \sum_{B \in \mathcal{B}} \sum_{j \leq J} \sum_{(b_1, b_2) \in \mathbb{Z}} d_{j,b_1,b_2}^B \psi_{j,b_1,b_2}^B(x, y) \quad (19)$$

where $\phi_{j,b_1,b_2}^{LL} = 2^{-j} \phi(2^{-j}x - b_1, 2^{-j}y - b_2)$, $\psi_{j,b_1,b_2}^B = 2^{-j} \psi^B(2^{-j}x - b_1, 2^{-j}y - b_2)$ and $\mathcal{B} = \{HL, LH, HH\}$. The HL, LH and HH are called the details, or wavelet, subbands. L and H represent the low and high pass band conjugate mirror filters, respectively h and g . Thus HL correspond to the vertical, LH to the horizontal and HH to the diagonal subband. LL is called the approximation, or scaling, subband.

The decomposition coefficients are expressed as :

$$a_{J,b_1,b_2} = \int_{\mathbb{R}^2} g(x,y) \phi_{J,b_1,b_2}^{LL} dx dy \quad (20)$$

and

$$d_{J,b_1,b_2}^B = \int_{\mathbb{R}^2} g(x,y) \psi_{J,b_1,b_2}^B dx dy \quad (21)$$

In the sequel we will use the notation V_J , W_j^V , W_j^D and W_j^H , respectively for the subbands LL,HL,LL and LH. W_j^V , W_j^D and W_j^H are respectively the vertical, diagonal and horizontal detail subbands. The corresponding wavelet filters are given by :

$$\begin{cases} W_j^V & \rightarrow \psi_j \phi_j[\vec{b}] \\ W_j^D & \rightarrow \psi_j \psi_j[\vec{b}] \\ W_j^H & \rightarrow \phi_j \psi_j[\vec{b}] \end{cases} \quad (22)$$

The decomposition of our observable g is done from the initial resolution $L = 0$ or 2^L up to the scale J or 2^J . Decomposing on two scale means that $J = 2$ and the scaling parameter is $j \in 0, 1, \dots, J$. The corresponding resolution to a scale is obtained by the inverse power 2^{-j} . The confusion between scale and resolution is easy and in the sequel we will try to stay as clear as possible in the description of the algorithm starting from the coarsest scale (the highest j).

With this notation, the decomposition of our observable \tilde{g} in the wavelet domain can also be expressed as :

$$\tilde{g}(g, V_j, W_j) = P(g, V_J) + \sum_{j=1}^J P(g, W_j) \quad (23)$$

The figures below describe the application of the wavelet orthogonal transform on one band of an hyperspectral (224 spectral bands) image of a satellite view. The figures 2 b) and c) show respectively the histograms of the approximation coarse subband and of the detail diagonal coarse subband. As previously said, the first histogram can be modeled by a mixture of several independent Gaussians, which motivates the segmentation of this approximation subband in a high number of labels. Conversely the second histogram of the detail diagonal subband can be modeled by a mixture of two independent Gaussians (which is more visible in the lin-log representation), one with a high variance (sparse high value coefficients) and one with a low variance (numerous low value coefficients) and the segmentation in all detail subbands is thus done with only two labels.

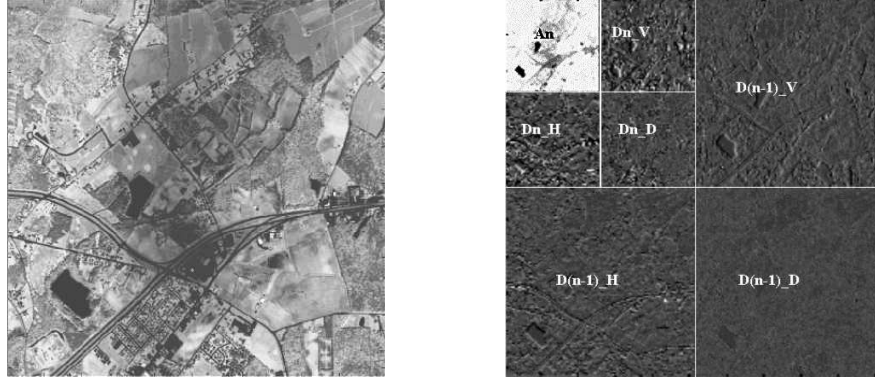


Figure 1: a) Original image of the Channel 100 of an hyperspectral satellite image composed of 224 channels. b) Pyramidal representation of the Fast Orthogonal Wavelet 2-levels decomposition applied to the the channel 100 of our hyperspectral image. Here the letter n can be replaced by J , the scale parameter at the coarsest scale 2^J bands.

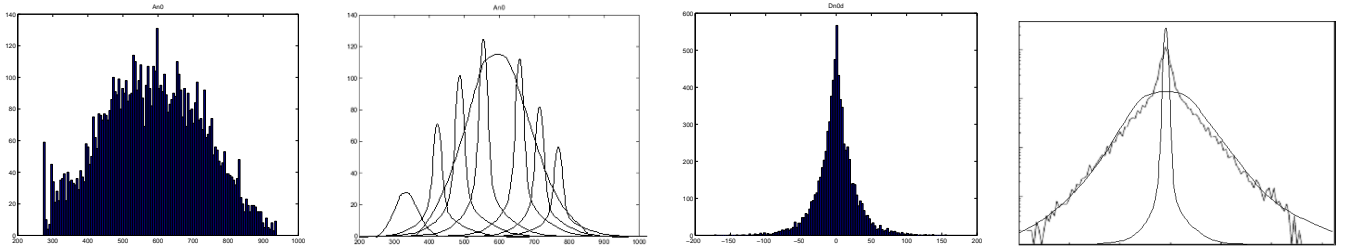


Figure 2: a) Histogram of the A_J (or A_n), the approximation coefficients at the coarsest scale. b) Histogram of A_J described by a mixture of multiple Gaussians, one for each class. c) Histogram of the D_J^D (or D_n^D), the detail diagonal coefficients at the coarsest scale. d) Linlog histogram of the D_J^D explaining the choice of a two independent Gaussians mixture model, with one large and one small variance.

4 Eight connexity, first and second order Potts-Markov random field

4.1 Model description

In order to find statistically homogeneous regions for the segments, our Markovian segmentation method in the direct domain uses a Potts-Markov random field [12] to define the spatial dependency of the labels in the HMM. In this direct model, the dependency of the pixel label is searched in a first order neighborhood. Nevertheless and due to the fact that the segmentation is done now in the wavelet domain, we have to take into account that this spatial dependency is quite different than in the direct domain due to the fact that the three “detail” subbands are oriented in the vertical, diagonal ($D_1 = \pi/4$ and $D_2 = 3\pi/4$) and horizontal directions. This prior knowledge is taken into account by introducing these orientations in a new “eight-connexity” (or neighborhood of orders 1 and 2) Potts-Markov Random Field. The new model of the PMRF can be written:

$$p(z(i, j), (i, j) \in \mathcal{R}) = \frac{1}{T(\alpha_V, \alpha_{D_1}, \alpha_{D_2}, \alpha_H)} \times \exp \left\{ \begin{aligned} & +\alpha_V \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i-1, j)) \\ & +\alpha_{D_1} \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i+1, j-1)) \\ & +\alpha_{D_2} \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i-1, j-1)) \\ & +\alpha_H \sum_{(i, j) \in \mathcal{R}} \delta(z(i, j) - z(i, j-1)) \end{aligned} \right\} \quad (24)$$

The parameters α_V , α_{D_1} , α_{D_2} and α_H respectively control the degree of spatial dependency of the z variable in the directions V , D_1 , D_2 and H . For the wavelet “diagonal” subband we gather the D_1 and D_2 dependencies in one diagonal $D_1 D_2$ dependency. For the first, coarse, approximation subband, we also gather the V and H dependency in a VH dependency which is equivalent to the first-order dependency (first order neighborhood) used in the direct-domain implementation.

Also, in order to accelerate the sampling of the image sites, we implement the MCMC-Gibbs algorithm “in parallel” [12]. If we consider, for the VH dependency, that the label value of one site is conditional to the label values of the sites of its first order neighboring ($p(z(i, j)|z(i, j-1), z(i-1, j))$), then all the pixel sites corresponding to the white (odd numbered) cases of a chessboard can be considered as independent (Fig. 4 a)). In a same way all the black cases (even numbered) of this chessboard can be considered as independent conditionally to the values of the white sites in their first order neighborhood. Thus we are able, once all the white cases are sampled, to sample all the black cases in only one iteration due to the independency of the sites. With this scheme, the whole image can be sampled in two successive iterations, one on the black sites and one on the white sites.

Similarly, for the $D_1 D_2$ dependencies, we can also use an implementation “in parallel” of the Gibbs sampler, but by using a different scheme for the independent sites. The image is then split in two sets of interlaced black and white sites which enables to consider each type of sites as independent conditionally to the knowledge of the sites of the second order neighboring.

The application of our new PMRF model to the wavelet subbands is done by deleting the PMRF terms that do not apply to the subband concerned. Moreover the α_V parameter used in this “vertical” term can be adjusted to any positive value, which means that we assign a specified, non null, dependency strength between the labels z of the region concerned. As we have seen in [12], this means that the higher the α parameter, the higher the *a priori* of a little number of great homogeneous regions. For the first, coarse, approximation subband, we also gather the V and H dependency in a VH dependency which is equivalent to the first-order dependency used in the direct-domain implementation.

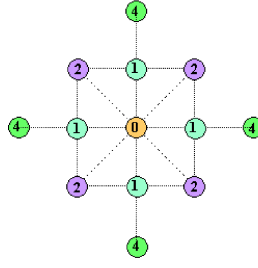


Figure 3: First, second and fourth order Markov fields (after [16]). The first order neighboring is used in the parallel implementation of the Gibbs sampler for the VH dependency (scaling subband). The second order neighboring is used in the parallel implementation of the D_1D_2 dependency (diagonal wavelet subband)

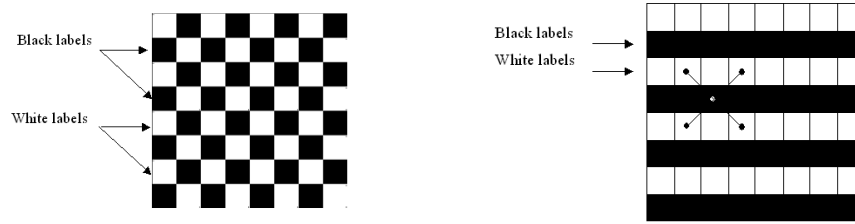


Figure 4: Parallel implementation of the Gibbs sampling algorithm. By considering the image as divided in two sets of independent white and black sites, one iteration of the Gibbs sampler for the whole image can done in only two times : one time on the black sites and the second on the white sites. The two sets of sites are build differently depending on the directions of the neighbor sites considered : a) set of independent sites for the vertical and horizontal directions (first order neighboring) b) set of independent sites for the diagonal directions (second order neighboring) .

4.2 Application to the wavelet decomposition

The application of this model to the wavelet subbands is done by deleting the PMRF terms, in expression 24, that do not apply to the subband concerned. For example if we want to segment the coefficients in the V subband, we apply only the corresponding PMRF term, i.e., the one that deals with the V dependency. Moreover the α_V parameter used in this “vertical” term can be adjusted to any value > 0 , which means that we assign a specified, non null, dependency strength between the labels z of the region concerned. As we have seen in [12], this means that the higher the α parameter, the higher the *a priori* of a little number of great homogeneous regions. For most of our applications, the number of classes is relatively low ($K < 10$) and we thus can take a value of α relatively high, i.e. equal to 1.

5 Bayesian segmentation in the wavelet domain

The initial operation to perform on the image to be segmented, is first to do its decomposition on an orthogonal wavelet basis. The decomposition used here is the classical Mallat pyramidal transform of complexity $O(N^2)$. This transform enables to get wavelet coefficients which are split in two main classes : the weak and the strong coefficients. Noticing this point on the wavelet coefficients has an important incidence for segmentation purposes : it enables to segment the “detail” images with only two classes ($K = 2$). On complex images, i.e. the first “approximation” subband, the segmentation will be led with a higher number of classes, e.g. $K = 8$.

Algorithm description

Our segmentation scheme in the wavelet domain is based on a coarse to fine scale scheme (Fig. 5). The Bayesian segmentation is performed on the wavelet subband coefficients and at all scales with a low value $K = 2$, except for the approximation coefficients. The segmentation starts from the g observation projected into the wavelet domain. The segmentation algorithm can be described with the following steps:

- 1) Wavelet decomposition to the order J , i.e. down to scale 2^J , on an orthogonal basis (Haar wavelet). This wavelet is chosen for its property w.r.t. image discontinuities, a property that we are going to use within the next steps.
- 2) We segment the approximation coefficients V_J at scale 2^J with the number of classes desired in the final segmentation, e.g. $K = 8$. For the Gibbs sampling, and with a large value of K , the iteration number is given a high value in order to assure the convergence.
- 3) In the segmented image of the approximations $z(V_J)$, we detect (by derivation) the regions exhibiting vertical, diagonal ($\pi/4$ and $3\pi/4$) and horizontal discontinuities.
- 4) At this same scale, the 3 detail subbands W_J^V , W_J^D and W_J^H are segmented and we respectively use, as an initialization of these segmentations, the 3 subsets of discontinuities $diff_V$, $diff_D$ and $diff_H$ computed at the former step. This step is realized with a weak number of classes ($K = 2$) and of iterations of the Gibbs sampling.
- 5) The 3 segmented detail subbands are upsampled by 2 in order to repeat the process at the next upper level.
- 6) We segment the 3 detail subbands $W_{(J-1)}^V$, $W_{(J-1)}^D$ and $W_{(J-1)}^H$ on the basis of the initialization obtained in the former step. The same process is then repeated up to the image resolution level.
- 7) We reconstruct the segmented image starting from the coarsest scale 2^J of the decomposition. The reconstruction uses:
 - for the initial approximation scale (level 2^J): the average of the original scaling coefficients within each region of the segmented approximation subband: $(\bar{V}_J(z_k))$ with $k \in 1, \dots, K$.
 - for all the detail subbands: the original wavelet coefficients, $W_{j \in \{1, \dots, J\}}^{(V,D,H)}$, for the segments that pertain to the class $k = 2$. The coefficients pertaining to the class $k = 1$ are cancelled.

- 8) We re-classify the histogram, in the direct domain, of the segmented image obtained in the former step. This is done by finding the thresholds Th_M^j , $j \in \{1 \dots (K + 1)\}$ of the modes of this histogram. This re-classification is necessary because there is no reason for the wavelet coefficients to pertain, after inversion, to a perfect classification in K classes. This step is easily realized by using the K means m_k computed at the coarsest approximation scale V_j and used now to find the histogram modes thresholds by simply taking the mid-point of the means for each couple of successive classes : $Th_M^j = \frac{m_{k+1} + m_k}{2}$. We also take, as value for the first and the last thresholds, Th_M^1 and Th_M^{K+1} , respectively the minimum and maximum values of the pixels.
- 9) The re-classification of the histogram is done in the same loop as the new labeling of the pixels in K successive classes, which provides at the same time the final segmented image in K classes with K starting from 1.

From a first point of view, this approach, based on the initialization of the segmentation of the coarse detail subbands from the discontinuities in the segmentation of the approximation subband, can be considered as a way to take into account the *intra-scale* dependency of the wavelet coefficients. From a second point of view, the initialization of the segmentation of detail subbands at a scale j , by the segmentation of the detail subbands at the immediate coarser scale $j + 1$, can also be considered as a way to take into account the *inter-scale* dependency of the wavelet coefficients.

Remark : the only parameters to choose before starting the computation are the maximum number of iterations, *itermax*, of the Gibbs sampling algorithm, and the number K of classes requested. *itermax* is generally taken large enough to assure the convergence of the segmentation, i.e. between 20 and 50. K must be taken at least equal to the number of classes in the image, if this value is known. The value of K will automatically decrease to fit the maximum number of classes in the image, but will not increase. Nevertheless, if the number of iterations is too low, the algorithm will not converge towards the correct number of classes and will give a value between the value asked and the real value (see Fig. 8). If the number of classes is unknown, then K must be chosen accordingly to the purpose of the application.

In a natural image, this number is generally taken between two and ten. Among other adjustable parameters are the α_{potts} parameter, i.e. α_V , α_W^V , α_W^D and α_W^H for the approximation and detail subbands. α_V , α_W^V , α_W^D and α_W^H are generally fixed to 1. Nevertheless, this parameter can be adjusted, especially for some natural images. The possibility to adjust separately α_V in the coarse approximation subband and α_W^V , α_W^D and α_W^H in the detail subbands enables to get smaller or larger homogeneous regions in the initial segmentation of the coarse approximation subband and to give more or less importance to the discontinuities in this subband and thus to the size of the regions in the wavelet subbands. These parameters, tested on some natural images, can be efficiently trimmed between 0.5 and 5. Another important initial parameter is the z_0 initial segmentation of the observable, if we own a liable knowledge of it. It is this same initialization parameter that we use, between subbands, in our algorithm, to improve the segmentation speed. The parameters m_k , v_k and v_ϵ can also be initially fixed if we have a

knowledge of them. In any other cases all these parameters can be left to their default initial values and will be automatically computed.

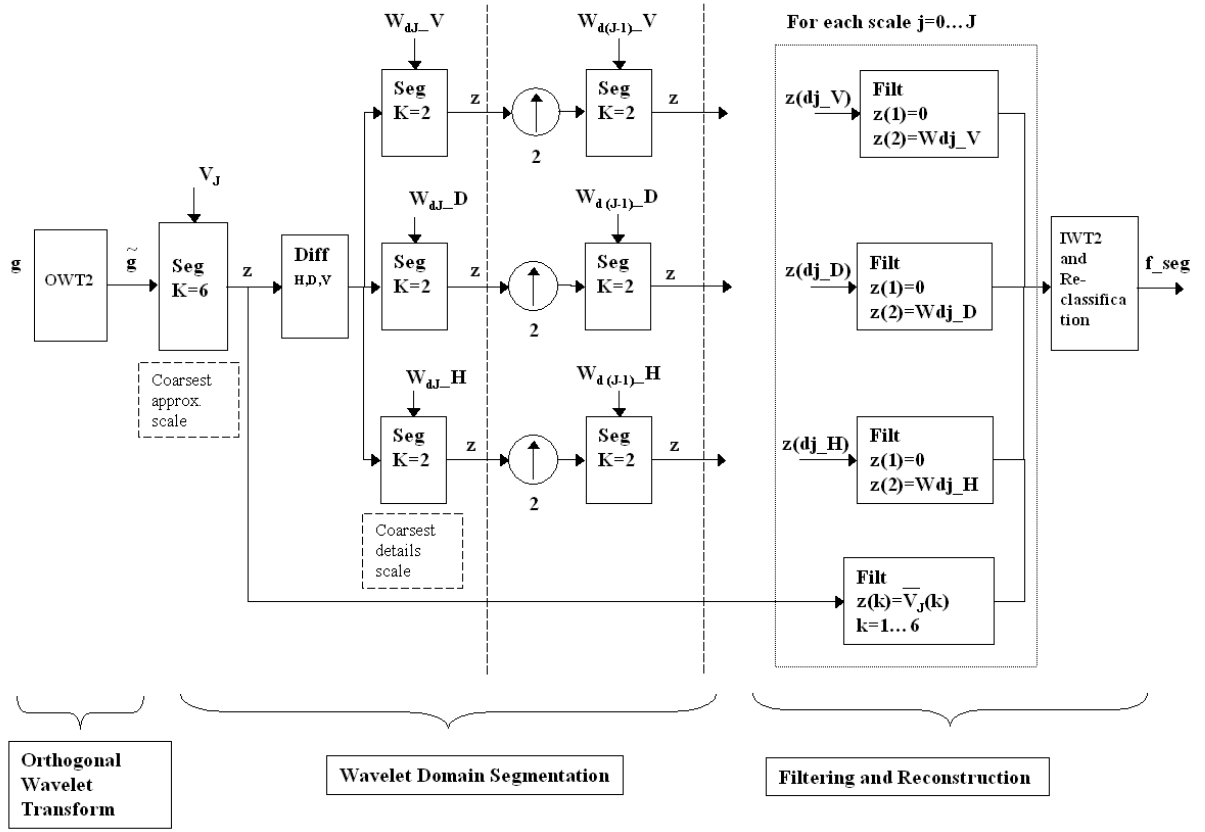


Figure 5: Wavelet domain Bayesian segmentation scheme. The observed data g is first decomposed in the wavelet domain (2 scales shown here). It is then segmented in this domain, with six labels for the approximation subband and two labels for all the detail subbands. The approximation subband is filtered by replacing the value $z(r) = k$ of each class k by the average of the initial scaling coefficients. All the wavelet subbands are filtered by zeroing the coefficients pertaining to the class $k = 1$ of “weak coefficients” and by leaving the coefficients of class $k = 2$ at their initial value. The final segmentation is obtained by reconstruction in the direct domain and histogram reclassification.

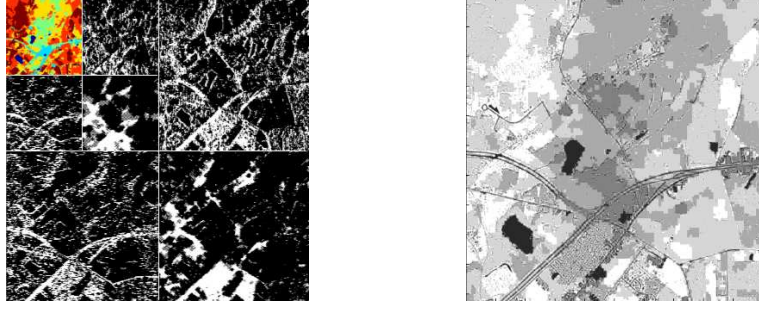


Figure 6: a) Mallat's pyramidal representation of the segmented image z . The approximation part is segmented in 6 classes (color part). The detail subbands are represented in black for the class of the weak coefficients, that are zeroed, and in white for the class of the strong coefficients that will be used for the final reconstruction. b) Reconstruction result of the wavelet segmentation shown in the direct domain.

6 Results and comparison

In this section we show comparative results on three examples of test images. The first test image is a synthetic mosaic image, "texmos3.s1024" that we upsampled from the texmos3.s512 image of the SIPI [22] database and that we have perturbed with additive Gaussian noises of different variances for the regions and for the whole image. The second test image is a natural image of the SIPI database representing the San Diego coast at Point Loma. The third image comes from one band of an hyperspectral AVIRIS satellite image of the earth and taken amongst the 224 contiguous channels of this "3D" image.

Four segmentation algorithms are compared : (1) the direct domain Bayesian Potts-Markov Segmentation (BPMS), (2) the Wavelet Bayesian Potts-Markov Segmentation (WBPMS) presented in this paper, (3) the HMT segmentation from [7] and (4) two versions of a clustering K-means algorithm with two different distance measures.

For the first example we have tested the algorithms 1, 2 and 4. For the second example we have tested the algorithms 1, 2 and 3. For the last example (our hyperspectral image), we have compared the methods 1 and 2.

To evaluate the segmentation quality, we use the numerical "Pa" criterion which is the *percentage* of pixels that are correctly classified, showing the *accuracy* of the segmentation [25, 24]. This test is used of course for synthetic images (our first test) with a known z classification.

We also indicate for each of the tests the computation time which is an important challenge for us. All computations have been done on a 1.6Mhz Pentium M, with 1MB of cache-RAM and 512MB System-RAM, which is not an optimal configuration for running image processing algorithms. This let us suppose that on a dedicated image processing machine, and a parallel architecture, we would reach much lower computation times.

6.1 First example

In this example we use the "texmos3.s1024" mosaic of the SIPI (Signal and Image Processing Institute) database of the USC (University of Southern California). This mosaic is a 512×512 , 8 bits, image, upsampled to 1024×1024 composed of 23 regions and 8 classes of homogeneous (not textured) regions. We have upsampled this image in order to demonstrate the WPMS on four scales of decomposition, which enables to start segmenting a not too small approximation subband of size 64×64 pixels. In order to model a more complex image, we have super-imposed additive Gaussian noises with different variances for each class to this texmos3s mosaic. Furthermore, a global Gaussian noise with another variance has been added to the previous image noised "by class". The figure 7 shows the original, the noisy 'observed' image g and the histogram of this observed image.

6.1.1 Test image : texmos3.s1024 SIPI Database image with additive Gaussian noises of different variances

The original histogram of this synthetic image is not represented here for it is composed of the eight bars ranging from one to eight, thus characterizing the eight classes of the original image. We show the image modified by various Gaussian noises and its histogram where the modes between the last third classes becomes difficult to detect.

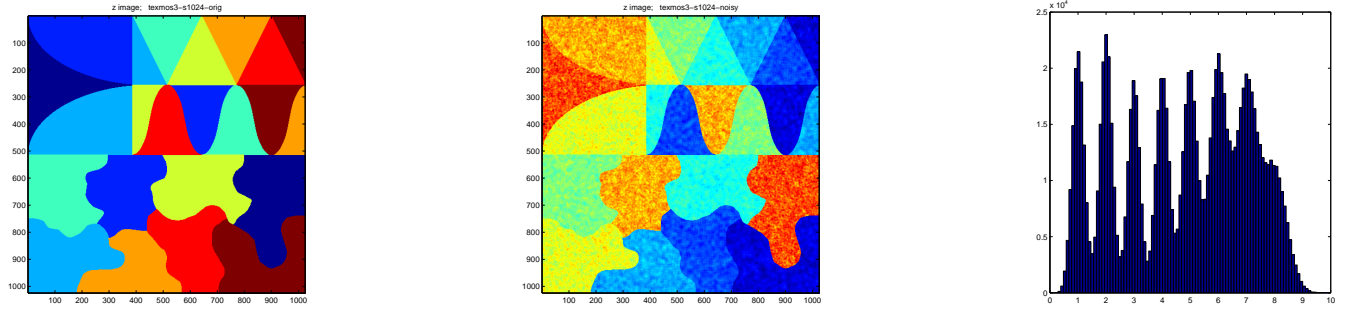


Figure 7: a) Original f image of the synthetic texmos3s mosaic of the "SIPI" database, upsampled from 512×512 and showing the $K = 8$ classes split in 23 regions. b) texmos3s mosaic perturbed by additive Gaussian noises with a different variance for each class and by a global gaussian noise of another variance value. c) Histogram of the noisy mosaic.

6.1.2 Result with the BPMS method : Bayesian Potts-Markov segmentation in the direct domain

The BPMS method has been tested with 20 iterations. The number of classes requested and obtained is 8. The computation time is 1805s and the percentage of accuracy is $Pa = 80.34\%$.

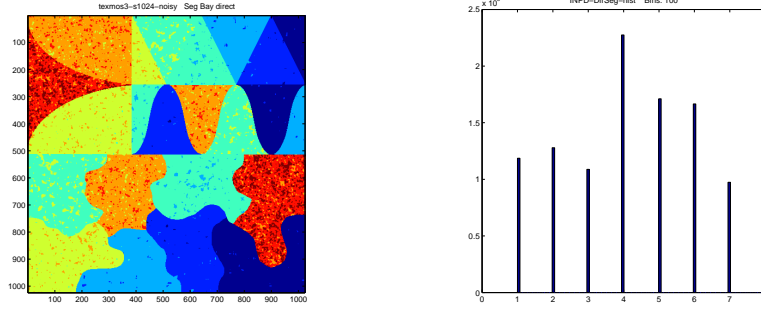


Figure 8: a) Result of BPMS method with parameters $K = 8$ and $itermax = 20$. The percentage of accuracy is $Pa = 80.34\%$. The computation time is 1805s. b) Final histogram with the $K = 8$ requested classes.

6.1.3 Result with the WBPMS method : wavelet Bayesian Potts-Markov segmentation

The WBPMS method has been tested with two scales of decomposition : $J = 4$ and $J = 3$. For both scales the number of iterations is 20 for the scaling and the wavelet subbands. With $J = 4$, we have obtained $Pa = 94.8\%$ in a total time $t = 260s$. With $J = 3$, we have obtained $Pa = 98.06\%$ in a total time $t = 384s$. These two results show a very good quantitative behaviour of our method for this test image.

- Results with a decomposition on 4 scales.

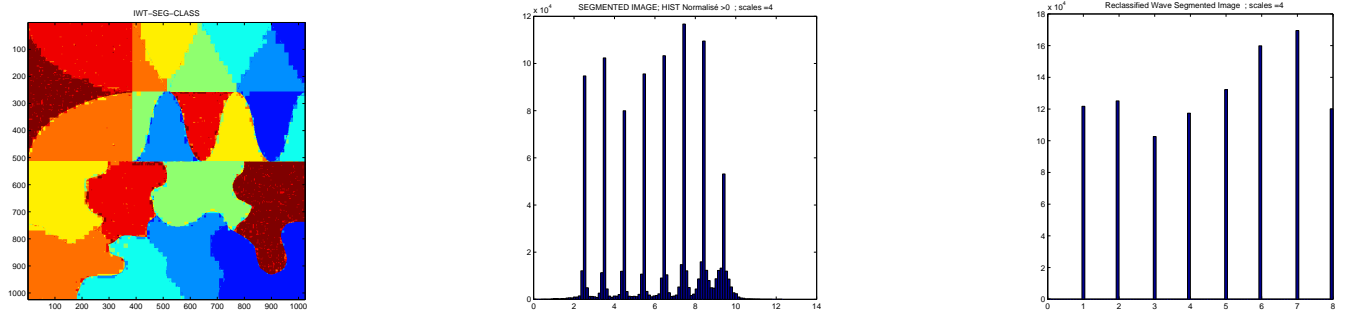


Figure 9: Segmentation of the texmos3.s1024 mosaic with added Gaussian noises. The segmentation parameters are : $J = 4$ scales, number of requested classes $K = 8$, iteration number $Itermax_{V_J} = 20$, $Itermax_W = 20$. The segmentation quality obtained is $Pa = 94.8\%$ in a time $t = 260s$ a) Final segmentation b) Non reclassified final histogram, after reconstruction, which shows a number of reconstructed pixels, coming from the wavelet subbands, and spread at the bottom of the main modes c) reclassified histogram in the direct domain and corresponding to the classification shown in a).

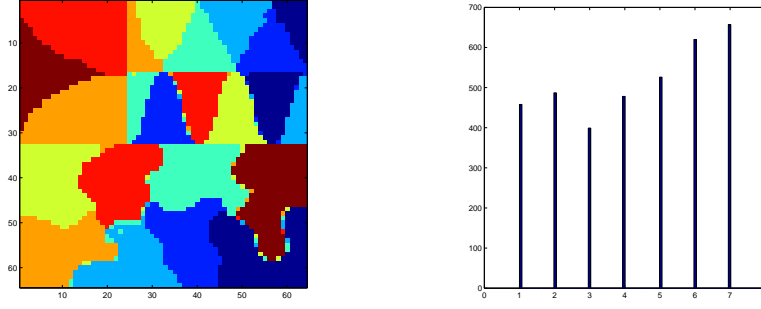


Figure 10: Following of Fig. 9. a) Segmentation of subband V_4 b) Histogram of subband V_4 showing that the 8 classes have been already found at the coarsest subband.

• Results with a decomposition on 3 scales.

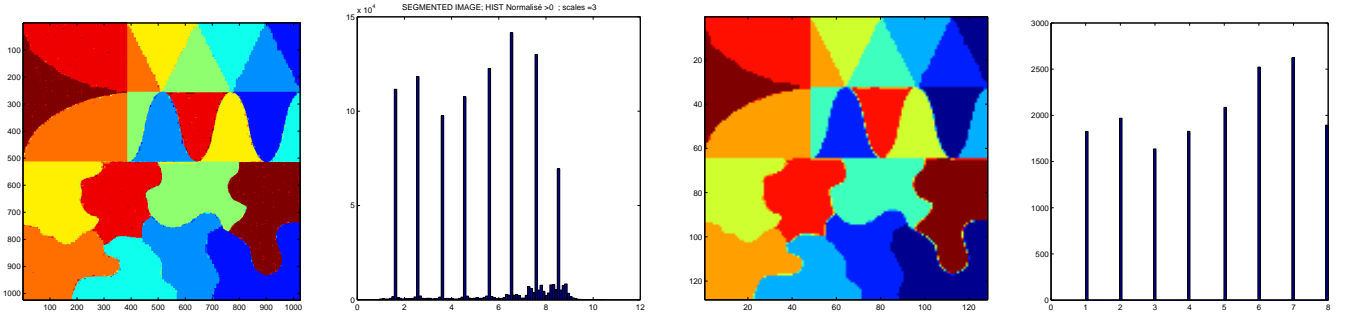


Figure 11: Result of the segmentation with $J = 3$ and $itermax = 20$ for both types of subbands. Here we reach an accuracy of 98.06% in a computation time of 384s. a) Final result of segmentation b) Histogram before final reclassification showing a pixel classification in 8 classes c) Segmentation of subband V_3 . d) Histogram of the classified subband V_3 .

6.1.4 Clustering (K-means) method

The clustering method algorithm we use is derived from the K-means method. The K-means clustering partitions the image into K clusters. Within each cluster C_k of pixels, we compute the means m_k and the variances v_k . Then, for every pixel D_{ij} , we compute either the L_1 or the L_2 distance, with :

$$L_1(k) = \frac{|(D_{ij} - m_k)|}{\sqrt{v_k}} \quad (25)$$

and

$$L_2(k) = \sqrt{\frac{\sum (D_{ij} - m_k)^2}{v_k}} \quad (26)$$

Then the pixel D_{ij} is classified in the class k which corresponds to the minimal distance.

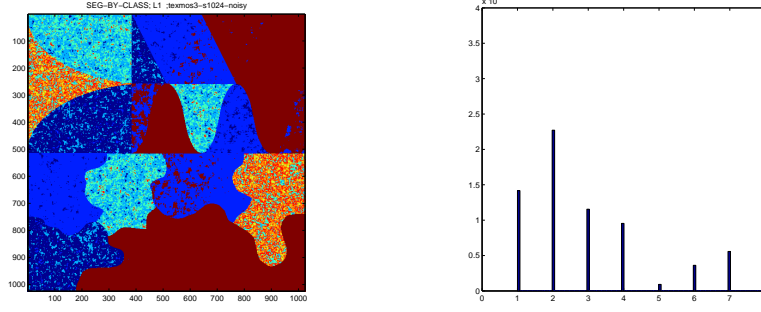


Figure 12: Classification results by the clustering method and distances L1 and L2. The only parameter is $K = 8$ classes requested (and obtained according to the histogram shown in part b) of this figure). For both distances the percentage of accuracy is $\simeq 50\%$ and the computation time is 116s a) Result of segmentation with distance L1 or L2. b) Final segmentation histogram in 6 classes.

6.1.5 Comparative results

In order to test the quality of the segmentation we count the number $N_m(\mathbf{r})$ of pixels mis-classified in the final segmented image w.r.t. the image of the z original image.

We can compare in the table below the results of the classification with the three methods and two different levels of decomposition for our WBPMS method. We obtain a very accurate classification compared to the K-means method and a much shorter computation time than with the BPMS method.

Method	Classes requested/obtained	Pixels mis-classified	Percentage of accuracy (P_a)	Total classif. time
BMPS (20 iter.)	8	206114	80.34%	1805s
WBMPS(J=4, 20 iter.)	8	55488	94.8%	260s
WBMPS(J=3, 20 iter.)	8	20336	98.06%	384s
K-Means (L_1 and L_2)	8	530196	49.44%	116s

Table 1: Comparison of three segmentation methods on the mosaic test image “texmos3” from the SIPI database. The original image, quantified on 8 bits, has been upsampled from 512×512 to 1024×1024 and has been perturbed by a Gaussian noise of different variance for each class, and a global additive Gaussian noise. The number of requested classes for the segmentation is $K = 8$ for all methods, which is the number of classes of the test image. The quality of segmentation is based on the number of mis-classified pixels, which gives the percentage of accuracy P_a as used in [25, 24].

6.2 Second example

In this second example three methods on a natural satellite image (512×512 and 8 bits) of the San Diego coast. We compare the segmentation results obtained with the following methods :

- The Bayesian Potts-Markov HM segmentation in the direct domain (BPMS).
- The Bayesian Potts-Markov segmentation in the wavelet model (WBPMS).
- The Bayesian semi-supervised multiscale segmentation based on a HMT model [7].

For the first two methods we have imposed a number of classes $K = 6$. The result with the HMT model is originated from [7] and is a binary segmentation, i.e. with only two classes, whose purpose is to make the distinction between the maritime and the terrestrial zones. So, first because we compare the results on natural images, second because it is with a different value of classes, and finally because the HMT method uses a semi-supervised algorithm based on textures discrimination, we will comment the methods on a qualitative point of view.

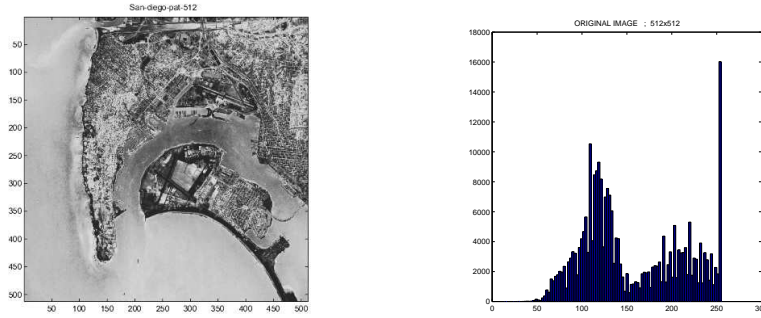


Figure 13: a) San-Diego coast test image 512×512 , 8 bits per pixel. b) Histogram of this test image leading to an interpretation of the image as a mixture of two main distributions, which are materialized by the very distinctive earth and the sea regions.

6.2.1 BPMS, Bayesian Potts-Markov segmentation in the direct domain

The BPMS method has been tested with $K = 6$ classes requested and 20 iterations. The number of classes obtained is 6. The computation time with these parameters is 114s.

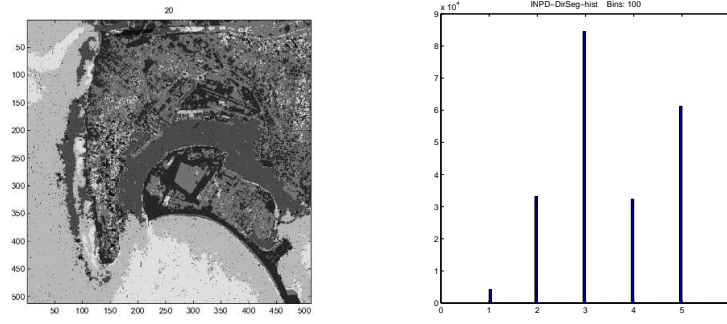


Figure 14: a) Bayesian segmentation in the direct domain with $K = 6$ classes and a number of iterations $itermax = 20$. Computation time = 114s. b) Histogram of the final segmentation.

6.2.2 WBPMS, wavelet Bayesian Potts-Markov segmentation

On the San Diego image, the WBPMS has been tested with $K = 6$ classes requested and 20 iterations as for the BPMS method. The number of classes obtained is 6. The number of scales for the decomposition is 2. The computation time with these parameters is 79s which is slightly faster (114s) than the BPMS method for the same number of iterations.

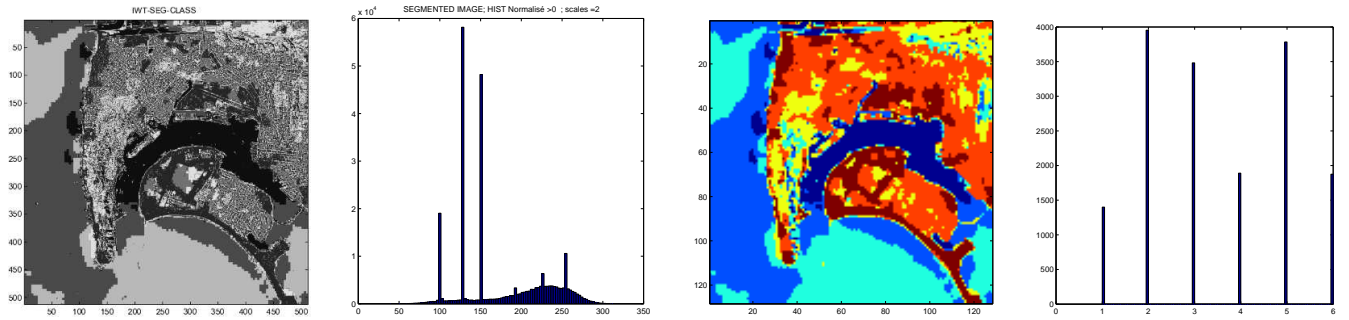


Figure 15: WBPMS segmentation of the San-Diego coast. Requested, and obtained, number of classes is $K = 6$. Number of iterations is 20 and number of scales is $J = 2$. The segmentation time is 79s. a) Final segmentation b) Raw histogram, before reclassification, exhibiting 6 classes and a number of spread, misclassified, pixels mainly due to the reconstruction of the wavelet subbands. c) Segmentation of the coarse approximation subband V_J d) Histogram of the coarse approximation subband showing the detection of the 6 classes at this level of decomposition.

6.2.3 HMT method

The goal of the segmentation for the authors was to segment the terrestrial zone from the maritime zone. The segmentation, based on a HMT, is thus realized with 2 classes. Using a HMT method, this means that the segmentation can take into account the textures. The two textures are in this case the maritime and the terrestrial texture. The HMT unsupervised segmentation has nevertheless to be initialized by a supervised learning of these two textures. The authors take two sub-images in the corners of the original whole image (1024×1024) that best represent the maritime and the terrestrial texture respectively and find the

model that best fit with these two textures. We can nevertheless notice that parts of the maritime zone (submarine harbor, top right of the sub-images in [7]) are reduced between step b) and inter-scale fusion, step and sub-image c). So if the result is good for the mountain chain which is classify in their result as a terrestrial zone, on the counterpart it is likely to reduce some maritime zones (harbors in the bay). We tend to encounter the same problem, with our method, when segmenting this image in two classes.



Figure 16: a) San Diego : original 1024x1024 image used for the learning phase of the two textures (ocean and earth), materialized by two 100×100 square sub-images. b) Result of a raw binary segmentation at the pixel level on a sub-image of size 256×256 . c) Binary segmentation result by inter-scale fusion.

6.2.4 Results

Segmentation is done with $K = 6$ classes.

- With the BPMS : Itermax = 20; Computation time is 114s.
- With the WBPMS : Itermax(Approx.) = 20; Itermax(Details) = 20; Total computation time is 79s with 2 scales.
- With the HMT method we do not know the computation time. The result, with this method, shows that the authors are able, after a learning phase of the maritime and terrestrial textures, to realize a classification in two classes : one representing only the terrestrial regions and a second representing only the maritime zones. In particular they have shown good results in segmenting the north-south mountains chain, on the left of the terrestrial zone, in an almost homogeneous region of class "earth". Nevertheless, in their paper the progression in the intercale-fusion based segmentation shows that the growth of the "earth" class in this mountain chain is done to the detriment of some maritime zones inside the bay.

6.3 Third example

The test image is a satellite 1-spectral band image taken from a "hyperspectral image" composed of 224 images taken in 224 contiguous spectral bands and already shown in Fig. 1. The interest here is again to classify, in a fully unsupervised way, this natural image in homogeneous regions accordingly, as much as possible, to the different objects or regions that compose the image.

We take the same number of iterations for the BPMS and WBPMS methods, i.e. 20. In the WBPMS method, we have one parameter for the maximum number of iterations in the scaling coarse subband and another parameter for all the wavelet

subbands. We fix both parameters to the same value, i.e. 20. We take 2 scales of decomposition in the WBPMS method, which, according to numerous trials, is a good value for a natural image which presents many details and also for an image of size $512 \times$. Working on two scales means that the coarse V_J subband is segmented on a sub-image of size 128×128 which stays large enough and does not affect small regions by a too large subsampling.

The computation time with the WBPMS method is 73s to compare with 110s with the BPMS method. This makes a slightly faster algorithm for the same number of iterations. But the most significant point of the test is the quality of the segmentation obtained with the WBPMS in comparison with the BMPS method, for only 20 iterations.

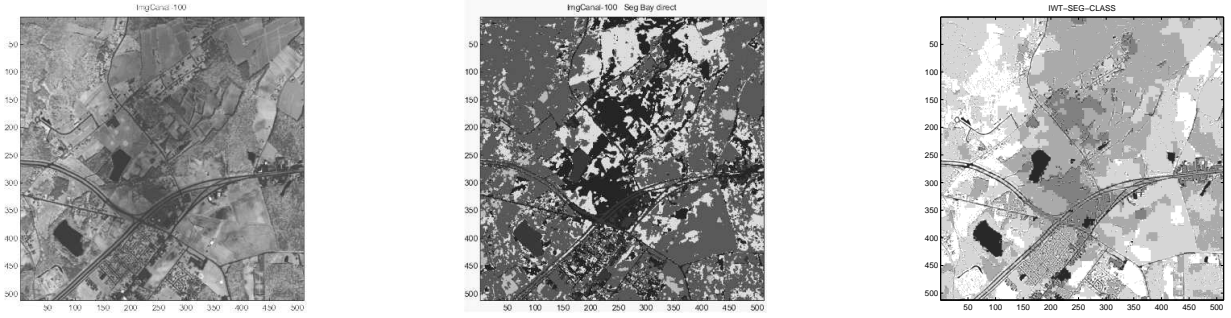


Figure 17: Comparison of the segmentation results with BPMS and WBPMS methods on a hyperspectral satellite image of size 512×512 . The segmentation parameters are $K = 6$ requested classes and 20 iterations (20 for the subbands V and W in the WBPMS method). Both methods lead to a final classification in the 6 requested classes (histograms not shown here). a) Original mono-band of a hyperspectral satellite image. b) Bayesian segmentation in the direct domain with the BPMS method; the segmentation time is 110s. c) Segmentation result with the WBPMS method; the number of scales is 2 and the segmentation time is 73s. We notice that the difference in computation time is slightly better with the WBPMS, but above all that the result with the BPMS method would need more iterations to reach a correct, more homogeneous, segmentation.

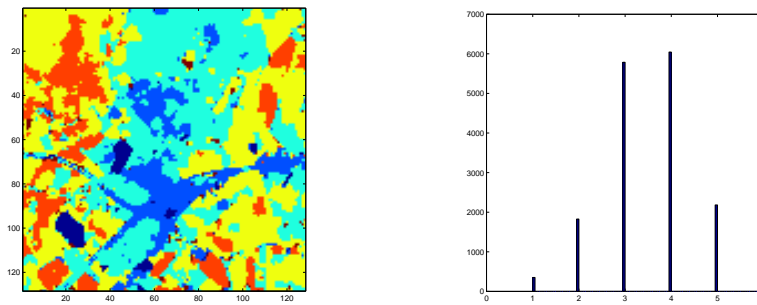


Figure 18: Segmentation detail of the coarse approximation subband V_J at level 2 and its histogram showing that the number of classes found at V_J already corresponds to the number $K = 6$ of requested classes and that subsequent detail subbands segments lead to a more detailed segmented image.

7 Conclusion

We have presented a new algorithm for the segmentation of images based on a Bayesian segmentation performed in the wavelet domain. The first originality of this work resides in the fact that we do the image segmentation in the wavelet domain, but rather than using the HMT property of wavelet coefficients, we concentrate on the initialization of a PMRF approach in a coarse to fine scale scheme. This wavelet-Bayesian segmentation is mainly based on the hypotheses of Gaussian distributions for the image, the noise and the regions segmented. Due to the fact that the segmentations are done with only two classes for all the Band-Pass subbands, we obtain a significant reduction of the complexity. The second originality of this work is that a second order, 8-connexity, Potts-Markov Random Field has been developed to fit to the three main orientation of the detail subbands of the wavelet decomposition.

This WBPMS scheme has led, depending on the test, to a reduction of the computation time by a factor of ten or even more, for the same classification quality. We have tested the WBPMS with different levels J of decomposition. The best results have been obtained for J between 2 and 4. The number J could exceed the value of 4 but for large images, i.e. above 1024×1024 . The reason is that small regions tend to disappear when the size of the coarse approximation subband is very small, like 64×64 or less. In this case, if the image contains many details, the regions become small and the number of classes in the V_J subband may become inferior to the number of requested classes.

We have shown that the quality of segmentation on a noisy synthetic test image (texmos3 mosaic) can exhibit a good accuracy of classification in a much shorter time than the other methods tested, especially in comparison to the same method performed in the direct domain (BPMS).

The main goal of our WBPMS scheme was to provide a new fully unsupervised algorithm for fast segmentation of still images. We think this goal has been reached. A second main application is to lower the segmentation speed of video sequences, as well as the motion estimation and the off-line video compression, a work that we have initiated in [4] and [3].

Acknowledgments

The authors are very grateful to both anonymous reviewers for their attentive and constructive remarks that helped improve the quality of the presentation.

References

- [1] H-H. Bock. Clustering methods - a review of classical and recent approaches. In *Proceedings of Modelling, Computation and Optimization in Information Systems and Management Sciences*, Metz, France, July 2004.
- [2] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, March 1994.
- [3] P. Brault. On the performances and improvements of motion-tuned wavelets for motion estimation. *WSEAS Transactions on Electronics*, 1(1):174–180, 2004.
- [4] P. Brault and A. Mohammad-Djafari. Bayesian segmentation of video sequences using a Markov-Potts model. *WSEAS Transactions on Mathematics*, 3(1):276–282, January 2004.
- [5] P. Brault and A. Mohammad-Djafari. Bayesian wavelet domain segmentation. In *Proceedings of the AIP, American Institute of Physics, for the International Workshop, MaxEnt, on Bayesian Inference and Maximum Entropy Methods*, pages 19–26, MaxPlanck Institute für Statistics, Garching, Germany, July 2004.
- [6] H. Cheng and C.A. Bouman. Multiscale Bayesian segmentation using a trainable context model. *IEEE Transactions on Image Processing*, 10(4):51–525, 2001.
- [7] H. Choi and R.G. Baraniuk. Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10(9):1309–1321, September 2001.
- [8] M.S. Crouse and R.G. Baraniuk. Contextual hidden Markov models for wavelet-domain signal processing. In *Proc. of the 31st Asilomar Conf. on Signals, Systems, and Computers*, volume 1, pages 95–100, Pacific Grove, CA., November 1997.
- [9] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, April 1998.
- [10] G. Fan and X.G. Xia. A joint multicontext and multiscale approach to Bayesian image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 39(12):2680–2688, December 2001.
- [11] G. Fan and X.G. Xia. Wavelet-based texture analysis and synthesis using hidden Markov models. *IEEE Transactions on Geoscience and Remote Sensing*, 40(1):229–229, January 2002.
- [12] O. Féron and A. Mohammad-Djafari. Image fusion and unsupervised joint segmentation using a HMM and MCMC algorithms. *Journal of Electronic Imaging*, 14(2), june 2005.
- [13] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of image. *IEEE PAMI*, 6(6):721–741, November 1984.

- [14] S. Geman, D. Geman, and C. Graffigne. Locating texture and object boundaries. In Ed. P.A. Devijver and J.Kittler, editors, *Pattern Recognition Theory and Application*, Heidelberg, 1987. Springer-Verlag.
- [15] G.G. Hazel. Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1199 – 1211, May 2000.
- [16] J. (Ed.) Idier. *Approche bayésienne pour les problemes inverses*. Hermes Science, 2001.
- [17] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, New-Jersey, 1988.
- [18] J. Pesquet, H. Krim, H. Leporini, and E. Hamman. Bayesian approach to the best basis selection. In *ICASSP*, Atlanta, May 1996.
- [19] J. Portilla, V. Strela, W. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. on Image Processing*, 12(11), 2003.
- [20] J.K. Romberg, H. Choi, and R.G. Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10(7):1056 – 1068, July 2001.
- [21] T. Simchony, R. Chellapa, and Lichtenstein Z. The graduated non-convexity algorithm for image estimation using compound Gauss-Markov field models. In *Proc. ICASSP89*, Glasgow, May 1989.
- [22] SIPI. Images and videos database. <http://sipi.usc.edu/publications.html>.
- [23] H. Snoussi and A. Mohammad-Djafari. Fast joint separation and segmentation of mixed images. *Journal of Electronic Imaging*, 13(2):349–361, April 2002.
- [24] X. Song and G. Fan. Unsupervised Bayesian image segmentation using wavelet-domain hidden Markov models. In *Proc. of ICIP, International Conference on Image Processin*, volume 2, pages 423–426, September 2003.
- [25] X. Song and G. Fan. Unsupervised image segmentation using wavelet-domain hidden Markov models. In *Proceedings of SPIE Wavelets X in applications in signal and image processing*, San Diego, 2003.
- [26] F.Y. Wu. The Potts model. *Review of Modern Physics*, 54(1):235–268, January 1982.
- [27] J. Zerubia and R. Chellapa. Mean field annealing using compound Gauss-Markov random fields for edge detection and image estimation. *IEEE Trans. on Neural Networks*, 4:703–709, 1993.



Patrice Brault graduated from the "Conservatoire National des Arts et Métiers (CNAM)" in Electrical Engineering. Before joining the "Centre National de la Recherche Scientifique (CNRS)" in 1998, he has been working in the telecommunications area, mainly for "Matra Communications", "Apple Europe Research" and the "Laboratoire d'Electronique Philips (LEP)", where he has participated to the development of the first MPEG2 digital television broadcast system. His main research interests are signal and image processing and in particular : fractals, wavelets and Bayesian methods applied to shape recognition, motion estimation, segmentation and video compression. He is presently finishing a Ph.D. on motion estimation and image segmentation at the "Laboratoire d'Electronique Fondamentale (IEF)", University of Orsay Paris-sud, in collaboration with the "Laboratoire des Signaux et Systemes (L2S)" at Supélec.



Ali Mohammad-Djafari received his BSc degree in electrical engineering from Polytechnique of Teheran, in 1975, his MSc (diploma degree) from Ecole Supérieure d'Electricité (Supélec), Gif sur Yvette, France, in 1977 and his "Docteur-Ingénieur" (PhD) degree and "Doctorat d'Etat" in Physics, from the Université Paris-Sud (UPS), Orsay, France, respectively in 1981 and 1987. He was associate professor at UPS for two years (1981 to 1983). Since 1984, he has a permanent position at "Centre National de la Recherche Scientifique (CNRS)" and works at "Laboratoire des Signaux et Systèmes (L2S)" at Supélec. From 1998 to 2002, he has been at the head of Signal and Image Processing division at this laboratory. Presently, he is "Directeur de recherche" and his main scientific interests are in developing new probabilistic methods based on information theory, maximum entropy and Bayesian inference approaches for inverse problems in general, and more specifically: image reconstruction, signal and image deconvolution, blind source separation, data fusion, multi and hyperspectral image segmentation. The main application domains of his interests are computed tomography (X rays, PET, SPECT, MRI, microwave, ultrasound, and eddy current imaging) either for medical imaging or for nondestructive testing (NDT) in industry.

List of Figures

1	a) Original image of the Channel 100 of an hyperspectral satellite image composed of 224 channels. b) Pyramidal representation of the Fast Orthogonal Wavelet 2-levels decomposition applied to the the channel 100 of our hyperspectral image. Here the letter n can be replaced by J , the scale parameter at the coarsest scale 2^J bands.	11
2	a) Histogram of the A_J (or A_n), the approximation coefficients at the coarsest scale. b) Histogram of A_J described by a mixture of multiple Gaussians, one for each class. c) Histogram of the D_J^D (or D_n^D), the detail diagonal coefficients at the coarsest scale. d) Linlog histogram of the D_J^D explaining the choice of a two independent Gaussians mixture model, with one large and one small variance.	11
3	First, second and fourth order Markov fields (after [16]). The first order neighboring is used in the parallel implementation of the Gibbs sampler for the VH dependency (scaling subband). The second order neighboring is used in the parallel implementation of the $D_1 D_2$ dependency (diagonal wavelet subband)	13
4	Parallel implementation of the Gibbs sampling algorithm. By considering the image as divided in two sets of independent white and black sites, one iteration of the Gibbs sampler for the whole image can done in only two times : one time on the black sites and the second on the white sites. The two sets of sites are build differently depending on the directions of the neighbor sites considered : a) set of independent sites for the vertical and horizontal directions (first order neighboring) b) set of independent sites for the diagonal directions (second order neighboring)	13
5	Wavelet domain Bayesian segmentation scheme. The observed data g is first decomposed in the wavelet domain (2 scales shown here). It is then segmented in this domain, with six labels for the approximation subband and two labels for all the detail subbands. The approximation subband is filtered by replacing the value $z(r) = k$ of each class k by the average of the initial scaling coefficients. All the wavelet subbands are filtered by zeroing the coefficients pertaining to the class $k = 1$ of “weak coefficients” and by leaving the coefficients of class $k = 2$ at their initial value. The final segmentation is obtained by reconstruction in the direct domain and histogram reclassification.	16
6	a) Mallat’s pyramidal representation of the segmented image z . The approximation part is segmented in 6 classes (color part). The detail subbands are represented in black for the class of the weak coefficients, that are zeroed, and in white for the class of the strong coefficients that will be used for the final reconstruction. b) Reconstruction result of the wavelet segmentation shown in the direct domain.	17
7	a) Original f image of the synthetic texmos3s mosaic of the ”SIPI” database, upsampled from 512×512 and showing the $K = 8$ classes split in 23 regions. b) texmos3s mosaic perturbed by additive Gaussian noises with a different variance for each class and by a global gaussian noise of another variance value. c) Histogram of the noisy mosaic.	18
8	a) Result of BPMS method with parameters $K = 8$ and $itermax = 20$. The percentage of accuracy is $Pa = 80.34\%$. The computation time is 1805s. b) Final histogram with the $K = 8$ requested classes.	19

9	Segmentation of the texmos3.s1024 mosaic with added Gaussian noises. The segmentation parameters are : $J = 4$ scales, number of requested classes $K = 8$, iteration number $Itermax_{V_J} = 20$, $Itermax_W = 20$. The segmentation quality obtained is $P_a = 94.8\%$ in a time $t = 260s$ a) Final segmentation b) Non reclassified final histogram, after reconstruction, which shows a number of reconstructed pixels, coming from the wavelet subbands, and spread at the bottom of the main modes c) reclassified histogram in the direct domain and corresponding to the classification shown in a).	19
10	Following of Fig. 9. a) Segmentation of subband V_4 b) Histogram of subband V_4 showing that the 8 classes have been already found at the coarsest subband.	20
11	Result of the segmentation with $J = 3$ and $itermax = 20$ for both types of subbands. Here we reach an accuracy of 98.06% in a computation time of 384s. a) Final result of segmentation b) Histogram before final reclassification showing a pixel classification in 8 classes c) Segmentation of subband V_3 . d) Histogram of the classified subband V_3	20
12	Classification results by the clustering method and distances L1 and L2. The only parameter is $K = 8$ classes requested (and obtained according to the histogram shown in part b) of this figure). For both distances the percentage of accuracy is $\simeq 50\%$ and the computation time is 116s a) Result of segmentation with distance L1 or L2. b) Final segmentation histogram in 6 classes.	21
13	a) San-Diego coast test image 512×512 , 8 bits per pixel. b) Histogram of this test image leading to an interpretation of the image as a mixture of two main distributions, which are materialized by the very distinctive earth and the sea regions.	22
14	a) Bayesian segmentation in the direct domain with $K = 6$ classes and a number of iterations $itermax = 20$. Computation time = 114s. b) Histogram of the final segmentation.	23
15	WPMS segmentation of the San-Diego coast. Requested, and obtained, number of classes is $K = 6$. Number of iterations is 20 and number of scales is $J = 2$. The segmentation time is 79s. a) Final segmentation b) Raw histogram, before reclassification, exhibiting 6 classes and a number of spread, misclassified, pixels mainly due to the reconstruction of the wavelet subbands. c) Segmentation of the coarse approximation subband V_J d) Histogram of the coarse approximation subband showing the detection of the 6 classes at this level of decomposition.	23
16	a) San Diego : original 1024×1024 image used for the learning phase of the two textures (ocean and earth), materialized by two 100×100 square sub-images. b) Result of a raw binary segmentation at the pixel level on a sub-image of size 256×256 . c) Binary segmentation result by inter-scale fusion.	24
17	Comparison of the segmentation results with BPMS and WPMS methods on a hyperspectral satellite image of size 512×512 . The segmentation parameters are $K = 6$ requested classes and 20 iterations (20 for the subbands V and W in the WPMS method). Both methods lead to a final classification in the 6 requested classes (histograms not shown here). a) Original mono-band of a hyperspectral satellite image. b) Bayesian segmentation in the direct domain with the BPMS method; the segmentation time is 110s. c) Segmentation result with the WPMS method; the number of scales is 2 and the segmentation time is 73s. We notice that the difference in computation time is slightly better with the WBMPMS, but above all that the result with the BPMS method would need more iterations to reach a correct, more homogeneous, segmentation.	25

18	Segmentation detail of the coarse approximation subband V_J at level 2 and its histogram showing that the number of classes found at V_J already corresponds to the number $K = 6$ of requested classes and that subsequent detail subbands segments lead to a more detailed segmented image.	25
----	---	----