REVUE DE STATISTIQUE APPLIQUÉE

W. PIECZYNSKI
J.-M. CAHEN

Champs de Markov cachés flous et segmentation d'images

Revue de statistique appliquée, tome 42, n^o 3 (1994), p. 13-31.
<http://www.numdam.org/item?id=RSA_1994__42_3_13_0>

© Société française de statistique, 1994, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/legal.php). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 1994, **XLII** (3), 13-31

CHAMPS DE MARKOV CACHÉS FLOUS ET SEGMENTATION D'IMAGES

W. Pieczynski, J.-M Cahen

Département Signal et Image, Institut National des Télécommunications
9 rue Charles Fourier 91011 Evry Cédex - France

RÉSUMÉ

Notre travail traite du problème de la segmentation statistique floue d'images. Nous proposons une modélisation markovienne du champ aléatoire flou des classes généralisant le modèle classique des champs aléatoires des classes dures. Nous montrons que l'échantillonneur de Gibbs peut être adapté au modèle flou et permet la simulation des réalisations du champs des classes selon la loi *a priori* et *a posteriori*. Cette possibilité permet d'introduire une méthode de segmentation floue généralisant la démarche utilisant le maximum des lois marginales a posteriori (MPM) dans le cas dur. Les images floues simulées sont plus riches visuellement et dans le cas où les données sont floues la méthode de segmentation floue proposée améliore la qualité de la segmentation classique.

Mots clés : *Champs aléatoires de Markov, champs aléatoires flous, segmentation statistique floue d'images, échantillonneur de Gibbs flou.*

SUMMARY

This work deals with the statistical fuzzy image segmentation. We propose a new fuzzy Markovian modelling generalizing the classical hard one. It is shown that the Gibbs sampler can be adapted and it is possible to simulate realizations of the fuzzy class field according to the prior distribution as well as to the posterior one. This possibility allows us to propose a statistical fuzzy segmentation method which generalizes the maximisation of posterior marginals (MPM) algorithm used in the hard case. Fuzzy images simulated are visually satisfactory. Furthermore, when the real data are fuzzy, the method proposed is more efficient that the classical hard one.

Key words : *Markov random fields, fuzzy random fields, fuzzy statistical image segmentation, fuzzy Gibbs sampler.*

1. Introduction

La segmentation fait partie des problèmes clés se posant en traitement d'images. Parmi les nombreuses méthodes la famille des approches statistiques s'avère être d'une remarquable efficacité dans certains cas. L'application d'une telle méthode suppose l'utilisation d'un modèle probabiliste : pour S l'ensemble des pixels on considère deux collections de variables aléatoires $X = (X_s)_{s \in S}, Y = (Y_s)_{s \in S}$

dites «champs aléatoires». L'image à segmenter est supposée être une réalisation $Y = y$ du champs Y et la segmentation recherchée est une réalisation invisible $X = x$ du champs X. Chaque Y_s est à valeurs dans R^d, avec d nombre naturel, et en segmentation classique, que nous appellerons «dure» dans la suite, chaque X_s est à valeurs dans un ensemble fini $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ des classes. La segmentation floue utilise une modélisation plus générale dans la mesure où elle permet de tenir compte de l'existence des classes «intermédiaires». Considérons, à titre d'exemple, le problème de segmentation d'une image satellitaire binaire : on cherche à associer à chaque pixel $s \in S$, à partir de $Y = y$, une étiquette dans l'ensemble $\Omega = \{\text{ville, forêt}\}$. La segmentation dure ne permet pas de déceler les pixels sur lesquels se trouvent simultanément des arbres et des maisons. La modélisation correspondant à la segmentation floue, dont l'objectif est de pouvoir retrouver de tels pixels, suppose que chaque X_s prend ses valeurs dans $[0, 1]$. $X_s = 0$ correspond à la classe «ville», $X_s = 1$ à la classe «forêt» et $X_s = x \in]0, 1[$ désigne les pixels où la proportion de la classe «ville» est $1 - x$ et celle de la classe «forêt» x. Dans le cas général l'ensemble $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ est ainsi remplacé par $[0, 1]^m$, chaque ω_i correspond à $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ où 1 se trouve à la $i$ème place.

La segmentation Bayésienne est l'outil de la segmentation dure généralement retenu. Il est possible de définir plusieurs méthodes, chacune d'entre elles nécessitant la connaissance de certains paramètres de la loi de (X, Y). La plus simple, dite aveugle, consiste en estimation de la réalisation de chaque X_s à partir du seul Y_s. Sa mise en œuvre est possible dès qu'on connaît la loi de (X_s, Y_s). Les méthodes contextuelles consistent en estimation de chaque X_s à partir de Y_V, la restriction de Y à un voisinage V de s et nécessitent la connaissance de la loi de (X_s, Y_V).

L'introduction des champs de Markov cachés a permis la conception des algorithmes approchant les solutions des méthodes globales : la réalisation de chaque X_s est estimée à partir de toute l'information disponible $Y = y$. Dans la modélisation utilisée dans cette démarche X est un champ markovien et les variables aléatoires (Y_s) sont indépendantes conditionnellement à X.

Notons N le cardinal de l'ensemble des pixels S et supposons, afin de simplifier, les Y_s à valeurs dans R. Dans le cas de la segmentation dure X prend ainsi ses valeurs dans Ω^N et dans le cas de la segmentation floue ces valeurs sont dans $[0, 1]^{mN}$, le champ Y prenant ses valeurs dans R^N dans les deux cas. Lorsqu'on cherche à généraliser les diverses segmentations dures au cas flou on est ainsi amené à définir la loi de (X, Y), donnée généralement par la loi de X et les lois de Y conditionnelles à X.

Les ensembles flous ont été appliqués avec succès à la classification d'images dans un cadre déterministe ([1], [5], [7], [8], [10], [13], [18]). Les modélisations faisant intervenir simultanément les ensembles flous et la théorie des probabilités ont également été proposées ([3], [4], [14], [22]). L'originalité des résultats exposés dans [3], [4] se situe au niveau de l'utilisation des mesures comportant des masses de Dirac et des parties continues. Les lois marginales (lois des X_s) sont modélisées dans [3], [4] par une mesure sur $[0, 1]$ comportant deux masses de Dirac en 0 et 1 et une partie continue, définie par une densité par rapport à la mesure de Lesbegue, sur $]0, 1[$. Cette modélisation a permis la conception des algorithmes de segmentation non supervisée floue (*i.e.* contenant un estimateur des paramètres nécessaires à la mise en œuvre de la segmentation) dont l'efficacité, dans le cas où les données réelles sont floues, améliore celle des algorithmes analogues de la segmentation dure. Nous étendons dans cet article ce type de modélisation à la loi globale de X en introduisant les champs markoviens flous. Cette modélisation permet la conception des méthodes globales floues, généralisant les méthodes de type «maximum a posteriori» (MAP) ou «maximum a posteriori des lois marginales» (MPM) en segmentation dure. Une de ces méthodes est décrite et nous montrons son bon comportement par rapport à l'algorithme de segmentation dure MPM.

L'organisation de l'article est la suivante :

Dans la section suivante nous introduisons les champs markoviens cachés flous et montrons que les champs markoviens cachés durs en sont un cas particulier. Nous présentons également une version floue de l'échantillonneur de Gibbs dont le bon comportement, connu dans le cas dur, est préservé.

Cet échantillonneur permet la conception d'une méthode globale de segmentation floue que nous présentons dans la section trois.

Les résultats des simulations se trouvent dans la quatrième section.

La cinquième et dernière section contient les conclusions.

2. Champs de Markov cachés flous

2.1. La loi de X

Considérons le cas de deux classes $\Omega = \{0, 1\}$, notons S l'ensemble des pixels et N son cardinal. Supposons X, qui est à valeurs dans Ω^N, markovien relativement aux quatre plus proches voisins et tel que $P[X = x] \neq 0$ pour tout x dans Ω^N. On peut alors écrire :

$$P[X = x] = ce^{-U(x)} \tag{1}$$

où U, dite «énergie», est une somme des fonctions définies sur les cliques, qui sont ici soit des singletons soit des couples des pixels voisins. Si on admet que ces fonctions ne dépendent pas de la position de la clique considérée dans l'ensemble des pixels l'énergie est entièrement déterminée par trois fonctions $\varphi_v : \Omega^2 \rightarrow R$ (v pour les voisins verticaux), $\varphi_h : \Omega^2 \rightarrow R$, ($h$ pour les voisins horizontaux) et $\varphi_1 : \Omega \rightarrow R$ (pour les singletons). L'énergie U s'écrit :

$$U(x) = \sum_{\substack{t,s \text{ voisins} \\ \text{verticaux}}} \varphi_v(x_t, x_s) + \sum_{\substack{t,s \text{ voisins} \\ \text{horizontaux}}} \varphi_h(x_t, x_s) + \sum_{s \in S} \varphi_1(x_s) \tag{2}$$

Notons que cette modélisation ne permet pas de tenir compte des anisotropies éventuelles de l'image selon les bissectrices. De telles anisotropies peuvent être prises en compte en considérant la markovianité par rapport aux huit plus proches voisins; apparaissent alors des cliques contenant des pixels voisins selon les bissectrices. Si on suppose que les fonctions définies sur les cliques de cardinal supérieur à deux sont nulles l'énergie U s'écrit :

$$\begin{aligned} U(x) = & \sum_{\substack{t,s \text{ voisins} \\ \text{verticaux}}} \varphi_v(x_t, x_s) + \sum_{\substack{t,s \text{ voisins} \\ \text{horizontaux}}} \varphi_h(x_t, x_s) \\ & + \sum_{\substack{t,s \text{ voisins} \\ \text{selon 1re} \\ \text{bissectrice}}} \varphi_{b_1}(x_t, x_s) + \sum_{\substack{t,s \text{ voisins} \\ \text{selon 2me} \\ \text{bissectrice}}} \varphi_{b_2}(x_t, x_s) + \sum_{s \in S} \varphi_1(x_s) \end{aligned} \tag{3}$$

Nous allons exposer le principe de notre généralisation de tels champs de Markov au cas flou dans le cas le plus simple, la démarche proposée étant immédiatement généralisable aux champs définis par (3).

Considérons l'énergie de la forme :

$$U(x) = \sum_{t,s \text{ voisins}} \varphi(x_t, x_s) \tag{4}$$

avec φ de la forme :

$$\varphi(x_t, x_s) = \begin{cases} -\alpha & \text{si} \quad x_t = x_s \\ \alpha & \text{si} \quad x_t \neq x_s \end{cases} \tag{5}$$

Le paramètre α détermine ainsi la loi de X.

L'énergie considérée est d'une grande simplicité, ce qui permet d'exposer le principe de passage aux champs flous dans un cadre utilisant peu de paramètres.

L'objectif de cette section est de généraliser ce modèle de façon à ce que chaque X_s prenne ses valeurs dans $\Omega_f = [1, 0]$(f pour «flou») et que sa loi soit de la forme :

$$P_{X_s} = h(\delta_0 + \delta_1 + \mu) = h\nu \tag{6}$$

où δ_0, δ_1 sont des masses de Dirac en 0 et 1, μ la mesure de Lesbegue sur $]0, 1[$ et h la densité de la loi P_{X_s} par rapport à la mesure $\nu = \delta_0 + \delta_1 + \mu$. Ainsi :

$$\begin{aligned} P_{X_s}[\{0\}] &= h(0) \\ P_{X_s}[\{1\}] &= h(1) \\ P_{X_s}[[a, b]] &= \int_a^b h(t)dt \end{aligned} \tag{7}$$

la dernière égalité ayant lieu pour $0 < a \leqslant b < 1$.

L'intérêt des modélisations markoviennes en segmentation dure est essentiellement dû à la possibilité de la simulation, par des procédures itératives de type de l'échantillonneur de Gibbs, des réalisations de X. Le cardinal de S étant trop grand la relation (1) n'est en effet pas directement exploitable (c inconnu). Notons que les difficultés sont encore plus grandes en segmentation floue. Si on cherche à définir la loi d'un couple (X_t, X_s), on est amené à considérer la densité $h(x_t, x_s)$ de la loi de (X_t, X_s) par rapport à la mesure

$$\nu^2 = \nu \otimes \nu = (\delta_0 + \delta_1 + \mu) \otimes (\delta_0' + \delta_1' + \mu') = \delta_0 \otimes \delta_0' + \delta_0 \otimes \delta_1' + \\ + \delta_0 \otimes \mu' + \delta_1 \otimes \delta_0' + \delta_1 \otimes \delta_1' + \delta_1 \otimes \mu' + \mu \otimes \delta_0' + \mu \otimes \delta_1' + \mu \otimes \mu'$$

La mesure ν^2 comporte ainsi quatre masses de Dirac aux points $(0,0), (0,1), (1,0), (1,1)$, quatre mesures de Lebesgue définies sur les côtés du carré unité et une mesure de Lebesgue sur le carré $]0,1[^2$. La loi d'un triplet (X_s, X_t, X_u) est définie par une densité par rapport à la mesure ν^3, cette dernière étant définie par huit masses de Dirac sur les sommets du cube unité, douze mesures de Lebesgue définies sur les segments reliant ces sommets, six mesures de Lebesgue définies sur les faces du cube et une mesure de Lebesgue définie sur son intérieur. On constate que ν^N est difficile à manipuler pour $N \geqslant 3$. Notons que tel n'est pas le cas en segmentation dure, ν est dans ce cas la mesure de dénombrement et ν^N est accessible pour $N \leqslant 7$ environ. Ce fait peut avoir une importance en segmentation contextuelle. Dans le cas dur il est possible de considérer plusieurs voisins dont le nombre peut aller, dans le cas de deux classes, jusqu'à quatre. En segmentation bayésienne floue il semble difficile de pouvoir tenir compte de plus d'un voisin.

Afin de généraliser le modèle ci-dessus considérons

$$P_X = h_f \nu^N \tag{8}$$

avec $\nu = \delta_0 + \delta_1 + \mu$ et h_f(f pour «flou») densité de P_X par rapport à ν^N. Comme dans le cas dur considérons h_f de la forme

$$h_f(x) = ce^{-\sum_{t,s \text{ voisins}} \varphi_f(x_t, x_s)} \tag{9}$$

φ «dure» est définie sur $\{0,1\}^2$ et φ_f «floue» ci-dessus sur $[0,1]^2$.

On montre, exactement comme dans le cas des champs durs, que la loi de X définie par (8) et (9) est une loi d'un champs markovien relativement aux quatre plus proches voisins : la loi de X_s conditionnelle à $(X_t)_{t \neq s}$ est égale à sa loi conditionnelle à X_1, X_2, X_3, X_4, où $1, 2, 3, 4$ sont les quatre plus proches voisins de s. Plus précisément, la loi de X_s conditionnelle à $(X_1, X_2, X_3, X_4,) = (x_1, x_2, x_3, x_4) = x_V$ (V pour «voisinage») est définie par la densité par rapport à la mesure ν donnée par :

$$h^{x_V}(x) = \frac{e^{-W_{x_V}(x)}}{e^{-W_{x_V}(0)} + e^{-W_{x_V}(1)} + \int_0^1 e^{-W_{x_V}(t)}\, dt} \tag{10}$$

avec $W_{x_V}(x) = \varphi_f(x, x_1) + \varphi_f(x, x_2) + \varphi_f(x, x_3) + \varphi_f(x, x_4)$.

Finalement, la fonction $\varphi_f : [0,1]^2 \rightarrow R$ définit P_X, de plus les lois conditionnelles données par (10) sont exploitables. On peut en effet effectuer des tirages selon cette loi de la façon suivante :

- Tirage dans $\{0, 1, F\}$ (F pour «flou») selon la probabilité $h^{xv}(0), h^{xv}(1)$, $1 - h^{xv}(0) - h^{xv}(1)$. Si le tirage est dans $\{0, 1\}$ on arrête, s'il vaut F :

- Tirage dans $]0, 1[$ selon la probabilité donnée par la densité par rapport à la mesure de Lebesgue :

$$g^{xv}(u) = \frac{e^{-W_{xv}(u)}}{\displaystyle\int_0^1 e^{-W_{xv}(t)}\,dt} \tag{11}$$

Dans la pratique le calcul de l'intégrale figurant dans (10) et (11) se fait par discrétisation : on choisit $x_1, x_2, \dots, x_n$ une subdivision équidistante de $]0, 1[$, avec $x_1 = a, x_{i+1} - x_i = a, x_n = 1 - a$, (donc $a = \frac{1}{n+1}$) et on remplace cette intégrale par

$$\frac{1}{n}\sum_{j=1}^{n} e^{-W_{xV}(x_j)} \tag{12}$$

Le tirage dans $]0, 1[$ selon la probabilité donnée par la densité (11) est par ailleurs remplacé par le tirage dans $\{x_1, x_2, \dots, x_n\}$ selon la loi discrète obtenue à partir de (11). On peut alors remplacer le tirage «théorique» en deux phases décrit ci-dessus par un tirage unique dans l'ensemble $\{0, x_1, x_2, \dots, x_n, 1\}$ selon la loi de probabilité :

$$\begin{aligned}
h^{xv}(0) &= \frac{e^{-W_{xV}(0)}}{e^{-W_{xV}(0)} + e^{-W_{xV}(1)} + \frac{1}{n}\sum_{j=1}^{n} e^{-W_{xV}(x_j)}} \\
h^{xv}(x_i) &= \frac{\frac{1}{n} e^{-W_{xV}(x_i)}}{e^{-W_{xV}(0)} + e^{-W_{xV}(1)} + \frac{1}{n}\sum_{j=1}^{n} e^{-W_{xV}(x_j)}} \quad \text{pour } 1 \leqslant i \leqslant n \\
h^{xv}(1) &= \frac{e^{-W_{xV}(1)}}{e^{-W_{xV}(0)} + e^{-W_{xV}(1)} + \frac{1}{n}\sum_{j=1}^{n} e^{-W_{xV}(x_j)}}
\end{aligned} \tag{13}$$

Nous disposons ainsi d'un modèle markovien de la loi de X tel que les lois marginales des X_s contiennent deux masses de Dirac en 0 et 1 et une partie continue sur $]0, 1[$, de plus la simulation des réalisations des X_s selon leurs lois conditionnelles est possible.

2.2. Échantillonneur de Gibbs flou

La possibilité d'effectuer des tirages selon les lois conditionnelles décrites ci-dessus permet la simulation des réalisations de X par un échantillonneur de Gibbs «flou», dont le principe est le même que celui de l'échantillonneur de Gibbs couramment utilisé dans le cas dur. Dans ce dernier cas la démarche est la suivante :

1. On se donne une configuration arbitraire X^0

2. On numérote les pixels $1, 2, \ldots, N$. La loi de X_1 conditionnelle à la réalisation de tous les autres X_i, donnés par la configuration X^0, est explicitement calculable : on effectue un tirage, selon cette loi, dans l'ensemble $\{0, 1\}$. On remplace la classe figurant sur le pixel 1 dans la configuration X^0 par la classe obtenue par le tirage (qui peut être, éventuellement, la même). On passe au pixel 2 et on calcule la loi de X_2 conditionnelle à tous les autres X_i, en tenant compte de la nouvelle valeur, éventuellement différente de la valeur précédente, de X_1 . Comme précédemment, on effectue un tirage, selon cette loi, dans l'ensemble $\{0, 1\}$. On remplace la classe figurant sur le pixel 2 dans la configuration X^0 par la classe obtenue par le tirage, et ainsi de suite. Après N tirages on aura effectué un «balayage complet» de l'image, soit X^1 la nouvelle configuration ainsi obtenue. On recommence la procédure en remplaçant X^0 par X^1. Dans le cas des images simples l'aspect de l'image se stabilise après une quinzaine de balayages.

Nous proposons exactement la même démarche en remplaçant $\{0, 1\}$ par $[0, 1]$.

Dans le cas dur on montre que la suite aléatoire des champs $X^1, X^2, \ldots, X^n, \ldots$ ainsi construite converge en loi vers la loi de probabilité donnée par (1), [9]. Le résultat reste valable lorsque l'espace des états, *i.e.* l'ensemble dans lequel chaque X_s prend ses valeurs, est compact, [23]. Notre modélisation se situe dans ce cadre, en effet l'espace des états est $[0, 1]$. Notons que ce résultat reste valable lorsqu'on considère la loi de X *a posteriori* définie dans la section suivante.

2.3. Forme de U_f

Le contenu des paragraphes 2.1, 2.2 est valable pour une forme quelconque de U_f . Lorsqu'on cherche à généraliser le cas dur simple défini par (4) et (5) on peut imposer à U_f de prendre la valeur $-\alpha$ sur les points $(0, 0), (1, 1)$ et la valeur α sur les points $(0, 1), (1, 0)$. L'aspect «flou» sera alors défini par la restriction de U_f à $[0, 1]^2 - \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Supposons que cette restriction dépende de deux paramètres β et δ, ce qui nous permettra de gérer le dégradé du flou dans l'image d'une part et son importance d'autre part.

Considérons U_f de la forme suivante :

$$U_f(x) = \begin{cases} -\alpha \quad \text{si } (x, y) = (0, 0) \text{ ou } (x, y) = (1, 1) \\ \alpha \quad \text{si } (x, y) = (0, 1) \text{ ou } (x, y) = (1, 0) \\ -\beta(-2x + 2y + 1) + \delta \text{ si } (x, y) \in [0, 1]^2 - \{(0, 0), (0, 1), (1, 0), (1, 1)\} \\ \qquad \text{et } x - y \geqslant 0 \\ -\beta(2x - 2y + 1) + \delta \quad \text{si } (x, y) \in [0, 1]^2 - \{(0, 0), (0, 1), (1, 0), (1, 1)\} \\ \qquad \text{et } x - y \leqslant 0 \end{cases} \tag{14}$$

La loi de X dépend ainsi de trois paramètres α, β, δ. Le paramètre α détermine la taille des taches, β détermine la forme du dégradé et δ l'importance du flou dans l'image.

Plus précisément, montrons que l'on retrouve le modéle markovien dur lorsque $\delta \to +\infty$. Considérons $E_f \subset [0,1]^N$ défini par

$$[x = (x_1, \ldots, x_N) \in E_f] \Leftrightarrow [\exists 1 \leqslant i \leqslant N \quad \text{tel que} \quad x_i \in]0,1] \tag{15}$$

et E_d son complémentaire :

$$[x = (x_1, \ldots, x_N) \in E_d] \Leftrightarrow [\forall 1 \leqslant i \leqslant N, x_i \in \{0,1\}] \tag{16}$$

Nous allons montrer que le champ de Markov dur est un cas particulier, dans le sens où il peut être obtenu par passage à la limite, d'un champ de Markov flou. Plus précisément, lorsque $\delta \to +\infty$ la probabilité définie par la vraisemblance donnée par (9) et (14) se concentre sur E_d avec vitesse exponentielle.

Proposition 1.

Il existe une constante $b > 0$ telle que pour tout $\delta \geqslant |\beta| - |\alpha|$

$$P_X[E_d] \geqslant 1 - be^{-4\delta} \tag{17}$$

Démonstration

Faisons apparaître δ dans la vraisemblance h_f donnée par la relation (9) :

$$h_{f,\delta}(x) = c(\delta)e^{-U_{f,\delta}(x)} \tag{18}$$

Montrons :

$$c(\delta) \leqslant e^{-2N(N-1)|\alpha|} \tag{19}$$

Pour $x \in E_d$ la quantité $U_{f,\delta}(x)$ ne dépend pas de δ, par ailleurs $h_{f,\delta}(x)$ est la probabilité de $X = x$. On peut écrire :

$$c(\delta)e^{-U_f(x)} \leqslant 1 \tag{20}$$

Supposons $\alpha > 0$. La majoration ci-dessus étant valable pour tout $x \in E_d$ considérons x en forme de damier. Etant donné qu'il y a $N(N-1)$ cliques horizontales et autant de cliques verticales (20) donne :

$$c(\delta)e^{2N(N-1)\alpha} \leqslant 1 \tag{21}$$

d'où (19). Si $\alpha < 0$ on considère x dont toutes les coordonnées sont d'une même classe et on arrive à la même équation que (21) avec $-\alpha$ à la place de α. D'où (19) pour tout α.

Considérons $x \in E_f$. Il existe au moins une coordonnée x_i de x dans $]0,1[$. Les valeurs de la fonction φ_f sur les quatre cliques contenant x_i sont alors minorées par $\delta - |\beta|$. Sur toutes les autres cliques (t,s) les valeurs de $\varphi_f(x_t, x_s)$ sont minorées par $-|\alpha|$ si x_t, x_s sont durs et par $\delta - |\beta|$ (donc encore par$-|\alpha|$ car $\delta \geqslant |\beta| - |\alpha|$)) si l'un d'eux est flou. Finalement on peut écrire pour tout $x \in E_f$:

$$h_{f,\delta}(x) \leqslant c(\delta) e^{-4(\delta - |\beta|) + (2N(N-1)-4)|\alpha|} \tag{22}$$

Etant donné la majoration (19) nous avons :

$$h_{f,\delta}(x) \leqslant e^{-4\delta + 4(|\beta| - |\alpha|)} \tag{23}$$

Etant donné que $\nu^N([0,1]^N) = 3^N$ et $\nu^N(E_d) = 2^N$ la majoration (23) donne ($h_{h,\delta}$ est une densité de P_X par rapport à ν^N) :

$$P_X(E_f) \leqslant (3^N - 2^N) e^{-4\delta + 4(|\beta| - |\alpha|)} \tag{24}$$

d'où (17) avec $b = (3^N - 2^N) e^{4(|\beta| - |\alpha|)}$, ce qui achève la démonstration.

On peut également montrer que lorsque $\delta \to -\infty$ la loi de X se concentre sur E_f avec vitesse exponentielle.

Proposition 2.

Il existe une constante $b' > 0$ telle que pour tout $\delta \leqslant |\alpha| - |\beta|$

$$P_X[E_f] \geqslant 1 - b' e^{4\delta} \tag{25}$$

Démonstration

Considérons $x \in E_f$. Il existe au moins une coordonnée x_i de x dans $]0,1[$. Les valeurs de la fonction φ_f sur les quatre cliques contenant x_i sont alors majorées par $\delta + |\beta|$. Sur toutes les autres cliques (t,s) les valeurs de $\varphi_f(x_t, x_s)$ sont majorées par $|\alpha|$ si x_t, x_s sont durs et par $\delta + |\beta|$ (donc encore par $|\alpha|$ car $\delta \leqslant |\alpha| - |\beta|$) si l'un d'eux est flou. Finalement on peut écrire pour tout $x \in E_f$:

$$e^{-U_{f,\delta}(x)} \geqslant e^{-4(\delta + |\beta|) - (2N(N-1)-4)|\alpha|} \tag{26}$$

par ailleurs nous avons :

$$\begin{aligned} \frac{1}{c(\delta)} = \int_{[0,1]^N} e^{-U_{f,\delta}(x)}\, d\nu^N(x) &\geqslant \int_{E_f} e^{-U_{f,\delta}(x)}\, d\nu^N(x) \\ &\geqslant (3^N - 2^N) e^{-4(\delta + |\beta|) - (2N(N-1)-4)|\alpha|} \end{aligned} \tag{27}$$

la dernière inégalité dans (27) ayant lieu en vertu de (26). Finalement

$$c(\delta) \leqslant b'' e^{4\delta} \tag{28}$$

Par ailleurs

$$P_X[E_d] = \sum_{x \in E_d} c(\delta) e^{-U_f(x)} \tag{29}$$

avec $U_f(x)$ indépendant de δ. (28) et (29) impliquent :

$$P_X[E_d] \leqslant b'' e^{4\delta} \sum_{x \in E_d} e^{-U_f(x)} \tag{30}$$

d'où (25) avec $b' = b'' \sum\limits_{x \in E_d} e^{-U_f(x)}$, ce qui achève la démonstration.

Les exemples d'images simulées par l'échantillonneur flou pour $\delta = 0$ et différentes valeurs de α, β sont présentés dans la section 3.

2.4. Champs cachés

L'objectif de cette sous-section est de définir la loi du couple (X, Y). Nous pourrons alors en déduire la loi «*a posteriori*», *i.e.* conditionnelle à Y, de X et proposer une méthode de segmentation globale floue. La loi de (X, Y) peut être définie, comme dans le cas dur, par la loi de X et les lois de Y conditionnelles à X. La loi de X est définie ci-dessus et afin de définir les lois de Y conditionnelles à X nous faisons, comme on le fait habituellement dans le cas dur, deux hypothèses simplificatrices suivantes :

(i) les variables aléatoires (Y_s) sont indépendantes conditionnellement à X.

(ii) la loi de chaque Y_s conditionnelle à X est égale à sa loi conditionnelle à X_s.

Avec les hypothèses (i), (ii) la loi de Y conditionnelle à X est définie par les lois des Y_s conditionnelles à X_s. En supposant les lois des (X_s, Y_s) indépendantes de s et en notant $N(m, \sigma^2)$ la loi normale de moyenne m et variance σ^2 nous prendrons pour la loi de Y_s conditionnelle à $X_s = \varepsilon \in [0, 1]$:

$$N((1-\varepsilon)m_0 + \varepsilon m_1, (1-\varepsilon)\sigma_0^2 + \varepsilon \sigma_1^2) \tag{31}$$

où $m_0, m_1, \sigma_0^2, \sigma_1^2$ sont donnés. Ainsi les paramètres $m_0, m_1, \sigma_0^2, \sigma_1^2$ déterminent toutes les lois de Y conditionnelles à X.

Pour $m_0, m_1, \sigma_0^2, \sigma_1^2$ fixés notons Ψ_ε la densité gaussienne définie ci-dessus. La densité Ψ de la loi de (X, Y) par rapport à la mesure $\nu^N \otimes \mu^N$ (ν étant la mesure définie sur $[0, 1]$ dans la sous-section 2.1, μ la mesure de Lebesgue sur R et N le nombre des pixels) est alors donnée par :

$$\Psi(x, y) = k e^{-U_f(x)} \prod_{s \in S} \Psi_{x_s}(y_s) \tag{32}$$

3. Loi de X *a posteriori* et segmentation globale floue

3.1 Loi de X a posteriori

En notant $\prod_{s \in S} \Psi_{x_s}(y_s) = e^{\sum_{s \in S} \mathrm{Log} \Psi_{x_s}(y_s)} = e^{-V_x(y)}$

et $W_y(x) = U_f(x) + V_x(y)$, (32) s'écrit :

$$\Psi(x, y) = ke^{-W_y(x)} \tag{33}$$

La densité de la loi de X *a posteriori (i.e.* conditionnelle à $Y = y$) par rapport à la mesure ν^N est alors donnée par :

$$\Psi^y(x) = \frac{ke^{-W_y(x)}}{\displaystyle\int_{[0,1]^N} ke^{-W_y(x)} \, d\nu^N(x)} \tag{34}$$

Le dénominateur de l'expression (34) ne dépend pas de x, y étant fixé on peut écrire :

$$\Psi^y(x) = k(y)e^{-W_y(x)} = k(y)e^{-(U_f(x)+V_x(y))} \tag{35}$$

Nous constatons que l'énergie apparaissant dans (35) est du même type que celle donnée par (2). La différence entre la loi *a priori* donnée par (2) et la loi *a posteriori* donnée par (35) est due à l'existence de l'énergie additionnelle $V_x(y)$, qui est ici fonction de x, y étant fixé. Rappelons que $V_x(y)$ est de la forme :

$$V_x(y) = -\sum_{s \in S} \mathrm{Log} \, \Psi_{x_s}(y_s) \tag{36}$$

où Ψ_{x_s} est une gaussienne de moyenne $(1 - x_s)m_0 + x_s m_1$ et de variance $(1 - x_s)\sigma_0^2 + x_s\sigma_1^2$. Comme dans le cas dur notre méthode de segmentation est fondée sur la possibilité de la simulation des réalisations de X selon sa loi *a posteriori*. L'énergie additionnelle donnée par la relation (36) étant une somme des fonctions rattachées aux singletons (rappelons que y_s est fixé et x_s variable) le caractère markovien du champ X, lorsqu'on considère sa distribution *a posteriori*, est préservé. On peut ainsi utiliser l'échantillonneur de Gibbs flou de la sous-section 2.2. Notons $h_{y_s}^{x_V}$ la densité, par rapport à la mesure ν, de la loi de X_s conditionnelle à $X_V = x_V$, où V est le voisinage de x_s composé des quatre plus proches voisins. La densité $h_{y_s}^{x_V}$ est donnée par une formule analogue à (10) qui correspond à la loi de X *a priori* : $W_{x_V}(x)$ utilisé dans (10) est remplacé par :

$$H_{x_V, y_s}(x_s) = W_{x_V}(x_s) - \mathrm{Log} \, \Psi_{x_s}(y_s) \tag{37}$$

Nous pouvons écrire :

$$h_{y_s}^{x_V}(x_s) = \frac{e^{-H_{x_V,y_s}(x_s)}}{e^{-H_{x_V,y_s}(0)} + e^{-H_{x_V,y_s}(1)} + \int_0^1 e^{-H_{x_V,y_s}(t)}\,dt} \tag{38}$$

Les simulations sont alors obtenues en utilisant la même procédure de discrétisation que celle décrite dans la sous-section 2.1.

3.2 Segmentation globale floue

La méthode proposée correspond à l'algorithme de Marroquin et *al.* ([16]) en segmentation dure, ce dernier permettant d'approcher la solution de la méthode MPM. La segmentation MPM s'intéresse aux lois des X_s conditionnelles à Y qui sont les lois marginales *a posteriori* du champ X. Pour chaque s dans S on choisit pour l'estimation de la réalisation invisible de X_s la classe dont la probabilité *a posteriori* est maximale.

Nous nous intéressons, comme dans le cas de la méthode MPM en segmentation dure, aux lois marginales *a posteriori* du champ X. Pour chaque s dans S la loi *a posteriori* de X_s est estimée à partir des réalisations simulées par l'échantillonneur de Gibbs flou. Elle est donnée par une densité par rapport à la mesure ν que nous noterons $h_{s,y}$, avec $Y = y$ l'image numérique à segmenter. Contrairement au cas dur il existe alors plusieurs possibilités de segmentations de types différents. Nous avons étudié dans [4] la segmentation floue aveugle, où la réalisation de chaque X_s est estimée à partir du seul Y_s. Ici notre démarche est différente car nous considérons les lois des X_s conditionnelles à Y au lieu des lois des X_s conditionnelles aux Y_s. Toutefois, au stade de la segmentation le problème est identique : on doit proposer une stratégie de recherche de la réalisation x_s de X_s à partir de la densité de la loi de X_s par rapport à la mesure ν. Nous avons proposé dans [4] quatre stratégies de segmentation : moindres carrés MC, espérance conditionnelle EC, maximum de vraisemblance MV et maximum de vraisemblance «adaptée» MVA. Nous retenons dans la suite la stratégie MVA dont l'efficacité s'est avérée compétitive vis-à-vis des trois autres méthodes dans le cas de segmentation aveugle. Sa démarche est la suivante :

(i) choix dans $\{0, 1, F\}$ (F pour «flou») selon la règle bayésienne classique : l'élément choisi maximise la probabilité $h_{s,y}(0), h_{s,y}(1), 1 - h_{s,y}(0) - h_{s,y}(1)$.

(ii) si l'élément choisi est dans $\{0, 1\}$ on arrête, sinon on choisit dans $]0, 1[$ l'élément maximisant la restriction de $h_{s,y}$ à cet ensemble.

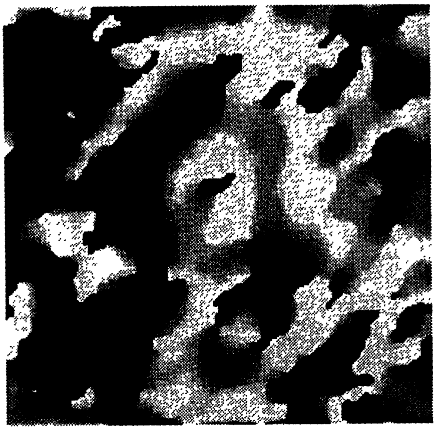
4. Expérimentations

4.1. Champs durs et champs flous

Dans cette sous-section nous présentons, afin d'illustrer l'apport visuel de la modélisation par champs flous, deux images simulées par l'échantillonneur de Gibbs dur et flou respectivement. Le paramètre α vaut 10 dans les deux cas et, dans le cas flou, les paramètres β, δ valent 1 et 0 respectivement. On obtient, après $M = 100$ itérations, les images Im1, Im2 respectivement. Dans les légendes M désigne le nombre d'itérations de l'échantillonneur de Gibbs (dur ou flou).

Champs de Gibbs dur avec $\alpha = 10$, $M = 100$

Im1



Champs de Gibbs flou avec $\alpha = 10$, $\beta = 1,\ \delta = 0,\ M = 100$.

Im2

On observe un bon dégradé, lorsqu'on passe d'une classe dure à une autre, dans l'image floue obtenue. Notons que des images floues d'aspect différent pourraient être obtenues en modifiant l'allure générale de U_f ou les paramètres β, δ.

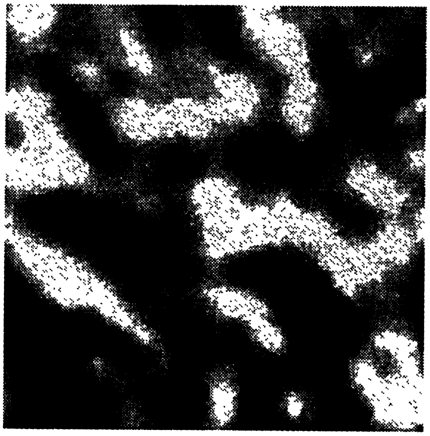
4.2. Images floues bruitées

Nous nous intéressons dans cette sous-section à l'aspect visuel des images floues bruitées. L'image floue Im3, obtenue par l'échantillonneur de Gibbs flou, correspond aux paramètres $\alpha = 5, \beta = 1, \delta = 0$. Nous présentons plusieurs bruitages par des bruits blancs ou corrélés. Notons que notre modélisation ne permet pas de tenir compte, au niveau de la segmentation, de la corrélation spatiale du bruit. Ainsi lorsqu'on segmente une image bruitée par un bruit corrélé par une méthode globale (dure ou floue) fondée sur une modélisation markovienne on est amené à ignorer la corrélation, ce qui revient à changer de modèle. L'étude des situations où le bruit est corrélé peut ainsi apporter des renseignements sur la robustesse d'une méthode donnée. L'étude de cette robustesse est importante car des études précédentes ont montré que dans certains cas des images réelles, comme les images satellitaires, la corrélation du bruit peut être importante ([15], [17]).

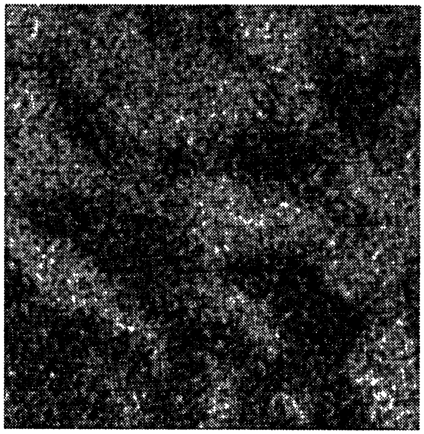
Les bruits corrélés sont obtenus à partir des bruits blancs en utilisant la notion de moyenne mobile. Si $B = (B_s)_{s \in S}$ est un bruit blanc gaussien on pose :

$$(C_s) = \gamma \sum_{t \in V_s} B_t \tag{39}$$

avec γ tel que Var (C_s) = Var (B_s). Nous considérons deux types de bruits corrélés : type 1 est obtenu à partir d'un bruit blanc en considérant les voisinages composés des 8 plus proches voisins et le type 3 correspond aux voisinages contenant 48 plus proches voisins. Pour t, s voisins la corrélation entre C_t et C_s est ainsi de $\frac{6}{9}$ dans le cas du type 1 et de $\frac{42}{49}$ dans le cas du type 3. Le type 2, que nous n'étudions pas ici, correspond au cas des 24 plus proches voisins, la corrélation entre C_t et C_s est alors de $\frac{20}{25}$.

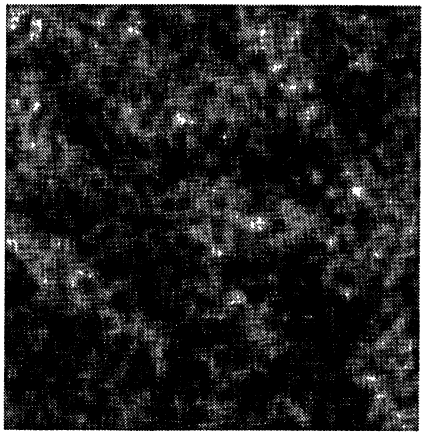


Champs de Gibbs flou avec $\alpha = 5$, $\beta = 1,\ \delta = 0,\ M = 100$
Im3

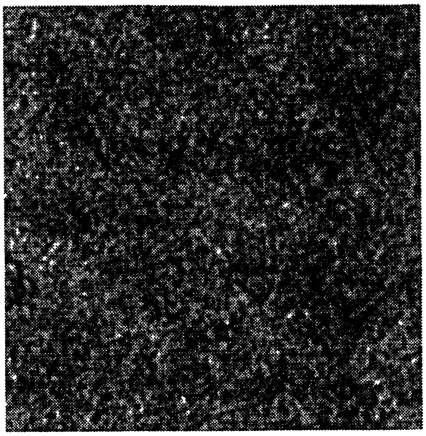


Im3 bruitée avec bruit corrélé de type1 :
$m_0 = 1,\ m_1 = 4,\ \sigma_0 = \sigma_1 = 1$
Im4

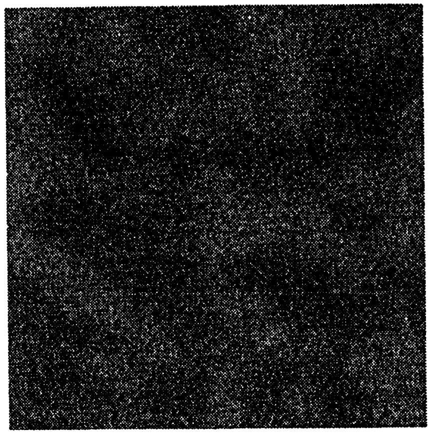
Im3 bruitée avec bruit indépendant
$m_0 = 1,\ m_1 = 4,\ \sigma_0 = \sigma_1 = 1$
Im5



Im3 bruitée avec bruit corrélé de type 3 :
$m_0 = 1,\ m_1 = 4,\ \sigma_0 = \sigma_1 = 1.$
Im6



Im3 bruitée avec bruit corrélé de type 1, $m_0 = 1, m_1 = 2, \sigma_0 = \sigma_1 = 1.$
Im7

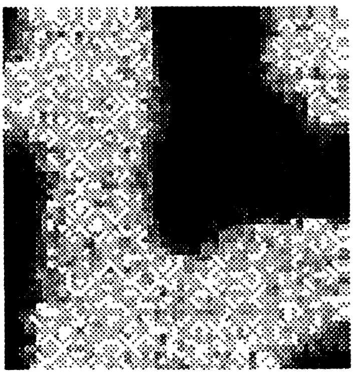


Im3 bruitée avec bruit indépendant, $m_0 = 1, m_1 = 2, \sigma_0 = \sigma_1 = 1.$
Im8

Nous constatons que l'aspect visuel des bruitages proposés est différent.

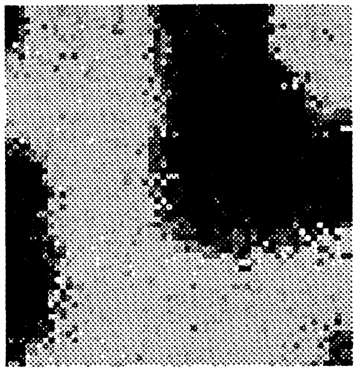
4.3. Segmentations dure et floue

Nous considérons dans cette sous-section les résultats de la segmentation dure et floue. La segmentation floue étant coûteuse en temps calcul nous considérons des images de taille plus réduite. L'image Im9 correspond aux paramètres $\alpha = 3, \beta = 1, \delta = 0$ et l'image Im14 aux paramètres $\alpha = 10, \beta = 5, \delta = 0$. M est le nombre d'itérations utilisé dans la synthèse de l'image par l'échantillonneur de Gibbs flou et L le nombre de balayages utilisés dans l'estimation des lois marginales *a posteriori*.



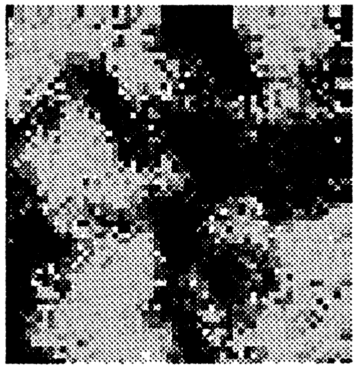
Champs de Gibbs flou avec $\alpha = 3, \beta = 1, \delta = 0, M = 150.$
Im9

Im9 restaurée, après bruitage indépendant
$m_0 = 1,\ m_1 = 2,\ \sigma_0 = \sigma_1 = 1,$
par le MPM dur avec L = 50.
Im10

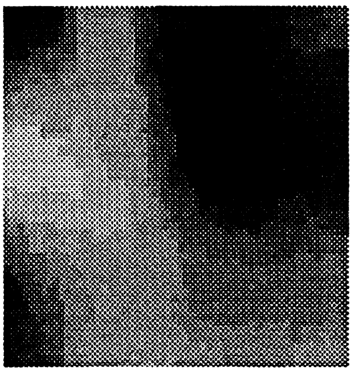


Im9 restaurée, après bruitage indépendant
$m_0 = 1,\ m_1 = 2,\ \sigma_0 = \sigma_1 = 1,$
par le MPM flou avec L= 50.
Im11

Im9 restaurée, après bruitage corrélé type 1
$m_0 = 1,\ m_1 = 2,\ \sigma_0 = \sigma_1 = 1,$
par le MPM dur avec $L = 50$.
Im12

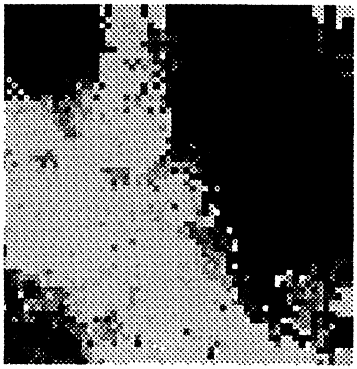


Im9 restaurée, après bruitage corrélé type 1
$m_0 = 1,\ m_1 = 2,\ \sigma_0 = \sigma_1 = 1,$
par le MPM flou avec $L = 50$
Im13



Champs de Gibbs flou avec $\alpha = 10, \beta = 5, \delta = 0, M = 100.$
Im14

Im14 restaurée, après bruitage corrélé de type 1, $m_0 = 1$, $m_1 = 2$, $\sigma_0 = \sigma_1 = 1$, par le MPM dur avec $L = 50$.
Im15



Im14 restaurée, après bruitage corrélé de type 1, $m_0 = 1$, $m_1 = 2$, $\sigma_0 = \sigma_1 = 1$, par le MPM flou avec $L = 50$.
Im16

Nous constatons sur ces exemples que la segmentation floue aboutit aux résultats nettement plus proches, visuellement, des images de départ que la segmentation dure. La qualité d'une segmentation est parfois évaluée par le pourcentage des pixels bien classés. $\widehat{x} = (\widehat{x}_s)$ étant le résultat de la segmentation et N le nombre des pixels on pose

$$ER = \frac{1}{N} \sum_{s \in S} |x_s - \widehat{x}_s| \tag{40}$$

On peut retenir le même critère de qualité dans le cas flou. En appliquant ce critère aux segmentations dures et floues des images Im9, Im14 bruitées par le bruit blanc de moyennes 1, 2 et de variance 1 on obtient

	Seg Dure	Seg Floue
Im9	0.13	0.11
Im14	0.32	0.22

L'amélioration est surtout sensible dans le cas de Im14, ce qui est normal étant donné que cette image contient une plus forte proportion de pixels flous que Im9.

5. Conclusion

La segmentation fait partie des problèmes clé en imagerie. Les méthodes statistiques, dont le développement date d'une quinzaine d'années, s'avèrent d'une remarquable efficacité dans certaines situations. Ces méthodes utilisent la théorie de la classification bayésienne et s'appuient, de ce fait, sur des modélisations probabilistes rigoureuses. Toutefois, ce cadre peut s'avérer insuffisant dans certaines situations réelles. Lorsqu'un pixel appartient naturellement à plusieurs classes une méthode de segmentation classique, ou dure, assimilera cette situation à un bruitage et assignera à ce pixel une seule classe.

Les modélisations probabilistes utilisant les ensembles flous s'avèrent utiles dans ce type de situations. Nous avons proposé dans ce travail un modèle markovien flou caché. Ce modèle apparaît comme une généralisation du modèle markovien classique dans la mesure où l'on retrouve ce dernier lorsqu'un paramètre tend vers l'infini. La technique classique MPM de segmentation dure est généralisable au modèle flou. Lorsque les données sont floues la méthode «MPM flou» permet d'obtenir de meilleurs résultats, aussi bien visuellement que numériquement.

Remerciements. Les auteurs remercient Alain Hillion, Professeur à L'École Nationale Supérieure des Télécommunications de Bretagne, dont les idées ont fortement contribué à la conception de ce travail.

Références

[1] BHARATHI DEVI B. & SARMA V.V.S. (1985) «Estimation of fuzzy memberships from histograms». Information Sciences 35, pp. 43-59.

[2] BRAATHEN B. & PIECZYNSKI W. & MASSON P. (1993) «Global and local methods of unsupervised Bayesian segmentation of images ». Machine Graphics & Vision, Vol. 2, No. 1, pp. 39-52.

[3] CAILLOL H. & PIECZYNSKI W. (1992) «Fuzzy statistical unsupervised image segmentation». Proceedings of IGARSS'92, Houston, Texas, Mai 92.

[4] CAILLOL H. & HILLION A. & PIECZYNSKI W. (1993) «Fuzzy random fields and unsupervised image segmentation». IEEE Transactions on Geoscience and Remote Sensing, Vol. 31, No 4, pp. 801-810.

[5] CASAS J. R. & HILLION A. & ROUX C. & TORRES L. & GASULL A. «Fuzzy classification of remote sensing images : a pseudocolor representation of fuzzy partitions». Conf. on Neural and Stochastic Methods in Image and Signal Processing, à paraître.

[6] CELEUX G. & DIEBOLD J. (1986) L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités. Revue de Statistique Appliquée, Vol. 34, No 2.

[7] FENG J. & LIN W. & CHEN C. (1991) Epicardial boundary detection using fuzzy reasoning. IEEE Transactions on MI, Vol. 10, No 2, pp. 187-199.

[8] GATH I. & GEVA A. B. (1989) Unsupervised optimal fuzzy clustering. IEEE transactions on PAMI, Vol. 11, No 7, pp. 773-781.

[9] GEMAN S. & GEMAN D. (1984) Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, pp. 721-741.

[10] GU T. & DUBUISSON B. (1990) Similarity of classes and fuzzy clustering. Fuzzy Sets and Systems 34, pp. 213-221.

[11] GUYON X. (1993) Champs aléatoires sur un réseau. Collection Techniques Stochastiques, Masson, Paris.

[12] HILLION A. (1992) Les approches statistiques pour la reconnaissance des images de télédétection. Atti della XXXVI Riunione Scientifica, SIS, Vol. 1, pp. 287-297.

[13] HUNSTBERGER T.L. & JACOBS C.L. & CANNON R.L. (1985) Iterative fuzzy image segmentation. Pattern Recognition Vol. 18, No 2 , pp. 131-138.

[14] KENT J.T. & MARDIA K.V. (1991) Spatial classification using fuzzy membership. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 10, No 5.

[15] MARHIC N. & MASSON P. & PIECZYNSKI W. (1991) Mélange de lois et segmentation non supervisée des données SPOT. Statistique et Analyse des Données, Vol. 16, No 2.

[16] MARROQUIN J. & MITTER S. & POGGIO T. (1987) Probabilistic solution of ill-posed problems in computational vision. Journal of the American Statistical Association - 82 - pp. 76-89.

[17] MASSON P. & PIECZYNSKI W. (1993) SEM algorithm and unsupervised segmentation of satellite images. IEEE Transactions on Geoscience and Remote Sensing, Vol. 31, No 3, pp. 618-633.

[18] PEDRYCZ W. (1990) Fuzzy sets in pattern recognition : methodologie and methods. Pattern Recognition, Vol. 23, No 1/2, pp. 121-146.

[19] PIECZYNSKI W. (1989) Estimation of context in random fields. Journal of Applied Statistics, Vol. 16, No 2, pp. 283-290.

[20] PIECZYNSKI W. (1992) Statistical image segmentation. Machine Graphics and Vision, Vol. 1, No 1/2, pp. 261-268.

[21] QUELLE H. C. & BOUCHER J. M. & PIECZYNSKI W. (1992) Local parameter estimation and unsupervised segmentation of SAR images. Proceedings of IGARSS'92, Houston, Texas, mai 92.

[22] WANG F. (1990) Fuzzy supervised classification of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. Vol. 28, No 2.

[23] YOUNES L. (1988) Problèmes d'estimation paramétrique pour des champs de Gibbs Markoviens. Application au Traitement d'images. Thèse, Université de Paris Sud.