

Variations on Markovian Quadtree Model for Multiband Astronomical Image Analysis

Christophe J.-F. Collet and Farid Flitti

LSIIT UMR CNRS 7005, Université Strasbourg 1 (ULP)
Bd S. Brant, BP 10413, 67412 Illkirch, France
collet@lsiit.u-strasbg.fr , flitti@lsiit.u-strasbg.fr

Abstract. This paper is concerned with the analysis of multispectral observations, provided by space or ground telescopes. The large amount and the complexity of heterogeneous data to analyse lead us to develop new methods for segmentation tasks, which aim to be robust, fast and efficient. Some prior knowledge on the information to be extracted from the original image is available, and Bayesian statistical theory is known to be a convenient tool to take this *a priori* knowledge into consideration. In this paper, we investigate the use of the Bayesian inference on Markovian quadtrees for some reduction, fusion, segmentation or restoration problems of great importance in multiband astronomical imagery.

Keywords: Markovian quadtree, Bayesian inference, fusion, data reduction, copulas, astronomy.

1 Introduction

This paper deals with the unsupervised segmentation, reduction, fusion or restoration of multiband images. These different tasks are developed in an astronomical multispectral imagery framework, and validated on raw data cubes. The main goal of this presentation, consists in showing different processing chains describing the power, the efficiency and the fruitfulness of hierarchical Markovian modeling based on a quadtree topology. We will see that such modeling allows to deal with a large varieties of data : missing data, multiresolution data, multiband data, strongly noised data. In particular, we show how such approach is general and how this tool is able to face with a large number of various image processing tasks. The paper is organized as follows. The Markovian quadtree model is described in section 2. In section 3 a reduction methods on large data cube is coupled with quadtree modeling, in order to provide a single segmentation map avoiding thus the curse of dimensionality phenomenon. Then, in the fourth section, we propose to process the wavelet coefficients on the raw data cube, and feed a Markovian quadtree with the multiscale coefficients of the wavelet transform. Indeed, the quadtree topology exhibits a suitable structure to deal with multiscale coefficients : in this way, it becomes possible to use the different multi-scale segmentation maps obtained along a quadtree to restore and fused multiband images. Particularly, the problem of between-channels correlation modeling in the non-Gaussian case is briefly presented.

2 Markovian quadtree and segmentation tasks

Statistical Markovian approaches have proved to be fruitful to design robust and efficient images analysis methods. In the context of multispectral images, handling

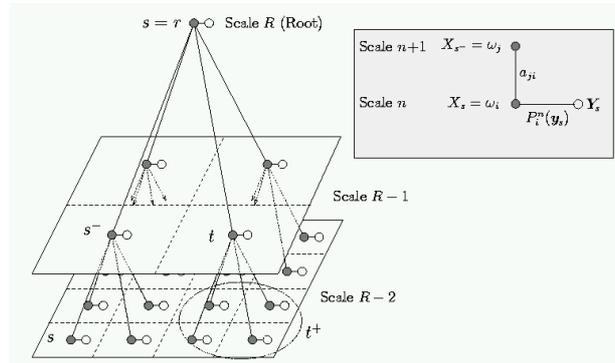


Fig. 1. Example of a dependency graph corresponding to a quadtree structure on a 16×16 lattice. Black circles represent labels and white circles represent multi-component observations. Each node t has a unique parent t^- and four “children” t^+ . a_{ij} stands for inter-scale probability of label transition, whereas $P_i^n(y_s)$ represents the likelihood to affect a label ω_i with observation y_s . Likelihood parameters and Markovian quadtree parameters (a_{ij} and root probabilities) can be estimated with, *e.g.*, an EM algorithm. The segmentation algorithm re-estimates iteratively the parameters of a given hidden in-scale-Markov model, to produce a new model which has a higher probability of generating the given observation sequence. This re-estimation procedure is continued until no more significant improvement in parameters can be obtained. The two-step computation of posterior marginals propagates available information all over the tree : on one hand, the bottom-up step spreads the influence of data to other levels up to the root, on the other hand the top-down step computes the posterior marginals taking into account this information. Thus, this proposed modeling scheme captures, over the quadtree, significant statistical dependencies and provides a robust scheme for segmentation.

correlated observed data requires a well-designed modeling framework. Resorting to a Bayesian scheme based on Markov models is indeed attractive when dealing with large amount of multispectral observations. Nevertheless, the well known Markov Field Models (MFM) lead to iterated optimization algorithms, not really well adapted [Geman and Geman, 1984, Graffigne *et al.*, 1995, Kato *et al.*, 1996] for many applications, even if some strategies to decrease the computing time have been proposed in the last decade (e.g., [Pérez *et al.*, 2000, Mignotte *et al.*, 2000]). This is due to the fact that most of Markov models are non-causal. As a consequence, inference must be conducted iteratively, which might turn prohibitively expensive. One way to circumvent this problem is to resort to a Markov model on a quadtree where in-scale causality[Laferté *et al.*, 2000, Provost *et al.*, 2003] permits non-iterative inference . A quadtree-based approach offers the well-known advantages of standard hierarchical techniques (improved robustness, ability to deal with multiresolution or missing data), while allowing for non-iterative inference as in the case of hidden Markov chains [Giordana and Pieczynski, 1997]. Let $G = (S, L)$ be a graph composed of a set S of nodes and a set L of edges. A tree is a connected graph with no cycle, where as a consequence, each node apart from the root r has

a unique predecessor, its "parent", on the path to the root. A quadtree, as illustrated in Fig. 1, is a special case of tree where each node, apart from the terminal ones, the "leaves", has four "children". The set of nodes S can be partitioned into "scales", $S = S^0 \cup S^1 \dots \cup S^R$, according to the path length from each node to the root. Thus, $S^R = \{r\}$, S^n involves 4^{R-n} sites, and S^0 is the finest scale formed by the leaves. We consider a labeling process X which assigns a class label X_s to each node of $G : X = \{X^n\}_{n=0}^R$ with $X^n = \{X_s, s \in S^n\}$ where X_s takes its values in the set $\Omega = \{\omega_1, \dots, \omega_K\}$, of the K classes. A number of conditional independence properties are assumed. First, X is supposed to be Markovian in scale, *i.e.*,¹ $P(x^n|x^k, k > n) = P(x^n|x^{n+1})$. It is also assumed that the probabilities of inter-scale transitions can be factorized in the following way [Laferté *et al.*, 2000]:

$$P(x^n|x^{n+1}) = \prod_{s \in S^n} P(x_s|x_{s^-}), \quad (1)$$

where s^- designates the father of site s , as illustrated in Fig. 1. Finally, the likelihood of the multiband/multisensor observations \mathbf{Y} conditionally to X is expressed as the following product (assuming conditional independence):

$$P(\mathbf{Y}=\mathbf{y}|x) = \prod_{n=0}^R P(\mathbf{y}^n|x^n) = \prod_{n=0}^R \prod_{s \in S^n} P(\mathbf{y}_s^n|x_s), \quad (2)$$

where $\forall s \in S^n, \forall n \in \{0, \dots, R\}, P(\mathbf{y}_s|x_s = \omega_i) \triangleq f_i(\mathbf{y}_s)$, captures the likelihood of the data y_s . Each site s of scale n can be associated with a label ω_i . If data are available at scale n , then the likelihood is expressed as $f_i^n(\mathbf{y}_s^n)$. Of course, if the data-driven terms do not follow a Gaussian law, the analytic expression of the multidimensional density $f_i^n(\mathbf{y}_s^n)$ is not always available. To overcome this difficulty, one may decorrelate bands via an adequate mapping, compute the multidimensional density of the decorrelated data as a simple product of the marginals and then obtain $f_i^n(\mathbf{y}_s^n)$ by Jacobian method [Provost *et al.*, 2003]. Another solution is to use copulas theory [Nelsen, 1998][Brunel *et al.*, 2005] (see Annexe). In section 3, we present a new way for multidimensional data-driven term computation, thanks to a regularized mixture of Probabilistic Principal Component.

Sometimes, the lack of observed data on some locations within the pictures leads to intricate segmentation problems but here, missing data can be easily inferred [Provost *et al.*, 2003]. In a general manner, we suppose the data available at different levels n , including the finest level ($n = 0$). On one hand, when no observation exists (for any given scale n), the likelihood $f_i^n(\mathbf{y}_s^n)$ is set to 1. On the other hand, if we have images of the same area at different levels of resolution, the quadtree structure can be still used and permits to properly consider all the available data. It is a way to conduct the segmentation while merging data. From these assumptions, it can be easily inferred that the joint distribution $P(x, \mathbf{y})$ can be factorized as follows :

$$P(x, \mathbf{y}) = P(x_r) \prod_{s \neq r} P(x_s|x_{s^-}) \prod_{n=0}^R \prod_{s \in S^n} P(\mathbf{y}_s|x_s). \quad (3)$$

One of the interests of this model lies in the possibility of computing exactly the posterior marginals $P(X_s|\mathbf{Y})$ and $P(X_s, X_{s^-}|\mathbf{Y})$ at each node s within two passes

¹ To simplify notation, we will denote the discrete probability $P(X = x)$ as $P(x)$.

in an unsupervised way [Delmas, 1997]. The segmentation label map \hat{x} to be determined is finally given by:

$$\hat{x}_{s,s \in S^n} = \arg \max_{\omega_i \in \Omega} P(X_s = \omega_i | \mathbf{Y} = \mathbf{y}). \quad (4)$$

Equation (4) shows that we obtain a labeling of each pixel at each level of the quadtree, even if observations only lie on the finest level and even if there is missing data.

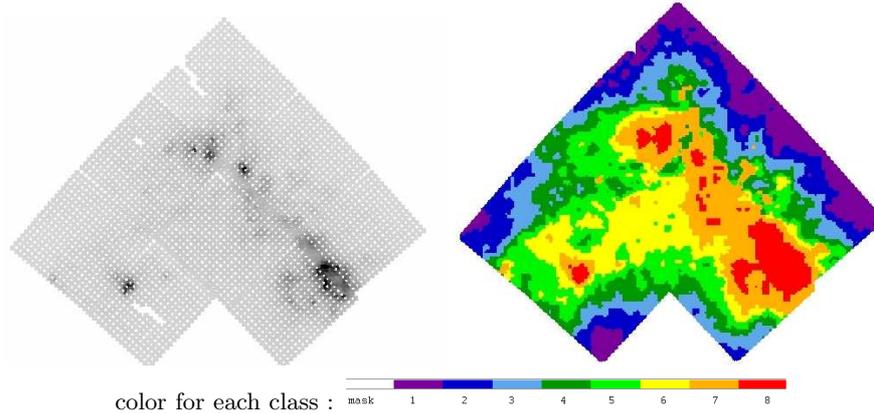


Fig. 2. On the left picture, composed of a mosaic of 9 observations, the missing pixel data, due to sampling adjustment problems, appear as a regular lattice of white dots. The Markovian quadtree allow to reconstruct a segmentation map without missing labels : the missing observations are labeled thanks to the Markovian *a priori* model.

3 Reduction/Segmentation on the Quadtree

Analysis of multicomponent data sets is a very hard task, due to the curse of dimensionality [Hughes, 1968]. Indeed, learning algorithms need a large diversity of observations to cover the behavior of the studied process. Especially, in the multidimensional case, the required number of samples grows quickly with the dimension, so that the process behavior becomes rapidly untractable in practice. This is the so-called Hughes phenomenon which corresponds to an important loss of accuracy in the process statistics estimation as dimensionality grows (more precisely the likelihood term in the quadtree). For example, for an observation size of $H \times W$ pixels by D spectral bands, one more channel observed adds $H \times W$ additional samples whereas the complexity deals with \mathbb{R}^{D+1} . To deal with this problem, one may carry out a space reduction step before classification [Landgrebe, 2003]. Fortunately, high dimensional observed data can often be described in a significantly smaller number of dimension than the original due to redundancy in data cube where neighboring bands are highly correlated. Many approaches were proposed to solve such analysis task. All seek a mapping on a reduced dimension space by maximizing a given criterion [Duda *et al.*, 2001]. More graceful solution consists

on combining reduction and classification by associating a generative model to the observations within each class to compute the corresponding likelihood. Thus the observations are modeled as a mixture of such generative models [Tipping and Bishop, 1999, Lee *et al.*, 2000]. In this paper we propose to use a Markovian *a priori* associated with such generative models to regularize multidimensional pixel classification. In the sequel this approach will be illustrated using the Probabilistic Principal Component analysis (PPCA) generative model.

3.1 Probabilistic Principal Component analysis (PPCA)

The PPCA [Tipping and Bishop, 1999] is based on a latent variable model which links each $D \times 1$ observed vector \mathbf{y} to $q \times 1$ latent vector \mathbf{t} , $q < D$, as follows:

$$\mathbf{y} = A\mathbf{t} + \mu + \epsilon \quad (5)$$

where A is a $D \times q$ matrix, μ the observed data mean and ϵ is a random variable following an Gaussian $\mathcal{N}(0, \sigma^2 I)$ noise, I being the identity matrix. Given \mathbf{t} and Eq. 5, the \mathbf{y} probability distribution is :

$$P(\mathbf{y}/\mathbf{t}) = (2\pi\sigma^2)^{-\frac{D}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y} - A\mathbf{t} - \mu\|^2\right\}. \quad (6)$$

Choosing Gaussian *prior* for \mathbf{t} , *i.e.*; $\mathcal{N}(0, I)$, the marginal distribution of \mathbf{y} is

$$P(\mathbf{y}) = (2\pi)^{-\frac{D}{2}} |C|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu)^t C^{-1}(\mathbf{y} - \mu)\right\} \quad (7)$$

with $C = \sigma^2 I + AA^t$ a $D \times D$ matrix. Using the Bayes rule, the *a posteriori* probability of \mathbf{t} is found to be [Tipping and Bishop, 1999] $\mathcal{N}(M^{-1}A^t(\mathbf{y} - \mu), \sigma^2 M^{-1})$ where $M = \sigma^2 I - A^t A$.

The maximization of the data log-likelihood $\mathcal{L} = \sum_{s \in S^0} \ln\{p(\mathbf{y}_s)\}$ gives the following parameter estimators :

$$\mu_{ML} = \frac{\sum_{s \in S^0} \mathbf{y}_s}{\text{card}(S^0)}; \quad \sigma_{ML}^2 = \frac{1}{D - q} \sum_{j=q+1}^D \lambda_j; \quad A_{ML} = U_q(A_q - \sigma^2 I)^{\frac{1}{2}} R. \quad (8)$$

where λ_j are the eigenvalues of the data covariance matrix $\Sigma_x = \frac{1}{\text{card}(S^0)} \sum_{s \in S^0} (\mathbf{y}_s - \mu)(\mathbf{y}_s - \mu)^t$ given in descending order ($\lambda_1 \geq \dots \geq \lambda_q$), A_q is a diagonal matrix of the q largest eigenvalues, U_q the matrix of the corresponding eigenvectors, and R is an arbitrary orthogonal rotation matrix.

3.2 Regularized mixture of Probabilistic Principal Component analyzers

A mixture of PPCA (MPPCA) was proposed in [Tipping and Bishop, 1999] to model complex data structures as a combination of local PCA. For a K component MPPCA, the observations are partitioned in K clusters (*i.e.*; classes) each one spanned by a local PPCA. Given this model, the distribution of the observations is $P(\mathbf{y}_s) = \sum_{i=1}^K \pi_i P(\mathbf{y}_s/x_s = \omega_i)$. Note that in this formulation the *prior* is the same for all $s \in S^0$ and thus, any information about the neighborhood is taken into account when classifying \mathbf{y}_s . We adapt this model by imposing a Markovian constraints via the quadtree modelling. The observation distribution become

$P(\mathbf{y}_s) = \sum_{i=1}^K P(x_s = \omega_i)P(\mathbf{y}_s/x_s = \omega_i)$, where X_s is drawn from a hierarchical Markovian process (Eq. 1) and

$$P(\mathbf{y}_s/x_s = \omega_i) = (2\pi)^{\frac{-D}{2}} |C_i|^{\frac{-1}{2}} \exp\left\{\frac{-1}{2}(\mathbf{y} - \mu_i)^t C_i^{-1}(\mathbf{y} - \mu_i)\right\}. \quad (9)$$

The matrix C_i is obtained in analog manner to Eqs. 7 and 8 by eigen-decomposition of the weighted covariance matrix $\Sigma_i = \frac{\sum_{s \in S^0} P(x_s = \omega_i/Y)(\mathbf{y}_s - \hat{\mu}_i)(\mathbf{y}_s - \hat{\mu}_i)^t}{\sum_{s \in S^0} P(x_s = \omega_i/Y)}$, where $\hat{\mu}_i = \frac{\sum_{s \in S^0} P(x_s = \omega_i/Y)\mathbf{y}_s}{\sum_{s \in S^0} P(x_s = \omega_i/Y)}$. The estimation of the *a priori* parameter remains the same as in the classical quadtree. To test our approach, we generate 3 sets of 3 images (2 classes (geometric shape and background) with Gaussian distribution (mean 120/120/128 and 136/136/128, standard deviation 16/16/16). Thus we obtain 9 images to segment. The obtained 4-classes segmentation map shows clearly the better behavior of our proposed approach towards MPPCA (cf. Fig. 3).

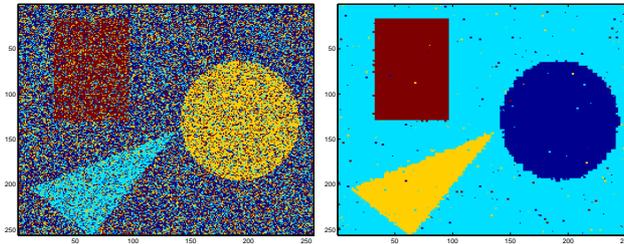


Fig. 3. Segmentation map obtained with the MPPCA on 9 images (left) remains noisy whereas the map obtained with the proposed technique (right) is well regularized.

4 Wavelet domain for restoration and fusion tasks

Fusion of multiband images is of great interest in astronomy, allowing to obtain an efficient summary of the whole multiband information in a single scene. Generally this task is more difficult for noisy observations. The wavelet domain is well adapted both for fusion [Zhang and Blum, 1999] and denoising [D.L.Donoho and Johnstone, 1994] tasks. Actually, wavelet coefficients measure local variations in the image and the sharper the discontinuity, the larger the coefficients. Intensity fluctuations corresponding to the noise, most of time considered as uncorrelated, are most important at the finest resolution and related wavelet coefficients decrease quickly as the scale increases. Real structures in the image will therefore lead to larger wavelet coefficient values at these coarsest resolutions. A threshold can be defined at each scale below which all the coefficients are discarded [D.L.Donoho and Johnstone, 1994]. Note that the result of such analysis depends strongly both on the wavelet used and on the thresholds chosen. Generally astronomical objects are diffuse and exhibit smooth edges so isotropic wavelet transforms are well adapted [Starck *et al.*, 1998]. We use the pyramidal algorithm with one wavelet which is an isotropic transform obtained by adapting the classical Laplacian pyramid [Starck *et al.*, 1998].

Few years ago [Crouse *et al.*, 1998], an efficient Markovian modeling of wavelets was introduced capturing interscale and spacial wavelet coefficient correlations. In

this paper we use a more general Markovian framework modeling not only spatial and interscale dependencies as the existent models but also interband correlation for multiband image fusion and denoising. Moreover, the multidimensional likelihood may be efficiently modeled using the copulas theory [Nelsen, 1998] allowing us to use any kind of marginal densities with a given interband correlation. The

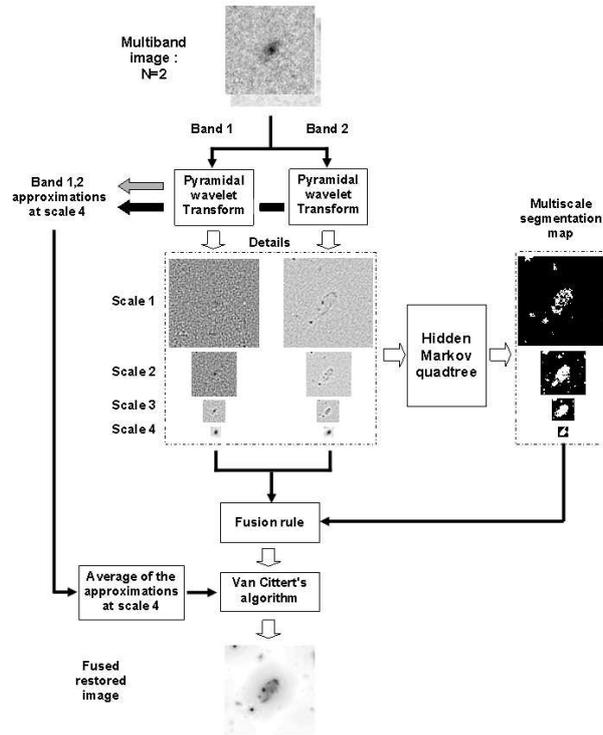


Fig. 4. Fusion-restoration algorithm illustrated for a bi-band image. A pyramidal wavelet transform analyzes the two spectral bands (on the top). This leads to a multiresolution pyramid of wavelet coefficients for each band, up to scale 4. Then, all wavelet pyramids are combined to carry out two-class multiresolution Markovian segmentation map (on the right). This segmentation map masks small coefficients at different scales. The remainder coefficients are fused using an appropriate rule. The result with the average of coarsest approximations feed an iterative reconstruction procedure to give a unique fused restored image.

proposed approach is summarized in Fig.4. For a multiband image \mathbf{Y} with D bands, a wavelet decomposition is carried out for each band b separately leading to a multiresolution pyramids \mathcal{W}^b , $b = 1, \dots, D$. These D pyramids are combined in unique Multiband-Multiresolution Pyramid (MMP, cf. Fig. 4 and 5) \mathcal{W} by considering details coefficients, $\mathcal{W}_{s \in S^j}^1, \dots, \mathcal{W}_{s \in S^j}^D$, for space location s at scale j as a

components of a unique vector $\mathcal{W}_j(s)$. The MMP is segmented in two-classes (*i.e.*; $\forall s \in S : x_s \in \{0, 1\}$) using a vectorial hidden Markov quadtree (Fig.5) to separate significant wavelet coefficients from those associated with the noise. The selection relies now not only on the sole coefficient magnitude but also takes into account its neighbors : in space, in scale and with wavelength. This classification scheme produces a multiresolution binary mask highlighting significant wavelet coefficients and removing the others, corresponding to the noise contribution. The fusion of the *cleaned* wavelet coefficients is operated using the following rule :

$$\forall s \in S^n : W_s^{fused} = \frac{\sum_{i=1}^D \sigma_i^n x_s \mathcal{W}_s^i}{\sum_{i=1}^D \sigma_i^n x_s}, \quad (10)$$

σ_i^n being the standard deviation of the i^{th} marginal of the likelihood associated with class kept at scale n . The structure W^{fused} does not correspond to a smooth image since all non significant coefficients are put to zero before fusion. We seek instead a smooth solution \hat{F}^{fused} which minimizes $\| (W^{fused} - O(\hat{F}^{fused})) \|$ where O is the wavelet transform operator. In practice we use the Van Cittert's algorithm [Starck *et al.*, 1998] to obtain the final restored-fused image (see Fig. 6).

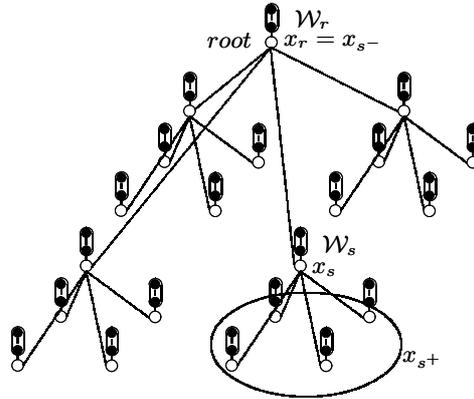


Fig. 5. Example of a dependency graph corresponding to a quadtree structure on a 4×4 lattice. White circles represent labels and black circles represent multiband observations $\mathcal{W}_s, s \in S$ in the wavelet domain.

Conclusion

This paper summarizes some variations around Markovian quadtree model, in order to show the efficiency of such a tool, to deal with unsupervised multiband image analysis, for *e.g.*, reduction, segmentation, restoration, fusion tasks. Our motivations for using such a model are to provide fast computations and efficient structures to process multispectral and multiresolution large observations. Indeed, computer vision and astronomers communities need efficient tools to analyse and interpret large data cubes : ground or on-board telescopes provide larger amount of multispectral/multiresolution data cube every year, that have to be processed in an efficient way.

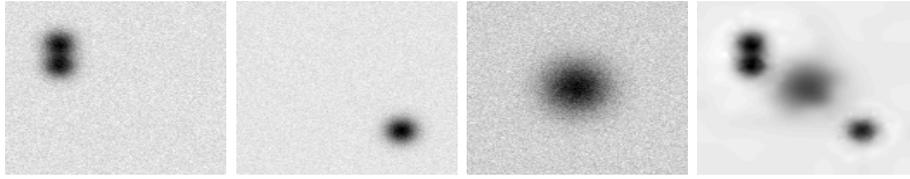


Fig. 6. Example of image fusion: from the left three simulated bands, the fusion result is on the right. All objects appearing in the three bands are present in fused image.

Annexe : Copulas for N-D likelihood computation

The basis of the copulas theory is Sklar's Theorem [Nelsen, 1998] which asserts the existence of a function C , called copula and defined on $[0, 1]^N$, binding the joint cumulative distribution function $F(\mathbf{y}_s^1, \dots, \mathbf{y}_s^N)$ to the marginal cumulative distribution functions $F^{[1]}(\mathbf{y}_s^1), \dots, F^{[N]}(\mathbf{y}_s^N)$ as follows : $F(\mathbf{y}_s^1, \dots, \mathbf{y}_s^N) = C(F^{[1]}(\mathbf{y}_s^1), \dots, F^{[N]}(\mathbf{y}_s^N))$. If the marginals $F^{[1]}, \dots, F^{[N]}$ are continuous, then C is unique. Moreover, if C is differentiable it is possible to define a copula density as [Nelsen, 1998]:

$$f(\mathbf{y}_s^1, \dots, \mathbf{y}_s^N) = f^{[1]}(\mathbf{y}_s^1) \times \dots \times f^{[N]}(\mathbf{y}_s^N) \times c(F^{[1]}(\mathbf{y}_s^1), \dots, F^{[N]}(\mathbf{y}_s^N)) \quad (11)$$

where $f^{[j]}(\mathbf{y}_s^j)$ is the probability density function corresponding to $F^{[j]}(\mathbf{y}_s^j)$ and $c = \partial C / (\partial F^{[1]}, \dots, \partial F^{[N]})$ is the copula density. For multivariate Gaussian copula C_G , the copula density is given by [Nelsen, 1998]:

$$\forall \mathbf{t} = (t^1, \dots, t^N)^T \in \mathbb{R}^N : c_G(\mathbf{t}) = |R|^{-\frac{1}{2}} \exp \left[-\frac{\tilde{\mathbf{t}}^T (R^{-1} - I) \tilde{\mathbf{t}}}{2} \right] \quad (12)$$

where $\tilde{\mathbf{t}} = (\Phi^{-1}(t^1), \dots, \Phi^{-1}(t^N))^T$ with $\Phi(\cdot)$ the standard Gaussian cumulative distribution, R is the $N \times N$ correlation matrix of $\tilde{\mathbf{t}}$ and I the same size identity matrix. To model non-Gaussian multivariate densities, we use Eq. 11 with a Gaussian copula density (Eq. 12) and Generalized Gaussian marginal densities [Provost *et al.*, 2003] each one characterized by three parameters namely the mean, the standard deviation and the shape parameter. This modeling allows us to cover Upper-Gaussian (shape parameter < 2), Gaussian (shape parameter $= 2$) and Sub-Gaussian (shape parameter > 2) multidimensional densities. See [Nelsen, 1998] for more details on copulas theory.

References

- [Brunel *et al.*, 2005] N. Brunel, W. Pieczynski, and S. Derrorde. Copulas in vectorial hidden Markov chains for multicomponent image segmentation. In *ICASSP'05*, Philadelphia, USA, March 19-23 2005.
- [Crouse *et al.*, 1998] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Image Processing*, 46(4), April 1998.

- [Delmas, 1997]J.-P. Delmas. An equivalence of the EM and ICE algorithm for exponential family. *IEEE-T-SP*, 45(3):2613–2615, October 1997.
- [D.L.Donoho and Johnstone, 1994]D.L.Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, September 1994.
- [Duda *et al.*, 2001]R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
- [Geman and Geman, 1984]S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984.
- [Giordana and Pieczynski, 1997]N. Giordana and W. Pieczynski. Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):465–475, 1997.
- [Graffigne *et al.*, 1995]C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle, and J. Zerubia. Hierarchical Markov random field models applied to image analysis : a review. In *SPIE Neural Morphological and Stochastic Methods in Image and Signal Processing*, volume 2568, pages 2–17, San Diego, 10-11 July 1995.
- [Hughes, 1968]G.F. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Information Theory*, 14(1):55–63, 1968.
- [Kato *et al.*, 1996]Z. Kato, M. Berthod, and J. Zerubia. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *Graphical Models and Image Processing*, 58(1):18–37, 1996.
- [Laferté *et al.*, 2000]J.-M. Laferté, P. Pérez, and F. Heitz. Discrete markov image modeling and inference on the quad-tree. *IEEE-T-IP*, 9(3):390–404, March 2000.
- [Landgrebe, 2003]D. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley and Sons, 2003.
- [Lee *et al.*, 2000]T.W. Lee, M.S. Lewicki, and T.J. Sejnowski. ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1078–1089, October 2000.
- [Mignotte *et al.*, 2000]M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Trans. on Image Processing*, 9(7):1–17, July 2000.
- [Nelsen, 1998]R. B. Nelsen. *An introduction to copulas*. Lecture Notes in Statistics. Springer, New York, 1998.
- [Pérez *et al.*, 2000]P. Pérez, A. Chardin, and J.-M. Laferté. Noniterative manipulation of discrete energy-based models for image analysis. *Pattern Recognition*, 33(4):573–586, April 2000.
- [Provost *et al.*, 2003]J.N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy. Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps. *Computer Vision and Image Understanding*, 93(2):155–174, December 2003.
- [Starck *et al.*, 1998]J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.
- [Tipping and Bishop, 1999]M. E. Tipping and C. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, (11):443–482, 1999.
- [Zhang and Blum, 1999]Z. Zhang and R. S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a

digital camera application. *Proceedings of the IEEE*, 87(8):1315 – 1326, August 1999.