UNSUPERVISED MULTICOMPONENT IMAGE SEGMENTATION COMBINING A VECTORIAL HMC MODEL AND ICA

Stéphane Derrode — GSM group, ENSPM, stephane.derrode@enspm.u-3mrs.fr

Grégoire Mercier — ITI department, ENST Bretagne, gregoire.mercier@enst-bretagne.fr

Wojciech Pieczynski — CITI department, INT, wojciech.pieczynski@int-evry.fr

ABSTRACT

This work extents the Hidden Markov Chain (HMC) model for the unsupervised segmentation of multicomponent images. Although the vectorial extension of the model is almost straightforward, we are faced to the problem of estimating a mixture of non-Gaussian multidimensional densities. In this work, we adopt an Independent Component Analysis (ICA) approach that allows the mutual dependance between the layers to be taken into account in the segmentation process. Classification results on a four bands SPOT-IV image illustrates the method. Also, a comparison is performed when only mutual independence or correlation between the components is assumed.

1. INTRODUCTION

The aim of this paper is to present an extension of the HMC model for the unsupervised segmentation of multicomponent images. Such vectorial image can be obtained, for example, from different channels (multispectral, color images), from several sensors (multisensor) or from image taken at various moments (multitemporal). Each component exhibits different characteristics of the spatial scene and the motivation for this work is to combined their respective information in order to improve the segmentation obtained when only one image is considered.

Precisely, in the one image case, Bayesian restoration in the framework of the hidden Markov models (HMM) is among the best known statistical classification methods. This success is mainly due to the fact that when the unobservable process X can be modeled by a finite Markov model, then X can be recovered from the observed process Y using different Bayesian classification criteria like Maximum A Posteriori (MAP) or Maximum Posterior Mode (MPM). Among HMMs, the Hidden Markov Chain (HMC) models, applied to a Hilbert-Peano scan of the image [1], constitute a fast and sometimes competitive alternative to hidden Markov fields, even though the latter provide a finer and more intuitive modeling of spatial relationships.

In the case of unsupervised classification, the statistical properties of the classes are unknown and the mixture estimation problem should first be solved. In the HMC context, iterative methods such as Expectation-Maximisation (EM) or Stochastic EM can be used. In this study, we limit ourselves to a third procedure called Iterative Conditional Estimation (ICE) [2], which has been successfully performed in several unsupervised contexts such as sonar, medical and radar images [3, 4, 5]. In such modalities, the noise is not necessarily Gaussian and several generalized mixture estimation algorithms have been proposed [1, 6].

When dealing with a M-layers multicomponent image, the extension of ICE is almost straightforward but stumble against the problem of estimating a generalized mixture of M-dimensional non-Gaussian densities from samples. The present work lies within the scope of this general problem. Partial solutions to the problem have been proposed in [1, 7] where, respectively, mutual independence and correlation between layers have been assumed. However, such assumptions can not be justified in most real applications. In this paper, we combine ICE estimation in a HMC context with an ICA procedure in order to totaly take into account the layers dependence in the classification process. This paper is organized as follows. In next section, the ICE principle is briefly recalled. Then, a generalized multidimensional density estimation algorithm based on ICA is presented in section 3. In section 4, the method is illustrated on a four bands SPOT-IV image and the segmentation result is compared to the ones obtained when only mutual independence or correlation is assumed. Finally, conclusions are drawn in section 5.

2. ICE IN A VECTORIAL HMC MODEL

The pixels of the M layers of a multicomponent images are first transformed into 1D chains using a Hilbert-Peano scan on each image. Hence, we get N series of M data, denoted by $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_N)$, where $\mathbf{y}_n = \left(y_n^1, ..., y_n^M\right)^t$, $1 \leq n \leq N$. The objective is to classify each $\mathbf{y}_n$ into a

This work was supported by the 'Groupement des Écoles des Télécommunications' under contract number GET RE-315. The authors are also grateful to L. Hubert-Moy (COSTEL) for providing the SPOT-IV image used for illustration.

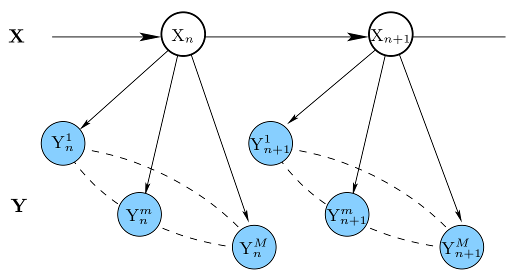

Fig. 1. Independence assumptions in the vectorial HMC model. The dash curves represent the mutual dependance of the observations along the layers.

set of K classes $\Omega = \{\omega_1, ..., \omega_K\}$ in order to obtain the segmented chain $\mathbf{x} = (x_1, ..., x_N)$. The segmented image is then reconstructed from $\mathbf{x}$ by using an inverse Hilbert-Peano scan.

The probabilistic approach developed here consists in assuming that x_n is a realization of a random variable X_n, and each $\mathbf{y}_n$ is a realization of a random vector $\mathbf{Y}_n = \left(\mathrm{Y}_n^1, ..., \mathrm{Y}_n^M\right)^t$. Thus the problem is to estimate the unobserved realization $\mathbf{x}$ of a random process $\mathbf{X} = (\mathrm{X}_1, ..., \mathrm{X}_N)$ from the observed realization $\mathbf{y}$ of a random process $\mathbf{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_N)$. The process $\mathbf{X}$ is supposed to be Markovian and stationary. We also consider the usual two following assumptions: the random vectors $(\mathbf{Y}_n)_{1\leq n\leq N}$ are independent conditionally on $\mathbf{X}$ and the distribution of each $\mathbf{Y}_n$ conditionally on $\mathbf{X}$ is equal to its distribution conditional on X_n. It is important to note that the random variables $(\mathrm{Y}_n^m)_{1\leq m\leq M}$ are not assumed to be mutually independent conditionally on X_n (see Fig. 1).

In the case of unsupervised segmentation, the distribution $P(\mathbf{X}, \mathbf{Y})$ is unknown and must first be estimated in order to apply a Bayesian classification technique. Therefore, we have to estimate the following sets of parameters $(1 \leq k, l \leq K)$:

1. the set $\mathbf{\Pi}$ characterizing the stationary Markov chain $\mathbf{X}$, i.e. the initial probability vector $\boldsymbol{\pi} = (\pi_{\omega_1}, ..., \pi_{\omega_K})$ and the transition matrix $\mathbf{A}$ with entries a_{ω_k,ω_l};

2. the set $\mathbf{\Delta}$ characterizing the observation densities, i.e. the parameters of the K M-dimensional distributions f_{ω_k}. In the Gaussian case, $\mathbf{\Delta}$ is composed of the means and the covariance matrix.

The estimation of all the parameters in $\mathbf{\Theta} = (\mathbf{\Pi}, \mathbf{\Delta})$ can be achieved using the general ICE algorithm. The ICE procedure is based on the conditional expectation of some estimators from the complete data $(\mathbf{x}, \mathbf{y})$ [2].ICE is an iterative method which produces a sequence of estimations θ^q of parameter θ as follows : (1) initialize θ^0; (2) compute $\theta^{q+1} = \mathrm{E}\left[\hat{\theta}(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}\right]$, where $\hat{\theta}(\mathbf{X}, \mathbf{Y})$ is an estimator of parameter θ. In practice, we interrupt the algorithm at iteration Q if $\theta^{Q-1} \approx \theta^Q$. This procedure leads to two different situations:

- For parameters in $\mathbf{\Pi}$, the expectation can be calculated analytically, by using the normalized Baum's Forward and Backward probabilities [8];

- For parameters in $\mathbf{\Delta}$, θ^{q+1} is not tractable. However, it can be estimated by computing the empirical mean of several estimates according to $\theta^{q+1} = \frac{1}{L}\sum_{l=1}^{L} \hat{\theta}\left(\mathbf{x}^l, \mathbf{y}\right)$, where $\mathbf{x}^l$ is a *a posteriori* realization of $\mathbf{X}$ conditionally on $\mathbf{Y}$[1].

We now present three methods for the estimation of the densities f_{ω_k}, $1 \leq k \leq K$.

3. MULTIDIMENSIONAL DENSITY ESTIMATION

Let $\mathbf{x}$ be a realization of $\mathbf{X}\,|\mathbf{Y}$ and let $\mathbf{z}$ denote the data of $\mathbf{y}$ attributed to a given class ω_k. The Multidimensional Density Estimation (MDE) algorithms presented below should be applied on all classes in Ω and at each ICE iteration.

If Gaussian densities are considered in a MDE problem, parameters estimation can be easily estimated from the first and second order moments of a M-dimensional sample. However, in a number of modalities, it has been shown that the noise can not be properly modeled by Gaussians. Moreover, the nature and the form of the distribution of each class in the different layers can vary (e.g. multisensor case). However, non-Gaussian multidimensional densities can be difficult to estimate. One solution is to decompose the problem into M one-dimensional density estimations. Several strategies are possible, depending on the assumptions made on the links between the components.

If independence is assumed [1], f_{ω_k} is the product of M densities $g^1_{\omega_k}$, ..., $g^M_{\omega_k}$ defined on $\mathbb{R}$:

$$f_{\omega_k}(\mathbf{z}_n) = \prod_{m=1}^{M} g^m_{\omega_k}(z_n^m). \qquad (1)$$

However, most of the time, multicomponent images can not be considered mutually independent. Hence, a more sophisticated solution consists in applying a Principal Component Analysis (PCA) algorithm on the data before densities estimation [7]. This can be done by projecting $\mathbf{z}$ onto an orthonormal system defined by $\mathbf{W}$ so that the new data $\mathbf{t}_n = \mathbf{W}\,\mathbf{z}_n$ are decorrelated[2]. Hence, we get the following estimation:

$$f_{\omega_k}(\mathbf{z}_n) = |\det(\mathbf{W})| \prod_{m=1}^{M} g^m_{\omega_k}(t_n^m). \qquad (2)$$

[1]Simulations are rather simple since $\mathbf{X}\,|\mathbf{Y}$ is a non stationary Markov chain whose transition matrix is tractable analytically.

[2]A solution is given by $\mathbf{W}^t = (\mathrm{C}(\mathbf{\Gamma_z}))^{-1}$ where C is the Choleski decomposition and $\mathbf{\Gamma_z}$ is the covariance matrix of the observed data.

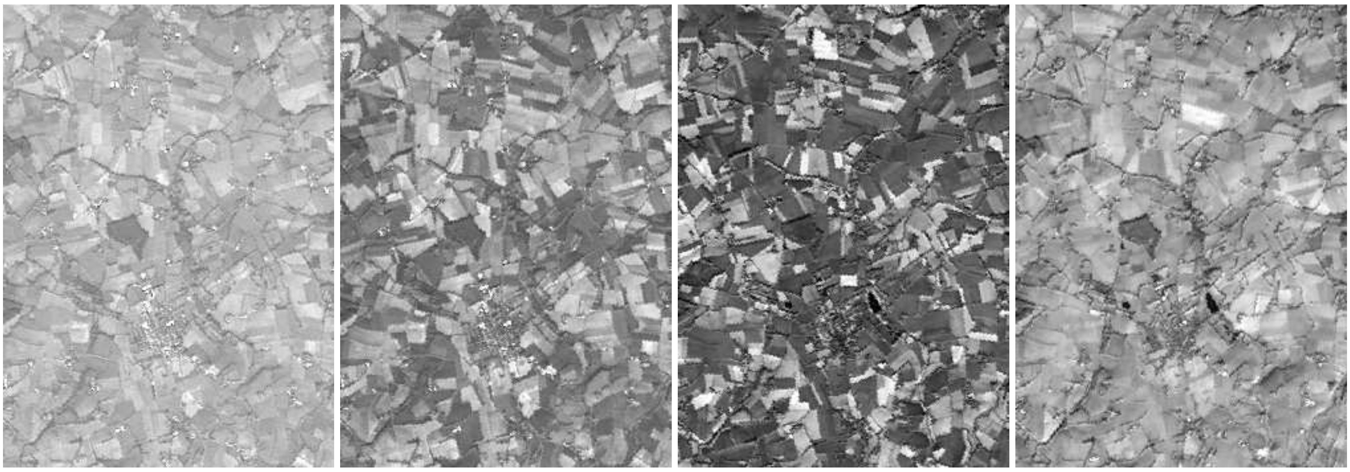

Fig. 2. Four bands SPOT-IV image to be segmented (size: 250×350, $N = 87500$, $M = 4$).

However, ideally, data should be independent and this naturally fits an ICA approach. The objective of ICA is to find a linear transformation $\mathbf{W}'$ so that the new data $\mathbf{s}_n = \mathbf{W}'\,\mathbf{z}_n$ are mutually independent. A solution to this problem can be find under the assumption that the s^m components are not Gaussian. This is an optimization problem that has no analytical solution. The computation of $\mathbf{W}'$ requires the optimisation of a 'non-gaussianity' criterion such as the kurtosis or the neguentropy. For sake of reduced complexity, we have chosen to adapt the one proposed by Hyvärinen in [9]. Then, the density f_{ω_k} can then be reconstructed using a formula similar to Eq. (2), replacing t_n^m by s_n^m.

It is important to note that no assumption has been done on the shape of the 1D densities so that parametric estimators (moments, maximum likelihood, ...) and non parametric estimators (kernel, orthogonal expansion) can be used.

4. SEGMENTATION RESULTS

Fig. 2 shows an extract of a SPOT-IV image of Brittany (France) with four spectral bands (red, green, near infrared and middle infrared). We decided to segment this multicomponent image into $K = 4$ classes and to estimate the mixture from the parametric family of exponential power distributions :

$$p(x; \mu, \alpha, \beta) = \frac{\beta}{2\alpha\,\Gamma\,(1/\beta)}\; e^{\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}}, \qquad (3)$$

where $\Gamma(\cdot)$ is the Gamma function, α and β are respectively the scale and shape parameters. This family includes a large variety of symmetrical shapes, including Laplace ($\beta = 1$) and Gaussian ($\beta = 2$) distributions. The mean was estimated from the first empirical moment of samples, whereas the shape and scale parameters were estimated using the maximum likelihood estimator.

The set $\mathbf{\Delta}$ is composed of the 3 parameters for the $M = 4$ images and the $K = 4$ classes ($card(\mathbf{\Delta}) = 48$). The set $\mathbf{\Pi}$ is composed of the 4 initial probabilities and the transition matrix of 16 entries ($card(\mathbf{\Pi}) = 20$). Hence, 68 parameters are estimated at each ICE iteration. The initial parameter values was computed using the vectorial K-means classification result as a realization of $\mathbf{X}$. The Bayesian MPM classification has been performed from the values obtained after 70 ICE iterations.

The class images obtained from the three MDE algorithms described above have been reported in Fig. 3. The segmentation results confirm the interest of a vectorial HMC model for the unsupervised classification of multicomponent images and the robustness of ICE in such a context. Nevertheless, segmentations exhibit some differences directly related to the dependence assumption made on the layers. No general conclusion on the thematic representation of each class can be drawn since ground truth data were not available. However, one can note that PCA and ICA based segmentations are very similar. Indeed, all the estimated parameters are closed to each other. However, ICA based segmentation furnishes more homogeneous regions and field frontiers seems to be better detected.

The computation times needed to segment this multicomponent image are respectively of 23, 30 and 80 minutes on a PC with 1.3 GHz processor running Linux. The complexity of the ICE based ICA algorithm is much more important than the two other methods since the estimation of $\mathbf{W}'$ is an iterative process done for each class and at each iteration. However, the computation time can be significantly reduced by performing an initial estimation from the segmented image obtained with ICE based PCA. Hence, only a small number of ICE iterations is necessary.

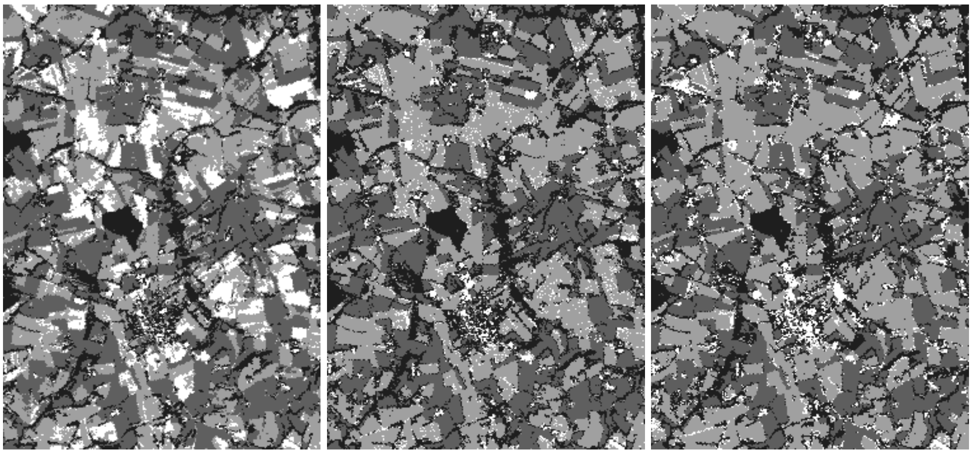

Fig. 3. Segmentation of images in Fig. 2. From left to right: mutual independence, PCA and ICA based methods.

5. CONCLUSION

In this works, we described an extent of a HMC model for the unsupervised segmentation of multicomponent images. The extension of the HMC model to vectorial data is almost straightforward. However, one difficulty arises from the estimation of non-Gaussian multidimensional distribution. In most applications, the layers hold redundant information and mutual independence can not be assumed. In this work, we incorporated an ICA approach in the general ICE procedure for parameters estimation. An example of a SPOT-IV image, and some others in radar imaging (multiscale [10], multispectrale [11] and multitemporal images [12]), confirms the interest of the method.

It is important to note that the multicomponent segmentation method presented here is valid when the number of components is small with respect to the number of pixels. Indeed, for example, the large number of spectral bands in a hyperspectral image induces non consistent estimations (Hughes phenomenon). A solution consists in reducing the dimension by projecting the data into a lower dimension space.

6. REFERENCES

[1] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor HMC and unsupervised image segmentation," *IEEE Trans. on PAMI*, vol. 19, no. 5, pp. 465–475, 1997.

[2] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. and Vis.*, vol. 1, pp. 261–268, 1992.

[3] M. Mignotte, C. Collet, P. Pérez, and P. Bouthémy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. on IP*, vol. 9, no. 7, pp. 1216–1231, 2000.

[4] M. Mignotte, J. Meunier, J.-P. Soucy, and C. Janicki, "Comparison of deconvolution techniques using a distribution mixture parameter estimation: Application in SPECT imagery," *J. of Electr. Imaging*, vol. 1, no. 1, pp. 11–25, 2002.

[5] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised segmentation of radar images using HMC and HMRF," *to appear in* IEEE Trans. on GRS, 2003.

[6] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixture and its application in image segmentation," *IEEE Trans. on IP*, vol. 6, no. 10, pp. 1364–1375, 1997.

[7] W. Pieczynski, J. Bouvrais, and C. Michel, "Estimation of generalized mixture in the case of correlated sensors," *IEEE Trans. on IP*, vol. 9, no. 2, pp. 308–311, 2000.

[8] P.A. Devijver, "Baum's Forward-Backward algorithm revisited," *Pat. Recog. Let.*, vol. 3, pp. 369–373, 1985.

[9] A. Hyvärinen and E. Oja, "ICA: algorithms and applications," *Neur. Net.*, vol. 13, pp. 411–430, 2000.

[10] G. Mercier, S. Derrode, and W. Pieczynski, "Multiscale oil slick segmentation with Markov chain model," in *IGARSS*, 2003.

[11] G. Mercier, S. Derrode, and M. Lennon, "Hyperspectral image segmentation with Markov chain model," in *IGARSS*, 2003.

[12] S. Derrode, G. Mercier, and W. Pieczynski, "Unsupervised change detection in SAR images using a multicomponent HMC model," in *MultiTemp*, 2003.