

Modèles markoviens d'images et algorithmes d'estimation non linéaire sur le quadarbre

Markov Image Models and Non-Linear Estimation Algorithms on the Quad-Tree

par Jean-Marc LAFERTÉ*, Fabrice HEITZ**, Patrick PÉREZ*

* IRISA / INRIA Campus de Beaulieu, 35042 Rennes Cedex

perez@irisa.fr

** LSIIT - UPRES-A CNRS 7005 / Université Strasbourg I, ENSPS, Bd Sébastien Brant, 67400 Illkirch

Fabrice.Heitz@ensps.u-strasbg.fr

résumé et mots clés

Les modèles markoviens spatiaux non causaux utilisés dans le domaine de l'analyse d'images conduisent à des algorithmes d'estimation itératifs, réputés pour leur complexité calculatoire. Dans cet article, nous considérons une classe de processus markoviens hiérarchiques et non linéaires, définis sur le quadarbre. Ces modèles markoviens présentent des propriétés de causalité en échelle qui permettent de construire des algorithmes non itératifs exacts, similaires à ceux existant pour les chaînes de Markov, dans le cas des signaux mono-dimensionnels. Nous présentons ainsi une version originale de l'algorithme de Viterbi sur le quadarbre, associé à une estimation exacte au sens du Maximum A Posteriori (MAP), ainsi que deux autres algorithmes d'estimation, respectivement associés à un critère du MAP modifié et au critère du MPM (mode de la marginale *a posteriori*). Deux nouveaux algorithmes EM, permettant une estimation non supervisée sur le quadarbre, sont également introduits pour cette classe de représentations.

Les propriétés de ces modèles et algorithmes sont illustrées et comparées, pour un problème simple de classification d'images, sur des images synthétiques et réelles.

Modèles hiérarchiques d'images, modèles markovien causaux, quadarbre, algorithmes non itératifs, classification supervisée et non supervisée

abstract and key words

Non-causal Markov Random Field (MRF) models are now widely used for representing images, but are known to yield iterative, often computational intensive, estimation algorithms. In this paper we consider a special class of Markov models which allow to circumvent the latter drawback : MRF attached to the nodes of a quad-tree. The specific structure of these models results in an appealing causality property through scale, which allows the design of exact, non-iterative inference algorithms which are similar to those used in the context of Markov chain models. We first introduce an original extension of the Viterbi algorithm for the exact computation of Maximum A Posteriori (MAP) estimates, along with two other algorithms respectively related to a modified MAP estimator, and to the Marginal Posterior Mode (MPM) estimator. The estimation of the parameters of the model is then addressed with two original Expectation-Maximization (EM)-type algorithms, allowing an unsupervised use of these models. The practical relevance of the different algorithms is investigated in the context of a standard image classification problem, both on synthetic and natural images.

Hierarchical image modeling, causal Markov models, quad-trees, non iterative inference, supervised and unsupervised image classification.

1. introduction

Les modèles markoviens non causaux spécifiés pour la représentation des processus spatiaux, en particulier dans le domaine de l'analyse d'images, sont usuellement définis sur des graphes correspondant à des grilles bi-dimensionnelles régulières. Ces représentations, abondamment exploitées comme modèles *a priori* dans le cadre classique de l'estimation statistique bayésienne et de la théorie de la régularisation, conduisent à des algorithmes d'estimation itératifs (relaxation déterministe ou stochastique), réputés pour leur complexité calculatoire.

D'autres structures de graphes ont été récemment considérées pour définir des modèles markoviens de processus aléatoires [16], parmi lesquelles l'arbre dyadique, (dans le cas des signaux monodimensionnels [2,12]) et le quadarbre, dans le cas des signaux bidimensionnels et des images [8, 17, 32, 33, 34]. Les processus aléatoires mis en jeu sont indicés par les sommets de l'arbre et possèdent la propriété d'être markoviens en échelle (et non plus spatialement). Le quadarbre permet, comme la grille bidimensionnelle régulière, de définir des modèles markoviens possédant des propriétés régularisantes, les interactions statistiques se définissant dans ce cas de façon hiérarchique plutôt que spatialement.

Le recours à un graphe hiérarchique tel que l'arbre est motivé par les considérations suivantes :

- la structure de l'arbre conduit de façon naturelle à des modèles causaux (en échelle) et à des algorithmes d'estimation *non itératifs*, qui s'apparentent à ceux développés dans le cas des chaînes de Markov ;

- comme l'ont montré Luetgen *et al.* dans [33], tout modèle markovien spatial (i.e., défini sur une grille 2D) peut être construit à partir d'un modèle en échelle défini sur un quadarbre¹. Les modèles sur les quadarbres, définissent donc, en ce sens, une sur-classe des modèles spatiaux usuels.

Ces modèles constituent, en particulier, une alternative intéressante aux modèles markoviens causaux plus classiques tels que les réseaux de Markov ou les champs de Pickard [13, 15], qui ont pour inconvénient de privilégier certaines orientations spatiales particulières. Ils définissent également un cadre plus général que les chaînes de Markov 1D, qui ont parfois été utilisées (dans le cadre de parcours de Hilbert-Peano) pour modéliser des images [6].

L'analogie entre modèles hiérarchiques sur le quadarbre et modèles markoviens 1-D causaux (chaînes de Markov) n'a toutefois pas été entièrement exploitée jusqu'à maintenant, et seuls des modèles hiérarchiques linéaires [2, 18, 27, 32] ou des modèles non linéaires, associés à des algorithmes d'estimation approchés [8] ont, à notre connaissance, été décrits.

Dans cet article nous présentons un ensemble d'algorithmes originaux, permettant l'estimation statistique (supervisée et non supervisée) par des modèles markoviens non linéaires définis sur le quadarbre. Plus précisément, dans le cas de l'estimation

supervisée (cas où les paramètres du modèle sont supposés connus) nous avons construit les algorithmes suivants :

- construction d'un algorithme de Viterbi [19] sur le quadarbre, permettant une estimation non itérative exacte au sens du MAP ;

- dérivation d'un second algorithme d'estimation, reposant sur un critère du MAP modifié (le SMAP), mieux approprié pour la structure du quadarbre. Cet algorithme approché vient généraliser l'algorithme proposé par Bouman *et al.* dans [8] ;

- dérivation d'un algorithme exact non itératif, pour une estimation au sens du mode de la marginale a posteriori (MPM), s'inspirant de l'algorithme de Baum « Forward-Backward » pour une chaîne de Markov [4].

Nous considérons également le problème de l'estimation non supervisée à travers l'algorithme EM, dont une version exacte sur le quadarbre ainsi qu'une extension stochastique, de type MCEM (Monte Carlo EM), sont définies. L'algorithme MCEM permet, en particulier, de s'affranchir très largement des problèmes d'initialisation posés par l'algorithme EM classique.

L'article est organisé comme suit : un bref état de l'art sur les modèles markoviens hiérarchiques est présenté dans la partie 2. La partie 3 définit les notations et pose les hypothèses de dépendances statistiques utilisées pour la définition des modèles markoviens causaux sur le quadarbre. L'algorithme de Viterbi (associé au critère du MAP), une extension de l'algorithme de Bouman (associé au critère du SMAP) ainsi qu'un algorithme de type « Forward-Backward » associé au MPM, sont successivement décrits en partie 4. Ces différents algorithmes sont comparés expérimentalement, pour un modèle simple de classification d'images. La partie 5 est dédiée aux algorithmes de type EM pour la classification non supervisée sur le quadarbre. Nous les comparons, au travers d'exemples synthétiques et réels, à des algorithmes EM qui ont été proposés dans le cas de modèles markoviens spatiaux classiques (EM gibbsien en particulier [10]).

2. modèles markoviens sur des graphes hiérarchiques

Les modèles considérés dans cet article se rattachent aux modèles markoviens hiérarchiques [22, 23], qui ont suscité une activité de recherche soutenue depuis la fin des années quatre-vingts. Le développement de ces modèles, qui s'est réalisé suivant des approches variées [23], a été dicté par les motivations suivantes [2] :

- nécessité en analyse d'images s'appuyant sur des modèles statistiques de disposer, comme en analyse numérique [41], de techniques multigrilles rapides et peu sensibles aux minima locaux ;

- besoin de modéliser, dans un cadre stochastique, des données intrinsèquement multirésolutions ou hiérarchisées (signaux fractals, phénomènes de turbulences, etc.) ;

- multiplication, en particulier en imagerie satellitaire [17], des capteurs, qui engendrent des volumes croissants de données

1. Notons toutefois que cette construction peut être arbitrairement complexe.

multispectrales, souvent définies à des résolutions différentes. Se pose alors le problème de définir des modèles (statistiques) aptes à fusionner des images de résolutions variables [29].

Les approches développées dans ce domaine se rattachent aux familles de méthodes suivantes (on trouvera dans [23] une classification, sur des bases mathématiques, des différentes approches ainsi qu'un descriptif plus détaillé de leurs caractéristiques) :

- approches par groupe de renormalisation ou restriction d'un modèle sur un sous-graphe [14, 21, 37];
- approches multigrilles par sous-espaces contraints [25, 27, 28, 36];
- modèles markoviens définis sur des graphes hiérarchiques (quadarbres [8, 29, 32] ou graphes pyramidaux [26]);
- familles hiérarchisées de modèles identiques [11];
- familles hiérarchisées multi-modèles [1, 38, 40].

Les modèles hiérarchiques que nous considérons dans cet article appartiennent à la famille des modèles markoviens définis sur des graphes hiérarchiques. Dans cette famille d'approches, un modèle markovien unique est défini sur un graphe hiérarchisé, *i.e.*, partitionné en échelles. Ceci conduit, contrairement aux autres approches, à la résolution d'un unique problème d'estimation statistique, sur le graphe hiérarchique.

Le graphe hiérarchique, que nous considérons ici est le quadarbre, qui se caractérise par sa simplicité et par les « bonnes » propriétés (causalité en particulier) qu'il induit sur les modèles qui lui sont associés. Les processus indicés sur ce graphe sont en particulier markoviens en échelle. Chou *et al.* [12] ont ainsi proposé, pour des signaux mono-dimensionnels, un cadre de modélisation *linéaire* (gaussien) sur l'arbre dyadique. Ce modèle est associé à des algorithmes d'estimation non-itératif, qui s'identifient à des schémas de filtrage de Kalman en échelle. Ces algorithmes permettent d'obtenir une estimée exacte, pour ce modèle linéaire, en deux passes sur l'arbre. Des modèles linéaires autorégressifs ont par ailleurs été étudiés, sur la même structure de graphe, par Basseville *et al.* [3]. Une extension du modèle de Chou *et al.* a été étudiée par Luetgen *et al.* [32, 33, 34] pour le quadarbre et appliquée à différents problèmes d'analyse d'images (estimation du mouvement [32], estimation de la profondeur des océans [17], etc.). Cette extension s'appuie toujours sur un modèle gaussien (donc linéaire). Bouman *et al.* ont proposé dans [8] un cadre de modélisation non linéaire sur le quadarbre, associé à un algorithme non itératif *approché*, pour la classification et la segmentation d'images [8]. Le critère d'estimation utilisé (le SMAP) est une extension originale du critère du MAP, mieux adaptée à la structure hiérarchique du modèle. L'algorithme associé au critère du SMAP repose également sur une exploration du quadarbre en deux passes, donnant une estimée approchée du processus.

Notons que les modèles markoviens définis sur le quadarbre présentent l'inconvénient d'introduire des artefacts (sous forme d'effets de blocs dans les estimées), comme l'ont noté de nombreux auteurs [12, 17, 29, 32]. Ces effets de bloc peuvent être atténués efficacement par des techniques heuristiques proposées récemment, (lissage *a posteriori* [32], arbres entrelacés [18], etc.). Ce point n'a toutefois pas été abordé dans notre étude.

La considération d'autres structures de graphes hiérarchiques permet également de pallier le problème des effets de blocs,

souvent au détriment du coût calculatoire. Bouman *et al.* considère ainsi dans [8] une structure de graphe déduite du quadarbre en introduisant des liens supplémentaires entre échelles, qui permettent, au moins partiellement, de briser la structure de blocs. Ce nouveau graphe conduit théoriquement à des algorithmes d'estimation itératifs, mais un estimateur approché non itératif est proposé. Kato *et al.* [26] définissent par ailleurs un modèle markovien sur un graphe plus complexe qui n'est autre que le quadarbre auquel on adjoint, à chaque échelle un système de voisinage spatial, déduit de la technique des sous-espaces contraints [36]. La perte de causalité du modèle défini sur cette structure de graphe, conduit à des algorithmes de relaxation itératifs, qui requièrent en pratique beaucoup plus d'itérations par pixel que les modèles spatiaux. Kato a donc proposé un recuit simulé modifié par rapport au recuit simulé standard. Il s'agit d'associer une température plus élevée aux échelles supérieures, celles-ci influant fortement ensuite sur les autres échelles. Ce type de recuit s'avère plus rapide que le recuit classique pour des résultats similaires. Ce modèle conduit à de très bons résultats en classification, mais demeure néanmoins très coûteux. Le bénéfice des approches hiérarchiques, en ce qui concerne le gain de temps, est donc ici perdu.

Nous présentons dans la suite un ensemble d'algorithmes qui conduisent, en des temps comparativement très courts, à des estimées sur le quadarbre (exactes pour le MAP et le MPM), qui, peuvent être considérées comme satisfaisantes dans de nombreuses applications.

Avant de décrire ces algorithmes, nous introduisons tout d'abord, dans la partie suivante, les notations et hypothèses structurelles concernant le modèle markovien non linéaire sur le quadarbre.

3. modèle markovien sur le quadarbre

3.1. notations

On considère le problème classique qui consiste à estimer un ensemble de variables dites cachées X , à partir d'observations Y , selon des critères que l'on précisera plus loin. X et Y sont des processus aléatoires indicés par les sommets d'un quadarbre (voir Figure 1), qui possède la propriété de ne pas présenter de cycles. Des graphes d'interaction plus complexes peuvent bien entendu être considérés, mais on s'aperçoit que l'introduction de cycles dans le graphe conduit à des algorithmes de relaxation (itératifs) en raison des interactions multiples entre sommets.

On désigne par S l'ensemble des sommets de l'arbre. On note s^- l'unique « père » d'un site s , l'ensemble des « fils » de t est appelé t^+ , $t^+ \triangleq \{s : t = s^-\}$. Plus généralement, $\mathcal{D}^0(s) \triangleq \{s\}$ et pour $n \geq 1$, $\mathcal{D}^n(s) \triangleq \cup_{t \in \mathcal{D}^{n-1}(s)} t^+$, *i.e.*, $\mathcal{D}^n(s)$ représente l'ensemble des descendants de s à la $n^{\text{ième}}$ génération. Par ailleurs l'ensemble $> s \triangleq \cup_{n>0} \mathcal{D}^n(s)$ désigne la « descendance stricte » de s (s non inclus) ou le « futur strict » de s . De même $\geq s$ désigne la

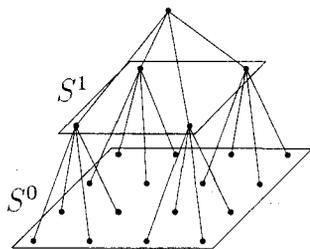


Figure 1. – Structure du modèle hiérarchique : le quadarbre.

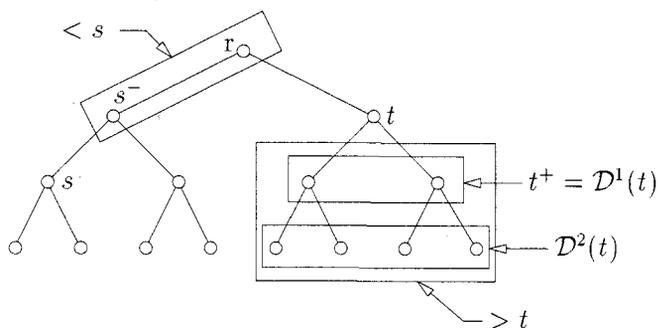


Figure 2. – Notations utilisées sur le quadarbre.

« descendance large » de s (s inclus), *i.e.*, $\geq s = \cup_{n \geq 0} \mathcal{D}^n(s)$. On définit de même l'ensemble des ancêtres de s par $< s \triangleq \{t : \exists n > 0, s \in \mathcal{D}^n(t)\} = \{s^-, s^{--}, \dots, r\}$, r désignant la racine. Les sommets de l'arbre appartenant à la même « génération » forment un ensemble noté S^n où n représente l'échelle, l'échelle 0 correspondant à la résolution la plus fine et l'échelle R (résolution dite la plus grossière) correspondant au singleton $S^R = \{r\}$ appelé « racine » de l'arbre (voir Figure 2). La restriction de X à S^n est notée $X^n \triangleq \{X_s, s \in S^n\}$, où X_s représente l'étiquette en un site s .

On définit de même les ensembles d'observations Y^n (et les variables Y_s). Les variables Y_s sont usuellement définies à la résolution la plus fine (échelle 0), mais peuvent également exister à d'autres niveaux de résolution, dans le cas d'observations multirésolutions [29]. Les observations Y^n lorsqu'elles existent, correspondent donc à une image (ou plusieurs images, dans le cas multispectral) à l'échelle n . Dans le cas général Y_s est donc un vecteur (Y^n est alors un d -uplet d'images définies à la même résolution). X_s peut de même être vectoriel.

3.2. hypothèses sur le modèle statistique

On considère le processus $X = (X^n)_{n=0}^R$ avec $X^n = \{X_s, s \in S^n\}$ et X_s à valeurs dans Λ , espace d'états supposé discret ici. Les variables observées Y_s sont supposées à valeurs dans un ensemble Υ discret ou continu.

Nous faisons les hypothèses suivantes sur les distributions intervenant dans la description des propriétés statistiques de ce modèle :

– l'hypothèse fondamentale consiste à considérer le processus aléatoire X markovien en échelle, *i.e.* :

$$P(x^n | x^k, k > n) = P(x^n | x^{n+1}) \quad (1)$$

– les probabilités de transitions inter-échelles sont supposées se factoriser de la façon suivante :

$$P(x^n | x^{n+1}) = \prod_{s \in S^n} P(x_s | x_{s^-}) \quad (2)$$

– d'autre part nous supposons une dépendance ponctuelle de Y relativement à X , c'est-à-dire :

$$P(y|x) = \prod_{n=0}^R P(y^n | x^n) = \prod_{n=0}^R \prod_{s \in S^n} f^n(y_s | x_s) = \prod_{s \in S} P(y_s | x_s) \quad (3)$$

(on dit que Y est « semi-markovien » par rapport à X). On prendra pour convention $f^n(y_s | \cdot) \equiv 1$ si s ne porte pas d'observation.

A partir des hypothèses (1), (2) et (3) on déduit aisément que (X, Y) est markovien sur le quadarbre². En effet la loi jointe s'écrit :

$$\begin{aligned} P(x, y) &= \prod_{s \in S} P(y_s | x_s) P(x_r) \prod_{s \in S} P(x_s | x_{s^-}) \quad (4) \\ &= \exp \left\{ \sum_{n=0}^R \sum_{s \in S^n} \log f^n(y_s | x_s) + \log P(x_r) \right. \\ &\quad \left. + \sum_{s \in S} \log P(x_s | x_{s^-}) \right\} \end{aligned}$$

Autrement dit (X, Y) suit une loi gibbsienne ce qui revient à dire, grâce au théorème d'Hammersley-Clifford [20], que $Z = (X, Y)$ est markovien sur le quadarbre (le graphe de dépendance³ dans le cas monodimensionnel - pour des raisons de clarté - est représenté Figure 3; il s'agit d'un arbre dyadique avec des arcs verticaux sur chaque sommet, symbolisant les dépendances ponctuelles de Y par rapport à X).

Nous ferons dans la suite largement usage des propriétés de séparabilité⁴ (ou d'indépendance conditionnelle) suivantes [16] (vraies si $P(z) > 0, \forall z$) :

$\forall A, B, C \subset S$, si B sépare A et C ,

$$\text{alors } P(z_A, z_C | z_B) = P(z_A | z_B) P(z_C | z_B) \quad (5)$$

qui est équivalente à :

$$P(z_A | z_B, z_C) = P(z_A | z_B) \quad (6)$$

cette dernière forme étant plus classique.

2. En toute rigueur il ne s'agit pas d'un quadarbre. Il faudrait en effet rajouter à partir de chaque sommet un arc symbolisant la dépendance ponctuelle de Y par rapport à X (voir Figure 3). Par contre X et $X|Y$ sont effectivement markoviens sur le quadarbre.

3. On dit que G est un graphe de dépendance pour X si et seulement si X est markovien sur G . Pour toutes ces notions, on renvoie le lecteur à [43].

4. On dit que B sépare A et C ssi pour tout site de A et tout site de C , toute chaîne joignant ces deux sites traverse B .

i.e. :

$$\bullet \text{ si } s \in S^0 : \begin{cases} \hat{P}_s(x_{s-}) \triangleq \max_{x_s} P(x_s, y_s | x_{s-}) \\ = \max_{x_s} f^0(y_s | x_s) P(x_s | x_{s-}) \\ \hat{x}_s(x_{s-}) = \arg \max_{x_s} \{ f^0(y_s | x_s) P(x_s | x_{s-}) \} \end{cases}$$

• si $s \in S^n, n < R$: $\hat{P}_s(x_{s-})$ est calculé à partir de la relation (10).

On mémorise simultanément les $\hat{x}_s(x_{s-})$, définis par la relation (11) et réalisant les maxima.

Passé descendante

Une seconde *passé descendante*, partant de la racine, « déroule » les estimées finales \hat{x}_s , à partir des maxima mémorisés dans la *passé montante*, et ce jusqu'aux feuilles :

$$\bullet \hat{x}_r = \arg \max_{x_r} \left\{ f^R(y_r | x_r) \prod_{s \in r^+} \hat{P}_s(x_r) \right\} \text{ (on suppose ici que } X_r \text{ suit une loi uniforme)}$$

• Pour $s \in S^n, 0 \leq n < R$, \hat{x}_s est l'argument qui réalise le maximum $\hat{P}_s(\hat{x}_{s-})$, i.e., $\hat{x}_s = \hat{x}_s(\hat{x}_{s-})$.

En pratique, pour éviter les problèmes de dépassements inférieurs de capacité, nous passons par le logarithme des expressions précédentes pour implanter l'algorithme.

4.2. estimateur du SMAP et algorithme de Bouman modifié

Le critère du MAP, considéré dans le paragraphe précédente, définit un estimateur pertinent mais assez mal adapté à la structure particulière du quadarbre. Considérons en effet l'équation (8). On s'aperçoit que la fonction de coût du MAP pénalise les erreurs indépendamment de l'échelle à laquelle elles se produisent. Autrement dit une erreur à l'échelle la plus grossière (la racine) est pénalisée de la même façon qu'une erreur à l'échelle la plus fine. Or il est clair que les erreurs aux échelles grossières devraient être pénalisées plus lourdement que les erreurs aux échelles fines puisqu'un site à une échelle $n+1$ « représente » 4 sites à l'échelle n . On pourra donc préférer à l'estimateur du MAP, des estimateurs (i.e., des fonctions de coût) mieux adaptés à la structure hiérarchique considérée, comme par exemple le critère du SMAP (« Sequential MAP ») de Bouman *et al.* [8] :

$$C_{SMAP}(x, x') \triangleq \sum_{n=0}^R 2^n C_n(x, x')$$

avec :
$$C_n(x, x') \triangleq 1 - \prod_{i=n}^R \delta(x^i - x'^i)$$

Bouman [8] montre que cette fonction pénalise une erreur en un site proportionnellement à la taille de son correspondant à la résolution la plus fine, i.e., au bloc constitué par ses descendants. Il propose, sur la base de cette fonction de coût, un algorithme d'estimation approché au sens du SMAP [8] sur le quadarbre. Cet algorithme, décrit de façon détaillée dans [8], est non itératif,

comme ceux que nous proposons, et se déroule en deux passes sur l'arbre. Il s'appuie sur des approximations et fournit donc une solution approchée.

Nous avons développé une extension de cet algorithme qui permet, comme les autres algorithmes présentés, de traiter le cas général de données multispectrales-multirésolutions (Bouman considérait en effet uniquement des observations scalaires au niveau le plus fin). Cette extension, proche de l'algorithme de Bouman, est décrite en détail dans [29, 31]. Nous ne la reprenons pas ici, par souci de concision.

4.3. estimateur du MPM et algorithme « Forward-Backward »

Le dernier algorithme développé permet d'obtenir, de façon *exacte et non itérative*, le mode de la marginale *a posteriori* sur le quadarbre (critère dit du « MPM »).

Il s'agit, rappelons-le, de maximiser en chaque site la marginale *a posteriori* :

$$\forall s, \hat{x}_s \triangleq \arg \max_{x_s} P(x_s | y)$$

On montre que la fonction de coût associée au MPM pénalise les erreurs proportionnellement à leur nombre. Le critère du MPM est, pour cette raison, souvent préféré au critère du MAP. La dérivation d'un algorithme de calcul de cet estimateur passe par le calcul des marginales *a posteriori* $P(X_s = i | y)$, que nous détaillons maintenant.

L'algorithme consiste essentiellement en deux passes comparables à celles correspondant au calcul des probabilités *forward* et *backward* dans l'algorithme classique de Baum pour une chaîne de Markov [4]. Avant d'explicitier les calculs, nous devons souligner le fait que nous raisonnons non sur des probabilités jointes, comme dans l'algorithme de Baum standard pour une chaîne de Markov cachée⁵, mais sur des probabilités *a posteriori*. Plus précisément au lieu de manipuler des probabilités monovariées du type $P(y_{\geq s} | x_s)$ d'autant plus petites que l'ensemble $\geq s$ est grand, nous manipulons des probabilités du type $P(x_s | y_{\geq s})$ (beaucoup plus grandes). Ainsi nous ne sommes pas affectés par le problème, crucial ici, de dépassement inférieur de capacité, que nous avons évoqué précédemment.

Passé montante

La première *passé montante*, propage les observations des feuilles de l'arbre jusqu'à la racine et consiste à calculer récursivement les marginales *a posteriori* partielles $P(x_s | y_{\geq s})$, $P(x_{s-} | y_{\geq s})$ et $P(x_s, x_{s-} | y_{\geq s})$.

On a :

$$P(x_s | y_{\geq s}) = \frac{P(x_s)P(y_{\geq s} | x_s)}{P(y_{\geq s})} = \frac{P(x_s)P(y_s | x_s) \prod_{t \in s^+} P(y_{\geq t} | x_s)}{P(y_{\geq s})}$$

avec :

$$P(y_{\geq t} | x_s) = \frac{P(x_s | y_{\geq t})P(y_{\geq t})}{P(x_s)}$$

5. Il est en effet possible de calculer les marginales *a posteriori* sous cette forme.

Donc

$$P(x_s|y_{\geq s}) = \frac{1}{Z} P(x_s)^{-3} P(y_s|x_s) \prod_{t \in s^+} P(x_s|y_{\geq t}) \quad (12)$$

où Z est une constante de normalisation.

La récurrence n'apparaît pas encore clairement. Il s'agit d'exprimer les $P(x_s|y_{\geq t})$ de la formule (12) en fonction des $P(x_t|y_{\geq t})$, ce qui se fait facilement par le biais des lois des couples $P(x_s, x_{s-}|y_{\geq s})$, dont nous aurons également besoin dans la passe descendante :

$$P(x_s, x_{s-}|y_{\geq s}) = P(x_{s-}|x_s) P(x_s|y_{\geq s}) \quad (13)$$

Ainsi en intégrant (13) sur x_s on obtient $P(x_{s-}|y_{\geq s})$:

$$P(x_{s-}|y_{\geq s}) = \sum_j P(x_{s-}|X_s = j) P(X_s = j|y_{\geq s}) \quad (14)$$

De (12) on tire alors la récurrence ascendante suivante :

$$P(x_s|y_{\geq s}) = \frac{1}{Z} P(x_s)^{-3} P(y_s|x_s) \prod_{t \in s^+} \sum_j P(x_s|X_t = j) P(X_t = j|y_{\geq t}) \quad (15)$$

avec :

$$Z = \sum_i \left\{ P(X_s = i)^{-3} P(y_s|X_s = i) \prod_{t \in s^+} \sum_j P(X_s = i|X_t = j) P(X_t = j|y_{\geq t}) \right\}$$

On voit apparaître dans cette formulation les probabilités *a priori* $P(x_s)$ (qui dépendent de s) et les probabilités de transition « inverse » (du fils vers le père) $P(x_s|x_t)$. Les premières se calculent aisément au préalable par *récurrence descendante* à l'aide de la loi initiale et des probabilités de transition :

$$P(X_s = i) = \sum_j P(X_s = i|X_{s-} = j) P(X_{s-} = j) \quad (16)$$

Les secondes se calculent à l'aide des premières et des probabilités de transition :

$$P(x_{s-}|x_s) = \frac{P(x_s|x_{s-}) P(x_{s-})}{P(x_s)} \quad (17)$$

Enfin à l'échelle 0, où commence la récurrence montante (15), on calcule :

$$P(x_s|y_s) = \frac{P(x_s) P(y_s|x_s)}{\sum_i P(X_s = i) P(y_s|X_s = i)} \quad (18)$$

Pour la seconde passe, à l'issue de la récurrence montante (15), on conserve la vraisemblance calculée à la racine de l'arbre, à savoir $P(x_r|y_{\geq r}) = P(x_r|y)$, ainsi que les autres vraisemblances $P(x_{s-}|y_{\geq s})$ et $P(x_s, x_{s-}|y_{\geq s})$ obtenues par les formules (14) et (13) respectivement.

Passé descendante

À la fin de la première passe on a calculé $P(x_r|y)$, qui constitue le début de la récurrence descendante destinée à calculer les $P(x_s|y)$, $\forall s \in S$. En fait on ne calcule pas directement les marginales *a posteriori*, on passe à nouveau par les lois des couples (X_s, X_{s-}) , cette fois par rapport à y tout entier :

$$\begin{aligned} P(x_s|y) &= \sum_j P(x_s, X_{s-} = j|y) \\ &= \sum_j P(x_s|X_{s-} = j, y) P(X_{s-} = j|y) \\ &= \sum_j P(x_s|X_{s-} = j, y_{\geq s}) P(X_{s-} = j|y) \\ &= \sum_j \frac{P(x_s, X_{s-} = j|y_{\geq s})}{P(X_{s-} = j|y_{\geq s})} P(X_{s-} = j|y) \quad (19) \end{aligned}$$

Il s'agit donc d'une récurrence *descendante* fournissant les lois locales marginales *a posteriori* à partir de celles de la racine (obtenue à l'issue de la passe précédente). Les lois de X_{s-} et (X_s, X_{s-}) conditionnellement à $y_{\geq s}$ nécessaires à cette récurrence ont été obtenues également lors de la passe montante⁶. Il suffit alors de maximiser les marginales *a posteriori*, obtenues par la récurrence (19), pour obtenir les estimées (étiquettes) recherchées.

4.4. Illustration : classification supervisée d'images

Afin de valider nos algorithmes et de dégager les caractéristiques générales de chaque estimateur, nous avons choisi de les comparer sur un modèle simple de classification supervisée. Nous illustrons également les capacités de notre modèle à intégrer (et fusionner) des observations multirésolutions.

L'objectif de la classification est de partitionner les images observées en classes, en s'appuyant sur un modèle d'observation gaussien. Une classe est un ensemble non nécessairement connexe de régions homogènes au sens, ici, du niveau de gris. Les étiquettes X_s sont à valeurs dans $\Lambda = \{1, \dots, L\}$ et représentent les numéros de classe. Le modèle d'observation gaussien à l'échelle n est défini par :

$$P(y_s|X_s = k) \triangleq \frac{\exp\{-1/2(y_s - \mu_k^n)^T (\Gamma_k^n)^{-1} (y_s - \mu_k^n)\}}{(2\pi)^{d^n/2} \det(\Gamma_k^n)^{1/2}}$$

où d^n est le nombre d'images à l'échelle n , $\mu_k^n \in \mathbb{R}^{d^n}$ est la moyenne de la classe k à l'échelle n et Γ_k^n est sa matrice de covariance. Le nombre de classes L est supposé fixé *a priori* et les paramètres de chaque classe (moyenne, matrices d'autocovariance) sont préalablement estimés sur des populations d'apprentissage.

À des fins de comparaison, nous choisissons pour la forme de la probabilité de transition celle adoptée par Bouman *et al.* dans [8] :

6. Notons que ces lois seront également utilisées dans l'algorithme EM sur le quadarbre, Section 5.

$$P(X_s = m | X_{s-} = k) \triangleq \begin{cases} \theta_n & \text{si } m = k \\ \frac{1-\theta_n}{L-1} & \text{sinon} \end{cases} \quad (20)$$

Ce modèle (non linéaire) permet (pour $\theta_n > 0$) de favoriser un étiquetage père-fils identique.

Nous désignerons par la suite par H-MAP (MAP hiérarchique) l'algorithme de Viterbi (associé au critère du MAP sur le quadarbre), par H-MPM l'algorithme correspondant au critère du MPM sur la structure hiérarchique et enfin par H-SMAP l'algorithme de Bouman modifié [8, 29]. Nous présenterons également, dans le cas d'observations monorésolutions, les résultats obtenus par l'algorithme de relaxation déterministe classique des modes conditionnels itérés [7] non hiérarchique (NH-ICM) et par un algorithme de recuit simulé non hiérarchique (NH-MAP) basé sur la dynamique de Metropolis et permettant d'approcher la solution du MAP. Ces deux derniers algorithmes non hiérarchiques minimisent une fonction d'énergie dérivée d'un modèle markovien spatial classique (associé à un 4-voisinage, avec des attaches aux données gaussiennes comme pour le modèle hiérarchique et un terme *a priori* correspondant à un modèle spatial de Potts [20]). Ces deux derniers algorithmes (itératifs) sont considérés comme ayant convergé lorsque le nombre de changements d'étiquettes entre deux itérations est inférieur au millième du nombre de sites. Pour le recuit simulé, une loi de descente en température exponentielle décroissante est spécifiée : $T = T_0 / (1.01)^i$ où T_0 est la température initiale fixée à 100 ici et i désigne l'itération.

Les performances (en terme de qualité d'estimation et de coût calculatoire) ont été évaluées par le taux de bonne classification (pour les images synthétiques) ainsi que par la mesure des temps CPU (obtenus pour une implantation en C sur une station SUN SPARC 10) (tableau 1). Nous présentons tout d'abord des résultats pour deux images synthétiques, présentant des rapports signal à bruit de -0,27dB et -2,16dB respectivement (Figure 4 et 5). Le nombre d'images à l'échelle 0 (pleine résolution) est fixé ici à $d^0 = 1$. L'image synthétique de la planche 4 est constituée de quatre classes correspondant à des formes circulaires de tailles différentes et d'une classe représentant le fond. Chaque classe est caractérisée par un niveau de gris constant (moyenne) auquel on a ajouté un bruit gaussien centré additif de même variance sur toute l'image. Ce n'est pas le cas de l'image 5 pour laquelle la variance du bruit varie selon la classe (il y en a 5).

L'algorithme des ICM non hiérarchique (déterministe) s'avère, comme attendu, très sensible au bruit et aux minima locaux. Quoique rapide, il conduit dans tous les cas au taux de bonne classification le plus faible (Tableau 1). Ses performances chutent très rapidement lorsque le rapport signal-à-bruit se dégrade. L'algorithme du recuit simulé, associé à l'estimateur du MAP non hiérarchique (NH-MAP) conduit à des résultats de bonne qualité lorsque le niveau de bruit n'est pas trop élevé, mais il a tendance (comme l'ont noté d'autres auteurs [35]) à favoriser des solutions trop homogènes lorsque le niveau de bruit augmente. Dans tous les cas, il conduit aux temps de calcul les plus élevés. Des solutions de meilleure qualité peuvent sans doute être obtenues pour NH-MAP en adoptant, pour le recuit simulé, des lois de descente en température plus proches de la loi optimale

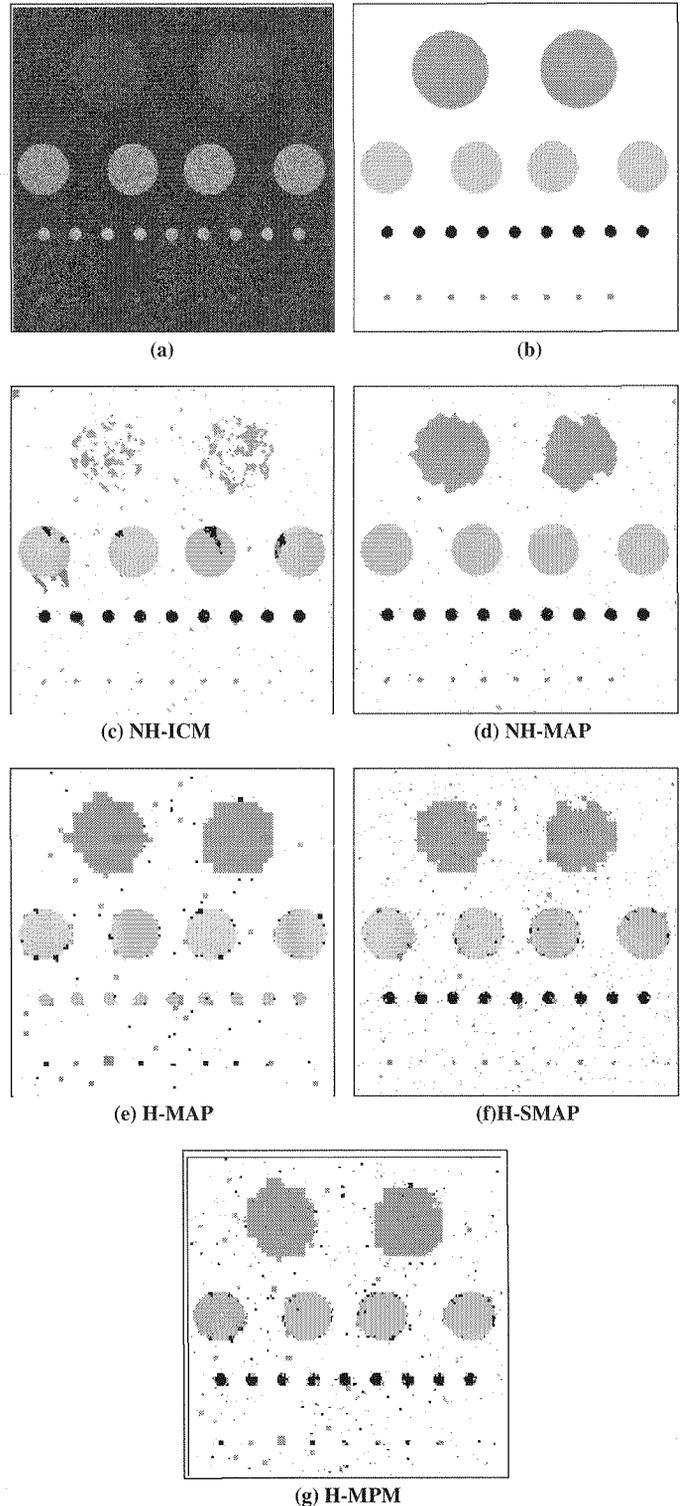


Figure 4. – a) Image 256×256 (rapport signal-sur-bruit -0,27 dB; les niveaux de gris, pour des raisons de lisibilité, ne correspondent pas nécessairement aux moyennes), b) Vérité-terrain, c) Estimateur NH-ICM (non hiérarchique), d) Estimateur NH-MAP (recuit simulé, non hiérarchique), e) Estimateur H-MAP (algorithme de Viterbi hiérarchique), f) Estimateur H-SMAP (algorithme de Bouman, hiérarchique), g) Estimateur H-MPM (algorithme « Forward-Backward » hiérarchique).

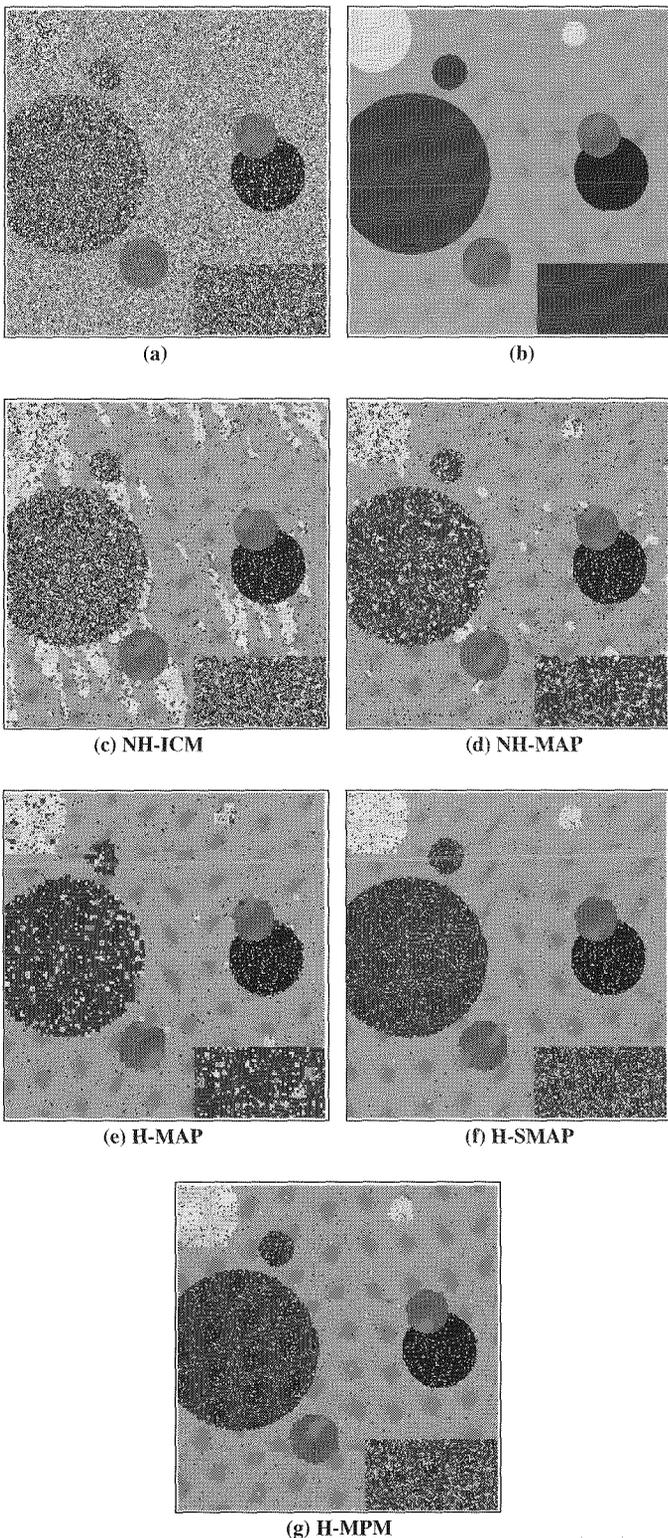


Figure 5. – a) Image 256×256 (rapport signal-sur-bruit $-2,16$ dB ; les niveaux de gris, pour des raisons de lisibilité, ne correspondent pas nécessairement aux moyennes), b) Vérité-terrain, c) Estimateur NH-ICM (non hiérarchique), d) Estimateur NH-MAP (recuit simulé, non hiérarchique), e) Estimateur H-MAP (algorithme de Viterbi hiérarchique), f) Estimateur H-SMAP (algorithme de Bouman, hiérarchique), g) Estimateur H-MPM (algorithme « Forward-Backward » hiérarchique).

Tableau 1. – Résultats (en haut pour l'image de la Figure 4a, en bas pour l'image de la Figure 5a). Le temps indiqué est le temps cpu sur une station Sun Spare 10.

Algorithme	Taux de bonne classification	Temps CPU
NH-ICM	89	10 s
NH-MAP	97.1	444 s
H-MAP	96.0	26 s
H-SMAP	97.4	36 s
H-MPM	95.5	16 s

Algorithme	Taux de bonne classification	Temps CPU
NH-ICM	57	11 s
NH-MAP	87	708 s
H-MAP	91	28 s
H-SMAP	89	19 s
H-MPM	93	9 s

(logarithmique), mais au prix de temps de calcul prohibitifs. Les estimateurs hiérarchiques sur le quadarbre (H-MAP, H-SMAP et H-MPM) permettent d'atteindre un bon compromis entre qualité de l'estimation et rapidité d'exécution. L'algorithme du H-MPM, lié à une fonction de coût qui pénalise de façon plus graduelle les erreurs locales que le H-MAP ou le H-SMAP, conduit en particulier à des résultats satisfaisants pour les temps de calcul les plus courts, dans un rapport de 30 à 80 avec ceux du recuit simulé (NH-MAP). Les algorithmes non hiérarchiques se sont avérés, de façon générale moins robustes au bruit (on trouvera dans [31] des résultats expérimentaux complémentaires). Comme attendu, on observe quelques effets de bloc pour les estimées obtenues par le modèle sur le quadarbre. Notons que ces effets, quoique visuellement gênants, n'ont qu'un effet marginal sur la qualité objective des estimées (mesurée par le taux de bonne classification). On voit en effet que les frontières des objets sont également mal estimées dans le cas des approches non hiérarchiques (grands disques pour NH-MAP, Figure 4d), mais leur caractère peu structuré rend ces imprécisions moins visibles dans ce cas. Rappelons également que des solutions ont été proposées récemment, dans le cas d'étiquettes continues, pour atténuer les artefacts apparaissant avec les modèles sur le quadarbre [18].

Pour illustrer les capacités du modèle à fusionner des observations multiples, définies à des résolutions différentes, nous présentons par ailleurs un exemple de classification, sur images aériennes. La scène présente (dans le visible) la région de Saint-Louis prise lors de la crue historique des fleuves Mississippi et Missouri en juillet 1993. Nous considérons ici 4 classes : une classe pour le Mississippi, une classe pour le Missouri, une classe pour la zone urbaine (routes, buildings ...) et enfin une classe pour les autres types de terrains (forêts, champs ...). Ce choix est visuellement justifié par l'effective homogénéité des niveaux de gris au sein de chaque classe. La figure 6 montre la structure pyramidale des observations et des étiquettes sur le quadarbre. Les observations sont définies à deux résolutions différentes (512×512

et 64×64). Ces images étant peu bruitées, les algorithmes hiérarchiques donnent visuellement des résultats très proches. On constate également que les effets de blocs sont peu visibles sur ce type d'images. Nous présentons la classification obtenue avec l'algorithme de Viterbi, associé au critère du MAP (H-MAP). On peut voir Figure 6 les étiquettes obtenues à chaque niveau de l'arbre.

L'algorithme affine graduellement la classification, en intégrant et fusionnant progressivement l'information. Le temps de calcul est de 1 mn 40 s sur la même station de travail que précédemment. Comme attendu, si l'on ne prend en compte qu'une seule image (Figure 7a), les deux fleuves ne sont pas discriminés. La fusion avec une seconde image permet ici cette discrimination.

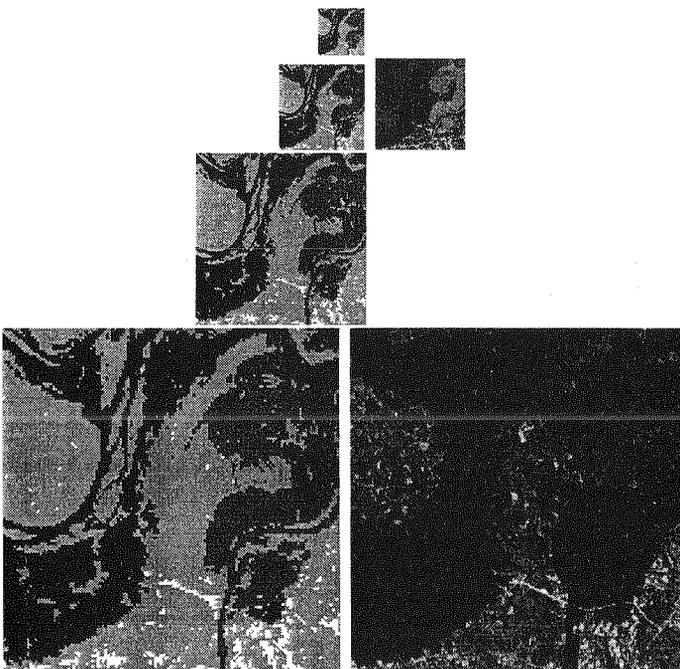


Figure 6. – Fusion par l'algorithme de Viterbi sur le quadarbre (H-MAP) d'images aériennes multirésolutions (observations à droite, cartes de classification à gauche).

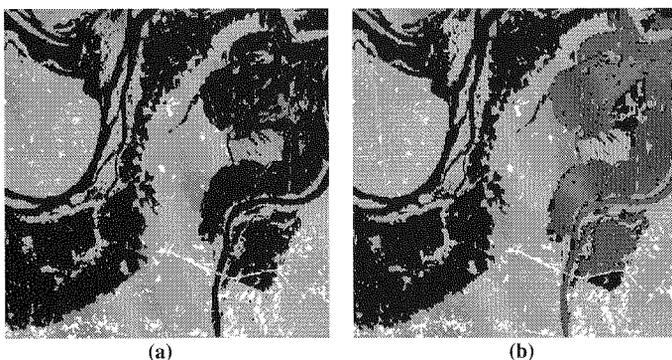


Figure 7. – a) H-MAP avec un seul champ d'observations à la résolution fine; b) H-MAP avec fusion des deux champs d'observations. On constate qu'un seul champ d'observation ne permet pas de discriminer les deux fleuves dans ce cas.

5. algorithmes EM d'estimation des paramètres du modèle sur le quadarbre

Les algorithmes considérés dans la partie précédente sont *supervisés*, c'est-à-dire supposent connus les hyperparamètres du modèle sur le quadarbre (probabilités de transition) ainsi que les paramètres liés au modèle d'observation gaussien (nombre de classes, moyennes et matrices de covariance). L'estimation de ces paramètres peut se faire par des techniques d'apprentissage. Elle est primordiale car elle conditionne fortement les résultats finaux.

Dans cette partie nous développons deux algorithmes EM (« Expectation-Maximization ») d'estimation *non supervisée* sur le quadarbre. L'algorithme EM [39] est un algorithme d'estimation itératif des paramètres qui a pour but de maximiser la vraisemblance des observations. Cet algorithme, bien que donnant de bons résultats (en particulier dans le cas d'un mélange de lois gaussiennes), a pour principal inconvénient d'être très coûteux dans le cas des modèles markoviens spatiaux [10]. Nous montrons comment l'utilisation de modèles hiérarchiques sur le quadarbre permet de réduire sensiblement ce coût, sans préjudice sur la qualité des estimées obtenues. Notons que Bouman propose dans [8] un algorithme de type EM pour estimer le paramètre θ_n , définissant les probabilités de transition de son modèle sur le quadarbre (voir paragraphe 4.4). Il n'estime toutefois pas les autres paramètres du modèle (supposés connus). Contrairement aux algorithmes EM présentés dans la suite, l'algorithme proposé dans [8], est donc spécifique et ne traite pas le cas général d'un modèle hiérarchique non linéaire à observations gaussiennes.

5.1. rappels sur l'algorithme EM et ses variantes

Soit Φ l'ensemble des paramètres à estimer dont dépend la distribution $P(x, y|\Phi)$. L'estimée au sens du maximum de vraisemblance des paramètres s'écrit :

$$\hat{\Phi} \triangleq \arg \max_{\Phi} P(Y = y|\Phi).$$

Le champ des étiquettes X étant caché, l'algorithme EM considère la maximisation itérative de l'espérance conditionnelle :

$$Q(\Phi|\Phi^{(k)}) = \mathbb{E}_{X|Y, \Phi} [\log P(X, y|\Phi)|y, \Phi^{(k)}] \quad (21)$$

où $\Phi^{(k)}$ représente le vecteur des paramètres estimés à l'étape k . L'algorithme EM consiste à répéter jusqu'à convergence les deux procédures suivantes :

- étape E (« Expectation ») Calcul de $Q(\Phi|\Phi^{(k)})$
- étape M (« Maximization ») $\Phi^{(k+1)} = \arg \max_{\Phi} Q(\Phi|\Phi^{(k)})$

On montre [39] que cet algorithme converge vers un point stationnaire, qui est soit un point de selle, soit un maximum local. En pratique, la mise en œuvre de l'algorithme EM est parfois délicate, car l'estimée finale dépend fortement de l'initialisation. L'algorithme peut en effet facilement se piéger dans un maximum local (proche de l'initialisation). Par ailleurs, il est généralement (sauf pour des problèmes de « petite » taille, ce qui est rarement le cas en imagerie) d'une lenteur rédhibitoire. Enfin le calcul de l'espérance, ainsi que la maximisation, sont difficiles à mener. Pour réduire la sensibilité de l'algorithme à la configuration initiale, des variantes stochastiques de l'algorithme EM ont été développées (un état de l'art sur ces méthodes peut être trouvé dans [9]). Certaines variantes utilisent ainsi des méthodes de Monte-Carlo pour estimer l'espérance ou maximiser cette espérance (par exemple les algorithmes MCEM - Monte-Carlo EM - [42] ou l'algorithme SEM - Stochastic EM - cas particulier d'algorithme MCEM [9]) Ces méthodes sont très coûteuses puisqu'elles sont de nature itérative et qu'elles s'imbriquent dans une procédure déjà itérative.

Lorsque le signal à modéliser est un processus markovien et que le graphe de dépendance est un arbre, il est par contre possible de développer des algorithmes de type EM, efficaces d'un point de vue calculatoire. Ainsi pour un signal monodimensionnel représenté par une chaîne de Markov, Baum et Welch [5] ont développé un algorithme permettant d'estimer les paramètres du modèle (probabilités de transition, vraisemblances, probabilités initiales) à l'aide de formules de récurrence simples. Les procédures **E** et **M** sont alors non itératives. Cet algorithme est classiquement utilisé en reconnaissance de la parole et de l'écriture. Dans le cas des images, des extensions de l'algorithme de Baum et Welch ont été développées pour certaines classes de modèles markoviens causaux (chaînes de Markov [6], réseaux de Markov, champs de Pickard [15]). Ces modèles markoviens causaux ne sont toutefois pas toujours aptes à capter toutes la richesse des statistiques spatiales d'une image, même si certains travaux récents ont montré leur réel intérêt en analyse d'images [6].

Dans le cas d'une image modélisée par un champ de Markov spatial non causal, il n'est malheureusement plus possible de dériver un algorithme efficace comme les algorithmes précédents. Chalmond [10] a toutefois proposé un algorithme approché (appelé « EM gibbsien »), dont nous rappelons brièvement les caractéristiques. Si (X, Y) est un processus markovien spatial, Chalmond considère sa loi gibbsienne [20] :

$$P(x, y|\Phi) = \frac{1}{Z(\Phi)} e^{-U(x, y|\Phi)}$$

où $Z(\Phi)$ est une constante de normalisation qui dépend de Φ . L'impossibilité, en pratique, de calculer la constante de normalisation [20], rend également impossible le calcul de Q . La solution adoptée par Chalmond [10] et Zhang *et al.* [44] est de remplacer la fonction de vraisemblance par la fonction de « pseudo-vraisemblance » : $\mathbb{P}(x, y|\Phi) \triangleq P(y|x, \Phi)\mathbb{P}(x|\Phi)$ avec $\mathbb{P}(x|\Phi) \triangleq \prod_s P(x_s|x_{\partial s}, \Phi)$ où ∂s désigne le voisinage spatial de s

et $P(y|x, \Phi) = \prod_{s \in S} P(y_s|x_s, \Phi)$. Le calcul de l'espérance devient alors possible. Dans le cas gaussien il est même possible d'obtenir

explicitement les formules de ré-estimation [10]. Celles-ci dépendent des marginales locales *a posteriori* qui peuvent être estimées à l'aide d'un échantillonneur de Gibbs [10, 20]. Conjointement à l'estimation des paramètres des classes et du modèle, le champ d'étiquettes est estimé à chaque itération selon le critère du MPM [10]. Cet algorithme possède plusieurs inconvénients, en particulier son coût calculatoire élevé dû essentiellement au nécessaire recours à un échantillonneur de Gibbs qui requiert un grand nombre d'itérations pour atteindre l'équilibre. En outre l'EM gibbsien utilise non pas la fonction de vraisemblance mais la fonction de pseudo-vraisemblance. Or la convergence de l'algorithme n'a pas été étudiée dans ce cas. Comme nous le verrons dans la suite, le modèle sur le quadarbre permet d'éviter ces problèmes, inhérents aux modèles spatiaux non causaux.

5.2. algorithme EM sur le quadarbre

On considère le processus (X, Y) markovien sur le quadarbre, présenté dans le paragraphe (3.2), où X_s est à valeurs dans $\Lambda = \{1, \dots, L\}$ et Y_s est à valeurs dans Υ^n pour $s \in S^n$. La vraisemblance conjointe de (X, Y) s'exprime alors par (cf. Eq. 4) :

$$P(x, y|\Phi) = \prod_{s \in S} P(y_s|x_s, \Phi) P(x_r|\Phi) \prod_{s \in >r} P(x_s|x_{s-}, \Phi) \quad (22)$$

Φ représente l'ensemble des paramètres à estimer, à savoir ici :

$$\Phi \triangleq \{a_{ij}, \pi_i, f_i^n(l), i, j \in \Lambda, l \in \Upsilon^n, n = 0, \dots, R\}$$

où :

a_{ij} désigne la probabilité de transition du père vers le fils, i.e., $a_{ij} \triangleq P(X_s = j | X_{s-} = i)$ (indépendant de s car le processus est supposé homogène);

π_i représente la loi *a priori* du processus à la racine, i.e., $\pi_i \triangleq P(X_r = i)$;

$f_i^n(l)$ désigne la vraisemblance des observations au niveau n , i.e. $f_i^n(l) \triangleq P(Y_s = l | X_s = i), s \in S^n$, que l'on supposera dépendre de l'échelle n .

Nous présentons tout d'abord un algorithme EM *exact*⁷ sur le quadarbre, pour le processus markovien caché (X, Y) . On introduit les notations suivantes :

$$\xi_s^{(k)}(i) \triangleq P(X_s = i | y, \Phi^{(k)}) \text{ et}$$

$$\Psi_s^{(k)}(i, j) \triangleq P(X_s = j, X_{s-} = i | y, \Phi^{(k)})$$

représentent les marginales *a posteriori*.

7. Aucune approximation n'est faite que ce soit lors du calcul de l'espérance ou de la maximisation de celle-ci.

Le calcul et la maximisation de l'espérance sont menés en Annexe A. On aboutit aux formules de ré-estimation suivantes :

$$\left\{ \begin{array}{l} f_i^{n(k+1)}(l) = \frac{\sum_{s \in S^n: Y_s=l} \xi_s^{(k)}(i)}{\sum_{s \in S^n} \xi_s^{(k)}(i)} \\ a_{ij}^{(k+1)} = \frac{\sum_{s>r} \Psi_s^{(k)}(i, j)}{\sum_{s>r} \xi_{s-}^{(k)}(i)} \\ \pi_i^{(k+1)} = \xi_r^{(k)}(i) \end{array} \right. \quad (23)$$

Nous avons montré précédemment (paragraphe 4.3) que les marginales *a posteriori* $\xi_s^{(k)}(i)$ pouvaient se calculer en deux passes sur l'arbre, l'une montante et l'autre descendante. Par ailleurs, les lois des couples $\Psi_s^{(k)}(i, j)$ se déduisent également des marginales *a posteriori* ainsi que de différentes autres marginales, qui sont obtenues lors du calcul des $\xi_s^{(k)}(\cdot)$ (voir paragraphe 4.3) :

$$\begin{aligned} P(X_s = j, X_{s-} = i | y, \Phi^{(k)}) \\ = P(X_s = j | X_{s-} = i, y_{\geq s}, \Phi^{(k)}) P(X_{s-} = i | y, \Phi^{(k)}) \\ = \frac{P(X_s = j, X_{s-} = i | y_{\geq s}, \Phi^{(k)})}{P(X_{s-} = i | y_{\geq s}, \Phi^{(k)})} \xi_{s-}^{(k)}(i) \end{aligned}$$

L'algorithme EM consiste donc à alterner, jusqu'à convergence, le calcul de ces marginales et la remise à jour des paramètres du modèle sur la base de ces probabilités (les procédures d'initialisation et les critères d'arrêt seront présentés dans le paragraphe 5.4). Il est ensuite bien entendu possible d'obtenir, à la fin de l'algorithme EM, sur la base des paramètres estimés à convergence, une estimation du processus caché X selon l'un des critères définis dans le paragraphe 4 (H-MAP, H-SMAP ou H-MPM). Le calcul des marginales *a posteriori* étant déjà réalisé dans l'algorithme EM, c'est l'estimateur H-MPM, qui est le moins coûteux dans ce cas (il permet également d'obtenir, avec un faible surcoût, une estimée de X après chaque itération de l'algorithme EM).

5.3. algorithme MCEM sur le quadarbre

L'algorithme MCEM (et l'algorithme SEM qui n'en est qu'un cas particulier) consiste, rappelons-le, à tirer des échantillons selon la loi $P(x|y, \Phi^{(k)})$ pour estimer l'espérance à calculer lors de l'étape E de l'algorithme EM [9, 42] (cf. Eq. 21). Il s'agit donc [42] de tirer, à l'itération k , et en chaque site s du quadarbre, M étiquettes $x_s^k(1), \dots, x_s^k(M)$ selon la loi $P(x|y, \Phi^{(k)})$ et d'approcher l'espérance par :

$$\mathbb{E}_{X|Y, \Phi} [\log P(X, y | \Phi) | y, \Phi^{(k)}] \simeq \underbrace{\frac{1}{M} \sum_{m=1}^M \log P(x^k(m), y | \Phi)}_{\hat{=} \tilde{Q}(\Phi | \Phi^{(k)})} \quad (24)$$

La maximisation (étape M) porte ensuite sur $\tilde{Q}(\Phi | \Phi^{(k)})$.

D'après (22), la loi $P(x|y, \Phi^{(k)})$ peut se décomposer sous la forme suivante :

$$P(x|y, \Phi) = P(x_r | \Phi) \prod_{s > r} P(x_s | x_{s-}, \Phi) \prod_{s \in S} P(y_s | x_s, \Phi) P(y | \Phi)^{-1} \quad (25)$$

La loi (25) étant causale, la procédure d'échantillonnage devient ici *non itérative*. On démarre ainsi à la racine r de l'arbre où l'on tire des échantillons selon :

$$P(X_r = i | y, \Phi^{(k)}) \quad (26)$$

puis, en se déplaçant vers les résolutions croissantes, on tire des échantillons selon les lois :

$$\begin{aligned} P(X_s = i | y, X_{s-} = j, \Phi^{(k)}) &= P(X_s = i | y_{\geq s}, X_{s-} = j, \Phi^{(k)}) \\ &\propto P(X_s = i | X_{s-} = j, \Phi^{(k)}) P(y_{\geq s} | X_s = i, \Phi^{(k)}) \end{aligned} \quad (27)$$

Les formules de ré-estimation sont obtenues comme pour l'algorithme EM sur le quadarbre présenté dans la partie précédente. Le détail des calculs est présenté en Annexe B. En redéfinissant les notations :

$$\left\{ \begin{array}{l} \xi_s^{(k)}(i) \hat{=} \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{x_s^k(m)=i} \\ \Psi_s^{(k)}(i, j) \hat{=} \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{x_s^k(m)=j, x_{s-}^k(m)=i} \end{array} \right. \quad (28)$$

on obtient des formules de ré-estimation similaires à celles obtenues pour l'algorithme EM (cf. Eq. 23) :

$$\left\{ \begin{array}{l} f_i^{n(k+1)}(l) = \frac{\sum_{s \in S^n: Y_s=l} \xi_s^{(k)}(i)}{\sum_{s \in S^n} \xi_s^{(k)}(i)} \\ a_{ij}^{(k+1)} = \frac{\sum_{s>r} \Psi_s^{(k)}(i, j)}{\sum_{s>r} \xi_{s-}^{(k)}(i)} \\ \pi_i^{(k+1)} = \xi_r^{(k)}(i) \end{array} \right. \quad (29)$$

Notons toutefois que cette version *exacte* de l'algorithme MCEM sur le quadarbre est très coûteuse même si, contrairement à l'algorithme EM gibbsien, [10, 44] le tirage aléatoire d'après la loi $P(x|y, \Phi^{(k)})$ peut se faire de façon non itérative. En effet, le tirage des échantillons (Eq. 26) et (27)) nécessite le calcul des probabilités $P(X_r = i | y, \Phi^{(k)})$ et $P(y_{\geq s} | X_s = i, \Phi^{(k)})$. Ces probabilités peuvent certes être calculées explicitement (comme dans l'algorithme H-MPM), mais ce calcul, doit être répété à chaque itération de l'algorithme. La première solution habituellement envisagée pour réduire la charge calculatoire, consiste à remplacer le conditionnement en $y_{\geq s}$ dans l'équation (27) par un conditionnement en y_s , simplement. Cette approximation, parfois utilisée pour les chaînes de Markov, n'est toutefois pas licite ici, dans la mesure où les seules observations disponibles dans notre modèle sont

généralement portées par les feuilles de l'arbre. Cette version approchée de l'algorithme MCEM reviendrait dans ce cas à ne prendre en compte les observations qu'à la fin de la procédure de tirage (Eq. (26) et (27)) qui, rappelons le, progresse de la racine vers les feuilles de l'arbre. Une autre version approchée de l'algorithme, qui a donné de bons résultats en pratique, consiste à construire au préalable une pyramide d'observations de sorte que l'on ait des observations en chaque site du quadarbre. On vient alors tirer une étiquette en un site en fonction de celle de son père et de l'observation (ponctuelle) créée en ce site. C'est cette solution qui a été retenue dans les expérimentations. Par rapport aux modèles markoviens spatiaux qui nécessitent des procédures d'échantillonnage itératives très coûteuses (comme l'échantillonneur de Gibbs) le gain de temps obtenu par cette approche est considérable. Le recours à des versions stochastiques de l'algorithme EM devient donc possible avec le modèle causal sur le quadarbre, conduisant à des résultats qualitativement et quantitativement meilleurs que pour l'algorithme EM gibbsien, en raison de la moindre sensibilité à l'initialisation.

5.4. illustration : classification non supervisée d'images

Nous présentons des résultats sur des images synthétiques et réelles dans le cas de la classification non supervisée. Il s'agit non seulement d'estimer les paramètres du modèle au sens du maximum de vraisemblance, mais également de trouver la meilleure classification au sens du MPM. Nous considérons les résultats obtenus par les deux algorithmes présentés dans cette partie (l'EM et le MCEM sur le quadarbre). Ces algorithmes sont comparés avec deux algorithmes non hiérarchiques classiques : l'EM gibbsien et le MCEM « aveugle » [42] (*i.e.*, lorsque y est considéré issu d'une loi de mélange sans introduire d'information contextuelle). La comparaison avec le MCEM aveugle permet d'apprécier l'apport des modèles markoviens et donc l'intérêt d'avoir développé des algorithmes de type EM pour de tels processus. Notons qu'il n'est pas envisageable, pour des raisons de complexité calculatoire de mettre en œuvre une version stochastique de l'algorithme EM gibbsien. L'EM gibbsien est appliqué à un modèle markovien spatial de Potts défini sur une grille 2-D régulière, associé à un 4-voisinage. On se place toujours dans le cas d'observations gaussiennes.

La procédure d'initialisation est la même pour tous les algorithmes et est fournie par une analyse grossière de l'histogramme de l'image à la résolution la plus fine. Le nombre de classes est estimé comme suit : il est initialisé à un majorant (en général assez grand), puis à chaque itération les classes i dont le nombre de représentants $\#\{s \in S : X_s = i\}$ est inférieur à un seuil sont éliminées. Le critère d'arrêt retenu est :

$$\max_{i \in A} \|(\mu_i, \Gamma_i)^{(k+1)} - (\mu_i, \Gamma_i)^{(k)}\| < \epsilon$$

avec $\|(\mu, \Gamma)\| \triangleq |\mu| + |\Gamma|$ dans le cas scalaire et avec $\epsilon = 1,0$ dans nos expériences.

Image synthétique

Cette image (Figure 9-b) représentant des formes géométriques variées a été dégradée par un bruit gaussien. Les classes sont caractérisées par leur moyenne et leur variance (tableau 2). L'histogramme de cette image est présenté Figure 8. On remarque qu'il est très difficile de distinguer plus de trois ou quatre modes alors que le nombre exact de classes est 5 (voir la vérité-terrain Figure 9-b). Les algorithmes EM sur l'arbre parviennent à estimer le bon nombre de classes dans ce cas. Cependant afin de faciliter la comparaison entre approches, les résultats sont présentés pour un nombre de classes fixé *a priori*. Les paramètres estimés par les différents algorithmes sont comparés tableau 2. Les cartes de classification apparaissent en Figure 9.c-f. Le tableau 3 permet de constater le gain significatif en temps de calcul apporté par les algorithmes hiérarchiques EM et MCEM par rapport à l'EM gibbsien. L'algorithme EM sur l'arbre est évidemment le plus rapide car il ne nécessite aucun tirage d'échantillons, par contre il requiert une bonne initialisation (ce qui était le cas ici). Les algorithmes hiérarchiques conduisent également à de meilleurs résultats qualitatifs et quantitatifs. En particulier l'EM gibbsien (associé au modèle markovien spatial) ne parvient pas à discriminer deux cercles correspondant à des classes différentes (Figure 9d et tableau 2). De ce fait les variances estimées par cet algorithme pour les classes 1 et 3 sont très élevées et l'estimée de la moyenne de la classe 1 est fortement biaisée. Le MCEM aveugle donne quant à lui des résultats très bruités, dus à l'absence d'information contextuelle.

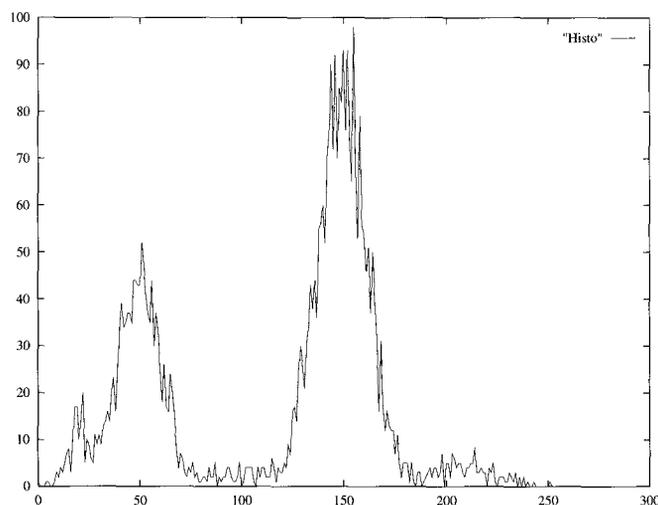


Figure 8. – Histogramme de l'image bruitée de la figure 9.

Image réelle

Nous présentons enfin Figure 10 une comparaison des différentes versions (hiérarchiques et non hiérarchiques) de l'algorithme EM sur une image de la base de donnée du *GDR-PRC ISIS*, représentant des cellules de fibre musculaire (d'autres résultats sont commentés dans [31]). Pour comparer ces algorithmes, nous les avons placés dans les mêmes conditions (le nombre de classes est fixé à 4, et l'initialisation (Figure 10b), obtenue par maximum de vraisemblance, est la même pour tous les algorithmes).

Tableau 2. – Moyennes et variances des classes estimées par les différentes versions de l'algorithme EM (voir image Figure 9)

Classe	Moyennes					Variances				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Vérité-terrain	20	50	100	150	210	25	100	400	144	225
MCEM aveugle	41,4	52,3	141,9	155,2	157,0	280,4	70,3	66,3	54,2	1180,3
EM gibbsien	60,4	49,6	109,4	148,1	197,0	2255,5	128,8	2150,3	276,8	1441,4
EM arbre	21,3	49,5	119,6	149,4	202,1	182,3	105,5	778,4	138,8	545,2
MCEM arbre	19,3	49,4	116,7	149,5	210,5	28,0	102,4	1896,0	134,3	194,9

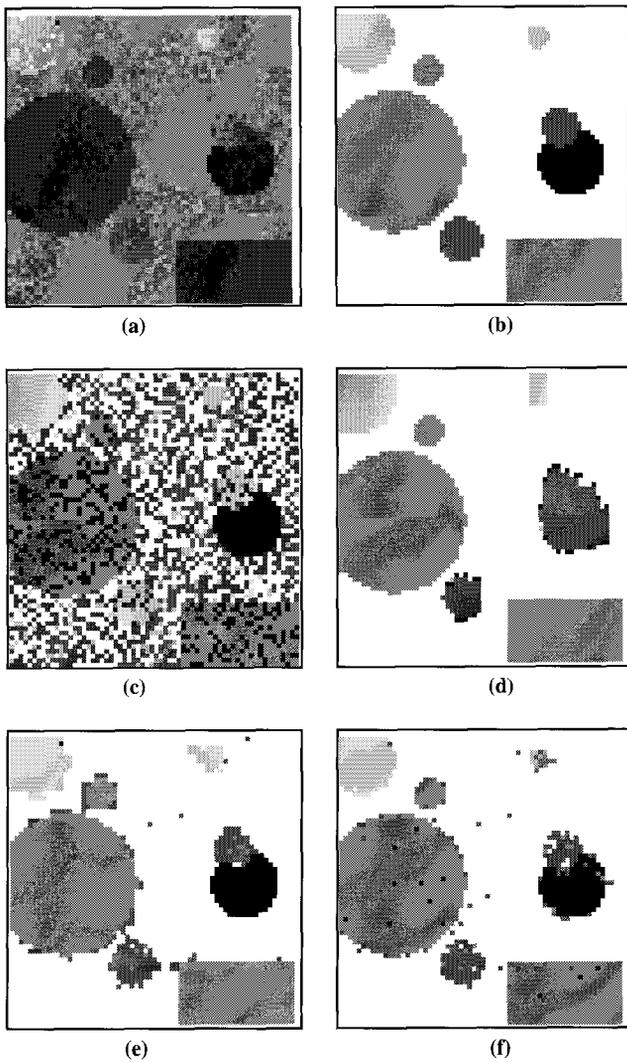


Figure 9. – Résultats de différentes versions de l'algorithme EM sur une image synthétique (les effets de blocs apparaissant sur ces images sont liés à l'agrandissement des images et non aux modèles). a) Image bruitée 64 × 64, b) Vérité-terrain, c) Classification (MCEM aveugle), d) Classification (EM gibbsien), e) Classification (EM sur l'arbre), f) Classification (MCEM sur l'arbre).

On remarque que l'EM gibbsien et l'EM sur le quadarbre ne parviennent pas à s'affranchir de la (relative) mauvaise initialisation (voir Figure 10b) et ne discriminent pas le fond (en blanc sur la Figure 10) et les cellules claires. Ceci n'est pas le cas de l'algorithme stochastique MCEM. Ce résultat est en concordance avec

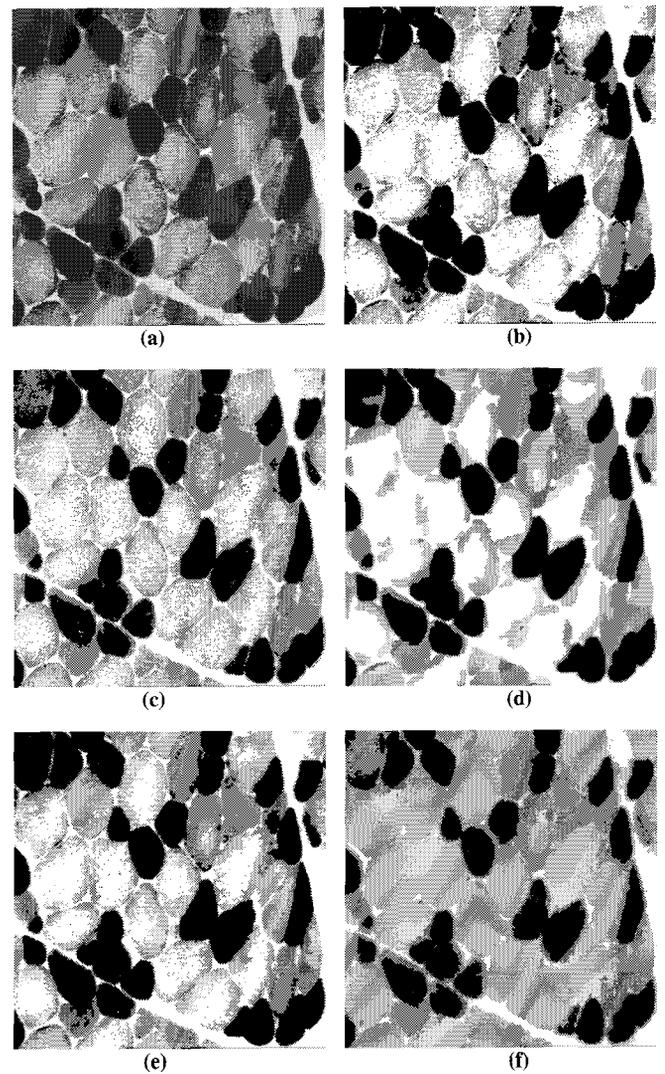


Figure 10. – Résultats des différentes versions de l'algorithme EM sur l'image « muscle ». a) Image originale 256 × 256, b) Initialisation (maximum de vraisemblance), c) Classification par le MCEM aveugle ($nb\ iter = 52, M = 10$). d) Classification par l'EM gibbsien ($nb\ iter = 41, M = 50$). e) Classification par l'EM sur l'arbre ($nb\ iter = 3$). f) Classification par le MCEM sur l'arbre ($nb\ iter = 228, M = 10$).

les propriétés des algorithmes EM qui sont très sensibles à l'initialisation. Là encore le MCEM « aveugle » produit des résultats plus bruités pour la simple raison qu'il ne prend pas en compte l'information contextuelle.

Tableau 3. – Résultats sur l'image synthétique 64×64 de la Figure 9a. (*nb iter* désigne le nombre d'itérations à convergence, *M* désigne le nombre d'échantillons tirés aléatoirement dans les algorithmes stochastiques).

Algorithme	Taux de bonne classification	M	nb iter	Temps cpu
MCEM aveugle	58	10	338	26 mn
EM gibbsien	95	50	341	3 h 12 mn
EM arbre	97	10	•	45 s
MCEM arbre	98	10	228	19 mn

6. conclusion

Dans cet article nous avons présenté, pour la classe des modèles markoviens non linéaires définis sur le quadarbre, un ensemble d'algorithmes originaux permettant l'estimation supervisée et non supervisée en analyse d'images. Différents estimateurs (MAP, SMAP, MPM) et techniques d'estimation non supervisée des paramètres de ces modèles (EM, MCEM) ont été décrits. Les propriétés de ces modèles et algorithmes ont été étudiées sur un exemple simple de classification d'images et comparées à celles de modèles spatiaux non causaux classiques. Les algorithmes sur le quadarbre ont montré, dans ce contexte, leur supériorité sur les algorithmes de relaxation spatiale. Cette supériorité est très significative en termes de complexité algorithmique mais apparaît également, de façon plus inattendue au niveau de la qualité des champs estimés. Les effets de blocs, inhérents à cette classe de modèle, sont en particuliers peu sensibles sur les images réelles.

Les résultats expérimentaux reproduits ici sont bien entendu préliminaires et ont été présentés à titre illustratif. Il conviendrait en particulier d'analyser, sur d'autres problèmes d'imagerie, et avec un nombre significatif d'images représentatives, les propriétés de ces modèles et algorithmes et de les comparer aux autres représentations markoviennes d'images : modèles spatiaux causaux (chaînes de Markov, réseaux de Markov, champs de Pickard), modèles spatiaux non causaux (champs de Gibbs).

Par ailleurs, le champ des modèles markoviens définis sur des graphes hiérarchiques plus complexes que le quadarbre n'a été que très partiellement exploré [8,26] et mérite qu'un effort de recherche plus important lui soit consacré.

En conclusion, nous pensons que les modèles hiérarchiques définis sur l'arbre devraient constituer une alternative très attractive aux modèles spatiaux classiques, en raison de la faible complexité calculatoire de leur mise en œuvre et de leur capacité naturelle à fusionner, suivant l'approche bayésienne, des informations se présentant à des résolutions différentes.

Annexe A : Algorithme EM sur le quadarbre

On considère un processus markovien (X, Y) défini sur le quadarbre et vérifiant les propriétés du paragraphe 3.2. X est le processus caché, Y est observé.

On rappelle que la vraisemblance conjointe de (X, Y) peut se mettre sous la forme suivante (cf. Eq. (22)) :

$$P(x, y|\Phi) = \prod_{s \in S} P(y_s|x_s, \Phi) P(x_r|\Phi) \prod_{s \in >r} P(x_s|x_{s-}, \Phi) \quad (30)$$

En reprenant les notations du paragraphe 5.2, les deux étapes de l'algorithme EM s'écrivent comme suit.

Etape E : calcul de l'espérance

Il s'agit de calculer $Q(\Phi|\Phi^{(k)})$. Il vient :

$$\begin{aligned} Q(\Phi|\Phi^{(k)}) &= \mathbb{E}[\log P(X, y|\Phi)|y, \Phi^{(k)}] \\ &= \sum_s \mathbb{E}[\log P(y_s|X_s, \Phi)|y, \Phi^{(k)}] \\ &\quad + \sum_{s>r} \mathbb{E}[\log P(X_s|X_{s-}, \Phi)|y, \Phi^{(k)}] \\ &\quad + \mathbb{E}[\log P(X_r|\Phi)|y, \Phi^{(k)}] \quad \text{d'après (30)} \\ &= \sum_s \sum_i P(X_s = i|y, \Phi^{(k)}) \log P(y_s|X_s = i, \Phi) \\ &\quad + \sum_{s>r} \sum_i \sum_j P(X_s = j, X_{s-} = i|y, \Phi^{(k)}) \\ &\quad \log P(X_s = j|X_{s-} = i, \Phi) \\ &\quad + \sum_i P(X_r = i|y, \Phi^{(k)}) \log P(X_r = i|\Phi) \\ &= \sum_{n=0}^R \sum_{s \in S^n} \sum_i P(X_s = i|y, \Phi^{(k)}) \log f_i^n(y_s) \\ &\quad + \sum_{s>r} \sum_i \sum_j P(X_s = j, X_{s-} = i|y, \Phi^{(k)}) \log a_{ij} \\ &\quad + \sum_i P(X_r = i|y, \Phi^{(k)}) \log \pi_i \end{aligned}$$

Etape M : maximisation de l'espérance

Il s'agit de trouver les paramètres $\Phi = \{f_i^n(l), a_{ij}, \pi_i\}$ qui maximisent $Q(\Phi|\Phi^{(k)})$. Ce problème d'optimisation (sous contraintes : $\sum_l f_i^n(l) = 1, \sum_j a_{ij} = 1, \sum_i \pi_i = 1$) se résout classiquement à l'aide de la technique des multiplicateurs de Lagrange. On obtient alors les formules de ré-estimation suivantes (calcul de $\Phi^{(k+1)}$ en fonction de $\Phi^{(k)}$) :

$$\left\{ \begin{aligned} f_i^{n(k+1)}(l) &= \frac{\sum_{s \in S^n: Y_s=l} P(X_s = i|y, \Phi^{(k)})}{\sum_{s \in S^n} P(X_s = i|y, \Phi^{(k)})} = \frac{\sum_{s \in S^n: Y_s=l} \xi_s^{(k)}(i)}{\sum_{s \in S^n} \xi_s^{(k)}(i)} \\ a_{ij}^{(k+1)} &= \frac{\sum_{s>r} P(X_s = j, X_{s-} = i|y, \Phi^{(k)})}{\sum_{s>r} P(X_{s-} = i|y, \Phi^{(k)})} = \frac{\sum_{s>r} \Psi_s^{(k)}(i, j)}{\sum_{s>r} \xi_{s-}^{(k)}(i)} \\ \pi_i^{(k+1)} &= P(X_r = i|y, \Phi^{(k)}) = \xi_r^{(k)}(i) \end{aligned} \right.$$

Annexe B : Algorithme MCEM sur le quadarbre

On considère un processus markovien (X, Y) défini sur le quadarbre et vérifiant les propriétés du paragraphe 3.2. X est le processus caché, Y est observé. L'algorithme MCEM consiste à tirer, à l'itération k , et en chaque site s du quadarbre, M étiquettes $x_s^k(1), \dots, x_s^k(M)$ selon la loi $P(x|y, \Phi^{(k)})$.

En reprenant les notations du paragraphe 5.3, les deux étapes de l'algorithme MCEM s'écrivent comme suit.

Etape E : calcul de l'espérance

Dans l'étape E, l'espérance est approchée par :

$$\mathbb{E}_{X|Y, \Phi} [\log P(X, y|\Phi)|y, \Phi^{(k)}] \simeq \underbrace{\frac{1}{M} \sum_{m=1}^M \log P(x^k(m), y|\Phi)}_{\tilde{Q}(\Phi|\Phi^{(k)})} \quad (31)$$

$\tilde{Q}(\Phi|\Phi^{(k)})$ se décompose comme suit :

$$\begin{aligned} \tilde{Q}(\Phi|\Phi^{(k)}) &= \frac{1}{M} \sum_{m=1}^M [\sum_s \log P(y_s|x_s^k(m), \Phi) \\ &+ \sum_{s>r} \log P(x_s^k(m)|x_{s-}^k(m), \Phi) \\ &+ \log P(x_r^k(m)|\Phi)] \\ &= \frac{1}{M} \sum_{m=1}^M [\sum_{n=0}^R \sum_{s \in S^n} \log f_{x_s^k(m)}^n(y_s) \\ &+ \sum_{s>r} \log a_{x_{s-}^k(m)x_s^k(m)} \\ &+ \log \pi_{x_r^k(m)}] \end{aligned}$$

M : maximisation de l'espérance

Il s'agit de trouver les paramètres $\Phi = \{f_i^n(l), a_{ij}, \pi_i\}$ qui maximisent $\tilde{Q}(\Phi|\Phi^{(k)})$. Ce problème d'optimisation (sous contraintes : $\sum_l f_i^n(l) = 1, \sum_j a_{ij} = 1, \sum_i \pi_i = 1$) se résout classiquement à l'aide de la technique des multiplicateurs de Lagrange. On obtient alors les formules de ré-estimation suivantes :

$$\left\{ \begin{aligned} f_i^{n(k+1)}(l) &= \frac{\sum_{s \in S^n: Y_s=l} \sum_{m=1}^M \mathbb{I}_{x_s^k(m)=i}}{\sum_{s \in S^n} \sum_{m=1}^M \mathbb{I}_{x_s^k(m)=i}} \\ a_{ij}^{(k+1)} &= \frac{\sum_{s>r} \sum_{m=1}^M \mathbb{I}_{x_s^k(m)=j, x_{s-}^k(m)=i}}{\sum_{s>r} \sum_{m=1}^M \mathbb{I}_{x_{s-}^k(m)=i}} \\ \pi_i^{(k+1)} &= \frac{1}{M} \mathbb{I}_{x_r^k(m)=i} \end{aligned} \right. \quad (32)$$

En introduisant les notations :

$$\left\{ \begin{aligned} \xi_s^{(k)}(i) &\triangleq \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{x_s^k(m)=i} \\ \Psi_s^{(k)}(i, j) &\triangleq \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{x_s^k(m)=j, x_{s-}^k(m)=i} \end{aligned} \right. \quad (33)$$

on obtient des formules de ré-estimation similaires à celles obtenues pour l'algorithme EM (cf. Annexe A) :

$$\left\{ \begin{aligned} f_i^{n(k+1)}(l) &= \frac{\sum_{s \in S^n: Y_s=l} \xi_s^{(k)}(i)}{\sum_{s \in S^n} \xi_s^{(k)}(i)} \\ a_{ij}^{(k+1)} &= \frac{\sum_{s>r} \Psi_s^{(k)}(i, j)}{\sum_{s>r} \xi_{s-}^{(k)}(i)} \\ \pi_i^{(k+1)} &= \xi_r^{(k)}(i) \end{aligned} \right. \quad (34)$$

Remerciement

Cette étude a été soutenue par le MENESR par l'attribution d'une allocation de recherche, ainsi que par le GDR PRC ISIS. Nous remercions Eric Fabre pour de nombreuses et fructueuses discussions.

BIBLIOGRAPHIE

- [1] R. Azencott and C. Graffigne, "Non supervised segmentation using multi-level Markov Random Fields," In *Int. Conf. Pattern Rec.*, pp. III.201-III.204, The Hague, The Netherlands, Sept. 1992.
- [2] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah, and A.S. WILLISKY, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Information Theory*, Vol. 38, No. 2, pp. 766-784, 1992.
- [3] M. Basseville, A. Benveniste, and A.S. Willisky, "Multiscale autoregressive processes - Part I : Schur-Levinson parametrizations," *IEEE Trans. Signal Processing*, Vol. 40, No. 8, pp. 1915-1934, 1992.
- [4] L.E. Baum, T. Petrie, G. Soules and N. Weiss, "A Maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *IEEE Ann. Math. Stat.*, Vol. 41, pp. 164-171, 1970.
- [5] L.E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," *Inequalities*, Vol. 3, pp. 1-8, 1972.
- [6] 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 : pages 433-454, 1995.
- [7] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statist. Soc.*, Vol. 48, Serie B, No. 3, pp. 259-302, 1986.
- [8] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, Vol. 3, No. 2, pp. 162-177, 1994.
- [9] G. Celeux, D. Chauveau and J. Diebolt, "On stochastic versions of the EM algorithm" Technical Report INRIA, No 2514, 1995.

- [10] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition*, Vol. 22, No. 6, pp.747–761, 1989.
- [11] P. Charbonnier, L. Blanc-Féraud, and M. Barlaud, "Noisy image restoration using multiresolution Markov random fields," *J. of Visual Com. and Image Representation*, Vol. 3, No. 4, pp. 338–346, 1992.
- [12] K.C. Chou, S.A. Golden, and A.S. Willsky, "Multiresolution stochastic models, data fusion and wavelet transforms," *Signal Processing*, Vol. 34, No. 3, pp. 257–282, 1993.
- [13] H. Derin and P. Kelly, "Discrete-Index Markov-Type Random Processes," *Proc. IEEE*, Vol. 77, No 10 : pages 1485–1510, 1989.
- [14] X. Descombes, M. Sigelle, and F. Prêteux, "Estimating Gaussian Markov random field parameters in a non-stationary framework : application to remote sensing imaging," Submitted to *IEEE Trans. Image Proc.*, 1994.
- [15] P.A. Devijver et M. Dokesel, "Champs aléatoires de Pickard et modélisation d'images," *Traitement du Signal*, Vol. 5, No. 5, pp. 131–150, 1988.
- [16] E. Fabre, *Traitement du signal multirésolution : conception de lisseurs rapides pour une famille de modèles*, Thèse de Doctorat, Université de Rennes 1, France, Dec. 1994.
- [17] P.W. Fieguth, W.C. Karl, A.S. Willsky and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of Topex/Poseidon Satellite Altimetry," *IEEE Trans. Geoscience Remote Sensing*, Vol. 33, No 2, pages 280–292, March 1995.
- [18] P.W. Fieguth, *Application of Multiscale Estimation to Large Multidimensional Imaging and Remote Sensing Problems*, PhD thesis, No LIDS-TH-2302, MIT Laboratory of Information and Decision Systems, May 1995.
- [19] G.D. Forney, "The Viterbi Algorithm," *Proc. IEEE*, Vol. 61, No. 3, pp. 268–278, 1973.
- [20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, pp. 721–741, 1984.
- [21] B. Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 2, pp. 164–180, Feb. 1989.
- [22] C. Graffigne, F. Heitz, F. Prêteux, M. Sigelle, and J. Zerubia, "Modèles markoviens hiérarchiques pour l'analyse d'images", Ed. CNRS-GdR TdSI, Novembre 1994.
- [23] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle, J. Zerubia, "Hierarchical Markov random field models applied to image analysis : a review," soumis à *IEEE Trans. Information Theory*, Mars 1997, également : In *SPIE Conference on Neural, Morphological and Stochastic Methods in Image and Signal Processing*, San Diego, USA, July 1995.
- [24] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov Random Fields," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 15, No. 12, pp. 1217–1232, 1993.
- [25] F. Heitz, P. Pérez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," *CVGIP : Image Understanding*, Vol. 59, No. 1, pp. 125–134, 1994.
- [26] Z. Kato, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model and multitemperature annealing for parallel image classification," *Graph. Model and Image Proc.*, Vol. 58, No. 1, pp. 18–37, 1996.
- [27] S. Krishnamachari and R. Chellappa, "Multiresolution Gauss-Markov Random Field Models for Texture Segmentation," *IEEE Trans. Image Proc.*, Vol. 6, No. 2, pp. 251–267, Feb. 1997.
- [28] J.-M. Laferté P. Pérez, and F. Heitz, "Global non-linear multigrid optimization for image analysis tasks," In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 533–536, Adelaïde, Australia, April 1994.
- [29] J.-M. Laferté, F. Heitz, P. Pérez, and E. Fabre, "Hierarchical Statistical Models for the Fusion of Multiresolution Image Data," In *Proc. Int. Conf. Computer Vision*, pp. 908–913, Cambridge, USA, June 1995.
- [Laferte96] J.-M. Laferté, F. Heitz and P. Pérez, "A multiresolution EM algorithm for unsupervised image classification," In *Proc. Int. Conf. Pattern Recognition ICPR 96*, Vienna, Austria, Aug. 1996.
- [31] J.-M. Laferté, Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques. Application à la fusion de données multirésolutions, Thèse de doctorat, Université de Rennes I, Octobre 1996.
- [32] M.R. Luetzgen, W. Karl, and A.S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, Vol. 3, No. 1, pp. 41–64, 1992.
- [33] M. Luetzgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, Vol. 41, No. 12, pp. 3377–3395, 1993.
- [34] M.R. Luetzgen and A.S. Willsky, "Likelihood calculation for a class of multiscale stochastic models with application to texture discrimination," *IEEE Trans. Image Processing*, Vol. 4, No 2 : pages 194–207, 1995.
- [35] J. Marroquin, S. Mitter and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. American Statistical Association*, Vol. 82, No 397 : pages 76–89, 1987.
- [36] P. Pérez et F. Heitz "Une approche multi-échelle à l'analyse d'images par champs markoviens," *Traitement du Signal*, Vol. 9, No. 6, pp. 459–466, 1992.
- [37] P. Pérez and F. Heitz, "Restriction of a Markov Random Field on a Graph and Multiresolution Statistical Image Modeling," *IEEE Trans. Information Theory*, Vol. 42, No 1 : pages 180–190, 1996.
- [38] F. Prêteux and X. Descombes, "Synthèse et analyse de textures par coopération de processus multi-échelles", In *Proc. 8ème congrès AFCET/INRIA Reconnaissance des Formes et Intelligence Artificielle*, pp. 1015–1022, Lyon, France, 1991.
- [39] R.A. Redner and H.F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, Vol. 26, No. 2, pp. 195–239, April 1984.
- [40] C.S. Regazzoni, F. Arduini, and G. Vernazza, "A multilevel GMRF-based approach to image segmentation and restoration," *Signal Processing*, Vol. 34, No. 1, pp. 43–67, 1993.
- [41] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No. 2, pp. 129–139, 1986.
- [42] G.C. Wei and M.A. Tanner, "A Monte-Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *J. American Stat. Assoc.*, Vol. 85, pp. 699-704, 1990.
- [43] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, 1990.
- [44] J. Zhang, J.W. Modestino and D.A. Langan, "Maximum-Likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Proc.*, Vol. 3, No 4, pp. 404–420, July 1994.

Manuscrit reçu le 22 juillet 1997.

Modèles markoviens d'images et algorithmes d'estimation

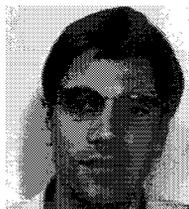
LES AUTEURS

Jean-Marc LAFERTÉ



Jean-Marc Laferté est né en 1968. Il a obtenu un Magistère de Mathématiques/Informatique, puis un Doctorat de l'Université de Rennes I en 1996. Ses travaux de thèse se sont déroulés à l'Irisa/Inria-Rennes (projet Temis) et ont porté sur la représentation des images par modèles markoviens causaux. Il a ensuite travaillé dans le projet Model de l'Irisa, sur un outil de génération de chaînes de Markov. Depuis mars 1998 il travaille comme ingénieur dans la société Grenat Productique à Nantes.

Fabrice HEITZ



Fabrice Heitz est né en 1961. Diplômé de Télécom Bretagne en 1984, il a soutenu sa thèse à Télécom Paris en 1988. Il a été de 1988 à 1994 Chargé de Recherches à l'INRIA de Rennes (dans le projet Temis). Il est actuellement Professeur à l'école Nationale Supérieure de Physique de Strasbourg et anime un groupe de recherche en analyse multi-images au sein du LSIT (Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection, UPRES-A CNRS 7005). F. Heitz est éditeur associé pour la revue IEEE Transactions on Image Processing depuis 1996. Ses domaines d'intérêt incluent la modélisation statistique et les modèles déformables appliqués à l'analyse multi-images et à l'imagerie médicale.

Patrick PÉREZ



Patrick Pérez est né en 1968. Diplômé de l'école Centrale Paris en 1990, il a effectué sa thèse à l'Irisa/Inria-Rennes. Après un an de séjour post-doctoral au Dpt. of Applied Mathematics à Brown University (Rhode-Island, USA), il a rejoint l'INRIA en 1994 en tant que Chargé de Recherches. Ses thèmes de recherches portent sur les modèles statistiques pour le traitement de problèmes inverses de grande dimension, en analyse d'images. Les domaines d'applications privilégiés sont l'analyse du mouvement apparent dans les séquences d'images d'une part, et la fusion multirésolution d'autre part (en télédétection et en imagerie sonar).