UNSUPERVISED IMAGE SEGMENTATION BASED ON A NEW FUZZY HMC MODEL

C. Carincotte, S. Derrode

GSM, Inst. Fresnel (CNRS - UMR 6133),
Dom. Univ. de St Jérôme,
F-13013 Marseille Cedex 20
cyril.carincotte@fresnel.fr

G. Sicot, J.M. Boucher

GET, ENST Bretagne,
TAMCIC, CNRS FRE 2658,
CS 83818, F-29238 Brest Cedex

ABSTRACT

In this paper, we propose a technique, based on a fuzzy Hidden Markov Chain (HMC) model, for the unsupervised segmentation of images. The main contribution of this work is to simultaneously use Dirac and Lebesgue measures at the class chain level. This model allows the coexistence of hard and fuzzy pixels in the same picture. In this way, the fuzzy approach enriches the classical model by adding a fuzzy class, which has several interpretations in signal processing. One such interpretation in image segmentation is the simultaneous appearance of several thematic classes on the same pixel (mixture). Model parameter estimation is performed through an extension of the Iterative Conditional Estimation (ICE) algorithm to take into account the fuzzy part. The fuzzy segmentation of a real image of clouds is studied and compared to the classification obtained with a "classical" hard HMC model.

1. INTRODUCTION

This work addresses fuzzy statistical unsupervised image segmentation. Image segmentation is one of the major problem in image processing. The aim is to try to restitute the ground truth image ($\boldsymbol{x}$) from a noisy observation ($\boldsymbol{y}$). To that goal, the HMC model has been used successfully [1], thanks to the use of a Hilbert-Peano scan that converts the 2D grid into a 1D sequence [2]. The success of HMC models is due to the fact that when the unobservable process $\boldsymbol{X}$ can be modelled by a finite Markov chain and when the noise is not too complex, then $\boldsymbol{X}$ can be recovered from the observed process $\boldsymbol{Y}$ using different Bayesian classification techniques like Maximum A Posteriori (MAP), or Maximal Posterior Mode (MPM).

Nevertheless, it is sometimes interesting to take into account, not only the uncertainty of the noisy observation, but also the imprecision of this observation. To this aim, fuzzy Markov chains and fuzzy HMC have recently been studied respectively in [3] and in [4]. However, by adding a fuzzy class in a statistical model, we obtain an original modelling, different from both probabilistic and fuzzy modellings. Indeed, it preserves the propriety (measure of uncertainty) and robustness of the statistical segmentation and enriches it with the fuzzy characteristic (measure of imprecision). This has already been done in unsupervised image segmentation in two different estimation contexts: blind and contextual in [5], and hidden Markov random field (HMRF) in [6]. Fuzzy HMRF have also been used to model the so-called partial volume effect encountered in medical images context [7]. In this work, we propose to adapt this point of view to the HMC context.

The paper is organized as follows: HMC structure is briefly recalled in Section 2. We specify in Section 3 the fuzzy HMC model used for image segmentation. The unknown parameters estimation, achieved with an extension of the ICE method [5]to the context considered here, is then briefly presented. Comparative results on a real image are presented in Section 5, whereas conclusions and perspectives are drawn in Section 6.

2. HMC MODEL

For notational brevity, $\boldsymbol{X_{1\to n}}$ will denote the sequence of random variables $\{X_1, \ldots, X_n\}$ and $\boldsymbol{x}$ a realization of process $\boldsymbol{X}$.

2.1. Markov chain model

The sequence $\boldsymbol{X} = \{X_n\}_{n\in\{1,\ldots,N\}}$ is a finite Markov chain of order one, with length N, if and only if:

$$P\left(X_n = x_n \mid \boldsymbol{X_{1\to n-1}} = \boldsymbol{x_{1\to n-1}}\right) = P\left(X_n = x_n \mid X_{n-1} = x_{n-1}\right), \quad (1)$$

with each X_n taking its value in the set of classes $\Omega = \{0, \ldots, K-1\}$.

We only consider here the homogeneous Markov chain, in which Eq. (1) does not depend on the position n in the sequence. The set of state transition probabilities matrix $\boldsymbol{T} = \left\{ t_{x_{n-1,n}} \right\}$ is defined by:

$$t_{j,i} = P\left(X_n = i \mid X_{n-1} = j\right),$$

$\forall\, i,j \in \Omega$ and $\forall\, n \in \{2, ..., N\}$ with the state transition coefficients having the properties: $t_{j,i} \geq 0$ and $\sum_{i=0}^{K-1} t_{j,i} = 1$. The initial state probabilities are defined by:
$\pi_i = P\left(X_1 = i\right), \forall i \in \Omega$.

2.2. Hidden Markov chain model

Usually, HMC based image segmentation methods consider the two following assumptions: $\boldsymbol{H_1}$: the random variables $Y_1, \ldots, Y_N$ are independent conditionally on $\boldsymbol{X}$ and $\boldsymbol{H_2}$: the distribution of each Y_n conditionally on $\boldsymbol{X}$ is equal to its distribution conditionally on X_n.

Let $\boldsymbol{X} = \boldsymbol{X_{1 \to N}}$ be an homogeneous Markov chain, corresponding to the unknown class image.
We get $P\left(\boldsymbol{X} = \boldsymbol{x}\right) = \pi_{x_1} \prod_{n=2}^{N} t_{x_{n-1}, x_n}$.

Assuming that distributions of (X_n, Y_n) are independent of n, each state x_n of the state space is associated with a distribution, characterizing the repartition of observations:

$$f_{x_n}(y_n) = P\left(Y_n = y_n \mid X_n = x_n\right). \tag{2}$$

Given an observed sequence $\boldsymbol{y} = \boldsymbol{y_{1 \to N}}$, we can compute the joint state-observation probability by:

$$P\left(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}\right) = \pi_{x_1} f_{x_1}(y_1) \prod_{n=2}^{N} t_{x_{n-1}, x_n} f_{x_n}(y_n).$$

In unsupervised classification, the distribution $P\left(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}\right)$ is unknown and must first be estimated in order to apply a Bayesian classification technique. Therefore the following sets of parameters need to be estimated:

1. The set $\boldsymbol{\Gamma}$ characterizing the homogeneous Markov chain, i.e. the initial probability vector $\boldsymbol{\pi} = (\pi_i)_{\forall i \in \Omega}$ and the transition matrix $\boldsymbol{T}$.
2. The set $\boldsymbol{\Delta}$ characterizing the K pdf presented in Eq. (2). In the Gaussian case, $\boldsymbol{\Delta}$ is composed of the means and the variances.

3. NEW FUZZY HMC MODEL

Let us consider the problem of segmenting a satellite image into two classes: "land" and "sea". There obviously may be some pixels with only "land" and others with only "sea", but there may also exist many pixels, as over the coast, in which "land" and "sea" are simultaneously present. Thus we have two hard classes, say 0 for "land" and 1 for "sea", and a fuzzy one. Let specify this fuzzy class by $\varepsilon \in]0, 1[$, which can be seen as the proportion of the area of class 1 ("sea") in the considered pixel, the quantity $1 - \varepsilon$ consequently represents the proportion of "land" in this pixel.

Let us consider the two classes case $\Omega = \{0, 1\}$, called "hard" in what follows.

3.1. Fuzzy Markov chain representation

As detailed in [5, 6], a simple way to introduce a fuzzy class in such a statistical model is to consider that X_n does not take its value in the set $\{0, 1\}$ anymore, but in the continuous interval $[0, 1]$. The new representation of X_n is then $X_n = \varepsilon_n$, with:

- $\varepsilon_n = 0$ if the pixel is from class "0",
- $\varepsilon_n = 1$ if the pixel is from class "1",
- $\varepsilon_n \in]0, 1[$ if the pixel is a fuzzy one.

3.2. Fuzzy Markov chain probabilities

The statistical approach requires a definition of *a priori* probability defined on $\Omega = \{0, 1\}$.

As stated previously, each componentX_n contains two types of components: two hard (discrete) components and a (continuous) fuzzy one. Let δ_0, δ_1 be Dirac weights on 0 and 1 and μ the Lebesgue measure on $]0, 1[$. By taking $\nu = \delta_0 + \delta_1 + \mu$ as a measure on $[0, 1]$, the distribution of X_n can be defined by a density h on $[0, 1]$ with respect to ν.

If we assume that $\boldsymbol{X}$ is homogeneous and the distribution of each X_n is uniform on the fuzzy class, $P\left(X_n = \varepsilon_n\right)$ can be written:

$$\begin{aligned} h(0) &= P\left(X_n = 0\right) = \pi_0, \\ h(1) &= P\left(X_n = 1\right) = \pi_1, \\ h(\varepsilon_n) &= P\left(X_n = \varepsilon_n\right) = 1 - \pi_0 - \pi_1, \forall \varepsilon_n \in]0, 1[. \end{aligned}$$

Let now detail the expression of the transition probabilities of the Markov chain:

$$\begin{aligned} &P\left(X_n = \varepsilon_n \mid X_{n-1} = \varepsilon_{n-1}\right) = P\left(X_n = 0 \mid \varepsilon_{n-1}\right) \delta_0(\varepsilon_n) \\ &+ P\left(X_n = \varepsilon_n \mid \varepsilon_{n-1}\right) 1_{]0,1[}(\varepsilon_n) + P\left(X_n = 1 \mid \varepsilon_{n-1}\right) \delta_1(\varepsilon_n). \end{aligned}$$

4. PARAMETERS ESTIMATION

In order to apply some Bayesian criterion, we need to define $\boldsymbol{X} \mid \boldsymbol{Y}$. We again consider assumptions $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$.

4.1. ICE procedure principle

For the estimation of the parameters in $\boldsymbol{\Theta} = \{\boldsymbol{\Gamma}, \boldsymbol{\Delta}\}$, we propose to use an adaptation of the general ICE algorithm [5], which can be seen as an alternative to well-known Estimation-Maximization (EM) algorithm. In fact, ICE does not refer to the likelihood, a notion which is difficult to handle in the context of our study, but it is based on the conditional expectation of some estimators from the complete data $(\boldsymbol{x}, \boldsymbol{y})$. It is an iterative method which produces a sequence of estimations θ^q of parameter θ as follows: (1) initialize θ^0, (2) compute $\theta^{q+1} = E_q[\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) \,\big|\, \boldsymbol{Y} = \boldsymbol{y}]$, where $\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y})$ is an estimator of θ. In practice, we stop the algorithm at iteration Q if $\theta^{Q-1} \approx \theta^Q$. This procedure leads to two different situations detailed in the next subsections.

4.2. Estimation of parameters in $\boldsymbol{\Gamma}$

As in the classical case, parameters in $\boldsymbol{\Gamma}$ can be calculated analytically by using the normalized Baum-Welch algorithm. In this new context, the forward and backward probabilities can be defined by:

$$\begin{aligned} \alpha_{n+1}(\xi) &\propto \int_{]0,1[} \alpha_n(\xi)\, t_{\zeta,\xi}\, f_\xi(y_{n+1})\, d\zeta, \qquad (3) \\ \beta_n(\xi) &\propto \int_{]0,1[} \beta_{n+1}(\zeta)\, t_{\xi,\zeta}\, f_\zeta(y_{n+1})\, d\zeta. \end{aligned}$$

These integrals can not be solved analytically. A numerical integration must be performed; the interval $]0, 1[$ can be partitioned into a given number of sub-intervals. We though obtained F "discrete fuzzy" classes, whose fuzzy value corresponds to the medium value of the considered sub-interval. The bigger F is, the closer it is from Eq. (3), which implies bigger computation time (see discussion in Section 5).

4.3. Estimation of parameters in $\boldsymbol{\Delta}$

The set $\boldsymbol{\Delta}$ has to be estimated in this new context.

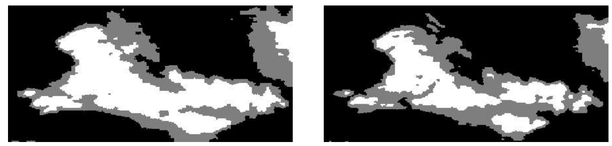
Denoting by $\mathcal{N}(m, \sigma^2)$ the normal distribution with mean m and variance σ^2, the pdf can then be expressed by:

$$\begin{aligned} \varepsilon_n = 0 :& \quad \mathcal{N} \quad (m_0, \sigma_0^2), \\ \varepsilon_n = 1 :& \quad \mathcal{N} \quad (m_1, \sigma_1^2), \\ \varepsilon_n \in]0,1[:& \quad \mathcal{N} \quad \begin{pmatrix} (1-\varepsilon_n) m_0 + \varepsilon_n m_1, \\ (1-\varepsilon_n)^2 \sigma_0^2 + \varepsilon_n^2 \sigma_1^2 \end{pmatrix}. \end{aligned}$$

For the parameters $\boldsymbol{\Delta} = \{m_0, m_1, \sigma_0, \sigma_1\}$, θ^{q+1} are not tractable. However, they can be estimated by computing the empirical mean of several estimates according to $\theta^{q+1} = \frac{1}{L} \sum_{l=1}^{L} \hat{\theta}(\boldsymbol{x}^l, \boldsymbol{y})$, where $\boldsymbol{x}^l$ is an *a posteriori* realization of $\boldsymbol{X}$ conditionally on $\boldsymbol{Y}$. It can be shown that $\boldsymbol{X} \mid \boldsymbol{Y}$ is a non homogeneous Markov chain whose parameters can be computed with the forward and backward probabilities in Eq. (3). Accordingly, the parameters of the fuzzy class can then be estimated.

Due to the numerical approximation, the fuzzy HMC model with two hard classes tends to be a HMC with more classes $(F + 2)$, but with the parameters of those pdf depending only on the two hard classes.

Fig. 1. Excerpt of a Space Shuttle Sensor photograph (432×208), acquired in February 2^{nd} 1984, near the Parana River in Southern Brazil.



(a) Classical HMC (b) Fuzzy HMC ($F = 1$)

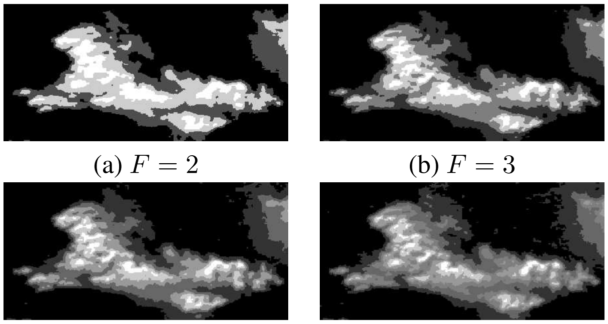
Fig. 2. Segmentations obtained with the classical HMC model and the new fuzzy HMC one ($F = 1$ fuzzy class).

5. SEGMENTATION

The new fuzzy HMC model has been tested on the clouds image in Fig. 1. This image is undoubtedly well suited to the fuzzy model presented here since the sky and the opaque cloud can be considered as hard classes, whereas the spots where the sky can be seen through clouds can be considered as the fuzzy class. It should be noted that the 2-classes segmentation task is really not obvious even for a human observer.

Both HMC and fuzzy HMC models parameters have been estimated with one hundred of ICE iterations. The image classification was performed with respect to the fuzzy MPM classifier, detailed in [6].

Fig. 2-(a) presents the segmentation obtained with a 3-classes classical HMC model. We can first see that the global shape of the cloud is not precisely segmented and is quite rough. Furthermore, we can constat that the hard class, corresponding to the opaque zone of the cloud, is also not precisely detected. Fig. 2-(b) shows the segmentation result obtained with the fuzzy HMC model and one fuzzy "discrete" class ($F = 1$, $\varepsilon = 0.5$ on $]0, 1[$). As we can first observe, the global shape of the cloud seems to be much more accurate. Furthermore, the hard class, corresponding to the opaque zone of the cloud, seems to be in respect with the real one. The resulting image confirms the interest of



(a) $F = 2$ (b) $F = 3$

(c) $F = 4$ (d) $F = 5$

Fig. 3. Segmentations obtained with the new fuzzy HMC model for different numbers of fuzzy classes F.

F fuzzy classes	0	1	2	3	4	5
Time (sec.)	41	72	114	178	250	327

Table 1. Computation time for different numbers of fuzzy classes F.

the fuzzy HMC model.

Fig. 3 presents segmentation results for different number of "discrete" fuzzy classes F. It implies different values of fuzzy measure, i.e. different values of ε, and so different corresponding fuzzy classes. For example, $F = 2$ implies $\varepsilon \in \{0.25, 0.75\}$, $F = 3$ implies $\varepsilon \in \{0.165, 0.495, 0.825\}$, ...

We can observe that bigger is the number of "discrete" fuzzy classes F, more accurate is the fuzzy segmentation. Each fuzzy class can be interpreted as the measure of imprecision between "sky" and "cloud" on the considered pixel. For example, the fuzzy class corresponding to $\varepsilon = 0.1$ is very close to the hard class 0, and could be considered as so. In this work, we are not concerned by the choice of a threshold or others hardening methods. Indeed, fuzzy HMC model seems to be useful in situation where the aim is to detect and characterize mixed areas.

Due to numerical approximations, the computational complexity involved in the model is quite lower than the one involved in the classical HMC. Table 1 presents the computational time according to the number of fuzzy classes F (50 ICE iterations), for the presented results.

Let us specify one possible application of such segmentation of clouds. An important problem in meteorology is to automatically classify clouds. One could imagine that different kinds of clouds would be characterized by different parameters. As the parameter estimation is automated, it becomes possible to perform an automated classification of clouds from the estimates so obtained.

6. CONCLUSION

In this work, we described a new fuzzy HMC model, with application to unsupervised image segmentation. The main contribution of this work is the simultaneous use of fuzzy and statistical methods in a HMC model and the use of the fuzzy extension of the Baum-Welch probabilities for parameters estimation.

Experiments on a real image confirm the interest of the fuzzy classification, which seems to be very performing in situation where the aim is to detect and characterize mixed area. As it has been explained in this work, fuzzy and hard segmentations are not competing but correspond to two different situations. This new fuzzy HMC model is able to cope with fuzzy situations, where the classical HMC failed.

From this work, we plan to study others densities h than the uniform one and try to apply this new model to synthetic aperture radar images. The extension of the model to K hard classes could also be of interest.

7. REFERENCES

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

[2] W. Skarbek, "Generalized Hilbert scan in image printing," in *Theoretical Foundations of Computer Vision*, R. Klette and W. G. Kropetsh, Eds. Akademik Verlag, Berlin, 1992.

[3] K.E. Avrachenkov and E. Sanchez, "Fuzzy Markov chains," *Fuzzy Optimization and Decision Making*, vol. 1, no. 2, pp. 143–159, June 2002.

[4] M.A. Mohamed and P. Gader, "Generalized hidden Markov models-Part I: Theorical frameworks," *IEEE Trans. on Fuzzy Syst.*, vol. 8, no. 1, pp. 67–81, 2000.

[5] H. Caillol, W. Pieczynski, and A. Hillion, "Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation," *IEEE Trans. on Im. Proc.*, vol. 6, no. 3, pp. 425–440, 1997.

[6] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graph. Mod. and Im. Proc.*, vol. 59, no. 4, pp. 205–220, July 1997.

[7] S. Ruan, B. Moretti, J. Fadili, and D. Bloyet, "Fuzzy Markovian segmentation in application of magnetic resonance images," *Comp. Vis. and Im. Under.*, vol. 85, pp. 54–69, 2002.