

Hyperspectral Image Segmentation with Markov Chain Model

Grégoire MERCIER*, Stéphane DERRODE[†], and Marc LENNON*[‡]

*GET - ENST Bretagne, dpt ITI, CNRS FRE 2658 TAMCIC, team TIME.

Technopole Brest-Iroise, CS 83818; 29238 Brest cedex - France

e-mail:gregoire.mercier@enst-bretagne.fr

[†]ENSPM, Laboratoire Fresnel (UMR 6133), team GSM.

Domaine universitaire de Saint Jérôme, 13013 Marseille Cedex 20 - France

[‡]AvelMor, technopole Brest-Iroise, Place Nicolas Copernic, 29 280 Plouzané - France

Abstract—The Hidden Markov Chain (HMC) model has been extended to take into consideration the multi-component representation of an hyperspectral data cube. Parameters estimation is performed using the general Iterative Conditional Estimation (ICE) method. The vectorial extension of the model is straightforward since the vectorial point of view joints the observation of each pixel as a spectral signature. Then, the segmentation procedure achieves an estimation of multi-dimensional correlated probability density functions (pdf). Multi-dimensional densities have been estimated by a set of 1D densities through a projection step that makes component independent and of reduced dimension.

Classifications have been applied to an image from the CASI sensor including 17 bands (from 450 to 950 nm) representing an intensive agricultural region (Brittany, France). Since, the intrinsic dimensionality of the observation has been estimated to 4, the multi-component HMC model has been applied to the CASI image reduced to 4 bands through an adapted projection pursuit method.

I. INTRODUCTION

Texture analysis plays an important role in image analysis. Various methods for feature extraction have been proposed in the literature. Nevertheless, texture analysis remains a difficult problem and even more when applied to color, multi- or hyperspectral images where each pixel takes its values in a multi-dimensional space.

Most feature extraction methods relies on the assumption that texture can be defined by the local statistics of pixel grey levels. First-order statistics may be derived from the local histogram and used as texture characterization [1]. Nevertheless, second-order statistics are required to outperform texture description. The popular co-occurrence matrix [2] may be used as an efficient tool for texture-based segmentation [3]; but some other techniques have been proposed with success: description through wavelet coefficient statistics [4], the Markov Random Field [5] or Markov Chain [6] models, fractal-based techniques [7], or multi-fractals [8] have been widely studied and applied on remotely-sensed images. Some filtering-based techniques may also be defined [9] to characterize the neighborhood of each pixels.

In this paper, a multi-component Hidden Markov Chain (HMC) has been developed for the segmentation of multi-

or hyperspectral data cube. The segmentation is based on the estimation of generalized mixture of multi-dimensional laws.

II. HIDDEN MARKOV CHAIN MODEL

In the context of HMC model, the remotely sensed data is considered as a *noisy* observation from which the segmentation has to be found. The 2D observation is first transformed into a 1D chain through a Hilbert-Peano scan on the image [10]. When the observation is an hyperspectral data cube, the Hilbert-Peano scan is applied spatially in order to yield a chain that contains each pixel y (*i.e.* each spectral signature) along the scan.

A. Overview of the scalar case

It is considered that observations y , which are the pixels of the image, are the noisy realization of a random process X that takes its values in $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$, K being the number of classes expected for the segmentation.

Several methods may be considered when the link between observation Y and segmentation X (*i.e.* $P(X, Y)$) is known. If it is not the case, $P(X, Y)$ has first to be estimated. Here, it is supposed that X is a stationary Markovian process and parameters estimation is achieved using ICE algorithm [11]. The ICE procedure is based on the conditional estimation of some estimators from the complete data (x, y) . It is an iterative method which produces a sequence of estimations θ^q of parameters θ as follows: 1) initialize θ^0 ; the first guess of X is achieved by a fuzzy C-means segmentation. 2) compute $\theta^{q+1} = E[\hat{\theta}(X, Y)|Y = y]$, where $\hat{\theta}(X, Y)$ is an estimator of θ . Usually, ICE is stopped when $\theta^{q+1} \approx \theta^q$. The parameters θ to estimate are of two kinds:

- 1) the set Π that characterizes the stationary Markov Chain X : the initial probability vector $\pi = (P(X = \omega_1), \dots, P(X = \omega_K)) = (\pi_{\omega_1}, \dots, \pi_{\omega_K})$ and the transition matrix A of components $P(X = \omega_k, X = \omega_l) = a_{w_k, w_l}$ ($1 \leq k, \ell \leq K$).

In an ICE iteration, the expectation of those parameters can be evaluated analytically along the Hilbert-Peano chain by using the normalized Baum's Forward and Backward probabilities [12].

2) the mixture parameters set Δ that characterizes the observation for each class ω_k : $P(Y|X = \omega_k)$ of pdf f_{ω_k} . In the Gaussian case, Δ is composed of means and variances; by using the Pearson's system of distributions [6], Δ needs the four first moments for each pdf.

For those parameters, θ^{q+1} is not tractable. But it can be estimated by empirical mean of several estimations according to $\theta^{q+1} = \frac{1}{L} \sum_{\ell} \hat{\theta}(x^{\ell}, y)$, where x^{ℓ} is an *a posteriori* realization of X conditionally on Y .

Finally, the restoration (that is estimation of X) is achieved by using the Maximum *a Posteriori* Mode (MPM) Bayesian segmentation rule.

B. Integration of multi-component observations

In the case of the segmentation of hyperspectral data cube, observation Y is becoming a multi-dimensional random variable. The dimension M is the number of spectral bands of the hyperspectral cube. Then, realizations \mathbf{y} of \mathbf{Y} take their value in M .

Then, pdf $f_{\omega_k}(\cdot)$ are becoming M -dimensional distributions to be estimated. Unfortunately, components of spectral signatures are very correlated and cannot be considered as independent. In order to make the estimation of $f_{\omega_k}(\cdot)$ easier by considering M estimations of independent 1D pdf, several transformations are applied to the observation \mathbf{Y} :

- 1) Dimension Reduction, in order to prevent from the Hughes phenomenon and to make estimations of the pdf more accurate (see section III). The transformed observation becomes of dimension M' .
- 2) Independent Component Analysis (ICA) of the projected observation. Thus, the pdf $f(\mathbf{y})$, where \mathbf{y} has been transformed by ICA, becomes a M' -dimensional distribution with self-independent components, *i.e.* $f(\mathbf{y}) = \prod_{m=1}^{M'} f_m(y_m)$.
- 3) Principal Component Analysis (PCA) applied on the data before estimation of $f_{\omega_k}(\mathbf{y})$, so that the M' components of $f_{\omega_k}(\mathbf{y})$ are becoming uncorrelated.

The first two steps may be viewed as pre-processing before applying HMC segmentation and will be considered together as explained in the next section. $f(\mathbf{y})$ is then considered of smaller dimension (that suits natural statistic estimations) and of independent components. But decorrelation does not induce independence and $f_{\omega_k}(\mathbf{y})$ is not of independent components even if it is the case for $f(\mathbf{y})$. Actually, it will be considered as independent and $f_{\omega_k}(\mathbf{y})$ will be estimated with M' 1D pdf.

It should be possible to apply ICA for each estimation of $f_{\omega_k}(\mathbf{y})$ at each iteration of the ICE algorithm; but, ICA is a non-orthogonal projection that tends to yield multi-modal distributions (actually, non-gaussianity acts as an independence criteria) that makes ICA procedure not converging. Although not rigorous, it makes sense to estimate $f_{\omega_k}(\mathbf{y})$ through a PCA at each iteration of the ICE algorithm, knowing that components $f(\mathbf{y})$ have been made independent first.

III. DIMENSION REDUCTION AND INDEPENDENCE

It is known that pattern recognition algorithms have poor capabilities in large dimension spaces but their performances are largely improved in lower dimension. In the study, several linear projections (such as Principal Component Analysis – PCA–, Maximum Noise Fraction Transform –MNF– or its equivalent as Noise Adjusted PCA, Projection Pursuit –PP–) and some non-linear projections (such as Curvilinear Component Analysis –CCA– or Curvilinear Distance Analysis –CDA–) have been evaluated [13].

Those transformations are used in order to reduce dimensionality with specific point of view:

- PCA: the transformation is linear and orthogonal. Components are ordered by axes of decreasing variance. The transformation matrix is evaluated by diagonalizing covariance matrix. Nevertheless, it appears that the components of smaller variance still contain interesting structural information.
- MNF: the transformation is based on an estimation of spatial noise before applying PCA-like transform. This transformation is interesting since components are ordered in signal-to-noise ratio and the components of smaller SNR do not contain anymore structural information. In this NA-PCA, the “noise” is defined as the spatial gradient in horizontal and vertical directions.
- PP: the transformation is non-orthogonal and the components are becoming independent and ordered relatively to a criteria of non-gaussianity.
- NAPP: this non-orthogonal transform is an extent of the PP by considering the same strategy as the MNF which is an extent of PCA, that is fully described in [14]. It yields independent components that are ordered in signal-to-noise ratio, but the initial data has first been transformed in order to decorrelate the signal from the spatial noise.
- CCA: this non-linear projection takes a geometrical point of view to project the data into a smaller dimension space. The benefit of this approach is to take into consideration the topology of the data points.
- CDA: this non-linear projection is very similar to the CCA but uses local topology (a distance evaluation that is not quadratic anymore but follows the links between data points themselves) to build the projection system. It induces projections that fit geometrical characterization of clusters and does not prove to have benefit with statistical clustering technique.

All those transformation techniques have the same goal of reducing data but the means used induce specific behavior relatively to HMC-based segmentation. CCA and CDA are non-linear projections that are based on the preservation of the local topology. Those induce multi-component pdf $f(\mathbf{y})$ to be of multi-modal densities for which the characterization by mixture of laws of Gaussian densities or from the Pearson's system of distributions is not relevant. Actually, it appears

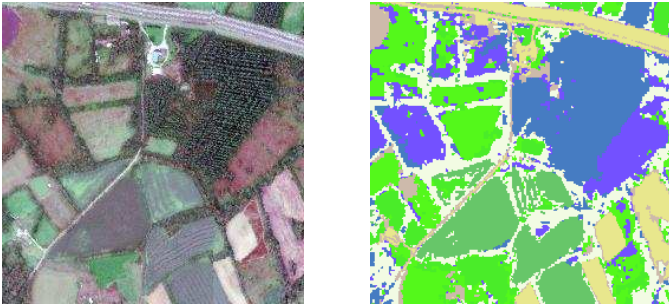


Fig. 1. Original CASI image and the multi-component Markov Chain segmentation.

that CCA and CDA are efficient transformations for structure-based processes [15]. On the contrary, transformations that are based on correlation or independence concepts keep a statistical point of view and suit HMC-based segmentation. Moreover, transformations that take into consideration spatial dispersion (MNF, NAPP) reduce the number of false alarms by making clusters spatially more homogeneous.

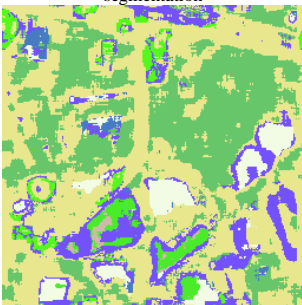
IV. EXPERIMENTS

Multi-component HMC model has been applied to a hyperspectral image from the airborne CASI sensor including 17 spectral bands from 450 to 950 nm. The ground resolution is two meters and the image has been calibrated to reflectance by the means of the empirical line method, as shown on Fig. 1.

The segmentation has been achieved for segmenting 8 classes that should correspond to artificial forest, several kinds of fields, water, roads and wasteland.

Comparisons with supervised methods show that the image segmented by multi-component HMC model (Fig.1) is much more relevant since it takes into consideration the spectral signatures of the observation (reduced by NAPP) but also the spatial organization of the texture. The Maximum Likelihood illustration (Fig.2-a) takes only in consideration the spectral signatures while the wavelet-based hyperspectral texture segmentation (Fig.2-b) does not take into consideration the coarse spectral signature. It results, with the unsupervised multi-component HMC, a better discrimination of the different areas

(a) Wavelet-based hyperspectral texture segmentation



(b) Maximum Likelihood segmentation

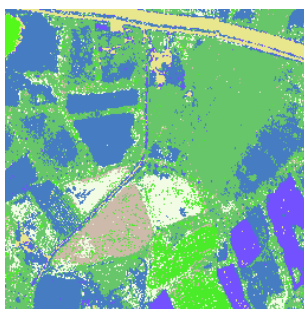


Fig. 2. Segmentation of the CASI image, reduced by NAPP, with 8 classes. (a) wavelet-based texture segmentation [15], (b) gaussian maximum likelihood.

and a more homogeneous characterization of the classes.

V. CONCLUSION

In this work, we described an extent of a HMC model for the unsupervised segmentation of hyperspectral data cube. The extension of the HMC model to multi-component data is almost straightforward. However, one difficulty arises from the estimation of non-gaussian multi-dimensional noise distribution. Here, we have considered the mixture parameters as independent through the true independence of $f(\mathbf{y})$ components and through the uncorrelation of $f_{\omega_k}(\mathbf{y})$ components. An illustration of the segmentation of CASI image proved the interest of the method.

ACKNOWLEDGMENT

The authors would like to thank Dr. Laurence Hubert-Moy from COSTEL for her generous offer of CASI data and for helpful and valuable comments and suggestions.

REFERENCES

- [1] L.-K. Soh and C. Tsatsoulis, "Automated sea ice segmentation (ASIS)," in *IGARSS*, 1998.
- [2] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, November 1973.
- [3] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE transactions on geoscience and remote sensing*, vol. 37, no. 2, pp. 780–795, March 1999.
- [4] M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, February 2002.
- [5] G. Hazel, "Multivariate gaussian MRF for multispectral scene segmentation and anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1199–1211, May 2000.
- [6] S. Derrode, G. Mercier, J.-M. LeCaillec, and R. Garello, "Estimation of sea-ice SAR clutter statistics from Pearson's system of distributions," in *IGARSS*, 2001.
- [7] J.-C. Liu, W.-L. Hwang, and M.-S. Chen, "Estimation of 2-D noisy fractional brownian motion and its applications using wavelets," *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1407–1419, August 2000.
- [8] H. Chen and W. Kinsner, "Texture segmentation using multifractal measures," in *Proceedings of the Wescanex Conference on Communications, Power and Computing*, 22-23 May 1997, pp. 222–227.
- [9] P. Kruijzinga and N. Petkov, "Nonlinear operator for oriented texture," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1395–1407, October 1999.
- [10] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chain and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 465–475, May 1997.
- [11] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. and Vis.*, vol. 1, pp. 261–268, 1992.
- [12] P. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recognition Letter*, vol. 3, pp. 369–373, 1985.
- [13] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy, "Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images," in *SPIE*, 2001.
- [14] M. Lennon and G. Mercier, "Noise-adjusted non orthogonal linear projections for hyperspectral data analysis," in *IGARSS*, 2003.
- [15] G. Mercier and M. Lennon, "On the characterization of hyperspectral texture," in *IGARSS*, 2002.