INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Classification d'Images Multibandes par Modèles Variationnels

Christophe Samson — Laure Blanc-Féraud — Gilles Aubert — Josiane Zerubia

N° 4010

Septembre 2000

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4010--FR

Classification d'Images Multibandes par Modèles Variationnels

Christophe Samson, Laure Blanc-Féraud, Gilles Aubert*, Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 4010 — Septembre 2000 — 59 pages

Résumé : Dans ce rapport, nous proposons deux modèles variationnels pour la classification d'images multibandes.
Le premier modèle présenté repose sur la minimisation d'une famille de critères dont la suite de solutions converge vers une partition des données composée de classes homogènes séparées par des contours réguliers.
Parallèlement à cette approche, nous avons développé un second modèle de classification mettant en jeu un ensemble de régions et contours actifs. Nous utilisons une approche par ensembles de niveaux pour définir le critère à minimiser. Le critère proposé contient des termes reliés à l'information sur les régions ainsi qu'à l'information sur les contours.
L'imagerie multispectrale permet de prendre en compte, et de combiner, l'information des différentes bandes spectrales renvoyée par un capteur satellitaire ou aérien. L'extension au cas multispectral intervient à des niveaux différents pour les deux modèles proposés dans ce rapport. Nous traitons une application réelle sur une scène SPOT en mode XS pour laquelle nous disposons d'une vérité terrain. Nous comparons les deux modèles variationnels que nous proposons à d'autres approches dont un modèle stochastique hiérarchique, récemment développé à l'IRISA au sein du projet VISTA.

Mots-clés : Modèles variationnels, classification d'images, imagerie multispectrale, Γ-convergence, formulation par ensembles de niveaux, régions actives, contours actifs.

* Laboratoire J.A. dieudonné UMR 6621 CNRS, Université de Nice-Sophia Antipolis, 06108 Nice Cedex 2, FRANCE

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Multiband Image Classification Using Variational Models

Abstract: Herein, we propose two variational models for multiband image classification.
The first model we propose herein is based on the minimization of a criterion family whose set of solutions is converging to a partition of the data set composed of homogeneous regions with regularized boundaries.
The second model we propose is based on a set of active regions and contours. We use a level set formulation to define the criterion we want to minimize. Each class and its associated set of regions and boundaries is defined thanks to a level set function.
The extension of these two models to the multispectral case is presented in this report. The extension of the dynamic model is quite straightforward whereas the one of the first model is more tricky.
We have conducted experiments on SPOT XS data whose ground truth is given. We compare the results we obtain with other approaches, in particular we compare the proposed models to a stochastic hierarchical model recently developed within the VISTA group from IRISA.

Key-words: Variational models, classification, multiband imaging, Γ-convergence, level set formulation, active regions, active contourss.

Table des matières

1 Position du problème

1.1 Introduction

Les modèles variationnels ainsi que les approches par équations aux dérivées partielles ont montré leur efficacité dans de nombreux problèmes tels que la restauration et la détection de contours [1, 19, 54, 59]. En revanche, la classification d'images n'a pas suscité beaucoup de travaux dans le cadre variationnel. La raison en est certainement que la notion discrète d'étiquetage n'est pas adaptée au formalisme continu. L'image originale est dégradée, par conséquent la classification doit être couplée avec un processus de restauration.

La classification concerne de nombreuses applications en traitement des images, et consiste à trouver une partition d'une image en régions homogènes : les classes. La classification est étroitement liée à la segmentation (voir par exemple [54, 55]) dans le sens où nous obtenons une partition de l'image, mais la classification présente un processus supplémentaire d'étiquetage. Chaque région de la partition obtenue est constituée d'un ensemble de pixels ayant la même étiquette. Nous supposerons que le nombre de classes ainsi que leurs paramètres respectifs sont connus (classification *supervisée*), grâce à une pré-estimation par exemple. L'attribut de l'image qui nous intéresse est la distribution spatiale d'intensité (ou niveau de gris), mais d'autres attributs peuvent être envisagés comme ceux concernant la texture par exemple (utilisation de filtres de Gabor [37] ou des matrices de cooccurence [21]).

De nombreux modèles de classification ont été développés, que ce soit par des approches structurales comme la croissance de régions [58], ou par des approches stochastiques comme dans [9, 14, 23, 22, 38, 45, 49].

Dans [64, 67, 65, 68], nous avons développé deux modèles de classification fondés sur des approches variationnelles. Le premier modèle que nous avons proposé [64, 67] repose sur la minimisation d'une famille de critères dont la suite de solutions converge vers une partition des données composée de classes homogènes séparées par des contours réguliers. Cette approche entre dans le cadre des problèmes à discontinuité libre (*free discontinuity problems*) et fait appel à des notions de convergence variationnelle telle que la théorie de la Γ-convergence [31, 52]. La famille de fonctionnelles que nous proposons de minimiser contient un terme de régularisation, ainsi qu'un terme de classification. Lors de la convergence de cette suite de critères, le modèle change progressivement de comportement en commençant par restaurer l'image avant d'entamer le processus d'étiquetage des pixels.

Parallèlement à cette approche, nous avons développé un second modèle de classification mettant en jeu un ensemble de régions et contours actifs [64, 65, 68]. Nous utilisons une approche par ensembles de niveaux pour définir le critère à minimiser, cette approche ayant déjà suscité de nombreux travaux dans le cadre de la segmentation d'images. Chaque classe, et son ensemble de régions et contours associé, est défini à travers une fonction d'ensemble de niveaux. Le critère contient des termes reliés à l'information sur les régions ainsi qu'à l'information sur les contours. Nous aboutissons à la résolution d'un système d'équations aux dérivées partielles couplées et plongées dans un schéma dynamique. L'évolution de chaque région est guidée par un jeu de forces permettant d'obtenir une partition de l'image composée de classes homogènes et dont les frontières sont lisses.

Dans ce rapport, nous proposons une extension de ces deux modèles variationnels de classification au cas de données multibandes, comme par exemple des données SPOT en mode XS constituée de trois images de la même scène prises à différentes longueurs d'ondes. Le couplage entre les bandes intervient à des niveaux différents pour les deux modèles. Concernant le modèle reposant sur la minimisation d'une suite de fonctionnelles, le couplage entre les bandes s'effectue au niveau du calcul du gradient sur les données, ainsi qu'au travers du potentiel de contrainte de niveaux. Pour le modèle de régions/contours actifs, la combinaison d'information fournie par les différentes bandes intervient au niveau du terme d'attache aux données.

1.2 Notations

Généralités

- Les inconnues des modèles présentés au cours de ce travail sont des images définies sur un ensemble Ω que nous considérons inclus dans $\mathbb{R}^n$, $n = 2,3$. La plupart du temps ces inconnues seront représentées par des fonctions à variables continues même s'il s'agit d'images qui, en pratique, renvoient à des notions discrètes.
  D'un point de vue discret, le support Ω d'une image est modélisé par une grille de N pixels (ou *sites*) de taille nbl lignes par nbc colonnes, tel que $nbl * nbc = N$. $S = \{s\}_{s=1...N}$ est l'ensemble des sites ou pixels de l'image discrète, et $card(S) = N$. Les coordonnées d'un point sur la grille seront représentées par le couple (m,n) élément de $([1...nbl],[1...nbc])$.

- Dans ce qui suit, nous nous intéressons à la radiométrie des images. On représentera par la fonction continue $I : \Omega \rightarrow \mathbb{R}$, avec $I \in L^2(\Omega,\mathbb{R})$, l'intensité lumineuse associée à chaque pixel des données collectées. Ω, le support de l'image de frontière $\partial\Omega$, est supposé ouvert borné dans $\mathbb{R}^n$, avec $n = 2{,}3$. La fonction I représente les données collectées (comme une image satellitaire ou aérienne par exemple).

- Dans un cadre continu, l'information contenue dans une image est modélisée par une fonction $f : \Omega \to \mathbb{R}$. Par défaut, $f \in L^\infty(\Omega,\mathbb{R})$. Dans le cas continu (respectivement discret), on notera $f(x)$ (resp. f_{mn}) la valeur de f au point $x \in \Omega$ (resp. la valeur de f au pixel de coordonnées (m,n)).

- Par abus de langage, et afin d'alléger les notations, nous confondrons parfois la fonction f avec sa valeur $f(x)$ comme par exemple dans l'expression des termes des fonctionnelles proposées où $\int_\Omega (f - I)^2 dx$ est utilisé en lieu et place de $\int_\Omega (f(x) - I(x))^2 dx$.

Définitions

- Id est la matrice identité.

- A^T, A^* et A^{-1} sont respectivement la transposée, l'adjointe et l'inverse de la matrice A.

- $\nabla = (\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n})^T$ est l'opérateur gradient.

- L'expression $f \sim \mathcal{N}(m,s)$ signifie que f suit une loi gaussienne de moyenne m et d'écart-type s.

- RSB est le Rapport Signal sur Bruit (SNR en Anglais) permettant de mesurer le rapport de variances de deux signaux s_a et s_b et défini par:

$$RSB = 10 \log_{10} \left[\frac{\sigma_a^2}{\sigma_{a-b}^2} \right],$$

avec σ_a^2 la variance du signal s_a et σ_{a-b}^2, la variance de la différence $s_a - s_b$. Ce rapport sera utilisé par la suite pour caractériser le niveau de bruit d'une image dont la version originale non bruitée sera représentée par s_a et la version bruitée par s_b.

- $div(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}$ est l'opérateur de divergence appliqué à f.

- $\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial {x_i}^2}$ est l'opérateur Laplacien appliqué à f.

- $W^{1,p}(\Omega)$ est l'espace de Sobolev des fonctions dont la dérivée (au sens des distributions) est dans $L^p(\Omega)$.
  On note $W^{1,2}(\Omega) = H^1(\Omega)$.

- La Variation Totale (TV) de f est définie par:

$$TV(f) = \int_\Omega |Df| = sup\{\int_\Omega f.div(g);\ g \in C_0^1(\Omega,\mathbb{R}^n)\ et\ |g(x)| \leq 1\}.$$

  si $f \in C^1(\Omega)$ alors $TV(f) = \int_\Omega |\nabla f|$, voir [32] par exemple.

- $BV(\Omega)$ est l'espace des fonctions de Variation Totale Bornée (BV) [32]:

$$BV(\Omega) = \{f \text{ telles que } TV(f) < +\infty\}.$$

  Cet espace contient les fonctions pouvant être discontinues le long des courbes ($n = 2$), ce qui n'est pas le cas des espaces de Sobolev, et par conséquent $BV(\Omega)$ est l'espace adéquat pour l'étude des problèmes impliquant des ensembles de discontinuités tels que ceux liés à la segmentation d'images.

- Le périmètre de l'ensemble $A \subset \Omega$, au sens de De Giorgi [31], est défini par $Per_\Omega(A) = TV(\chi_A)$ où χ_A est la fonction caractéristique de A.

- $\mathcal{H}^{n-1}(\partial A)$ est la mesure $n-1$ dimensionnelle de Hausdorff de l'ensemble ∂A, et si la frontière ∂A de A est lisse, nous avons:

$$\mathcal{H}^{n-1}(\partial A \cap \Omega) = Per_\Omega(A) = TV(\chi_A).$$

- **Γ-convergence** [31, 52]
  Définition: Soit X un espace métrique, Y ⊂ X.
  Soit g_ε: Y $\longrightarrow$ [0;+∞] une famille de fonctions dépendant de ε, $\varepsilon > 0$,

alors :
g_ε **Γ-converge** vers g quand $\varepsilon \to 0$ si :

$$\forall x \in X \begin{cases} \forall x_\varepsilon \to x, & \liminf_{\varepsilon\to 0} g_\varepsilon(x_\varepsilon) \geq g(x) \\ \exists x_\varepsilon \to x, & \limsup_{\varepsilon\to 0} g_\varepsilon(x_\varepsilon) \leq g(x). \end{cases}$$

Acronymes et abréviations

- EDP est l'acronyme d'"Equation aux Dérivées Partielles".
- ENZ est l'acronyme de "Ensembles de Niveaux Zéro" désignant l'ensemble:

$$\bigcup_{i=1}^{K} \{x \in \Omega / \Phi_i(x) = 0\},$$

  pour Φ_i la fonction associée à la $\mathrm{i}^{\text{ième}}$ classe.
- Le terme $p.p$ signifie "presque partout", i.e. partout sauf sur un ensemble de mesure de Lebesgue nulle.

Hypothèses et notations spécifiques à la classification

Nous définissons ici quelques notations utilisées ultérieurement en ce qui concerne la classification.

- L_s représente l'étiquette attribuée au site s.
- Λ est l'ensemble des étiquettes admissibles, $card(\Lambda) = K$ et $L_s \in \Lambda$.
- Lorsque cela n'est pas précisé, les indices i et j relatifs à la classification appartiennent à l'ensemble $[1,...,K]$.
- C_i est la $i^{ème}$ classe caractérisée par une étiquette dans Λ.
- Chaque pixel est caractérisé par son niveau d'intensité observé I_s ($I(x)$ en variable continue).

- La distribution est gaussienne à l'intérieur de chaque classe $C_{i;i=1..K}$, de moyenne μ_i et d'écart type σ_i. Dans le cas monospectral, nous utiliserons la notation vectorielle suivante :

  $$\vec{\mu} = \{\mu_i\}_{i=1..K} \quad \text{et} \quad \vec{\sigma} = \{\sigma_i\}_{i=1..K}.$$

  Dans le cas multispectral, où les données sont constituées de P bandes, $\vec{I} = [I^1,...,I^P]^T$, nous utiliserons les notations:

  $$\vec{\mu_i} = \{\mu_i^p\}_{p=1...P} \quad \text{et} \quad \vec{\sigma_i} = \{\sigma_i^p\}_{p=1...P} \quad \text{pour} \quad i = 1,...,K.$$

- Le nombre K et les vecteurs de paramètres des classes sont connus ou estimés au préalable (i.e. classification *supervisée*).

2 Classification par approche variationnelle

Nous rappelons ici les principales caractéristiques des deux modèles variationnels que nous avons développés dans [65, 67].

2.1 Classification avec restauration

Une grande partie des signaux collectés, en l'occurrence les images satellitaires ou aériennes, ne constituent qu'une version dégradée du signal original. Le signal émis par la source est soumis à deux principaux types de dégradation: l'influence de la réponse impulsionnelle du système physique d'acquisition du signal (ajout de flou sur les données, transformation des données...), et l'information parasite (ou *bruit*) venant se greffer sur l'information originelle. La réponse impulsionnelle du système physique sera notée par un opérateur $\mathcal{R}$ de $L(\Omega,\mathbb{R})$ l'ensemble des fonctions linéaires définies sur Ω et à valeurs réelles. Le bruit possède différentes natures (gaussien/impulsionnel,..., additif/multiplicatif,...) ainsi que différentes origines (capteur défectueux, sources parasites...). Le but de la restauration consiste à supprimer cette information parasite ainsi qu'à corriger l'effet de la réponse impulsionnelle du système, en évitant de dégrader l'information pertinente.

Nous formulons l'hypothèse suivante: une image est formée de zones homogènes séparées par des ensembles de contours francs.

Il existe plusieurs approches pour modéliser la présence d'un contour, la plus simple étant de détecter un contour par la présence d'un contraste élevé de l'intensité. C'est le modèle de contour que nous utilisons ici. Il existe des modèles de contours plus élaborés dont un qui sera abordé plus en détails dans le paragraphe suivant. Afin de modéliser le processus de dégradation de l'image source par du bruit, nous noterons par:

$$\begin{cases} f : \Omega \subset \mathbb{R}^n \to \mathbb{R} & \text{l'image source,} \\ \nu : \Omega \subset \mathbb{R}^n \to \mathbb{R} & \text{le bruit,} \\ I : \Omega \subset \mathbb{R}^n \to \mathbb{R} & \text{les données bruitées observées.} \end{cases}$$

Le modèle de dégradation considéré s'écrit

$$I(x) = \mathcal{R}f(x) + \nu(x), \forall x \in \Omega. \tag{1}$$

Par la suite, nous ne prendrons pas en compte la réponse impulsionnelle du système, i.e. $\mathcal{R} = Id$.

Dans [67], nous avons proposé le modèle variationnel de classification suivant:

$$J_\varepsilon(f) = \underbrace{\int_\Omega \Big(f(x) - I(x)\Big)^2 dx}_{\text{attache aux données}} + \varepsilon\lambda^2 \underbrace{\int_\Omega \varphi(|\nabla f(x)|)dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_\Omega W(f(x))dx}_{\text{contrainte de niveaux}} , \tag{2}$$

et le problème associé consiste à trouver $\tilde{f}$ vérifiant:

$$\tilde{f} = \lim_{\varepsilon \to 0^+} \Big[\arg\min_f J_\varepsilon(f)\Big]. \tag{3}$$

Les réels λ et η sont fixés. La fonction φ est une fonction régularisante parmi celles présentées dans le tableau 1. Lorsque le réel ε est fixé, les deux premiers termes de J_ε dans (2) induisent une restauration de l'image permettant de préserver les contours si la fonction régularisante φ est sous-quadratique en l'infini. Ce type de fonctionnelle a donné lieu à de nombreux travaux et le lien avec la fonctionnelle proposée par Mumford et Shah [55] a également été étudié [2, 3, 4, 15, 19, 50, 51, 54, 73].

Le potentiel parabolique par morceaux W ($W \in C^1(\mathbb{R})$) est défini par

$$W(y) = \begin{cases} \frac{w_1}{{\sigma_1}^2}(y - \mu_1)^2 & \text{si } y \leq \alpha_1 \\ \frac{w_K}{{\sigma_K}^2}(y - \mu_K)^2 & \text{si } y \geq \beta_{K-1} \\ \frac{w_i}{{\sigma_i}^2}(y - \mu_i)^2 & \text{si } \beta_{i-1} \leq y \leq \alpha_i \quad i = 1,...,(K-1) \\ k_i - c_i(y - p_i)^2 & \text{si } \alpha_i \leq y \leq \beta_i \quad i = 1,...,(K-1) \end{cases} \tag{4}$$

	$\varphi(t)$	$\frac{\varphi'(t)}{2t}$	convexité
Tikhonov [74]	t^2	1	oui
Total Variation [63]	$\lvert t\rvert$	$\frac{1}{2\lvert t\rvert}$ (si $t \neq 0$)	oui
Geman & McClure [30]	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	non
Hebert & Leahy [35]	$\log(1+t^2)$	$\frac{1}{1+t^2}$	non
Green [34]	$\log(\cosh(t))$	$\frac{\tanh(t)}{2t}$ (si $t \neq 0$)	oui
Hyper Surfaces [19]	$2\sqrt{1+t^2}-2$	$\frac{1}{\sqrt{1+t^2}}$	oui

TAB. 1 – *Quelques fonctions φ proposées dans la littérature.*

Les réels w_i sont fixés. Cette contrainte de niveaux est inspiré des travaux menés dans le cadre de la théorie de Van der Waals-Cahn-Hilliard sur les transitions de phases [5, 8, 25, 53, 70, 71]. Le potentiel W restreint l'espace des valeurs prises par $f_\varepsilon(x)$, minimum de $J_\varepsilon(f)$, à celui des K étiquettes des classes en tenant compte du caractère gaussien de la distribution d'intensité sur chaque classe. D'après les conventions choisies, W attire les valeurs de $f_\varepsilon(x)$ vers celles des étiquettes $\{\mu_1,...,\mu_K\} = \Lambda$.
La convergence sur ε induit une évolution du poids relatif des deux termes. L'influence de la contrainte de niveaux évolue inversement à celle du terme de régularisation. La valeur initiale de ε est suffisamment importante[1] pour rendre le terme de contrainte de niveaux négligeable par rapport à celui de régularisation. Au cours de la décroissance de ε, l'influence du potentiel W augmente progressivement en même temps que diminue celle de la régularisation. On peut dire grossièrement qu'on assiste à un changement progressif du comportement du modèle au cours de la convergence sur ε.
Initialement, le modèle est de type restauration plutôt que de type classification, et lors de la décroissance de ε, la tendance s'inverse progressivement. Schématiquement, l'étiquetage des pixels de l'image n'a lieu qu'une fois l'image suffisamment restaurée.
Selon le niveau de bruit de l'image, il s'avèrera utile de laisser plus de temps à la phase de restauration et donc d'opter pour une décroissance lente du paramètre ε.
Les travaux menés dans le cadre de la théorie de Van der Waals-Cahn-Hilliard ont donné lieu à des résultats concernant l'expression de la Γ-limite de la suite de fonctionnelles J_ε, et concernant les propriété de régularité de la suite f_ε des solutions

1. "importante" dans le sens "non négligeable"

de $inf J_\varepsilon$. En résumé, les auteurs de [5, 8, 25, 53, 70, 71] ont montré que, pour $\varphi(t) = t^2$, la solution de $\lim_{\varepsilon\to 0^+}\left[\arg\min_f J_\varepsilon(f)\right]$ est une partition des données I constituée de régions homogènes séparées par des interfaces de longueurs minimales. De plus, le niveau de chaque région est une des étiquettes de Λ. Nous renvoyons le lecteur à [64, 67] pour plus de détails concernant ce modèle de classification appliqué à l'imagerie.

2.2 Classification par modèle dynamique

Cette approche est différente de la précédente dans le sens où le problème n'est plus posé sous la forme d'un problème de restauration sous contrainte de classification, mais sous la forme d'un problème de partitionnement des données observées en fonction des classes prédéfinies. L'originalité de cette approche repose sur le fait que la partition optimale cherchée est définie à l'aide d'ensembles de niveaux. Nous renvoyons le lecteur aux travaux [16, 17, 40, 56] pour de plus amples informations concernant les contours actifs et la formulation par ensemble de niveaux. Un certain nombre de travaux ont été effectués dans le cadre général de la segmentation, mettant en œuvre des modèles actifs fondés sur des approches de type région. Parmi ces travaux, nous trouvons également des modèles actifs combinant l'information au niveau des contours aux statistiques des régions délimitées par ces contours [18, 57, 81, 85]. Le modèle que nous avons proposé dans [64, 65, 66] est inspiré des travaux effectués par Zhao *et al.* [84].

Nous cherchons une partition du support de l'image Ω:

$$\Omega = \bigcup_{i=1}^{K}\left(\Omega_i \bigcup \Gamma_i\right) \quad \text{et} \quad \Omega_i \bigcap_{i\neq j} \Omega_j = \varnothing. \tag{5}$$

On appelle $\Gamma_i = \partial\Omega_i \cap \Omega$ l'intersection de l'ensemble des frontières de Ω_i avec l'ouvert Ω. L'interface entre Ω_i et Ω_j est notée par:

$$\Gamma_{ij} = \Gamma_{ji} = \Gamma_i \cap \Gamma_j, \quad \forall i \neq j. \tag{6}$$

De plus, nous avons:

$$\Gamma_i = \bigcup_{j\neq i} \Gamma_{ij}. \tag{7}$$

CONDITION A: $\{\Omega_i\}_{i=1...K}$ est une partition de Ω:

$$\Omega = \bigcup_i \Big(\Omega_i \bigcup \Gamma_i\Big) \quad \text{avec} \quad \Omega_i \bigcap_{i \neq j} \Omega_j = \varnothing.$$

CONDITION B: La partition $\{\Omega_i\}_{i=1...K}$ doit prendre en compte les caractéristiques de chaque classe, à savoir la moyenne et la variance des classes compte tenu des hypothèses formulées sur la distribution (gaussienne) de l'intensité pour chaque classe. Cette condition est traduite à travers la minimisation du terme suivant:

$$\text{minimiser} \sum_i \int_{\Omega_i} \left(\frac{I(x) - \mu_i}{\sigma_i}\right)^2 dx \quad \text{par rapport aux variables } \Omega_i.$$

CONDITION C: La partition obtenue doit être régulière dans le sens où la longueur totale des interfaces Γ_{ij} est minimale:

$$\text{minimiser} \sum_{i,j} \xi_{ij} |\Gamma_{ij}| \quad \text{par rapport aux variables } \Gamma_{ij} \quad (\xi_{ij} \in \mathbb{R}).$$

TAB. 2 – *Propriétés souhaitées sur la partition, solution du modèle de classification*

On note $|\Gamma_i| = \mathcal{H}^1(\Gamma_i)$ la mesure unidimensionnelle de Hausdorff de l'ensemble Γ_i vérifiant:

$$|\Gamma_i| = \sum_{j \neq i} |\Gamma_{ij}| \text{ avec } |\varnothing| = 0. \tag{8}$$

La partition que nous cherchons doit vérifier les trois conditions exposées dans le tableau 2.

Soit $\Phi_i : \Omega \to \mathbb{R}$ une fonction lipschitzienne continue associée à chaque région Γ_i telle que:

$$\begin{cases} \Phi_i(x) > 0 & \text{si } x \in \Omega_i \\ \Phi_i(x) = 0 & \text{si } x \in \Gamma_i \\ \Phi_i(x) < 0 & \text{sinon .} \end{cases} \tag{12}$$

L'ensemble des régions Ω_i ainsi que l'ensemble des frontières associées Γ_i sont entièrement décrites par la fonction Φ_i.

Remarque: Par la suite, nous omettrons la variable $x \in \Omega$ dans l'expression $\Phi_i(x)$

lorsque cela ne prête pas à confusion.

Nous définissons les approximations des distributions de Dirac δ_α et de Heaviside H_α, avec $\alpha \in \mathbb{R}^+$, par :

$$\delta_\alpha(s) = \begin{cases} \frac{1}{2\alpha}\Big(1+\cos(\frac{\pi s}{\alpha})\Big) & \text{si } |s| \leq \alpha \\ 0 & \text{si } |s| > \alpha \end{cases} \tag{13}$$

$$H_\alpha(s) = \begin{cases} \frac{1}{2}\Big(1+\frac{s}{\alpha}+\frac{1}{\pi}\sin(\frac{\pi s}{\alpha})\Big) & \text{si } |s| \leq \alpha \\ 1 & \text{si } s > \alpha \\ 0 & \text{si } s < -\alpha \end{cases} \tag{14}$$

et nous avons:

$$\begin{cases} \delta_\alpha \overset{\mathcal{D}'(\Omega)}{\longrightarrow} \delta & \text{lorsque } \alpha \to 0^+ \\ H_\alpha \overset{\mathcal{D}'(\Omega)}{\longrightarrow} H & \text{lorsque } \alpha \to 0^+ \end{cases}$$

où $\mathcal{D}'(\Omega)$ est l'espace des distributions défini sur Ω. D'après (12), (13) et (14) on obtient:

$$\{x \in \Omega / \lim_{\alpha \to 0^+} H_\alpha\Big[\Phi_i(x)\Big] = 1\} = \Omega_i \tag{15}$$

$$\{x \in \Omega / \lim_{\alpha \to 0^+} \delta_\alpha\Big[\Phi_i(x)\Big] \neq 0\} = \Gamma_i. \tag{16}$$

Les K ensembles Ω_i et Γ_i sont ainsi modélisés à l'aide des fonctions d'ensemble de niveaux Φ_i. Nous allons ainsi pouvoir traduire les contraintes sur la solution de partition (exprimées dans le tableau 2) à travers une fonctionnelle définie sur l'ensemble des fonctions Φ_i. La fonctionnelle que nous proposons de minimiser s'écrit:

$$\begin{aligned} F_\alpha(\Phi_1,...,\Phi_K) = & \underbrace{\sum_{i=1}^{K} e_i \int_\Omega H_\alpha(\Phi_i)\frac{(I(x)-\mu_i)^2}{\sigma_i^2}dx}_{\text{attache aux données}} \\ & + \underbrace{\sum_{i=1}^{K} \gamma_i \int_\Omega g(|\nabla I|)\delta_\alpha(\Phi_i)|\nabla\Phi_i|dx}_{\text{minimisation de la longueur des contours}} \\ & + \underbrace{\frac{\lambda}{2}\int_\Omega\Big(\sum_{i=1}^{K} H_\alpha(\Phi_i)-1\Big)^2 dx}_{\text{contrainte de partition}}. \end{aligned} \tag{17}$$

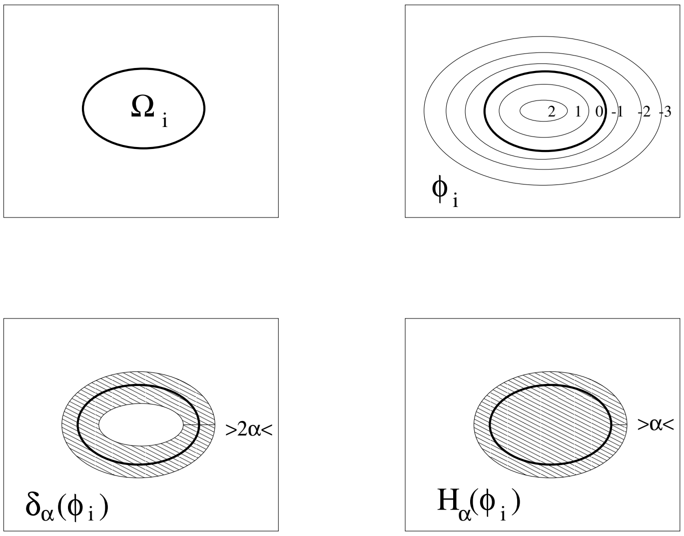


FIG. 1 – *A chaque ensemble de régions Ω_i est associée une fonction de distance signée Φ_i. Les régions hachurées représentent les zones non nulles lorsque l'on applique une approximation de la distribution de Dirac δ_α ou de Heaviside H_α sur la fonction Φ_i.*

où $g : \mathbb{R} \rightarrow \mathbb{R}^+$ est définie par $g(|\nabla I|) = \frac{1}{1+|\nabla I(x)|^2}$.

Les trois termes de la fonctionnelle traduisent les trois conditions exprimées dans le tableau 2. Les paramètres e_i,γ_i et λ sont des réels fixés. La minimisation de cette fonctionnelle par rapport à chaque fonction Φ_i conduit à une évolution des régions et de leurs contours guidée par des forces permettant de tenir compte des caractéristiques gaussiennes des classes et de la minimisation de la longueur des interfaces. La contrainte de partition impose une et une seule étiquette par pixel (pénalisation des régions vides et du chevauchement de régions). Nous présentons de nombreuses applications dans [64, 66].

3 Classification de données multispectrales

Nous introduisons dans un premier temps les particularités du cas multispectral, en insistant sur les points concernés par l'extension des deux modèles présentés au sein des paragraphes précédents au cas de données multiples. L'extension au cas multispectral intervient à des niveaux différents pour ces deux modèles. Nous commençons par exposer les différentes notations utilisées dans le cas multispectral, ainsi que par une présentation du modèle d'estimation du nombre et des paramètres des classes que nous utilisons au préalable, avant d'aborder l'extension de chacun des deux modèles au cas multispectral.

3.1 Le cas multispectral

Ce paragraphe concerne la classification d'images multispectrales à l'aide des modèles proposés dans les paragraphes précédents. Il s'agit de combiner l'information retournée dans les différentes bandes spectrales par les capteurs. Nous ne traiterons pas de façon approfondie de la redondance d'information entre les bandes (voir par exemple [76, 77, 79]) pour laquelle il s'agit de prendre en compte la physique des capteurs, même si nous savons par exemple qu'il existe une forte redondance d'information entre les bandes pour les images multispectrales fournies par les satellites SPOT. La corrélation entre les bandes est prise en compte par l'intermédiaire de matrices de covariances.

Nous supposerons que les données observées sont représentées par une fonction constituée de P composantes: les P signaux renvoyés par les différents capteurs des satellites. Nous utiliserons les notations suivantes:

- P est le nombre de bandes considéré.
- Les observations sont représentées par la fonction $\vec{I} : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^P$, soit $\vec{I} = [I^1,...,I^P]^T$. Nous supposons que $\vec{I} \in L^2(\Omega,\mathbb{R}^P)$.
- Nous supposons que le nombre de classes K est le même dans chaque bande, chaque bande représentant la même scène dans différents domaines spectraux.
- Les K classes sont caractérisées par les paramètres gaussiens suivants:

$$\vec{\mu_i} = \{\mu_i^p\}_{p=1...P} \quad \text{et} \quad \vec{\sigma_i} = \{\sigma_i^p\}_{p=1...P} \quad \text{pour} \quad i = 1,...,K.$$

- Pour $i = 1,...,K$, nous noterons Σ_i la matrice de variance-covariance définie sur $\mathbb{R}^P \times \mathbb{R}^P$ par:

$$\Sigma_i = \begin{bmatrix} \sigma_i^{11} & \sigma_i^{12} & \cdots & \sigma_i^{1P} \\ \sigma_i^{21} & \sigma_i^{22} & \cdots & \sigma_i^{2P} \\ \vdots & & \ddots & \vdots \\ \sigma_i^{P1} & \sigma_i^{P2} & \cdots & \sigma_i^{PP} \end{bmatrix}.$$

  où $\sigma_i^{mn} = cov(I^m(x),I^n(x))$ et avec $[\sigma_i^p]^2 = \sigma_i^{pp}$ la variance associée à la classe C_i pour la bande p.

Bien que nous n'étudions pas en détails la redondance d'information entre les P bandes, les covariances sont, en quelques sortes, une mesure de la corrélation existant entre les différentes bandes pour chacune des K classes. Une valeur élevée de σ_i^{mn} signifie que l'information sur la classe C_i dans la bande m est fortement corrélée à celle dans la bande n. Une matrice Σ_i diagonale signifie que les bandes sont complètement décorrélées concernant la classe C_i.

Etant données les P bandes spectrales, il s'agit de trouver une partition des données, composée de K régions homogènes séparées par des interfaces de longueurs minimales. Nous attendons ainsi les mêmes propriétés sur les solutions de classification fournies par les deux modèles présentés dans les paragraphes précédents, que ce soit dans les cas monospectral ou multispectral.

3.2 Estimation du nombre de classes et de leurs paramètres

L'étape d'estimation du nombre de classes et de leurs paramètres peut être effectuée au préalable à l'aide d'un algorithme de *Fuzzy Cmeans* [10, 43, 62] dont le critère a été modifié par A. Lorette dans [46, 47] par l'ajout d'un terme d'entropie. Cette méthode entre dans le cadre de la partition d'ensembles flous tels qu'ils ont été définis dans [82]. Dans ce contexte, il s'agit de découper l'image en un ensemble de Cl nuages de points (les ensembles *flous*, ou encore les ensembles de régions associés aux Cl classes) et d'attribuer à chaque site $s \in S$ de l'image, Cl nombres compris entre 0 et 1 représentant le degré d'appartenance à chacun des nuages. La partition de l'image est alors caractérisée par une matrice $A = [a_s^i]$ où $a_s^i \in [0;1]$ représente le degré d'appartenance du site s au nuage $i \in [1,2,...,Cl]$.
Un terme d'entropie est utilisé pour estimer le nombre de classes (voir [46, 47] pour plus de détails). Le critère obtenu s'écrit:

$$J = \sum_{i=1}^{Cl} \sum_{s \in S} (a_s^i)^m d^2(I_s, c_i) - \alpha \sum_{i=1}^{Cl} p_i log(p_i), \tag{18}$$

sous la contrainte

$$\sum_{i=1}^{Cl} a_s^i = 1 \ \forall s \in S.$$

Le paramètre $m \in [0;\infty]$ caractérise le degré de flou. En pratique, il est fréquent de fixer $m = 2$ pour bon nombre d'applications (pour $m = 0$, on retrouve la méthode classique du *K means* [48], plus un terme d'entropie). Les valeurs $c_i = (\vec{\mu_i},\vec{\sigma_i})$ représentent les vecteurs caractéristiques des centres de gravité des nuages et $d^2(I_s,c_i)$ est la distance euclidienne entre l'observation I_s et c_i. La probabilité a priori p_i de la i$^{\text{ème}}$ classe est définie par

$$p_i = \frac{1}{N}\sum_{s\in S} a_s^i, \quad \text{avec} \quad N = card(S).$$

Le premier terme de J caractérise l'homogénéité des nuages; il est minimum lorsqu' un nuage ne caractérise qu'un seul niveau de gris. Le second terme, entropique, est lié au nombre de nuages, il est minimum lorsque tous les pixels sont dans le même nuage et que tous les autres sont vides.
En injectant la contrainte $\sum_{i=1}^{Cl} a_s^i = 1 \ \forall s \in S$ sous forme de pénalité dans le critère, nous sommes conduits à la minimisation par rapport aux c_i et par rapport aux a_s^i du critère:

$$E = \sum_{i=1}^{Cl}\sum_{s\in S}(a_s^i)^m d^2(I_s,c_i) - \alpha\sum_{i=1}^{Cl} p_i log(p_i) - \sum_{s\in S}\lambda_s[\sum_{i=1}^{Cl} a_s^i - 1]. \tag{19}$$

Les paramètres λ_s représentent les multiplicateurs de Lagrange dans le cadre de l'optimisation sous contrainte.
Nous alternons les minimisations de E par rapport aux c_i (avec les a_s^i fixés) et par rapport aux a_s^i (pour les c_i fixés), en annulant les dérivées partielles associées.
Il est préférable de commencer l'algorithme avec une grande valeur de α afin de déterminer dès les premières itérations le nombre optimal de nuages de points (i.e. le nombre de classes K que nous allons considérer dans les modèles multispectraux détaillés dans les paragraphes suivants), puis de tenir compte peu à peu du terme de classification. Pour ce faire, α décroît de manière exponentielle [26] à chaque itération de la manière suivante : $\alpha(n) = \alpha(0)\exp\left(-\frac{n}{\tau}\right)$, où n est le nombre d'itérations et τ est une constante de temps.
On évite ainsi que le terme de classification ne soit biaisé par le terme d'entropie puisque $\alpha(n)$ tend vers 0 quand n tend vers ∞.
En outre, $\alpha(0)$ dépend fortement du nombre de pixels dans l'image. Afin de s'affranchir des données, nous posons $\alpha(0) = \alpha_0 N$ (où N est le nombre de pixels de l'image). Dans ce cas, le seul paramètre que nous avons à fixer est α_0. Ce paramètre détermine la finesse de la classification que l'on souhaite obtenir. Pour une application donnée, nous fixons une fois pour toute α_0 et τ alors que le nombre de nuages de points initial Cl peut être différent suivant la scène traitée.

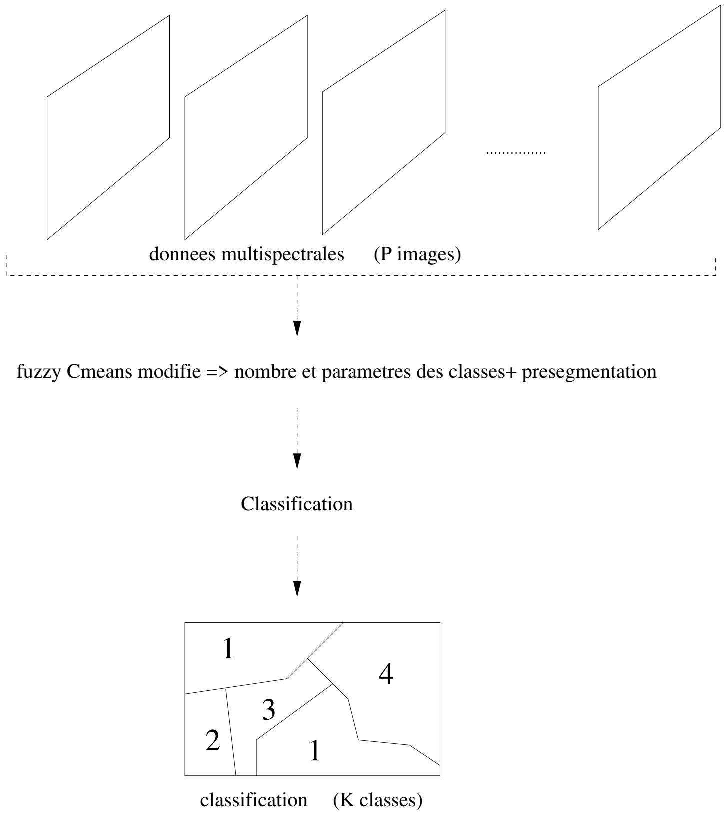


FIG. 2 – *Principe de la classification automatique dans le cas multispectral.*

Après avoir mis à jour les centres de gravité et les degrés d'appartenance, nous estimons les probabilités a priori de chacun des nuages de points. Nous éliminons les nuages de points dont la probabilité est inférieure à un certain seuil ϵ. L'algorithme de *Fuzzy Cmeans* modifié est présenté dans le tableau 3.

L'algorithme de *Fuzzy Cmeans* modifié (issu des travaux de [46]) permet donc de trouver le nombre de classes K, ainsi qu'une première classification sur laquelle nous pouvons estimer les paramètres gaussiens de chaque classe. Le nombre de classes, ainsi que leur paramètres sont ensuite donnés en entrée des modèles multispectraux que nous proposons pour obtenir une classification finale, comme l'illustre la figure 2. En pratique, nous utilisons l'algorithme de *Fuzzy Cmeans* modifié pour l'estimation

```
0 - Initialiser les centres de gravité des nuages de points, n = 0.
1 - Calculer la matrice A^(n)
2 - Calculer les centres de gravité des nuages de points
3 - Calculer les probabilités a priori p_i de chaque nuage de points:
    si p_i < ϵ le nuage de points i est supprimé
4 - Si ||A^(n) - A^(n-1)||>ξ alors:
    n = n + 1
    mettre à jour α
    retourner à l'étape 1
  - Sinon aller à l'étape 5
5 - Assigner à chaque pixel l'étiquette du nuage de points auquel
   il a la plus grande probabilité d'appartenir.
```

TAB. 3 – *Algorithme de Fuzzy Cmeans modifié.*

des paramètres K, $\vec{\mu_i}$ et $\vec{\sigma_i}$, sans généralement utiliser le résultat de classification fourni par cet algorithme comme initialisation de nos modèles de classification multispectrale car nous initialisons le premier modèle proposé par les données observées et le second par une initialisation automatique dont le principe est présenté sur la figure 6.

3.3 Extension du premier modèle

3.3.1 Introduction

Nous proposons une extension du modèle de restauration avec contrainte de classification présenté au paragraphe 2.1 pour le cas de données $\vec{I} : \mathbb{R}^2 \to \mathbb{R}^P$.

Dans le cadre des travaux portant sur la théorie de Van der Waals-Cahn-Hilliard, nous rappelons le problème P_ε défini dans [5, 8, 25, 53, 70, 71]. Considérons un système physique constitué d'un fluide dont l'*énergie libre de Gibbs* par unité de volume est une fonction notée W dépendant de la distribution de densité u du fluide et possédant K minima. Dans le cas où $u : \mathbb{R}^2 \to \mathbb{R}$ et pour $W : \mathbb{R} \to \mathbb{R}^+$:

$$P_\varepsilon \begin{cases} \inf_{u \in H^1(\Omega;\mathbb{R})} E_\varepsilon(u), \\ E_\varepsilon(u) = \int_\Omega \Big[\varepsilon|\nabla u(x)|^2 + \frac{1}{\varepsilon} W(u(x))\Big] dx, \\ \text{avec la contrainte suivante sur } u: \quad \int_\Omega u(x)dx = m, \end{cases} \tag{20}$$

L'étude théorique de la Γ-convergence de E_ε lorsque $\varepsilon \to 0^+$ et de la caractérisation des suites de solutions u_ε a été étendue au cas où $u : \mathbb{R}^2 \to \mathbb{R}^P$ et pour $W : \mathbb{R}^P \to \mathbb{R}^+$ par [5].
Dans le cas vectoriel, les conditions imposées sur le potentiel W sont de même nature que celles nécessaires dans le cas scalaire, à savoir:

- W s'annule sur les K minima $\vec{\mu_i}$ de $\mathbb{R}^P$,
- W est localement quadratique autour de chaque minimum,
- W croît au moins linéairement en l'infini.

Les résultats montrés par S. Baldo [5] dans le cas vectoriel sont du même type que dans le cas scalaire, en particulier la solution limite $\lim_{\varepsilon \to 0^+} u_\varepsilon$ est géométriquement caractérisée par une partition en régions homogènes de niveaux μ_i^p pour la bande p, séparées par des interfaces de longueurs minimales.

Dans le cas vectoriel et pour K minima, l'expression de E_0, la Γ-limite de E_ε définie dans (20) est exprimée dans [5] par:

$$E_0(u) = \sum_{i,j=1}^{K} d(\vec{\mu_i},\vec{\mu_j})\mathcal{H}^1(\Gamma_{ij}) \text{ si } E_0(u) < +\infty,\, u \in BV(\Omega),\, W(u(x)) = 0\, p.p. \quad (21)$$

avec Γ_{ij} l'interface entre les régions Ω_i de la classe i et Ω_j. La distance d est définie par:

$$\begin{aligned} d(\vec{\mu_i},\vec{\mu_j}) = inf_g \Big\{ \int_0^1 \sqrt{W(g(s))}|g^{'}(s)|ds; \\ g \in C^1([0;1],\mathbb{R}^P) \text{ et } g > 0,\, g(0) = \vec{\mu_i},\, g(1) = \vec{\mu_j} \Big\} \end{aligned} \quad (22)$$

Ainsi, le problème limite de minimisation de E_ε est un problème de minimisation de la longueur pondérée de l'ensemble des interfaces entre les régions associées à chaque classe.
Si u_ε, solution de P_ε, converge dans $L^1(\Omega;\mathbb{R}^P)$ vers u_0 lorsque $\varepsilon \to 0^+$, alors:

$$u_0(x) = \sum_{i=1}^{K} \vec{\mu_i} \chi_{\Omega_i}(x) \quad (23)$$

avec χ_{Ω_i} la fonction caractéristique de l'ensemble Ω_i.
Par conséquent, d'après (23), la suite de solutions u_ε de la famille de problèmes P_ε définie par (23) est une partition régulière de Ω qui minimise la quantité $\sum_{i,j=1}^{K} d(\vec{\mu_i},\vec{\mu_j})\mathcal{H}^1(\Gamma_{ij})$.

D'où les propriétés de régularité sur la solution du problème limite annoncées plus haut et démontrées par S. Baldo.

Nous allons faire un parallèle entre la théorie de Van der Waals-Cahn-Hilliard et le traitement d'images moyennant quelques extrapolations dues aux spécificités des images et du traitement de données réelles (comme l'utilisation d'un terme de perturbation sous-quadratique à la place de $|\nabla u(x)|^2$).

3.3.2 Modèle multispectral

Les principales difficultés liées à l'extension du modèle (2) au cas multispectral proviennent de l'expression du module du gradient de f dans $\mathbb{R}^P$, ainsi que de la construction du potentiel W sur $\mathbb{R}^P$. La famille de fonctionnelles que nous proposons de minimiser par rapport à $\vec{f}:\mathbb{R}^2 \rightarrow \mathbb{R}^P$ s'écrit:

$$J_\varepsilon(\vec{f}) = \underbrace{\int_\Omega \left|\vec{f}(x) - \vec{I}(x)\right|^2_{\mathbb{R}^P} dx}_{\text{attache aux données}} + \varepsilon\lambda^2 \underbrace{\int_\Omega \varphi(|\vec{\nabla} f(x)|_{\mathbb{R}^P}) dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_\Omega W(\vec{f}(x)) dx}_{\text{contrainte de niveaux}} \quad , \tag{24}$$

et le problème associé consiste à trouver $\tilde{f}$ vérifiant:

$$\tilde{f} = \lim_{\varepsilon \rightarrow 0^+} \left[\arg\min_{\vec{f}} J_\varepsilon(\vec{f})\right]. \tag{25}$$

avec $|.|_{\mathbb{R}^P}$ désignant une norme dans $\mathbb{R}^P$.
Le potentiel W prenant comme argument le vecteur $\vec{f}$, s'annule aux K points $\vec{\mu_i}$, minima du potentiel, et W possède une forme gaussienne (dans $\mathbb{R}^P$) autour de chaque minima, comme l'illustre la figure 3.

Comme pour le cas monospectral, nous utilisons une fonction de régularisation φ, non nécessairement quadratique, ce qui constitue un écart par rapport au cadre théorique défini dans [5].

L'information des différentes bandes est combinée au niveau du gradient appliqué à $\vec{f}$, ainsi qu'à travers le potentiel de classification W.

Calcul des normes dans $\mathbb{R}^P$

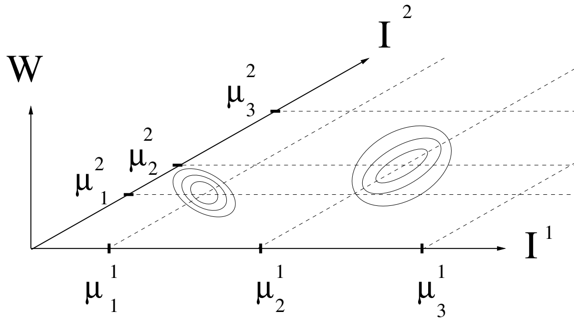


FIG. 3 – *Projection du potentiel vectoriel W sur le plan (I^1,I^2) dans le cas de 2 bandes et pour 3 classes.*

Les normes de $\mathbb{R}^P$ définies dans (24) ne sont pas de même nature, qu'il s'agisse du terme d'attache aux données ou de l'argument de la fonction régularisante:

- $\left|\vec{f}(x) - \vec{I}(x)\right|^2_{\mathbb{R}^P}$: La norme portant sur le terme d'attache aux données est la norme $L^2(\mathbb{R}^P)$.

- $|\nabla \vec{f}(x)|_{\mathbb{R}^P}$: nous supposons pour le moment que nous utilisons la norme $L^2(\mathbb{R}^P)$.

Extension du potentiel W

Le potentiel W, reliant la solution $\vec{f}$ aux classes et à leurs caractéristiques, doit tenir compte de la multi-dimensionalité des moyennes et variances dans le cas multispectral.
Nous allons prendre en compte les covariances des classes en plus des moyennes et variances caractéristiques, afin de combiner l'information sur les paramètres des classes à travers les différentes bandes. Ainsi, après avoir pré-estimé le nombre de classes K par un algorithme de *fuzzy Cmeans* modifié par exemple, nous calculons leurs moyennes $\vec{\mu_i} \in \mathbb{R}^P$ et écart-types $\vec{\sigma_i} \in \mathbb{R}^P$ pour $i = 1,...,K$, ainsi que les covariances σ_i^{mn} pour m et $n = 1,...,P$.

Comme nous l'avons précisé précédemment, W est localement quadratique autour de chaque minimum $\vec{\mu_i}$, et s'annule en ces points. Nous définissons alors l'expression locale de W par:

$$W_{loc}(\vec{f}) = [\vec{f}(x) - \vec{\mu_i}]^T \Sigma_i^{-1} [\vec{f}(x) - \vec{\mu_i}] \tag{26}$$

avec Σ_i^{-1} désignant l'inverse de la matrice de variance-covariance. Les covariances permettent de prendre en compte l'information sur une classe à travers les différentes bandes. Nous combinons ainsi l'information des différentes composantes de $\vec{f}$ à travers les covariances.

Pour construire le potentiel W en fonction de chaque valeur de $\vec{f}(x)$, il s'agit de déterminer le plus *proche* minimum $\vec{\mu_i}$ de x afin de trouver l'expression locale (26). Pour cela, à chaque valeur de $\vec{f}(x)$, nous calculons l'indice i tel que:

$$i = \min_{k=1,\dots,K} [\vec{f}(x) - \vec{\mu_k}]^T \Sigma_k^{-1} [\vec{f}(x) - \vec{\mu_k}]. \tag{27}$$

La figure 4 illustre graphiquement, dans le cas d'une bande et pour 3 classes, la méthode utilisée d'après (27) permettant de sélectionner le plus *proche* minimum $\vec{\mu_i}$ pour chaque valeur de $\vec{f}(x)$, et d'en déduire l'expression locale de W (voir 26). Sur la figure 4, nous trouvons $i = 1$, par conséquent, en ce point x, nous avons:

$$W_{loc}(\vec{f}) = [\vec{f}(x) - \vec{\mu_1}]^T \Sigma_1^{-1} [\vec{f}(x) - \vec{\mu_1}].$$

Nous remarquons que W n'est pas dérivable aux points de jonctions des paraboles multidimensionnelles. Il faudrait pour cela effectuer, par exemple, des jonctions $C^1(\mathbb{R}^P)$ afin d'éviter des éventuelles instabilités numériques pour des valeurs de $\vec{f}(x)$ se trouvant au niveau des points de jonction, ce que nous n'avons pas rencontré pour le moment au cours des applications traitées. La construction d'un potentiel W qui soit élément de $C^1(\mathbb{R}^P)$ fait partie des travaux à effectuer dans un proche avenir.

Fonctionnelle proposée

Compte tenu des calculs de normes multidimensionnelles adoptés et des expressions de W en fonction des valeurs de $\vec{f}(x)$, nous proposons de minimiser la fonctionnelle suivante par rapport à $\vec{f} \in \mathbb{R}^P$:

$$J_\varepsilon(\vec{f}) = \int_\Omega \left| \vec{f}(x) - \vec{I}(x) \right|^2_{\mathbb{R}^P} dx + \varepsilon\lambda^2 \int_\Omega \varphi(|\vec{\nabla} f|_{\mathbb{R}^P}) dx + \frac{\eta^2}{\varepsilon} \int_\Omega W(\vec{f}(x)) dx, \tag{28}$$

avec W le potentiel de contrainte de niveaux dont l'expression locale est déterminée par (26).

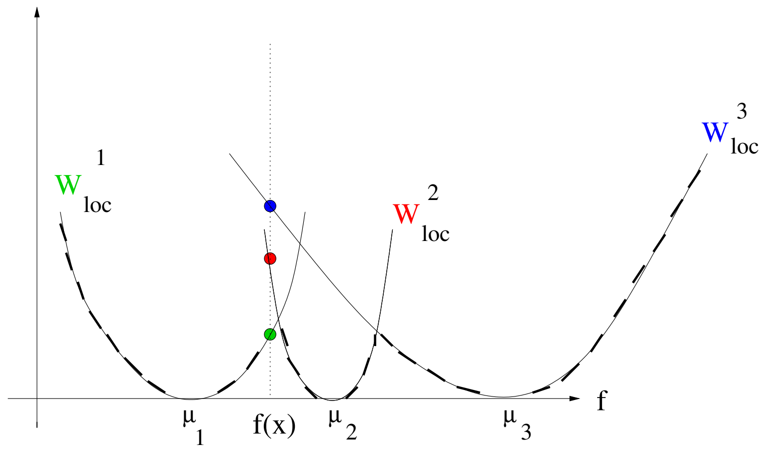


FIG. 4 – *Sélection de $i = \min_{k=1,\ldots,K}[\vec{f}(x) - \vec{\mu_k}]^T \Sigma_k^{-1} [\vec{f}(x) - \vec{\mu_k}]$ dans le cas où $P = 1$ et $K = 3$. Ici nous en obtenons $i = 1$ d'après (27). Le potentiel W final est la courbe représentée en pointillés.*

Cette fonctionnelle combine l'information des différentes bandes au niveau du terme régularisant pondéré par ε ainsi qu'à travers le potentiel W. Nous utiliserons quelques notions de géométrie différentielle pour le calcul de $|\vec{\nabla f}|_{\mathbb{R}^P}$ uniquement au niveau des EDP d'Euler-Lagrange.

3.3.3 Minimisation et algorithme

Nous utilisons la méthode de *régularisation semi-quadratique* [19, 29] permettant de surmonter la difficulté liée à la non linéarité des P EDP résultant des équations d'Euler-Lagrange pour la minimisation de (28). Les conditions que doit respecter la fonction φ pour pouvoir utiliser cette méthode sont les mêmes que celles présentées dans [19, 67].
En introduisant la variable auxiliaire $b : \mathbb{R}^2 \to \mathbb{R}$, nous ramenons la minimisation de (28) à celle de $J_\varepsilon{}^*(\vec{f},b)$ par rapport au couple $(\vec{f},b)$, avec:

$$J_\varepsilon{}^*(\vec{f},b) = \int_\Omega \left| \vec{f}(x) - \vec{I}(x) \right|^2_{\mathbb{R}^P} dx + \varepsilon\lambda^2 \int_\Omega \left[b(x) |\vec{\nabla} f(x)|^2_{\mathbb{R}^P} + \psi(b(x)) \right] dx + \frac{\eta^2}{\varepsilon} \int_\Omega W(\vec{f}(x)) dx, \tag{29}$$

avec ψ une fonction convexe en b dont l'expression analytique est déduite de celle de φ.
Par le biais des équations d'Euler-Lagrange, nous sommes alors conduits à résoudre alternativement les équations suivantes pour chaque valeur de ε:

$$\begin{cases} b = \frac{\varphi'(|\vec{\nabla} f|_{\mathbb{R}^P})}{2|\vec{\nabla} f|_{\mathbb{R}^P}} \in \mathbb{R} \text{ avec } \vec{f} \text{ fixé,} \\ \text{pour p=1,...,P:} \quad f^p - \lambda^2 \varepsilon div(b \nabla f^p) + \frac{\eta^2}{2\varepsilon} W'_p(\vec{f}) = I^p, \text{ avec } b \text{ fixé.} \end{cases} \tag{30}$$

L'indice $p \in [1; P]$ désigne la $p^{\text{ième}}$ composante des vecteurs concernés dans les P EDP du schéma (30), et nous notons:

$$W'(\vec{f}) = [w'_1, ..., w'_P]^T \quad \text{et} \quad W'_p(\vec{f}) = w'_p.$$

Pour chaque valeur de ε, nous trouvons une solution $(\vec{f}, b)$ qui nous sert d'initialisation pour le système (30) lors de l'étape suivante de résolution pour une valeur plus faible du paramètre ε. En résolvant (30) pour chaque valeur de ε, nous obtenons une solution vectorielle $\vec{f}$ dont chaque composante converge vers une solution de restauration et classification.

L'algorithme qui en découle est présenté dans le tableau 4. Dans ce tableau, Les P EDP de l'étape `ii`) issues de la minimisation de $J^*_{\varepsilon_l}$ par rapport à $\vec{f}$ sont résolues par un algorithme de gradient conjugué. L'indice l est relatif à la boucle sur ε, alors que l'indice q est relatif aux étapes concernant les minimisations alternées (30). En pratique, l'initialisation $\vec{f_0}$ de $\vec{f_\varepsilon}$ est telle que $\vec{f_0} = \vec{I}$.

Remarque concernant le calcul de $|\vec{\nabla} f|_{\mathbb{R}^P}$

Afin de combiner l'information des différentes bandes au niveau du calcul desvariations de $\vec{f}$, nous utilisons des notions de géométrie riemannienne au lieu de la norme $L^2(\mathbb{R}^P)$ appliquée à $|\vec{\nabla} f|_{\mathbb{R}^P}$. Ce calcul intervient au niveau des EDP pour la mise à jour de la variable b dans (30) et les expérimentations ont donné de meilleurs résultats lorsque l'on utilise des notions de géométrie riemannienne pour calculer les variations de $\vec{f}$, même si on ne peut plus dire que le système (30) résulte de la minimisation de (29).
Cette approche a initialement été proposée par [83], on la retrouve également dans [12, 13, 69, 75, 78]. Le principe de cette méthode consiste à extraire les directions

```
initialiser f⃗0
initialiser ε0
• repéter:
  ε(l+1) = ρ εl avec ρ < 1
  f⃗0_ε(l+1) = f⃗min_εl
  ∘ repéter:
      f = f⃗q_εl
      b = bq_εl
      incrémenter q
      i) résoudre min_b J_εl*(f,b)  avec f  fixé:
```

$$b = \frac{\varphi'(|\vec{\nabla} f|_{\mathbb{R}^P})}{2|\nabla f|_{\mathbb{R}^P}}$$

```
      ii) résoudre min_f J*_εl(f,b)  avec b  fixé:
```

$$\text{pour p=1,...,P} \quad f^p - \lambda^2 \varepsilon div(b\nabla f^p) + \frac{\eta^2}{2\varepsilon} W_p'(f) = I^p$$

```
  ∘ jusqu'à convergence sur f⃗q_εl → f⃗min_εl
• faire décroître εl  jusqu'à convergence sur f⃗min_εl
```

TAB. 4 – *Algorithme associé à la minimisation de (29) par rapport à* $(\vec{f},b)$.

de plus grandes variations d'une fonction à travers les caractéristiques de la forme fondamentale qui lui est associée.

Première forme fondamentale:

Soit $v : (x_1,x_2) \in \mathbb{R}^2 \rightarrow v(x_1,x_2) \in \mathbb{R}^P$ une fonction multidimensionnelle continue, dérivable.
Le *première forme fondamentale* associée à v s'écrit (voir par exemple [42]):

$$dv^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\partial v}{\partial x_i} \cdot \frac{\partial v}{\partial x_j} dx_i dx_j \tag{31}$$

ou encore:

$$dv^2 = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}_{[G]} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \tag{32}$$

avec

$$g_{ij} = \sum_{p=1}^{P} \frac{\partial v^p}{\partial x_i} \frac{\partial v^p}{\partial x_j} \tag{33}$$

La première forme fondamentale dv^2 de v est une forme quadratique dont les coefficients g_{ij} permettent de trouver les directions ainsi que les valeurs extrémales des variations de v. La matrice $[G]$ est symétrique. Les valeurs propres λ_+ et λ_- sont données par:

$$\lambda^{\pm} = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2}, \tag{34}$$

et les vecteurs propres associés par:

$$\theta_+ = \frac{1}{2} \arctan\left(\frac{2g_{12}}{g_{11} - g_{22}}\right), \quad \textit{modulo } \pi$$

$$\theta_- = \theta_+ + \frac{\pi}{2}$$

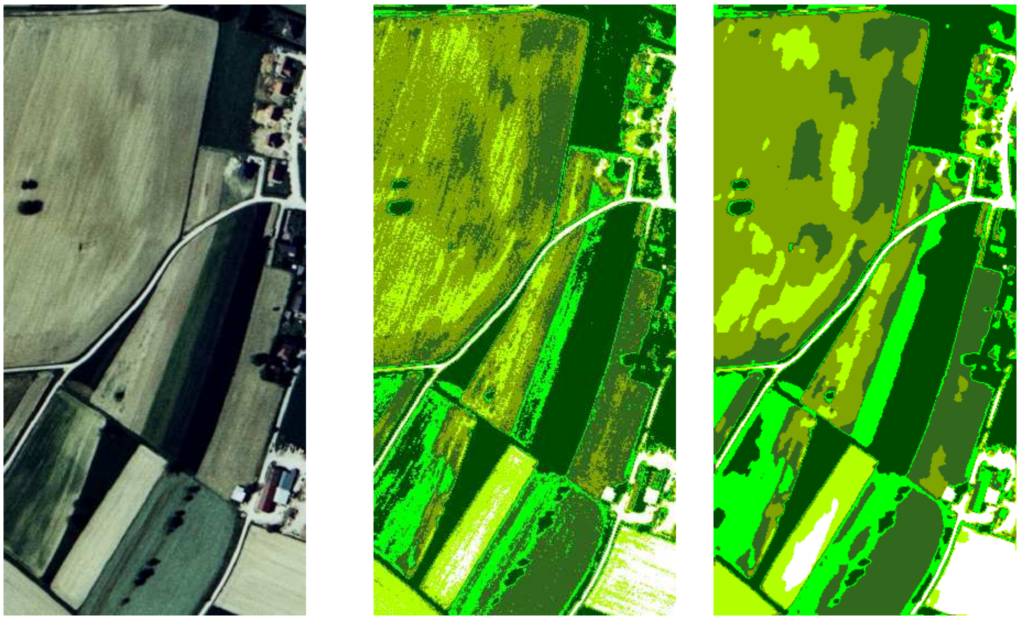
θ_+ est la direction de plus grande variation de v (avec λ_+ la valeur de cette variation), alors que θ_- est la direction de plus faible variation (avec λ_- la valeur de cette variation). On remarque que pour $P = 1$, c'est-à-dire dans le cas monospectral, nous avons $\lambda_+ = |\nabla v|^2$ et $\lambda_- = 0$.

Cette approche concernant le calcul du gradient, a été utilisée dans des schémas d'EDP pour la diffusion anisotrope permettant de lisser le bruit sur des images multispectrales, en combinant l'information des différentes bandes, et sans endommager les contours. Par exemple, un schéma de type:

$$\frac{\partial v}{\partial t} = h(\lambda_+,\lambda_-) \frac{\partial^2 v}{\partial {\theta_-}^2} \tag{35}$$

peut être considéré, où h est par exemple une fonction continue, décroissante de $(\lambda_+ - \lambda_-)$, un choix "naturel" étant $h(\lambda_+,\lambda_-) = \sqrt{\lambda_+ - \lambda_-}$.
L'EDP (35) permet de lisser les données dans les directions de plus faibles variations de l'intensité (si v représente l'intensité lumineuse des différentes bandes), voir par exemple [12, 13, 69].

Notons que si le calcul de $|\nabla \vec{f}|_{\mathbb{R}^P}$ est remplacé par des calculs reposant sur des notions de géométrie riemannienne, c'est-à-dire en remplaçant $\varphi(|\nabla \vec{f}|_{\mathbb{R}^P})$ par $\varphi(\sqrt{\lambda_+ - \lambda_-})$



a) données aériennes couleur b) fuzzy Cmeans modifié c) classification obtenue

FIG. 5 – *Résultats de classification sur des données aériennes en couleur (fig. a) obtenus par l'algorithme de fuzzy Cmeans modifié (fig. b) et par l'algorithme présenté dans le tableau 4.*

dans (28), l'expression des équations d'Euler-Lagrange n'est pas exactement celle donnée dans (30) car il faut alors dériver les valeurs propres extrémales par rapport aux composantes f^p. L'expression exacte de ces dérivées est donnée dans [28]. Ici, nous utilisons les valeurs propres extrémales au niveau des EDP d'Euler-Lagrange pour calculer un gradient multispectral, et non au niveau de la fonctionnelle.

3.3.4 Exemple: imagerie couleur

A titre indicatif, nous montrons un exemple de résultat obtenu sur une image couleur aérienne, à 0.5 mètre de résolution, fournie par l'Institut Géographique National (IGN). En effet, l'approche que nous avons adoptée permet d'effectuer une classification sur des images couleurs, les trois canaux RVB (rouge, vert et bleu) jouant le rôle de trois bandes spectrales. Nous avons au préalable estimé le nombre de classes ainsi que leurs paramètres à l'aide de l'algorithme de *fuzzy Cmeans* avec terme d'entropie que nous avons présenté précédemment. Le nombre initial de classes Cl est de 9 et au bout de 10 itérations de l'algorithme de *fuzzy Cmeans* modifié, nous obtenons un nombre optimal de classes K égal à 6. Le poids α de la contrainte d'entropie dans (18) est de 5.10^6. Nous présentons les données initiales sur la figure 5 (fig. a) et le premier résultat de classification donné par l'algorithme de *fuzzy Cmeans* (fig. b). Le résultat de classification obtenu par le modèle multispectral défini par (29) est présenté sur la figure 5 (fig. c), après 60 itérations sur la décroissance de ε (cf. algorithme dans le tableau 4). La valeur initiale de $\vec{f}$ est celle des données $\vec{I}$, nous n'utilisons donc pas le résultat préliminaire de classification fourni par l'algorithme du *fuzzy Cmeans* modifié (fig. b).

3.4 Extension du deuxième modèle

Le modèle de classification par régions et contours actifs présenté au paragraphe 2.2 s'étend au cas de données $\vec{I}$: $\mathbb{R}^2 \to \mathbb{R}^P$. Les propriétés de régularité de la solution étant les mêmes que dans le cas monospectral, et compte tenu de l'expression du critère monospectral, seul le terme d'attache aux données sera affecté par un passage au multispectral. C'est à travers ce terme que sera combinée l'information des différentes bandes spectrales, les autres termes jouant sur la partition du support Ω qui est unique pour toutes les bandes.

3.4.1 Fonctionnelle multispectrale

Rappelons l'expression de la fonctionnelle que nous avons définie dans le cas monospectral:

$$\begin{aligned} F_\alpha(\Phi_1,...,\Phi_K) = & \underbrace{\sum_{i=1}^{K} e_i \int_\Omega H_\alpha(\Phi_i) \frac{(I(x)-\mu_i)^2}{\sigma_i^2} dx}_{\text{attache aux données}} \\ & + \underbrace{\sum_{i=1}^{K} \gamma_i \int_\Omega g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx}_{\text{minimisation de la longueur des contours}} \\ & + \underbrace{\frac{\lambda}{2} \int_\Omega \Big(\sum_{i=1}^{K} H_\alpha(\Phi_i) - 1 \Big)^2 dx}_{\text{contrainte de partition}} . \end{aligned}$$

Lorsque $\vec{I} \in L^2(\Omega,\mathbb{R}^P)$, nous cherchons une partition des données vectorielles composées d'un ensemble de K régions Ω_i séparées par des interfaces de longueurs minimales Γ_i. Les conditions A et C du tableau 2 imposées sur la solution de partition dans le cas monospectral restent identiques dans le cas multispectral: nous voulons une et une seule étiquette par pixels et nous souhaitons obtenir des frontières régularisées. Seule la condition B doit prendre en compte le caractère vectoriel des données et combiner l'information des P bandes à travers la matrice de variance-covariance Σ_i définie pour chaque classe C_i. Chaque classe C_i est caractérisée par une distribution gaussienne multidimensionnelle de l'intensité $\vec{I}$ sur l'ensemble des régions Ω_i:

$$\Omega_i = \{x \in \Omega / \vec{I}(x) \sim \mathcal{N}_{\mathbb{R}^P}(\vec{\mu}_i,\Sigma_i)\}$$

En approchant l'ensemble Ω_i à l'aide de la fonction d'ensemble de niveaux Φ_i : $\Omega \rightarrow \mathbb{R}$, définie comme une fonction de distance signée, à laquelle est appliquée une approximation de Heaviside (14), nous définissons l'attache aux données par:

$$F_\alpha^{\text{DONNÉES}}(\Phi_1,...,\Phi_K) = \sum_{i=1}^{K} e_i \int_\Omega H_\alpha(\Phi_i)[\vec{I} - \vec{\mu_i}]^T \Sigma_i^{-1}[\vec{I} - \vec{\mu_i}]dx. \tag{36}$$

Une matrice Σ_i diagonale, signifie que les bandes sont complètement décorrélées concernant la classe C_i.
Comme pour le cas monospectral, remarquons que dans le cas où les classes ont les mêmes moyennes et des variances différentes, le terme d'attache aux données défini dans (36) n'est pas discriminant. On peut alors envisager un terme supplémentaire tenant compte du fait que l'intensité $\vec{I}$ suit une loi de distribution gaussienne sur les classes:

$$\begin{aligned} F_\alpha^{\text{DONNÉES}}(\Phi_1,...,\Phi_K) = \sum_{i=1}^{K} e_i \int_\Omega H_\alpha(\Phi_i)\Big\{&[\vec{I} - \vec{\mu_i}]^T \Sigma_i^{-1}[\vec{I} - \vec{\mu_i}] \\ &+log\Big(\sqrt{(2\pi)^P det(\Sigma_i)}\Big)\Big\}dx. \end{aligned} \tag{37}$$

Nous avons négligé le terme logarithmique par la suite car son influence est très faible dans les applications traitées.

Les contraintes de partition et sur la longueur des interfaces restant inchangées dans le cas multispectral, nous proposons de minimiser par rapport aux fonctions $\{\Phi_i\}_i$

le critère multispectral suivant:

$$F_\alpha(\Phi_1,\ldots,\Phi_K) = \underbrace{\sum_{i=1}^{K} e_i \int_\Omega H_\alpha(\Phi_i)[\vec{I}-\vec{\mu_i}]^T \Sigma_i^{-1}[\vec{I}-\vec{\mu_i}]dx}_{\text{attache aux données multispectrales}}$$
$$+ \underbrace{\sum_{i=1}^{K} \gamma_i \int_\Omega g(|\nabla I|)\delta_\alpha(\Phi_i)|\nabla\Phi_i|dx}_{\text{minimisation de la longueur des contours}}$$
$$+ \underbrace{\frac{\lambda}{2}\int_\Omega \Big(\sum_{i=1}^{K} H_\alpha(\Phi_i) - 1\Big)^2 dx}_{\text{contrainte de partition}}\,. \qquad (38)$$

Les différents termes du critère F_α doivent permettre, lors de la minimisation par rapport aux fonctions Φ_i, d'obtenir une partition régulière des P bandes tenant compte des caractéristiques gaussiennes des classes.

3.4.2 Schéma algorithmique

La minimisation de la fonctionnelle (38) repose sur l'annulation des K dérivées de F_α rapport aux fonctions Φ_i. Nous plongeons les EDP d'Euler-Lagrange dans un schéma dynamique où t représente le paramètre de temps, et nous obtenons les K schémas suivants, pour $i = 1,\ldots,K$:

$$\Phi_i^{t+1} = \Phi_i^t - dt\Big\{\delta_\alpha(\Phi_i^t)\Big[e_i[\vec{I}-\vec{\mu_i}]^T\Sigma_i^{-1}[\vec{I}-\vec{\mu_i}] - \gamma_i g(|\nabla I|)div\Big(\frac{\nabla\Phi_i^t}{|\nabla\Phi_i^t|}\Big)$$
$$-\gamma_i\frac{\nabla g\nabla\Phi_i^t}{|\nabla\Phi_i^t|} + \lambda\Big(\sum_{i=1}^{K} H_\alpha(\Phi_i^t) - 1\Big)\Big]\Big\}. \qquad (39)$$

Ces K EDP sont couplées au niveau du terme de partition $\Big(\sum_{i=1}^{K} H_\alpha(\Phi_i^t) - 1\Big)$. L'évolution de chaque fonction d'ensemble de niveaux Φ_i au cours du temps, et par là même celle des régions Ω_i et des contours associés Γ_i, est guidée par un jeu de forces contraignant la solution de partition finale à respecter les propriétés de régularité souhaitées. L'algorithme qui en découle est présenté dans le tableau 5. Nous souhaitons que les fonctions Φ_i restent des fonctions de distance signée au cours de leur évolution à travers le schéma (39). Comme dans le cas monospectral, afin de préserver la contrainte $|\nabla\Phi_i| = 1$ (cf. [6, 24]), nous devons régulariser fréquemment chaque fonction Φ_i au cours du temps par l'EDP:

$$\frac{\partial \Phi_i}{\partial t} = signe(\Phi_i)(1 - |\nabla\Phi_i|), \tag{40}$$

proposée par Sussman *et al.* [72].

```
0 - initialiser Φ_i^0 pour i = 1...K
1 - t ← t + 1
  2 - pour i = 1...K résoudre les K EDP couplées de (39)
  3 - toutes les n itérations de (39), régulariser les
      fonctions Φ_i à l'aide de l'EDP proposée dans [72].
4 - aller à l'étape 1 pour incrémenter le temps t
```

TAB. 5 – *Algorithme associé au schéma (39).*

3.4.3 A propos du choix des paramètres et de l'initialisation

Les différents paramètres du modèle sont ajustés par l'expérimentateur. Il s'agit de régler le poids accordé aux différents termes de la fonctionnelle que nous proposons. Ces poids sont choisis en fonction des données, selon la taille et la morphologie des objets de l'image (poids du terme de longueur des contours), selon le bruit dans l'image (poids du terme d'attache aux données). Comme l'algorithme que nous utilisons est plongé dans un schéma dynamique, nous devons choisir un pas de temps dt que nous fixons en général à 0.2. Nous réinitialisons les fonctions Φ_i toutes les 5 itérations par 5 étapes de l'EDP proposée dans [72]. Le paramètre d'approximation des distributions de Dirac et de Heaviside est fixe, sa valeur est fixée à 3.0 dans toutes les applications suivantes.

L'initialisation des fonctions Φ_i est opérée manuellement ou de façon automatique comme l'illustre la figure 6.
De manière générale, nous choisissons des ensembles de niveaux circulaires. A partir du cercle constituant l'ENZ d'une fonction Φ_i, nous propageons les autres niveaux de cette fonction de distance signée. Une initialisation correcte nécessite d'avoir au moins un pixel de la classe C_i à l'intérieur de la région délimitée par l'ENZ de la fonction

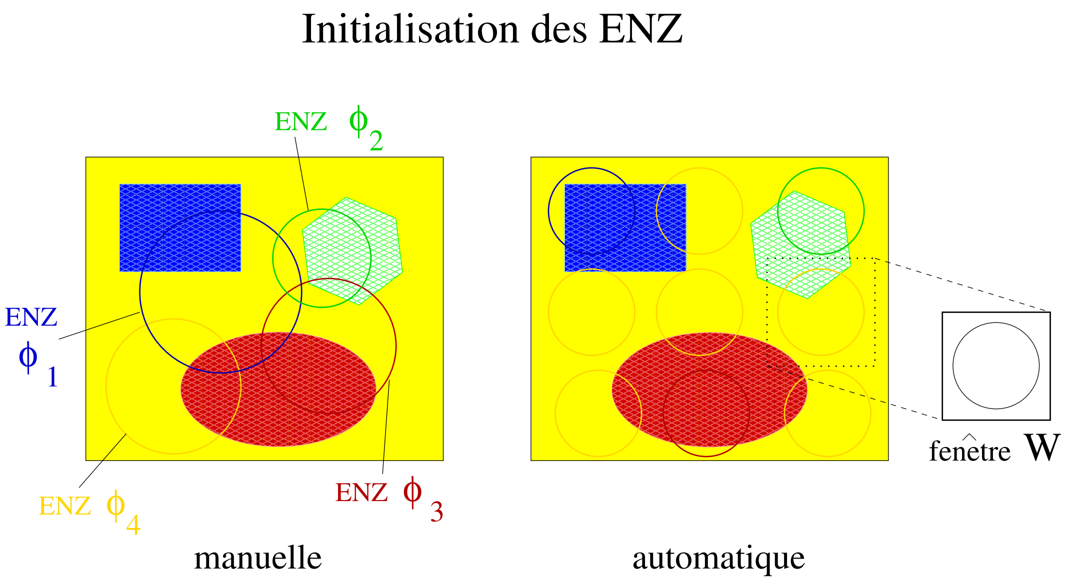


FIG. 6 – *Les deux types d'initialisation des ENZ: manuelle (à gauche) ou automatique (à droite).*

Φ_i. Lorsque l'initialisation est manuelle, l'opérateur positionne lui-même les fonctions Φ_i sur l'image. Pour une initialisation automatique, le principe consiste à découper l'image I en N_W fenêtres $W_{p,\,p=1..N_W}$ de taille fixée. Dans chaque fenêtre W_p, nous calculons la moyenne m_p et l'écart-type σ_p des données I, puis nous cherchons la classe la plus proche. La classe la "plus proche" est déterminée en trouvant l'indice k tel que $k = arg\min_j d_B(\mathcal{N}_{(m_p,\sigma_p)},\mathcal{N}_{(\mu_j,\sigma_j)})$, où d_B est la distance de Bhattacharyya [7, 11, 27, 41] mesurant la distance entre deux distributions gaussiennes $\mathcal{N}_{(\mu_a,\sigma_a)}$ et $\mathcal{N}_{(\mu_b,\sigma_b)}$ par la relation:

$$d_B(\mathcal{N}_{(\mu_a,\sigma_a)},\mathcal{N}_{(\mu_b,\sigma_b)}) = \frac{(\mu_b-\mu_a)^2}{4(\sigma_a+\sigma_b)} + \frac{1}{2} log \frac{|\sigma_a+\sigma_b|}{2\sqrt{|\sigma_a\sigma_b|}}.$$

Ensuite, nous initialisons la fonction de distance signée Φ_k sur chaque fenêtre W_p. Les fenêtres $W_{p,\,p=1..N_W}$ ne se chevauchent pas et ne contiennent qu'une et une seule fonction Φ_k, évitant un recouvrement initial des régions Ω_k définies par les fonctions Φ_k. La taille des fenêtres W_p est liée à la résolution des objets que l'on peut détecter. Une fenêtre trop petite ne contient pas suffisamment d'échantillons de données pour pouvoir obtenir une estimation significative des paramètres m_p et σ_p. Les avantages de cette méthode d'initialisation par rapport à l'initialisation manuelle sont les suivantes: elle est automatique (seule la taille des fenêtres W_p doit être fixée), elle accélère la vitesse de convergence car les contours des régions actives sont plus proches des contours des classes à détecter, elle permet d'être moins sensible au bruit à partir du moment où nous calculons des moyennes sur les données pour initialiser les régions actives sans lisser les frontières des objets à détecter.

Remarque: Les contours montrés sur les différentes figures, visualisent les frontières *internes* des régions actives. Ainsi, à convergence, la position réelle des interfaces entre les régions est située entre ces contours, d'où la présence de pixels sans étiquette au niveau des interfaces. On peut ensuite attribuer une étiquette à ces pixels de la façon suivante: à la fin de la convergence de l'algorithme, à chaque site s sans étiquette, on assigne l'étiquette $L_s \in \Lambda$ telle que:

$$L_s = arg \max_{i=1\ldots K} H_\alpha(\Phi_i(s)) \tag{41}$$

Lorsque les données synthétiques ne sont pas bruitées, nous prenons $\sigma_i = 1.0$, $\forall i = 1..K$.

3.4.4 Exemple: imagerie couleur

A titre indicatif, nous montrons un exemple de résultat obtenu sur l'image couleur aérienne déjà présentée sur la figure 5. Nous présentons sur la figure 7 les données initiales (fig.a) et le premier résultat de classification donné par l'algorithme de *fuzzy Cmeans* (fig. b), dont les paramètres choisis (valeur de α, nombre initial de classes, nombre d'itérations) sont les mêmes que ceux de la figure 5.
Nous utilisons deux types d'initialisation des fonctions Φ_i pour l'algorithme donné dans le tableau 5: une initialisation automatique par des fonctions Φ_i circulaires dans des fenêtres de taille 3×3 (cf. figure 6), donnant le résultat de classification présenté sur la figure 7 (fig. c), et une initialisation des Φ_i par le résultat de classification obtenu à l'aide de l'algorithme de *fuzzy Cmeans* (fig. b), aboutissant au résultat final de classification présenté sur la figure d. Nous constatons sur la figure 7 la régularisation supplémentaire obtenue par le modèle de régions/contours actifs que nous proposons (fig. c et d) par rapport au résultat de classification du *fuzzy Cmeans* (fig. b), indépendamment du type d'initialisation choisi. Les classes que nous obtenons sont bien plus homogènes et les frontières finales sont relativement lisses. Nous avons délibérément choisi une forte régularisation (poids élevé du terme de contrainte sur la longueur des contours dans (38)) pour bien insister sur le pouvoir régularisant de ce modèle.
Nous constatons une légère différence sur les résultats de classifications selon le type d'initialisation choisie (fig.c et fig. d). Il est difficile d'en déduire le meilleur type d'initialisation en visualisant les résultats obtenus. Cependant, une initialisation des Φ_i par une classification erronée de l'algorithme de *fuzzy Cmeans* pourrait conduire à une mauvaise classification par régions et contours actifs. Dans ce cas, il semble préférable d'utiliser uniquement l'algorithme de *fuzzy Cmeans* modifié pour l'estimation du nombre et des paramètres des classes, par précaution. Le problème ne se pose pas pour le paragraphe suivant, car nous n'utilisons par le modèle de *fuzzy Cmeans* modifié puisque le nombre de classes ainsi que leurs paramètres ont été estimés par un autre algorithme dont les résultats nous ont été fournis par le projet VISTA de l'IRISA.
Nous montrons sur la figure 8 l'évolution des régions et de leurs contours (ENZ) au cours des itérations de l'algorithme du tableau 5 pour une initialisation automatique des Φ_i, correspondant au résultat de classification c) de la figure 7. Sur la figure 9, nous présentons l'évolution des régions et de leurs contours pour une initialisation des Φ_i par le résultat de classification donné par l'algorithme de *fuzzy Cmeans*, et aboutissant au résultat de classification d) de la figure 7.


a) données aériennes couleur   b) fuzzy Cmeans modifié


c) classif. pour init. 1   d) classif. pour init. 2

FIG. 7 – *Résultats de classification sur des données aériennes en couleur.*

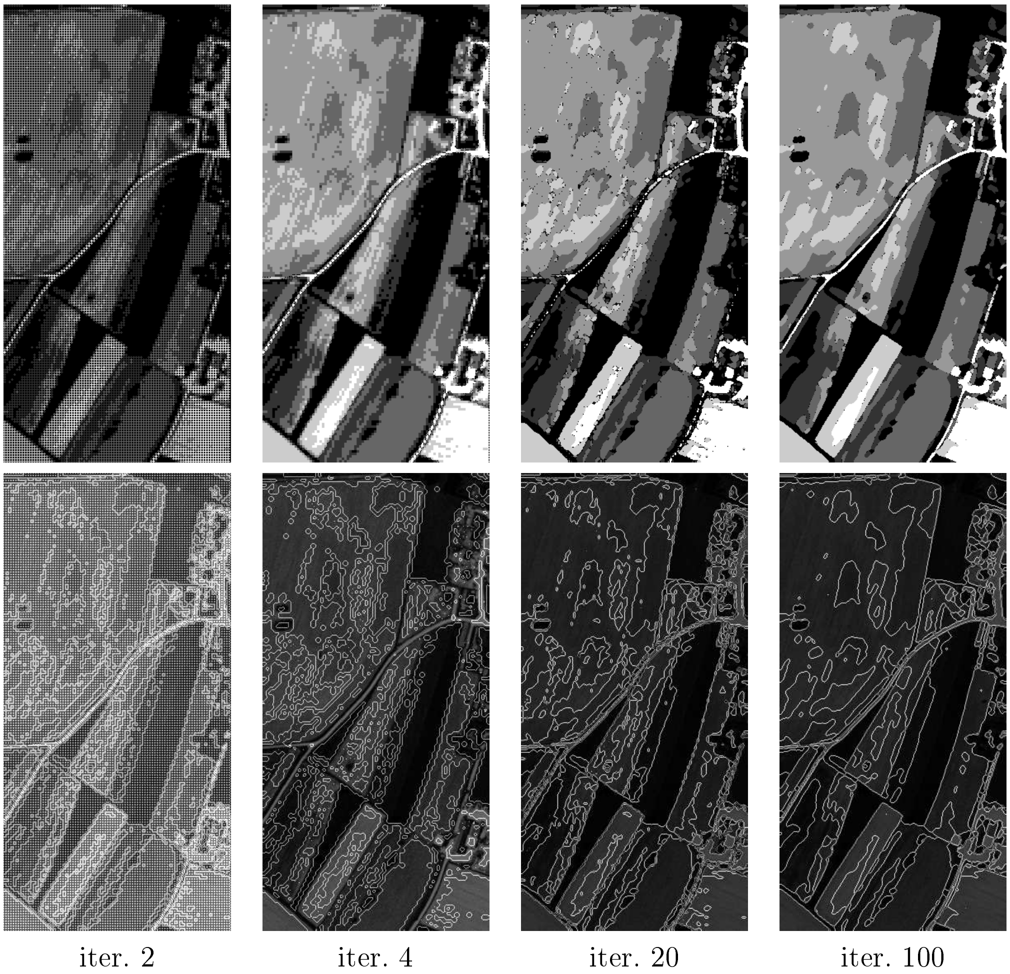


FIG. 8 – *Image couleur de l'IGN: évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des* Φ_i *dans des fenêtres 3×3.*

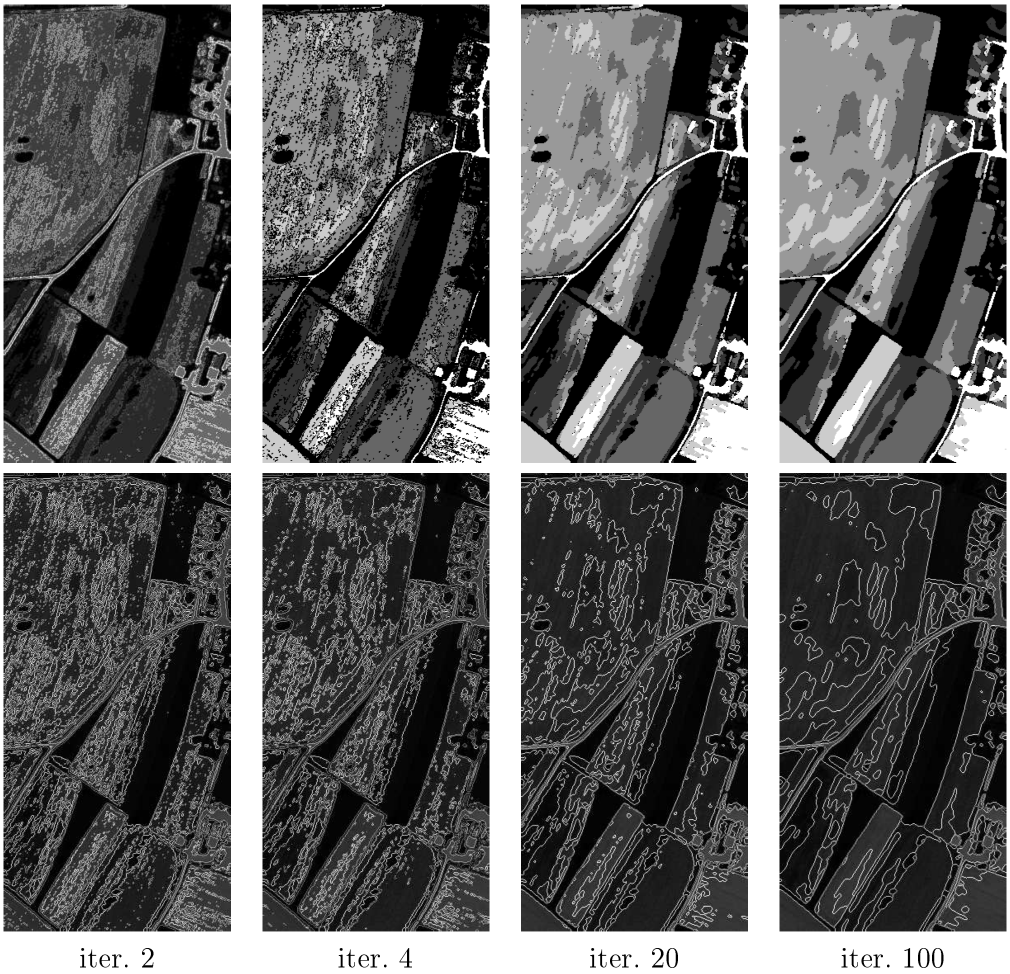


FIG. 9 – *Image couleur de l'IGN: évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des* Φ_i *par le résultat de classification obtenu par fuzzy Cmeans modifié.*

4 Une application sur des données SPOT

4.1 Présentation des données

Les données multispectrales SPOT 2 de la baie de Lannion présentées sur la figure 11 ont fait l'objet d'une étude approfondie dans la thèse d'Annabelle Chardin [20] menée à l'IRISA dans le projet VISTA. Cette étude entre dans le cadre d'une collaboration entre plusieurs laboratoires de la région Bretagne impliqués dans la télédétection et regroupés au sein du GSTB (Groupement Scientifique pour la Télédétection en Bretagne). Ces données permettent de mesurer l'impact de la culture intensive sur l'évolution de l'utilisation des sols dans cette baie. Une étude de l'occupation des sols à partir d'images prises à différentes époques (images SPOT 2 en mode XS prises en hiver 1996 et été 1997) a permis d'analyser cette évolution. A. Chardin a développé un algorithme de classification fondé sur des champs de Markov hiérarchiques que nous présentons par la suite. Les résultats qui ont été obtenus dans le cadre de sa thèse ont été comparés à deux autres modèles, le maximum de vraisemblance MV ainsi qu'à l'"ICM niveau 0" qui correspond au premier niveau du modèle hiérarchique mentionné ci-dessus. De plus, les géographes du COSTEL (Climat, Occupation des Sols par TELédétection) de l'Université de Rennes II ont fourni une vérité terrain permettant de valider ces résultats. Tous ces résultats ont été gracieusement mis à notre disposition par le projet VISTA de l'IRISA Rennes que nous tenons à remercier, et les données SPOT de la scène d'été 1997 ont été acquises auprès de la société SPOTIMAGE. Notre objectif n'est pas de réaliser une étude approfondie de l'utilisation des sols, comme cela a pu être le cas dans [20], mais de pouvoir comparer nos résultats de classification avec d'autres approches sur des donnés satellitaires multispectrales pour lesquelles nous disposons d'une vérité de terrain établie par des experts géographes.

4.2 Modèle markovien hiérarchique sur arbre tronqué

L'approche adoptée dans [20] entre dans le cadre des modèles markoviens hiérarchiques. Le lecteur intéressé par ces approches et algorithmes associés pourra trouver de plus amples informations sur le sujet dans la synthèse [33] ou dans les travaux [14, 36, 39, 38, 44, 60, 61]. L'originalité du modèle hiérarchique proposé dans [20] repose sur l'introduction d'une structure hiérarchique permettant de combiner des a priori spatiaux et hiérarchiques: il s'agit d'un quadrarbre tronqué dont les noeuds du niveau le plus grossier supportent un champ de Markov spatial sur une grille.

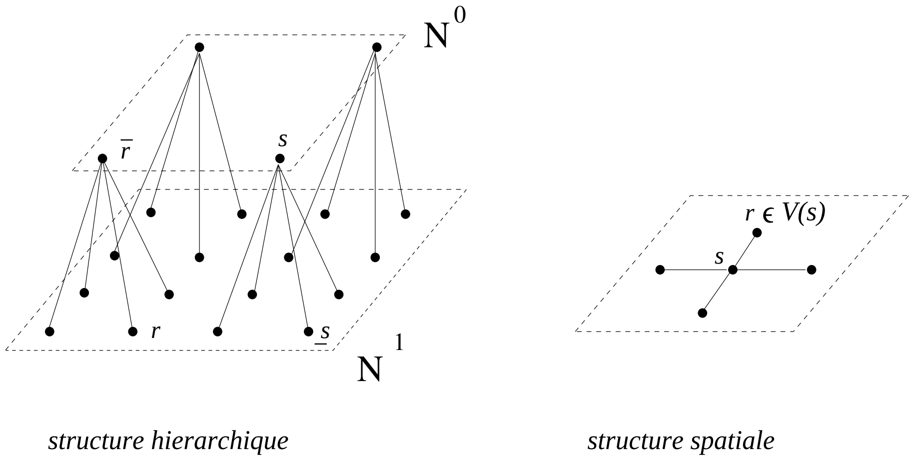


FIG. 10 – *Structure de l'arbre tronqué.*

Dans un cadre stochastique, on cherche à estimer un étiquetage représenté par la variable L, étant données les observations représentées par la variable O. Nous ne détaillerons pas ce modèle, mais nous nous contenterons d'expliciter les points essentiels à sa compréhension générale. Tous les détails relatifs à cette approche sont clairement présentés dans [20].

On considère une structure pyramidale de $M+1$ niveaux de résolution. Soit N^m l'ensemble des noeuds du niveau m, avec $m \in [0; M]$, où N^M est le niveau de résolution le plus fin et N^0 le plus grossier. L'arbre tronqué est un quadrarbre dont la racine N^0 n'est pas un pixel, mais une grille spatiale rectangulaire munie d'un système de voisinage du premier ordre et dont chaque noeud est la racine d'un quadrarbre. On note $\underline{s}$ l'ensemble des enfants de s, $\bar{r}$ le parent du site r (voir l'illustration sur la figure 10) et $\underline{\underline{s}}$ l'ensemble des sites formant l'arbre de racine s.

Dans le cadre de l'estimateur du MAP (Maximum A Posteriori), l'estimé $\hat{L}$ de l'étiquetage L est obtenu par:

$$\hat{L} = \arg\min_L E(L), \quad \text{où } E \text{ est l'énergie globale définie par:}$$

$$E(L) = \sum_{s \in N^0} g_s(L_s, O_s) + \sum_{<r,s> \in N^0} \underbrace{f_{r,s}(L_r, L_s)}_{\text{Potts spatial}}$$

```
Passe montante
    - Feuilles s, s ∈ N^M :
        V_s(L_s̄) = min_{L_s}[h_s(L_s,L_s̄) + g_s(L_s,O_s)]
        L*_s(L_s̄) = arg min_{L_s}[h_s(L_s,L_s̄) + g_s(L_s,O_s)]
    - Récurrence pour m = M − 1,...,1 et s ∈ N^m :
        V_s(L_s̄) = min_{L_s}[h_s(L_s,L_s̄) + g_s(L_s,O_s) + Σ_{r∈s̲} V_r(L_s)]
        L*_s(L_s̄) = arg min_{L_s}[h_s(L_s,L_s̄) + g_s(L_s,O_s) + Σ_{r∈s̲} V_r(L_s)]
ICM sur la grille grossière N^0
    - Initialisation:
        ∀s ∈ N^0, L_s^(0) = arg min_λ[g_s(λ,O_s) + Σ_{r∈s̲} V_r(λ)]
    - Effectuer plusieurs mises à jour de tous les sites de N^0,
      en faisant décroître l'énergie
        Σ_{<r,s>∈N^0} f_{r,s}(L_r,L_s) + Σ_{s∈N^0}[g_s(L_s,O_s) + Σ_{r∈s̲} V_r(L_s)]
      pour obtenir L̂_s, ∀s ∈ N^0
Passe descendante
    pour m = 1,...,M, s ∈ N^m
        L̂_s = L*_s(L̂_s̄)
```

TAB. 6 – *Algorithme de minimisation.*

$$+ \sum_{s \in N \setminus N^0} \Big[\underbrace{h_s(L_s,L_{\bar{s}})}_{\text{Potts hiérarchique}} + g_s(L_s,O_s) \Big] \tag{42}$$

avec g_s une fonction quadratique prenant en compte la répartition gaussienne de l'intensité lumineuse dans les classes (i.e. attache aux données). Les deux termes d'a priori $f_{r,s}$ et h_s sont des modèles de Potts 4-connexes [80], $f_{r,s}$ étant un a priori de Potts *spatial* (en 4-connexité), alors que h_s est un a priori de Potts *hiérarchique*.
On pose:

$$V_s(L_{\bar{s}}) = \min_{L_{\underline{s}}} \Big[\sum_{r \in \underline{s}} [h_r(L_r,L_{\bar{r}}) + g_r(L_r,O_r)] \Big] \tag{43}$$

et l'énergie E peut alors s'écrire:

$$E(L) = \sum_{<r,s> \in N^0} f_{r,s}(L_r,L_s) + \sum_{s \in N^0} \Big[g_s(L_s,O_s) + + \sum_{r \in \underline{s}} V_r(L_s) \Big]. \tag{44}$$

L'algorithme hiérarchique d'optimisation de (44), issu de [20], est présenté dans le tableau 6. Il se décompose en trois étapes: une passe montante sur les sous-arbres dont la racine est un site de N^0, un algorithme ICM sur la grille grossière N^0 et une passe descendante sur les sous-arbres parcourus lors de la passe montante.

4.3 Résultats et comparaisons

4.3.1 Présentation des résultats

Nous avons conduit des expérimentations sur des images SPOT 2 en mode XS de la baie de Lannion. Ces images datent du 15 Août 1997, et nous avons choisi de les acquérir pour pouvoir comparer nos résultats de classification avec ceux fournis par le projet VISTA.
Les images originales issues des 3 bandes spectrales ont une taille de 1480×1024 pixels, et nous avons mené nos expérimentations sur une portion de la scène de taille 400×400 pixels. Cette zone a été choisie pour la diversité morphologique et thématique des objets qui la composent. L'intérêt de cette zone provient également du fait que la densité d'échantillons de la vérité terrain fournie par le COSTEL est importante, comparée au reste de la scène globale. Nous présentons sur la figure 11 les 3 scènes XS1, XS2 et XS3 considérées.

Par la suite, nous désignerons les différents résultats de classification par:

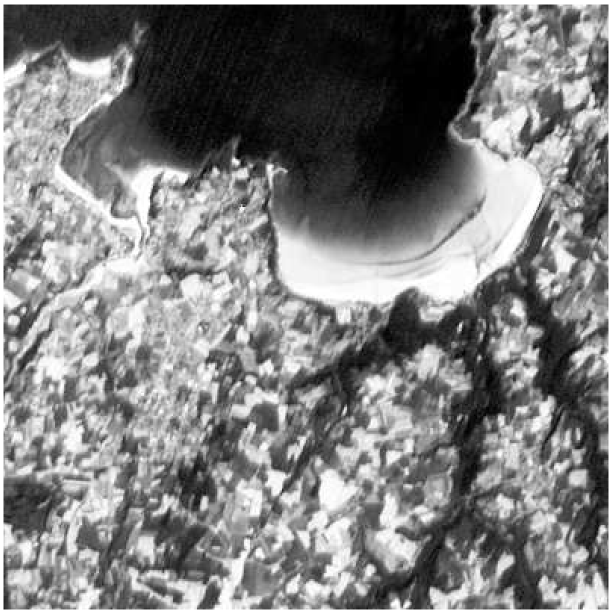
- **MV**, pour le Maximum de Vraisemblance
- **ICM-N0**, pour le modèle (42) proposé par A. Chardin, mais sans aspect hiérarchique, c'est-à-dire pour $N = 0$
- **H-MAP**, pour le modèle hiérarchique sur arbre tronqué proposé par A. Chardin (42), avec pour niveau de résolution le plus fin $N = 3$
- **M1X**, pour la version multispectrale du premier modèle que nous avons proposé (cf. algorithme dans le tableau 4)
- **M2X**, pour la version multispectrale du second modèle que nous avons proposé (cf. algorithme dans le tableau 5)

Les résultats de classification MV, ICM-N0 et H-MAP proviennent des travaux de [20]. Notons que ces trois algorithmes ont été expérimentés sur les scènes complètes de taille 1480×1024 pixels et sur une machine différente de la nôtre, nous ne comparerons donc pas les temps de calcul. Concernant les résultats H-MAP, la taille initiale de la grille de niveau $N = 0$ est de 185×128 pixels.

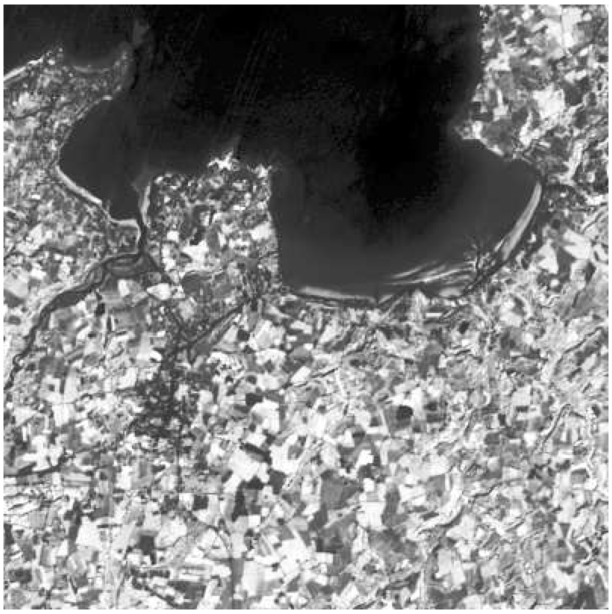
Le nombre de classes est fixé par les thématiciens et géographes du laboratoire COSTEL, et la légende des thèmes d'occupation des sols est représentée sur la figure 13. Les géographes ont sélectionné des zones différentes de l'image, sous forme de petites parcelles rectangulaires, pour l'estimation des paramètres des 8 classes d'une part (parcelles d'apprentissage), et pour valider les résultats de classification d'autre part (parcelles de vérité terrain). La figure 12 présente ces parcelles sur des images de taille 400×400 pixels extraites des images originales dont la taille est de 1480×1024 pixels. Notons au passage que les parcelles d'apprentissage et de validation ne sont pas les mêmes. Les paramètres des classes que nous avons utilisés ont été estimés sur les parcelles d'apprentissage de l'image complète 1480×1024 et sont présentés dans [20], alors que nous effectuons la validation des résultats obtenus par les différents modèles à partir des extraits 400×400 de la vérité terrain.
Concernant l'estimation des paramètres, le nombre d'échantillons diffère selon les classes. En particulier pour la classe "légume" (classe 7) il n'y a que 38 pixels d'échantillons pour l'apprentissage: les résultats pour cette classe sont par conséquent peu fiables, voire insignifiants, d'autant que, dans la portion d'image 400×400 que nous avons choisi de traiter, la vérité terrain ne comporte que 5 pixels de validation pour cette classe. Il en est quasiment de même pour la classe 4 ("bois et landes") puisque la vérité terrain ne comporte que 14 pixels dans la portion 400×400 que nous avons sélectionnée. La vérité terrain contient au total 1466 pixels, permettant de localiser et de compter le nombre de pixels bien/mal classés sur les petites zones rectangulaires représentées sur la figure 12. Le nombre de pixels par classe pour la vérité terrain est comptabilisé sur la dernière ligne du tableau 8.

Nous présentons dans un premier temps le résultat de classification obtenu par MV sur la figure 13, c'est-à-dire sans aucune régularisation sur la solution cherchée. Nous constatons que la classification comporte un grand nombre de petites régions et que les classes sont peu homogènes.
Sur la figure 14, nous présentons les quatre résultats de classification obtenus par les modèles ICM-N0, H-MAP, M1X et M2X. Pour le modèle M1X, nous avons effectué 40 itérations sur la décroissance de ε. Concernant le modèle M2X, nous avons effectué 60 itérations pour ce qui est de l'évolution dynamique des régions et de leurs contours, et nous avons initialisé les régions associées aux fonctions Φ_i ($i = 1 ... 8$) dans des fenêtres de taille 5×5.

canal XS1

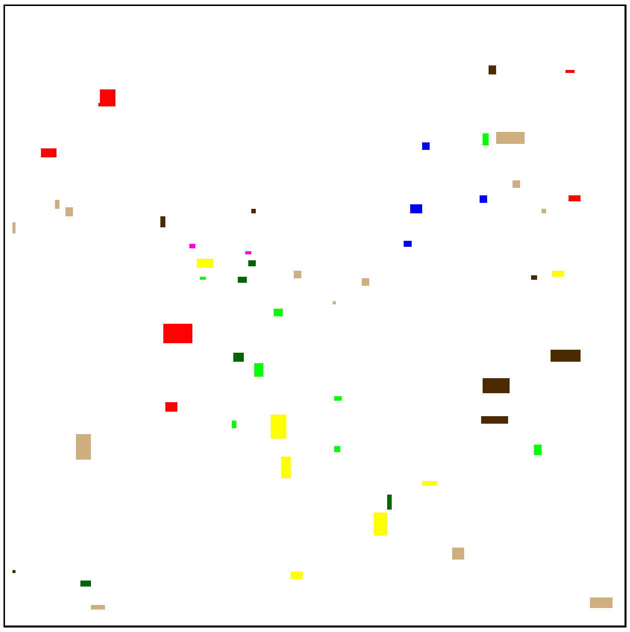


canal XS2

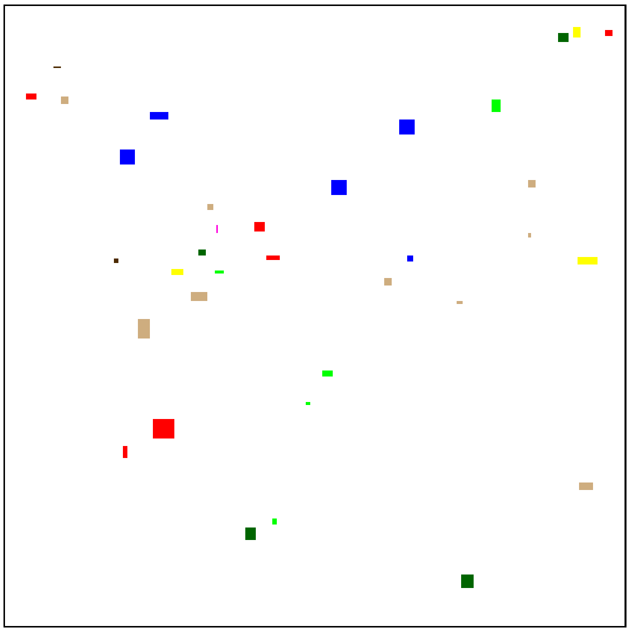


canal XS3

FIG. 11 – *Données multispectrales SPOT 2 de la baie de Lannion (Août 1997) après égalisation de l'histogramme.*



zones d'apprentissage



zones de validation

FIG. 12 – *Zones d'apprentissage (2719 pixels) dans lesquelles les paramètres des 8 classes ont été estimés et zones servant de validation (1466 pixels) pour les résultats de classification obtenus (vérité terrain).*

1 - mer et eau

2 - sable et sols nus ou peu couverts

3 - zones urbaines

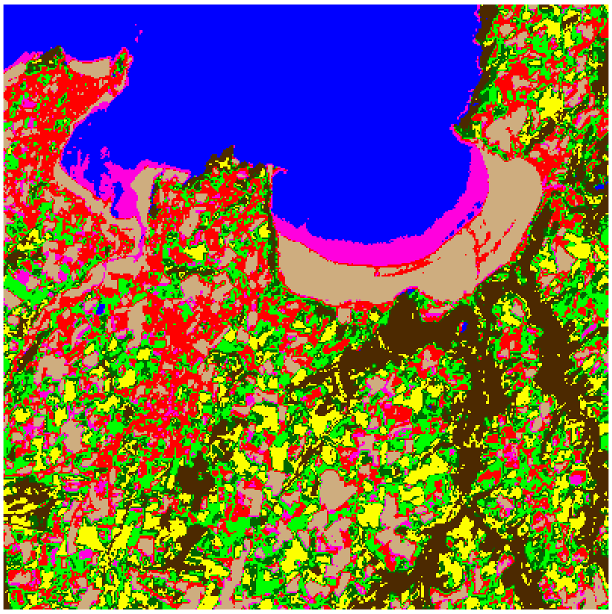
4 - bois et landes

5 - prairies naturelles

6 - prairies artificielles

7 - légumes

8 - maïs



classification par MV

FIG. 13 – *Légende des thèmes d'occupation des sols et résultat de classification pour le modèle MV.*

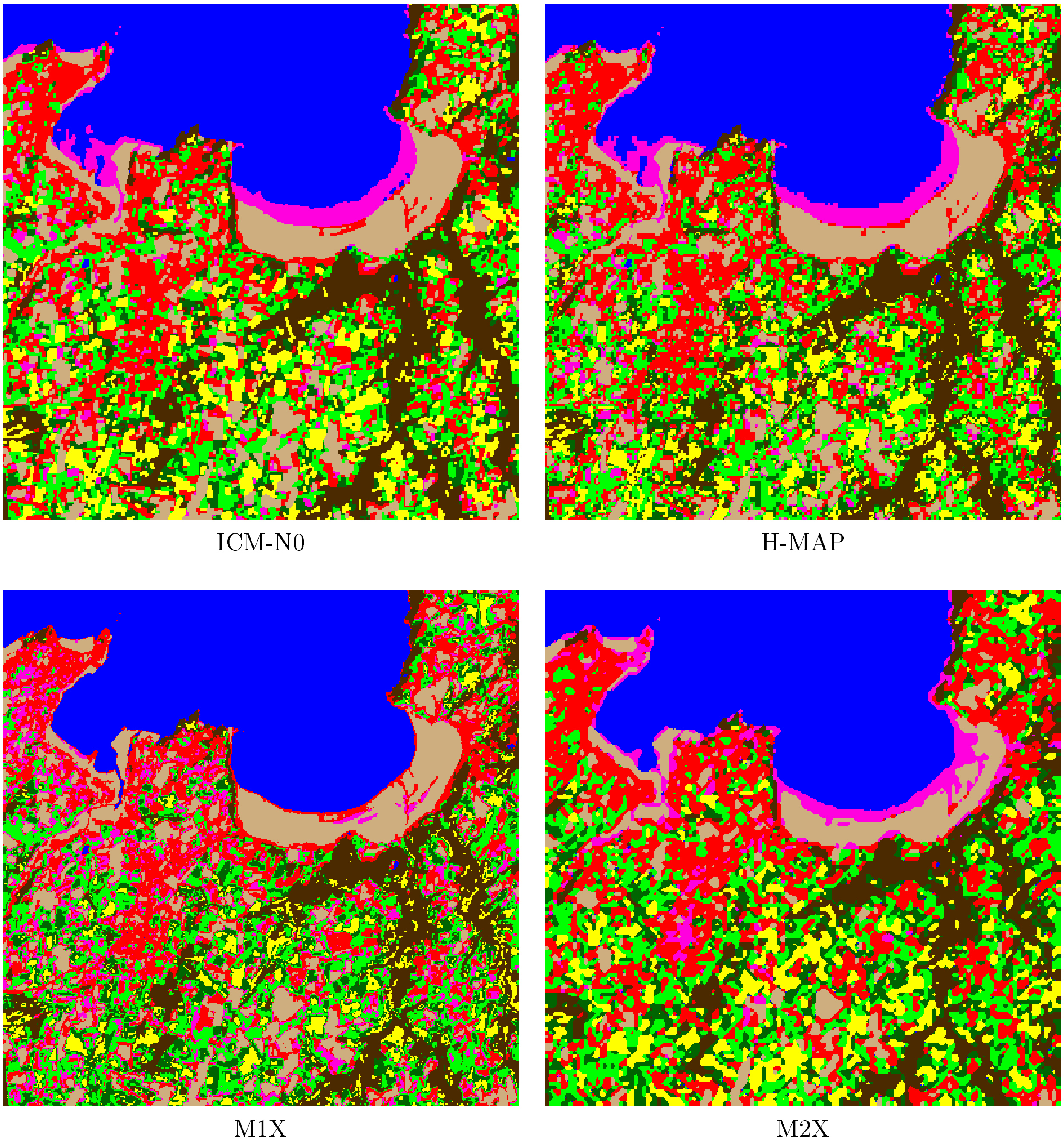


FIG. 14 – *Résultats de classification pour les modèles ICM-N0, H-MAP, M1X et M2X.*

modèle	**nombre d'erreurs**	**pourcentage de réussite**
MV	512	65.07%
ICM-N0	498	66.03%
H-MAP	433	70.46%
M1X	421	71.28%
M2X	440	69.99%

TAB. 7 – *Nombre d'erreurs commises et pourcentages de réussite pour les différents modèles de classification*

4.3.2 Commentaires

A première vue, nous pouvons constater sur la figure 14 que la régularité de la solution est différente selon les modèles de classification. Le modèle M2X est celui qui fournit le plus de régularisation dans le sens où les régions obtenues sont très homogènes avec des frontières relativement lisses. Cette régularité est le résultat conjoint du terme de contrainte sur la longueur des contours de la fonctionnelle proposée (38) et surtout de la procédure automatique d'initialisation des fonctions Φ_i dans des fenêtres 5×5, limitant la résolution spatiale du modèle. La procédure automatique semble avoir plus d'impacts sur la morphologie des régions que le terme de minimisation de longueur des contours dans le sens où les régions obtenues n'ont pas nécessairement une faible courbure puisque nous trouvons un certain nombre de "coins".
Nous pouvons constater que les résultats des modèles ICM-N0 et H-MAP conduisent à de nombreuses régions dont les contours sont horizontaux ou verticaux. Ce phénomène a sans doute pour origine la 4-connexité des modèles de Potts utilisés dans ces deux modèles.
Nous remarquons que la classe 7 ("légumes"), dont nous avons mentionné les problèmes en termes de validité, est gérée de façon très similaire par les modèles ICM-N0, H-MAP et M2X, alors que le modèle M1X conduit à des résultats assez différents. Même si cette classe n'est pas significative pour la validation des résultats, il est intéressant de noter le comportement original du modèle M1X par rapport aux trois autres modèles, dont le terme de classification (contrainte de niveaux W) est très différent des termes d'attache aux données gaussiens des autres modèles.
Le tableau 7 présente les résultats obtenus par les différents modèles en fonction de la vérité terrain. Dans ce tableau, nous donnons le nombre n de pixels mal classés (appelé "erreurs" par la suite) ainsi que le pourcentage p de réussite selon les modèles. Les nombres n et p sont liés par la relation $p = 100.(1 - \frac{n}{N})$. Les modèles MV

et ICM-N0 offrent les moins bonnes performances et les trois autres modèles ont des performances du même ordre. Le lecteur intéressé pourra consulter le travail d'A. Chardin [20] pour une étude comparative approfondie des modèles MV, ICM-N0 et H-MAP.
Nous présentons le détail, classe par classe, des performances des différents modèles dans le tableau 8.
Pour la classe "mer et eau", les performances des différents modèles sont bonnes, et très légèrement meilleures pour ce qui est des modèles variationnels M1X et M2X. Le modèle M2X est moins performant que les autres pour la classe "sable et sols nus". Une analyse plus fine montre que cet excès d'erreurs, produit par le modèle M2X par rapport aux autres modèles, provient des petites parcelles de la vérité terrain pour la classe "sable et sols nus" (cf. figure 12). Ces parcelles possèdent une taille inférieure à la résolution maximale du modèle M2X, laquelle est de l'ordre de 5×5 pixels. Il en est de même pour la classe "bois et landes" dont la vérité terrain contient (seulement) 14 pixels et est constituée de deux parcelles de tailles respectives 5×1 et 3×3 pixels. Le modèle H-MAP est celui qui offre les meilleurs performances en ce qui concerne la classe "zones urbaines". Les villes sont très texturées, et une extraction performante des zones urbaines nécessite de prendre en compte des paramètres de texture comme cela a été effectué dans [46]. Les différents modèles de classification que nous avons comparés ne prennent pas de tels paramètres de texture en compte, cependant il semble que la structure hiérarchique du modèle H-MAP soit plus robuste que les autres en ce qui concerne la classification des zones urbaines. Les prairies artificielles et naturelles ont un fort chevauchement spectral et sont difficiles à séparer. Il en résulte de mauvaises performances pour ces deux classes, en particulier pour la classe "prairies artificielles" dont les éléments ont tendance à être attribués à la classe "prairie naturelle". Les performances sur la classe "maïs" sont les meilleures pour le modèle M2X qui dépasse les performances des autres modèles d'environ 20% ou plus.

5 Conclusion et perspectives

Nous avons étendu les modèles variationnels de classification développés dans [64, 67, 65, 68] au cas des données multispectrales. Cette extension intervient à différents niveaux pour ces deux modèles.
L'extension du modèle combinant restauration et classification concerne tous les termes de la fonctionnelle définie dans le cas monospectral, dont le terme de contrainte de niveaux (potentiel W) pour lequel l'extension au cas multispectral est assez déli-

cate à effectuer. Les résultats théoriques issus de la théorie de Van de Waals-Cahn-Hilliard restent valables pour le cas multispectral.
L'extension du modèle dynamique est quasi-immédiate, puisqu'elle ne concerne qu'un seul terme de la fonctionnelle définie dans le cas monospectral, le terme d'attache aux données, la seule difficulté consistant à calculer une norme dans un espace muti-dimensionnel.
Nous avons testé ces deux modèles sur des images aériennes en couleur, dont les canaux RVB ont été assimilés à différentes bandes spectrales, comme les bandes XS1, XS2 et XS3 des satellites SPOT.
Nous avons mené une expérience sur des données multispectrales SPOT pour lesquelles nous disposions d'une série de résultats ainsi que d'une vérité terrain. Nous avons comparé et analysé en détails les résultats obtenus par les différents modèles. Notons que les résultats fournis le projet VISTA ont été obtenus sur des images taille 1480×1024 pixels, alors que nous avons testé nos modèles sur des portions 400×400 pixels afin d'accélérer les calculs.
Nous avons évalué les performances des différents modèles en terme de pixels mal classés, grâce à la vérité terrain fournie par les géographes du laboratoire COSTEL. Il est important de souligner le fait que l'avis d'un expert doit faire partie du processus de validation. En effet, d'une part la vérité terrain ne contient que 1466 pixels sur 160000, ce qui est relativement peu pour en tirer des conclusions générales, d'autre part nous pouvons constater que les résultats sont visuellement différents au niveau de la régularité morphologique des régions. La connaissance d'un expert est donc d'une grande importance pour juger de la qualité des résultats et pour déterminer le modèle de classification le mieux adapté pour une application donnée.

Remerciements

Nous remercions l'Institut Géographique National (IGN) pour les données aériennes traitées dans ce rapport. Nous tenons à remercier sincèrement le projet VISTA (VIsion Spatio-Temporelle et Active) de l'IRISA Rennes, ainsi que le laboratoire COSTEL (Climat, Occupation des Sols par TELédétection) de l'Université de Rennes II pour la vérité de terrain ainsi que les différents résultats fournis sur les données SPOT de la baie de Lannion.

CLASSES

	1 mer eau	**2** sable sols nus	**3** zones urbaines	**4** bois landes	**5** prairies naturelles	**6** prairies artificielles	**7** légumes	**8** maïs	**total**
MV	16 (95.7%)	77 (75.9%)	119 (63.2%)	1 (92.9%)	54 (49.1%)	182 (4.2%)	5 (0%)	58 (56.1%)	512 (65.07%)
ICM-N0	16 (95.7%)	73 (77.2%)	103 (68.1%)	1 (92.9%)	50 (52.8%)	189 (0.5%)	5 (0%)	61 (53.8%)	498 (66.03%)
H-MAP	16 (95.7%)	76 (76.3%)	63 (80.5%)	1 (92.9%)	41 (61.3%)	177 (6.8%)	5 (0%)	54 (59.1%)	433 (70.46%)
M1X	0 (100%)	79 (75.3%)	93 (71.2%)	2 (85.7%)	52 (50.9%)	146 (23.2%)	4 (20%)	44 (66.7%)	421 (71.28%)
M2X	0 (100%)	105 (67.2%)	101 (68.8%)	9 (35.7%)	41 (61.32%)	160 (15.8%)	5 (0%)	19 (85.6%)	440 (70.0%)
nb pix. pour VT	376	320	323	14	106	190	5	132	1466

TAB. 8 – *Pour chaque classe: nombre de pixels* **mal** *classés et pourcentage de* **réussite** *selon les modèles. La dernière ligne comptabilise le nombre de pixels d'échantillons par classe pour la vérité terrain (VT).*

Références

[1] L. Alvarez, P.-L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. of Numerical Analysis*, 29(3):845–866, 1992.

[2] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Commun. Pure Appl. Math.*, 43:999–1036, 1990.

[3] L. Ambrosio and V. Tortorelli. On the approximation of functionals depending on jumps by quadratic, elliptic functionals. *Boll. Un. Mat. Ital.*, 46-B:105–123, 1992.

[4] G. Aubert and L. Vese. A variational method in image recovery. *SIAM J. of Numerical Analysis*, 34(5):1948–1979, 1997.

[5] S. Baldo. Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids. *Ann. Inst. Henri Poincaré*, 7:67–90, 1990.

[6] G. Barles, H. M. Soner, and P.E. Souganidis. Front propagation and phase field theory. *SIAM J. Control and Optimization*, 31:439–479, 1993.

[7] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal processing*, 18:349–369, 1985.

[8] G. Bellettini, M. Paolini, and C. Verdi. Numerical minimization of geometrical type problems related to calculus of variations. *Calcolo*, 27:251–278, 1991.

[9] M. Berthod, Z. Kato, S. Yu, and J. Zerubia. Bayesian image classification using Markov random fields. *Image and Vision Computing*, 14(4):285–293, 1996.

[10] J.C. Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New-York, 1981.

[11] C. Bhattacharyya. A simple method of resolution of a distribution into Gaussian components. *Biometrics*, 23(4):115–135, 1967.

[12] P. Blomgren. *Color TV: Total Variation Methods for Restoration of Vector-Valued Images*. PhD thesis, University of California, Los Angeles, 1998.

[13] P. Blomgren and T.F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Trans. on Image Processing*, 7(3):304–309, Mars 1998. Special issue on partial differential equations and geometry driven diffusion in image processing and analysis.

[14] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3:162–177, Mars 1994.

[15] A. Braides. *Approximation of free-discontinuity problems*. Lecture notes in mathematics, number 1964, Springer, 1998.

[16] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours. *Numerische Mathematik*, 66:1–31, 1993.

[17] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International J. of Computer Vision*, 22(1):61–79, 1997.

[18] T. Chan and L. Vese. An active contour model without edges. In *Proc. of IEEE ICCV, Corfu, Grèce*, pages 141–151, Septembre 1999.

[19] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. on Image Processing*, 6(2):298–311, février 1997.

[20] A. Chardin. *Modèles énergétiques hiérarchiques pour la résolution des problèmes inverses en analyse d'images.* PhD thesis, Université de Rennes I, France, 2000.

[21] R. Conners, M. Trivedi, and C. Harlow. Segmentation of high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 25:273–310, 1984.

[22] X. Descombes, R. Morris, and J. Zerubia. Quelques améliorations à la segmentation d'images bayésiennes - Deuxième partie : classification. *Traitement du Signal*, 14(4):383–395, 1997.

[23] X. Descombes, R. Morris, and J. Zerubia. Quelques améliorations à la segmentation d'images bayésiennes - Première partie : modélisation. *Traitement du Signal*, 14(4):373–382, 1997.

[24] L.C. Evans and J. Spruck. Motion of level sets by mean curvature. part II. *Trans. of the American Mathematical Society*, 330(1):321–332, 1992.

[25] I. Fonseca and L. Tartar. The gradient theory of phase transitions for systems with two potential wells. *Proc. of the Royal Society of Edinburgh*, 111A(11):89–102, 1989.

[26] H. Frigui and R. Krishnapuram. A robust clustering algorithm based on competitive agglomeration and soft rejection of outliers. In *CVPR*, pages 550–555, San Francisco, Etats-Unis, juin 1996.

[27] K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, New York, 1972.

[28] S. Gadat. Restauration d'images satellitaires multispectrales. Rapport interne, projet ARIANA, INRIA Sophia Antipolis, France, 1999.

[29] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1992.

[30] S. Geman and D.-E. McClure. Bayesian image analysis : an application to single photon emission tomography. *Proc. Statist. Comput. Sect.*, Washington DC, Amer. Statist. Assoc.:11–18, 1985.

[31] E. De Giorgi. Convergence problems for functionals or operators. *Proc. of the International Meeting on Recent Methods in Nonlinear Analysis, Pitagoria, Ed. Bologna*, 1978.

[32] E. Giusti. *Minimal surfaces and functions of bounded variation.* Birkhäuser, 1984.

[33] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle, and J. Zerubia. Hierarchical markov random field models applied to image analysis : a review. In *Proc. SPIE Conf. on neural, morphological, stochastics methods in image and signal, San Diego, Etats-Unis*, Juillet 1995.

[34] P.-J. Green. Bayesian reconstruction from emission tomography data using modified EM algorithm. *IEEE Trans. on Medical Imaging*, 9(1):84–93, 1990.

[35] T. Hebert and R. Leahy. A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, 8(2):194–202, 1989.

[36] F. Heitz, P. Pérez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP: Image Understanding*, 59(1):125–134, 1994.

[37] A. Jain and F. Farrokhnia. Usupervised texture segmentation using Gabor filters. *Pattern recognition*, 24(12):1167–1186, 1991.

[38] Z. Kato. *Multiresolution Markovian modeling for computer vision. Application to SPOT image segmentation.* PhD thesis, Université de Nice-Sophia Antipolis, France, 1994.

[39] Z. Kato, J. Zerubia, and M. Berthod. Unsupervised image classification using markovian models. *Pattern Recognition*, 32(4):591–604, 1999.

[40] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi Jr. Conformal curvature flows : from phase transitions to active vision. *Arch. Rational Mech. Anal.*, 134:275–301, 1996.

[41] J. Kittler. Mathematical methods of feature selection in pattern recognition. *Int. Journal Man-Machine Studies*, 7:609–637, 1975.

[42] E. Kreyszig. *Differential geometry.* University of Toronto Press, Toronto, Canada, 1959.

[43] R. Krishnapuram and J. M. Keller. *Fuzzy and possibilistic clustering methods for computer vision*, volume 12. SPIE institute series, in neural and fuzzy systems, 1994.

[44] J.-M. Laferté. *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques: application à la fusion de données multirésolutions.* PhD thesis, Université de Rennes I, France, 1996.

[45] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:799–813, Août 1989.

[46] A. Lorette. *Analyse de texture par méthodes Markoviennes et par morphologie mathématique: application a l'analyse des zones urbaines sur des images satellitaires.* PhD thesis, Université de Nice-Sophia Antipolis, France, 1999.

[47] A. Lorette, X. Descombes, and J. Zerubia. Texture analysis through a Markovian modelling and fuzzy classification: application to urban area extraction from satellite images. *International Journal of Computer Vision*, 36(3):221–236, 2000.

[48] J. MacQueen. Some methods of classification and analysis of multivariate observations. In *Proc. 5^{th} Berkeley Symposium on math. stat. and prob.*, Univ. of California Press, Berkeley, Etat-Unis, 1967.

[49] B. Manjunath and R. Chellappa. Unsupervised texture segmentation using Markov random fields models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:478–482, Mai 1991.

[50] R. March. Visual reconstruction using variational methods. *Image and Vision Computing*, 10:30–38, 1992.

[51] R. March and M. Dozio. A variational method for the recovery of smooth boundaries. *Image and Vision Computing*, 15:705–712, 1997.

[52] G. Dal Maso. *Introduction to Γ-convergence.* Birkhäuser, 1992.

[53] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Arch. Rational Mech. Anal.*, 98:123–142, 1987.

[54] J.-M. Morel and S. Solimini. *Variational methods in image segmentation.* Birkhäuser, 1995.

[55] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, Etats-Unis, 1985.

[56] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed : algorithms based on the Hamilton-Jacobi formulation. *J. of Computational Physics*, 79:12–49, 1988.

[57] N. Paragios. *Geodesic active regions and level set methods: contributions and applications in artificial vision*. PhD thesis, Université de Nice Sophia Antipolis, France, Janvier 2000.

[58] T. Pavlidis and Y.-T. Liow. Integrating region growing and edge detection. In *Proc. of IEEE CVPR*, 1988.

[59] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[60] P. Pérez. *Champs markoviens et analyse multirésolution: application à l'analyse du mouvement*. PhD thesis, Université de Rennes I, France, 1993.

[61] P. Pérez and F. Heitz. Restriction of a markov random field on a graph and multiresolution statistical image modeling. *IEEE Trans. Information Theory*, 42(1):180–190, 1996.

[62] P. Rostaing, J.-N. Provost, and C. Collet. Unsupervised multispectral image segmentation using generalized Gaussian noise model. In *Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, York, G.B.*, Juillet 1999.

[63] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based removal algorithm. *Physica D*, 60:259–268, 1992.

[64] C. Samson. *Contribution à la classification d'images satellitaires par approche variationnelle et équations aux dérivées partielles*. PhD thesis, Université de Nice-Sophia Antipolis, France, Septembre 2000.

[65] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A level set model for image classification. In *Proc. of Scale-Space Theories in Computer Vision, Corfu, Grèce*, Septembre 1999.

[66] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. Multiphase evolution and variational image classification. Rapport de Recherche INRIA RR-3662, (http://www.inria.fr/RRRT/RR-3662.html), Avril 1999.

[67] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A variational model for image classification and restoration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(5):460–472, Mai 2000.

[68] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. Une approche variationnelle pour la classification d'images par re'gions actives. In *Proc. RFIA, Paris, France*, Février 2000.

[69] G. Sapiro and D.L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. on Image Processing*, 5(11):1582–1586, Novembre 1996.

[70] P. Sternberg. Vector-valued local minimizers of nonconvex variational problems. *J. of Mathematics*, 21:799–807, 1991.

[71] P. Sternberg and W.P. Zeimer. Local minimisers of a three-phase partition problem with triple junctions. *Proc. of the Royal Society of Edinburgh*, 124(A):1059–1073, 1994.

[72] M. Sussman, P. Smereka, and S. Osher. A level set approach for computing solutions to incompressible two-phase flow. *J. of Computational Physics*, 114:146–159, 1994.

[73] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Variational approach for edge-preserving regularization using coupled PDE's. *IEEE Trans. on Image Processing*, 7(3):387–397, 1998.

[74] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems.* Winston and Wiley, 1977.

[75] V. Torre and T. Poggio. On edge detection. *IEEE Trans. on Pattern analysis and machine intelligence*, 8(3):147–163, 1986.

[76] E. Volden, G. Giraudon, and M. Berthod. Information in Markov random fields and image redundancy. In *Lecture Notes in Computer Science: information theory and applications*, Septembre 1996.

[77] E. Volden, G. Giraudon, and M. Berthod. La redondance d'images: étude et application. In *Proc. RFIA, Rennes, France*, Janvier 1996.

[78] R. Whitaker and G. Gerig. *Vector-valued diffusion.* computational imaging and vision: geometry driven diffusion in computer vision, Kluwer, 1994.

[79] A. Winter. *Entropie et représentations multi-échelles pour l'interprétation automatique d'images satellitaires et aériennes.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, France, 1997.

[80] F. Wu. The Potts model. *Reviews of modern physics*, 54(1):235–268, 1982.

[81] A. Yezzi, Jr., A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Proc. of ICCV, Corfu, Grèce*, pages 898–903, Septembre 1999.

[82] L. A. Zadeh. Fuzzy sets. *Information and control*, 8:338–353, 1965.

[83] S. Di Zenzo. A note on the gradient of a multi-image. *CVGIP*, 33:116–125, 1986.

[84] H-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *J. of Computational Physics*, 127:179–195, 1996.

[85] S. C. Zhu and A. Yuille. Integrating region growing and edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399