AN INTERSCALE MULTIVARIATE STATISTICAL MODEL FOR MAP MULTICOMPONENT IMAGE DENOISING IN THE WAVELET TRANSFORM DOMAIN

Amor Elmzoughi [(1)], *Amel Benazza-Benyahia* [(1)], *Jean-Christophe Pesquet* [(2)]

[(1)] URISA, Ecole Supérieure des Communications de Tunis,
Route de Raoued 3.5 Km, 2083 Ariana, Tunisia.
[(2)] IGM and UMR-CNRS 8049, Université de Marne-la-Vallée,
5, Bd Descartes, Champs sur Marne, 77454 Marne la Vallée Cédex 2, France.
e-mail: ben.yahia@planet.tn,pesquet@univ-mlv.fr

ABSTRACT

The objective of this paper is to design a multivariate statistical approach for multicomponent image denoising in the wavelet transform domain. To this respect, we extend an appealing approach that we have recently proposed, where the wavelet coefficients of all the image channels at the same spatial position, in a given orientation and at the same resolution level, are grouped into a vector and a multivariate Bernoulli-Gaussian distribution is used as a prior model. The contribution of the paper is to develop low-complexity maximum a posteriori rules that exploit jointly the intra- and inter-scale redundancies between the wavelet coefficients. Experimental results carried out on remote sensing multispectral images show that the proposed procedure improves the state-of-the-art wavelet-based denoising methods.

1. INTRODUCTION

Most of the images suffer from noise due to the imperfections of imaging systems and transmission channels. This noise is strongly undesirable since it renders more difficult image interpretation tasks. Hence, denoising is a required step before image analysis. A great number of denoising methods have been reported [1] that may be classified into non-Bayesian and Bayesian methods. In the first class, the unknown image is often considered as deterministic whereas in the second one, it is viewed as a realization of a random field with a given prior probability distribution. Obviously, a delicate operation is the selection of a realistic prior distribution. To this respect, the sparseness and the decorrelation properties of the Wavelet Transform (WT) [3] facilitate the derivation of statistical priors in the transform domain [2]. For instance, the wavelet coefficients of the “clean” image have been considered as independent variables distributed according to a Generalized Gaussian Distribution (GGD) [4, 5, 6] or Gaussian mixtures [7, 8, 9]. More sophisticated priors such as hidden Markov models [10] or Markov random field models in the WT domain [11, 12] have been used in order to take into account the inter-scale dependencies related to the fact that large-magnitude coefficients tend to occur at the same positions in subbands at adjacent scales [13]. Recently, the scale dependencies were also captured by a non-Gaussian bivariate joint distribution for each coefficient and its parent and, the corresponding Maximum A Posteriori (MAP) estimator was obtained [14]. In parallel to these works, a great effort was performed for denoising multichannel images acquired by sensors operating in different spectral ranges. In this context, two alternatives may be envisaged: denoise separately each component or denoise simultaneously all the spectral components following a multivariate approach that exploits the spectral similarities. In this respect, we have proposed an efficient MAP estimation of the wavelet coefficients based on a *multivariate* Bernoulli-Gaussian (BG) model [15, 16]. A similar approach has also been envisaged in the WT domain [17]. Recently, we have proposed to exploit the inter-scale redundancies in order to improve the performances of our multivariate BG estimation procedure [18]. The objective of this paper is to design more sophisticated methods that capture jointly the intra and inter-scale dependencies in the different spectral channels. The paper is organized as follows. In Section 2, we briefly present the theoretical background. In Section 3, we describe two new multivariate estimators which take into account the inter-scale dependencies. Interestingly, the second approach does not introduce any markovianity assumption on the hidden variables, unlike most existing interscale models. Then, we address the problem of the hyperparameters estimations in Section 4. Finally, some simulation results are provided and, some conclusions are drawn in Section 5.

2. MAP FOR MULTIVARIATE BG PRIORS

2.1. The observation model

The unknown multichannel image consists of $B \in \mathbb{N}^*$ spectral components $s^{(b)}$ of size $L \times L$ with $b \in \{1, \dots, B\}$. At every spatial position $\mathbf{m} \in \{1, \dots, L\}^2$, each spectral component $s^{(b)}(\mathbf{m})$ is corrupted by an additive noise $n^{(b)}(\mathbf{m})$. The noisy observation vector $\mathbf{r}(\mathbf{m})$=$\mathbf{s}(\mathbf{m}) + \mathbf{n}(\mathbf{m})$ is such as the noise vector $\mathbf{n}(\mathbf{m}) = (n^{(1)}(\mathbf{m}), \dots, n^{(B)}(\mathbf{m}))^T$ is iid $\mathcal{N}(\mathbf{0}, \mathbf{R}^{(n)})$, independent of $\mathbf{s}(\mathbf{m}) = (s^{(1)}(\mathbf{m}), \dots, s^{(B)}(\mathbf{m}))^T$. It is worth noting that a non diagonal matrix $\mathbf{R}^{(n)}$ indicates that inter-spectral correlations exist between noise samples. Then, a dyadic separable WT over J stages is applied separately to each component $r^{(b)}$ of $\mathbf{r}$. At each resolution j, 3 wavelet subbands $r_j^{(b,o)}$ of size $L/2^j \times L/2^j$ oriented horizontally ($o = 1$), vertically ($o = 2$) or diagonally ($o = 3$) are produced. An observation vector $\mathbf{r}_j^{(o)}(\mathbf{k})$ in the WT domain is defined as $\mathbf{r}_j^{(o)}(\mathbf{k}) \stackrel{\triangle}{=} (r_j^{(1,o)}(\mathbf{k}), \dots, r_j^{(B,o)}(\mathbf{k}))^T$. As the WT is linear, $\mathbf{r}_j^{(o)}(\mathbf{k}) = \mathbf{s}_j^{(o)}(\mathbf{k}) + \mathbf{n}_j^{(o)}(\mathbf{k})$ where $\mathbf{s}_j^{(o)}(\mathbf{k})$ and $\mathbf{n}_j^{(o)}(\mathbf{k})$ are defined similarly to $\mathbf{r}_j^{(o)}(\mathbf{k})$. Besides, we can easily check that $\mathbf{n}_j^{(o)}(\mathbf{k}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_j^{(n,o)})$ where $\mathbf{\Gamma}_j^{(n,o)} = \mathbf{R}^{(n)}$.

2.2. Prior model

In the conventional Bayesian framework, an appropriate prior distribution for the coefficients $s_j^{(b,o)}$ is usually chosen for every channel b. We have generalized this approach to the vector $\mathbf{s}_j^{(o)}$ by defining *multivariate* Bernoulli-Gaussian priors $p_j^{(o)}$ able to reflect the parsimony of the wavelet representation. We have then

$$\forall \mathbf{u} \in \mathbb{R}^B, \quad p_j^{(o)}(\mathbf{u}) = (1-\epsilon_j^{(o)})\delta(\mathbf{u}) + \epsilon_j^{(o)} g_{\mathbf{0},\mathbf{\Gamma}_j^{(s,o)}}(\mathbf{u}), \quad (1)$$

where δ is the Dirac distribution, $g_{\mathbf{0},\mathbf{\Gamma}_j^{(s,o)}}$ denotes the probability density of a zero-mean multivariate Gaussian vector with covariance matrix $\mathbf{\Gamma}_j^{(s,o)}$ and, $\epsilon_j^{(o)} \in [0,1]$ is the mixture parameter. In order to avoid degenerate MAP estimates, the mixture model is coupled with hidden Bernoulli random variables $q_j^{(o)}(\mathbf{k})$ such that:

$$\begin{array}{rcl} p(\mathbf{s}_j^{(o)}(\mathbf{k})/q_j^{(o)}(\mathbf{k})=0) & = & \delta(\mathbf{s}_j^{(o)}(\mathbf{k})) \\ p(\mathbf{s}_j^{(o)}(\mathbf{k})/q_j^{(o)}(\mathbf{k})=1) & = & g_{\mathbf{0},\mathbf{\Gamma}_j^{(s,o)}}(\mathbf{s}_j^{(o)}(\mathbf{k})) \end{array}, \quad (2)$$

with $\mathsf{P}(q_j^{(o)}(\mathbf{k}) = 1) = \epsilon_j^{(o)}$. Conditionally to the hidden variables, the coefficients $\mathbf{s}_j^{(o)}(\mathbf{k})$ will be assumed independent. As a consequence, a generic denoising procedure consists in estimating $q_j^{(o)}(\mathbf{k})$ and computing the MAP estimator $\hat{\mathbf{s}}_j^{(o)}(\mathbf{k})$ of $\mathbf{s}_j^{(o)}(\mathbf{k})$. Finally, the spectral bands in the spatial domain are obtained by applying the inverse WT to the components $\hat{s}_j^{(b,o)}$.

2.3. MAP estimation of the wavelet coefficients

It is worth pointing out that the key-step is the estimation of the hidden variables. Indeed, a MAP estimate of the signal can be easily derived from the estimated values of $\hat{q}_j^{(o)}(\mathbf{k})$. More precisely, if $\hat{q}_j^{(o)}(\mathbf{k}) = 0$, we decide that the observation reduces to noise and then $\hat{\mathbf{s}}_j^{(o)}(\mathbf{k}) = \mathbf{0}$. At the opposite, if $\hat{q}_j^{(o)}(\mathbf{k}) = 1$, a MAP estimator of the signal of interest is given by

$$\hat{\mathbf{s}}_j^{(o)}(\mathbf{k}) = \arg\max_{\mathbf{u}} p_{\mathbf{s}_j^{(o)}}(\mathbf{u} \mid \mathbf{r}_j^{(o)}(\mathbf{k}), q_j^{(o)}(\mathbf{k}) = 1). \quad (3)$$

In this case, the pair of vectors $(\mathbf{r}_j^{(o)}(\mathbf{k}), \mathbf{s}_j^{(o)}(\mathbf{k}))$ is a Gaussian vector whose posterior distribution is also Gaussian and

$$\mathrm{E}[\mathbf{s}_j^{(o)}(\mathbf{k}) \mid \mathbf{r}_j^{(o)}(\mathbf{k}), q_j^{(o)}(\mathbf{k}) = 1] = \mathbf{Q}_j^{(o)}\mathbf{r}_j^{(o)}(\mathbf{k}) \quad (4)$$

where $\mathbf{Q}_j^{(o)} \triangleq \mathbf{\Gamma}_j^{(s,o)}(\mathbf{\Gamma}_j^{(s,o)} + \mathbf{\Gamma}_j^{(n,o)})^{-1}$. Finally, we obtain a shrinkage rule that performs a tradeoff between a linear estimation in the sense of a minimum mean square error and a hard thresholding:

$$\hat{\mathbf{s}}_j^{(o)}(\mathbf{k}) = \begin{cases} \mathbf{Q}_j^{(o)}\mathbf{r}_j^{(o)}(\mathbf{k}) & \text{if } \hat{q}_j^{(o)}(\mathbf{k}) = 1 \\ \mathbf{0} & \text{otherwise} \end{cases}. \quad (5)$$

Different denoising methods may however be obtained depending on the estimation of the hidden variables. In our preliminary work concerning this multivariate framework, the hidden variables were assumed to be iid. We will now investigate extensions taking into account scale dependencies.

3. ESTIMATION OF THE HIDDEN VARIABLES

3.1. Previous works

In the case of binary random variables, the MAP estimator is the Bayesian estimator corresponding to a hit-or-miss cost. In [15], we have firstly used an estimate $\hat{q}_j^{(o)}(\mathbf{k}) = 1$ of $q_j^{(o)}(\mathbf{k})$ based on an iid assumption. This leads to

$$p(\mathbf{r}_j^{(o)}(\mathbf{k})/q_j^{(o)}(\mathbf{k}) = 0) < p(\mathbf{r}_j^{(o)}(\mathbf{k})/q_j^{(o)}(\mathbf{k}) = 1). \quad (6)$$

In [18], we have introduced an inter-scale dependence in the modelling of $q_j^{(o)}(\mathbf{k})$. More precisely, recall that each father wavelet coefficient at position $\mathbf{k}_\mathrm{f} = (k_{(1,\mathrm{f})}, k_{(2,\mathrm{f})})$ has four children located in the next finest subband at positions $\mathbf{k}$ that are $(2k_{(1,\mathrm{f})}, 2k_{(2,\mathrm{f})})$, $(2k_{(1,\mathrm{f})}, 2k_{(2,\mathrm{f})}+1)$, $(2k_{(1,\mathrm{f})}+1, 2k_{(2,\mathrm{f})})$ or $(2k_{(1,\mathrm{f})}+1, 2k_{(2,\mathrm{f})}+1)$. If a father wavelet coefficient has a significant magnitude, its children are also likely to be significant and *vice-versa*. In [18], a simple estimator was proposed so as to exploit this fact. An estimate $\hat{q}_j^{(o)}(\mathbf{k}) = 1$ of each of the four children $q_j^{(o)}(\mathbf{k})$ was obtained by using the value of its father $q_{j+1}^{(o)}(\mathbf{k}_\mathrm{f})$ as follows:

$$\begin{aligned} &\mathsf{P}(q_j^{(o)}(\mathbf{k}) = 0/q_{j+1}^{(o)}(\mathbf{k}_\mathrm{f}), \mathbf{r}_j^{(o)}(\mathbf{k})) \\ &\qquad < \mathsf{P}(q_j^{(o)}(\mathbf{k}) = 1/q_{j+1}^{(o)}(\mathbf{k}_\mathrm{f}), \mathbf{r}_j^{(o)}(\mathbf{k})). \end{aligned} \quad (7)$$

In the sequel, we will designate by Inter I the inter-scale method described by Equation (7). We now investigate two alternative methods for the estimation of the hidden variables.

3.2. Inter II

This method relies on the use of the observation vectors of all the *ancestors* located at the coarser resolutions. In other words, it is decided that $\hat{q}_j^{(o)}(\mathbf{k}) = 1$ if:

$$\begin{aligned} &\mathsf{P}(q_j^{(o)}(\mathbf{k}) = 0/(\mathbf{r}_{j'}^{(o)}(\mathbf{k}_{\mathrm{f}_{j'}})_{j'})_{j' \geq j}) \\ &\qquad < \mathsf{P}(q_j^{(o)}(\mathbf{k}) = 1/(\mathbf{r}_{j'}^{(o)}(\mathbf{k}_{\mathrm{f}_{j'}})_{j'})_{j' \geq j}), \end{aligned} \quad (8)$$

where $\mathbf{k}_{\mathrm{f}_{j'}}$ denotes the ancestor position at scale level j' of the wavelet coefficient located at position $\mathbf{k}$ at resolution level j. To proceed further, we assume that the hidden variables $q_j^{(o)}$ form a Markov chain from coarse-to-fine scales. After some manipulations, it can be shown that the MAP rule is equivalent to maximize the following conditional probability $\mathcal{P}_j$:

$$\mathcal{P}_j \triangleq \sum_{q_{j+1}^{(o)},\dots,q_J^{(o)}} \Big(\prod_{\ell \geq j} p(\mathbf{r}_\ell^{(o)}/q_\ell^{(o)})\Big)\Big(\prod_{j \leq \ell < J} P(q_\ell^{(o)}/q_{\ell+1}^{(o)})\Big)\mathsf{P}(q_J^{(o)}), \quad (9)$$

where we have deliberately dropped the spatial indices for the sake of clarity. At resolution level j, a MAP estimate $\hat{q}_j^{(o)} = 1$ of $q_j^{(o)}$ is obtained iff:

$$p(\mathbf{r}_j^{(o)}/q_j^{(o)} = 0)O_j^{(o)} < p(\mathbf{r}_j^{(o)}/q_j^{(o)} = 1)U_j^{(o)}, \quad (10)$$

where

$$O_j^{(o)} \triangleq \sum_{q_{j+1}^{(o)},\dots,q_J^{(o)}} \mathsf{P}(q_j^{(o)}=0/q_{j+1}^{(o)})\Big(\prod_{j<\ell\leq J} p(\mathbf{r}_\ell^{(o)}/q_\ell^{(o)})\Big) \Big(\prod_{j<\ell<J} \mathsf{P}(q_\ell^{(o)}/q_{\ell+1}^{(o)})\Big)\mathsf{P}(q_J^{(o)}), \tag{11}$$

$$U_j^{(o)} \triangleq \sum_{q_{j+1}^{(o)},\dots,q_J^{(o)}} \mathsf{P}(q_j^{(o)}=1/q_{j+1}^{(o)})\Big(\prod_{j<\ell\leq J} p(\mathbf{r}_\ell^{(o)}/q_\ell^{(o)})\Big) \Big(\prod_{j<\ell<J} \mathsf{P}(q_\ell^{(o)}/q_{\ell+1}^{(o)})\Big)\mathsf{P}(q_J^{(o)}). \tag{12}$$

Let $\mathbf{M}_j^{(o)}$ denote the definite positive matrix:

$$\forall j=1,\dots,J \quad \mathbf{M}_j^{(o)} = (\mathbf{\Gamma}_j^{(n,o)})^{-1} - (\mathbf{\Gamma}_j^{(s,o)}+\mathbf{\Gamma}_j^{(n,o)})^{-1}. \tag{13}$$

After some simple calculation, we obtain:

$$\hat{q}_j^{(o)} = \begin{cases} 1 & \text{if } (\mathbf{r}_j^{(o)})^T\mathbf{M}_j^{(o)}\mathbf{r}_j^{(o)} > \chi_j^{(o)}, \\ 0 & \text{else,} \end{cases} \tag{14}$$

$$\text{where } \chi_j^{(o)} \triangleq 2\ln\left(\frac{|\,\mathbf{\Gamma}_j^{(n,o)}+\mathbf{\Gamma}_j^{(s,o)}\,|^{1/2}\, O_j^{(o)}}{|\,\mathbf{\Gamma}_j^{(n,o)}\,|^{1/2}\, U_j^{(o)}}\right). \tag{15}$$

It is worth noticing that the variables $O_j^{(o)}$ and $U_j^{(o)}$ can be computed recursively.

3.3. Inter III

The motivation of this method is to use the hidden variables of all the ancestors of a given wavelet node, in addition to the observation vector at the current resolution level. A main advantage of this approach compared to Inter II is to relax the markovianity assumption for the hidden variables.

At resolution level j, we assume that the values of the hidden variables are available at coarser resolutions $j+1,\dots,J$. (In practice, the values of these hidden variables are estimated at the considered resolution level using a scale-recursive approach, starting from the coarsest resolution.) Hence, $q_j^{(o)}$ is estimated by 1 if:

$$\mathsf{P}(q_j^{(o)}(\mathbf{k})=0/(q_{j'}(\mathbf{k}_{f_{j'}})^{(o)})_{j'>j},\mathbf{r}_j^{(o)}(\mathbf{k})) < \mathsf{P}(q_j^{(o)}(\mathbf{k})=1/(q_{j'}(\mathbf{k}_{f_{j'}})^{(o)})_{j'>j},\mathbf{r}_j^{(o)}(\mathbf{k})). \tag{16}$$

The Bayes theorem states that:

$$\mathsf{P}(q_j^{(o)}/(q_{j'}^{(o)})_{j'>j},\mathbf{r}_j^{(o)}) \propto \mathsf{P}(q_j^{(o)})\mathsf{P}((q_{j'}^{(o)})_{j'>j},\mathbf{r}_j^{(o)}/q_j^{(o)}). \tag{17}$$

As conditionally to $q_j^{(o)}$, the observation $\mathbf{r}_j^{(o)}$ is assumed to be independent of the hidden variables $q_{j'}^{(o)}$ for $j'>j$, we can deduce that:

$$\mathsf{P}((q_{j'}^{(o)})_{j'>j},\mathbf{r}_j^{(o)}/q_j^{(o)})\mathsf{P}(q_j^{(o)}) = p(\mathbf{r}_j^{(o)}/q_j^{(o)})\mathsf{P}(q_j^{(o)}/(q_{j'}^{(o)})_{j'>j}) \mathsf{P}((q_{j'}^{(o)})_{j'>j}). \tag{18}$$

Therefore, an estimate $\hat{q}_j^{(o)}=1$ of $q_j^{(o)}$ is obtained if:

$$\mathsf{P}(q_j^{(o)}=0/(q_{j'}^{(o)})_{j'>j})p(\mathbf{r}_j^{(o)}/q_j^{(o)}=0) < \mathsf{P}(q_j^{(o)}=1/(q_{j'}^{(o)})_{j'>j})p(\mathbf{r}_j^{(o)}/q_j^{(o)}=1). \tag{19}$$

After some basic manipulations, the following MAP estimator for $q_j^{(o)}$ is derived:

$$\hat{q}_j^{(o)} = \begin{cases} 1 & \text{if } (\mathbf{r}_j^{(o)})^T\mathbf{M}_j^{(o)}\mathbf{r}_j^{(o)} > \xi_j^{(o)} \\ 0 & \text{elsewhere} \end{cases}, \tag{20}$$

where the positive threshold $\xi_j^{(o)}$ is defined as:

$$\xi_j^{(o)} = 2\ln\left(\frac{\mathsf{P}(q_j^{(o)}=0/(q_{j'}^{(o)})_{j'>j})}{\mathsf{P}(q_j^{(o)}=1/(q_{j'}^{(o)})_{j'>j})}\right) + \ln\left(\frac{|\mathbf{\Gamma}_j^{(s,o)}+\mathbf{\Gamma}_j^{(n,o)}|}{|\mathbf{\Gamma}_j^{(n,o)}|}\right). \tag{21}$$

There are $J-j$ hidden variables corresponding to the resolution levels coarser than the j-th one. Therefore, the $J-j$-uplet $(q_{j'})_{j'>j}$ of the Bernouilli variables can take 2^{J-j} different values. Then, we can deduce that the threshold $\xi_j^{(o)}$ also can take 2^{J-j} different values according to the values taken by the considered $J-j$-uplet.

4. ESTIMATION OF THE HYPERPARAMETERS

All the described methods need a determination of some hyperparameters. Those of the Bernoulli-Gaussian prior ($\mathbf{\Gamma}_j^{(s,o)}$ and $\epsilon_j^{(o)}$) can be estimated thanks to the method of moments that we have already proposed in [15]. Estimates of the diagonal terms of the covariance matrix of the noise can be obtained by the classical median estimator [19]: $[\widehat{\mathbf{\Gamma}^{(n)}}]_{b,b} = \big(\hat{\sigma}[r_1^{(b,3)}]\big)^2$, where

$$\hat{\sigma}[r_1^{(b,3)}] = \frac{1}{0.6745}\mathrm{median}[r_1^{(b,3)}(\mathbf{k}),\ \mathbf{k}\in\{1,\dots,L_1\}^2]. \tag{22}$$

A robust estimator of the off-diagonal terms can also be performed [20]:

$$[\widehat{\mathbf{\Gamma}^{(n)}}]_{b,b'} = \frac{1}{4\alpha\beta}\Big(\big(\hat{\sigma}[\alpha r_1^{(b,3)}+\beta r_1^{(b',3)}]\big)^2 + \big(\hat{\sigma}[\alpha r_1^{(b,3)}-\beta r_1^{(b',3)}]\big)^2\Big) \tag{23}$$

where $\alpha = \big(\hat{\sigma}[r_1^{(b,3)}]\big)^{-1}$ and $\beta = \big(\hat{\sigma}[r_1^{(b',3)}]\big)^{-1}$. Concerning the Inter II method, the knowledge of the probability transitions $\mathsf{P}(q_j^{(o)}=1/q_{j+1}^{(o)}=1)$ and $\mathsf{P}(q_j^{(o)}=0/q_{j+1}^{(o)}=0)$ is additionally required. In [18], we have derived an efficient method for computing these quantities by considering the inter-covariance matrix $E[|\mathbf{r}_j^{(o)}|\,|\mathbf{r}_{j+1}^{(o)}|^T]-E[|\mathbf{r}_j^{(o)}|]E[|\mathbf{r}_{j+1}^{(o)}|]^T$ where $|\mathbf{u}|$ denotes the vector whose components are the absolute values of the components of $\mathbf{u}$. Inter III needs to compute the 2^{J-j} transition probabilities $\mathsf{P}(q_j^{(o)}/(q_{j'}^{(o)})_{j'>j})$. We propose to compute the 2^{J-j} absolute moments: $E(|\mathbf{r}_j^{(b,o)}|)$, $E(|\mathbf{r}_j^{(b,o)}||\mathbf{r}_{j+1}^{(o)}|^T)$, $E(|\mathbf{r}_j^{(o)}||\mathbf{r}_{j+2}^{(o)}|^T)$, …, $E(|\mathbf{r}_j^{(o)}||\mathbf{r}_J^{(o)}|^T)$, $E(|\mathbf{r}_j^{(o)}|\otimes|\mathbf{r}_{j+1}^{(o)}|\otimes|\mathbf{r}_{j+2}^{(o)}|)$, $E(|\mathbf{r}_j^{(o)}|\otimes|\mathbf{r}_{j+1}^{(o)}|\otimes|\mathbf{r}_J^{(o)}|)$,…, $E(|\mathbf{r}_j^{(o)}|\otimes|\mathbf{r}_{J-1}^{(o)}|\otimes|\mathbf{r}_J^{(o)}|)$, …, $E(|\mathbf{r}_j^{(o)}|\otimes|\mathbf{r}_{j+1}^{(o)}|\otimes|\mathbf{r}_{j+2}^{(o)}|\otimes\ldots\otimes|\mathbf{r}_J^{(o)}|)$ where $\otimes$ designates the Kronecker product. By conditioning w.r.t to all the hidden variables $q_j^{(o)}$, $q_{j+1}^{(o)}$,…, $q_J^{(o)}$ and reminding that observation at resolution j' only depends on the hidden variable $q_{j'}^{(o)}$, we can derive that every absolute moment can be expressed linearly w.r.t to the transition probabilities. For instance, $\mathcal{M}_j \triangleq E(|\mathbf{r}_j^{(o)}| \otimes |\mathbf{r}_{j+1}^{(o)}| \otimes |\mathbf{r}_{j+2}^{(o)}|)$ is expressed as

$$\begin{aligned}
&\mathcal{M}_j \\
&= \sum_{(q_{j'}^{(o)})_{j'\geq j}} E(|\mathbf{r}_j^{(o)}| \otimes |\mathbf{r}_{j+1}^{(o)}| \otimes |\mathbf{r}_{j+2}^{(o)}|/(q_{j'}^{(o)})_{j'\geq j})\mathsf{P}((q_{j'}^{(o)})_{j'\geq j}) \\
&= \sum_{(q_{j'}^{(o)})_{j'>j}} \mathsf{P}(q_{j'}^{(o)})_{j'>j})\Big(E(|\mathbf{r}_j^{(o)}|/q_j^{(o)}=1)\mathsf{P}(q_j^{(o)}=1/(q_{j'}^{(o)})_{j'>j})) \\
&+ E(|\mathbf{r}_j^{(o)}|/q_j^{(o)}=0)\mathsf{P}(q_j^{(o)}=0/(q_{j'}^{(o)})_{j'>j}))\Big) \otimes E(|\mathbf{r}_{j+1}^{(o)}|/q_{j+1}^{(o)}) \\
&\otimes E(|\mathbf{r}_{j+2}^{(o)}|/q_{j+2}^{(o)}).
\end{aligned}$$

Besides the b-th component $r_{j'}^{(b,o)}/q_{j'}^{(o)}$ has an univariate normal distribution, which allows us to write:

$$\begin{aligned}
E[|r_{j'}^{(b,o)}|/q_{j'}^{(o)}=0] &= \left(\tfrac{2}{\pi}[\mathbf{\Gamma}_{j'}^{(n,o)}]_{b,b}\right)^{\frac{1}{2}} \\
E[|r_{j'}^{(b,o)}|/q_{j'}^{(o)}=1] &= \left(\tfrac{2}{\pi}[\mathbf{\Gamma}_{j'}^{(s,o)}+\mathbf{\Gamma}_{j'}^{(n,o)}]_{b,b}\right)^{\frac{1}{2}}
\end{aligned} \quad . \qquad (24)$$

Therefore, if the joint probabilities $\mathsf{P}((q_{j'}^{(o)})_{j'>j}))$ are available, the unknown probabilities $\mathsf{P}(q_j^{(o)}=1/(q_{j'}^{(o)})_{j'>j})$ satisfy a *linear* system of, at least, 2^{J-j} equations. Then, the joint probabilities are estimated recursively by following a coarse-to-fine strategy.

5. EXPERIMENTAL RESULTS

Simulations were performed on multispectral SPOT images of size 512×512. Realizations of a zero-mean Gaussian multivariate were artificially added to these components. In our simulations, WTs based on Symmlets of order 8 have been used. However, additional simulations carried out with other orthogonal wavelets yield the same conclusions. The performances are expressed in terms of a Signal-to-Noise Ratio (SNR) averaged over the B components and they have been compared to those obtained with the most up-to-date interscale wavelet-based methods. All of the latter methods are applied separately to each spectral component. In particular, we have tested the Hidden Markov Tree (HMT) [10] and the locally bivariate shrinkage method [14]. In our experiments, it appeared that Inter II did not lead to significant improvements compared with Inter I. Hence, this method has been discarded in our subsequent comparisons.

Table 1 provides the resulting SNRs for a four stage decomposition. These results indicate that both the intra scale and the Inter I methods perform better than the efficient bivariate method. A multivariate approach should therefore be preferred to a monovariate one in the case of multispectral images. Furthermore, it can be observed that Inter III outperforms all the tested methods. It yields an average gain of 0.14 dB w.r.t. Inter I and, it improves the SNR up to 0.78 dB w.r.t the Bivariate method. Visual inspection of the denoised images shows that the proposed method preserves the edges better than the intrascale MAP-BG and there are less granular artifacts.

Table 1. Kairouan ($B = 4$): performances in terms of SNR (in dB) of some wavelet-based denoising methods (symmlet8).

Initial	HMT	Bivariate	MAP-BG intra	MAP-BG inter I	MAP-BG inter III
7.01	11.41	11.57	12.00	12.15	**12.30**
8.02	12.04	12.17	12.55	12.77	**12.90**
9.00	12.65	12.79	13.14	13.38	**13.51**
10.00	13.23	13.43	13.74	14.01	**14.10**
11.00	13.86	14.10	14.38	14.66	**14.76**
12.02	14.49	14.80	15.05	15.33	**15.42**
13.03	15.13	15.51	15.72	16.00	**16.09**
14.00	15.74	16.21	16.37	16.66	**16.77**

6. REFERENCES

[1] I. Pitas, A. Venetsanopoulos, *Nonlinear digital filters*, Kluwer, Dodrecht, 1990.

[2] B. Vidakovic, *Statistical modeling by wavelets*, John Wiley & Sons, New-York, 1999.

[3] S. G. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, 1998.

[4] P. Moulin, J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. on IT,* vol. 45, no. 3, pp. 909-919, April 1999.

[5] A. Antoniadis, D. Leporini, J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statistica Neerlandica*, vol. 56, issue 4, pp. 434-453, December 2002.

[6] P. Ishwar, P. Moulin, "On the equivalence of set-theoretic and maxent MAP estimation," *IEEE Trans. on SP,* vol. 51, no. 3, pp. 698-713, March 2003.

[7] H.A. Chipman, E.R. Kolaczyk, R.E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *JASA*, vol. 92, no. 440, pp. 1413-1421, December 1997.

[8] F. Abramovich, T. Sapatinas, B.W. Silverman, B.W., "Wavelet thresholding via a Bayesian approach," *J. of the Royal Stat. Soc.*, Series B, vol. 60, pp. 725-749, 1998.

[9] D. Leporini, J.-C. Pesquet, H. Krim,"Best basis representations based on prior statistical models," in *Bayesian Inference in Wavelet Based Models*, P. Müller and B. Vidakovic editors, Springer Verlag, 1999.

[10] J.K. Romberg, H. Choi, R.G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. on IP*, vol. 10, no. 7, pp. 1056-1068, July 2001.

[11] A. Pizurika, W. Philips, I. Lemahieu, M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. on IP*, vol. 11, no. 5, pp. 545-556, May 2002.

[12] M. Malfait, D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. on IP*, vol. 6, no. 4, pp. 549-565, April 1997.

[13] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on SP,* vol. 41, no. 12, pp. 3445-3462, December 1993.

[14] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. on SP*, vol. 50, no. 11, pp. 2744-2756, November 2002.

[15] A. Benazza-Benyahia, J.-C. Pesquet, "Wavelet-based multispectral image denoising with Bernouilli-Gaussian model," *Proc. of NSIP'03*, Grado, Italy, 8-11 June 2003.

[16] A. Benazza-Benyahia, J.-C. Pesquet, "An extended SURE approach for multicomponent image denoising", *Proc. of the ICASSP'2004*, Montreal, Canada, May 2004.

[17] P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. on IP,* vol. 13, no. 4, pp. 411-416, April 2004.

[18] A. Benazza-Benyahia, J.-C. Pesquet, "An Interscale multivariate MAP estimation of multispectral images," *Proc. of the EUSIPCO'04*, Wien, Austria, September 2004.

[19] D.L. Donoho, I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. of the American Stat. Assoc.*, vol. 90, no. 432, pp. 1200-1224, December 1995.

[20] P. J. Huber, *Robust statistics*, John Wiley & Sons, New-York, NJ, USA, 1981.