SIViP (2007) 1:191–201
DOI 10.1007/s11760-007-0010-y

ORIGINAL PAPER

Markovian regularization of latent-variable-models mixture for New multi-component image reduction/segmentation scheme

F. Flitti · Ch. Collet

Received: 15 October 2006 / Revised: 28 February 2007 / Accepted: 28 February 2007 / Published online: 11 April 2007

Abstract This paper proposes a new framework for multi-component images segmentation which plays an increasing role in many imagery applications like astronomy, medicine, remote sensing, chemistry, biology etc. In fact, inference on such images is a very difficult task when the number of components increases due to the well-known Hughes phenomenon. A common solution is to reduce dimensionality, keeping only relevant information before segmentation. Linear models usually fail with complex data structure, and mixture of linear models, each of which modeling a local cluster of the data, is more suitable. Moreover, a probabilistic formulation based on linear latent variable models allows efficient solution using a maximum-likelihood-based decision to recover the clusters. However, for multi-component image classification, this is not enough because it completely neglects the spatial positions of the multi-dimensional pixels on the lattice. Therefore, we propose to consider the neighborhood by introducing a Markovian *a priori* to efficiently regularize pixel classification. As a consequence, segmentation and reduction are performed simultaneously in an efficient and robust way. In this paper, we focus on the Probabilistic Principal Component Analysis (PPCA) as a latent variable model, and the Hidden Markov quad-Tree (HMT) as an *a priori* for regularization. The method performs well both on synthetic and real remote sensing and Stokes–Mueller images.

F. Flitti · Ch. Collet (✉)
LSIIT UMR 7005, Strasbourg University, Bd S. Brant,
BP 10413, 67412 Illkirch, France
e-mail: collet@lsiit.u-strasbg.fr

F. Flitti
Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology HKUST,
Kowloon, Hong Kong
e-mail: eeflitti@ust.hk

Keywords Bayesian approach · Hyperspectral images · Markovian quadtree · Image reduction · Mixture of probabilistic PCA

1 Introduction

Multi-component image segmentation represents a great challenge in many imagery applications in medicine, remote sensing, astronomy, non-destructive control, etc. Indeed, today's new systems and new sensor technologies allow acquisition of spatially resolved high-dimensional data with a huge quantity of information. The segmentation of high-dimensional multi-component (i.e., multi-wavelength, multi-modal, multi-variate) images remains a difficult task and is more complicated when observations are corrupted with noise. Segmentation algorithms learn data structure to gather pixels according to a given measure, under some smoothness constraints. They require sufficient observations to correctly estimate model parameters. For multi-component images the required number of samples grows quickly along with the dimension (i.e., number of components) so that the segmentation accuracy decreases rapidly in practice. This is the curse of dimensionality (Hughes phenomenon) [1] which consists of an important loss of parameter estimation accuracy as dimensionality grows. To address this problem, one may carry out a space dimensionality reduction as a preprocessing step [2]. Fortunately, high-dimensional observations can often be described in a significantly smaller dimension than the original due to redundancy and correlation between components. Thus, many approaches were proposed in the past decade, all of them seek a mapping onto a reduced dimensional space by maximizing various criteria [3–5]. Data with complex structures require non-linear mapping [6], and several works have been proposed to develop non-linear

models. One attractive way is to use a collection of locally linear models, so that each observation is modeled using either a single local model [7] or a mixture of all local models [6,8,9]. In the first case, observations are hardly partitioned into clusters (i.e., classes), each one spanned by a local linear model. In the second case, each observation is shared by all clusters. When local models are coupled with probabilistic modeling, the observation likelihood is given by a density mixture model [6,8,9]. Nevertheless, for D-components images, the clustering established by such mixture models is based only on the closeness of the observations (pixels) in the D-dimensional space and do not take into account their proximity in the image lattice. We propose to use a Markovian *a priori* associated with such model to regularize D-dimensional pixel classification. In this way, the pixel classification is based on their closeness both in the $\mathbb{R}^D$ space and the image lattice. Thus, segmentation and reduction are performed simultaneously avoiding the Hughes phenomenon. In this work, we focus on the Probabilistic Principal Component Analysis (PPCA) as a latent variable model, and the Hidden Markov quad-Tree (HMT) as a Markovian *a priori*. However, our method may be extended to other latent variable models like factor analysis and independent component analysis and other Markov models such as Hidden Markov Chain and Markov Random Field. Our technique was applied on synthetic and real remote sensing and Mueller data. The observed results were very promising.

The paper is organized as follows. The Hidden Markov quad-Tree (HMT) model is described in Sect. 2. In Sect. 3, the PPCA and the Mixture of Probabilistic Principal Component Analyzers (MPPCA) are presented and linked to the HMT. In Sect. 4, the advantages of the method are discussed and thoroughly analyzed. Section 5 presents the experiments on synthetic and real multi-component images. Finally, a conclusion outlining possible extensions of the method is presented in Sect. 6.

2 Hidden Markov quad-Tree

In the past decade, Hidden Markov models have proved to be robust and efficient image analysis methods for many detection, denoising, segmentation, classification and pattern recognition tasks. Nowadays, in the multi-component image context, handling correlated and redundant observed data requires particular modeling framework, and poses again the problem of constructing relevant models. Resorting to a Bayesian scheme based on Markov models is indeed attractive when dealing with large amount of multi-component observations, because of their parsimonious properties. Nevertheless, the well-known Markov Random Fields (MRF) leads to iterated optimization algorithms (due to the fact that most of Markov models are non-causal) with high computation cost [10], even for one-dimensional image. As a consequence, the exact inference is not computable and has to be iteratively approximated, which might turn prohibitively expensive. Although some multi-grid strategies to decrease the computation time have been proposed in the past decade [11–14], the processing of high-dimensional data remains problematic. One way to circumvent this problem is to resort to a Markov model on a chain [15] or a quadtree where in-scale causality allows non-iterative inference [16,17] as in the case of hidden Markov chains [15].

A Hidden Markov quad-Tree (HMT) is an acyclic graph $G = (S, L)$ with a set of nodes S and a set of edges L. S is partitioned into "scales", i.e.; $S = S^0 \cup S^1 \ldots \cup S^R$, such that $S^R = \{r\}$ is the root, S^n involves 4^{R-n} nodes, and S^0 is the finest scale formed by the leaves. Each node s, except the root r, has a unique predecessor, its "parent" s^-. Each node s, expect the "leaves", has four "children" $s^+ = \{u \in S : u^- = s\}$. We note also s^{++} all the descendent of s.

Let the hidden process[1] X which assigns to each node $s \in S$ a hidden state X_s chosen from the label set $\Omega = \{\omega_1, \ldots, \omega_K\}$ of the K classes. X is assumed Markovian in scale, i.e., :

$$P(x^n|x^k, k > n) = P(x^n|x^{n+1}); \quad x^n = \{x_s : s \in S^n\}. \tag{1}$$

Moreover, $X_s, s \in S^n$, is independent from all $X_u, u \in S^{n+1}$, given its parent and the inter-scale transition probability can be factorized in the following way [16]:

$$P(x^n|x^{n+1}) = \prod_{s \in S^n} P(x_s|x_{s^-}) \tag{2}$$

The hidden process X is called Markov tree because it verifies [16]:

$$P(x) = P(x_r) \prod_{n=0}^{R-1} \prod_{s \in S^n} P(x_s/x_{s^-}) \tag{3}$$

The multi-component observations $\boldsymbol{Y}$ are introduced at the scale S^0 so that each D-dimensional pixel $\mathbf{y}_s$ is linked to the hidden state X_s (Fig. 1). The HMT assumes $\mathbf{y}_s$ independent from all the quad-Tree given its hidden state, which is formulated as follows :

$$P(\mathbf{y}_s/x, \mathbf{y} - \{\mathbf{y}_s\}) = P(\mathbf{y}_s/x_s). \tag{4}$$

Thus the probability of $\boldsymbol{Y}$ conditionally to X is expressed as the following product:

$$P(\mathbf{y}|x) = \prod_{s \in S^0} P(\mathbf{y}_s|x_s), \tag{5}$$

where $\forall\ s \in S^0$, $P(\mathbf{y}_s|x_s = \omega_i)$, called data-driven term, captures the likelihood of the observation $\mathbf{y}_s$ w.r.t the class

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

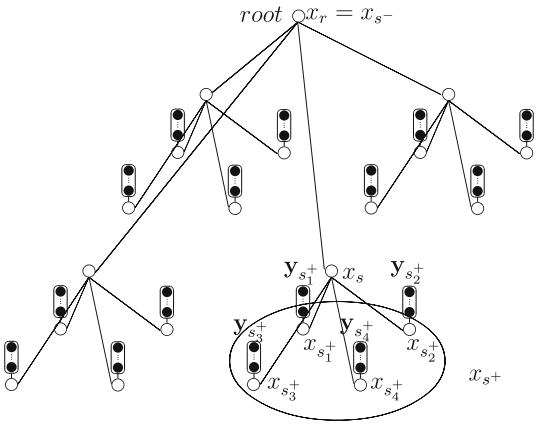

Fig. 1 Example of a dependency graph corresponding to a quadtree structure on a 4×4 lattice. *White circles* represent labels and *black circles* represent multi-band observations $\mathbf{y}_s$, $s \in S$. Each node s has a unique *parent* s^-, and four *children* $s^+ = \{s_1^+, \ldots, s_4^+\}$

ω_i. If no data is available at a given site s in the image, i.e., missing or masked data, the likelihood at this site is set to 1. In Sect. 3, we link MPPCA to the HMT so that each class is spanned by a local PPCA model. Thus $P(\mathbf{y}_s|x_s = \omega_i)$ is computed as the likelihood of $\mathbf{y}_s$ w.r.t the local PPCA modeling of the class ω_i.

From the assumptions above, the joint distribution $P(\mathbf{x}, \mathbf{y})$ can easily factorized as follows :

$$P(\mathbf{x}, \mathbf{y}) = P(x_r) \prod_{s \neq r} P(x_s|x_{s^-}) \prod_{s \in S^0} P(\mathbf{y}_s|x_s). \tag{6}$$

The HMT parameters are :

- Φ_x the *a priori* parameters regrouping :
  - $\{\pi_i = p(x_r = \omega_i)\}_{i=1,\ldots,K}$ the probabilies at the root,
  - $\{a_{ij} = p(x_s = \omega_j / x_{s^-} = \omega_i)\}_{i,j=1,\ldots,K}$ the parent/child transition probabilities,
- Φ_y the parameters of the likelihoods $\{P(.|x_s{=}\omega_i)\}_{i=1,\ldots,K}$.

One of the interests of this model is 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 in two passes on the quad-Tree (Algorithm 1).

The EM algorithm used for the estimation of the *a priori* parameters Φ_x, leads to an iterative procedure with the followings updates [16] :

$$\begin{aligned} a_{ij}^{[c+1]} &= \frac{\sum_{s \in S^n,\, n \neq r} p^{[c]}(x_s = \omega_j, x_{s^-} = \omega_i/\mathbf{y})}{\sum_{s \in S^n,\, n \neq r} p^{[c]}(s_{s^-} = \omega_i/\mathbf{y})} \\ \pi_i^{[c+1]} &= p^{[c]}(x_r = \omega_i/\mathbf{y}) \end{aligned} \tag{7}$$

where $[c]$ stands for the current iteration and $p^{[c]}(x_s = \omega_i/\mathbf{y})$ and $p^{[c]}(x_s = \omega_j, x_{s^-} = \omega_i/\mathbf{y})$ are computed by way of the two passes of Algorithm 1 using the current parameters. These parameters are initialized as mentioned in Algorithm 2.

When converged, i.e., the difference between successive updates is small enough or the maximum of number of iteration is reached, the Marginal *a Posteriori* Mode criterion (MPM) is used to obtain the segmentation map :

$$\forall\, s \in S^0,\ \hat{x}_s = \arg\max_{x_s \in \Omega} p(x_s/\mathbf{y}) \tag{8}$$

The estimation of likelihood parameters Φ_y in the case of regularized MPPCA is presented in the next section.

3 Regularization of the Mixture of Probabilistic Principal Component Analyzers (MPPCA)

3.1 Probabilistic Principal Component analysis (PPCA)

The PPCA is a statistical modeling of the well-known PCA, introduced by Tipping and Bishop [18]. It is based on a latent variable model which links each $D \times 1$ observed vector $\mathbf{y}_s$ to $q \times 1$ latent vector $\mathbf{t}_s$, $q < D$, as follows:

$$\mathbf{y}_s = A\mathbf{t}_s + \mu + \epsilon \tag{9}$$

where A is a $D \times q$ matrix, μ the observed data mean and ϵ is an isotropic Gaussian noise, i.e., $\mathcal{N}(0, \sigma^2 I)$, I being the $D \times D$ identity matrix.

Thus, the probability distribution of $\mathbf{y}_s$ given $\mathbf{t}_s$ is :

$$P(\mathbf{y}_s/\mathbf{t}_s) = \mathcal{N}(\mathbf{y}_s; A\mathbf{t}_s + \mu, \sigma^2 I) \tag{10}$$

Choosing Gaussian *prior* for $\mathbf{t}_s$, i.e., :

$$p(\mathbf{t}_s) = (2\pi)^{\frac{-d}{2}} \exp\left\{-\frac{1}{2}\mathbf{t_s}^t \mathbf{t_s}\right\} \tag{11}$$

the marginal distribution of $\mathbf{y}_s$ is:

$$P(\mathbf{y}_s) = \mathcal{N}(\mathbf{y}_s; \mu, C) \tag{12}$$

with $C = \sigma^2 I + AA^t$ is a $D \times D$ matrix [6]. Bayes rule gives the *a posteriori* probability of $\mathbf{t}_s$ [6]:

$$p(\mathbf{t}_s/\mathbf{y}_s) = \mathcal{N}(\mathbf{t}_s; M^{-1}A^t(\mathbf{y}_s - \mu), M^{-1}) \tag{13}$$

where $M = \sigma^2 I - A^t A$ is a $q \times q$ matrix.

The maximization of the data log-likelihood $\mathcal{L} = \sum_{s \in S^0} \ln\{p(\mathbf{y}_s)\}$ gives the following parameter estimators [6]:

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{s \in S^0} \mathbf{y}_s}{\mathrm{card}(S^0)} \\ \hat{\sigma}^2 &= \frac{1}{D-q} \sum_{j=q+1}^{D} \lambda_j \\ \hat{A} &= U_q(\Lambda_q - \sigma^2 I)^{\frac{1}{2}} R. \end{aligned} \tag{14}$$

where λ_j are the eigenvalues of the data covariance matrix $\Sigma = \frac{1}{\mathrm{card}(S^0)} \sum_{s \in S^0} (\mathbf{y}_s - \mu)(\mathbf{y}_s - \mu)^t$ given in descending order ($\lambda_1 \geq \cdots \geq \lambda_q$), Λ_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.

Algorithm 1 Two passes on the quadtree for *posterior* computation given HMT parameters $\{\Phi_x, \Phi_y\}$.

• Evaluation of the partial posterior marginals at the bottom of the quadtree :

$$\forall s \in S^0,\ P(x_s = \omega_i/\mathbf{y}_{s^{++}}) = P(x_s = \omega_i/\mathbf{y}_s) = \frac{P(x_s = \omega_i)P(\mathbf{y}_s/x_s = \omega_i)}{\sum_{\omega_j} P(x_s = \omega_j)P(\mathbf{y}_s/x_s = \omega_j)},$$

where $P(x_s = \omega_i)$ is recursively evaluated through a top-down pass, given the *prior* probability $P(x_r = \omega_i) = \pi_i$ as follows :
for $n = R-1, \cdots, 0$ **do**
  for all $s \in S^n$ **do**
    $P(x_s = \omega_i) = \sum_{\omega_j} P(x_s = \omega_i/x_{s^-} = \omega_j)P(x_{s^-} = \omega_j)$
  end for
end for.
• Upward pass :
for $n = 1, \cdots, R$ **do**
  for all $s \in S^n$ **do**
    $P(x_s = \omega_i/\mathbf{y}_{s^{++}}) = \frac{1}{Z} P(x_s = \omega_i) \prod_{t \in s^+} \sum_{\omega_j} \frac{a_{ij} P(x_t = \omega_j/\mathbf{y}_{t^{++}})}{P(x_t = \omega_j)}$
  end for
end for
where $a_{ij} = P(x_t = \omega_j/x_{t^-} = \omega_i)$ is the parent/child transition probability
and Z is a normalizing factor such that $\sum_{\omega_i} P(x_s = \omega_i/\mathbf{y}_{s^{++}}) = 1$. Note that at the top of quadtree we obtain $P(x_r = \omega_i/\mathbf{y})$
• Downward pass :
for $n = R-1, \cdots, 0$ **do**
  for all $s \in S^n$ **do**

$$P(x_s = \omega_j, x_{s^-} = \omega_i/\mathbf{y}) = P(x_{s^-} = \omega_i/\mathbf{y}) \frac{P(x_s = \omega_j/\mathbf{y}_{s^{++}}) a_{ij} P(x_{s^-} = \omega_i)/P(x_s = \omega_j)}{\sum_{\omega_l} P(x_s = \omega_l/\mathbf{y}_{s^{++}}) a_{il} P(x_{s^-} = \omega_i)/P(x_s = \omega_l)},$$

$$P(x_s = \omega_j/\mathbf{y}) = \sum_{\omega_i} P(x_s = \omega_j, x_{s^-} = \omega_i/\mathbf{y})$$

  end for
end for

The sum $\sum_{j=q+1}^{D} \lambda_j$ represents the squared error of the approximation of D-dimensional vector $\mathbf{y}_s$ by $\mathbf{t}_s$ on the q-dimensional space, $q \leq D$. Thus, it can be efficiently used to estimate the local dimension q.

3.2 Mixture of Probabilistic Principal Component Analyzers (MPPCA)

The main difficulty with PPCA (or PCA) algorithm is that it considers the observed data distribution as a multi-variate Gaussian (Eq. 12) which masks all local structures present in the data. Taking into account this drawback, Tipping and Bishop [6] introduced the mixture of Probabilistic Principal Component Analyzers (MPPCA) to model complex data structures as a mixture of local PPCA.

For a K component MPPCA, the observations are shared by K clusters (i.e., classes) each one spanned by a local PPCA. Given this model, the distribution of the observations is

$$P(\mathbf{y}_s) = \sum_{i=1}^{K} \Pi_i P(\mathbf{y}_s/x_s = w_i) \tag{15}$$

where the local PPCA corresponding to the classe ω_i is characterized by the mean μ_i, the variance σ_i^2, the projection matrix A_i and the prior Π_i.

Note that in this formulation the *prior* is the same for all $s \in S^0$ and thus no information about the neighborhood is taken into account when classifying $\mathbf{y}_s$, whereas the data are organized on a regular lattice. We think that taking into account the data topology during the segmentation step will be of great importance.

The *a posteriori* responsibility of the component i for generating the vector $\mathbf{y}_s$ is given by :

$$\Upsilon_{si} = P(x_s = w_i/\mathbf{y}_s) = \frac{P(\mathbf{y}_s/x_s = w_i)\Pi_i}{P(\mathbf{y}_s)} \tag{16}$$

The EM algorithm is used to iteratively estimate the mixture parameters [6] :

$$\hat{\Pi}_i = \frac{1}{\text{card}(S^0)} \sum_{s \in S^0} \Upsilon_{si} \tag{17}$$

$$\hat{\mu_i} = \frac{\sum_{s \in S^0} \Upsilon_{si} \mathbf{y}_s}{\sum_{s \in S^0} \Upsilon_{si}} \tag{18}$$

and $\hat{A}_i$ et $\hat{\sigma}_i^2$ are given, in the same way of Eq. 14, by eigen decomposition of the *a posteriori* responsibility-weighted covariance matrix:

$$\Sigma_i = \frac{\sum_{s \in S^0} \Upsilon_{si} (\mathbf{y}_s - \hat{\mu_i})(\mathbf{y}_n - \hat{\mu_i})^t}{\sum_{s \in S^0} \Upsilon_{si}} \tag{19}$$

The MPPCA is a powerful well-formulated tool to capture data local structures in the D-dimensional space, even if the dimensionality reaches large values. However, its use in multi-component images segmentation gives no smooth maps because no information about observation location in the image is taken into account in this modeling : one infers each data in a blind way. In the next subsection we propose a new method linking the MPPCA advantages and HMT regularization properties to segment high-dimensional data cube.

3.3 Regularized mixture of Probabilistic Principal Component Analyzers

We integrate MPPCA model by imposing a Markov constraints via the quad-tree modeling. The observation probability becomes :

$$P(\mathbf{y}_s) = \sum_{i=1}^{K} P(x_s = \omega_i) P(\mathbf{y}_s / x_s = \omega_i) \tag{20}$$

where X is a Markov tree (Eq. 3) and each class w_i is spanned by a local PPCA. In this way the likelihood $P(\mathbf{y}_s/x_s = \omega_i)$ is computed as the likelihood of $\mathbf{y}_s$ w.r.t. the local PPCA corresponding to the class ω_i. As Eq. 12 suggests, this likelihood is given by:

$$P(\mathbf{y}_s/x_s = \omega_i) = \mathcal{N}(\mathbf{y}_s; \mu_i, C_i) \tag{21}$$

The matrix C_i is obtained in analog manner to Eq. 14 by eigen-decomposition of the weighted covariance matrix

$$\Sigma_i = \frac{\sum_{s \in S^0} P(x_s = \omega_i/\mathbf{y})(\mathbf{y}_s - \hat{\mu}_i)(\mathbf{y}_s - \hat{\mu}_i)^t}{\sum_{s \in S^0} P(x_s = \omega_i/\mathbf{y})} \tag{22}$$

where

$$\hat{\mu}_i = \frac{\sum_{s \in S^0} P(x_s = \omega_i/\mathbf{y})\mathbf{y}_s}{\sum_{s \in S^0} P(x_s = \omega_i/\mathbf{y})} \tag{23}$$

The estimation of the *a priori* parameter remains the same as in the classical quad-tree Eq. 7. The whole algorithm is given in Algorithm 2.

Algorithm 2 Regularized MPPCA procedure

- Computation of an initial segmentation map, using K-means algorithm for example, and estimation of the means and variances within each class ($\Phi_y^{[0]}$).
- Initialization of the prior parameters $\Phi_x^{[0]}$:

$$\begin{cases} \pi_i = 1/K, \ i = 1, \cdots, K \\ a_{ii} = 3/4 \ and \ a_{ij} = 1/4(K-1); \ i, j = 1, \cdots, K \end{cases}$$

- c=1

repeat

- Perform a two passes on the quadtree (Algorithm 1) using $\{\Phi_x^{[c]}, \Phi_y^{[c]}\}$.
- Compute $\{\Phi_x^{[c+1]}, \Phi_y^{[c+1]}\}$: Eq.7, Eq. 23, Eq. 22 and Eq. 14.
- c=c+1

until the difference between $\{\Phi_x^{[c]}, \Phi_y^{[c]}\}$ and $\{\Phi_x^{[c+1]}, \Phi_y^{[c+1]}\}$ is small enough or the maximum number of iterations is reached

- Obtain the segmentation map using Eq. 8.

4 Comparison with the multi-variate Gaussian

Actually, Regularized MPPCA model is equivalent to use multi-variate Gaussian for the data driven term (Eq. 21). However, it brings two main advantages :

1. Less parameters for the covariance matrix : the covariance matrix of a D- dimensional multivariate Gaussian exhibits $D(D+1)/2$ independent parameters, while the covariance matrix $C_i = \sigma_i^2 I + A_i A_i^t$ (Eq. 21) presents only $D\,q_i + 1 - q_i(q_i - 1)/2$ independent parameters, where q_i is the reduced dimensionality of the local PPCA and $q_i(q_i - 1)/2$ corresponds the number of parameters required to specify the rotation R_i associated to A_i (Eq. 14). Thus, the number of parameters depends on q_i : a general Gaussian is recovered for $q_i = D - 1$ and isotropic Gaussian for $q_i = 0$ [6].
2. Additional informations : the Regularized MPPCA exhibits two kind of latent variables which are the label process X and the hidden vectors $\mathbf{t}_s^i$, $i = 1, \ldots, K$, corresponding to the projection of the observations $\mathbf{Y}_s$ using the local PPCAs (Eq. 9). Thus, we may obtain a set of $N = \sum_{i=1}^{K} q_i$ representative images instead of the D original ones where the parameter q_i equals to the number of dominant eigen values of the ith local PPCA. As a consequence, when the original images are multi-spectral ones, the eigen vectors of the PPCA associated with a given class, are eigen spectra which represent a signature of this class. This is very useful when the classes correspond to physical structures like astronomical objects, lands or water.

5 Experiments

Synthetic images

To test our approach, we generated three sets of three **256** × **256** correlated images each, with constant correlation ratio $\rho = 0.8$. Each image contains two Gaussian classes representing a geometric shape and background as detailed

Table 1 Parameters of the two Gaussian classes

	Class 1		Class 2	
	μ_1	σ_1	μ_2	σ_2
Band 1	7.5	1	8.5	1
Band 2	7.5	1	8.5	1
Band 3	8	1	8	1

The third band is completely corrupted with the noise

in Table 1. Thus, we obtain 9 images to segment (Fig. 2, top). The segmentation maps with four classes using the maximum likelihood classifier based on MPPCA and regularized MPPCA are shown in the bottom of Fig. 2. We choose two for the reduced dimensionality for each local PPCA for both MPPCA and our technique. It is very easy to note the effect of the Markov regularization (on the right) compared to the maximum likelihood classification (on the left). More than 98% of the pixels are correctly classified.

We also observe that the contours of the rectangle form are very good exhibited in the segmentation map (Fig. 2, bottom right). This is due to the well-known bloc effect drawback of the Hidden Markov Tree (HMT). By construction (each node has four children), it favorites the rectangular forms. Many solutions have been proposed to overcome this problem [19–21]. However, in our case comparing the obtained performances and additional complexity induced by the solutions, we found that the use of the HMT is a good compromise. If for a specific application the performance decreases, the a priori may be changed by a more appropriate one like Hidden Markov Chain or MRF.

Remote sensing images

We also applied our algorithm on two sets of real remote sensing images of Hartheim area (near Strasbourg city), each one of 16 bands and of size $\mathbf{256} \times \mathbf{256}$ (wavelengths between 477 and 2,425 nm), shown in Figs. 3 and 4. The images are obtained from the hyperspectral HyMap sensor, resolution at soil 4m. Each set of images includes one strongly noised band, number 9. We perform 10-class segmentation using regularized MPPCA. The results are presented at the bottom of Figs. 3 and 4. In the map of Fig. 3, we clearly distinguish the Rhine river on the right top, the road in the bottom left

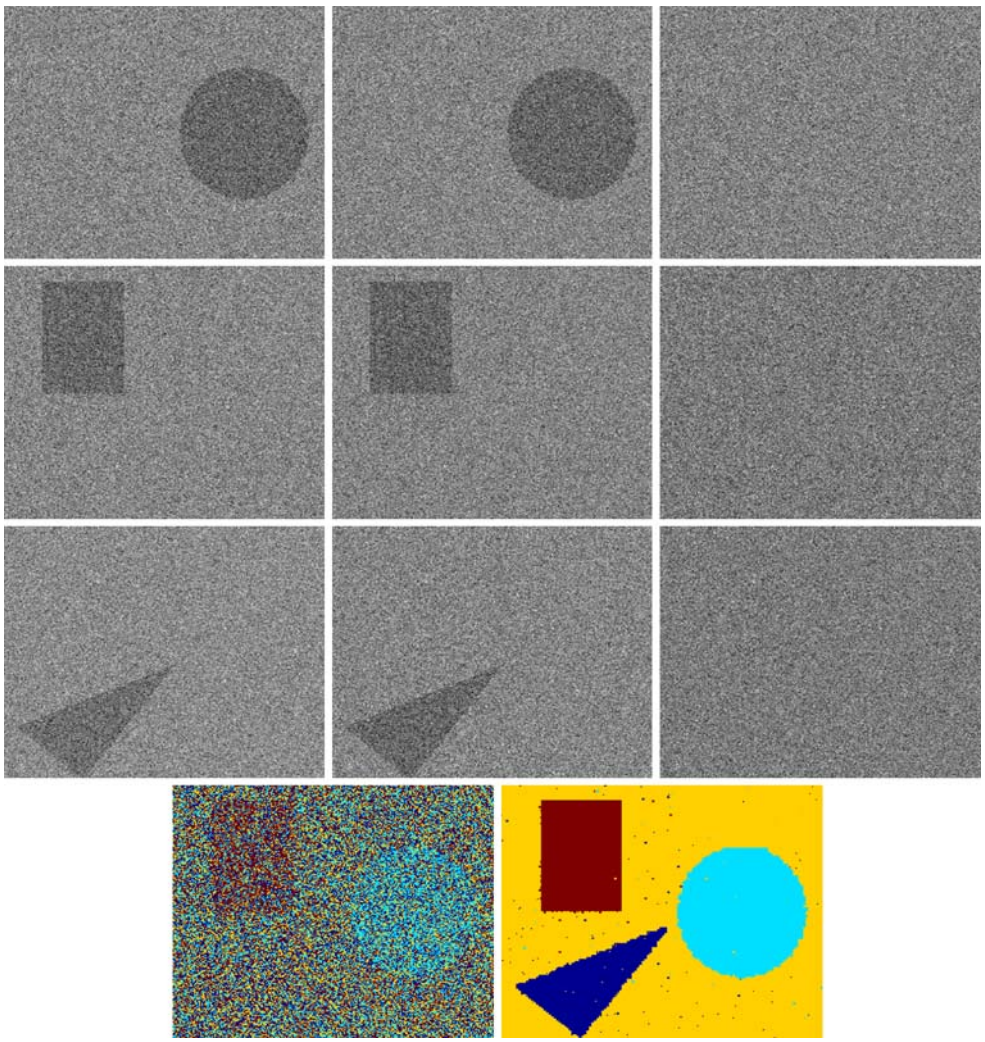

Fig. 2 Synthetic images of size $\mathbf{256} \times \mathbf{256}$ on the *top*. Each 3 images of the *same line* are simulated using the parameter of Table 1. On the *bottom*, segmentation map obtained with the MPPCA (*left*) is very noisy, whereas the map obtained with the proposed technique (*right*) is well regularized (1.2 % of misclassification)

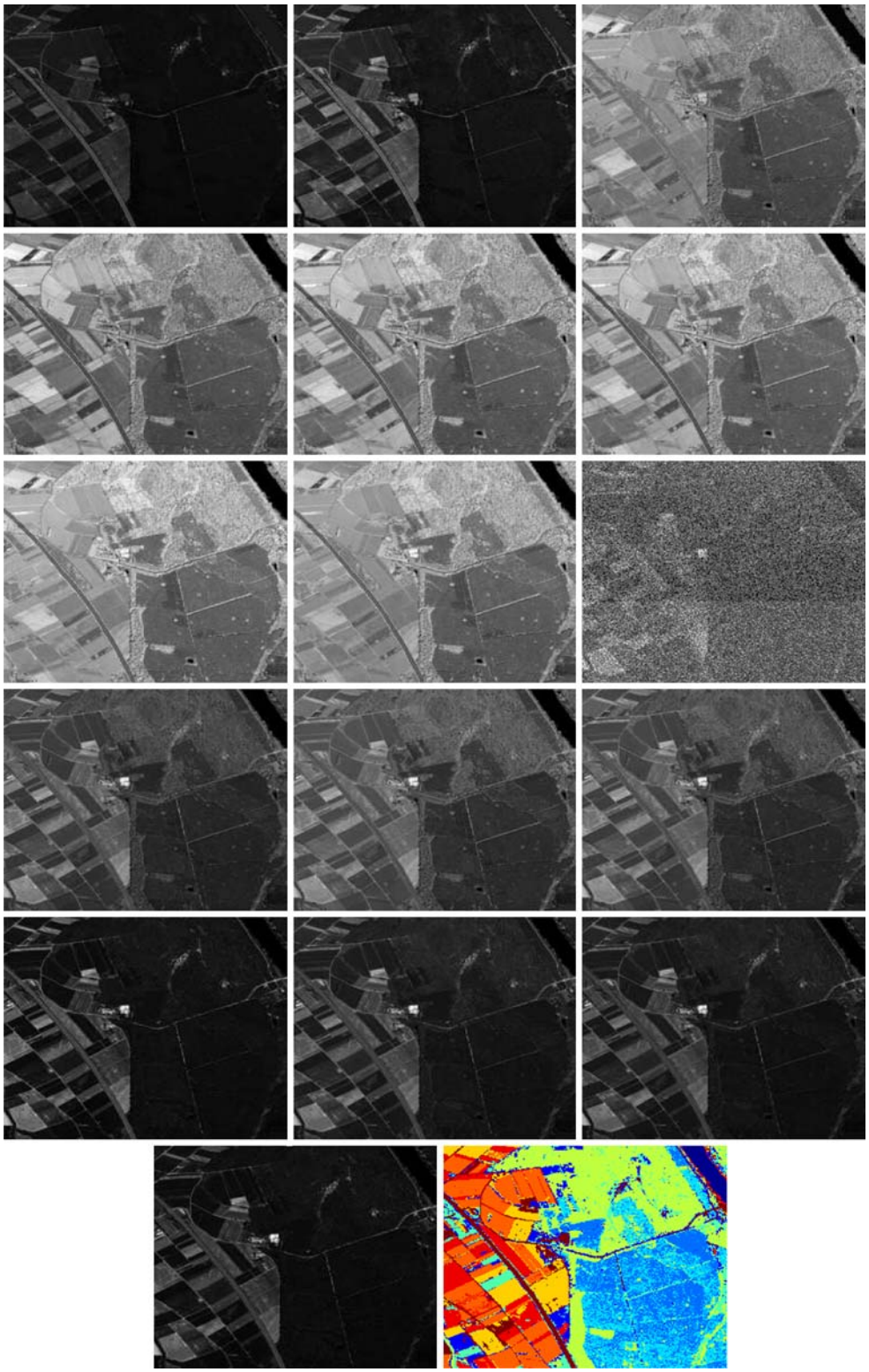

Fig. 3 Segmentation on 10 classes of remote sensing images using regularized MPPCA : 16 bands of size **256** × **256** of Hartheim area (near Strasbourg city), with wavelengths between 477 and 2,425 nm

corner, the different crop fields on the left side and vegetation on the right side. Similar remarks may be pointed for the map of Fig. 4. We distinguish easily the central road, the vegetation on the right side, the crop fields on the left side and few ponds, especially on the top and the bottom right. In both simulations, we choose 3 as reduced dimensionality for each

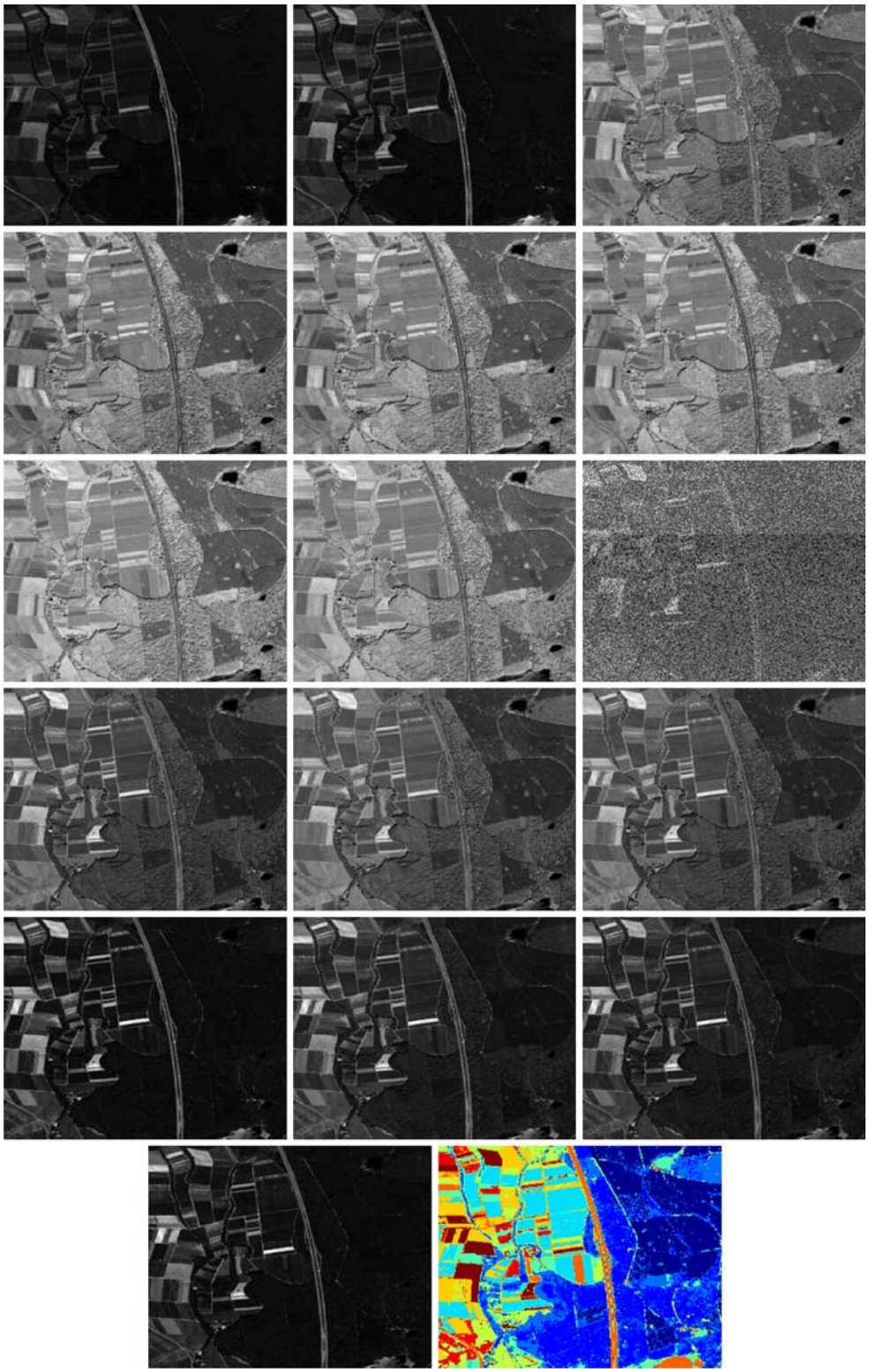

Fig. 4 Segmentation on 10 classes of remote sensing images using regularized MPPCA : 16 bands of size **256** × **256** of Hartheim area (near Strasbourg city), with wavelengths between 477 and 2,425 nm

local PPCA. Therefore, only 47 independent parameters are needed for the covariance matrix of each class instead of 136 when the classical multi-variate Gaussian density is used. This significant reduction is the success key to the method to transcend Hugh's phenomenon. The computation time, using non-optimized matlab code implementation and executed on

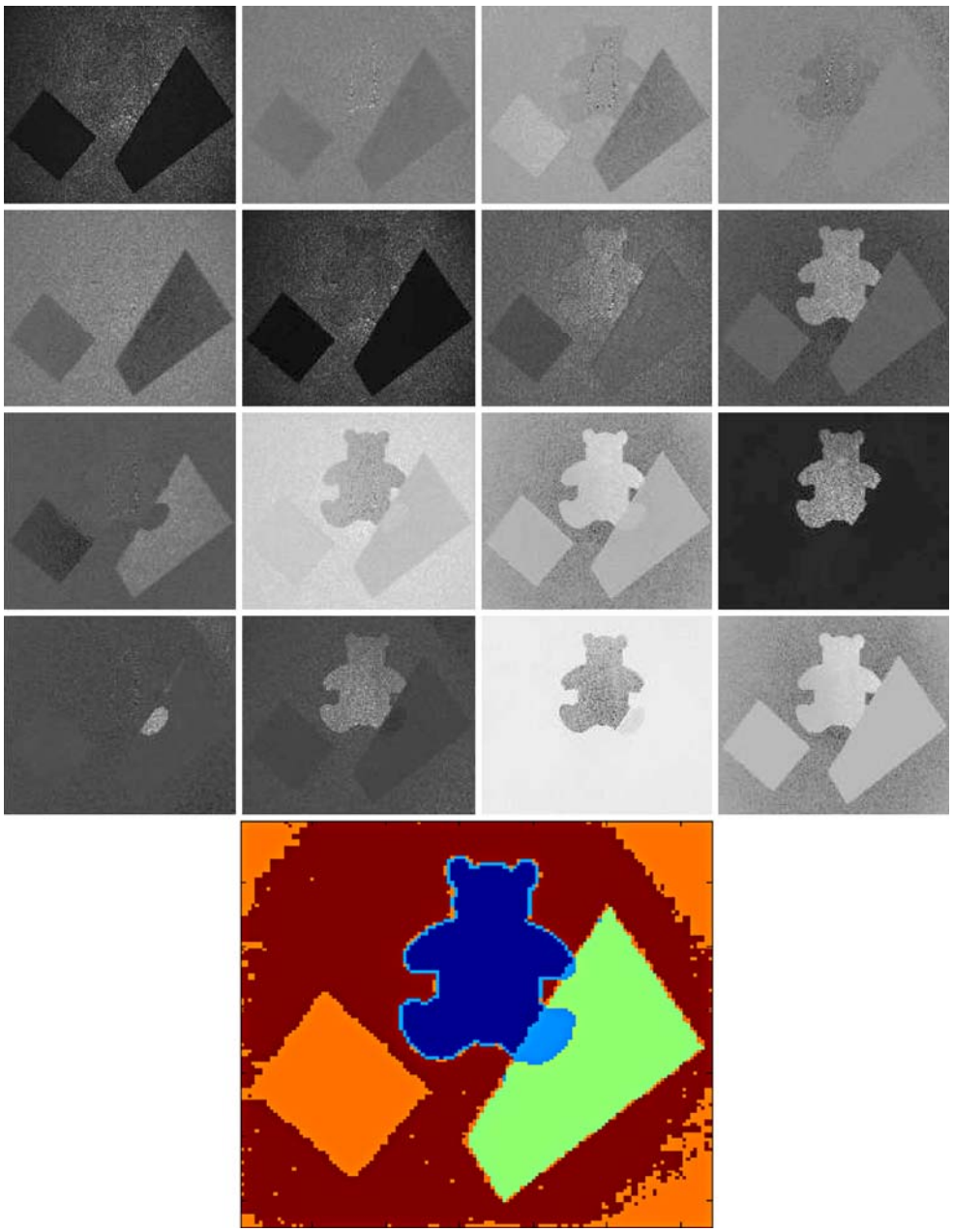

Fig. 5 Label map obtained with MPPCA-Markovian quadtree algorithm applied simultaneously on the 16 Mueller images. Five classes are obtained, each of them in a specific color. One observes that in spite of strong noise in the different channels, the regularized MPPCA algorithm manages correctly the various objects within the scene. The problem of analyzing polarization-encoded images and to explore the potential of this information for classification issues is now under study

a bi-processor Pentium 4 with 1.4 GHz frequency and 3 GB memory, is less then 2 mn for each iteration. A few iterations are needed to convergence (approximatively between three and six, therefore computation time is less then 15 mn for data cube of size $\mathbf{256} \times \mathbf{256} \times 16$ wavelength and 10 classes.

Stokes–Mueller imagery

The method proposed in this paper can be used in many other applications where multi-component (and not only multi-wavelength) images are available. For example, we tested our approach on another framework, in optical imaging, where the Stokes–Mueller formalism allows the definition of light polarization parameters in terms of real quadratic observables (intensities) which are directly sensed by CCD detectors. This allows extending classical intensity-wise imaging systems to acquire polarization parameters images through the use of Polarization State Modulators. Although many such systems using different polarization modulation techniques have been built in recent years for many application

areas ranging from metrology to biomedical imaging, remote sensing the analysis of Stokes–Mueller multi-component images remains difficult [22]. Nevertheless, the relevance of polarization-encoded images that comes from the rich set of physical information they carry about the local nature of the target incited us to test this new method, taking into account both spatial regularization and high-dimensional data in the segmentation process [22]. Hence, polarization-sensitive imaging systems are emerging as a very attractive vision technique that provides insightful understanding of the elements that constitute the object based on their polarimetric properties, i.e., birefringence, dichroism, depolarizing properties, transmittances. This is particularly useful among other applications to probe the constituent elements organization in biological tissues and to detect defects or outliers in optical elements. In the framework of Mueller parameters imaging, polarization-encoded images have 16 channels which make physical interpretation of such multi-dimensional structures hard to grasp at once. Furthermore, the information content is intricately combined in the parameters channels which involves the need for a proper tool that allows the analysis and understanding of this kind of images. Figure 5 shows real measured images acquired by a full Mueller imaging polarimeter. The used object to carry out these measurements consists of two dichroic patches (Polaroid) and a cellophane shape contacted on a diffusing glass slide with a drop of water. The overall mount was backlight illuminated by a polarized beam, and the intensity images were sensed by a four probing states rotating quarter-wave-plate analyzer in front of a 12-bits CCD scientific camera. We note that the different orientations were given to transmission axes of the Polaroid shapes to obtain different signatures at the output.

We perform a 5 classes classification using the proposed method. One observes that in spite of strong noise in the different channels, the regularized MPPCA algorithm manages correctly the various objects within the scene. The problem of analyzing polarization-encoded images and to explore the potential of this information for classification issues is now under study.

6 Conclusion

A new Markov regularization of the Mixture of Probabilistic Principal Component Analyzers (MPPCA) for multi-component image joint reduction/segmentation has been presented in this paper. This method intrinsically avoided the well-known curse of dimensionality when handling large number of components. The model parameter estimation is performed using the EM algorithm. The results presented on synthetic image prove the interest to link MPPCA advantages and HMT regularization properties to segment high-dimensional data cube. Results presented on real images in the framework of remote sensing and Mueller imagery are very satisfactory and show the high potential of the proposed method. In comparison with a Maximum likelihood classification process based on MPPCA, our classification shows clearly the benefit of the smoothing effect of the Markov regularization. This method is promising and can be easily extended to other mixtures of latent variable models when a statistical modeling is available. Thus, one may use the regularized mixture of factor analyzers based on the work of Ghahramani et al. [9], or the regularized mixture independent component analyzers based on the work of Roberts et al. [23]. The method may also be extended to other Markov models such as Hidden Markov Chain and Markov Random Field depending on the on hand application.

References

1. Hughes, G.F.: On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory **14**(1), 55–63 (1968)
2. Landgrebe, D.: Signal Theory Methods in Multispectral Remote Sensing. J Wiley, New York (2003)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. J Wiley, New York (2001)
4. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. J Wiley, New York (2001)
5. Huber, P.J.: Projection pursuit with discussion. Ann. Stat. **13**(2), 435–525 (1985)
6. Tipping, M.E., Bishop, C.: Mixtures of Probabilistic Principal Component Analysers. Neural Comput. **11**, 443–482 (1999)
7. Kambhatla, N., Leen, T.K.: Dimension reduction by local principal component analysis. Neural Comput. **9**, 1493–1516 (1997)
8. Lee, T.W., Lewicki, M.S., Sejnowski, T.J.: ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1078–1089 (2000)
9. Ghahramani, Z., Hinton, G.E.: The EM algorithm for mixtures of factor analyzers. Tech. Rep. CRG-TR-96-1, University of Toronto. http://www.gatsby.ucl.ac.uk/∼zoubin/papers.html (1996)
10. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transa. Pattern Anal. Mach. Intell., PAMI- **6**(6), 721–741 (1984)
11. Graffigne, C., Heitz, F., Pérez, P., Prêteux, F., Sigelle, M., Zerubia, J.: Hierarchical Markov random field models applied to image analysis : A review. In: SPIE Neural Morphological and Stochastic Methods in Image and Signal Processing. San Diego, 10-11 July 1995, vol. 2568, pp. 2–17 (1995)
12. Kato, Z., Berthod, M., Zérubia, J.: A hierarchical Markov random field model and multitemperature annealing for parallel image classification. Graph. Models Image Process. **58**(1), 18–37 (1996)
13. Pérez, P., Chardin, A., Laferté, J.-M.: Noniterative manipulation of discrete energy-based models for image analysis. Pattern Recognit. **33**(4), 573–586 (2000)
14. Mignotte, M., Collet, C., Pérez, P., Bouthemy, P.: Sonar image segmentation using an unsupervsised hierarchical mrf model. IEEE Trans. Image Process. **9**(7), 1–17
15. Giordana, N., Pieczynski, W.: Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **19**(5), 465–475 (1997)
16. Laferté, J.-M., Pérez, P., Heitz, F.: Discrete markov image modeling and inference on the quad-tree. **9**(3), 390–404 (2000)
17. Provost, J.N., Collet, C., Rostaing, P., Pérez, P., Bouthemy, P.: Hierarchical markovian segmentation of multispectral images for

the reconstruction of water depth maps. Comput. Vis. Image Underst. **93**(2), 155–174 (2004)
18. Tipping, M.E., Bishop, C.: Probabilistic Principal Component Analysers. J. R. Stat. Soc. B **61**, 611–622, Part 3
19. Chardin, A., Pérez, P.: Semi-iterative inferences with hierarchical energy-based models for image analysis. Proceedings of international Workshop EMMCVPR'99 : Energy Minimisation Methods in Computer Vision and Pattern Recognition. Lecture Notes in Computer Science, vol. 1654 (1999)
20. Williams, C.K.L., Adams, A.J.: Advances in Neural Information Processing Systems. chapter DTs: Dynamic trees, MIT Press (1999)
21. Rehrauer, H., Seidel, K., Datcu, M.: Multiscale image segmentation with a dynamic label tree. In: Stein, T.I., (ed.) Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 1772–1774 (1998)
22. Collet, Ch., Zallat, J., Takakura, Y.: Clustering of mueller matrix images for skeletonized structure detection. Opt. Express **12**(7), 1271–1280 (2004)
23. Roberts, S., Penny, W.: Mixtures of independent component analysers. In: Proceedings of ICANN, Vienna (2001)