

Champs de Markov flous pour images multispectrales

Fuzzy Markov random fields for multispectral images

F. Salzenstein¹, Ch. Collet², M. Petreman²

¹Laboratoire PHASE, UPR CNRS 292 Université Strasbourg 1 (ULP)
Fabien.Salzenstein@phase.c-strasbourg.fr

²LSIIT UMR CNRS 7005, Université Strasbourg 1 (ULP)
collet@lsiit.u-strasbg.fr

Manuscrit reçu le 15 octobre 2003

Résumé et mots clés

Dans cet article, nous proposons un nouveau mode de segmentation fondé sur une modélisation statistique markovienne floue multispectrale. La segmentation étant non supervisée, nous nous intéressons à l'estimation des hyper-paramètres en combinant un algorithme du gradient stochastique avec les moments empiriques, en supposant que les lois d'attache aux données sont gaussiennes. Sous l'hypothèse de corrélation des champs observés, nous établissons les relations qui relient les matrices de variances-covariances relatives aux sites flous avec celles qui caractérisent les sites durs ou homogènes de l'image. Nous comparons les résultats des segmentations mettant en jeu le critère du MPM (Mode des Marginales *a Posteriori*) et la technique ICM (Mode Conditionnel Itératif) correspondant au critère du MAP (Maximum *a Posteriori*). Dans un premier temps, nous validons la procédure sur des images de synthèse, puis sur des données astronomiques multispectrales.

Segmentation Floue Multispectrale, Champs de Markov Flous, Gradient Stochastique, MPM, ICM, Images Multispectrales Astronomiques.

Abstract and key words

This paper deals with a new statistical segmentation based on fuzzy multispectral markovian random fields model. We propose to solve the problem of parameter estimation, applying a stochastic gradient algorithm and empirical moment method, in order to estimate respectively the *a priori* parameters of the hidden Markovian field and the conditional densities of the observed data. Under correlated spectral band assumption, we introduce a model to express the variance-covariance matrix related to the fuzzy classes, by means of the ones related to the hard classes. We compare the results applying MPM (Mode of Posterior Marginales) and ICM (Iterated Conditional Mode) algorithms. We validate our procedure on synthetic images and test this approach on real multispectral astronomical data.

Fuzzy Multipectral Markovian Segmentation, Markovian Random Field, Stochastic Gradient Algorithm, MPM, ICM, Multispectral Astronomical Images.

Remerciements

Ce travail de recherche a été partiellement financé par le Ministère de la Recherche dans le cadre de l'Action Concertée Incitative GRID IDHA (2001-2004). Les auteurs remercient G. Lagache (Institut d'Astrophysique Spatiale d'Orsay) et A. Lançon (Observatoire de Strasbourg) pour la fourniture d'images astronomiques.

1. Introduction

1.1. Contexte en Imagerie Astronomique

L'astronomie produit de grandes masses de données, réparties et hétérogènes : archives des observatoires sol et spatiaux, catalogues et cartographie numérique du ciel, articles dans les journaux électroniques. Savoir identifier un même objet (étoile, galaxie) détecté dans des catalogues à des longueurs d'ondes différentes, ou à des époques différentes, est d'une importance cruciale en astronomie pour l'interprétation physique des observations. Il convient donc de poursuivre le développement de méthodes et d'outils pour l'identification croisée entre de grands ensembles de données hétérogènes et de développer de nouvelles approches en analyse d'images (segmentation, classification, ...), adaptées aux caractéristiques de l'imagerie astronomique, *i.e.*, grande dynamique des luminances (codées en flottant) et objets astronomiques intrinsèquement multirésolutions et multispectraux. Étant donnée la quantité d'information générée par les multiples campagnes d'observations, il est aujourd'hui indispensable de proposer des traitements automatiques de données astronomiques multibandes et de développer de nouveaux outils permettant l'analyse de cubes de données multispectrales, voire hyperspectrales dans le cadre des projets d'Observatoires Virtuels [1].

1.2. Imprécisions et incertitudes sur les observations multispectrales

Une autre particularité de l'imagerie astronomique, outre la grande dynamique des observations, est l'imprécision liée à la nature même des objets observés, intrinsèquement multi-échelles. Ainsi, lorsque l'on raisonne en terme de classes homogènes et parfaitement différenciées, on peut dans certains cas, éliminer artificiellement les imprécisions liées à la physique de l'observation (nature du terrain en télédétection ou nature des tissus en imagerie cérébrale par exemple). La simple représentation par classes thématiques occulte alors la complexité de l'information réelle et son aspect continu. En effet, le passage d'une classe homogène (ou « dure ») à l'autre ne se fait pas nécessairement de manière abrupte et nécessite parfois la prise en compte explicite d'un continuum de variations. Dans cette perspective, au lieu d'indexer les pixels par des ensembles discrets, une voie peut consister à les indexer par leur degré d'appartenance aux différentes régions homogènes. Pour modéliser ce type d'approche, certains auteurs ont appliqué des méthodes de segmentations ([2], [3], [4]) fondées sur la théorie des ensembles flous dont l'idée initiale a été proposée par Zadeh [5]. Le principe consiste à décrire un objet au moyen d'un intervalle de valeurs. Il est désormais établi que la logique floue et les outils probabilistes peuvent être combinés en un modèle statistique flou, y compris à l'aide de modèles markoviens [6]. Dans

la *segmentation floue* chaque pixel appartient simultanément aux éléments de l'ensemble discret à K éléments, alors que dans un schéma de *segmentation dure* chaque pixel ne peut appartenir qu'à une classe unique.

Soit une image x , réalisation d'un champ aléatoire X composé d'un ensemble de variables aléatoires x_s indexées sur la grille discrète $s \in S$, à valeurs dans $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ où ω_i désigne une étiquette. Soit une observation multispectrale \mathbf{y} , réalisation d'un champ de vecteurs aléatoires \mathbf{Y} composé d'un ensemble de vecteurs \mathbf{y}_s de variables aléatoires $y_s^{(i)}$ indexées sur la même grille discrète $s \in S$ pour chacune des D bandes spectrales $(i)_{i \in \{1, 2, 3, \dots, D\}}$ observées. Une modélisation possible consiste à associer en chacun des sites s de la carte de segmentation, un vecteur $(\varepsilon_1(s), \varepsilon_2(s), \dots, \varepsilon_K(s))$ représentant les degrés d'appartenance aux classes ω_i avec la condition de normalisation : $\varepsilon_1(s) + \varepsilon_2(s) + \dots + \varepsilon_K(s) = 1$. La segmentation dure est un cas particulier pour lequel tous les $\varepsilon_j(s)$ sont nuls, sauf une composante égale à 1. Il est à noter que ce modèle combine à la fois l'imprécision et l'incertitude :

- le concept de « flou » modélise l'*imprécision* portant sur l'appartenance d'un pixel à une région de l'image. Il suppose qu'il peut être impossible de choisir précisément la région d'appartenance ;
- le concept de « probabilité » rend compte de l'*incertitude* sur la classification des pixels ([7], [8]). Il suppose que le pixel est clairement identifié, mais les données sont bruitées.

Il existe également d'autres outils de traitement des données multispectrales, combinant incertitude et imprécision, telle que la théorie de la fusion de Dempster Shafer [9]. Celle-ci élargit le cadre probabiliste classique en introduisant des jeux de masses modélisant l'imprécision sur l'appartenance d'un site à une ou plusieurs classes simultanément. Il est toutefois possible de combiner les modèles markoviens probabilistes et la théorie de l'évidence de Dempster Shafer, ou encore d'utiliser des modèles markoviens dits « évidentiels » où les jeux de masses remplacent les probabilités [10]. Les techniques floues que nous étudions ici sont en fait une extension au cas continu de modèles utilisés dans le cadre markovien discret.

1.3. Modélisation markovienne floue multispectrale

L'introduction des champs de Markov en imagerie [11] marque le point de départ de nombreux traitements fondés sur les modèles markoviens [12]. Par la suite, les modélisations par les chaînes ou les arbres de Markov ont permis le développement d'algorithmes rapides. Ces algorithmes ont été appliqués aux observations mono-bandes, puis étendus aux observations multi-bandes ([13], [14], [15]). Ces approches ont en commun l'hypothèse suivante : en chaque site, la probabilité conditionnellement à tous les autres sites est égale à la probabilité conditionnellement à un voisinage (qui peut être situé à un autre niveau de résolution, comme c'est le cas dans les arbres). Ceci permet une prise de décision locale, à partir de toute l'image. Il

a été proposé dans [6] de combiner ce modèle probabiliste avec l'approche floue en un modèle markovien flou. À notre connaissance, le problème de l'identification de classes en imagerie astronomique multispectrale n'a jamais été abordé sous cet angle. Or nous pensons que l'on gagnerait à introduire cette notion d'imprécision lors de l'observation de zones particulières de l'Univers constituées de nébuleuses, d'amas gazeux ou d'amas de galaxies lointaines à faible rayonnement par exemple, pour lesquels une délimitation stricte entre les régions est délicate (cf. Figure 9). Nous proposons dans cet article un nouveau type d'approche floue fondée sur une statistique markovienne appliquée à des images multispectrales : jusqu'à présent il n'avait été proposé que de traiter des images floues monobandes [16], et les différentes approches multispectrales fondées sur une statistique markovienne n'ont portées que sur des classes discrètes. Nous proposons donc ici une extension des approches multispectrales discrètes fondées sur les champs de Markov [12] à la segmentation floue non supervisée à deux classes dures. En outre, l'introduction d'observations multispectrales (de 3 à une dizaine de bandes spectrales) pose d'emblée deux difficultés : 1) la ressource informatique nécessaire à la prise en compte simultanément de toutes les observations spectrales dans un contexte markovien délivrant au final une unique carte de segmentation ; 2) la nécessaire mise en œuvre d'une chaîne de traitement non supervisée (le nombre de paramètres à estimer dépasse souvent la centaine) qui se heurte rapidement (au delà d'une dizaine de bandes typiquement) au phénomène de Hughes [19].

1.4. Étapes pour une segmentation non supervisée

Plus précisément, le contexte local, spectral et spatial, autour de chaque site est pris en compte globalement en attribuant aux données non bruitées $X = x$ une loi de probabilité markovienne, tandis que le bruit multidimensionnel présent dans l'observation multibande $\mathbf{Y} = \mathbf{y}$ est modélisé au moyen de lois gaussiennes conditionnellement à x . Il nous faut donc estimer préalablement les paramètres liés à la statistique, *i.e.*, les lois d'attache aux données et paramètres du champ de Markov flou caché.

1.4.1. Étape d'estimation des paramètres

L'étape d'estimation des hyper-paramètres met en jeu l'algorithme ICE qui est une méthode générale d'estimation dans le cas des données incomplètes [21]. Nous la combinons avec le gradient stochastique pour obtenir les paramètres *a priori* liés aux données cachées. Nous cherchons à améliorer les performances de cet algorithme en modifiant le pas d'adaptation. Pour calculer les paramètres des lois d'attache aux données, nous estimons les paramètres des distributions statistiques des classes dures, étant donné que ceux des classes floues s'en déduisent linéairement.

1.4.2. Étape de segmentation

Le problème de la segmentation consiste alors à retrouver la réalisation \hat{x} , donnée par un continuum de valeurs à la différence des modèles discrets, qui approxime x au sens d'un critère préalablement choisi (*maximum a posteriori (MAP), mode marginal a posteriori (MPM), champ moyen (MF),...*). Cette étape mettra en jeu la méthode ICM [11] (Iterative Conditional Mode) ou la solution MPM adaptée au cas flou [23].

1.4.3. Organisation de l'article

L'article est organisé en 6 parties. Dans le paragraphe suivant, nous présentons le modèle de champ de Markov à deux classes dures (modèle *a priori*), puis en section 3 nous explicitons les lois liées aux observations et le modèle de champ de Markov caché, dans le cas multispectral. En 4^{ème} partie, nous décrivons la méthode d'estimation des paramètres dans le cadre multispectral. La section 5 traite du problème d'inférence bayésienne (critères MAP et MPM) pour la segmentation. Enfin, nous validerons ces approches (section 6) en présentant les résultats obtenus sur images de synthèse et images réelles multispectrales, pour lesquelles une modélisation du flou traduisant l'imprécision des observations se révèle pertinente.

2. Champs de Markov flous

2.1. Expression de la loi de X

Le champ des étiquettes $X = x$ est la réalisation d'un champ de variables aléatoires. Chaque pixel est susceptible d'appartenir à l'une des deux classes dures «0» ou «1», ou à une classe floue intermédiaire entre ces classes :

$$\forall s \in S, x_s = (\varepsilon_1(s), \varepsilon_2(s))$$

avec $\varepsilon_1(s) + \varepsilon_2(s) = 1$ (1)

$$\text{et } (\varepsilon_1(s), \varepsilon_2(s)) \in [0, 1]^2$$

Ainsi $X \in \Omega_f^{\#S} = [0, 1]^{\#S}$ où $\#S$ désigne le cardinal de S . La loi de chaque variable X_s est donnée par une densité h relativement à la mesure $\nu = \delta_0 + \delta_1 + \mu$, incluant à la fois des composantes dures et une composante continue (mesure de Lebesgue sur $]0, 1[$) [20]. L'information dure se rapporte aux composantes discrètes tandis que l'information floue est introduite par les composantes continues. En supposant le champ stationnaire, h vérifie en chaque site la condition de normalisation :

$$h(0) + h(1) + \int_0^1 h(t) dt = 1$$
 (2)

Dans le cadre de notre étude X est un champ markovien, dont la densité h_f par rapport à la mesure $v^{\#S}$ est donnée par :

$$h_f(x) = C \exp^{-U_f(x)} \quad (3)$$

U_f est l'énergie exprimée par une somme de fonctions définies sur les cliques de cardinal i sur $[0, 1]^i$:

$$U_f(x) = \sum_i \sum_{c \in C_i} \Phi_i(x_c) \text{ avec } x_c \in [0, 1]^i \quad (4)$$

Les fonctions Φ_i sont donc des applications de $[0, 1]^i$ sur l'ensemble \mathbb{R} . À l'instar du modèle en dur, une réalisation du champ $X = x$ est obtenue par tirage aléatoire au moyen de l'échantillonneur de Gibbs.

2.2. Modèle de fonction d'énergie

Nous nous limiterons ici aux cliques de cardinal $i = 2$, selon les différentes directions horizontales, verticales et diagonales. Pour une clique binaire (x_s, x_t) , la valeur de la fonction potentiel $\Phi_{i=2}(x_s, x_t)$ que l'on notera dorénavant $\Phi(\cdot)$ dépend de la nature des pixels (deux pixels durs ou au moins un pixel flou) et de leur orientation. Le modèle énergétique retenu est le suivant [22] :

- si $(x_s, x_t) \in \{0, 1\}^2$, *i.e.*, les deux pixels sont durs, alors

$$\Phi(x_s, x_t) = \begin{cases} -\alpha_d \text{ si } x_s = x_t \\ +\alpha_d \text{ si } x_s \neq x_t \end{cases} \quad (5)$$

- si $(x_s, x_t) \in [0, 1]^2$, *et* au moins un pixel est flou, alors

$$\Phi(x_s, x_t) = -\alpha_f \varphi(x_s, x_t) \text{ avec } 0 \leq |\varphi(x_s, x_t)| \leq 1 \quad (6)$$

On suppose ici implicitement que $\{\alpha_d, \alpha_f\} \in (\mathbb{R}^+)^2$. La fonction φ doit être compatible avec les conditions d'homogénéité des régions floues et les propriétés de continuité d'une classe dure à l'autre. Il est peu probable, par exemple, que deux sites voisins appartiennent à deux régions « dures » différentes. Ceci implique les propositions suivantes [23] :

$$\begin{cases} x_s = x_t \Rightarrow \varphi(x_s, x_t) = 1 \\ x_s = 1 \text{ et } x_t \rightarrow 0 \Rightarrow \varphi(x_s, x_t) \rightarrow -1 \\ x_s = 0 \text{ et } x_t \rightarrow 1 \Rightarrow \varphi(x_s, x_t) \rightarrow -1 \end{cases} \quad (7)$$

Nous optons pour les fonctions continues de la forme :

$$\varphi(x_s, x_t) = 1 - 2|x_s - x_t|^r \text{ avec } r \in \mathbb{R}^+ \quad (8)$$

Des réalisations d'un champ de Markov associé à différentes fonctions sont proposées en section 6.1. Finalement, les simulations ont été réalisées uniquement à partir d'une seule fonction

φ avec $r = 1$. Nous prenons en compte les voisinages verticaux, horizontaux et diagonaux (selon les deux diagonales). Au total, ceci définit un ensemble de 8 paramètres *a priori* (4 pour les sites durs, et 4 pour les sites flous) :

$$\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_1^f, \alpha_2^f, \alpha_3^f, \alpha_4^f) \quad (9)$$

3. Distribution du couple (X, Y)

Les lois d'attache aux données sont modélisées par des gaussiennes. Cette hypothèse, liée au théorème de la limite centrale, permet de supposer que les observations sont issues de capteurs indépendants, lorsqu'elles ont été préalablement décorrélées, par exemple par une méthode basée sur l'analyse en composantes principales [24]. Dans les zones floues de l'image, les paramètres de ces lois associées aux données observées sont déduites par combinaison linéaire des paramètres associés aux zones homogènes.

3.1. Capteurs non corrélés

Supposons que l'on dispose d'observations issues de D capteurs, réalisations $(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(D)})$ de la variable aléatoire $\mathbf{Y} = (\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(D)})$. Chaque réalisation $\mathbf{y}^{(i)} = \{y_s^{(i)}\}_{s \in S}$ est à valeur dans $\mathbb{R}^{\#S}$. Ainsi, une observation au site s de l'image est défini par un vecteur $\mathbf{y}_s = [y_s^{(1)}, y_s^{(2)}, \dots, y_s^{(D)}]^t$. De la même manière que dans les problèmes classiques de segmentation multispectrale, l'objectif est de retrouver une réalisation du champ caché X à partir de l'ensemble des observations bruitées $(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(D)})$. Nous ferons dès lors les hypothèses suivantes :

- indépendance spectrale : les observations conditionnellement à une classe donnée sont spectralement indépendantes, autrement dit les variables aléatoires $(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(D)})$ sont indépendantes conditionnellement à X ;

- indépendance spatiale : pour un capteur i donné, les variables aléatoires $Y_s^{(i)} \Big|_{s \in S}$ sont indépendantes conditionnellement à X ;

- la loi de $Y_s^{(i)}$ conditionnellement à X est égale à la loi de $Y_s^{(i)}$ conditionnellement à X_s .

En supposant que les distributions de $(Y_s^{(i)}, X_s)$ soient indépendantes de s (hypothèse de stationnarité du champ aléatoire) et en notant $\mathcal{N}(\mu, \sigma^2)$ la loi normale de moyenne μ et de variance σ^2 , nous supposons que la loi de distribution de la densité de $Y_s^{(i)}$ conditionnellement à X_s s'écrit $f_{X_s}^{(i)}$ où :

$$\forall x_s \in [0, 1], f_{X_s}^{(i)}(y_s^{(i)}) = \mathcal{N}((1 - x_s)\mu_0^{(i)} + x_s\mu_1^{(i)}, (1 - x_s)^2(\sigma_0^{(i)})^2 + x_s^2(\sigma_1^{(i)})^2) \quad (10)$$

où $(\mu_0^{(i)}, (\sigma_0^{(i)})^2)$ et $(\mu_1^{(i)}, (\sigma_1^{(i)})^2)$ représentent les paramètres associés aux densités conditionnelles lorsque X_s appartient à l'une des deux classes dures «0» ou «1». L'hypothèse (10) est compatible avec la définition des degrés flous comme combinaison linéaire des deux classes dures ([16], [25]). Par conséquent l'ensemble des valeurs $(\mu_0^{(i)}, \mu_1^{(i)}, \sigma_0^{(i)}, \sigma_1^{(i)})_{i \in \{1, D\}}$ définit les lois d'attache aux données. Compte tenu des hypothèses précédentes, exprimons la densité f des variables aléatoires (X, \mathbf{Y}) par rapport à la mesure $\nu^{\#S} \otimes (\mu^{\#S})^{(1)} \otimes (\mu^{\#S})^{(2)} \otimes \dots \otimes (\mu^{\#S})^{(D)}$ où $(\mu^{\#S})^{(i)}$ est une mesure de Lebesgue sur $\mathbb{R}^{\#S}$:

$$f(x, \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(D)}) = h_f(x) f(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(D)} | x) \\ = C \exp^{-U_f(x)} \prod_{i=1}^D \prod_{s=1}^{\#S} f_{x_s}^{(i)}(y_s^{(i)}) \quad (11)$$

puisque nous faisons l'hypothèse de bruits gaussiens, il revient au même de parler d'indépendance ou de non corrélation.

3.2. Capteurs corrélés

Généralement, l'hypothèse de non corrélation entre capteurs est trop restrictive et on est amené à supposer que la loi de la variable $\mathbf{y}_s = [y_s^{(1)}, y_s^{(2)}, \dots, y_s^{(D)}]^t$ sachant $X_s = \varepsilon_s \in [0, 1]$ est une loi gaussienne multidimensionnelle de vecteur moyenne $\mu_{\varepsilon_s} = [\mu_{\varepsilon_s}^{(0)}, \mu_{\varepsilon_s}^{(1)}, \dots, \mu_{\varepsilon_s}^{(D)}]^t$ et de matrice de covariance Γ_{ε_s} de taille $\mathbb{R}^D \times \mathbb{R}^D$:

$$f_{x_s}(y_s) = \frac{1}{2\pi^{D/2} (\det \Gamma_{\varepsilon_s})^{1/2}} \exp\left(-\frac{1}{2} (y_s - \mu_{\varepsilon_s})^t \Gamma_{\varepsilon_s}^{-1} (y_s - \mu_{\varepsilon_s})\right) \quad (12)$$

le terme général de la matrice de covariance est donné par $\Gamma_{\varepsilon_s}(i, j) = E\left[\left(y_s^{(i)} - \mu_{\varepsilon_s}^{(i)}\right)\left(y_s^{(j)} - \mu_{\varepsilon_s}^{(j)}\right)\right]$. Notre hypothèse consiste à généraliser le cas des capteurs non corrélés en exprimant les covariances et moyennes des sites flous en fonction des covariances et moyennes correspondant aux classes «dures». Soient respectivement (μ_0, μ_1) et (Γ_0, Γ_1) les couples de vecteurs moyennes et de matrices de covariances associés aux deux classes dures. En un site flou de valeur $X_s = \varepsilon_s$, les paramètres μ_{ε_s} et Γ_{ε_s} se déduisent de (μ_0, μ_1) et (Γ_0, Γ_1) par les relations :

$$\mu_{\varepsilon_s} = (1 - \varepsilon_s) \mu_0 + \varepsilon_s \mu_1 \\ \text{et } \Gamma_{\varepsilon_s}(i, j) = (1 - \varepsilon_s)^2 \Gamma_0(i, j) + \varepsilon_s^2 \Gamma_1(i, j) \quad (13)$$

où $(i, j) \in \{1, 2, \dots, D\}$. La modélisation proposée est une extension de (10).

3.3. Distribution de X a posteriori

Dans le cas de capteurs indépendants, on écrit

$$\prod_{i=1}^D \prod_{s=1}^{\#S} f_{x_s}^{(i)}(y_s^{(i)}) = \exp\left(\sum_{i=1}^D \sum_{s=1}^{\#S} \log f_{x_s}^{(i)}(y_s^{(i)})\right) = \exp^{V_x(\mathbf{y})} \quad (14)$$

et il vient en posant $W_{\mathbf{y}}(x) = U(x) - V_x(\mathbf{y})$,

$$f(x, \mathbf{y}) = C' \exp^{-W_{\mathbf{y}}(x)} \quad (15)$$

où C' désigne le facteur de normalisation de la densité de probabilité de la loi jointe.

Lorsque les capteurs sont corrélés, $V_x(\mathbf{y})$ correspond à $V_x(\mathbf{y}) = \sum_{i=1}^D \sum_{s=1}^{\#S} \log f_{x_s}^{(i)}(y_s^{(i)})$, où

$$f_{x_s}(\mathbf{y}_s) = \left[f_{x_s}^{(1)}(y_s^{(1)}), f_{x_s}^{(2)}(y_s^{(2)}), \dots, f_{x_s}^{(D)}(y_s^{(D)}) \right]^t$$

est donnée par (12). Comme dans le cas de l'approche markovienne monobande, nous préservons la nature markovienne du champ X a posteriori et il est donc possible d'en simuler des réalisations au moyen de l'échantillonneur de Gibbs.

4. Estimation des hyperparamètres par l'algorithme ICE

4.1. L'algorithme ICE

La procédure de classification des pixels nécessite la connaissance des hyper-paramètres qui sont de deux natures différentes : les paramètres définissant l'énergie du champ markovien *a priori*, et ceux qui définissent les lois gaussiennes d'attache aux données. Notons α le paramètre qui correspond aux fonctions de cliques et β le paramètre correspondant aux termes d'attache aux données. Le paramètre $\theta = (\alpha, \beta)$ définit donc la distribution de la variable (X, Y) . Pour estimer θ à partir de la réalisation $Y = y$, nous utilisons la procédure ICE qui est un algorithme général d'estimation dans le cas de données incomplètes ([21]). En supposant que l'on dispose de la réalisation du champ caché X , un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ consistera à calculer séparément $\alpha = \hat{\alpha}(x)$ et $\beta = \hat{\beta}(x, y)$. La procédure ICE de calcul du paramètre θ à partir d'un estimateur $\hat{\theta}$ se déroule de manière itérative :

- Initialisation : $\theta^{[0]} = \theta_{\text{initialisation}}$;
- à l'itération courante $\theta^{[p+1]} = E_{\theta^{[p]}}[\hat{\theta}(X, Y) | Y = y]$.

À l'étape courante de la procédure, le calcul de l'espérance conditionnelle $E_{\theta^{[p]}}[\hat{\theta}(X, Y) | Y = y]$ est contourné par approximation de type Monte-Carlo en simulant N réalisations cachées du champ selon la loi *a posteriori*. Cela donne :

- Initialisation : $\theta^{[0]} = \theta_{\text{initialisation}}$
- calcul de $\theta^{[p+1]} = (\alpha^{[p+1]}, \beta^{[p+1]})$:

$$\alpha^{[p+1]} = \frac{1}{N} \left(\widehat{\alpha} \left(x_1^{[p]} \right), \widehat{\alpha} \left(x_2^{[p]} \right), \dots, \widehat{\alpha} \left(x_N^{[p]} \right) \right)$$

$$\beta^{[p+1]} = \frac{1}{N} \left(\widehat{\beta} \left(x_1^{[p]}, \mathbf{y} \right), \widehat{\beta} \left(x_2^{[p]}, \mathbf{y} \right), \dots, \widehat{\beta} \left(x_N^{[p]}, \mathbf{y} \right) \right)$$

où les $x_i^{[p]}$ sont des réalisations du champ caché à l'étape p obtenues au moyen du paramètre courant $\theta^{[p]} = (\alpha^{[p]}, \beta^{[p]})$.

4.2. Estimation du paramètre α à partir d'une réalisation observable de X

Nous explicitons dans cette section l'estimateur du paramètre α à partir d'une réalisation x du champ caché X . L'estimateur $\widehat{\alpha}(x)$ correspond à l'algorithme du gradient stochastique de Younes [26]. Cette procédure itérative se déroule de la manière suivante :

- Initialisation : $\alpha^{[0]}$;
- calcul du paramètre à l'itération n :

$$\alpha^{[n+1]} = \alpha^{[n]} + \frac{c}{n+1} [U'(x^{[n+1]}) - U'(x)] \quad (16)$$

où $U'(x)$ est le gradient de $U(x)$ selon le paramètre α . $x^{[n+1]}$ est une réalisation de X simulée par l'échantillonneur de Gibbs flou selon la distribution *a priori*, au moyen des paramètres $\alpha^{[n]}$ ([16], [25]). Comme dans le cas dur, un choix approprié du facteur c permet d'assurer la convergence de la méthode. Une valeur inversement proportionnelle au nombre de pixels de l'image garantit une bonne convergence. Remarquons qu'il s'agit ici de la méthode du gradient stochastique à pas décroissant. L'un de ses inconvénients majeurs est la perte rapide de la capacité d'adaptation. Un algorithme à pas fixe (avec $c = 1/N$ où $N = \#S$) permettrait une plus grande souplesse, mais sans garantir une convergence stricte. De récentes contributions [27] ont proposé un algorithme à décroissance plus lente, par moyennage à chaque itération des valeurs estimées lors des pas précédents. Nos tests ont montré que l'algorithme le plus efficace reste l'algorithme à pas décroissant avec $c = 1/N$: nous retenons donc cette méthode proposée.

4.3. Estimation du paramètre β à partir d'une réalisation observable de (X, Y)

4.3.1. Cas mono-spectral: estimation pour une observation $Y = y^{(1)}$

En classification dure, et pour des lois gaussiennes, l'estimation des paramètres associés (moyennes et variances) par le critère du maximum de vraisemblance équivaut à la méthode des moments empiriques. Dans le cadre flou, l'estimation des paramètres β se ramène à l'estimation des moyennes et variances des

classes dures associées, étant donné que pour les pixels flous les lois d'attache aux données s'en déduisent. Certaines techniques d'estimation de ces paramètres en segmentation markovienne floue non supervisée ([16], [25]) utilisent la méthode des moments empiriques avec les mêmes formules de ré-estimation qu'en classification dure. En fait, l'application du maximum de vraisemblance au cas flou conduit à un algorithme plus complexe. Considérons une seule observation $Y = y$ et supposons que l'on dispose de la réalisation cachée $X = x$ (simulée selon la loi *a posteriori*). Notons $(\varepsilon_0 = 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_Q, \varepsilon_{Q+1} = 1) \subset [0, 1]$, l'ensemble correspondant aux Q niveaux de flou discrétisés présents dans l'image $X = x$. Soit N_q le nombre de sites associés au niveau q . On a donc nécessairement: $\#S = \sum_{q=0}^{Q+1} N_q$. La procédure d'estimation de $\beta = (\mu_0, \sigma_0; \mu_1, \sigma_1)$ est la suivante :

- Initialisation par une méthode de coalescence de type K-moyennes : $\beta^{[0]} = (\mu_0^{[0]}, \sigma_0^{[0]}; \mu_1^{[0]}, \sigma_1^{[0]})$
- ré-estimation des moyennes à l'étape courante p :

$$\mu_0^{[p]} = \frac{\sum_{q=0}^{Q+1} \frac{1-\varepsilon_q}{(1-\varepsilon_q)^2 (\sigma_0^{[p-1]})^2 + \varepsilon_q^2 (\sigma_1^{[p-1]})^2} \sum_{i=1}^{N_q} (y_{i|x_i=\varepsilon_q} - \varepsilon_q \mu_1^{[p-1]})}{\sum_{q=0}^{Q+1} \frac{(1-\varepsilon_q)^2 N_q}{(1-\varepsilon_q)^2 (\sigma_0^{[p-1]})^2 + \varepsilon_q^2 (\sigma_1^{[p-1]})^2}} \quad (17)$$

$$\mu_1^{[p]} = \frac{\sum_{q=0}^{Q+1} \frac{\varepsilon_q}{(1-\varepsilon_q)^2 (\sigma_0^{[p-1]})^2 + \varepsilon_q^2 (\sigma_1^{[p-1]})^2} \sum_{i=1}^{N_q} (y_{i|x_i=\varepsilon_q} - (1-\varepsilon_q) \mu_0^{[p-1]})}{\sum_{q=0}^{Q+1} \frac{\varepsilon_q^2 N_q}{(1-\varepsilon_q)^2 (\sigma_0^{[p-1]})^2 + \varepsilon_q^2 (\sigma_1^{[p-1]})^2}} \quad (18)$$

- ré-estimation des variances à l'étape courante p , $\forall q \in \{0, 1, \dots, Q+1\}$

$$(1-\varepsilon_q)^2 (\sigma_0^{[p]})^2 + \varepsilon_q^2 (\sigma_1^{[p]})^2 = \frac{1}{N_q} \sum_{i=1}^{N_q} (y_{i|x_i=\varepsilon_q} - \varepsilon_q \mu_1^{[p-1]} - (1-\varepsilon_q) \mu_0^{[p-1]})^2 \quad (19)$$

On estime ainsi le paramètre β en maximisant la vraisemblance qui est la densité conditionnelle de $Y = y$ par rapport à la variable $X = x$.

$$\widehat{\beta} = \arg \max_{\beta} [f_{X=x}(Y = y | x, \beta)] \quad (20)$$

L'annulation de la dérivée de cette expression en utilisant les hypothèses d'indépendance faites au paragraphe indépendance conduit immédiatement aux expressions (17) et (18). Remarquons que μ_0 dépend des paramètres $\sigma_0, \mu_1, \sigma_1$, d'où la nécessité d'un algorithme récursif qui utilise à l'étape p les paramètres calculés à l'étape $p-1$. Remarquons d'autre part,

que la dérivation de (20) par rapport à σ_0 ou σ_1 conduit à des expressions plus complexes qui ne permettent pas d'estimer simplement ces deux paramètres. Le système d'équations (19) à deux inconnues $(\sigma_0^{[p]}, \sigma_1^{[p]})$, résulte de l'application de la formule des moments empiriques pour les variances appliquées à chaque niveau de flou. Il peut être résolu par une technique des moindres carrés. On montre aisément, que si l'on considère uniquement les niveaux ε_0 et ε_1 , les équations (17), (18) et (19) redonnent les formules de ré-estimation classiques de la classification en « dur ».

4.3.2. Estimation pour D observations non corrélées $\mathbf{Y} = \mathbf{y}$

Compte tenu des hypothèses d'indépendance énoncées au paragraphe 3.1, la généralisation des formules d'estimation des paramètres de lois d'attache aux données est aisée. On maximise le \log de l'expression (11) p. 7 par rapport à β ce qui permet d'appliquer (17), (18) et (19) pour chacune des réalisations.

4.3.3. Estimation pour D observations corrélées $\mathbf{Y} = \mathbf{y}$

L'objectif est alors d'estimer l'ensemble les paramètres $\beta = \{(\mu_0, \Gamma_0; \mu_1, \Gamma_1)\} = \left\{ \left(\mu_0^{(i)}, \Gamma_0^{(i,j)}; \mu_1^{(i)}, \Gamma_1^{(i,j)} \right) \right\}$ où i et j désignent deux bandes spectrales avec : $1 < i, j < D$. Deux approches sont proposées.

La première consiste à estimer les matrices des moyennes et covariances relatives aux pixels des classes « durs » en utilisant les formules des moments empiriques, basées sur les sites durs uniquement. En reprenant les notations précédentes :

$$\mu_0^{(i)[p]} = \frac{1}{N_0} \sum_{s=1}^{N_0} \left(y_{s|_{x_s=0}}^{(i)} \right) \text{ et } \mu_1^{(i)[p]} = \frac{1}{N_1} \sum_{s=1}^{N_1} \left(y_{s|_{x_s=1}}^{(i)} \right) \quad (21)$$

La covariance interbandes spectrales (i, j) s'écrit pour chaque classe dure

$$\Gamma_0^{(i,j)[p]} = \frac{1}{N_0} \sum_{s=1}^{N_0} \left(y_{s|_{x_s=0}}^{(i)} - \mu_0^{(i)[p]} \right) \left(y_{s|_{x_s=0}}^{(j)} - \mu_0^{(j)[p]} \right) \quad (22)$$

et

$$\Gamma_1^{(i,j)[p]} = \frac{1}{N_1} \sum_{s=1}^{N_1} \left(y_{s|_{x_s=1}}^{(i)} - \mu_1^{(i)[p]} \right) \left(y_{s|_{x_s=1}}^{(j)} - \mu_1^{(j)[p]} \right) \quad (23)$$

Les paramètres des lois d'attache aux données relatifs aux sites flous se calculent par l'expression (11).

D'autre part, de manière à tenir compte des sites flous, nous proposons une deuxième approche, basée sur la méthode des moindres carrés, qui consiste à calculer, pour les sites flous, les moyennes et covariances par les moments empiriques. On aboutit alors à deux systèmes de D équations

$$\mu_{\varepsilon_q}^{(i)[p]} = (1 - \varepsilon_q) \mu_0^{(i)[p]} + \varepsilon_q \mu_1^{(i)[p]} = \frac{1}{N_q} \sum_{s=1}^{N_q} \left(y_{s|_{x_s=\varepsilon_q}}^{(i)} \right) \quad (24)$$

et

$$\begin{aligned} & (1 - \varepsilon_q)^2 \Gamma_0^{(i,j)[p]} + \varepsilon_q^2 \Gamma_1^{(i,j)[p]} \\ &= \frac{1}{N_q} \sum_{s=1}^{N_q} \left(y_{s|_{x_s=\varepsilon_q}}^{(i)} - \mu_{\varepsilon_q}^{(i)[p]} \right) \left(y_{s|_{x_s=\varepsilon_q}}^{(j)} - \mu_{\varepsilon_q}^{(j)[p]} \right) \end{aligned} \quad (25)$$

avec $i \in \{1, 2, \dots, D\}$. On résout les deux systèmes par la méthode des moindres carrés, en affectant éventuellement des poids plus importants aux pixels des classes durs. Le nombre de sites durs étant significatifs dans les images traitées, nous avons appliqué la première approche.

4.4. Synoptique général de l'algorithme d'estimation des hyperparamètres

- **initialisation** : $\theta^{[0]} = (\alpha^{[0]}, \beta^{[0]})$.

Le paramètre $\beta^{[0]}$ est estimé en supposant que nous avons au départ une image avec des classes dures. L'estimation se fait par la méthode de l'histogramme ou des moyennes et variances globales calculées pour chacune des observations. Le paramètre de la loi *a priori* est mis à zéro.

- **itération courante** : $\theta^{[p+1]} = (\alpha^{[p+1]}, \beta^{[p+1]})$ est déduit de $\theta^{[p]}$ et des observations $\mathbf{Y} = \mathbf{y}$ par la procédure suivante :

a) Simuler par l'échantillonneur de Gibbs N_{ICE} réalisations $\left\{ x_i^{[p]} \right\}_{1 < i < N_{ICE}}$ du champ de Markov caché $X = x$ selon la distribution *a posteriori* sachant $\mathbf{Y} = \mathbf{y}$ et correspondant au paramètre $\theta^{[p]}$. On choisit généralement $N_{ICE} = 1$ en pratique [17]. Au lieu de N_{ICE} réalisations *a posteriori* simulées par l'échantillonneur de Gibbs, on pré-segmente ici l'image par l'algorithme ICM [11], plus rapide (voir en section 5.3).

b) Pour chaque champ *a posteriori* $x_i^{[p]}$, on procède à l'estimation du paramètre *a priori* associé $\hat{\alpha}^{[p]}$ au moyen de l'algorithme du gradient stochastique flou (cf. Eq. 16 p. 9).

c) Pour les N_{ICE} couples et $(x_i^{[p]}, \mathbf{Y} = \mathbf{y})$, on estime les $\hat{\beta}^{[p]}$ par l'expression (20) p. 11, sous les deux hypothèses (dépendance ou indépendance).

d) $\theta^{[p+1]} = (\alpha^{[p+1]}, \beta^{[p+1]})$ est donné par $\alpha^{[p+1]} = \frac{1}{N_{ICE}} \sum_{k=1}^{N_{ICE}} \hat{\alpha}^{[p]}$ et $\beta^{[p+1]} = \frac{1}{N_{ICE}} \sum_{k=1}^{N_{ICE}} \hat{\beta}^{[p]}$.

e) Lorsque la séquence $\theta^{[p]}$ devient stable, nous stoppons l'algorithme.

L'utilisation de la segmentation ICM, lors de l'étape intermédiaire d'estimation, ne dégrade pas les performances de l'estimateur, tout en améliorant les temps de calcul. Nous avons donc opté pour cette approximation rapide de l'ICE.

5. Segmentation floue

À l'étape d'estimation des paramètres succède une prise de décision pour obtenir l'image finale «vérité-terrain». Deux algorithmes de segmentation ont été retenus pour tester le modèle markovien flou sur des données multispectrales. La démarche est fondée sur une stratégie bayésienne. Rappelons en brièvement les principes : considérons une fonction de coût définie par $L^* : [0,1]^S \times [0,1]^S \rightarrow \mathbb{R}^+$. Cette fonction représente la gravité de l'erreur commise, en attribuant à la réalisation la valeur \hat{x} au lieu de x . Étant donné un espace d'observations $Y = y$, la stratégie bayésienne associée \hat{s}_B est celle qui minimise le risque moyen $E[L^*(X, \hat{s}(Y))]$ associé à la décision $\hat{s}(Y)$. La stratégie du MAP consiste, dans le cas dur, à pénaliser de la même manière toutes configurations différentes de la réalisation réelle «vérité-terrain» [28] (fonction de perte «0-1»). Une solution du MAP est l'algorithme du recuit simulé, construit sur un schéma de descente en température combiné à l'échantillonneur de Gibbs. Nous optons pour la méthode ICM qui constitue une approximation du recuit simulé, (Iterative Conditional Mode [11]). Cet algorithme largement décrit dans la littérature présente l'avantage d'être rapide, même s'il reste sensible en terme d'initialisation. Son adaptation au cas continu, nécessite donc l'emploi de l'échantillonneur de Gibbs flou, la procédure générale restant la même. Toutefois, cet algorithme risque fortement d'être piégé dans à un minimum local de la fonction d'énergie à minimiser. Nous comparerons la solution ICM à la stratégie désormais classique en segmentation, basée sur le critère MPM [29]. On associe à chaque site la fonction de perte $L(x_s, \hat{x}_s)$. La stratégie bayésienne consiste à minimiser le risque $E[L(X_s, \hat{s}_s(Y))]$. D'où le caractère «global» du MPM : on prend une décision en chaque site, en tenant compte de toute l'information. Pratiquement, on montre que MPM est la solution qui minimise l'espérance conditionnelle (26) pour chaque réalisation $Y = y$:

$$E[L(X_s, \hat{s}_s(Y)) | Y = y] \quad (26)$$

Le calcul de (26) requiert la connaissance de la distribution *a posteriori* en chaque site s . Notons h_s^y les densités des distributions *a posteriori* par rapport à la mesure ν . Si on suppose ces densités connues, (26) s'écrit :

$$E[L(X_s, \hat{s}_s(Y)) | Y = y] = h_s^y(0) \cdot L(0, \hat{s}_s(y)) + h_s^y(1) \cdot L(1, \hat{s}_s(y)) + \int_{t=0}^{t=1} h_s^y(t) \cdot L(t, \hat{s}_s(y)) dt \quad (27)$$

Nous attribuons donc en chaque pixel la valeur $\hat{s}_s(y) \in [0,1]$ qui minimise (27). La recherche de la classe optimale s'effectuera numériquement en approchant l'intégrale de (27) par une somme de Riemann. Les densités des lois conditionnelles marginales sont estimées par les fréquences empiriques, en simulant plusieurs réalisations du champ caché selon sa loi *a posteriori*. Nous étudierons la stratégie fondée sur la fonction de perte «écart en valeur absolue» : $L_s(x_s, \hat{x}_s) = |x_s - \hat{x}_s|$.

Enfin, le fait de raisonner à partir d'ensembles continus permet une grande flexibilité dans le choix des fonctions de coût [23]. Et si l'on suppose le champ X stationnaire, l'erreur moyenne correspondant à la prise de décision est donnée par :

$$E[L(X_s, \hat{X}_s)] = \frac{1}{S} \sum_{s=1}^{s=S} L(x_s, \hat{x}_s) \quad (28)$$

Pour établir un critère de comparaison commun aux deux méthodes de segmentation étudiées, nous calculons pour chacune d'entre elles, le taux d'erreur (28) à partir du coût «écart en valeur absolue» : $L_s(x_s, \hat{x}_s) = |x_s - \hat{x}_s|$.

6. Résultats sur images multispectrales

Nous présentons ci-après une série de résultats portant dans un premier temps sur des images de synthèse en multibande, puis sur une série d'images multispectrales de la base de données de l'Observatoire de Strasbourg¹. Nous procédons dans un premier temps à l'évaluation de la méthode du gradient stochastique.

6.1. Images de synthèse floues multispectrales à 2 classes dures

6.1.1. Champs de Markov flous simulés à deux classes dures

Nous présentons trois champs de Markov flous à deux classes dures, dont les paramètres de la loi *a priori* figurent dans le tableau 1. Les paramètres sont définis par (9) au paragraphe 2.2. Nous calculons pour chacune des réalisations, le taux de flou (tableau 1) correspondant au ratio du nombre de pixels flous sur le nombre total de pixels. Les images respectivement obtenues étiquetées 1(a), 1(b), 1(c) diffèrent par la quantité de flou (1(a) contient plus de pixels flous), par le nombre de niveaux discrets entre les niveaux «0» (classe «0») et «1» (classe «1») ou privilégient une direction donnée (les paramètres de cliques horizontaux d'ordre deux pour l'image 1(c)). Nous choisirons de discrétiser le flou en 10 niveaux de manière à trouver un bon compromis entre la préservation de la continuité et la rapidité des calculs. Nous remarquons que les réalisations sont sensibles aux paramètres des cliques, en terme de *valeurs relatives* des cliques «flous» par rapport aux cliques «durs» (un écart de 0.5 suffit pour augmenter considérablement le niveau de flou).

1. <http://astro.u-strasbg.fr/Obs.html>

Tableau 1. Paramètres de la loi a priori

Images	α_1	α_2	α_3	α_4	α_1^f	α_2^f	α_3^f	α_4^f	niveaux discrets	taux de flou
1(a)	4	4	4	4	4.5	4.5	4.5	4.5	10	44.16 %
1(b)	3	3	3	3	3.5	3.5	3.5	3.5	16	37.79 %
1(c)	3	3	22	3.5	3.5	22.5	3.5	3.5	10	34.13 %

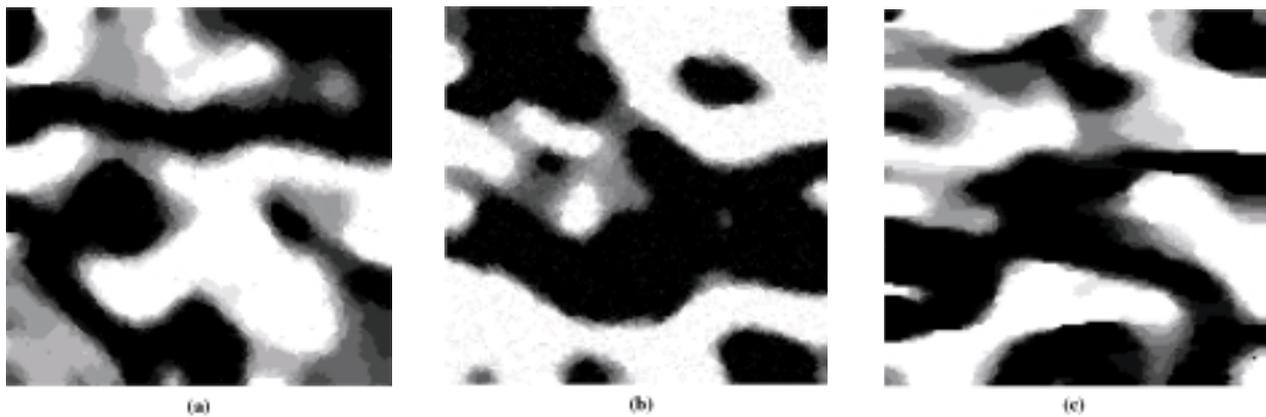


Figure 1. Champs markoviens simulés par les paramètres du tableau 1

Tableau 2. Paramètres de la loi a priori

Images	α_1	α_2	α_3	α_4	α_1^f	α_2^f	α_3^f	α_4^f	niveaux discrets	% flou
2(a)	4	4	4	4	4.5	4.5	4.5	4.5	10	44.16 %
2(b)	5	5	5	5	5	5	5	5	10	0.79 %

6.1.2. Algorithme du gradient stochastique appliqué aux champs markoviens flous

Nous avons testé l'algorithme du gradient dans deux situations (un champ «flou» et un champ «dur» représentés Figure 2 a-b). Ces champs de Markov ont été simulés par les paramètres du tableau 2 au moyen de l'échantillonneur de Gibbs flou: le champ «dur» contient donc également une faible proportion de sites flous. Le tableau 3 récapitule les résultats des paramètres estimés et les champs correspondant (colonne de gauche). Nous avons également indiqué la quantité de flou présente dans les réalisations des champs initiaux (colonne de droite du tableau 2) et celle des champs simulés avec les paramètres estimés (colonne de droite du tableau 3). La Figure 3 représente des réalisations aléatoires obtenues au moyen des paramètres estimés. Pour estimer les paramètres du champ 2(a), nous avons successivement choisi les valeurs initiales $\alpha = (4, 4, 4, 4, 4, 4, 4, 4)$,

puis $\alpha = (2, 2, 2, 2, 2, 2, 2, 2)$. En conséquence, l'image 3(a) est plus homogène que l'image 3(b), donc plus proche de la réalisation correspondante 2(a). Cette propriété résulte de l'application d'un algorithme à pas non constant décroissant (cf. section 4.2) qui assure une convergence plus rapide, au détriment de l'aspect stochastique. Des valeurs d'initialisation correspondant à un champ Markovien «dur» homogène, donnent en général de bons résultats. Remarquons que *ce qui importe est non pas tant la valeur des paramètres*, que le rapport entre les paramètres. En effet ce rapport détermine tout autant l'homogénéité du champ, que les proportions respectives de la classe floue et de la classe dure dans l'image. Ainsi, pour le champ initial 2(a), l'écart entre les coefficients estimés flous et durs reste de 0.5 quel que soit le mode d'initialisation (Tab. 2 et 3). D'autre part, nous avons mis en évidence (image 2(b)) la capacité de l'algorithme à bien identifier le cas dur et le cas flou. En effet la réalisation obtenue par les paramètres estimés à partir de l'image 2(b) donne une image

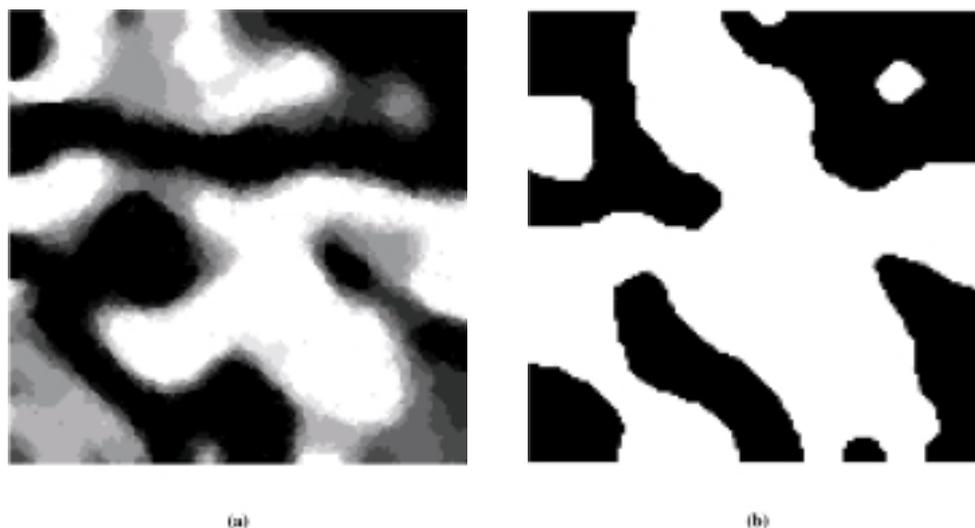


Figure 2. Champs markoviens simulés par les paramètres du tableau 2

Tableau 3. Paramètres estimés avec différentes initialisations

Image Initiale	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_1^f$	$\hat{\alpha}_2^f$	$\hat{\alpha}_3^f$	$\hat{\alpha}_4^f$	Image Finale	% flou image finale
2(a)	4.72	4.75	4.66	4.70	5.23	5.20	4.5	4.5	3(a)	45.54%
2(a)	1.95	1.99	1.94	1.93	2.39	2.38	2.43	2.38	3(b)	35.59%
2(b)	4.17	4.14	4.03	4.13	3.99	3.99	3.98	3.99	3(c)	0.92%

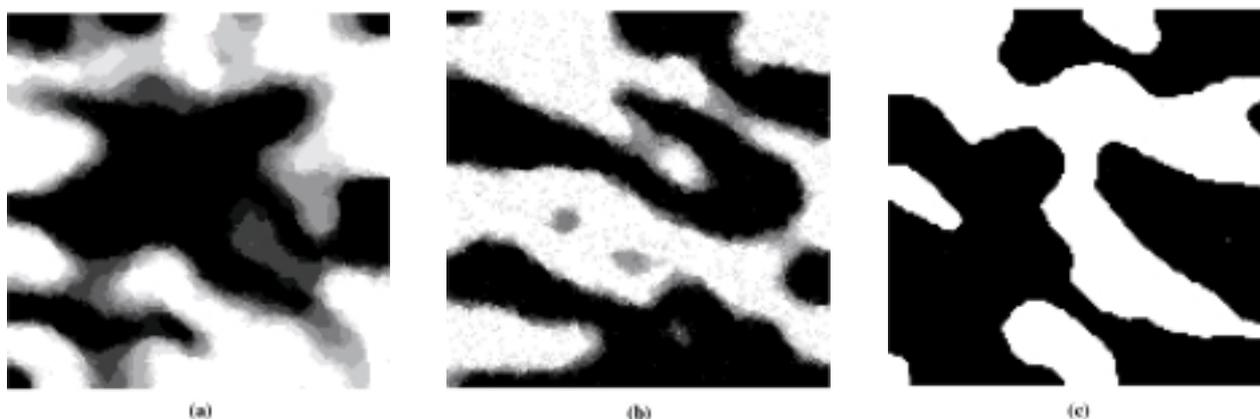


Figure 3. Champs markoviens simulés par les paramètres estimés du tableau 3

à dominante dure (coefficients des cliques durs égaux aux coefficients des cliques flous). Nous retrouverons cette propriété dans le cas non supervisé. Remarquons enfin que cette procédure d'estimation est la principale cause d'augmentation du temps de calcul puisque nous appliquons à chaque itération l'échantillonneur de Gibbs.

6.1.3. Résultats de segmentation trispectrale par champs markoviens flous en mode non-supervisé

Nous avons bruité l'image floue 2(a) à 10 niveaux, au moyen de trois bruits différents à moyennes discriminantes et à variances

identiques (dits bruits «MD» voir tableau 4). D'après (10), les paramètres des moyennes et variances relatives aux classes dures suffisent à caractériser les lois d'attache aux données. Les trois bandes (bandes 1,2 et 3) sont représentées respectivement par les images 4(b), 4(c), 4(d) de la figure 4. Le paramètre de la loi *a priori* est celui de l'image 4(a) $\alpha^c = \{4,4,4,4,4.5,4.5,4.5,4.5\}$. Nous présentons tableau 4 les résultats des estimations à partir des observations $\mathbf{Y} = \mathbf{y}$. L'algorithme d'estimation des lois d'attache aux données est donné en section 4.3.3. Les matrices de variances covariances des classes flous sont calculées d'après (13). Pour la segmentation, nous avons comparé un algorithme de type ICM avec une procédure MPM «écart en valeur abso-

Tableau 4. Estimation des paramètres de lois d'attache aux données des images 4 b-d

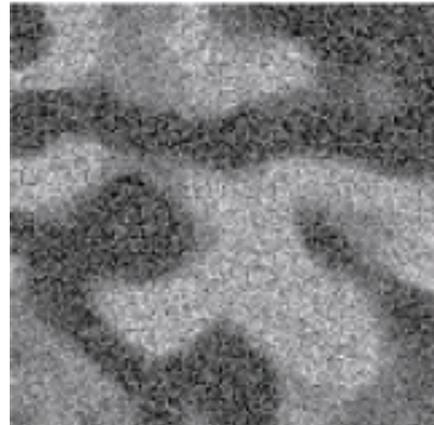
Image	μ_0	μ_1	σ_0	σ_1	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$
1: 4(b)	104	152	16	16	106.21	148.77	15.89	16.24
2: 4(c)	120	152	16	16	121.19	149.92	15.53	15.68
3: 4(d)	120	136	16	16	120.64	135.43	15.35	15.41

Tableau 5. Estimation des paramètres de la loi a priori

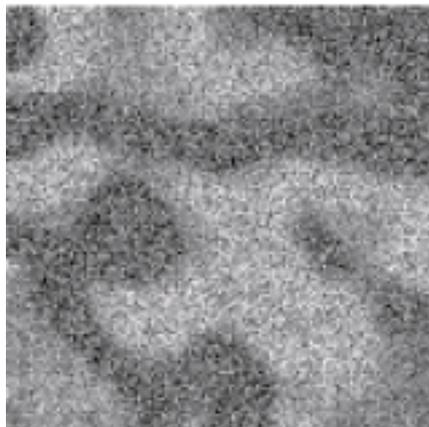
	α_1	α_2	α_3	α_4	α_1^f	α_2^f	α_3^f	α_4^f
valeurs réelles	4	4	4	4	4.5	4.5	4.5	4.5
valeurs estimées	1.76	1.82	1.68	1.75	2.12	2.09	2.14	2.10



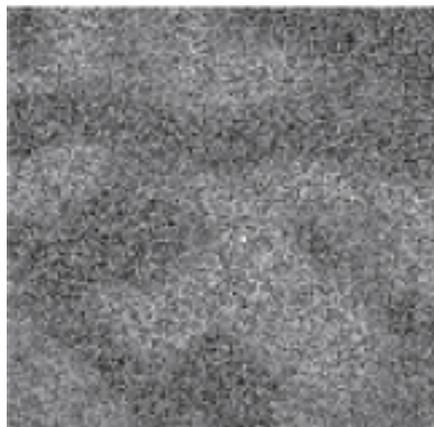
(a) Image initiale



(b)



(c)



(d)

Figure 4. Images initiales a) et bruitées dans trois bandes (b,c,d)

lue» décrite en section 5.2. Les images segmentées résultantes sont présentées Figure 5. ICM est plus efficace en terme de taux d'erreur. Cet algorithme conserve mieux l'information floue que la segmentation MPM (voir tableau 6), même si dans les deux cas, on observe une perte de l'information floue par rapport à l'image initiale (troisième colonne du tableau 6). Cette perte d'information est certainement due aux limitations du nombre d'itérations en raison d'un temps de calcul prohibitif (pour le

MPM: environ une heure pour une image de taille 128×128). Un autre résultat intéressant est illustré par la figure 6: nous avons segmenté en mode non supervisé un champ «dur» bruité de manière identique. L'image finale obtenue ne contient quasiment pas de flou (1.17%) tout comme dans l'image initiale (1.42%). Ceci montre que la méthode ne génère pas de classes floues «artificielles». Une segmentation floue sur une image réelle n'induit donc pas de classes floues parasites.

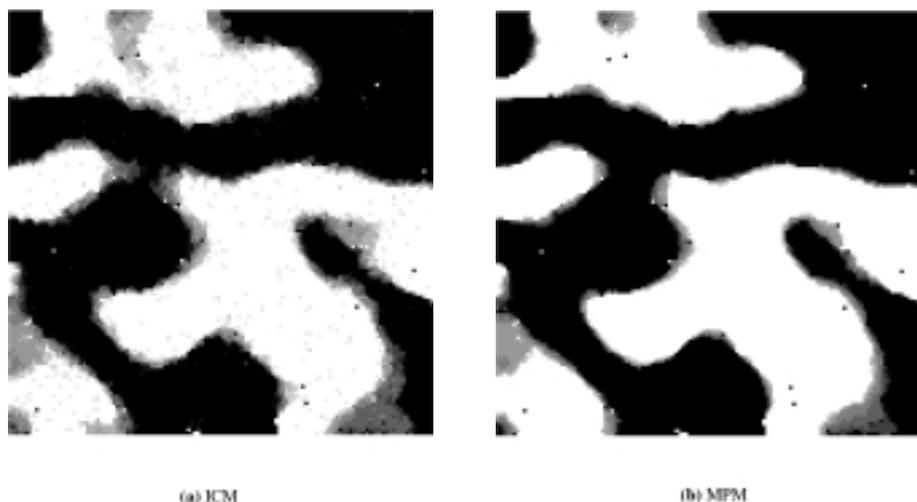


Figure 5. Images segmentées par ICM et MPM

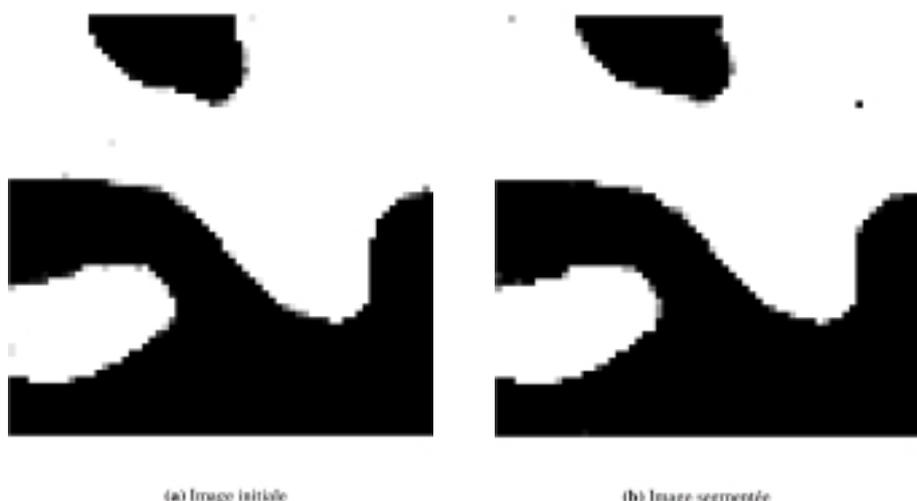


Figure 6. Segmentation d'un champ dur par une méthode floue multispectrale

Tableau 6. Comparaison des méthodes MPM et ICM

méthode	% flou image initiale	% flou image segmentée	erreur
ICM	44.16	21.77	8.24 %
MPM	44.16	9.63	10.00 %

6.2. Images multibandes astronomiques

6.2.1. Résultats de segmentation multispectrale d'une image astronomique en mode non-supervisé

Nous avons appliqué notre approche à une image astronomique sur un nombre plus élevé de bandes (6 bandes spectrales) en

mode non-supervisé. Ces images d'une galaxie présentées figure 7, sont extraites de la base de données de l'Observatoire de Strasbourg. Remarquons la différence d'aspect visuel entre les deux premières bandes et les quatre autres. Ceci illustre la capacité de notre algorithme non seulement à mettre en évidence l'aspect flou du terrain mais aussi à exploiter la complémentarité de l'information entre les bandes. Les tableaux 7 et 8 récapitulent les valeurs estimées des hyperparamètres. La figure 8 montre l'image segmentée. La carte de segmentation finale offre un aspect relativement homogène et conserve la nature « dure » des zones centrales (avec notamment les amas situés sur la gauche en bas de la figure) et périphériques (espace interstellaire). Remarquons que les valeurs des variances $\hat{\sigma}_0$ dans les deux premières bandes sont proches de zéros, ce qui n'affecte pas notre méthode en terme de convergence.

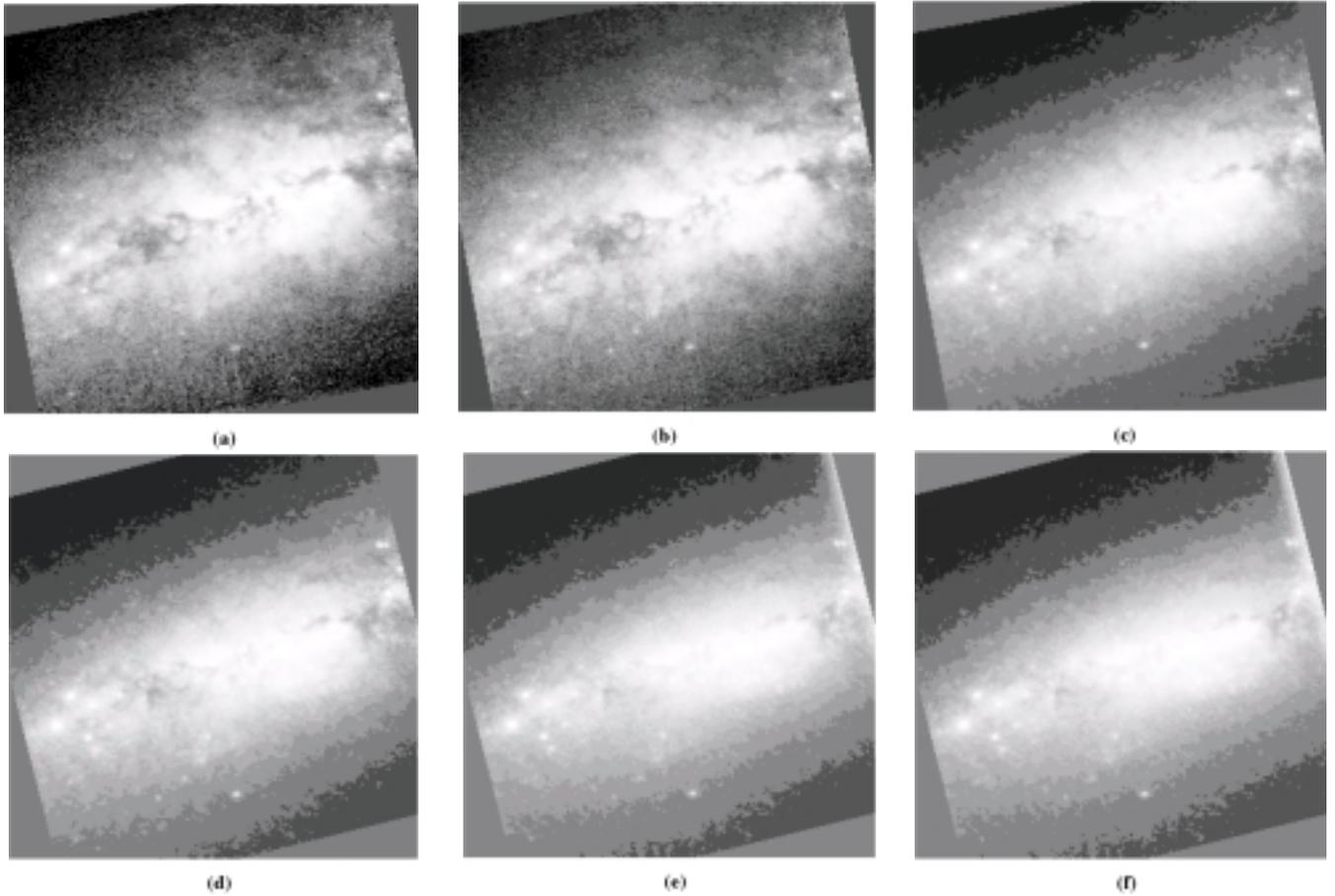


Figure 7. Images multibandes astronomiques de la galaxie M82. Il s'agit d'images de Hubble prises à diverses longueurs d'ondes infrarouges (1.08, 1.13, 1.64, 1.66, 2.12 et 2.15 microns). Les images ont été prétraitées et mosaïquées. Les temps de poses sont de l'ordre de 960 secondes. Il peut être intéressant de traiter ensemble les bandes 1.13, 1.66, et 2.15 qui correspondent au continuum stellaire puis toutes les bandes qui correspondent à des raies d'émission de gaz.

Tableau 7. Estimation des paramètres d'attache aux données

Images	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$
bande 1	0.03	1.60	0.06	0.84
bande 2	0.04	2.12	0.08	1.19
bande 3	0.13	16.61	0.37	4.33
bande 4	0.12	15.81	0.34	4.53
bande 5	0.14	29.02	0.46	10.15
bande 6	0.12	27.02	0.41	9.26

Tableau 8. Estimation des paramètres de la loi a priori

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_1^f$	$\hat{\alpha}_2^f$	$\hat{\alpha}_3^f$	$\hat{\alpha}_4^f$
2.07	2.17	1.98	2.04	2.46	2.49	2.43	2.44



Figure 8. Résultats de la segmentation de la galaxie M82 par ICM. On cherche à mettre en évidence dans ces images les vents stellaires (sur les images à raie d'émission), les amas stellaires à distinguer du fond de la galaxie et des étoiles du champ... Le problème est compliqué par l'extinction liée aux poussières plus ou moins fortes suivant les longueurs d'ondes et que l'on cherche à analyser en intégrant la notion de flou dans le modèle a priori. Les zones de fortes extinctions sont aussi intéressantes à détecter. Les premiers résultats sont jugés pertinents car ils permettent d'étudier l'image par zones appartenant à la même classe (au lieu de toute l'image). Néanmoins les jets stellaires ainsi que les zones d'extinction ne sont pas correctement détectées. Sur ces images, les résultats obtenus dans [33] avec une modélisation en 13 classes dures sont de meilleure qualité quant à l'interprétation physique

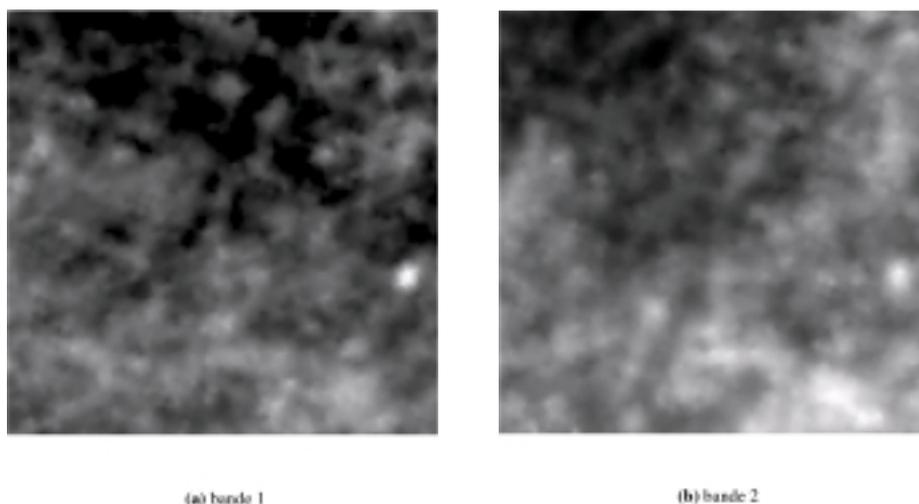


Figure 9. Images astronomiques sur 2 bandes à 100 et 160 microns. Il s'agit de champs de haute altitude de notre galaxie dans le domaine infrarouge, pour lesquels on cherche à mesurer les fluctuations induites par la présence de galaxies extérieures à la voie lactée, non résolues, dans le cadre d'une étude cosmologique. Des fluctuations, provoquées par la poussière interstellaire de notre galaxie, gênent l'observation. Cette poussière interstellaire possède une structure filamentaire que nous souhaitons mettre en évidence par une analyse multibande (en bas à droite).

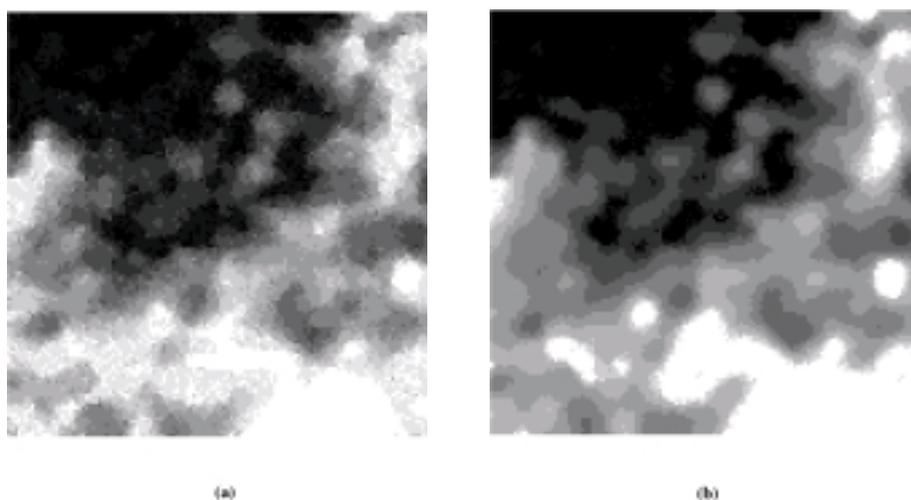


Figure 10. Segmentation par ICM (a) et MPM (b) des images 9(a) et 9(b). L'objectif astronomique était de faire apparaître des zones filamentaires. Les premiers résultats sont jugés pertinents car ils permettent d'étudier l'image par zones appartenant à la même classe (au lieu de toute l'image), et d'en extraire des paramètres cosmologiques par classes, dures ou floues.

6.2.2. Résultats de segmentation d'images astronomiques bi-spectrales en mode non-supervisé

La figure 9 montre deux images qui se prêtent bien « visuellement » au modèle flou. Nous avons comparé les résultats d'une segmentation sur critère MPM et MAP (algorithme ICM). À l'instar des résultats obtenus sur des images de synthèse, la segmentation MPM semble donner des contours plus réguliers et

restitue une image segmentée plus homogène : les zones floues intermédiaires présentent ainsi un aspect lisse. Dans ce cas, la segmentation MPM rend mieux compte de la gradation floue entre les zones dures. Remarquons que l'écart entre les paramètres des cliques durs et des cliques flous est plus important que précédemment, dénotant une plus grande quantité de flou dans les images Figure 10.

Tableau 9. Paramètres estimés des images Figure 9

Images	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$
bande 1	0.37	0.49	0.03	0.04
bande 2	0.59	0.91	0.04	0.08

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_1^f$	$\hat{\alpha}_2^f$	$\hat{\alpha}_3^f$	$\hat{\alpha}_4^f$
1.55	1.59	1.51	1.54	2.23	2.24	2.19	2.16

7. Conclusion

Au cours de cette étude, nous avons étendu aux données multispectrales la segmentation statistique non supervisée basée sur un modèle markovien flou. Le modèle de champ s'appuie sur une densité de probabilité par rapport à la mesure constituée d'une composante continue (mesure de Lebesgue) et de composantes discrètes (masses de Dirac représentant les classes dures). Nous avons supposé les capteurs corrélés, en exprimant les lois du bruit par des gaussiennes : les matrices des variances-covariances correspondant aux sites flous s'expriment en fonction des matrices de variances-covariances des classes dures. La segmentation proprement dite est précédée d'une étape d'estimation des hyperparamètres : l'algorithme du gradient stochastique sert à déterminer les paramètres de la loi a priori (pour des cliques d'ordre deux), tandis que la méthode des moments empiriques permet d'extraire les paramètres des lois gaussiennes au moyen des observations et des réalisations du champ caché obtenues à chaque itération. Nous avons proposé plusieurs méthodes d'estimation des vecteurs moyennes et matrices de variances-covariances s'appuyant sur des techniques de moindres carrés, résolvant ainsi le problème d'une forte présence de flou dans l'image. Lors de la segmentation, nous avons comparé l'algorithme MPM, avec la technique ICM – plus rapide. De manière à établir un bon compromis entre le temps de calcul et la continuité de l'information, nous avons travaillé sur 10 niveaux de flou. Dans un premier temps, les méthodes ont été testées sur des images de synthèse trispectrales, bruitées au moyen d'un bruit « MD ». S'il s'avère que la méthode ICM donne de meilleurs résultats en terme de taux d'erreur, la méthode MPM – si elle reste prohibitive en terme de temps de calcul – donne une image à l'aspect plus lisse, conforme à la continuité de l'information floue de l'image « vérité-terrain ». D'autre part, nous avons mis en évidence le fait que le modèle n'introduit pas de classes floues parasites, ce qui atteste de la présence de ces classes dans les images réelles traitées. Enfin, nous avons validé notre modèle et les traitements correspondants sur des données astronomiques, constituées de deux et six bandes respectivement. Les résultats encourageants montrent l'intérêt des méthodes non supervisées floues dans le champ de l'astronomie, en restituant ainsi une gradation floue entre les zones homogènes du ciel. Concernant

les lois a priori, on peut envisager d'étendre notre technique à d'autres modèles plus performants en terme de temps de calcul, comme les chaînes ou les arbres de Markov [32, 33]. Il est également envisageable d'étudier d'autres lois paramétriques [30, 31], pour caractériser les données bruitées.

Références

- [1] D. Egret, « Et le ciel devint virtuel », *La recherche*, n° 369, pp. 58-61, Novembre 2003. <http://wwwusr.obspm.fr>
- [2] J.C. Bezdek, « Pattern Recognition and Fuzzy Objective Function Algorithm », *Plenum Press*, New York, 1981.
- [3] I. Gath, A.B. Geva, « Unsupervised Optimal Fuzzy Clustering », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, n° 7, pp. 773-781, 1989.
- [4] W. Pedrycz, « Fuzzy Sets in Pattern Recognition: Methodology and Methods », *Pattern Recognition*, Vol. 23, n° 1/2, pp. 121-146, 1990.
- [5] L.A. Zadeh, « Fuzzy Sets », *Information and Control*, Vol. 8, pp. 338-352, 1965.
- [6] J.T. Kent, K.V. Mardia, « Spatial Classification Using Fuzzy Membership », *IEEE, Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, n° 5, pp. 659-671, 1988.
- [7] I. Bloch, H. Maître, « Fusion de données en traitement d'images : modèles d'informations et décisions », *Traitement du Signal*, Vol. 11 – n° 6, pp. 435-446, 1994.
- [8] B. Bouchon Meunier, « La Logique Floue et ses Applications », *Adison-Wasley-France*, Paris, 1995.
- [9] G. Shafer, « A Mathematical Theory of Evidence », *Princeton University Press*, 1976.
- [10] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, « Multisensor Images Segmentation Using Dempster-Shafer Fusion in Markov Field Context », *IEEE Trans. on Geosciences and Remote Sensing*, Vol. 39, n° 8, pp. 1789-1798, 2001.
- [11] J. Besag, « Spatial Interaction and the Statistical Analysis of lattice Systems », *Journal of Royal Statistical Society*, Ser. B 36, pp. 192-242, 1974.
- [12] W. Pieczynski, « Modèles de Markov en Traitement d'Image », *Traitement du Signal*, Vol. 20, n° 3, pp. 255-278, 2003.
- [13] N. Giordana, W. Pieczynski, « Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, n° 5, pp. 465-475, 1997.
- [14] J.-M. Laferté, P. Pérez, F. Heitz, « Discrete Markov Image Modeling and Inference on the Quad-tree », *IEEE Trans. on Image Processing*, Vol. 9, n° 3, pp. 390-404, March 2000.
- [15] J.-N. Provost, « Classification Bathymétrique en Imagerie Multi-spectrale Spot », *Thèse de l'Université de Bretagne Occidentale*, <ftp://picabia.u-strasbg.fr/pub/www/collet/PhD/>, 2001.
- [16] F. Salzenstein, W. Pieczynski, « Parameter Estimation in Hidden Fuzzy Markov Fields and Image Segmentation », *Graph. Models Process.*, Vol. 59, n° 4, pp. 205-220, 1997.
- [17] Ch. Collet, J.-N. Provost, P. Rostaing, P. Pérez, P. Bouthemy, « Segmentation bathymétrique d'images multispectrales SPOT », *Traitement du signal*, Vol. 18, n° 1, pp. 1-14, septembre 2001.
- [18] J.-N. Provost, Ch. Collet, P. Rostaing, P. Pérez, P. Bouthemy, « Hierarchical Markovian Segmentation of Multispectral Images for the Reconstruction of Water Depth Maps », *Computer Vision and Image Understanding*, Vol. 93, Issue 2, pp. 155-174, February 2004.
- [19] G.-F. Hugues, « On the mean accuracy of statistical pattern recognizers », *IEEE Trans. Information Theory*, Vol. 14(1), pp. 55-63, 1968.
- [20] H. Caillol, A. Hillion, W. Pieczynski, « Fuzy Random fields and Unsupervised Image Segmentation », *IEEE Trans. on Geosciences and Remote Sensing*, GE-31, n° 4, pp. 801-810, 1993.

- [21] W. Pieczynski, «Champs de Markov Cachés et Estimation Conditionnelle Itérative», *Traitement du Signal*, Vol. 11, n°2, pp. 141-153, 1994.
- [22] W. Pieczynski, J.M. Cahen, «Champs de Markov Cachés Flous et Segmentation d'Images», *Revue Statistique Appliquée*, Vol. 42, n°2, pp. 13-31, 1994.
- [23] F. Salzenstein, «Modèles Markoviens Flous et Segmentation Statistique Non Supervisée d'Images», *Thèse de l'Université de Rennes 1*, 1996.
- [24] R.-O. Duda, P.E. Hart, «Pattern classification and scene analysis», *Wiley Interscience*, 1st Edition, 1973.
- [25] S. Ruan, B. Moretti, J. Fadili, D. Bloyet, «Fuzzy Markovien Segmentation in Application of Magnetic Resonance Images», *Computer Vision and Image Understanding*, Vol. 85, pp. 54-69, 2002.
- [26] L. Younes, «Parametric Inference for Imperfectly Observed Gibbsian Fields», *Probability Theory and Related Fields*, Vol. 82, pp. 625-645, 1989.
- [27] F. Gustafsson, «Adaptive filtering and Change Detection», *Wiley*, Londres, 2000.
- [28] S. Geman, D. Geman, «Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images», *IEEE, Trans. On Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721-741, 1984.
- [29] J. Maroquin, S. Mitter, T. Poggio, «Probabilistic Solution of Ill-posed Problems in Computational Vision», *Journal of The American Statistical Association*, Vol. 82, pp. 76-89, 1987.
- [30] Y. Delignon, A. Marzouki, W. Pieczynski, «Estimation of Generalized Mixtures and its Application in Image Segmentation», *IEEE Trans. on Image Processing*, Vol. 6(10), pp. 1364-1375, 1997.
- [31] W. Pieczynski, J. Bouvrais, C. Michel, «Estimation of Generalized Mixtures in the Case of Correlated Sensors», *IEEE Trans. on Image Processing*, Vol. 9, n°2, pp. 308-311, 2000.
- [32] M.S. Crouse, R.D. Novak, R.G. Baraniuk, «Wavelet-Based Signal Processing using Hidden Markov Models», *IEEE Trans. on Signal Processing*, Vol. 46, n° 4, pp. 886-902, 1998.
- [33] C. Collet, M. Louys, J.-N. Provost and A. Oberto, «Fusion of Astronomical Multiband Images on a Markovian Quadtree», *Information Fusion*, Annapolis, Maryland, USA, 2002.





Fabien **Salzenstein**

Fabien SALZENSTEIN est diplômé de l'ENST Bretagne en 1990. Après avoir exercé comme ingénieur dans le domaine du trafic aérien à Thomson CSF, il reçoit le titre de doctorant de l'Université de Rennes 1 en 1996. Le sujet de thèse portait sur la segmentation des images : comparaison de méthodes statistiques et approche markovienne floue. Il est actuellement maître de conférence de l'Université Louis Pasteur, Strasbourg et poursuit ses travaux de recherche au laboratoire PHASE UPR CNRS 292. Ses domaines de recherche couvrent la segmentation statistique floue appliquée à l'imagerie astronomique, la fusion de données et l'analyse temps-fréquence pour la caractérisation de surfaces en microscopie interférométrique.



Christophe **Collet**

Christophe COLLET est né en 1966 en France. Après avoir obtenu son DEA en traitement du signal à l'Université Paris-Sud-Orsay en 1989, il a poursuivi en thèse à l'ENST de Bretagne et à l'École Navale. Nommé maître de conférences à l'Université de Bretagne Occidentale en 1993, il dirige le Groupe de Traitement du Signal de l'École Navale de 1994 à 2000 où il s'intéresse aux techniques d'inférence bayésienne et à la classification markovienne d'images SONAR en collaboration avec l'IRISA. Il soutient son Habilitation à diriger des Recherches en 2000 et est nommé Professeur à l'École Nationale Supérieure de Physique de Strasbourg en 2001. Il poursuit actuellement ses activités de recherche au sein du LSIIT UMR CNRS 7005 dans l'équipe Modèles, Images, Vision. Ses travaux concernent la segmentation et la classification hiérarchique multi-images (imagerie médicale multimodale, imagerie astronomique hyperspectrale, imagerie polarimétrique), l'inférence bayésienne, l'estimation robuste, les modélisations markoviennes.

<http://picabia.u-strasbg.fr/lisit/perso/collet/Perso/ColletPerso.htm>



Matthieu **Petremand**

Matthieu PETREMAND a obtenu son DEA d'informatique en 2003 à l'Université Louis Pasteur, STRASBOURG. Son stage de DEA, portait sur la segmentation d'images astronomiques par approche markovienne floue. Il est actuellement doctorant en première année au laboratoire LSIIT UMR CNRS 7005 à Strasbourg sous la direction de Christophe COLLET (LSIIT) et de Françoise GENOVA (Observatoire de Strasbourg). Il travaille sur le traitement d'images astronomiques hyperspectrales, en collaboration avec l'observatoire de Strasbourg, notamment sur les problèmes de visualisation, de réduction de données et de segmentation.

