

Modèles de Markov en traitements d'images Markov models in image processing

Wojciech Pieczynski
Institut National des Télécommunications
Département CITI, CNRS UMR 5157
9, rue Charles Fourier, 91000 Evry

Résumé

L'objet de l'article est de présenter divers aspects des traitements statistiques des images utilisant des modèles de Markov. En choisissant pour cadre la segmentation statistique nous rappelons brièvement la nature et l'intérêt des traitements probabilistes et présentons les modèles de Markov cachés classiques : champs, chaînes, et arbres. Les méthodes bayésiennes de segmentation sont décrites, ainsi que les grandes familles des méthodes d'apprentissage. Quelques modèles ou méthodes de traitements plus récents comme les modèles de Markov Couple et Triplet, la fusion de Dempster-Shafer dans le contexte markovien, ou l'estimation des mélanges généralisés sont également présentés. Nous terminons par une liste non exhaustive des divers prolongements des méthodes et modèles vers l'imagerie multidimensionnelle, multisenseurs, multirésolution. Des liens avec les modèles graphiques généraux sont également brièvement décrits.

Mots clés : champs de Markov, chaînes de Markov, arbres de Markov, segmentation bayésienne d'images, Espérance-Maximisation, Estimation conditionnelle itérative, segmentation floeu, fusion Dempster-Shafer, theorie de l'évidence, Markov couple, Markov triplet.

Abstract

The aim of this paper is to present some aspects of Markov model based statistical image processing. After a brief review of statistical processing in image segmentation, classical Markov models (fields, chains, and trees) used in image processing are developed. Bayesian methods of segmentation are then described and different general parameter estimation methods are presented. More recent models and processing techniques, such as Pairwise and Triplet Markov models, Dempster-Shafer fusion in a Markov context, and generalized mixture estimation, are then discussed. We conclude with a nonexhaustive description of candidate extensions to multidimensional, multisensor, and multiresolution imagery. Connections with general graphical models are also highlighted.

Key words : Markov fields, Markov chains, Markov trees, Bayesian image segmentation, Expectation-Maximization, Iterative Conditional Estimation, Fuzzy segmentation, Dempster-Shafer fusion, theory of evidence, pairwise Markov models, triplet Markov models.

1. Introduction

L'objet de cet article est de discuter l'intérêt que peuvent présenter les modélisations probabilistes et les traitements statistiques en imagerie, avec un accent particulier mis sur l'utilisation des modèles Markoviens. Le cadre illustratif choisi est celui de la segmentation. Ce problème est simple dans sa formulation et important dans les applications, ce qui nous semble bien adapté pour exposer la nature et l'intérêt profonds des méthodes de traitement probabilistes en imagerie.

La segmentation des images figure donc parmi les problèmes les plus importants en imagerie. Dans certaines situations le résultat d'une segmentation fournit directement la solution à un problème posé ; dans d'autres, il sert de point de départ à d'autres traitements. L'essor des méthodes markoviennes date de l'introduction des champs de Markov cachés [11, 37, 63], qui ont permis de contourner, de manière rigoureuse et particulièrement élégante, les problèmes liés aux temps de calcul prohibitifs, ces derniers étant simplement dus au nombre généralement important des pixels dans une image. L'intérêt de telles méthodes est de pouvoir prendre des « décisions » (classifications, estimations, filtrages, restaurations, ...) localement à partir de toute l'information disponible dans l'image. Le succès de ce type de méthode est dû à leurs aptitudes à produire, lorsque les divers bruits présents dans l'image considérée sont importants et lorsque les données correspondent bien au modèle utilisé, des résultats spectaculaires, dépassant parfois les capacités de l'œil humain.

La segmentation statistique des images trouve de nombreuses applications dans les domaines de traitement d'images les plus divers. Il existe désormais une riche bibliographie, qui pourra être, en particulier, consultée dans les ouvrages généraux comme [18, 111]. En particulier, en imagerie satellitaire ou aérienne de telles méthodes ont été utilisées à diverses fins comme, à titre d'exemple, la classification des zones ou l'extraction des structures, l'interprétation des scènes, ... [8, 10, 13, 15, 19, 22, 24, 27, 28, 30, 33, 35, 42, 47, 49, 55, 56, 58, 59, 62, 65, 74, 78, 87, 90, 91, 98, 106, 107]. En imagerie sonar, où les images obtenues sont souvent relativement bruitées, les méthodes statistiques de traitement fondées sur des modèles de Markov sont également bien adaptées [66, 67, 69, 113] et il en est de même dans de nombreuses situations se présentant en imagerie médicale [20, 26, 46, 68, 70, 75, 94, 95, 112]. Citons également des exemples d'application en vision par ordinateur [36, 39, 45, 73], ou encore imagerie industrielle [12, 48].

L'organisation de l'article est la suivante :

La nature de la classification bayésienne et son intérêt sont présentés de manière intuitive, dans un cadre général et en dehors des modèles de Markov, dans la section suivante. La section 3 contient une description des modèles de Markov cachés classiques : champs, chaînes, et arbres, ainsi que celle des méthodes bayésiennes associées. La section 4 est consacrée à un modèle récent dit champs de Markov Couple. L'importante question de l'estimation des paramètres, permettant de rendre les traitements automatiques, est abordée dans la section 5 avec la description, en particulier, de trois familles générales de démarches (Espérance-Maximisation, Estimation Conditionnelle Itérative, et les méthodes « pleinement bayésiennes »). Diverses extensions aux modèles plus complexes (multisenseur, multirésolution, imagerie tridimensionnelle, séquences d'images, modèles graphiques généraux, fusion de Dempster-Shafer, modèles de Markov flous,...) sont discutées dans la section 6. La dernière section, section 7, contient les conclusions et quelques perspectives.

2. Modèle probabiliste et classification bayésienne

Les modélisations probabilistes et les traitements statistiques peuvent s'avérer d'une remarquable efficacité dans le traitement de certains problèmes se posant en imagerie. De manière très générale, on peut envisager de faire appel à des méthodes probabilistes de traitement lorsque l'on se trouve dans une situation réunissant les deux conditions suivantes : (i) on souhaite déterminer des caractéristiques de l'image, non observables directement et représentées par un élément x d'un espace X , à partir des caractéristiques de l'image observables et représentées par un élément y d'un espace Y ; (ii) il n'existe manifestement pas de liens déterministes entre les y et les x (en d'autres termes, plusieurs x possibles peuvent correspondre à un y observé). Le deuxième point semble interdire définitivement tout traitement sérieux du problème ; cependant, un traitement statistique est souvent possible. De plus, ce type de traitement peut bénéficier de trois qualités majeures : généralité, optimalité et souplesse.

Précisons les sens donnés à ces trois qualités dans le contexte considéré.

Il n'existe donc pas de liens déterministes entre les y et les x ; cependant, l'observation de y doit apporter une certaine information sur x . Dans le contexte qui nous intéresse, cette information est donnée par une loi de probabilité sur X et peut être appréhendée, grâce à la loi des grands nombres, de la façon suivante : si la situation se reproduit un certain nombre de fois (à y donné), la probabilité $p(A|y)$ (qui dépend donc de y) d'un ensemble $A \subset X$ est la proportion des x se trouvant dans A . La « généralité » des traitements statistiques est alors assurée par la richesse des possibilités de définir les lois de probabilités, conditionnelles à y , sur X . A y fixé, cette « généralité » se traduit par la possibilité de modélisation des situations extrêmement diverses, allant du lien déterministe à l'indépendance totale entre y et x . En effet, dans le premier cas la probabilité $p(\cdot|y)$ est une masse de Dirac et dans le deuxième cas cette probabilité ne dépend pas de y . Il est ainsi possible de modéliser des liens « plus au moins » forts entre y et x . De plus, la situation où les liens entre y et x sont déterministes apparaît comme un cas particulier des modélisations probabilistes.

Concernant l'optimalité considérons une fonction $L : X^2 \rightarrow R^+$ modélisant la gravité des erreurs : $L(x^1, x^2)$ représente le coût de l'erreur « on pense que x est x^1 , alors que le vrai x est x^2 ». Notons que la fonction L , qui est appelée « fonction de perte », ou encore « fonction de coût », est indépendante des modélisations probabilistes éventuellement retenues. Supposons que l'on dispose d'une règle de décision $d : Y \rightarrow X$, qui associe donc à chaque y observé un x inobservable, et que l'on souhaite évaluer sa qualité. Connaissant la probabilité $p(\cdot|y)$, on peut calculer le « risque » défini par

$$R(L, d, y) = \int_X L(d(y), x)p(x|y)dx \quad (2.1)$$

Observons que le « risque » est une quantité très « parlante » pour un praticien, même s'il n'a aucune connaissance particulière de la théorie des probabilités. En effet, $R(L, d, y)$ désigne la « perte moyenne » : si l'on utilise la règle d un nombre n suffisamment grand de fois, la perte globale subie est le produit de n par $R(L, d, y)$. A y observé, on peut alors rechercher $d_B^L(y) \in X$ qui minimise l'intégrale (2.1) ($d_B^L(y)$ est dite « décision

Bayésienne » liée à L), et qui minimisera la perte globale subie après n décisions. Il reste à préciser une dernière généralisation. Lorsque l'on est confronté à n expériences, les n observations y peuvent ne pas être toutes égales et évoluer de manière aléatoire en suivant une loi de probabilité définie sur Y . Précisons dès à présent que cette nouvelle généralisation ne modifie pas la solution optimale d_B^L , qui reste celle qui minimise la perte globale. Cependant, préciser la manière dont on prend en compte l'aspect aléatoire de y nous permettra de terminer la modélisation probabiliste du problème. Les quantités x et y étant aléatoires, on peut les considérer comme réalisations des variables aléatoires X et Y . La loi $p(x, y)$ du couple (X, Y) est donnée par la loi $p(y)$ de Y et les lois $p(x|y)$ de X conditionnelle à Y . Le risque moyen associé à L et d est alors défini par

$$R(L, d) = E[L(d(Y), X)] \quad (2.2)$$

et la décision Bayésienne associée à L est la décision qui minimise ce risque moyen :

$$E[L(d_B^L(Y), X)] = \min_d E[L(d(Y), X)] \quad (2.3)$$

Finalement, la qualité d'optimalité peut se résumer de la manière suivante. Si la loi $p(x, y)$ du couple (X, Y) est connue, la règle de décision d_B^L qui consiste à minimiser, pour chaque y observé, le risque défini par (2.1), est celle parmi toutes les règles possibles, d'origine probabiliste ou pas, qui minimise la perte globale subie lorsque l'on traite un nombre suffisamment grand des y . De plus, le calcul de $E[L(d_B^L(Y), X)]$ nous donne le montant de la perte minimale.

Enfin, les méthodes probabilistes de traitement présentent une « souplesse » dans le sens où l'optimalité discutée ci-dessus est adaptable à des préoccupations particulières. En effet, devant une réalité « objective » donnée, modélisée par la loi $p(x, y)$ du couple (X, Y) , deux clients peuvent avoir des préoccupations différentes – voire contradictoires – modélisées par deux fonctions de perte différentes L^1, L^2 , qui donnent le montant de l'argent perdu, à long terme, par chacun des clients. Ces fonctions vont définir deux règles de décisions différentes d^1, d^2 . Ainsi d^1 est optimale pour le premier client, qui perdrait plus d'argent en appliquant d^2 , et d^2 est optimale pour le deuxième client, qui perdrait plus d'argent en appliquant d^1 .

Exemple

Considérons le problème d'établissement d'une carte d'une région comportant deux classes « eau », notée E , et « forêt », notée F . Supposons que l'on souhaite travailler « pixel par pixel ». Cela signifie que sur chaque pixel on observe un niveau de gris y et que l'on doit lui attribuer un x dans l'ensemble $X = \{E, F\}$. On se trouve dans un contexte probabiliste si un niveau de gris donné peut aussi bien correspondre à de l'eau qu'à de la forêt, ce qui arrive fréquemment dans des images fortement « bruitées », fournies par des capteurs radar par exemple. La loi $p(x, y)$ du couple (X, Y) peut alors être donnée par la loi $p(x)$ de X , qui est une probabilité sur $X = \{E, F\}$ et désigne simplement la proportion de deux classes dans la région considérée, et par les deux lois $p(y|E), p(y|F)$ de Y conditionnelles à X .

Notons que ces deux dernières lois modélisent simultanément la « variabilité naturelle » de deux classes et diverses dégradations dues à l'acquisition, la transmission, et la visualisation des données. Supposons que deux « clients » souhaitent se porter acquéreur de logiciels de segmentation d'images satellite comportant de l'eau et de la forêt. Le premier client, pour lequel la surévaluation de la classe « eau » est plus grave que celle de la classe « forêt », perd 1 euro quand un pixel « eau » est classé à tort comme étant « forêt », et 5 euros quand un pixel « forêt » est classé à tort comme étant « eau ». Un deuxième client, pour qui la surévaluation de la classe « forêt » est plus grave que celle de la classe « eau », perd 1 euro quand un pixel « forêt » est classé à tort comme étant « eau », et 10 euros quand un pixel « eau » est classé à tort comme étant « forêt ». Nous avons donc $L^1(E, E) = L^1(F, F) = 0$, $L^1(F, E) = 1$, $L^1(E, F) = 5$, et $L^2(E, E) = L^2(F, F) = 0$, $L^2(F, E) = 10$, $L^2(E, F) = 1$. Les deux minimisations de (2.1), qui s'écrit ici $L(d(y), E)p(E|y) + L(d(y), F)p(F|y)$, donnent $d^1(y) = E$ si $p(E|y) \geq 5p(F|y)$, $d^1(y) = F$ si $p(E|y) \leq 5p(F|y)$, et $d^2(y) = E$ si $10p(E|y) \geq p(F|y)$, $d^2(y) = F$ si $10p(E|y) \leq p(F|y)$ (rappelons que $p(E|y)$ et $p(F|y)$ sont calculées à partir de $p(x)$, $p(y|E)$, et $p(y|F)$ par

$$p(E|y) = \frac{p(E)p(y|E)}{p(E)p(y|E) + p(F)p(y|F)} \quad (2.4)$$

$$p(F|y) = \frac{p(F)p(y|F)}{p(E)p(y|E) + p(F)p(y|F)}.$$

Pour un nombre de pixels suffisamment grand – ce qui est pratiquement toujours le cas en imagerie – la règle d^1 assure ainsi la perte minimale au premier client et la règle d^2 assure la perte minimale au deuxième client. Si le modèle probabiliste correspond parfaitement à la réalité, ces règles sont optimales parmi toutes les règles opérant « pixel par pixel », d'origine probabiliste ou pas.

3. Modèles de Markov cachés

3.1 Méthodes locales contextuelles

L'utilisation des décisions Bayésiennes précisées dans le paragraphe précédent peut être mise en défaut lorsque l'ensemble X est de cardinal important. En reprenant la situation de l'exemple ci-dessus, notons S l'ensemble des pixels et considérons les « champs aléatoires » $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$ (notons que X et Y de l'exemple 2.1 deviennent respectivement X_s et Y_s). Nous avons vu qu'il était possible d'établir une carte de manière optimale en procédant « pixel par pixel ». Pour chaque s dans S , les variables X_s et Y_s sont à valeurs dans $X = \{E, F\}$ et $Y = R$ respectivement. La réalisation $X_s = x_s$ peut être estimée de manière optimale à partir de la réalisation $Y_s = y_s$; cependant, il est légitime d'imaginer que d'autres $Y_t = y_t$ contiennent de l'information utile à la classification du pixel s . En d'autres termes, pour attribuer une classe à s , regarder les niveaux de gris sur s et sur son voisinage devrait être plus efficace que de le regarder uniquement sur s . On

peut effectivement montrer que lorsque l'on utilise un voisinage V_s de s , l'utilisation de $Y_{V_s} = (Y_t)_{t \in V_s}$ à la place de Y_s permet d'améliorer l'efficacité des stratégies Bayésiennes : pour toute fonction de perte L , la perte moyenne minimale $E[L(d_B^L(Y_{V_s}), X_s)]$ est inférieure ou égale à la perte moyenne minimale $E[L(d_B^L(Y_s), X_s)]$. De telles méthodes de traitement, dites « locales contextuelles », ont été utilisées et donnent dans certaines situations des résultats de qualité suffisante ; cependant, on se heurte rapidement à des problèmes de calcul lorsque la taille de V croît. Pour deux classes, il est ainsi difficile de prendre en compte des voisinages de taille supérieure à huit. Le nombre de pixels étant couramment de l'ordre de cinquante mille, on est en droit d'imaginer qu'une partie non négligeable de l'information disponible ne peut être exploitée par les méthodes locales.

3. 2 Champs de Markov cachés

L'introduction des champs de Markov cachés [11, 37, 63] a permis de contourner, de manière particulièrement élégante, les problèmes de calcul mentionnés dans le paragraphe précédent. Afin de fixer les idées, nous considérerons le problème de la segmentation statistique d'image. Pour S l'ensemble des pixels, on considère deux ensembles de variables aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, appelés « champs aléatoires ». Chaque X_s prend ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque Y_s prend ses valeurs dans R . Nous considérons ici la variable cachée à valeurs discrètes car nous avons choisi la segmentation d'image comme cadre illustratif ; cependant, d'autres problèmes comme déconvolution et divers filtrages [14, 48, 72, 118], auraient pu être choisis, auquel cas la variable cachée pourrait éventuellement être à valeurs continues. En admettant que l'ensemble S contient N pixels, nous avons donc $X = \Omega^N$ et $Y = R^N$, ce qui rend les calculs explicites des décisions Bayésiennes du paragraphe précédent impossibles pour les tailles courantes des images. Comment manier les processus $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, qui peuvent comporter, chacun, plus d'un million de composantes ? Cela devient possible en adoptant pour la loi du couple (X, Y) le modèle dit « champ de Markov caché ». La loi du couple (X, Y) étant donnée par $p(x)$, la loi de X , et $p(y|x)$, la loi de Y conditionnelle à X , explicitons ces deux lois dans le cadre d'un tel modèle. Le champ aléatoire X est un champ de Markov par rapport à un système de voisinage $V = (V_s)_{s \in S}$, avec la forme des V_s indépendante de s , si sa loi est de la forme

$$p(x) = \gamma \exp \left[- \sum_{c \in C} \varphi_c(x_c) \right] \quad (3.1)$$

avec C l'ensemble des cliques, une clique étant soit un singleton, soit un ensemble des pixels mutuellement voisins au sens de V (la fonction $U(x) = \sum_{c \in C} \varphi_c(x_c)$ est dite « énergie »). Les fonctions φ_c , dites « fonctions de potentiel », traduisent ainsi une connaissance a priori d'une prédominance de certaines configuration part rapport à d'autres. A titre d'exemple, on utilise fréquemment les voisinages composés de quatre plus proches voisins, où les cliques sont les singletons, les ensembles contenant deux pixels voisins horizontalement, et les ensembles contenant deux pixels voisins verticalement. Notons qu'un champ X dont la loi est de forme vérifiant (3.1) vérifie l'hypothèse de

markovianité suivante : pour tout $s \in S$, la loi de X_s conditionnelle à $(X_t)_{t \neq s}$ est égale à sa loi conditionnelle à $(X_t)_{t \in V_s}$: la loi de X_s conditionnelle à toutes les autres réalisations ne dépend que des réalisations sur le voisinage. En d'autres termes, X_s et $(X_t)_{t \notin V_s}$ sont indépendants conditionnellement à $(X_t)_{t \in V_s}$. Cependant, notons que cette indépendance n'a plus lieu lorsque l'on supprime le conditionnement ; dans un champ de Markov, deux variables X_s, X_t sont toujours dépendantes, quelle que soit la position des pixels s, t dans S .

La formule (3.1) ne permet pas le calcul de $p(x)$ car la constante γ est inconnue ; cependant, et c'est cette propriété qui est fondamentale pour la suite, la loi de X_s conditionnelle à la restriction de X au voisinage V_s est calculable par

$$p(x_s | x_{V_s}) = \frac{\exp[-\sum_{c \in C, s \in c} \varphi_c(x_s, x_{\bar{c}})]}{\sum_{\omega \in \Omega} \exp[-\sum_{c \in C, s \in c} \varphi_c(x_s = \omega, x_{\bar{c}})]} \quad (3.2)$$

où $\bar{c} = c - \{s\}$ (la somme $\sum_{c \in C, s \in c} \varphi_c(x_s, x_{\bar{c}})$ est prise sur les cliques $c \in C$ telles que $s \in c$, le

pixel s étant fixé). Pour des voisinages de taille raisonnable (quatre, huit, voire douze voisins) le nombre des cliques contenant un pixel donné reste raisonnable et le calcul explicite de (3.2) est possible. Ce dernier permet alors de proposer des méthodes itératives de simulation de X selon sa loi Markovienne (3.1). Parmi les différentes méthodes, mentionnons l'échantillonneur de Gibbs, dont le déroulement est le suivant. Considérons un $x^0 \in \Omega^N$ quelconque et, en faisant parcourir à s l'ensemble S , dans un balayage « ligne par ligne » par exemple, faisons, pour chaque s , un tirage aléatoire dans Ω selon la loi (3.2). Dans ces tirages, on tient compte, pour chaque s , des nouvelles valeurs de certains pixels de son voisinage obtenues dans les tirages précédents, éventuellement différentes de celles qui composent x^0 . Notons x^1 l'élément de Ω^N (qui est une image de classes, ou une carte) obtenu après le balayage complet de S et recommençons, avec x^1 à la place de x^0 , et ainsi de suite On obtient une suite de réalisations $X^1 = x^1, X^2 = x^2, \dots$ dun processus aléatoire X^1, X^2, \dots qui est une chaîne de Markov à valeurs dans Ω^N . On montre alors, en utilisant des propriétés élémentaires des chaînes de Markov, que le processus X^1, X^2, \dots converge en loi vers (3.1).

Il est ainsi possible de simuler, et c'est ce fait là qui est à l'origine des possibilités du contournement des problèmes de calcul mentionnés dans le paragraphe précédent, des réalisations de X selon la loi donnée par (3.1).

Remarquons que le calcul de la constante γ est généralement impossible à cause du nombre trop important des réalisations possibles de X ; cependant, il existe des techniques de son approximation déterministes [6] ou stochastiques [88]. La connaissance de la valeur approchée de γ peut alors être utile dans la recherche des solutions d'un certain nombre de problèmes d'estimation des paramètres ou de tests [88], ces derniers servant à un choix, dans une famille des modèles admissibles, de celui le mieux adapté aux données en présence.

L'échantillonneur de Gibbs est un cas particulier des démarches générales connues sous l'appellation « Méthodes de Monte Carlo par Chaînes de Markov » (MCMC [92, 111]), dont la problématique est la suivante. On souhaite simuler une loi $p(x)$ trop compliquée

pour que l'on puisse le faire directement. On construit alors des probabilités de transitions $p(x^{i+1}|x^i)$, suffisamment simples pour permettre leurs simulations, définissant une chaîne de Markov dont les lois marginales tendent vers $p(x)$. Cette construction permet de faire des simulations de $p(x)$ « à la limite », en faisant un nombre suffisant de simulations selon les transitions. Notons qu'en dehors de l'échantillonneur de Gibbs il existe une autre technique, dite de « Hasting-Metropolis », faisant parties des méthodes MCMC et couramment utilisée dans le contexte des champs de Markov. La loi $p(x)$ étant définie par (3.1), il reste à définir les lois $p(y|x)$. Leur forme la plus simple, couramment utilisée, est (on suppose les f_{x_s} , qui sont les densités des lois de Y_s conditionnelles à $X_s = x_s$, strictement positives) :

$$p(y|x) = \prod_{s \in S} f_{x_s}(y_s) = \exp\left[\sum_{s \in S} \text{Log}(f_{x_s}(y_s))\right] \quad (3.3)$$

La loi de (X, Y) est alors

$$p(x, y) = p(x)p(y|x) = \gamma \exp[-U(x, y)] = \gamma \exp\left[-\sum_{c \in C} \varphi_c(x_c) + \sum_{s \in S} \text{Log}(f_{x_s}(y_s))\right] \quad (3.4)$$

et la loi de X a posteriori $p(x|y)$, qui est proportionnelle, à une constante dépendant de y près, à $p(x, y) = p(x)p(y|x)$, s'écrit

$$p(x|y) = \gamma(y) \exp[-U(x, y)] = \gamma(y) \exp\left[-\sum_{c \in C} \varphi_c(x_c) + \sum_{s \in S} \text{Log}(f_{x_s}(y_s))\right] \quad (3.5)$$

Nous constatons que (3.5) diffère de (3.1) par la présence d'une somme sur les pixels ; les singletons étant toujours des cliques, la structure Markovienne donnée par (3.1) est ainsi préservée (notons que $\gamma(y) = \frac{1}{\sum_{x \in \Omega^N} p(x, y)}$, avec $p(x, y)$ donnée par (3.4)).

3.3 Segmentation bayésienne avec les champs de markov cachés

Il est ainsi possible de simuler des réalisations de X selon sa loi a posteriori (3.5) en utilisant les lois conditionnelles (3.2), avec $\varphi_c(x_s, x_{\bar{c}}) - \text{Log}(f_{x_s}(y_s))$ à la place de $\varphi_c(x_s, x_{\bar{c}})$, et l'échantillonneur de Gibbs, ce qui permet la mise en place des différentes stratégies Bayésiennes de segmentation. Parmi ces dernières, les stratégies « Maximum des Marginales a Posteriori » (MPM [63]) et le « Maximum a Posteriori » (MAP [37]), qui correspondent aux fonctions de perte $L_1(x^1, x^2) = \sum_{s \in S} 1_{[x_s^1 \neq x_s^2]}$ et $L_2(x^1, x^2) = 1_{[x^1 \neq x^2]}$, sont le plus fréquemment utilisées. Notons la différence intuitive entre ces deux fonctions de perte : la première mesure la perte par la proportion des pixels mal classés, alors que la deuxième pénalise de la même manière toutes les images de classes différentes de la vraie image, même si la différence porte sur un seul pixel. A priori, la fonction L_2 peut sembler quelque peu brutale car si la différence porte sur un seul pixel une telle erreur est généralement négligeable ; ainsi la fonction L_1 peut apparaître comme plus souple et

mieux adaptée aux besoins classiques. Cependant, cette dernière n'est pas exempte de critiques. Supposons que la vraie image x^2 soit un damier « noir et blanc » et l'image estimée x^1 est obtenue par le décalage d'un pixel sur la droite (par exemple) : tous les pixels sont alors différents et donc $L_2(x^1, x^2)$ est maximale, ce qui modélise mal le fait que les deux damiers x^1 et x^2 apparaîtront, à l'œil, comme identiques. Bien entendu, toute autre fonction de perte pourrait être utilisée ; cependant, cet aspect des méthodes bayésiennes de segmentation ne semble pas susciter un intérêt des utilisateurs. Cet état de choses est probablement dû au fait, relativement curieux, que l'utilisation de L_1 mène à des résultats souvent très comparables à ceux obtenus en utilisant L_2 .

La stratégie MPM s'écrit :

$$\hat{x}^{MPM} = d_1(y), \text{ avec } \hat{x}_s^{MPM} = \arg \max_{\omega \in \Omega} p(x_s = \omega | y) \quad (3.6)$$

Le calcul explicite des marginales a posteriori $p(x_s | y)$ figurant dans (3.6) n'est pas possible ; cependant, en désignant par x^1, \dots, x^n les images simulées selon $p(x | y)$, elle peuvent être aisément estimées par

$$\hat{p}(x_s = \omega | y) = \frac{1_{[x_s^1 = \omega]} + \dots + 1_{[x_s^n = \omega]}}{n} \quad (3.7)$$

ce qui permet la mise en place de MPM. Nous voyons que (3.6) répond aux interrogations du sous-paragraphe précédent et permet d'effectuer la recherche de chaque \hat{x}_s à partir de toute l'information disponible $Y = y$. Le calcul explicite des marginales a posteriori $p(x_s | y_s)$ (voir (2.4) de l'exemple 2.2) a été remplacé par une estimation, utilisant une technique de type MCMC, des $p(x_s | y)$.

La stratégie MAP s'écrit :

$$\hat{x}^{MAP} = d_2(y) = \arg \max_x p(x | y) \quad (3.8)$$

et peut être approchée soit par le très élégant algorithme de « recuit simulé » [37], soit par des algorithmes rapides comme l'algorithme « Iterated Conditional Mode » (ICM [11]). Dans le premier, on introduit dans l'énergie U de (3.5) la « température » $T > 0$ en posant $U_T(x, y) = U(x, y)/T$. En choisissant une suite (T_n) décroissante et tendant vers 0, notons x_n une réalisation de X simulée selon (3.5), avec U_{T_n} à la place de U . On montre alors que si la décroissance de (T_n) est suffisamment lente, la suite (x_n) tend vers \hat{x} donné par (3.8) (lorsque plusieurs \hat{x} vérifient (3.8), $p(x_n | y)$ tend vers le maximum absolu sans que la suite (x_n) converge nécessairement vers un élément donné parmi ces \hat{x}). Dans les cas des modèles complexes, l'algorithme du recuit simulé peut être éventuellement accéléré par des techniques de parallélisation [3, 52, 105, 116]. L'algorithme ICM ressemble à l'échantillonneur de Gibbs : on effectue des balayages de l'ensemble de s pixels et pour chaque $s \in S$ on remplace la valeur courante \hat{x}_s par une nouvelle valeur \hat{x}_s' (avec, éventuellement, $\hat{x}_s = \hat{x}_s'$) obtenue par la maximisation de la probabilité $p(x_s | \hat{x}_{V_s}, y)$ (dans

l'échantillonneur de Gibbs, la nouvelle valeur est obtenue par le tirage aléatoire selon cette même probabilité). On obtient ainsi une suite déterministe $\hat{x}^1, \dots, \hat{x}^n$ telle que la suite $p(\hat{x}^1|y), \dots, p(\hat{x}^n|y)$ est croissante ; cependant, sa convergence vers \hat{x}^{MAP} donné par (3.8) n'est pas assurée.

Exemple

Considérons le cas simple d'un champ de Markov caché (modèle d'Ising) relativement à quatre plus proches voisins, à k classes $\Omega = \{\omega_1, \dots, \omega_k\}$. Les fonctions potentiel φ_c dans (3.1) sont nulles sur les singletons et s'écrivent sur les cliques d'ordre deux (formées par les couples (s, t) de voisins « horizontaux » ou « verticaux ») $\varphi_{(s,t)}(x_s, x_t) = \alpha(1 - 2\delta(x_s, x_t))$, avec $\delta(x_s, x_t) = 1$ si $x_s = x_t$ et $\delta(x_s, x_t) = -1$ si $x_s \neq x_t$. La formule (3.2), qui permet les tirages selon les lois conditionnelles et donc l'utilisation de l'échantillonneur de Gibbs, prend alors une forme simple suivante : en notant $n(\omega_i)$ le nombre des pixels $t \in V_s$ (variant de 0 à 4) tels que $x_t = \omega_i$, nous avons $p(x_s = \omega_i | x_{V_s}) = \exp \alpha [2n(\omega_i) - 4] / \sum_{1 \leq j \leq k} \exp \alpha [2n(\omega_j) - 4]$. En notant f_1, \dots, f_k les densités des lois de Y_s conditionnelles à $X_s = \omega_1, \dots, X_s = \omega_k$, la formule analogue correspondant à la loi a posteriori $p(x|y)$ est $p(x_s = \omega_i | x_{V_s}, y) = \exp[\alpha(2n(\omega_i) - 4) + \text{Log} f_i(y_s)] / \sum_{1 \leq j \leq k} \exp[\alpha(2n(\omega_j) - 4) + \text{Log} f_j(y_s)]$.

Notons que la forme simple (3.3) des lois $p(y|x)$ est équivalente aux hypothèses

(H_1) $p(y_s|x) = p(y_s|x_s)$ et

(H_2) les variables (Y_s) sont indépendantes conditionnellement à X .

Ces hypothèses peuvent être affaiblies ; en effet, si l'on remplace (H_1) par $p(y_s|x) = p(y_s|x_{c'})$, avec c' sous-ensemble de S de petite taille et contenant s , la Markovianité de la loi a posteriori (3.5), suffisante pour utiliser l'échantillonneur de Gibbs qui permet la mise en place des décisions Bayésiennes, est préservée. Cependant, l'ensemble des cliques C est enrichi par l'ensemble C' de toutes les « cliques » c' et donc la Markovianité de X a posteriori est relative à un système de voisinage de taille plus importante que sa Markovianité a priori [44]. Notons également que des champs de Markov cachés plus perfectionnés permettent de traiter le problème de détection des bords dans le cas des classes texturées [38].

3. 4 Chaînes de Markov cachées

Les chaînes de Markov cachées sont parmi les modèles de Markov cachés les plus simples et les plus anciens. Supposons que S est ordonné et considérons, comme précédemment, un processus d'intérêt caché $X = (X_1, \dots, X_N)$ et un processus observé $Y = (Y_1, \dots, Y_N)$. Comme précédemment, chaque X_i prend ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque Y_i prend ses valeurs dans R . Le processus X est une chaîne de Markov si pour tout $1 \leq n \leq N - 1$, $p(x_{n+1}|x_1, \dots, x_n) = p(x_{n+1}|x_n)$. La loi de X est alors

donnée par la loi de X_1 et les transitions $p(x_{n+1}|x_n)$. En posant les mêmes hypothèses (H_1) $p(y_i|x) = p(y_i|x_i)$ et (H_2) les variables (Y_i) sont indépendantes conditionnellement à X , nous avons

$$p(x, y) = p(x_1)p(y_1|x_1)\prod_{n=2}^N p(x_n|x_{n-1})p(y_n|x_n) \quad (3.9)$$

A la différence des champs de Markov cachés, où les différentes quantités d'intérêt doivent être estimées via des méthodes de type MCMC, dans le cas de chaînes de Markov cachées ces quantités peuvent être calculées. Plus précisément, posons $\alpha(x_i) = p(x_i, y_1, \dots, y_i)$, $\beta(x_i) = p(y_i, \dots, y_N|x_i)$.

Ces quantités, appelées respectivement « probabilités forward » et « probabilités backward », sont calculables par les récursions suivantes [35] :

$$\alpha(x_1) = p(x_1)p(y_1|x_1), \text{ et}$$

$$\alpha(x_{i+1}) = \sum_{x_i \in \Omega} \alpha(x_i)p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}), \text{ pour } 1 \leq i \leq N-1 ;$$

$$\beta(x_N) = 1, \text{ et}$$

$$\beta(x_i) = \sum_{x_{i+1} \in \Omega} \beta(x_{i+1})p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}), \text{ pour } 1 \leq i \leq N-1.$$

La loi $p(x|y)$ de X a posteriori est alors la loi d'une chaîne de Markov donnée par :

$$\begin{aligned} p(x_1|y) &= \alpha(x_1)\beta(x_1) / \sum_{x_1' \in \Omega} \alpha(x_1')\beta(x_1') \\ p(x_{i+1}|x_i, y) &= \beta(x_{i+1})p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}) / \beta(x_i) \end{aligned} \quad (3.10)$$

De plus, les marginales a posteriori sont données par :

$$p(x_i|y) = \alpha(x_i)\beta(x_i) \sum_{x_i' \in \Omega} \alpha(x_i')\beta(x_i') \quad (3.11)$$

ce qui permet le calcul de la décision bayésienne MPM. Par ailleurs, la solution de la décision bayésienne du MAP est également calculable par l'algorithme de Viterbi [32].

Notons que le calcul récursif des quantités "forward" et "backward" est possible grâce à la présence d'un ordre total dans l'ensemble des indices de deux processus X, Y . Nous retrouverons l'importance de cette présence dans les modèles graphiques plus généraux mentionnés dans la sous-section 6.7.

En résumé, le modèle par chaînes de Markov cachées permet un calcul direct des quantités d'intérêt, alors que celui par champs de Markov cachés fait appel à des simulations itératives, souvent coûteuses en temps calcul, qui permettent d'estimer ces quantités.

Bien entendu, l'utilisation directe des chaînes de Markov en segmentation d'images se heurte à l'absence d'un ordre naturel dans l'ensemble bi-dimensionnel des pixels S . Une des possibilités de transformation de S en une suite consiste en l'emploi des parcours de type Hilbert-Peano [100]. De tels parcours ont été utilisés dans le cadre de la segmentation

des images fixes et animées [9], multisenseur [42], ou encore multirésolution [35], où le caractère fractal de ces parcours peut être exploité. On obtient ainsi des méthodes concurrentes, plus rapides mais s'appuyant sur un modèle moins satisfaisant au plan intuitif, de celles utilisant les champs de Markov cachés. Laquelle de ces deux familles doit-on utiliser dans une situation donnée ? On trouvera dans [97] quelques éléments de réponses dans des situations simples ; il apparaît que la famille des méthodes rapides peut être compétitive. Par ailleurs, en situations complexes où le temps d'exécution des méthodes fondées sur les champs de Markov cachés devient prohibitif, les résultats obtenus avec les chaînes peuvent servir des initialisations, aussi bien au niveau de l'estimation des paramètres qu'au niveau de la segmentation proprement dite, jouant ainsi un rôle d'accélérateur.

3. 5 Arbres de Markov cachés

Soit S un ensemble d'indices et $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$, deux processus aléatoires, chaque X_s prenant ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque Y_s prenant ses valeurs dans R . Soit S^1, \dots, S^n une partition de S représentant différentes « générations ». A chaque $s \in S^i$ correspond $s^+ \subset S^{i+1}$ appelé l'ensemble des « enfants » de s , les ensembles des enfants des éléments de S^i formant une partition de S^{i+1} . Par ailleurs, on suppose que S^1 admet un seul élément s^1 appelé « racine », on note par s^{++} l'ensemble de tous les descendants de s , et par s^- son unique « père ». En unissant chaque pixel à ses enfants par une flèche on obtient ainsi un « graphe orienté », qui sera noté G . Le processus X est un arbre de Markov sur G si sa loi est donnée par :

$$p(x) = p(x_{s^1}) \prod_{s \in S - S^n} \prod_{t \in s^+} p(x_t | x_s) \quad (3.12)$$

En supposant une nouvelle fois les variables (Y_s) indépendantes conditionnellement à X et $p(y_s | x) = p(y_s | x_s)$, la loi $p(x, y)$ du couple (X, Y) s'écrit

$$p(x, y) = p(x_{s^1}) p(y_{s^1} | x_{s^1}) \prod_{s \in S - S^n} \prod_{t \in s^+} p(x_t | x_s) p(y_t | x_t) \quad (3.13)$$

On montre alors que la loi de X a posteriori reste une loi d'arbre de Markov. De plus, les transitions « père-fils » a posteriori sont calculables ainsi que les solutions explicites des décisions Bayésiennes MPM et MAP [19, 58, 59, 93, 110]. Les arbres de Markov généralisent ainsi les chaînes de Markov, que l'on retrouve dans le cas où les générations sont réduites aux singletons.

Notons que les observations Y_t peuvent exister pour certains t et ne pas exister pour d'autres. Pour les t « sans observation » on pose $p(y_t | x_t) = 1$ et la formule (3.13) reste valable. Nous nous trouvons ainsi en présence d'un modèle capable de modéliser un grand nombre de situations en imagerie.

En particulier, il peut être appliqué à la segmentation d'images traitée dans les paragraphes précédents (une image à segmenter avec l'ensemble des pixels S) seul S^n a ici une existence physique (on pose $p(y_s | x_s) = 1$ pour $s \in S - S^n$). Les ensembles S^1, \dots, S^{n-1} ,

ainsi que les variables $X^{S^{n-1}} = (X_{S^1}, \dots, X_{S^{n-1}})$ sont artificiels et servent uniquement en tant qu'outils de calcul. On a ainsi trois processus aléatoires $X^{S^{n-1}}, X_{S^n}, Y_{S^n}$, avec $Y_{S^n} = y_{S^n}$ image observée, $X_{S^n} = x_{S^n}$ image recherchée, et $X^{S^{n-1}}$ variables annexes.

Remarque

Dans le cas mentionné ci-dessus, la loi de (X_{S^n}, Y_{S^n}) est ainsi la loi marginale de la loi de $(X^{S^{n-1}}, X_{S^n}, Y_{S^n})$ et sa structure, très complexe, est également très différente de celle des champs de Markov cachés donnée par (3.4). Cependant, les résultats obtenus avec ces deux modèles sont relativement proches, et chacun d'eux améliore sensiblement les résultats des segmentations obtenues avec la démarche « pixel par pixel », cette dernière expression signifiant que chaque X_s est estimé à partir du seul Y_s . Ainsi tout se passe comme si l'important était de prendre en compte le contexte spatial par le biais d'un modèle markovien, la nature exacte de ce dernier jouant un rôle secondaire.

Un autre exemple est celui de deux images avec deux résolutions différentes : $(X_{S^{n-1}}, Y_{S^{n-1}})$ et (X_{S^n}, Y_{S^n}) (les cardinaux des ensembles S^{n-1}, S^n sont différents). Les segmentées $(\hat{x}_{S^{n-1}}, \hat{x}_{S^n})$ sont alors obtenues à partir des images observées $(y_{S^{n-1}}, y_{S^n})$ (notons que chacune des images $\hat{x}_{S^{n-1}}, \hat{x}_{S^n}$ est obtenue à partir de deux images observées $(y_{S^{n-1}}, y_{S^n})$, ce qui doit améliorer la qualité de la méthode consistant en recherches séparées de $\hat{x}_{S^{n-1}}$ à partir de $y_{S^{n-1}}$, et de \hat{x}_{S^n} à partir de y_{S^n}), les variables $X^{S^{n-2}} = (X_{S^1}, \dots, X_{S^{n-2}})$ n'ayant pas d'existence physique. Enfin, on peut imaginer des cas encore plus généraux où pour chaque « résolution » S^i certains Y_t sont observables et d'autres pas (chacune des images $y_{S^i} = (y_t)_{t \in S^i}$ comporte des zones non observées). Là encore, on peut estimer $X = (X_s)_{s \in S}$, qui donne l'ensemble des variables $(X_s)_{s \in S}$, correspondant aux vraies images ou aux variables artificielles.

Les arbres de Markov cachés apparaissent ainsi comme concurrents des champs de Markov cachés, avec pour atouts une plus grande vitesse d'exécution, aussi bien des méthodes d'estimation des paramètres que des méthodes de segmentation, et une excellente adéquation au traitement des images multirésolution. Si une certaine analogie des efficacités des deux modélisations a pu être constatée dans des cas d'images relativement simples, des études comparatives plus approfondies seraient nécessaires pour pouvoir décider, a priori, quel type de modélisation est le mieux approprié à une situation donnée.

Notons également les récents modèles « mixtes », utilisant un champ de Markov à une résolution grossière et un arbre aux résolutions plus fines, visant à remédier aux « effets de bloc » pouvant apparaître dans les arbres. Ces derniers sont dus au fait que la loi de (X_s, X_t) , pour $s, t \in S^n$, est dépendante du lien de parenté entre s et t (père commun, grand père commun et père différent, arrière grand père commun et grand père différent, ..). Pour des différences de résolution de deux ou trois ($X_{S^{n-2}}$, ou $X_{S^{n-3}}$ est un champ de Markov) ces modèles peuvent atténuer les effets de blocs tout en gardant des temps de calcul très compétitifs par rapports à ceux nécessaires à l'utilisant des champs de Markov [47, 78].

3.6 Choix de modèle et de méthode

Lorsque l'on est en présence des données bruitées que l'on souhaite restaurer, ici une image que l'on souhaite segmenter, on se trouve ainsi devant un grand nombre de possibilités de choix. On peut opter pour les champs, les chaînes, ou pour les arbres de Markov cachés ; de plus, divers modèles « hybrides » faisant intervenir simultanément ces diverses markovianités sont envisageables. Dans chacun de ces modèles se pose, par ailleurs, la question du choix du voisinage définissant la markovianité. Lorsque l'on a opté pour un modèle et un voisinage donné, il se pose encore le problème du choix du bruitage : bruit gaussien additif, bruit gaussien général, bruit quelconque, bruit dont la forme varie avec la classe De plus, le nombre de classes peut également être considéré, dans certaines situations, comme inconnu. Nous constatons immédiatement qu'il est impossible, du fait de la trop grande diversité des modélisations possibles, d'imaginer une étude comparative présentant un quelconque caractère exhaustif. Cependant, dans une situation donnée, on peut envisager de proposer un mode de sélection, parmi une famille limitée de modèles, de celui le mieux adapté. Dans le cadre du choix parmi plusieurs modèles par champs de Markov cachés on peut utiliser des approximations de type Monte Carlo et la pseudo-vraisemblance pour la sélection de champs de Markov lorsque ce dernier est observé [51], ou des approximations de type pseudo-vraisemblance [103] et champs moyen [31], lorsque ce dernier est caché. Notons également l'important problème de la détermination, à partir de l'image observée, du nombre de classes ; en particulier, ce problème se pose dans les études des séquences d'images où ce nombre peut évoluer au cours du temps. Une des techniques récentes pour traiter ce problème, que l'on peut assimiler soit à un problème de sélection de modèle, soit à celui de l'estimation des paramètres, consiste en l'utilisation des méthodes de Monte Carlo par chaînes de Markov « à sauts réversibles » (JMCMC, RJ pour « reversible jump » en anglais), où le nombre des classes apparaît comme un paramètre que l'on estime. A titre d'exemple, citons [5] pour l'application de telles techniques aux images à niveau de gris et [53] pour leur application aux images couleur. Enfin, mentionnons également une méthode de sélection automatique du meilleur modèle, parmi les champs de Markov, les chaînes de Markov, et un modèle local, étudiée dans des situations simples dans [97].

4. Modèles de Markov Couple

4.1 Textures

Considérons le cas d'une seule classe dans l'image présentant une « texture », que nous allons considérer, dans le cadre probabiliste retenu, comme la réalisation d'un champ aléatoire dont les composantes sont spatialement dépendantes. Si l'on suppose les champs aléatoires stationnaires au second ordre, il existe un certain nombre de modèles comme les modèles de Markov, CAR, AR, ou ARMA [44]. Les modèles de Markov sont parmi les modèles les plus simples. A titre d'exemple, supposons que l'image représente une forêt dont la texture est modélisée par un champ aléatoire $Y = (Y_s)_{s \in S}$ gaussien et markovien relativement aux quatre plus proches voisins. Sa loi s'écrit :

$$\begin{aligned}
 p(y) &= \frac{1}{\sqrt{\det(2\pi\Gamma)}} \exp\left[-\frac{y^T \Gamma^{-1} y}{2}\right] \\
 &= \frac{1}{\sqrt{\det(2\pi\Gamma)}} \exp\left[-\sum_{(t,u) \in C^1} \beta^1 (y_t - y_u)^2 - \sum_{(t,u) \in C^2} \beta^2 (y_t - y_u)^2\right]
 \end{aligned} \tag{4.1}$$

où C^1 est l'ensemble des pixels voisins horizontalement et C^2 est l'ensemble des pixels voisins verticalement.

4.2 Markov cachés et Markov couple

Considérons le problème de segmentation d'une image en deux classes texturées « eau » et « forêt ». En adoptant le modèle champs de Markov cachés, considérons X de Markov relativement aux quatre plus proches voisins et supposons que les deux textures sont modélisées par les lois de type (4.1). La loi de (X, Y) s'écrit

$$p(x, y) = p(x)p(y|x) = \tag{4.2}$$

$$\begin{aligned}
 &\frac{\gamma}{\sqrt{\det(2\pi\Gamma(x))}} \exp\left[-\sum_{c \in C^1 \cup C^2} \phi_c(x_c) - \sum_{(t,u) \in C^1} \beta_{x_t, x_u}^1 (y_t - y_u)^2 - \sum_{(t,u) \in C^2} \beta_{x_t, x_u}^2 (y_t - y_u)^2\right] = \\
 &\gamma(x) \exp\left[-\sum_{c \in C^1 \cup C^2} \phi_c(x_c, y_c)\right]
 \end{aligned}$$

On reconnaît l'écriture markovienne sous l'exponentielle ; cependant, il n'est pas certain que $\gamma(x) = \frac{\gamma}{\sqrt{\det(2\pi\Gamma(x))}}$ admette la même forme d'écriture. Il n'est donc pas certain que

$p(x|y)$ soit de Markov. Cela pose problème car la markovianité de $p(x|y)$ est indispensable pour mettre en œuvre aussi bien les procédures d'estimation des paramètres que les segmentations bayésiennes. Lorsque l'on utilise les champs de Markov cachés (ce qui signifiera, dans la suite, que le champ de classes X est de Markov), on est alors amené à faire diverses approximations, qui souvent reviennent à considérer que $\gamma(x)$ dans (4.2) ne dépend pas de x [55, 65]. En effet, la loi « approchée » $p'(x|y)$ est alors markovienne ce qui rend les divers traitements, mentionnés dans la section précédente, possibles.

Considérer un champ de Markov Couple consiste à poser directement la markovianité du couple (X, Y) [82]. En reprenant l'exemple précédent, posons directement :

$$p(x, y) = \lambda \exp\left[-\sum_{c \in C^1 \cup C^2} \phi_c(x_c, y_c)\right] \tag{4.3}$$

avec les ϕ_c définies dans (4.2). Nous constatons que dans le modèle Markov Couple la loi $p(x|y)$ (qui est égale à la loi approchée $p'(x|y)$ obtenue dans le modèle Markov caché) est markovienne, ce qui rend les traitements possibles sans approximation du modèle. Notons

que dans le modèle de Markov Couple la nature éventuellement markovienne de la loi du processus X n'est nullement évidente ; en effet, en multipliant et en divisant (4.3) par $\gamma(x)$ et en intégrant par rapport à y on trouve $p(x) = \frac{\lambda}{\gamma(x)} \exp[-\sum_{c \in C^1 \cup C^2} \phi_c(x_c)]$. Ainsi, par rapport au modèle de Markov caché classique on perd la markovianité de $p(x)$, qui n'est pas nécessaire aux traitements, et on gagne la markovianité de $p(x|y)$, qui est indispensable.

De manière analogue, on peut introduire les chaînes de Markov Couple [84]. Il est alors possible de préciser des conditions sous lesquelles la chaîne X n'est pas de Markov ; il en résulte que le modèle « chaîne de Markov couple » $Z = (X, Y)$ est strictement plus général que le modèle « chaîne de Markov cachée ».

Au plan intuitif, cette généralisation est visible lorsque l'on considère les transitions $p(z_{i+1}|z_i) = p(x_{i+1}, y_{i+1}|x_i, y_i)$. Dans une chaîne de Markov cachée classique vérifiant (3.9) cette transition est donnée par $p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1})$, et dans une chaîne de Markov Couple elle est donnée par la formule générale $p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i)$. On peut ainsi considérer que la première est obtenue à partir de la seconde en approchant $p(x_{i+1}|x_i, y_i)$ par $p(x_{i+1}|x_i)$ et en approchant $p(y_{i+1}|x_{i+1}, x_i, y_i)$ par $p(y_{i+1}|x_{i+1})$.

Les premières études montrent que l'utilisation des chaînes de Markov Couple peut améliorer les divers résultats de segmentations, en particulier ceux de la segmentation non supervisée des images utilisant le parcours de Hilbert-Peano, obtenus avec les chaînes de Markov cachés [25]. Enfin, les arbres de Markov cachés peuvent également être généralisés aux arbres de Markov Couple [85], dans lesquels tous les traitements et tous commentaires concernant les arbres de Markov cachés restent valables.

Notons également les récents modèles dits chaînes de Markov « Triplet » [86] et champs de Markov « Triplet » [87]. Ils sont obtenus en introduisant un processus artificiel $U = (U_s)_{s \in S}$ et en considérant la markovianité du triplet (X, U, Y) . On montre alors que les chaînes Triplet sont plus généraux que les chaînes Couple et autorisent néanmoins les traitements Bayésiens MPM [86]. Enfin, il existe un lien entre les modèles de Markov Triplet et la fusion de Dempster-Shafer [1, 99, 101], qui est précisé dans la section 6 ci-dessous.

4.3 Markov cachés avec un bruit corrélé

Considérons un processus de Markov (champ, chaîne, ou arbre) $X = (X_s)_{s \in S}$ dont la loi est $p(x)$. On peut alors montrer que l'hypothèse simplificatrice selon laquelle $p(y|x) = \prod_{s \in S} p(y_s|x_s)$ n'est pas une hypothèse nécessaire pour que $p(x|y)$ soit de Markov.

En d'autres termes, on peut considérer $Y = (Y_s)_{s \in S}$ tel que les (Y_s) ne sont pas indépendants conditionnellement à X , avec $p(x)$ et $p(x|y)$ de Markov.

On distingue ainsi quatre familles de modèles de généralité strictement croissante : les Markov cachés avec bruit indépendant, les Markov cachés avec bruit corrélé, les Markov Couple, et les Markov Triplet.

5. Apprentissage des Modèles de Markov

5.1 Champs de Markov

Supposons que la loi de X donnée par (3.1) dépende d'un ensemble des paramètres α , qui apparaissent dans les fonctions « potentiel » φ_c qui deviennent φ_c^α , et le problème est son estimation à partir de X . La méthode générale du Maximum de Vraisemblance (MV), dont le principe est :

$$\hat{\alpha}(x) = \arg \max_{\alpha} \gamma(\alpha) \exp \left[- \sum_{c \in C} \varphi_c^\alpha(x_c) \right] \quad (5.1)$$

ne peut être appliquée directement car $\gamma(\alpha)$ est inconnue. La première idée est de remplacer la vraisemblance par la « pseudo-vraisemblance » $pv(x)$, qui est le logarithme du produit par rapport à $s \in S$ des lois conditionnelles (3.2) [11]:

$$pv(x) = \text{Log} \left[\prod_{s \in S} p(x_s | x_{V_s}) \right] = \text{Log} \left[\prod_{s \in S} \frac{\exp \left[- \sum_{c \in C, s \in c} \varphi_c(x_s, x_c) \right]}{\sum_{\omega \in \Omega} \exp \left[- \sum_{c \in C, s \in c} \varphi_c(x_s = \omega, x_c) \right]} \right] \quad (5.2)$$

La fonction $\gamma(\alpha)$ disparaît et on peut, sous certaines conditions, calculer ou approcher le Maximum de la Pseudo-Vraisemblance (MPV). De plus, MPV jouit de bonnes propriétés asymptotiques [39]. Notons que l'emploi de la pseudo-vraisemblance peut paraître quelque peu surprenant de prime abord; en effet, les bonnes propriétés du comportement asymptotique sont surtout connues dans le cas de l'estimateur du maximum de vraisemblance (MV), ce dernier étant égal à MPV lorsque les données sont indépendantes, donc en l'absence de la markovianité. Cependant, cette incohérence n'est qu'apparente car les bonnes propriétés asymptotiques demeurent lorsque l'on remplace le MV par le maximum d'une fonction « score » quelconque, dès que cette dernière devient suffisamment « proche » de la vraisemblance lorsque la taille des données augmente [21].

Une deuxième méthode, que nous appellerons « Estimateur Empirique » (EE) et qui peut être appliquée lorsque la taille du voisinage et le nombre de classes ne sont pas trop importants, consiste à calculer la fréquence d'apparition de chacune des classes conditionnellement à toutes les configurations du voisinage, et à rechercher α qui ajuste ces fréquences aux lois conditionnelles données par (3.2). Lorsque l'énergie dépend linéairement du vecteur $\alpha = (\alpha_1, \dots, \alpha_m)$, à savoir

$$U_\alpha(x) = \sum_{1 \leq i \leq m} \alpha_i \left[\sum_{c \in C_i} \varphi_c(x_c) \right] = \sum_{1 \leq i \leq m} \alpha_i U_i(x), \text{ cet ajustement peut être fait au sens des}$$

moindres carrés [44, 77].

Enfin, toujours sous l'hypothèse de la dépendance linéaire, on peut montrer que :

$$\frac{\partial p_\alpha(x)}{\partial \alpha_i} = U_i(x) - E_\alpha[U_i(X)], \quad (5.3)$$

l'espérance $E_\alpha[U_i(X)]$ pouvant être approchée par la moyenne empirique calculée à partir des simulées de X . Il est alors possible de mettre en place des méthodes dites du gradient stochastique (GS) qui permettent de trouver, sous certaines conditions, l'estimateur du MV [114].

5.2 Champs de Markov cachés

Notons α l'ensemble des paramètres donnant la loi $p_\alpha(x)$ de X , β l'ensemble des paramètres donnant les lois conditionnelles $p_\beta(y|x)$, et $\theta = (\alpha, \beta)$ l'ensemble des paramètres que l'on cherche à estimer. Nous nous intéressons à leur estimation à partir de Y ; une telle estimation sera dite « estimation dans le cas des données incomplètes », alors que celle effectuée à partir de (X, Y) sera dite « estimation dans le cas des données complètes ». Nous décrivons brièvement trois méthodes générales d'estimation dans le cas des données incomplètes.

La vraisemblance de la loi de Y , qui s'écrit $p_\theta(y) = \sum_{x \in \Omega^N} p_\alpha(x) p_\beta(y|x)$, est trop complexe

pour que l'on puisse envisager le calcul direct de l'estimateur MV. Le principe de la méthode « Expectation-Maximization » (EM [64]), qui permet de définir une suite (θ^n) telle que la suite $(p_{\theta^n}(y))$ est croissante, est le suivant :

- (i) on se donne un paramètre initial θ^0 ;
- (ii) θ^{q+1} est défini à partir de θ^q et y par

$$\theta^{q+1} = \arg \max_{\theta} E_{\theta^q} [\text{Log}(p_\theta(X, Y) | Y = y)] \quad (5.4)$$

Notons que l'appellation EM vient du fait que (5.4) donne généralement θ^{q+1} en deux temps : calcul de l'espérance E_{θ^q} (en anglais « expectation »), et sa maximisation (en anglais « maximization »). La méthode EM est très largement utilisée et donne généralement des résultats satisfaisants ; cependant, dans le contexte des champs de Markov cachés son application est malaisée et l'on doit s'écarter du principe général en utilisant diverses approximations. Parmi ces dernières on peut citer des approximations fondées sur des simulations stochastiques [17, 18, 64, 77, 89, 119], ou encore sur les « champs moyens » [16, 29, 118]. Mentionnons également la difficulté de montrer la convergence de la suite définie par EM, même dans le cas où les deux phases peuvent être traitées par des calculs exacts, vers le maximum absolu de vraisemblance. Les seuls résultats généraux sont du type « sous de bonnes conditions la suite converge vers un des points stationnaires », un point stationnaire étant un point annulant le gradient de la vraisemblance. Lorsque la vraisemblance est suffisamment régulière, on obtient les résultats du type « lorsque l'initialisation est suffisamment proche du maximum de vraisemblance absolu, alors la suite converge vers ledit maximum » [64]. On constate qu'il est très difficile de vérifier de telles hypothèses dans des situations réelles ; en effet, pour être sûr de trouver la bonne solution on se doit de savoir où elle se trouve. Cela dit, ce type d'hypothèse ne semble pas souvent nécessaire et la méthode EM donne généralement de bons résultats dans la pratique.

La deuxième méthode générale, appelée « Iterative Conditional Estimation » (ICE [79, 80]) est applicable dès que l'on dispose d'un estimateur de $\theta \in \Theta$ à partir des données

complètes $\hat{\theta} = \hat{\theta}(X, Y)$ et d'une méthode de simulation de X selon $p_{\theta}(x|y)$. Le déroulement de ICE est le suivant :

- (i) on se donne un paramètre initial θ^0 ;
- (ii) θ^{q+1} est défini à partir de θ^q et y par :

$$\theta^{q+1} = E_{\theta^q}[\hat{\theta}(X, Y)|Y = y] \quad (5.5)$$

lorsque cette espérance est calculable, et par :

$$\theta^{q+1} = \frac{\hat{\theta}(x^1, y) + \dots + \hat{\theta}(x^l, y)}{l} \quad (5.6)$$

avec x^1, \dots, x^l simulés selon $p_{\theta^q}(x|y)$, lorsqu'elle ne l'est pas.

Remarquons qu'il peut arriver d'appliquer (5.5) pour certaines composantes de θ et (5.6) pour les composantes restant.

Nous avons vu que X pouvait être simulé selon $p_{\theta}(x|y)$ par l'échantillonneur de Gibbs, l'existence de $\hat{\theta} = \hat{\theta}(X, Y)$ est ainsi la seule condition à vérifier. C'est une condition très faible car, si l'on ne trouve pas un estimateur à partir des données complètes (x, y) il est illusoire d'en chercher un à partir des seules incomplètes y . Ainsi si $p_{\beta}(y|x)$ est donnée par (3.3) et si les k densités f_1, \dots, f_k sont paramétrées, on peut poser $\hat{\theta} = \hat{\theta}(X, Y) = (\hat{\alpha}(X), \hat{\beta}(X, Y))$, avec pour $\hat{\alpha}(X)$ l'un des estimateurs du paragraphe 5.1 et pour $\hat{\beta}(X, Y)$ les estimateurs des moments empiriques, valables dans les cas des densités classiques sur R . Bien que souffrant de l'absence de résultats théoriques concernant son comportement asymptotique, ICE ne fait pas appel à la notion de vraisemblance et peut utiliser tout estimateur $\hat{\theta} = \hat{\theta}(X, Y)$, ce qui la rend relativement souple et générale. De par cette souplesse, ICE semble bien adaptée aux champs de Markov cachés et a été utilisée avec succès dans de nombreuses situations [8, 9, 13, 22, 30, 54, 66-70, 96, 97]. Par ailleurs, des liens existent entre EM et ICE dans le cadre du modèle exponentiel, où ICE apparaît comme plus générale [23].

La famille des méthodes dites « pleinement bayésiennes » (« fully bayesian » en anglais) constitue une troisième famille générale de méthodes possibles d'estimation des paramètres [46, 71]. Sachant que seul $Y = y$ est observé et que l'on recherche θ pour pouvoir estimer $X = x$, on peut considérer que l'on recherche le couple (θ, x) à partir de $Y = y$. Des méthodes bayésiennes de cette recherche peuvent être envisagées dès que l'on dispose d'une loi de probabilité sur $(\Theta, \Omega^{Card(S)})$, qui sera supposée être la loi du couple de variables aléatoires (W, X) . Pour disposer d'une telle loi il suffit de se donner une loi $p(\theta)$ sur l'ensemble des paramètres Θ ; on a alors $p(\theta, x) = p(\theta)p(x|\theta)$, où $p(x|\theta)$ est la loi de X correspondant à θ . La loi $p(\theta)$ doit être choisie de manière que $p(\theta, x)$ ait une structure markovienne ; la loi $p(\theta, x|y)$ garde alors la même structure markovienne et les réalisations de (W, X) sont simulables selon une des procédures MCMC, comme échantillonneur de Gibbs ou l'algorithme de Hasting-Métropolis. On peut alors utiliser ces simulations pour estimer $p(\theta, x|y)$ et procéder à une estimation bayésienne simultanée,

utilisant la loi estimée, du couple (θ, x) . On peut également s'intéresser uniquement à l'estimation de θ ; en effet, les réalisations de (W, X) donnent des réalisations de W , qui peuvent être utilisées pour l'estimation de $p(\theta)$, cette dernière servant à l'estimation bayésienne de θ . Notons que dans les deux cas (estimation de (θ, x) ou estimation de θ) les estimations utilisent des fonctions de perte qui peuvent être choisie de manière arbitraire, ce qui constitue un avantage de ce type d'estimation. A titre d'exemple, si l'on a choisi d'estimer θ avec pour objectif final l'estimation de x à partir de y (nous dirons la segmentation), on peut considérer une fonction de perte favorisant l'estimation des composantes de θ par rapport auxquelles la segmentation est peu robuste. Mentionnons également la gestion de la non stationnarité de X comme une autre possibilité d'utilisation des méthodes « pleinement » bayésiennes. En effet, pour un champs de Markov X non stationnaire les paramètres θ dépendent des cliques et l'on peut alors les considérer comme une réalisation d'un champs aléatoire. Ainsi la variable aléatoire W , qui était ci-dessus à valeurs dans Θ , devient un champs aléatoire $W = (W_c)_{c \in C}$, avec C l'ensemble des cliques défini par la markovianité de X . On peut alors considérer, pour des lois de W suffisamment simples, une structure markovienne pour $p(w, x|y)$, ce qui permet la simulation de (W, X) selon cette dernière loi, et donc également la simulation de W selon $p(w|y)$. Lesdites simulations permettant des restaurations bayésiennes de W , ces dernières donnent des champs des paramètres définissant des non stationnarités de X [2]. Notons que les méthodes pleinement bayésiennes résolvent le problème de l'estimation de θ en en posant un autre : la loi $p(\theta)$ dépendant nécessairement des paramètres, comment déterminer ces derniers ? Ce problème semble avoir moins d'influence sur la qualité des résultats ; la loi $p(\theta)$ est généralement choisie parmi les lois simples, comme, à titre d'exemple, les lois uniformes sur des ensembles compacts.

Mentionnons également d'autres méthodes d'estimation, comme celle dérivant du gradient stochastique [115] qui présente une étude mathématique rigoureuse, celle utilisant le gradient pour accélérer les méthodes stochastiques [117], ou les méthodes utilisant alternativement l'estimation des paramètres et la restauration du champs x sur la bse des paramètres courants, comme celle de Besag [11], ou celle de Lakshmanan *et al.* [60].

Signalons la possibilité de l'estimation des modèles plus généraux, où la nature des densités f_1, \dots, f_k n'est pas connue ; cependant, chacune appartient à un ensemble connu de formes admissibles. A titre d'exemple, imaginons que chacune des densités f_1, \dots, f_k est gaussienne ou Gamma, mais on ne sait pas dans quel cas, parmi les 2^k cas possibles, l'on se trouve. On peut alors proposer des procédures qui estiment α , donnent le type pour chacune des f_1, \dots, f_k , et estiment les paramètres β_i fixant f_i dans le type donné. De telles méthodes, qui peuvent être des extensions de EM ou ICE, sont appelés estimateurs de « mélanges généralisés » [8, 22, 42, 81]. Mentionnons également une méthode récente [120] permettant une estimation non paramétrique des densités f_1, \dots, f_k .

Exemple

Considérons un exemple, extrait de [22], d'une segmentation non supervisée « pixel par pixel » d'une part, et d'une segmentation non supervisée utilisant le modèle par champ de Markov caché, d'autre part (Figure 1). L'image « cible » à deux classes a été bruitée en utilisant deux densités de probabilité très proches (admettant, en particulier, la même

moyenne et la même variance). L'œil humain ne distingue pas les classes dans l'image bruitée et, conformément à la théorie, la segmentation « pixel par pixel » donne des résultats bien médiocres (pourtant ces résultats sont proches de ceux, optimaux, obtenus avec les vrais paramètres). Le résultat obtenu avec les champs de Markov cachés, tout à fait correct, montre alors l'étendue des possibilités d'amélioration des traitements par la prise en compte, au travers des modèles de Markov, de l'information spatialement contextuelle.

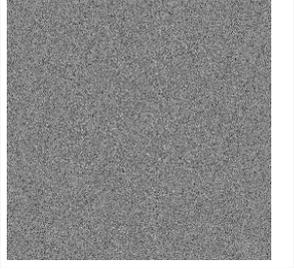
			
Image des classes $X = x$	Image observée $Y = y$	Segmentation « pixel par pixel »	Segmentation par champs de Markov

Figure 1

Image originale, la version bruitée (mêmes moyennes, mêmes variances), segmentation par une méthode « pixel par pixel » et une méthode fondée sur le modèle par champ de Markov caché.

Nous constatons la très grande richesse des possibilités de proposer une méthode d'estimation des paramètres des champs de Markov cachés. Ces méthodes sont généralement itératives et coûteuses en temps de calcul. Par ailleurs, le temps d'exécution peut être très variable en fonction de la complexité des images traitées, cette dernière influant sur le nombre des diverses itérations. Notons qu'il serait sans doute périlleux de vouloir les classer en fonction de leurs qualités, ces dernières étant fonction des différents types d'images.

5.3 Champs de Markov Couple

Mentionnons brièvement la possibilité de l'utilisation de ICE dans le contexte des champs de Markov Couple. Le processus $Z = (X, Y)$ est de Markov, sa loi $p_\theta(x, y)$ dépend des paramètres θ que l'on cherche à estimer à partir de $Y = y$. Nous devons d'abord considérer un estimateur à partir des données complètes $\hat{\theta} = \hat{\theta}(X, Y) = \hat{\theta}(Z)$, ce qui est possible en adaptant l'un des estimateurs MPV, EE, ou GS brièvement décrits dans le paragraphe 5.1 (notons bien que l'on considère ici $Z = (X, Y)$ à la place de X considéré dans le paragraphe 5.1). Par ailleurs, X étant Markovien conditionnellement à $Y = y$, ses simulations selon $p_\theta(x|y)$ sont possibles. Les deux conditions suffisantes de l'application de l'ICE sont ainsi remplies.

5.4 Chaînes et arbres de Markov cachés et couple.

Dans les cas classiques des chaînes ou arbres de Markov cachés avec du bruit gaussien, l'estimation des paramètres ne pose pas de problème majeur [9, 58, 59]. A titre d'exemple,

les formules de réestimation (5.4) définissant l'algorithme EM peuvent être explicitées analytiquement, ce qui rend l'estimation rapide. Par ailleurs, diverses études montrent que dans le cas des bruits gaussiens l'algorithme EM est, lorsqu'il s'agit d'utiliser les paramètres estimés à des fins de classification Bayésienne, tout à fait performant. En d'autres termes, la qualité des segmentations effectuées sur la base des paramètres estimés par EM est proche de celle des segmentations effectuées à partir des vrais paramètres et cela jusqu'à un niveau élevé du bruit. De plus, EM est très robuste dans le sens où, lorsque les données ne suivent pas nécessairement un modèle de Markov caché, la qualité des segmentations effectuées sur la base des paramètres estimés à partir des seules données observées par EM est proche de celle des segmentations effectuées sur la base des paramètres estimés à partir des données complètes. Notons que ICE est également applicable et donne des résultats proches ; cependant, EM présente un avantage par rapport à ICE au niveau du temps de calcul : dans les itérations, les paramètres α de la loi de X sont réestimés par les mêmes formules analytiques mais les paramètres du bruit, réestimés par des formules analytiques dans EM, le sont à partir des simulations dans ICE.

Les chaînes de Markov cachées peuvent être appliquées directement à la segmentation d'images, en transformant au préalable l'ensemble bi-dimensionnel des pixels en un ensemble mono-dimensionnel en utilisant le parcours de Hilbert-Peano (Figure 2). On obtient alors des méthodes non supervisées rapides pouvant donner des résultats comparables à ceux obtenus avec les champs de Markov [97]. Ces méthodes rapides peuvent également servir à l'initialisation des méthodes fondées sur les champs de Markov [30].

Le principe de la méthode ICE s'applique directement aux modèles par chaînes et arbres Couple ; en particulier, il est possible de proposer des variantes de ICE adaptées aux bruits corrélés et non Gaussiens et les premiers résultats numériques sont encourageants [25]. Enfin, le problème de l'estimation des paramètres dans les Markov Triplet est identique à celui de l'estimation des paramètres dans les Markov Couple ; en effet, le Triplet (X, U, Y) est également un Couple (V, Y) , avec $V = (X, U)$.

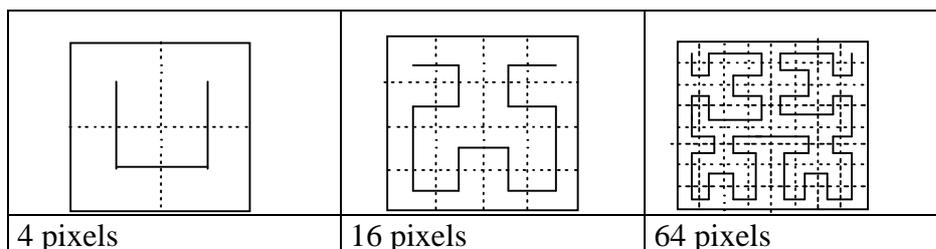


Figure 2

Construction du parcours fractal de Hilbert-Peano

5.5 Choix de méthode non supervisée de segmentation

Lorsque l'on est confronté au problème de segmentation statistique non supervisée d'une image, on a ainsi à sa disposition un certain nombre de modélisations markoviennes et il est possible d'appliquer à chacune d'elle une certaine stratégie bayésienne, après avoir préalablement estimé les paramètres par l'une de nombreuses méthodes possibles. Comment faire un bon choix en face d'un problème concret ? La richesse des méthodes

répondant à la richesses des images, le problème est bien complexe et il n'existe pas, à notre connaissance, d'études poussées à ce sujet.

Signalons cependant une étude simple présentée dans [97] dont une des conclusions est qu'une méthode locale très simple peut donner, dans des situations particulières, de meilleurs résultats que les méthodes markoviennes (voir Figure 3, extraite de [97]).

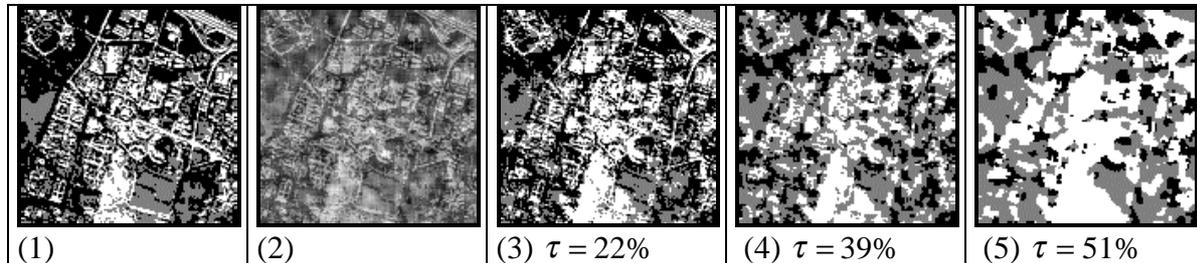


Figure 3

Image à trois classes obtenue à partir d'une image réelle par seuillage (1), sa version bruitée avec du bruit corrélé (2), segmentation par une méthode locale « pixel par pixel », avec l'estimation des paramètres adaptative sur une fenêtre mobile (3), segmentation par les chaînes de Markov cachées (4), et par les champs de Markov cachés (5). τ : pourcentage des pixels mal classés.

6. Extensions

La robustesse des modèles de Markov cachés, aussi bien en ce qui concerne la robustesse des traitements par rapport à l'adéquation des modèles aux données qu'en ce qui concerne les traitements non supervisés (paramètres estimés à partir des données incomplètes) par rapport aux traitements supervisés (paramètres estimés à partir des données complètes), permet de proposer leur utilisation dans des situations toujours plus complexes. Nous présentons brièvement ci-après une liste non exhaustive des situations plus générales auxquelles les modélisations et les traitements précédents peuvent s'adapter sans difficultés majeures.

6.1 Images multisenseurs.

Dans le cas de plusieurs senseurs chaque Y_s prend ses valeurs dans R^m , situation à laquelle les différents modèles de Markov cachés se généralisent aisément lorsque les paramètres sont connus. Lorsque les bruits sont gaussiens, ou lorsqu'ils ne sont pas gaussiens mais que les senseurs sont indépendants, l'estimation des paramètres ne pose pas de problème particulier, même lorsque la nature exacte des composantes du mélange n'est pas connue [42]. Lorsque les bruits sont non nécessairement gaussiens et que les senseurs sont dépendants, l'estimation est plus compliquée, notamment dans le cas où les types de densités ne sont pas connues [81]. Lorsque les senseurs opèrent dans des bandes spectrales complémentaires et lorsque leur nombre dépasse la dizaine, on est en présence de l'imagerie « hyperspectrale » [61, 90, 98, 104]. L'utilisation des champs de Markov dans ce contexte se heurte à la croissance rapide du nombre des paramètres ; cependant, l'utilisation des approximations pertinentes ou d'autres modèles Markoviens constituent des perspectives intéressantes de recherche et apportera sans doute des réponses efficaces

aux problèmes se posant dans un certain nombre de situations. Notons que le problème de la segmentation des images couleur entre dans cette problématique (trois bandes « rouge », « vert », et « bleu ») [53].

6.2 Images 3D et séquences d'images.

Lorsque l'ensemble des pixels se trouve dans R^3 , on les appelle alors « voxels », on a une image 3D, ou une séquence d'images. Les champs de Markov cachés se généralisent aisément à cette situation et peuvent être utilisés dans des situations très diverses. Citons, à titre d'exemple, des applications en séquences d'images [45, 57], en imagerie satellitaire [4] ou médicale [26, 75]. Cependant, le nombre de paramètres croît rapidement et la durée des divers traitements utilisant des techniques MCMC peut devenir rédhibitoire. Comme dans le cas bidimensionnel, on peut alors imaginer d'utiliser les chaînes de Markov, après avoir converti l'ensemble des voxels en une suite, par des parcours de type Hilbert-Peano [100]. L'introduction des modèles Couple ou Triplet est possible, de même que les traitements des images 3D – ou des séquences d'images – multisenseurs. Enfin, l'estimation des paramètres, dont le nombre cependant croît rapidement, ne pose pas de problèmes autres que ceux liés à la complexité de la programmation informatique et à la durée de l'exécution des programmes.

Exemple

Considérons un exemple, extrait de [9], d'utilisation des chaînes de Markov cachées dans la segmentation des séquences d'images. En construisant un parcours de Hilbert-Peano dans l'ensemble des pixels de l'image courante et dans celui de l'image précédente (Figure 4), on obtient une chaîne de longueur égale à deux fois le nombre de pixels (notons l'existence de deux matrices de transition différentes : une matrice « spatiale » qui gère les transitions entre deux variables indicées par les pixels dans une même image, et une matrice « temporelle » qui gère les transitions entre deux variables indicées par les pixels se trouvant dans deux images différentes). Nous constatons que la segmentation utilisant un tel parcours (Figure 5, séquence 2) est plus efficace que celle utilisant les images de manière indépendante (Figure 5, séquence 1).

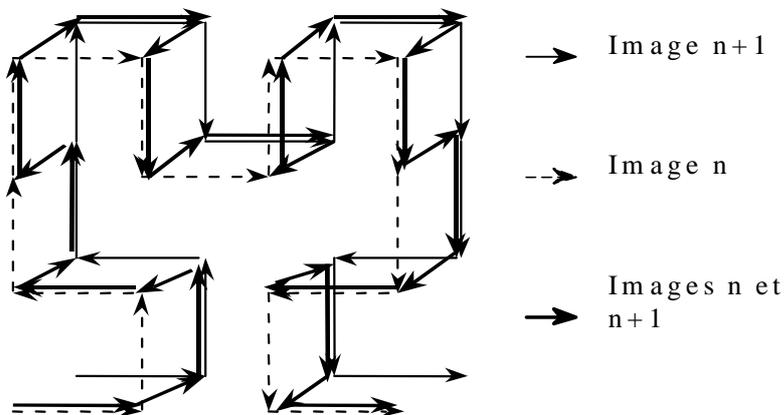


Figure 4

Construction du parcours fractal de Hilbert-Peano spatio-temporel (image courante et image précédente)



Figure 5

Séquence réelle, segmentation à partir de l'image courante (séquence 1), segmentation à partir de l'image courante et de l'image précédente (séquence 2).

6.3 Images multirésolution

Reprenons la situation du paragraphe 3.4, où S^1, \dots, S^n est une partition de S représentant diverses « générations ». Il existe alors une grande diversité des cas possibles. Les variables X_s peuvent avoir une existence réelle pour certaines générations et fictive pour d'autres, leurs ensembles d'arrivée pouvant par ailleurs varier avec la génération. La markovianité du processus de $X = (X_s)_{s \in S}$ peut alors être de type « arbre », avec des traitements utilisant des calculs directs et s'apparentant aux calculs classiques dans les chaînes de Markov, ou de type « champ », où les calculs directs ne sont pas possibles et l'on doit faire appel à des techniques de simulation, généralement de type MCMC [43, 76]. Notons que le panachage de deux types de modèles est possible ; en effet, ainsi que mentionné ci-dessus, on peut imaginer une pyramide où la restriction de X à S^1 , notée X^1 , qui est de forme rectangulaire, est un champ de Markov et la loi de sa restriction à $S - S^1$ est un arbre de Markov [47, 78]. La loi de X est alors donnée par (3.12), avec la loi d'un champ Markovien $p(x^1)$ à la place de $p(x_1)$. En adoptant les hypothèses classiques de l'indépendance des (Y_s) conditionnellement à X et $p(y_s|x) = p(y_s|x_s)$, $p(x, y)$ garde la structure de $p(x)$, et il en est de même de $p(x|y)$. Cette dernière propriété rend possible soit le calcul direct de diverses quantités d'intérêt, soit leur estimation à partir des simulées

de X selon $p(x|y)$ via des méthodes MCMC. Les quantités d'intérêt en question permettent alors aussi bien l'estimation des paramètres par des méthodes comme EM ou ICE que les segmentations bayésiennes correspondant à différentes fonctions de perte. Enfin, les généralisations aux modèles Couple, où on attribuerait à $Z = (X, Y)$ la structure markovienne de X , et aux modèles Triplet, où cette structure serait attribuée à $T = (X, U, Y)$, sont envisageables. Notons également l'existence des règles générales permettant de calculer, lorsque les deux processus X et Y sont indexés par des éléments d'un graphe quelconque, des lois marginales d'une partie des variables aléatoires, ou de leur loi conditionnelle à une autre partie des variables aléatoires [7, 50].

On trouvera dans [110] une synthèse récente des résultats concernant l'utilisation des modèles de Markov en multirésolution.

6.4 Fusion de Dempster-Shafer

Cette sous-section est consacrée à la description des possibilités de généralisation des modèles de Markov cachés en utilisant la théorie de l'évidence de Dempster-Shafer. En se plaçant sur un pixel et reprenant l'exemple d'eau et forêt de la section 2, nous avons vu que la probabilité a posteriori, qui permet la segmentation bayésienne, est $p(E|y_s) \propto p(E)p(y_s|E)$, $p(F|y_s) \propto p(F)p(y_s|F)$. En considérant la probabilité $q(E) \propto p(y_s|E)$, $q(F) \propto p(y_s|F)$ (la probabilité q est ainsi définie uniquement par l'observation y_s), nous constatons que la probabilité a posteriori est obtenue à partir de la probabilité a priori $p(E)$, $p(F)$ et la probabilité q par la « fusion » $p(E|y_s) \propto p(E)q(E)$, $p(F|y_s) \propto p(F)q(F)$. Il s'agit là d'un cas particulier de la fusion de Dempster-Shafer (DS), qui s'applique dans des modélisations plus générales proposées dans le cadre de la « théorie de l'évidence » [1, 99, 101]. Afin d'illustrer l'intérêt de la fusion DS dans certains problèmes pouvant survenir en segmentation d'images continus l'exemple ci-dessus en supposant que des nuages apparaissent dans l'image « eau » et « forêt ». Nous sommes ainsi en présence de la classe « nuages » notée N , avec $p(y_s|N)$, ce qui donne une probabilité q définie sur $\{E, F, N\}$ par $q(E) \propto p(y_s|E)$, $q(F) \propto p(y_s|F)$, et $q(N) \propto p(y_s|N)$. Si nous sommes intéressés uniquement par la détection de l'une des classes E , F , observer N ne nous apporte aucune information utile. Dans le cadre de la théorie de l'évidence ce fait est modélisé en assimilant N à $\{E, F\}$, ce qui nous mène à considérer que q est définie sur $\{E, F, \{E, F\}\}$ (dans le langage de la théorie de l'évidence q est dite « fonction de masse »). On peut alors dire, dans un certain sens, qu'une fonction de masse généralise la probabilité, cette dernière étant obtenue pour les fonctions de masses nulles sur $\{E, F\}$ (lorsque les nuages disparaissent, la fonction de masse redonne une probabilité classique). La fusion DS de q avec la probabilité a priori $p(E)$, $p(F)$ donne alors $p(E|y_s) \propto p(E)[q(E) + q(N)]$, $p(F|y_s) \propto p(F)[q(F) + q(N)]$, qui est une probabilité « généralisant » la probabilité a posteriori (on obtient la probabilité a posteriori classique lorsque les nuages disparaissent) et qui peut être utilisée à des fins de classification.

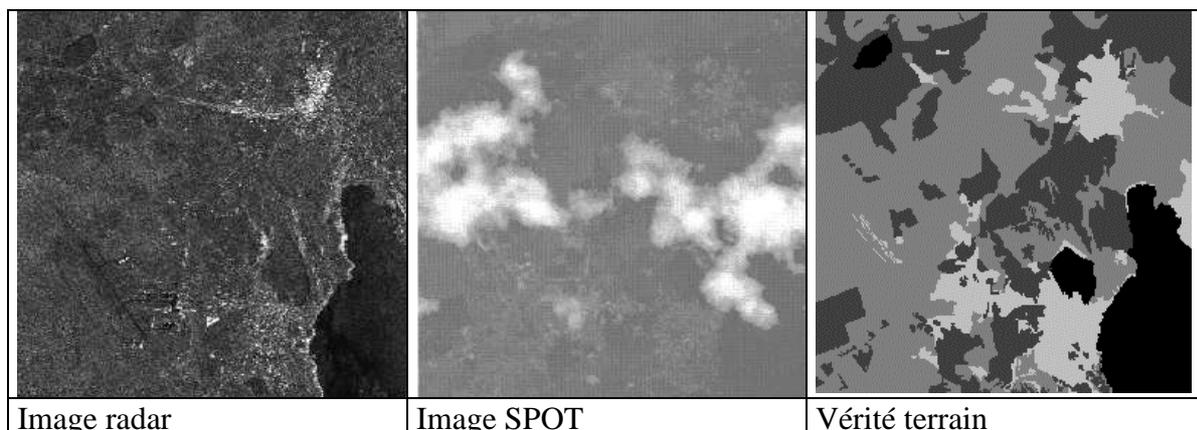
Cette démarche se généralise au cas de m senseurs indépendants, certains d'entre eux pouvant être sensibles à d'autres classes que celles d'intérêt, d'autres pouvant être insensibles à certaines classes d'intérêt ... Ce type de modélisation peut être appliqué dans

les modèles de Markov. Les premières études, relativement récentes, concernent les champs ou chaînes de Markov cachés [8, 33, 34, 107]. En particulier, on montre que lorsque l'information a priori est donnée par une loi de probabilité markovienne et lorsque les senseurs sont indépendants conditionnellement à X , le résultat de la fusion de Dempster-Shafer est une loi de probabilité markovienne, qui redonne la loi classique a posteriori lorsque l'information fournie par les senseurs est de type probabiliste (dans l'exemple ci-dessus, la fusion en présence des nuages donne une loi markovienne, cette dernière devenant la loi markovienne a posteriori classique lorsque les nuages disparaissent) [8]. Par ailleurs, on peut également considérer le cas où l'information a priori est markovienne et « évidentielle » : on garde l'écriture markovienne, avec les fonctions de masse à la place des probabilités. La fusion de Dempster-Shafer d'une telle « fonction de masse markovienne » avec la probabilité $q(x) = \prod_{s \in S} q_s(x_s)$ (où, comme précédemment, q_s

est définie sur chaque pixels à partir de l'observation y_s par $q(x_s) \propto p(y_s|x_s)$) détruit alors la markovianité ; cependant, on montre que les segmentations bayésiennes sont possibles car le résultat de la fusion est un modèle de Markov Triplet [86]. Cela s'étend à un cas de m capteurs, dès qu'un au moins parmi eux est de type probabiliste. La probabilité q_s est dans ce cas obtenue par la fusion DS, sur chaque pixel, des masses (ou probabilités) q_s^1, \dots, q_s^m associées aux observations produites par les m capteurs. Enfin, signalons également la possibilité d'extension de ce type de démarche au cas des capteurs corrélés [83].

Exemple

Considérons l'exemple suivant extrait de [8]. On dispose d'une image radar, d'une image optique, et d'une vérité terrain avec quatre classes. L'image optique contient des nuages, qui constituent une cinquième classe sans intérêt. Les résultats de trois segmentations non supervisées (« pixel par pixel » après fusion DS de deux capteurs, markovienne classique à partir de la seule image radar, et markovienne après fusion DS de deux capteurs) sont présentés à la Figure 6. Il apparaît que la troisième méthode aboutit à un taux d'erreur le plus intéressant.



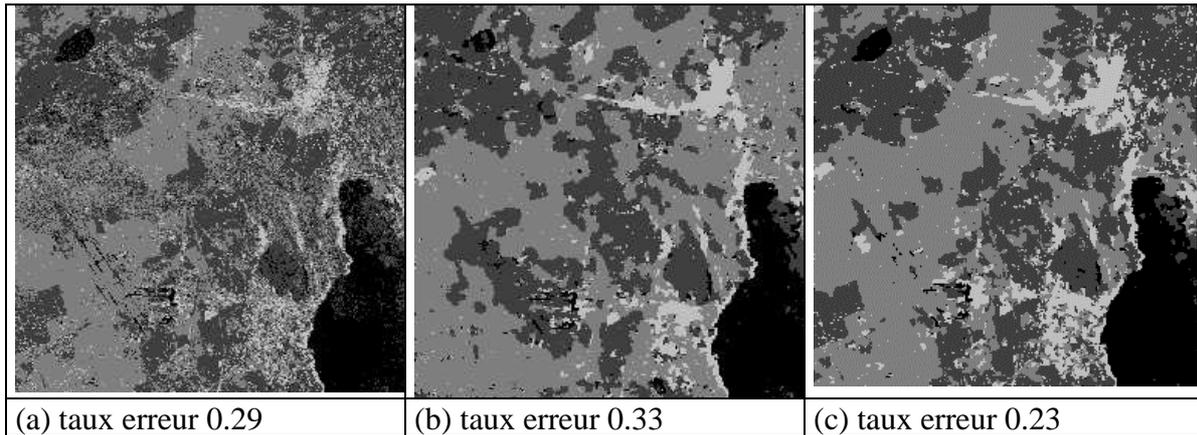


Figure 5

Images radar et optique, vérité terrain, segmentation « pixel par pixel » après fusion DS de deux capteurs (a), segmentation markovienne classique à partir de la seule image radar (b), et segmentation markovienne après fusion DS de deux capteurs (c).

6.5 Données incomplètes

Dans ce qui précède nous avons supposé tous les (Y_s) observés et tous les (X_s) recherchés, situation qui porte parfois le nom du problème des “données cachées”. La situation des “données incomplètes” plus générale, peut encore être traitée en utilisant les modèles de Markov. Dans une telle situation, supposons que X est observé sur $S_1 \subset S$ et Y est observé sur $S_2 \subset S$ (on retrouve la situation précédente avec $S_1 = \emptyset$ et $S_2 = S$). On peut alors mettre en place, dès que l’on dispose d’un modèle de Markov caché pour la loi de (X, Y) , des méthodes Bayésiennes d’estimation de X_{S-S_1} à partir de Y_{S_2} . Notons que cela est en particulier vrai pour $S_1 = \emptyset$ et $S_2 \neq S$: on peut ainsi estimer les réalisations des X_s même pour les pixels éloignés des éléments de S_2 sur lequel on observe les Y_t (rappelons que dans un champ de Markov $X = (X_s)_{s \in S}$ les variables X_t, X_u sont toujours dépendantes, quelles que soient les positions de t et u dans S).

6.6 Modèles de Markov flous

Nous avons implicitement admis qu’un pixel donné ne pouvait appartenir qu’à une seule classe dans l’ensemble $\Omega = \{\omega_1, \dots, \omega_k\}$. En considérant l’exemple de l’eau et la forêt, on peut imaginer, en imagerie satellitaire, l’existence des pixels contenant simultanément de l’eau et des arbres, sans qu’aucune des deux classes ne s’impose. On peut alors poser $X_s = 0$ si s est « eau », $X_s = 1$ si s est « forêt », et $X_s = x_s \in]0, 1[$ si la proportion de la surface de la forêt dans le pixel est $x_s \in]0, 1[$. On obtient une appartenance « floue » du pixel (on dit que le pixel est « purement flou » si $x_s \in]0, 1[$ et qu’il est « dur » si $x_s \in \{0, 1\}$) et on peut considérer cette appartenance comme aléatoire en définissant une loi de X_s sur $[0, 1]$. Une des possibilités est de définir cette loi par une densité par rapport à la mesure $\nu = \delta_0 + \delta_1 + \mu$, où δ_0, δ_1 sont des masses de Dirac en 0 et 1, et μ est la mesure de Lebesgue sur $]0, 1[$. On a alors $P_{X_s} = h\nu$, avec h positive telle que

$h(0) + h(1) + \int_0^1 h(t)dt = 1$. Notons que l'utilisation des masses de Dirac permet de retrouver

le modèle classique en considérant h nulle sur $]0,1[$. Un champ de Markov flou est alors obtenu en considérant un système de voisinage, l'ensemble C des cliques correspondant, et l'ensemble des fonctions $\varphi_c : [0,1]^{Card(c)} \rightarrow R$. Sa loi s'écrit $P_X = gV^{\otimes N}$, avec N le

cardinal de S et g densité, définie sur $[0,1]^N$, de la forme $g(x) = \gamma \exp\left[-\sum_{c \in C} \varphi_c(x_c)\right]$

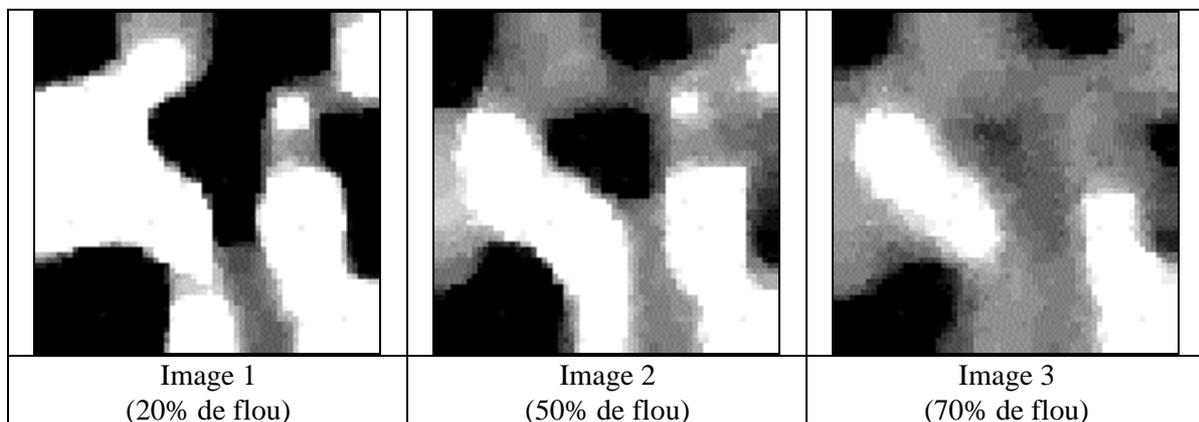
(écriture analogue à (3.1)). Si f_E et f_F sont les deux gaussiennes modélisant les bruits « eau » et « forêt », avec les moyennes m_E , et m_F et les variances σ_E^2 et σ_F^2 , on peut considérer (à titre d'exemple) que la densité f_{x_s} de la loi de Y_s conditionnelle à $X_s = x_s$ est une gaussienne de moyenne $m_{x_s} = (1-x_s)m_E + x_s m_F$ et de variance $\sigma_{x_s}^2 = (1-x_s)\sigma_E^2 + x_s \sigma_F^2$ (en procédant ainsi on retrouve les gaussiennes f_E et f_F lorsque le pixel considéré est dur). En définissant classiquement la loi de (X,Y) par $p(x,y) = g(x) \prod_{s \in S} f_{x_s}(y_s)$, on montre que la loi de X a posteriori est de Markov, ce qui

permet l'utilisation de l'échantillonneur de Gibbs et, par conséquent, rend possible la mise en œuvre de différentes stratégies bayésiennes (notons la grande richesse des possibilités pour les fonctions de perte).

De tels modèles de Markov « flous » ont été introduits dans [96], et des modèles semblables ont été récemment utilisés avec succès dans la segmentation floue des images de cerveau [95]. Notons l'utilisation des champs aléatoires cachés flous, dans le cadre des modélisations plus simples, sans masses de Dirac, est plus ancienne [56].

Nous présentons sur la Figure 7 quelques exemples visuels, extraits de [96], de simulations et de segmentation non supervisée (estimation des paramètres avec ICE) des champs de Markov flous. Par ailleurs, il est à noter que l'algorithme de segmentation ne confond pas, du moins dans des cas simples étudiés, le bruit avec le flou : la segmentation d' une image binaire (sans flou) donne une image "presque" binaire (avec très peu de flou) [96].

Ce type de démarche peut être également utilisé pour généraliser les chaînes et les arbres de Markov classiques aux chaînes et arbres de Markov flous sans difficulté autre que celle de la complexité algorithmique croissante. Un autre type de généralisation possible serait de considérer $\Omega = [0,1]^k$, sans que la somme des appartenances floues à chacune des classes soit nécessairement égale à un comme dans [95, 96].



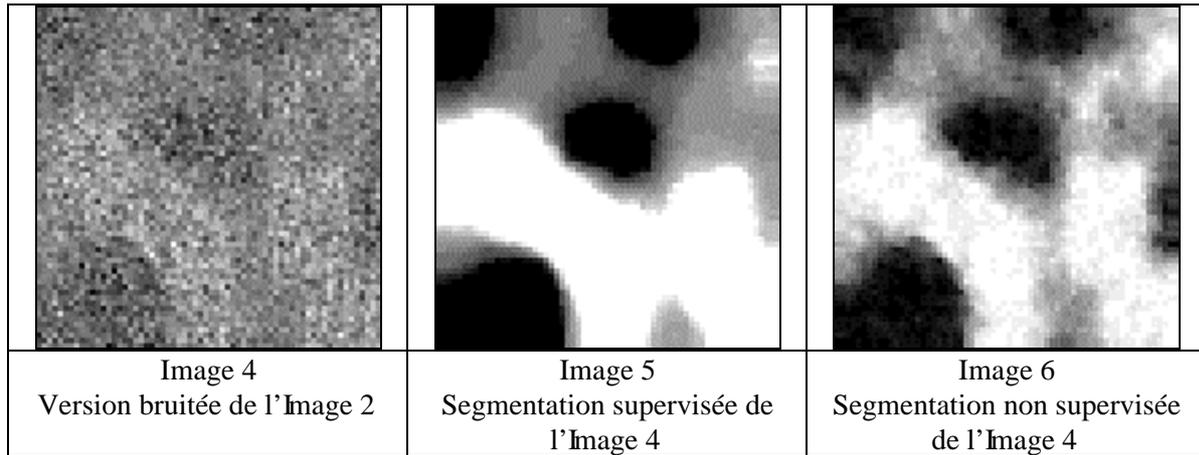


Figure 7

Simulations, par échantillonneur de Gibbs, des réalisations de champs de Markov flous (Images 1, 2, et 3). Version bruitée de l'Image 2 et les segmentations supervisée et non supervisée (respectivement Images 4, 5 et 6).

6.7 Modèles graphiques

Les champs, chaînes, et arbres de Markov sont des cas particuliers des « modèles graphiques » plus généraux. Pour un ensemble des pixels S on appelle graphe le couple $G = (S, A)$, où A est un ensemble d'arêtes (segments liant deux pixels). On peut alors traduire, de façon parfois élégante et facile à appréhender intuitivement, les différentes dépendances des variables aléatoires (X_s) indicées sur S en présence ou absence d'arêtes [7, 40, 41, 50, 102, 108]. Considérons, à titre d'exemple, le cas des réseaux bayésiens. On dit que le graphe est orienté si chaque arête est une flèche ; un graphe orienté est dit acyclique s'il n'est pas possible de revenir à un pixel en suivant des arêtes dans le sens des flèches. Un réseaux bayésien est un processus $X = (X_s)_{s \in S}$ tel que pour un graphe orienté acyclique $G = (S, A)$ sa loi s'écrit $p(x) = \prod_{s \in S} p(x_s | x_{pa(s)})$, avec $pa(s)$ l'ensemble, éventuellement vide, liés à s par une flèche orientée vers s ($pa(s)$ est dit « parents de s »). Lorsque chaque pixel, à l'exception de la racine, admet un parent unique on retrouve les arbres de Markov de la section 3.4, les chaînes de Markov étant un cas particulier des arbres, et donc un exemple très simple d'un réseau bayésien. On peut alors considérer que le processus X n'est pas observable et l'on observe un processus $Y = (Y_s)_{s \in S}$, dont la loi conditionnelle à X est donnée par $p(y|x) = \prod_{s \in S} p(y_s | x_s)$. La structure de réseau bayésien est alors conservée par la loi de X a posteriori $p(x|y)$ et le problème du calcul des différentes lois d'intérêt liées à $p(x|y)$, comme les lois marginales $p(x_s|y)$, peut être abordé. En absence de boucles, une boucle étant un parcours dans l'ensemble des pixels, commençant et finissant sur un pixel donné, et suivant les arêtes dans le sens des flèches ou dans le sens inverse, les solutions sont analogues à celles présentées dans les cas des chaînes et les arbres. En présence des boucles, les solutions sont plus compliquées : on doit procéder à la construction d'un nouveau réseau dont les pixels sont des ensembles des

pixels originels : de telles méthodes sont dites de « regroupement » [7]. Le nouveau réseau ainsi obtenu est, sous de bonnes conditions, sans boucles et les méthodes simples précédentes s'appliquent. Notons également que les champs de Markov sont des cas particuliers des processus sur graphes non orientés (il n'y a pas de flèches). Dans ces derniers, on peut s'intéresser à la « propagation de l'information » (« belief propagation » en anglais), laquelle est très souvent modélisée par une probabilité. Un certain nombre de résultats obtenus dans un cadre différent peut alors être transposé à la problématique de la segmentation des champs de Markov cachés ; citons, à titre d'exemple, l'utilisation d'une telle propagation pour approcher les lois marginales de X a posteriori [108]. Notons également divers résultats permettant de se ramener d'un graphe non orientés à un graphe orienté [102].

7. Conclusions et Perspectives

Les traitements statistiques d'images fondés sur des modèles Markoviens peuvent présenter des qualités exceptionnelles. L'avantage de ces modèles par rapport à des modèles « locaux » découle de leur aptitude à prendre en compte, de façon souvent élégante et mathématiquement rigoureuse, l'ensemble de l'information disponible. De plus, les diverses études semblent indiquer qu'une extraordinaire robustesse s'ajoute aux qualités classiques des méthodes statistiques que sont la souplesse et l'optimalité. Cette robustesse permet d'envisager des complexifications croissantes des modèles : séquences d'images, images 3D, multirésolutions, multisenseurs avec senseurs corrélés et non nécessairement Gaussiens, utilisation de la théorie de l'évidence et du flou, modèles de Markov Couple et Triplet,

Mentionnons, de manière non exhaustive, quelques perspectives se dégageant de cet article.

La production toujours croissante d'images nécessite la conception des méthodes toujours plus rapides, plus générales, plus automatiques. Une des perspectives pouvant être utile en situations complexes impliquant de grandes masses des données, comme séquences d'images 3D multirésolutions et multisenseurs à chaque résolution, est de considérer une suite des modèles de Markov de complexité croissante, chacun d'eux utilisant les résultats du modèles précédent comme initialisation.

Nous avons noté que les champs, chaînes, ou arbres de Markov présentés faisaient partie des « modèles graphiques » plus généraux, faisant actuellement l'objet d'une activité de recherche importante. Les divers résultats obtenus en dehors des préoccupations liées aux traitements d'images seront sans doute parmi les outils les plus stimulants dans la recherche des diverses généralisations des modèles utilisés en imagerie. Plus loin, les prolongements vers les réseaux de neurones (citons, à titre d'exemple, [109]) pourraient également constituer un axe intéressant de recherches dans les années à venir.

Références

- [1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, *Revue Scientifique et Technique de la Défense*, No. 11, pp. 27-40, 1991.
- [2] R. G. Aykroyd, Bayesian estimation for homogeneous and inhomogeneous Gaussian random fields, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 5, pp. 533 –539, 1998.

- [3] R. Azencott, *Simulated annealing : parallelization techniques*, Wiley, 1992.
- [4] C. Baillard and H. Maître, 3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy, *Computer Vision and Image Understanding*, Vol. 76, No. 3, pp. 244-258, 1999.
- [5] S. A. Barker and P. J. W. Rayner, Unsupervised image segmentation using Markov random field model, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lectures Notes in Computer Science, 1223, pp. 165-178, 1997.
- [6] R. J. Baxter, *Exactly solved models in statistical mechanics*, London, England: Academic, 1982.
- [7] A. Becker et P. Naïm, *Les réseaux bayésiens*, Eyrolles, Paris, 1999.
- [8] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 39, No. 8, pp. 1789-1798, 2001.
- [9] B. Benmiloud et W. Pieczynski, Estimation des paramètres dans les chaînes de Markov cachées et segmentation d' images *Traitement du Signal*, Vol. 12, No. 5, pp. 433-454, 1995.
- [10] M. Berthod, Z. Kato, Y. Shan, and J. Zerubia, Bayesian image classification using Markov random fields, *Image and Vision Computing*, Vol. 14, No. 4, pp. 285-295 1996.
- [11] J. Besag, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society*, Series B, 48, pp. 259-302, 1986.
- [12] A. Bonin, B. Chalmond, and B. Lavayssière, Monte-Carlo simulation of industrial radiography images and experimental designs, *NDT & E International*, Vol. 35, Issue 8, pp. 503-510, 2002.
- [13] J.-M. Boucher and P. Lena, Unsupervised Bayesian classification, application to the forest of Paimpont (Brittany), *Photo Interprétation*, Vol. 32, No. 1994/4, 1995/1-2, pp. 79-81, 1995.
- [14] E. Bratsolis and M. Sigelle. Image relaxation using Potts model with a fast deterministic method. *Journal of the Optical Society of America*, A, pp. 1033-1043, 1997.
- [15] L. Bruzzone and D. F. Prieto, Automatic analysis of the difference image for unsupervised change detection, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 38, No. 3, pp. 1171-1182, 2000.
- [16] G. Celeux, F. Forbes, and N. Peyrard, EM procedures using mean field-like approximations for Markov model-based image segmentation, *Pattern Recognition*, Vol. 36, No. 1, pp. 131-144, 2003.
- [17] B. Chalmond, An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recognition*, 22, pp. 747-761, 1989.
- [18] B. Chalmond, *Eléments de modélisation pour l'analyse d'images*, Société de Mathématiques Appliquées et Industrielles, Springer, 2000.
- [19] C. Collet, J.-N. Provost, P. Rostaing, P. Pérez, and P. Bouthemy. Segmentation bathymétrique d' images multispectrales SPOT, *Traitement du Signal*, Vol. 18, No. 1, pp. 1-16, 2001.
- [20] O. Coulon, J. -F. Mangin, J. -B. Poline, M. Zilbovicius, D. Roumenov, Y. Samson, V. Frouin and I. Bloch, Structural group analysis of functional activation maps, *NeuroImage*, Vol. 11, No. 6, pp. 767-782, 2000.
- [21] D. Dacuhna-Castelle et M. Duflo, *Probabilités et Statistiques, tome 2, Problèmes à temps mobile*, Masson, 1983.

- [22] Y. Delignon, A. Marzouki, and W. Pieczynski, Estimation of generalized mixtures and its application in image segmentation, *IEEE Trans. on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.
- [23] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, *IEEE Trans. on Signal Processing*, Vol. 45, No. 10, pp. 2613-2615, 1997.
- [24] S. Derrode, G. Mercier, J.-M. Le Caillec, and R. Garello, Estimation of sea-ice SAR clutter statistics from Pearson' s system of distributions, *Proceedings of International Geoscience and Remote Sensing Symposium, IGARSS '01*, Vol. 1, pp. 190-192, 2001.
- [25] S. Derrode and W. Pieczynski, SAR image segmentation using using generalized Pairwise Markov Chains, *Proceedings of SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.
- [26] X. Descombes, F. Kruggel, and D. Y. Von Cramon, Spatio-temporal fMRI analysis using Markov random fields, *IEEE Trans. on Medical Imaging*, Vol. 17, No. 6, pp. 1028-1039, 1998.
- [27] X. Descombes, M. Moctezuma, H. Maître and J-P. Rudant, Coastline detection by a Markovian segmentation on SAR images, *Signal Processing*, Vol. 55, Issue 1, pp. 123-132, 1996.
- [28] X. Descombes, M. Sigelle, and F. Preteux, Estimating Gaussian Markov random field parameters in a nonstationary framework: application to remote sensing imaging , *IEEE Trans. on Image Processing*, Vol. 8 No. 4, pp. 490 –503, 1999.
- [29] A. P. Dunmur and D. M. Titterington, Mean fields and two-dimensional Markov random fields in image analysis, *Pattern Analysis and Applications*, 1, pp. 248-260, 1998.
- [30] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 41, No. 3, pp. 675-686, 2003.
- [31] F. Forbes and N. Peyrard, Hidden Markov model selection criteria based on mean-like field approximations, INRIA Rhône-Alpes, Research Report No. 4371, 2002.
- [32] G. D. Fornay, The Viterbi algorithm, *Proceedings of the IEEE*, Vol. 61, No. 3, pp. 268-277, 1973.
- [33] S. Foucher, M. Germain, J.-M. Boucher, and G. B. Benié, Multisource classification using ICM and Dempster–Shafer theory, *IEEE Trans. on Instrumentation and Measurement*, Vol. 51, No. 2, pp. 277-281, 2002.
- [34] L. Fouque, A. Appriou, and W. Pieczynski, An evidential Markovian model for data fusion and unsupervised image classification, *Proceedings of 3rd International Conference on Information Fusion, FUSION 2000*, Vol. 1, July 10th-13th, 2000, Paris, France, pp. TuB4-25 - TuB4-31.
- [35] L. Fouque, A. Appriou, and W. Pieczynski, Multiresolution hidden Markov chain model and unsupervised image segmentation, *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'2000)* , 2-4April 2000, Austin, Texas, United States, pp. 121-125.
- [36] Gelgon and P. Bouthemy, A region-level motion-based graph representation and labeling for tracking a spatial image partition, *Pattern Recognition*, Vol. 33, No. 4, pp. 725-740, 2000.
- [37] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721-741, 1984.

- [38] D. Geman, S. Geman, C. Graffigne, and P. Dong, Boundary detection by constrained optimization, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, pp. 609 –628, 1990.
- [39] S. Geman and C. Graffigne, Markov random field image models and their applications to computer vision, *In Proceedings of the International Congress of Mathematicians*, Berkeley, 1987.
- [40] Z. Ghahramani and M. Beal, Graphical models and variational methods, In *Advances in mean field methods: Theory and Practice*, MIT Press, 2000.
- [41] Z. Ghahramani, An introduction to hidden Markov models and Bayesian networks, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 15, No. 1, pp. 9-42, 2001.
- [42] N. Giordana and W. Pieczynski, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 465-475, 1997.
- [43] C. Graffigne, F. Heitz, P. Pérez, F. Preteux, M. Sigelle, and J. Zerubia, Hierarchical Markov random field models applied to image analysis: a review, *In SPIE Conference on neural, morphological, stochastic methods in image and signal processing*, 1995.
- [44] X. Guyon, *Random fields on network. modeling, statistics, and applications*. Springer-Verlag, Probability and its Applications, New York, 1995.
- [45] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 12, pp. 1217-1232, 1993.
- [46] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, T. G. Gilland, and R. J. Jaszczak, Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data, *IEEE Trans. on Medical Imaging*, Vol. 16, No. 5, pp. 516 –526, 1997.
- [47] L. Hubert-Moy, A. Cotonnec, L. Le Du, A. Chardin, and P. Pérez. A comparison of parametric classification procedures of remotely sensed data applied on different landscape units, *Remote Sensing Environment*, Vol. 75, No. 2, pp. 174-187, 2001.
- [48] J. Idier, *Approches bayésiennes pour les problèmes inverses*, Traitement du Signal et de l'Image, Information – Commande – Communication, ouvrage collectif, Hermès, 2001.
- [49] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, Hyperparameter estimation for satellite image restoration using a MCMC maximum-likelihood method, *Pattern Recognition*, Vol. 35, No. 2, pp. 341-352, 2002.
- [50] F. V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, 2000.
- [51] C. Ji and L. Seymour, A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood, *Annals of Applied Probability*, 6, pp. 423-443, 1996.
- [52] Z. Kato, M. Berthod, and J. Zerubia, A hierarchical Markov random field model and multitemperature annealing for parallel image classification, *Graphical Models and Image Processing*, Vol. 58, No. 1, pp. 18-37, 1996.
- [53] Z. Kato, Bayesian color image segmentation using Reversible Jump Markov Chain Monte Carlo, Rapport de recherche ERCIM R005, 1999.
- [54] Z. Kato, J. Zerubia, and M. Berthod, Unsupervised parallel image classification using Markovian models, *Pattern Recognition*, Vol. 32, pp. 591-604, 1999.
- [55] P. A. Kelly, H. Derin, and K. D. Hartt, Adaptive segmentation of speckled images using a hierarchical random field model, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 10, pp. 1628-1641, 1988.

- [56] J. T. Kent and K. V. Mardia, Spatial classification using fuzzy membership, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 5, pp. 659-671, 1991.
- [57] C. Kervrann and F. Heitz, A hierarchical Markov modeling approach for the segmentation and tracking of deformable shapes, *Graphical Models and Image Processing*, Vol. 60, No. 3, pp.173-195, 1998.
- [58] J.-M. Laferté, F. Heitz, and P. Pérez, Modèles markoviens d'images et algorithmes d'estimation non linéaire sur le quadarbre, *Traitement du Signal*, Vol. 15, No. 3, pp. 213-230, 1998.
- [59] J.-M. Laferté, P. Pérez, and F. Heitz, Discrete Markov image modeling and inference on the quadtree, *IEEE Trans. on Image Processing*, Vol. 9, No 3, pp. 390-404, 2000.
- [60] S. Lakshmanan, and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 8, pp. 799-813, 1989.
- [61] D. Landgrebe, Hyperspectral image data analysis, *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp. 17-28, 2002.
- [62] H. Maître, *Traitement des images de RSO*, Traitement du Signal et de l'Image, Information – Commande – Communication, ouvrage collectif, Hermès, 2001.
- [63] J. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.
- [64] G. J. McLachlan and T. Krishnan, *EM Algorithm and Extensions*, Wiley, 1997.
- [65] D. Melas and S. P. Wilson, Double Markov random fields and Bayesian image segmentation, *IEEE Trans. on Signal Processing*, Vol. 50, No. 2, pp. 357-365, 2002.
- [66] M. Mignotte, C. Collet, P. Pérez and P. Bouthemy, Three-Class Markovian segmentation of high-resolution sonar images, *Computer Vision and Image Understanding*, Vol. 76, No. 3, pp. 191-204, 1999.
- [67] M. Mignotte, C. Collet, P. Pérez, and P. Bouthémy, Sonar image segmentation using an unsupervised hierarchical MRF model, *IEEE Trans. on Image Processing*, Vol. 9, No. 7, pp. 1216-1231, 2000.
- [68] M. Mignotte, J. Meunier, J.-P. Soucy, and C. Janicki. Comparison of deconvolution techniques using a distribution mixture parameter estimation: Application in single photon emission computed tomography imagery, *Journal of Electronic Imaging*, Vol. 11, No. 1, pp. 11-25, 2001.
- [69] M. Mignotte, C. Collet, P. Pérez and P. Bouthemy, Markov random field and fuzzy logic modeling in sonar imagery: application to the classification of underwater floor, *Computer Vision and Image Understanding*, Vol. 79, No. 1, pp. 4-24, 2000.
- [70] M. Mignotte and J. Meunier, Three-dimensional blind deconvolution of SPECT images, *IEEE Trans. on Biomedical Engineering*, Vol. 47, No. 2, pp. 274-280, 2000.
- [71] R. Morris, X Descombes, et J. Zerubia, Fully Bayesian image segmentation-an engineering perspective, *Proceedings of International Conference on Image Processing (ICIP'97)*, Vol. 3, pp. 26-29, 1997.
- [72] M. Nikolova, J. Idier, and A. Mohammad-Djafari, Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF, *IEEE Trans. on Image Processing*, Vol. 7, No. 4, pp. 571-585, 1998.
- [73] J.-M. Odobez and P. Bouthemy, Direct incremental model-based image motion segmentation for video analysis, *Signal Processing*, Vol. 66, No. 2, pp. 143-155, 1998.
- [74] S. K. Pal and P. Mitra, Multispectral image segmentation using the rough-set-initialized EM algorithm, *IEEE Trans. on Geoscience and Remote Sensing*, to appear.

- [75] C. Pellot, A. Herment, M. Sigelle, P. Horain, H. Maitre, and P. Peronneau, A 3D reconstruction of vascular structures from two X-ray angiograms using an adapted simulated annealing algorithm, *IEEE Trans. on Medical Imaging*, Vol. 13, No. 1, pp. 48-60, 1994.
- [76] P. Pérez and F. Heitz, Restriction of a Markov random field on a graph and multiresolution statistical image modeling, *IEEE Trans. on Information Theory*, Vol. 42, pp. 180-190, 1996.
- [77] P. Pérez. Markov random fields and images. *CWI Quarterly*, Vol. 11, No. 4, pp. 413-437, 1998.
- [78] P. Pérez, A. Chardin, and J.-M. Laferté, Noniterative manipulation of discrete energy-based models for image analysis, *Pattern Recognition*, Vol. 33, No. 4, pp. 573-586, 2000.
- [79] W. Pieczynski, Statistical image segmentation, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.
- [80] W. Pieczynski, Champs de Markov cachés et estimation conditionnelle itérative, *Traitement du Signal*, Vol. 11, No. 2, pp. 141-153, 1994.
- [81] W. Pieczynski, J. Bouvrais, and C. Michel, Estimation of generalized mixture in the case of correlated sensors, *IEEE Trans. on Image Processing*, Vol. 9, No. 2, pp. 308-311, 2000.
- [82] W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Machine Graphics and Vision*, Vol. 9, No. 3, pp. 705-718, 2000.
- [83] W. Pieczynski, Unsupervised Dempster-Shafer fusion of dependent sensors, *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'2000)*, Austin, Texas United States, pp. 247-251, 2-4 April 2000.
- [84] W. Pieczynski, Pairwise Markov chains, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 5, pp. 634-639, 2003.
- [85] W. Pieczynski, Arbres de Markov Couple, *Comptes Rendus de l'Académie des Sciences – Mathématiques*, Paris, Ser. I 335, pp. 79-82, 2002.
- [86] W. Pieczynski, Chaînes de Markov Triplet, *Comptes Rendus de l'Académie des Sciences – Mathématiques*, Paris, Ser. I 335, pp. 275-278, 2002.
- [87] W. Pieczynski, D. Benboudjema, and P. Lanchantin, Statistical image segmentation using Triplet Markov Fields, *SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.
- [88] G. Potamianos and J. Goutsias, Stochastic approximation algorithm for partition function estimation of Gibbs random fields, *IEEE Trans. on Information Theory*, Vol. 43, No. 6, pp. 1948-1968, 1997.
- [89] W. Qian, and D. M. Titterington, Stochastic relaxations and EM algorithms for Markov random fields, *J. Statist. Comput. Simulation*, 40, pp. 55-69, 1992.
- [90] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, Analyse de texture hyperspectrale par modélisation markovienne, Rapport de Recherche INRIA 4479, Sophia Antipolis, Projet Ariana, juin 2002.
- [91] G. Rellier, X. Descombes, and J. Zerubia, Local registration and deformation of a road cartographic database on a SPOT satellite image, *Pattern Recognition*, Vol. 35, No. 10, pp. 2213-2221, 2002.
- [92] C. P. Robert, *Méthodes de Monte Carlo par chaînes de Markov*, Economica, Paris, 1996.
- [93] P. Rostaing, J.-N. Provost, and C. Collet, Unsupervised Multispectral Image Segmentation using Generalized Gaussian Noise Model, *Proceedings of Energy*

- Minimization Methods in Computer Vision and Pattern Recognition*, York, England, pp. 142-156, July 1999.
- [94] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet, Brain tissue classification of magnetic resonance images using partial volume modeling, *IEEE Trans. on Medical Imaging*, Vol. 19, No. 12, pp. 1179-1187, 2000.
- [95] S. Ruan, B. Moretti, J. Fadili, and D. Bloyet, Fuzzy Markovian segmentation in application of magnetic resonance images, *Computer Vision and Image Understanding*, 85, pp. 54-69, 2002.
- [96] F. Salzenstein and W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, *Graphical Models and Image Processing*, Vol. 59, No. 4, pp. 205-220, 1997.
- [97] F. Salzenstein et W. Pieczynski, Sur le choix de méthode de segmentation statistique d' images *Traitement du Signal*, Vol. 15, No. 2, pp. 119-128, 1998.
- [98] S. M. Schweizer and J. M. F. Moura, Efficient detection in hyperspectral imagery, *IEEE Trans. on Image Processing*, Vol. 10, No. 4, pp. 584- 597, 2001.
- [99] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, 1976.
- [100] W. Skarbek, Generalized Hilbert scan in image printing, *Theoretical Foundations of Computer Vision*, R. Klette and W. G. Kropetsh, editors, pp. 45-57, Akademik Verlag, 1992.
- [101] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, Vol. 9, pp. 1-35, 1993.
- [102] P. Smyth, D. Heckerman, M. Jordan, Probabilistic independence networks for hidden Markov probability models, *Neural Computation*, 9, pp. 227-269, 1997.
- [103] D. Stanford, Fast automatic unsupervised image segmentation and curve detection in spatial point processes, PhD thesis, Department of Statistics, University of Washington, Seattle, 1999.
- [104] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, Anomaly detection from hyperspectral imagery, *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp. 58-69, 2002.
- [105] T. Szirányi, J. Zerubia, L. Czúni, D. Geldreich and Z. Kato, Image segmentation using Markov random field model in fully parallel cellular network architectures, *Real-Time Imaging*, Vol. 6, No. 3, pp. 195-211, 2000.
- [106] F. Tupin, H. Maitre, J.-F. Mangin, J.-M. Nicolas, and E. Pechersky, Detection of linear features in SAR images: application to road network extraction, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 36 No. 2, pp. 434 –453, 1998.
- [107] F. Tupin, H. Maitre, and I. Bloch, A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 37, No. 3, pp. 1327-1343, 1999.
- [108] Y. Weiss, Comparing the mean field method and belief propagation for approximate inference in MRFs, *In Advances in mean field methods: Theory and Practice*, MIT Press, 2000.
- [109] P. Wilinski, B. Solaiman, A. Hillion, and W. Czarnecki, Toward the border between neural and Markovian, Systems, *IEEE Trans. on Man and Cybernetics*, Part B, Vol. 28 No. 2, pp. 146 –159, 1998.
- [110] A. S. Willsky, Multiresolution Markov models for signal and image processing, *Proceedings of IEEE*, Vol. 90, No. 8, pp. 1396-1458, 2002.

- [111] G. Winkler, *Image analysis, random fields and and Markov Chain Monte Carlo Methods: a mathematical introduction*, Springer, 2003.
- [112] J.-H. Xue, S. Ruan, B. Moretti, M. Revenu, and D. Bloyet, Knowledge-based segmentation and labeling of brain structures from MRI images, *Pattern Recognition Letters*, Vol. 22, No. 3-4, pp. 395-405, 2001.
- [113] K. C. Yao, M. Mignotte, C. Collet, P. Galerne and G. Burel, Unsupervised segmentation using a self-organizing map and a noise model estimation in sonar imagery, *Pattern Recognition*, Vol. 33, No. 9, pp. 1575-1584, 2000.
- [114] L. Younes, Estimation and annealing for gibbsian fields, *Annales de l'Institut Henri Poincaré*, Vol. 24, No. 2, pp. 269-294, 1988.
- [115] L. Younes, Parametric inference for imperfectly observed gibbsian fields, *Probability Theory and Related Fields*, 82, pp. 625-645, 1989.
- [116] L. Younes, Synchronous random fields and image restoration, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 10, pp. 2570-2583, 1998.
- [117] Y. Yu and Q. Cheng, MRF parameter estimation by an accelerated method, *Pattern Recognition Letters*, Vol. 24, No. 9-10, pp. 1261-1269, 2003.
- [118] J. Zhang, The mean field theory in EM procedures for Markov random fields, *IEEE Trans. on Signal Processing*, Vol. 40, No. 10, pp. 2570-2583, 1992.
- [119] J. Zhang, J. W. Modestino, and D. A. Langan, Maximum likelihood parameter estimation for unsupervised stochastic model-based image segmentation, *IEEE Trans. on Image Processing*, Vol. 3, No. 4, pp. 404-420, 1994.
- [120] M. Zribi and F. Ghorbel, An unsupervised and non-parametric Bayesian classifier, *Pattern Recognition Letters*, Vol. 24, No. 1-3, pp. 97-112, 2003.