Multiband Segmentation Based on a Hierarchical Markov Model

Christophe Collet (1) and Fionn Murtagh (2)*
(1) LSIIT – UMR CRNS 7005 – Bd. Sébastien Brant
BP 10413, 67412 Illkirch, France
(2) School of Computer Science, Queen's University Belfast,
Belfast BT7 1NN, Northern Ireland

April 7, 2004

Abstract

We develop a new multiscale Markov segmentation model for multiband images. Using quadtree multiple resolution analysis of a multiband image, we use both inter-scale and intra-scale spatial Markov statistical dependencies. Bayesian inference is used to assess the appropriate number of segments. We exemplify the excellent results which can be obtained with this approach using synthetic images, and in two case studies involving multiband astronomical image sets.

Keywords: multispectral image, multiband image, multiresolution, multiscale, quadtree, Markov random field, generalized Gaussian distribution, Bayesian inference, Bayes factor, Bayes information criterion.

1 Introduction

Image segmentation is a classical problem in computer vision. High quality image segmentation remains a difficult issue in spite of much work devoted

*Corresponding author. Fax +44 28 9097-5666. *E-mail addresses:* christophe.collet@ensps.u-strasbg.fr, f.murtagh@qub.ac.uk

to it. Astronomy image processing typifies such a need, given the demanding signal retrieval and noise obfuscation properties of images in this field.

Motivated by the importance of utilizing information at various scales, multiresolution approaches have been proposed for astronomical image segmentation [1, 2], mainly to benefit from the computational efficiency that can be achieved by use of such models.

Statistical approaches have proven fruitful for the design of robust and efficient segmentation methods [3, 4, 5, 6]. In the context of multispectral images, handling correlated observed data requires a well-designed modeling framework [7, 8, 9, 10]. Resorting to a Bayesian scheme based on hierarchical Markov models is attractive when dealing with a large amount of multispectral observations, because of its ability to take into account simultaneously the whole set of observed data and its capacity to learn model parameters. Computationally efficient hierarchical Markov models have been recently proposed [8, 10, 11, 12]. In this work we take hierarchical Markov models further, in order to address the multispectral image segmentation task.

On the one hand, multiresolution models based on wavelet transforms provide nonredundant and directionally sensitive features at different scales, and hence they allow more selective feature extraction in the space-frequency domain. Such approaches use observation-based transforms at different resolutions in order to better describe the information provided by the observed data. On the other hand, hierarchical Markovian models incorporate regularization constraints, arising from a priori knowledge about acceptable solutions, directly into the solution space. In our opinion, these two approaches are very complementary, as a number of recent papers show: see [13, 14].

A separate difficulty to be addressed concerns the quantity of observations met with in astronomical multispectral imagery. In addition, we have to take into account explicitly the spectral correlation between observed wavelengths and within them (spatial redundancy). The lack of clearly defined edges in astronomical images, and the well-known Hughes phenomenon [15] due to sparseness of observation space, imply difficulties in interpreting the resulting segmentation.

In this article, the proposed modeling scheme captures, over a pyramidal graph and under a Markovian-in-scale assumption, significant intra-scale and inter-scale statistical dependencies. Our motivation for using such a model is to provide fast computation, improved robustness and an effective interpretational framework for processing large multispectral and multiresolution images.

2 Hierarchical Markovian Model

The introduction of spatial dependencies, that have to be inferred, between hidden random variables was first rigorously formalized in the context of spatial statistics [16] and image restoration [17], using the so-called Gibbs distribution. Such a distribution which stems from statistical physics, allows probability distributions to be specified and manipulated in a flexible way over huge sets of random variables that interact with each other on a local basis. From Bayes' theorem such prior distributions are easily combined with data models similar to those used in parametric classifiers (and thereby used for segmentation).

From a statistical point of view, the introduction of a contextual prior makes each pixel labeling dependent on all others, which might appear as an unbearable constraint. However, this prior is usually formulated in such a way that the conditional probability of some pixel label, given the labeling of all the other pixels, reduces to the conditional probability given only the labeling of the neighboring pixels. This is the so-called Markov property. The Markov property then allows simple iterative inference based on local propagation of information: each pixel is visited a number of times, and it is each time assigned to the class that is the most probable given the spectral information it bears and the current assignment of neighboring pixels.

Such Markov models are non-causal [3]. As a consequence, inference must be conducted iteratively, which might turn prohibitively expensive when dealing with large data images, such as multispectral images.

More recent studies in image analysis have suggested replacing the purely spatial priors with hierarchical priors, in which interactions between variables are not supported by the grid over the image pixels but are defined from scale to scale. In that case, one is no longer handling a single classification but a set of classifications at various scales. Such hierarchical approaches can also provide substantial gains in speed and quality of results. In the particular case where the underlying interaction structure is a tree whose leaves fit the pixels of the image [18, 19, 10, 11], MPM (modes of posterior marginal) inference is performed exactly, in a noniterative way, by propagating information first from the leaves to the root, and then in the reverse direction. The final assignment of each node of the tree is eventually obtained based on all the data. In the remainder of this section, we will describe the MPM inference approach.

Let $G = (S, L)$ be a graph composed of a set of nodes S and a set of edges,

L. A tree is a connected graph with no cycles, 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 child nodes. The set of nodes S can be partitioned into scales, $S = S^0 \cup S^1 \ldots \cup S^R$, according to the path length of 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:

$$x = \{x^n\}_{n=0}^{R} \text{ with } x^n = \{x_s, s \in S^n\}, \tag{1}$$

where x_s takes its values in the set $\Delta = \{\omega_1, \ldots, \omega_K\}$, of the K classes or segments. A number of conditional independence properties are assumed.

First, x is supposed to be Markovian in scale, i.e.

$$P\left(x^n \left| x^k, k > n\right.\right) = P\left(x^n \left| x^{n+1}\right.\right). \tag{2}$$

To simplify notation, we denote the discrete probability $P(X = x)$ as $P(x)$. It is also assumed that the probabilities of inter-scale transitions can be factorized in the following way [10, 20]:

$$P\left(x^n \left| x^{n+1}\right.\right) = \prod_{s \in S^n} P(x_s \left| x_{s^-}\right.). \tag{3}$$

where s^- designates the single father of a site s, as illustrated in Fig. 1.

Finally, the likelihood of the observations $\boldsymbol{y}$ conditionally on x is expressed as the following product (assuming conditional independence):

$$P(\boldsymbol{y}|x) = \prod_{n=0}^{R} P(\boldsymbol{y}^n|x^n) = \prod_{n=0}^{R} \prod_{s \in S^n} P(\boldsymbol{y}_s|x_s), \tag{4}$$

where $\forall s \in S^n$, $\forall n \in [0, \ldots, R]$, $P(\boldsymbol{y}_s|x_s = \omega_i) \stackrel{\triangle}{=} P_i^n(\boldsymbol{y}_s)$ captures the likelihood of the data y_s, formed by the vector of values at site or pixel s of scale n, given label ω_i.

We obtain a labeling of each pixel at each level of the quadtree, even if signal of interest to us only lies on the finest level. This is due to the fact that the two-step computation of posterior marginals propagates available

information all over the tree. The bottom-up step spreads the influence of data to other levels up to the root. Then, the top-down step computes the posterior marginals taking into account this information.

Of course, if observations are available for other nodes of the tree, this algorithm merges all the observations to obtain a more accurate segmentation result. The root of the quadtree is also a node of the pyramid.

In this article, data are only available at the finest level ($n = 0$), and each pixel is associated with a multiband vector. When no observation exists ($n > 0$) or when the site at scale 0 does not belong to an area which is considered classifiable (e.g. when there is a known erroneous pixel or missing observation), the likelihood $P_i^n(\boldsymbol{y}_s)$ is set to 1.

From these assumptions, it can be easily inferred that the joint distribution $P(x, \boldsymbol{y})$ follows a Gibbs law, the expression of which is given by [10]:

$$P(x, \boldsymbol{y}) = P(x_r) \prod_{s \neq r} P(x_s | x_{s^-}) \prod_{s \in S} P(\boldsymbol{y}_s | x_s) \tag{5}$$

One of the interests of this model lies in the possibility of computing exactly the posterior marginals $P(x_s|\boldsymbol{Y})$ and $P(x_s, x_s^-|\boldsymbol{Y})$ at each node s with two passes. These computed expressions will be first used in the iterative parameter estimation step, as described in the next section. The segmentation label map $\hat{x}$ to be determined is finally given by:

$$\hat{x}_s = \arg \max_{\omega_i \in \Delta} P(x_s = \omega_i | \boldsymbol{Y} = \boldsymbol{y}). \tag{6}$$

This approach on a quadtree structure exhibits similarities with Baum-Welch algorithm usage for hidden Markov chains, which is an iterative update algorithm that re-estimates parameters of a given hidden 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 probability can be obtained [9].

Starting from an initial label map given by a maximum likelihood based algorithm, parameters of which are estimated by a C-means algorithm [21], this technique converges to the maximum of the a posteriori pdf which is highly multi-modal; therefore good initialization is needed to guarantee fast convergence of the procedure to the global optimum.

This approach is robust and accurate as shown with the label map obtained on some synthetic 3-band images (Fig. 2 and Table 1: see results in Fig. 3). Two conclusions can be drawn from this analysis: first, the hierarchical Markovian assumption is able per se to provide a homogeneous segmentation map in the presence of strong noise; second, the usefulness of an in-scale Markovian model stems from the ability to learn its parameters through re-estimation on the quadtree (bottom-up step spreading the influence of data to other levels up to the root; then, the top-down step computing the posterior marginals taking into account this information), and to provide a form of context handling in such pattern recognition problems.

3 Parameter Estimation

Segmenting an image into different regions or classes and estimating all the model parameters are intricate problems. Fixing or estimating correctly model parameters is crucial in practice to get an accurate and reliable segmentation solution. Here, we are dealing with an unsupervised labeling problem. We have first to estimate the set of parameters $\Phi = \{\Phi_x, \Phi_y\}$, where:

- Φ_y denotes the parameters involved in the data model. The conditional data likelihood (i.e., data driven term) is entirely defined by $P_i^0(\boldsymbol{y}_s) = P(\boldsymbol{Y}_s = \boldsymbol{y}_s | x_s = \omega_i, s \in S^0)$ since the only observations available are associated with the leaves of the label tree x. In this context, several models of image noise could be used. We consider the generalized Gaussian distribution (GG) model as described in Section 4, in order to cope with a large variety of situations.

- $\Phi_x = \left((a_{ij})_{i,j=1}^K, (\pi_i)_{i=1}^K\right)$ is the set of the prior Markov model parameters where $a_{ij} \stackrel{\triangle}{=} P(X_s = \omega_j \,|X_{s^-} = \omega_i)$ and $\pi_i \stackrel{\triangle}{=} P(X_r = \omega_i)$.

We estimate these two sets of parameters using an estimation method [9] termed Iterative Conditional Estimation (ICE). This method requires finding two estimators, namely $\widehat{\Phi}_x = \Phi_x(X)$ and $\widehat{\Phi}_y = \Phi_y(X, \boldsymbol{Y})$. When X is however unobservable (i.e., incomplete data), the iterative ICE procedure introduces the estimates $\widehat{\Phi}_x^{[k+1]}$ and $\widehat{\Phi}_y^{[k+1]}$ at step $k+1$ defined as the conditional expectations of Φ_x and Φ_y given $\boldsymbol{Y} = \boldsymbol{y}$, and the current parameter values $\widehat{\Phi}_x^{[k]}$ and $\widehat{\Phi}_y^{[k]}$:

$$\widehat{\Phi}^{[k+1]} = E\left[\widehat{\Phi} \mid \boldsymbol{Y} = \boldsymbol{y}, \widehat{\Phi}^{[k]}\right]. \tag{7}$$

They are the best approximations of Φ_x and Φ_y with respect to the mean square error. When $E\left[\widehat{\Phi}/\boldsymbol{Y} = \boldsymbol{y}\right]$ is not calculable, but sampling from the posterior distribution $P(X|\boldsymbol{Y}, \Phi)$ is feasible, one can use stochastic approximation (in our case, the tree structure considered enables the exact computation of the posterior marginals). It is thus possible to determine the exact update of the prior parameter values, and $\widehat{\Phi}_x$ is the maximum likelihood (ML) estimator:

$$\begin{cases} a_{ij} = \frac{\sum_{s \neq r} \Psi_s(i,j)}{\sum_{s \neq r} \xi_{s^-}(i)} \\ \pi_i = \xi_r(i) \end{cases} \tag{8}$$

with

$$\begin{cases} \xi_s(i) \stackrel{\triangle}{=} P\left(X_s = \omega_i / \boldsymbol{Y} = \boldsymbol{y}, \Phi^{[k]}\right), \\ \Psi_s(i,j) \stackrel{\triangle}{=} P(X_s = \omega_j, X_{s^-} = \omega_i \,/\boldsymbol{Y} = \boldsymbol{y}, \Phi^{[k]}). \end{cases} \tag{9}$$

The complete procedure including the two-pass computation of posterior marginals Ψ_s and ξ_s is presented in the Appendix. The ICE algorithm results in a deterministic estimation of the a priori parameters (Equation 8), whereas the data-driven parameters Φ_y are updated in a stochastic way through sampling from $P\left(X|\boldsymbol{Y} = \boldsymbol{y}, \Phi^{[k]}\right)$.

4 Noise Modeling

As an image noise model, we consider the family of generalized Gaussian (GG) distributions $\mathcal{G}$ which is well-suited to a large variety of correlated multispectral data, and is defined as follows:

$$\mathcal{G}(z \mid \mu, \sigma^2, p) = [2\Gamma(1/p)]^{-1} \eta(p)\, p \, \exp\left[-(\eta(p)|z-\mu|)^p\right] \tag{10}$$

where $\Gamma(.)$ is the gamma function, p is a positive shape parameter governing the rate of decay ($p = 1$ for Laplacian noise, $p = 2$ for Gaussian noise, $p > 8$ for nearly uniform noise), μ is the mean, and

$$\eta(p) \triangleq \left[\frac{\Gamma(3/p)}{\sigma^2\Gamma(1/p)}\right]^{1/2} \tag{11}$$

where σ^2 is the variance. For small values of p (i.e., $p < 2$), this probability density function has a heavier tail than a Gaussian density. This noise model is very appropriate for modeling physical data properties, and has been recently used in many different contexts (watermarking [22], underwater acoustics [23], multiresolution analysis [24], multispectral analysis [11]). In particular, it incorporates Gaussian noise, impulse noise, and mixtures of these.

4.1 Hypothesis

In order to validate the use of GG marginal distributions for image segmentation, the hierarchical Markov algorithm was, firstly, used on 3-spectral-band synthetic images (Fig. 2) and secondly, on 4- and 6-band multispectral astronomical images. The observed random vector $\boldsymbol{Y}_s$, conditional on $X_s = \omega_i$, described by its positive definite symmetric covariance matrix, Σ_i^y ($i = 1, \ldots, K$), admits the Crout (Cholesky) factorization:

$$\Sigma_i^y = L_i\, L_i^T \tag{12}$$

where L_i is a $C \times C$ unique lower triangular matrix. Then, the correlated vector $\boldsymbol{Y}_s$ can be transformed into an uncorrelated vector:

$$\boldsymbol{Z}_s = A_i \boldsymbol{Y}_s \qquad \text{with } A_i = L_i^{-1}. \tag{13}$$

Directly from equation (13), we have $\boldsymbol{\mu}_i^z = A_i \boldsymbol{\mu}_i^y$ and the pdfs of $\boldsymbol{Z}_s$ and $\boldsymbol{Y}_s$ are related via:

$$\begin{aligned} P_{\boldsymbol{Y}_s|X_s}(\boldsymbol{y}_s|x_s = \omega_i) &= |A_i| P_{\boldsymbol{Z}_s|X_s}(A_i\boldsymbol{y}_s|x_s = \omega_i) \\ &= |A_i| \prod_{c=1}^{C} = g_i^c(\underbrace{[A_i\boldsymbol{y}_s]^{(c)}}_{z_s^{(c)}}) \end{aligned} \tag{14}$$

where $|.|$ denotes the determinant.

4.2 Moment Based Estimators of the GG Parameters

Let us consider N random samples z_i, $i = 1, \ldots, N$, assumed to be independent and identically distributed (iid) from the pdf given in equation (10). Three parameters are required to characterize the GG pdf from the samples of z_i. First, the empirical mean $\widehat{\mu}$ is defined:

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} z_i \tag{15}$$

Secondly, the estimator $\widehat{\sigma}^2$ of σ^2, based on the second order moment, is simply the empirical variance:

$$\widehat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{\mu})^2 \tag{16}$$

Finally, the estimation of the shape parameter p is based on the estimation of the centered fourth order moment $\mu(4)$ of the GG pdf given by the general relation:

$$\mu(n) = E[(Z - E[Z])^n] = \frac{\Gamma(\frac{n+1}{p})}{\Gamma(1/p)\eta(p)^n} \quad \text{for } n \text{ even.} \tag{17}$$

By substituting equation (11) in equation (17) for $n = 4$ and using estimators of mean $\widehat{\mu}$, variance $\widehat{\sigma}^2$ and centered fourth-order moment $\widehat{\mu}(4)$, we obtain the following relation:

$$\widehat{\mu}(4) = \widehat{\sigma}^4 \, \frac{\Gamma(5/p)\Gamma(1/p)}{\Gamma(5/p)^2} \triangleq \widehat{\sigma}^4 \, h(p) \tag{18}$$

where

$$\widehat{\mu}(4) = \frac{1}{N} \sum_{i=1}^{N} (z_i - \widehat{\mu})^4 \tag{19}$$

Equation (18) is numerically solved in order to obtain the estimator $\widehat{p}$ of the shape parameter p. The result is a unique estimate because of the monotonic decreasing property of function $h(p)$ for $p > 0$.

5 Model Selection

Through Bayesian inference we will assess output segmentations to decide on the best number of segments to be retained. There is no need to consider all parameters resulting from the hierarchical Markov modeling. Instead we will abstract out the following simpler, more focused and easily analyzed model: the segmentation maps, X are used; as observed data, Y, we take the principal component image of the bands used; the conditional likelihood $p(y_i \mid x_i = \omega_k)$ is taken as Gaussian; and a spatial Markov random field model is used, unlike the tree-based Markov model used up to now.

An approximate Bayes factor decision approach will be used to determine the best segmentation size (i.e., number of classes in the segmentation map). We take the segmentation, described as above, with a fixed number of classes K, as our prior. We denote this overall model, which we fit to our data, as M_K. M_K is defined from the distributions used and all parameters, Φ_K. We now wish to investigate one such model versus another, i.e. M_K versus $M_{K'}$ for two choices of numbers of classes, K and K'. The posterior probability of model M_K, with pixel labeling X, is

$$p(M_K \mid X) = \frac{p(X \mid M_K)p(M_K)}{\sum_{l=1}^{K} p(X \mid M_l)p(M_l)} \tag{20}$$

We can ignore $p(M_K)$ and the influence of M_l if each model is equilikely a priori. The Bayes factor is the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are equal: $p(X \mid M_K)/p(X \mid M_{K'})$. The term $p(X \mid M_K)$ is the integrated likelihood rather than the maximized likelihood, given by

$$p(X \mid M_K) = \int p(X \mid \Phi_K, M_K)p(\Phi_K)d\Phi_K \tag{21}$$

We have that $p(X \mid \Phi_K, M_K)$ is the usual likelihood. Finally $p(\Phi_K)$ is the prior, which we will assume as equilikely for all M_K. A good approximation to the integrated likelihood is given in terms of BIC, Bayes information criterion [25, 26]:

$$\text{BIC} = 2 \log p(X \mid \widehat{\Phi}_K, M_K) - \mid \Phi_K \mid \log N \tag{22}$$

where $\widehat{\Phi}_K$ is the maximum likelihood estimator of Φ_K, i.e. the result of the Gaussian mixture fitting. N is the dimensionality of the observation vectors,

and $\mid \Phi_K \mid$ is the cardinality of the parameter set. Finally the Bayes factor is approximated by the difference of BIC terms, which in turn are the maximized likelihood results of model fits for different numbers of classes, K and K':

$$2 \log \frac{p(X \mid M_K)}{p(X \mid M_{K'})} \approx \mathrm{BIC}(K) - \mathrm{BIC}(K') \tag{23}$$

We can also derive the BIC term as a parsimony or minimum information measure [27, 28]. In the Markov random field spatial interaction situation, a combinatorial explosion comes about with use of the likelihood. BIC rests on the likelihood of the observed data:

$$\sum_j p(y_i \mid x_i = \omega_j, K) p(x_i = \omega_j \mid K)$$

Since there are K^N possible configurations for N pixels we use the more tractable pseudo-likelihood instead of the likelihood. Rather than summing over all possible configurations of x, we consider configurations close to the target estimate of x, $\widehat{x}$. We consider each pixel vector y_i in turn and condition on $\widehat{x}_{Ni}$, i.e. $\widehat{x}$ excluding the value at i. Then, instead of the likelihood,

$$\sum_{j=1}^{K} p(y_i \mid x_i = \omega_j) p(x_i = \omega_j)$$

we use the pseudo-likelihood,

$$\sum_{j=1}^{K} p(y_i \mid x_i = \omega_j) p(x_i = \omega_j \mid N_i)$$

The first term of this expression is a Gaussian distribution, and the second term is the conditional distribution used with the Potts prior density.

The pseudo-likelihood of the image (the y_i terms are independent, conditional on the underlying hidden states) is then (with ϕ the Potts spatial dependency value):

$$\prod_i p(y_i \mid x_{Ni}, \phi) = \prod_i \sum_{j=1}^{K} p(y_i \mid x_i = \omega_j) p(x_i = \omega_j \mid N_i, \phi) \tag{24}$$

This is the likelihood integrated over the approximate posterior distribution of a set of models near the MAP estimate of x. Denoting equation (24) as $\mathrm{PL}(Y \mid K)$, BIC which addresses the computational implications of the spatial model becomes [29, 30]:

$$BIC_{PL}(K) = 2 \log \mathrm{PL}(Y \mid K) - \mid \Phi_K \mid \log N \tag{25}$$

BIC_{PL}, or BIC updated to take a Markov spatial model into account, is used in the same way as BIC, i.e. we vary $K = 2, 3, \ldots$ and find a first relative maximum. The curves produced do not usually increase to a plateau. Alternative approximations of BIC that take spatial contiguity into account are studied in [31].

The model selection procedure described provides a principled framework, which we show in the following sections to be scalable to real problems involving multiband (and potentially the multi-temporal, multisource data, with ancillary GIS data, used in [32]).

6 Results

Our motivation for using the hierarchical Markov model is to provide fast computation and to have an effective framework for processing large multispectral images. Astronomy is an example of just one application field with burgeoning needs in this direction. More generally, data grids have become a direction of research which is now seeing considerable attention. Data grids involve firstly and foremostly the federation of large distributed data stores. The work described in this paper provides a solution for some problems in this area.

Our segmentation method was recently applied with success to multispectral images, typically with 32 bit floating point pixel values in each band, available at Strasbourg Astronomical Observatory. We used this approach on astronomical images of the M82 region, the nearest starburst galaxy at a distance of 11 million light years from Earth. Massive stars are forming and expiring in M82 (nickname: the cigar galaxy) at a rate ten times higher than in our galaxy. Images come from the Nicmos archive, an instrument on board the Hubble Space Telescope. The mode and filters used were: NIC3 in F164N-F166N, FeII band and nearby continuum at wavelengths 1080, 1130, 1640, 1660, 2120 and 2150 nm. In Fig. 4 we show these graylevel images. The astronomy interest in these 6 bands is to separate the emission of the

stellar component and of the gas and to detect evidence of starburst activity in the central region of the galaxy.

The Bayes factor approach to model selection provides an objective measure of a number of classes versus another. In section 5, we described how the Bayes factor can be approximated by the Bayes information criterion, BIC, and furthermore how it can be changed to take spatial information into account. Taking spatial information into account is important, given that a Markov modeling procedure was used to find class assignments. The resulting criterion is called BIC_{PL} [30]. A recent application to multiband data segmentation (not availing of a multiple resolution scale model) is in [33].

To determine the optimal number of classes, we determined BIC_{PL} for various numbers of classes, $\mathrm{K} = 3, 4, \ldots, 15$. We did this with respect to the first principal component image as an optimal low dimensional representative of the image, and to avoid instability problems which become common in higher dimensional, sparse spaces. Fig. 5 shows the outcome. We find that $\mathrm{K} = 12$ provides for the best segmentation: this corresponds to the largest value of BIC_{PL}. The segmentation map obtained with $K = 12$ classes is shown in Fig. 6.

In a second study, we applied our algorithm to 3 infrared images of the Orion nebula from the Two Micron All Sky Survey, 2MASS, in 3 filter bands: J (1250 nm), H (1650 nm) and Ks (2170 nm). A color composition of the 3 bands can be seen in Fig. 7(a) and can be compared with the 3 original image bands in Fig. 7(b,c,d). BIC_{PL} values are shown for varying numbers of segments, K, in Fig. 8, pointing to particular interest in the $\mathrm{K} = 14$ result. The segmentation map in 14 classes is displayed in Fig. 7(e).

These results are now under evaluation and further discussion with astronomers at Strasbourg Astronomical Observatory. Each detected class is associated with astronomical regions or objects, using geometrical, statistical and spectral properties. Of particular interest are faint plume features. Interpretation of classes requires astronomer knowledge and the results of the classification algorithm.

The hierarchical modeling approach described is new, and is very promising in the field of astronomical image processing [34, 35, 6] because it takes into account a scale decomposition of the information. Simultaneously the multiband aspects are managed in a general manner, allowing all observed information to be taken into account during the segmentation task. In this work, we have additionally seen how the use of a Bayes factor criterion provides a way to objectively decide on the number of segments.

There is a burgeoning need for segmentation in this context of multiband image data. The number of filters and channels provided by detectors in astronomy is increasing continually. The fusion of all observed data in a general model of segmentation can be contrasted with the difficulties involved in less multidimensional processing [35]. "Multiple wavelength astronomy" has become the order of the day, with such major national and international initiatives as the virtual observatory (e.g., National Virtual Observatory: www.us-vo.org). The approach described in this article is a response to this growing need.

We conclude this section with a note on the very favorable computational requirements of our algorithms. The quadtree-based segmentation for images in Fig. 2, comprising 3 bands, and yielding 2 classes, was about 1 minute. For Fig. 4 comprising 6 bands, and yielding 12 classes, we required about 20 minutes. For Fig. 7, comprising 3 bands, and yielding 14 classes, we required about 8 minutes. These timings were on a Pentium 4, 2 GHz, machine. For images comprising 3 to 6 bands, and yielding 12 to 14 classes, without the quadtree processing, a number of hours of computational time were required on a similar machine.

7 Conclusion and Perspectives

This article has addressed the problem of unsupervised multispectral image segmentation using a quadtree structure which is Markovian in scale. Results obtained are promising: the relevance and efficiency of the ICE estimation on the quadtree has been exemplified on real astronomical images, in the context of multispectral segmentation (three and six bands respectively, with 12 and 14 classes). The quadtree approach is computationally faster than the usual methods based on spatial Markov fields. There is, in fact, no iterative process in the updating step. Last but not least, we show how this approach opens the way to multichannel segmentation (with an arbitrary number of spectral bands), within a well defined Bayesian inference framework. Important contributions of this work include taking into account the correlation coefficients between the different channels for each segmentation class within an ICE procedure for parameter estimation. The segmentation is thus easily obtained, according to the well-known Bayesian MPM criterion.

Acknowledgments

The authors thank A. Lançon for providing the M82 images, A. Oberto for software development and F. Bonnarel for discussions at the CDS, Strasbourg Astronomical Observatory. This research was supported by the French government ACI-Grid/IDHA project: Action Concertée Incitative – Globalisation des Ressources Informatiques et des Données, "Images Distribuées Hétérogènes pour l'Astronomie", 2001-2003; and by COST Action 283, "iAstro: Computational and Information Infrastructure in the Astronomical Data-Grid".

References

[1] A. Bijaoui and F. Rué, "A multiscale vision model adapted to the astronomical images," *Signal Processing*, vol. 46, pp. 345–362, 1995.

[2] E.J. Candès J.-L Starck and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[3] 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 Proceedings, Neural Morphological and Stochastic Methods in Image and Signal Processing*, San Diego, 10-11 July 1995, vol. 2568, pp. 2–17.

[4] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Bayesian inference and optimization strategies for some detection and classification problems in sonar imagery," in *Nonlinear Image Processing – SPIE Conference 3646*, San Jose, USA, January 1999, vol. 3646-02, pp. 14–27.

[5] R. Molina, J. Núñez, F. Cortijo, and J. Mateos, "Image restoration in astronomy. A Bayesian review," *IEEE Signal Processing Magazine*, vol. 18, no. 2, pp. 11–29, 2001.

[6] D.A. Van Dyk, "Hierarchical models, data augmentation and MCMC," in *Statistical Challenges in Modern Astronomy III*, J.G. Babu and E.D. Feigelson, Eds. 2002, Springer-Verlag.

[7] P. Rostaing, J.-N Provost, and C. Collet, "Unsupervised multispectral image segmentation using generalized Gaussian noise model," in *International Workshop EMMCVPR'99: Energy Minimisation Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science*. 1999, vol. 1654, pp. 141–156, Springer Verlag.

[8] P. Pérez, A. Chardin, and J.-M. Laferté, "Noniterative manipulation of discrete energy-based models for image analysis," *Pattern Recognition*, vol. 33, no. 4, pp. 573–586, April 2000.

[9] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 465–475, 1997.

[10] J.-M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 390–404, March 2000.

[11] J.-N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy, "Multispectral SPOT images analysis using generalized Gaussian modeling: Application to water depth mapping," *Computer Vision and Image Understanding*, vol. submitted, January 2002.

[12] C. Kraaikamp M. Dekking, A. Elfeki and J. Bruining, "Multi-scale and multi-resolution stochastic modeling of subsurface heterogeneity by tree-indexed Markov chains," *Computational Geosciences*, vol. 5, pp. 47–60, 2001.

[13] H Noda, MN Shirazi, and E Kawaguchi, "MRF-based texture segmentation using wavelet decomposed images," *Pattern Recognition*, vol. 35, pp. 771–782, 2002.

[14] S. Ruan, B. Moretti, L. Fadili, and D. Bloyet, "Fuzzy Markovian segmentation in application of magnetic resonance images," *Computer Vision and Image Understanding*, vol. 85, pp. 54–69, 2002.

[15] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, pp. 55–63, 1968.

[16] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, pp. 179–195, 1977.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

[18] S. Lakshmanan and D. Grimmer, "A deformable template approach to detecting straight edges in radar images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 438–443, 1996.

[19] M.R. Luettgen, W.C. Karl, A.S. Willsky, and R. Tenney, "Multiscale representation of Markov random fields," *IEEE Transactions on Image Processing*, vol. 41, no. 12, pp. 3377–3395, December 1993.

[20] M. Luettgen, *Image Processing with Multiscale Stochastic Models*, PhD thesis, MIT Laboratory of Information and Decision Systems, May 1993.

[21] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley Interscience, 1973.

[22] C. Lu and H. Liao, "Oblivious watermarking using generalized Gaussian," *3rd Inter. Conf. on Computer Vision, Pattern Recognition and Image Processing*, vol. II, pp. 260–263, Atlantic City, USA, Feb. 28-Mar. 2000.

[23] A. Tesei and C.S. Regazzoni, "The asymmetric generalized Gaussian funcion: a new HOS-based model for generic noise pdfs," *IEEE Workshop on Signal processing, Corfu*, 1996.

[24] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," *Proc. IEEE TFTS Symposium*, pp. 633–636, Pittsburgh, PA, Oct. 6-9 1998.

[25] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[26] R.E. Kass and A.E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.

[27] M.H. Hansen and Bin Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, vol. 96, pp. 746–774, 2001.

[28] J. Rissanen, "Stochastic complexity and modeling," *Annals of Statistics*, vol. 14, pp. 1080–1100, 1986.

[29] D.C. Stanford, *Fast Automatic Unsupervised Image Segmentation and Curve Detection in Spatial Point Patterns*, PhD thesis, Department of Statistics, University of Washington, 1999.

[30] D.C. Stanford and A.E. Raftery, "Approximate Bayes factors for image segmentation: the pseudolikelihood information criterion (PLIC)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1517–1520, 2002.

[31] F. Forbes and N. Peyrard, "Hidden Markov model selection criteria based on mean field-like approximations," technical report 4371, INRIA Rhône-Alpes, 2001.

[32] A.H. Schistad Solberg, T. Taxt, and A.K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 100–113, 1996.

[33] F. Murtagh, D. Barreto, and J. Marcello, "Decision boundaries using Bayes factors: the case of cloud masks," *IEEE Transactions on Geoscience and Remote Sensing*, 2003, in press.

[34] A.S. Szalay, A.J. Connolly, and G.P. Szokoly, "Simultaneous multicolor detection of faint galaxies in the Hubble Deep Field," *The Astronomical Journal*, vol. 117, pp. 68–74, January 1999.

[35] J.L. Starck and F. Murtagh, "Astronomical image and signal processing," *IEEE Signal Processing Magazine*, pp. 1–10, March 2001.

Appendix: Evaluation of Posterior Marginals

In this Appendix we describe the two-pass computation of posterior marginals on the quadtree. We use notation from Fig. 1. The algorithm to be described provides evaluation of the partial posterior marginals at full resolution ($n = 0$).

Notes:
The symbol $\geq s$ denotes all the descendents of s, s included.
The symbol s^+ denotes the 4 children (level $n-1$) of the site s (level n).
The symbol s^- denotes the single parent (level $n+1$) of the site s (level n).

$$P(X_s = \omega_i | \boldsymbol{y}_{\geq s}) = P(X_s = \omega_i | \boldsymbol{y}_s) = \frac{P(X_s = \omega_i) P_i^0(\boldsymbol{y}_s)}{\sum_j P(X_s = \omega_j) P_j^0(\boldsymbol{y}_s)} \qquad \forall s \in S^0$$

where $P(X_s = \omega_i)$ is recursively evaluated through a top-down pass according to:

$$\begin{array}{l} \text{for } n = R-1 \ldots 0 \\ \quad \left| \; P(X_s = \omega_i) = \sum_j a_{ji} P(X_{s^-} = \omega_j) \quad \forall s \in S^n \right. \end{array}$$

Note that the a priori probability $P(X_r = \omega_i) = \pi_i$ initializes the procedure (cf. Fig.1).

1. **Bottom-up step on partial posterior marginals**

$$\begin{array}{l} \text{for } n = 1 \ldots R \\ \quad \left| \begin{array}{l} P(X_s = \omega_i | \boldsymbol{y}_{\geq s}) = \frac{1}{Z} P(X_s = \omega_i)^{-3} P_i^n(\boldsymbol{y}_s) \prod_{t \in s^+} \sum_j a_{ij} \frac{P(X_s = \omega_i)}{P(X_t = \omega_j)} P(X_t = \omega_j | \boldsymbol{y}_{\geq t}) \\ \forall s \in S^n \end{array} \right. \end{array}$$

where Z is a normalizing factor satisfying the condition $\sum_i P(X_s = \omega_i | \boldsymbol{y}_{\geq s}) = 1$. At the end of the recursion, we obtain the a posteriori probability for the root $\xi_r(i) = P(X_r = \omega_i | \boldsymbol{y})$, which initializes the next step.

2. **Top-down step on posterior marginals**

$$\begin{array}{l} \text{for } n = R-1 \ldots 0 \\ \quad \left| \begin{array}{lcll} \Psi_s(i,j) & = & P(X_s = \omega_j, X_{s^-} = \omega_i / \boldsymbol{y}) & \\ & = & \xi_{s^-}(i) \frac{P(X_s = \omega_i | \boldsymbol{y}_{>s}) a_{ij} P(X_{s^-} = \omega_i) / P(X_s = \omega_j)}{\sum_l P(X_s = \omega_l | \boldsymbol{y}_{>s}) a_{il} P(X_{s^-} = \omega_i) / P(X_s = \omega_l)} & \forall s \in S^n \\ \xi_s(j) & = & P(X_s = \omega_j | \boldsymbol{y}) = \sum_i \Psi_s(i,j) & \end{array} \right. \end{array}$$

(Image) class	Mean of each spectral band	Std. dev. of each spectral band
(b) e_0	120,120,120	16,16,16
(b) e_1	136,136,136	16,16,16
(c) e_0	124,124,128	16,16,16
(c) e_1	132,132,128	16,16,16
(d) e_0	128,128,128	12,12,24
(d) e_1	128,128,128	24,24,24

Table 1: Means and standard deviations of synthetic three-band images (Fig. 2).

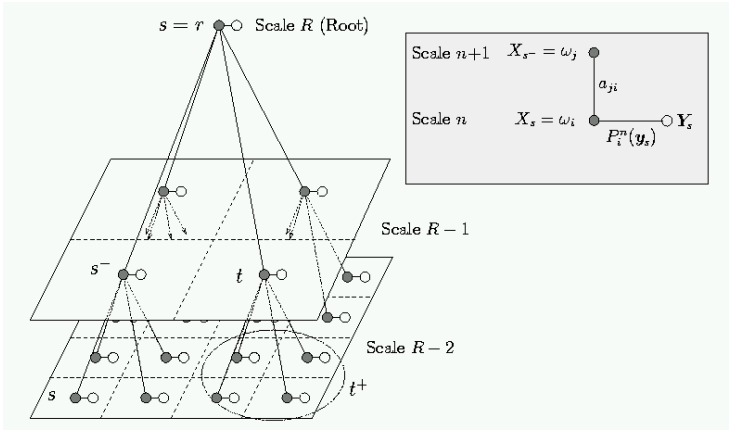
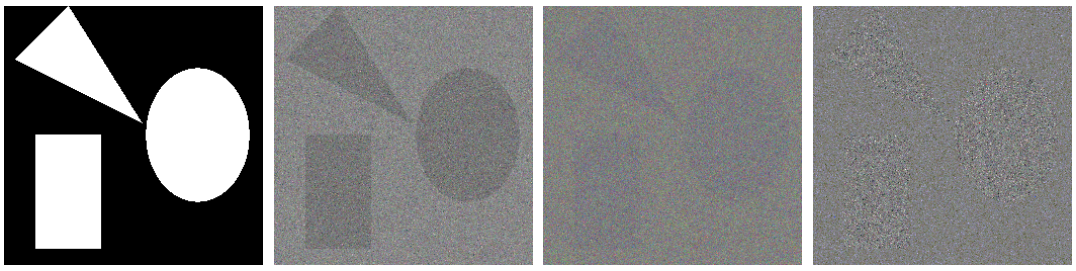


Figure 1: Dependency graph corresponding to a quadtree structure. Black circles represent labels and white circles represent observations.



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

Figure 2: (a) Original image, 2 classes (background and patterns). (b,c,d) Observed image in three spectral bands with mean-discriminant generalized Gaussian noise (b,c) and with variance-discriminant generalized Gaussian noise (d) (see section 4 for noise model and Table 1 for noise parameters in each spectral band). The segmentation of these images is reported on in Fig. 3.

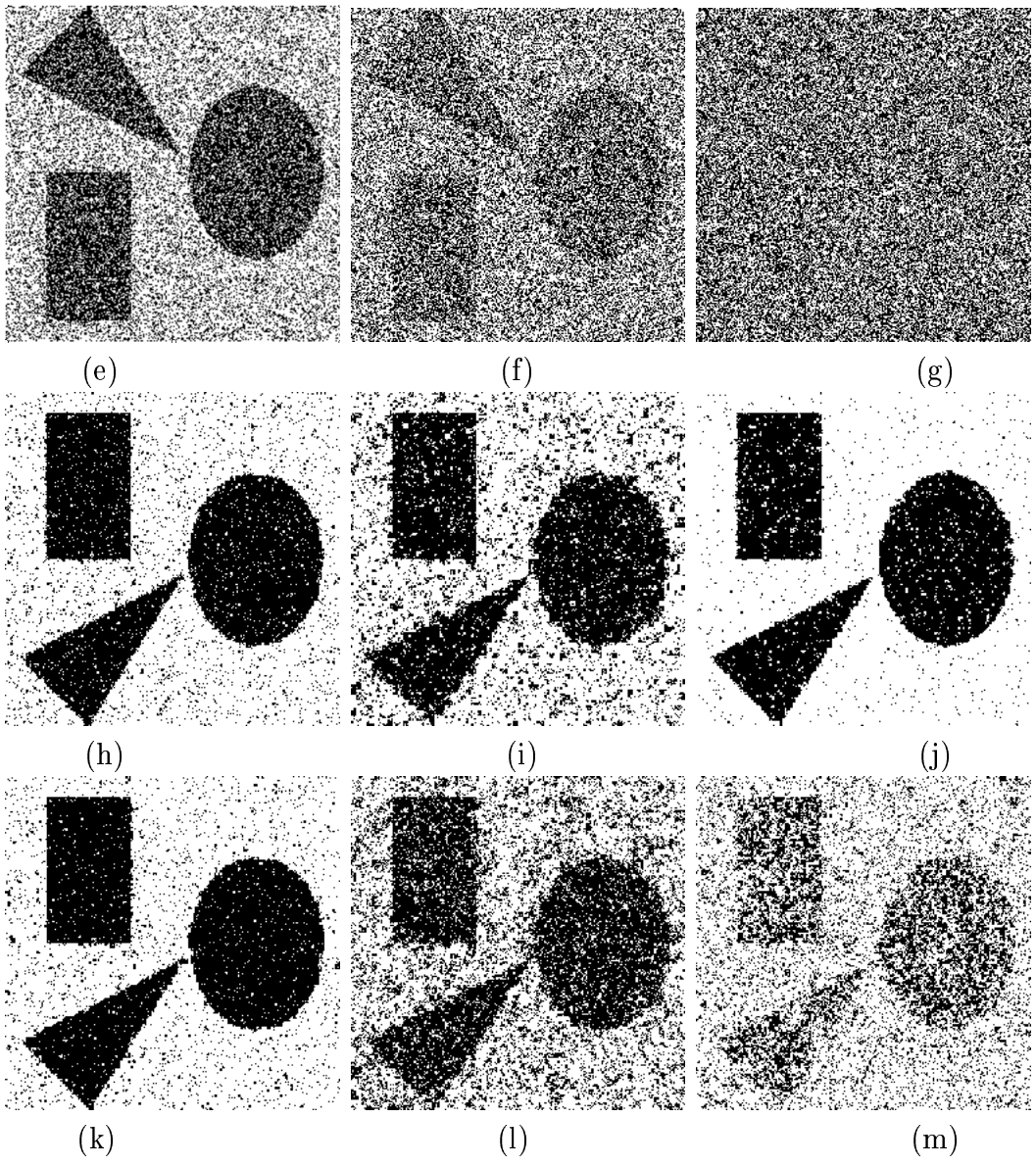


Figure 3: (e,f,g) Segmentation map based on maximum likelihood criterion for multiband images (Fig. 2 b,c,d) respectively. (h,i,j) Segmentation map obtained with Markovian-in-scale algorithm presented in section 3. (k,l,m) Segmentation map obtained on the larger eigenvalue image, following PCA of Fig. 2 b,c,d respectively, followed by a Markovian-in-scale segmentation.

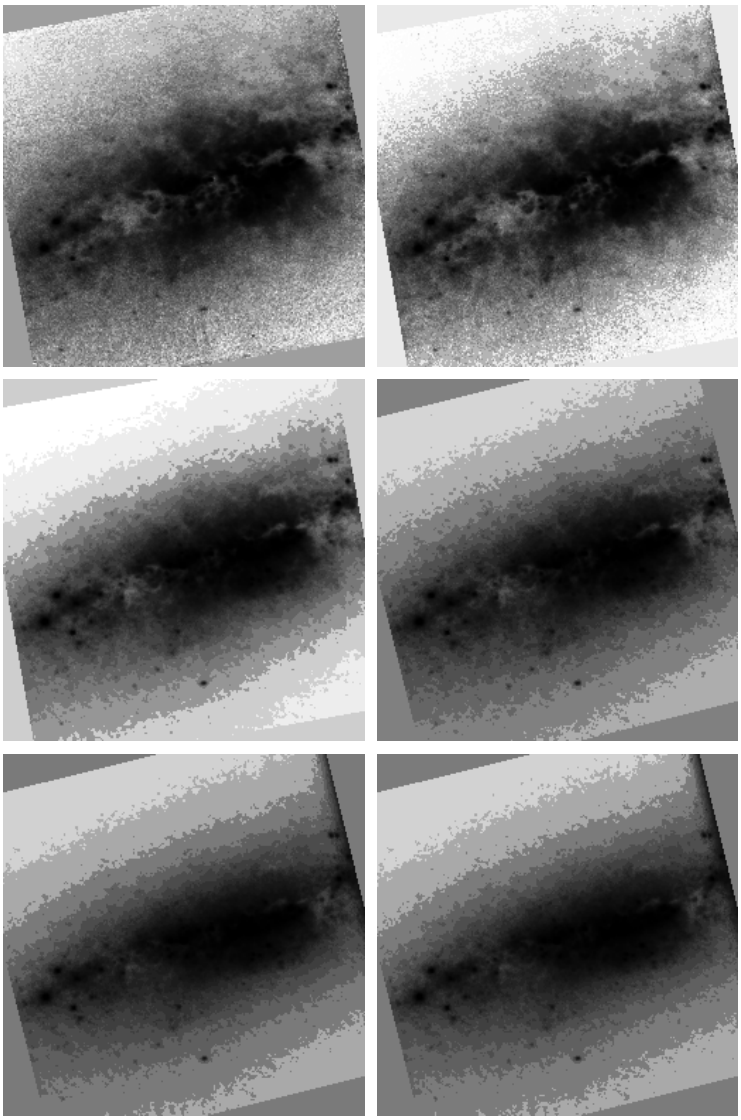


Figure 4: Astronomical image bands from the M82 starburst galaxy M82.

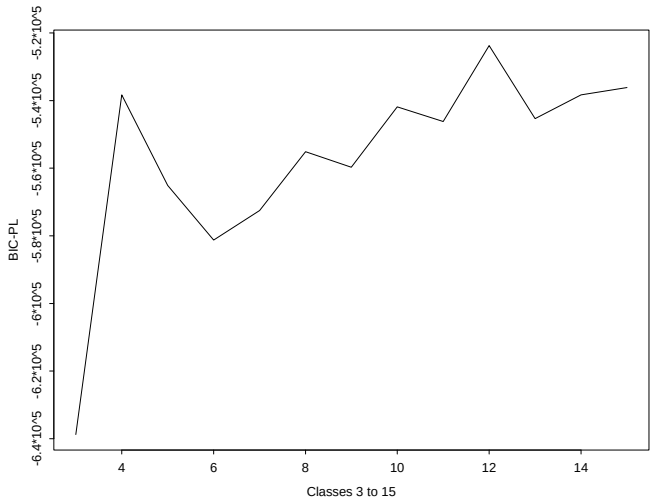


Figure 5: BIC_{PL} values for varying number of segments, for M82 multiband image. The peak value corresponding to 12 classes was used.

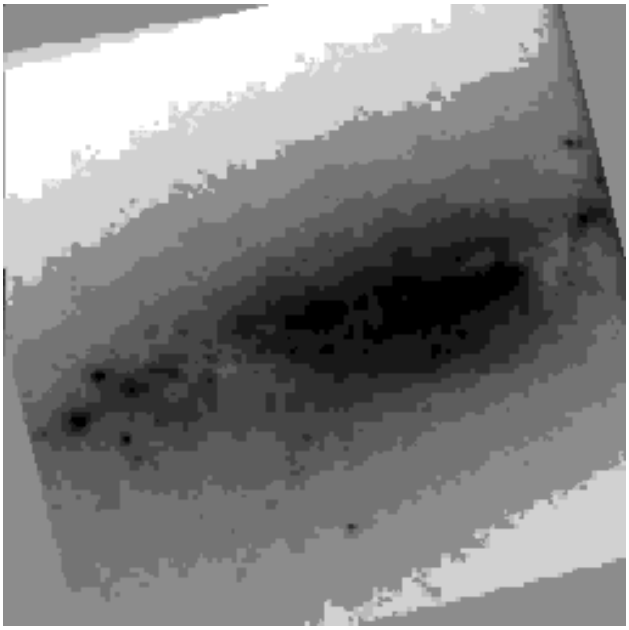


Figure 6: Hierarchical Markovian segmentation of astronomical 6-spectral band images from M82 region, into 12 classes. Reversed grayscale display used.

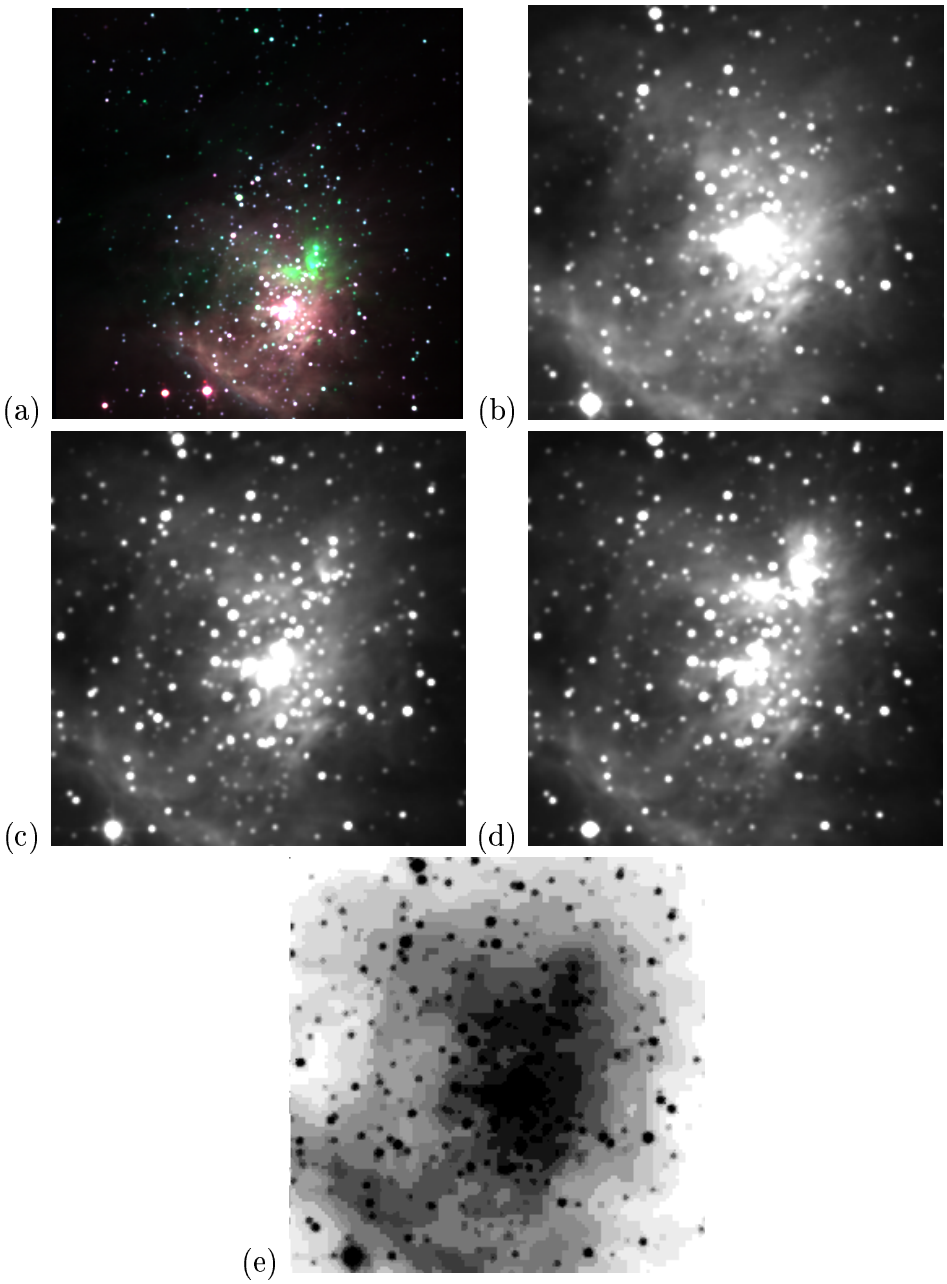


Figure 7: (a) Orion Nebula, observed in three spectral bands and displayed in red, green, blue colors. (b-d) Orion Nebula, of the 2MASS survey, displayed in graylevels for each band J, H, K (b,c,d). (e) Hierarchical Markovian segmentation into 14 classes of images of Orion Nebula, observed in three spectral bands (display uses reversed grayscale).

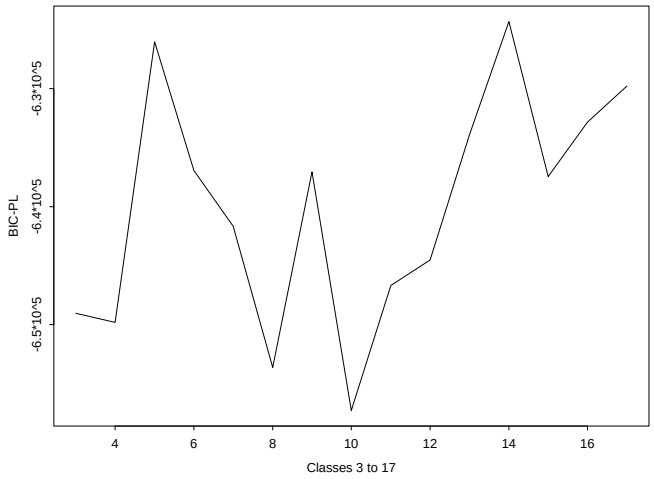


Figure 8: BIC_{PL} values for varying number of segments, for Orion multiband image. The peak value corresponding to 14 classes was used.

Biographies

Christophe Jean-Franois Collet was born in 1966 in France. He graduated from the Université Paris-Sud Orsay in 1989 (Master in Signal Processing, DEA) and received a PhD in Image Processing from the University of Toulon in 1992. His thesis research was carried out at the Ecole Nationale Supérieure des Télécommunications de Bretagne (ENST-Br) and at the French Naval Academy. He spent 8 years at the French Naval Academy and was the chairman of the laboratory GTS "Groupe de Traitement du Signal" from 1994 to 2000, where he developed hierarchical Markovian approaches for SONAR image segmentation. Since 2001 he holds a Full Professor position at Strasbourg University (LSIIT UMR CNRS 7005). His major research interests include multi-image segmentation and classification with hierarchical approaches (wavelet decomposition, multigrid optimization, multiscale modeling), Bayesian inference, Markovian approaches for pattern recognition, Bayesian networks, with a particular focus on astronomy (hyperspectral) and medical (multimodal) images.

Fionn Murtagh holds engineering science, mathematics and computer science degrees (BA, BAI, MSc) from Trinity College Dublin, a PhD in mathematical statistics from Université P. & M. Curie, Paris, and an Habilitation in computational astronomy from Université Louis Pasteur, Strasbourg. In the past he worked for the Space Science Department of the European Space Agency on the Hubble Space Telescope. He is currently Full Professor in Computer Science at Queen's University Belfast. Among recent books he is author with J.L. Starck (CEA, Saclay) and A. Bijoaui (Nice) of "Image and Data Analysis: The Multiscale Approach", Cambridge University Press, 1998, and is author with J.L. Starck of "Astronomical Image and Data Analysis", Springer-Verlag, 2002.