



Dans le cadre de l'Ecole Doctorale SSTO
Sciences des Systèmes Technologiques et Organisationnels
En partenariat avec l'Université de Technologie de Troyes

①

Spécialité : Optimisation et Sécurité des Systèmes

Par

Mlle Dalila Benboudjema

Thèse présentée pour l'obtention du grade de Docteur
de l'Institut National des Télécommunications

CHAMPS DE MARKOV TRIPLETS ET SEGMENTATION BAYESIENNE NON SUPERVISEE D'IMAGES

Soutenue le 12 Décembre 2005 devant le jury composé de :

Rapporteurs

M. Fabrice Heitz Professeur à l'Université de Strasbourg
M. Alain Hillion Professeur à Télécom Bretagne

Examineurs :

M. Gérard Govaert Professeur à l'UTC
M. Patrick Pérez Directeur des Recherches à l'IRISA
M. Wojciech Pieczynski Professeur à INT, directeur de la thèse
M. André Smolarz Maître de Conférence à l'UTT
Mme Florence Tupin Maître de Conférence à Télécom Paris, invitée



Dans le cadre de l'Ecole Doctorale SSTO
Sciences des Systèmes Technologiques et Organisationnels
En partenariat avec l'Université de Technologie de Troyes

①

Spécialité : Optimisation et Sûreté des Systèmes

Par

Mlle Dalila Benboudjema

Thèse présentée pour l'obtention du grade de Docteur
de l'Institut National des Télécommunications

CHAMPS DE MARKOV TRIPLETS ET SEGMENTATION BAYESIENNE NON SUPERVISEE D'IMAGES

Soutenue le 12 Décembre 2005 devant le jury composé de :

Rapporteurs

M. Fabrice Heitz	Professeur à l'Université de Strasbourg
M. Alain Hillion	Professeur à Télécom Bretagne

Examineurs :

M. Gérard Govaert	Professeur à l'UTC
M. Patrick Pérez	Directeur des Recherches à l'IRISA
M. Wojciech Pieczynski	Professeur à INT, directeur de la thèse
M. André Smolarz	Maître de Conférence à l'UTT
Mme Florence Tupin	Maître de Conférence à Télécom Paris, invitée

*À mes parents,
À tous ceux qui me sont chers . . .*

Remerciements

Le travail de cette thèse a été effectué au département Communications, Images, et Traitement de l'Information (CITI) de l'Institut National des Télécommunications d'Evry (INT-Evry) sous la direction de Monsieur le Professeur Wojciech Pieczynski, directeur du département CITI.

J'adresse mes très vifs remerciements à mon directeur de thèse Wojciech Pieczynski pour m'avoir accueillie, dès mon stage de DEA, au sein de son laboratoire. Mon parcours, depuis et durant toutes mes années de recherche, a été grandement facilité par sa disponibilité et son ouverture d'esprit. Ses conseils et orientations judicieux m'ont permis de bénéficier pleinement de son expérience et de son savoir, ce qui m'a aidée à évoluer avec attrait dans le monde passionnant de l'image. Aussi, je prie Monsieur Wojciech Pieczynski de trouver, dans ces remerciements, l'expression de ma profonde reconnaissance et ma sincère sympathie. Je suis très heureuse de le compter parmi les membres du jury.

J'exprime ma reconnaissance à Monsieur Gérard. Govaert, professeur à Université de Technologique de Compiègne pour m'avoir honorée en sa qualité de président du jury.

Mes remerciements vont également à Messieurs les Professeurs Fabrice Heitz, directeur du laboratoire des Sciences de l'Image et de l'Informatique et de la Télédétection à l'Université Louis Pasteur de Strasbourg, et Alain Hillion, Directeur Scientifique de l'Ecole Nationale Supérieure de Télécommunications de Bretagne (ENST-Bretagne), qui ont accepté ce travail.

Je suis également très honorée de compter parmi les membres du jury :

- Monsieur Patrick Pérez, directeur des recherches à l'IRISA ;
- Monsieur André Smolarz, maître de conférence à l'Université de Technologie de Troyes ;
- Madame Florence Tupin, maître de conférence à l'ENST-Paris.

que je remercie très vivement.

Enfin, je remercie tous les membres du département Communications, Images, et Traitement de l'Information de l'INT-Evry ainsi que toutes les personnes avec qui j'ai eu l'occasion de collaborer.

Résumé

La segmentation non supervisée d'images est parmi les problèmes clé en traitement d'images. Parmi les différents modèles et approches développées, une partie des méthodes statistiques, couramment utilisées, sont fondées sur le modèle par champs de Markov Cachés (CMC). Ce succès est principalement dû à l'aptitude du modèle de prendre en compte des dépendances spatiales des variables aléatoires, même lorsqu'elles sont en très grand nombre, pouvant dépasser le million. Dans un tel modèle le champ caché X est supposé markovien et doit être estimé à partir du champ observé Y . Un tel traitement est possible du fait de la markovianité de X conditionnellement à Y . Ce modèle a été ensuite généralisé aux champs de Markov couples (CMCouples), où l'on suppose directement la markovianité du couple (X, Y) , qui offrent les mêmes possibilités de traitements que les CMC et permettent de mieux modéliser le bruit ce qui permet, en particulier, de mieux prendre en compte l'existence des textures. Par la suite, les CMCouples ont été généralisés aux champs de Markov triplet (CMT), où la loi du couple (X, Y) est une loi marginale d'un champ de Markov triplet $T = (X, U, Y)$, avec un champ auxiliaire U . L'objet de cette thèse est d'étudier les CMT. Deux variantes originales sont présentées : les champs de Markov évidentiels (CME), permettant la modélisation des incertitudes sur les paramètres ; et les champs de Markov Triplets Adaptés (CMTA), permettant la modélisation des différentes stationnarités de l'image cachée. Pour une segmentation non supervisée, deux méthodes originales d'estimation des paramètres sont proposées. La première est fondée sur le principe du gradient stochastique, et la seconde est fondée sur le principe de l'estimation conditionnelle itérative (ECI) et les moindres carrés. Cette dernière est ensuite généralisée aux cas des images non stationnaires avec du bruit non Gaussien corrélé, et à marginales quelconques. Elle permet de rechercher la forme même des différentes lois marginales en utilisant le système de Pearson, ainsi que d'estimer tous les paramètres du modèle. L'intérêt des différentes méthodes non supervisées obtenues est attesté par des simulations informatiques ainsi que par les premiers traitements des images réelles.

Mots clés : Champs de Markov Cachés (CMC), Champs de Markov Couples (CMCouples), Champs de Markov Triplets (CMT), Champs de Markov Evidentiels (CME), segmentation statistique non supervisée, estimation des paramètres, Estimation Conditionnelle Itérative (ECI), Gradient Stochastique (GS), Système de Pearson, images non stationnaires.

Abstract

Image segmentation is a fundamental and yet difficult task in machine vision. Several models and approaches have been proposed, and the ones which have probably received considerable attention are hidden Markov fields (HMF) models. In such model the hidden field X which is assumed Markovian, must be estimated from the observed –or noisy- field Y . Such processing is possible because the distribution X conditional on the observed process Y remains Markovian. This model has been generalized to the Pairwise Markov field (PMF) which offer similar processing and superior modelling capabilities. In this model we assume directly the Markovianity of the couple (X, Y) . Afterwards, triplet Markov fields (TMF) which are the generalization of the PMF, have been proposed. In such model the distribution of the couple (X, Y) is the marginal distribution of a Markov field $T = (X, U, Y)$, where U is latent process.

The aim of this thesis is to study the TMF models. Two original models are presented: the Evidential Markov field (EMF) allowing to model the evidential aspects of the prior information and the adapted triplet Markov field (ATMF), allowing to model the simultaneous presence of different stationarities in the class image. For the unsupervised processing, two original approaches of estimation the model's parameters have been proposed. The first one is based on the stochastic gradient and the second one is based on the iterative conditional estimation (ICE) and the least square method, as well. The latter, have then been generalized to the non stationary images with non Gaussian correlated noise, which uses the Pearson system to find the natures of margins of the noise, which can vary with the class. Experiments indicate that the new models and related processing algorithms can improve the results obtained with the classical ones.

Index Terms : Hidden Markov Field (HMF), Pairwise Markov Field (PMF), Triplet Markov Field (TMF), Evidential Markov Field (EMF), unsupervised statistical segmentation, Parameters estimation, Iterative conditional estimation (ICE), Stochastic Gradient (SG) Pearson system, non stationary images.

Table des matières

Introduction	5
I Généralités	9
I.1 Champ de Markov et distribution de Gibbs	10
I.1.1 Système de voisinage et cliques	10
I.1.2 Champs de Markov	12
I.1.3 Champs de Gibbs	12
I.1.4 Equivalence champ de Markov – champ de Gibbs	13
I.1.5 Exemple : Modèle de Potts	14
I.2 Simulation d'un champ de Markov	16
I.2.1 Algorithme de Metropolis	16
I.2.2 Echantillonneur de Gibbs	17
I.2.3 Algorithme de Swendsen – Wang	19
I.2.4 Conclusion	20
I.3 Modèle markovien caché pour la segmentation d'images	20
I.4 Apprentissage	21
I.5 Estimation bayésienne	22
I.5.1 Estimateur maximum a posteriori	23
I.5.2 Algorithme du recuit simulé	25
I.5.3 Estimateur du mode conditionnel itéré	26
I.5.4 Estimateur du mode des marginales a posteriori	27
I.6 Conclusion	28

II	Modélisation markovienne	29
II.1	Champs de Markov cachés	30
II.1.1	Champs de Markov cachés avec bruit indépendant	31
II.1.2	Champs de Markov cachés avec bruit corrélé	33
II.1.3	Champs de Markov cachés généraux	34
II.2	Champs de Markov couples	35
II.2.1	Simulation des CM Couples gaussiens	41
II.3	Champs de Markov triplets	42
II.4	Champs de Markov évidentiels	46
II.4.1	Notions de bases de la théorie de l'évidence	46
II.4.2	Fusion de Dempster – Shafer	48
II.4.3	Fusion DS dans les champs de Markov	49
II.4.4	Critique de la règle de la fusion de Dempster – Shafer	51
II.5	La non stationnarité	53
II.5.1	Champs de Markov triplets adaptés à bruit indépendant	53
II.5.2	Champs de Markov triplets adaptés	56
II.6	Conclusion	58
III	Estimation des paramètres	59
III.1	Apprentissage à données complètes	60
III.1.1	Méthode de codage	60
III.1.2	Le gradient stochastique	62
III.1.3	Méthode de Derin et Elliott	63
III.2	Apprentissage à données incomplètes	66
III.2.1	Algorithme Expectation Maximisation	66
III.2.2	Algorithme Stochastique Expectation Maximisation	67
III.2.3	Algorithme Estimation conditionnelle itérative	68
III.2.4	Algorithme du gradient stochastique	69
III.3	Estimation des paramètres du champs de Markov triplet	70
III.3.1	Gradient stochastique adapté	71
III.3.2	Nouvelle méthode	72
III.3.3	Analyse de performances	76
III.4	Conclusion	80

IV Simulation et segmentation non supervisée d'images	81
IV.1 Segmentation couple et triplet	81
IV.1.1 Simulations	82
IV.1.2 Image de synthèse	86
IV.1.3 Image réelle	88
IV.2 Segmentation évidentielle	90
IV.3 Segmentation triplet adaptée	94
IV.3.1 Champ de Markov triplets adapté à bruit indépendant	94
IV.3.1.1 Simulations	95
IV.3.1.2 Image de synthèse	100
IV.3.1.3 Images réelles	101
IV.3.2 Champs de Markov adapté	103
IV.3.2.1 Simulation	103
IV.3.2.2 Image de synthèse	105
IV.3.2.3 Image réelle	106
IV.4 Conclusion	107
V Extension au cas non gaussien et mélange généralisé	109
V.1 Modélisation statistique	110
V.2 Système de Pearson	112
V.3 NM généralisée	114
V.4 Expérimentations	115
V.5 Conclusion	120
Conclusion et perspectives	121
Annexe	123
Liste de notations	127
Bibliographie	131

Introduction

Les champs de Markov cachés ont connu un essor considérable, en traitement d'images, à partir des années quatre-vingts. Leur intérêt est de pouvoir prendre en compte l'information contextuelle dans une image, de manière mathématiquement rigoureuse et pour les contextes de très grande taille. En segmentation d'images, problématique traitée dans cette thèse, on considère deux champs aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, où S est l'ensemble des pixels. Chaque X_s prend ses valeurs dans un ensemble fini de $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque Y_s prend ses valeurs dans l'ensemble des nombres réels R . Le problème de la segmentation est celui de l'estimation de $X = (X_s)_{s \in S}$ à partir de $Y = (Y_s)_{s \in S}$. Ce problème ne peut alors être traité directement à cause du nombre trop grand de réalisations possibles de $X = (X_s)_{s \in S}$.

En effet, si N est la taille de S , qui est couramment de l'ordre de 256×256 (mais peut également atteindre 1024×1024), le nombre de réalisation de $X = (X_s)_{s \in S}$ est de k^N , ce qui ne peut pas être traité directement pour toute loi du couple (X, Y) . On est ainsi obligé de considérer des lois (X, Y) particulières et les champs de Markov cachés sont largement utilisés pour cette modélisation. Après les articles fondateurs [Bes 74, Bes 86, DeE 87, GeG 84, MMP 87], des centaines de publications ont été présentées sur diverses applications des champs de Markov cachés, en imagerie et il est impossible de les citer toutes. Mentionnons les livres suivants [Cha 00, Guy 95, Li 95, Mai 03, Win 03], incluant chacun une riche bibliographie, ou encore les articles de synthèse [Per 98, Pie 03]. Citons également quelques exemples d'articles traitant des généralités ou des applications particulières en imagerie médicale [MMS 02], en vision [KeH 98, HPB 94], en télédétection [Ric 05, PSZ 05], en imagerie radar [DMP 97, TNT 04, TuR 05], ou encore en imagerie sonar [MCP 00, RPB 03]. De plus, diverses publications montrent que les résultats obtenus avec ce type de modèles peuvent être, dans certaines situations, spectaculaires.

Le succès des champs de Markov cachés est dû au fait qu'il est possible, dans leur cadre, de mettre en place des estimateurs bayésiens de X , à partir de Y , grâce à des techniques générales de simulation dites "*méthodes de Monté Carlo par chaînes de Markov*" (MCMC de l'anglais "*Monte Carlo Markov Chains*" [Rob 91], parmi d'autres). Les techniques MCMC

sont applicables à condition que la loi de X conditionnelle à $Y = y$ (sa loi "*a posteriori*") soit de Markov. Il a été remarqué que cette condition est vérifiée dès que le couple (X, Y) est de Markov (c'est le modèle "*champs de Markov couples*" [PiT 00]). On arrive ainsi à un modèle plus général, permettant de mieux modéliser le "*bruit*", qui est représenté par la loi de Y conditionnelle à $X = x$. Il est à noter que dans un champ de Markov couple, la loi de X qui est la loi marginale d'un champ de Markov (X, Y) , n'est plus nécessairement de Markov. Dans un certain sens, les champs de Markov couples marquent une rupture par rapport aux modèles classiques, dans lesquels le champ X est toujours markovien.

Notre travail concerne les champs de Markov triplets (CMT) qui sont des modèles généralisant les champs de Markov couples. Dans un CMT, on considère un champ aléatoire auxiliaire $U = (U_s)_{s \in S}$, chaque U_s prend ses valeurs dans un ensemble fini de $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, et on pose la markovianité du triplet $T = (X, U, Y)$ [PBL 02]. C'est un modèle plus général que les champs de Markov couples. En effet, (X, Y) n'est plus nécessairement de Markov. Cependant, les segmentations bayésiennes analogues à celles effectuées dans les cas classiques restent possibles. On arrive à une famille très riche de modèles dont les différents potentiels commencent seulement à être explorés. Nous proposons trois types de contributions :

- nous proposons d'utiliser des CMT particuliers (avec le bruit gaussien) pour modéliser des champs aléatoires non stationnaires. Deux manières différentes sont étudiées : (i) modéliser les différentes stationnarités présentes par $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, et (ii) utiliser la théorie de l'évidence ;
- nous proposons une méthode originale d'estimation des paramètres d'une famille des CMT. Appliquée à l'estimation des paramètres des modèles du premier point ci-dessus, elle permet, en particulier, la segmentation non supervisée des images non stationnaires ;
- d'une part, nous généralisons le modèle (i) aux bruits à marginales quelconques. D'autre part, lorsque ces marginales sont dans le système de Pearson, nous généralisons la méthode du deuxième point en une méthode permettant de traiter les bruits corrélés non gaussiens, avec les formes des marginales pouvant varier avec les classes (notre méthode généralise celle proposée dans le cas des bruits spatialement indépendants dans [DMP 97]).

Le plan de la thèse est le suivant :

Le premier chapitre est consacré à des généralités sur les champs de Markov cachés. La problématique générale est précisée et les méthodes classiques de simulation, de segmentation et d'estimations des paramètres, sont évoquées. Il permet d'introduire les différentes notions et alléger l'exposé des chapitres suivants, même si certains points sont repris par la suite.

Le chapitre II traite des champs de Markov cachés, couples et triplets. Trois niveaux de généralités dans les champs de Markov cachés sont précisés, dont le troisième, où aucune hypothèse classique n'est posée, est original. Deux modèles CMT particuliers originaux, permettant de prendre en compte les non stationnarités dans les modèles classiques, sont décrits.

Le chapitre III est consacré au problème de l'estimation des paramètres. Nous proposons deux méthodes originales dont l'une est une extension aux CMT de la méthode du gradient stochastique de Younes [You 89] et une deuxième est fondée sur le principe général de "*Iterative Conditional Estimation*" (ICE [Pie 92]). Diverses simulations ont été effectuées dont certaines feront l'objet d'une comparaison.

Dans le chapitre IV nous présentons une étude dont le but est de montrer l'intérêt des méthodes non supervisées de segmentation fondées sur les modèles triplets originaux du chapitre deux, et sur la nouvelle méthode d'estimation des paramètres proposée dans le chapitre trois. Nous étudions successivement des simulations du modèle, des images synthétiques bruitées artificiellement et des images réelles.

Dans le chapitre V nous étendons les modèles triplets $T = (X, U, Y)$, considérés "*gaussiens*" dans les chapitres précédents, au cas "*non gaussien*" et ce, en supposant les marginales $p(y_s | x_s, u_s)$ quelconques. Nous présentons également une méthode permettant la recherche de leurs formes et l'estimation de leurs paramètres, lorsqu'elles se trouvent dans le système de Pearson.

Chapitre I

Généralités

L'introduction des méthodes statistiques, pour l'analyse d'images a commencé avec la segmentation d'images satellitaires, dont l'un des pionniers est Switer (1969) [Swi 69]. Les champs de Markov introduits par Lévy en 1948 [Lev 48], ont montré leur intérêt dans la modélisation de l'image à partir des années 70 [Bes 74, Jai 81], notamment par les travaux de Geman et Geman [GeG 84]. Les champs de Markov introduisent la prise en compte de la notion de dépendance spatiale dans l'image et permettent de pondérer la portée de l'influence de ces dépendances, offrant ainsi un cadre mathématique cohérent pour divers traitements. En particulier, les techniques d'acquisition de données (optique, radar, sonar, scanner,...) créent des superpositions entre les mesures faites sur des pixels voisins, engendrant ainsi une dépendance spatiale qui rendent d'autant plus attrayante et nécessaire l'utilisation des modèles markoviens. Ainsi, l'intérêt de l'utilisation des champs de Markov en segmentation d'images est de mieux modéliser l'image. Ceci permet d'obtenir de bons résultats de segmentation qui, associée aux champs de Markov et à la théorie bayésienne, se ramène, comme nous le verrons par la suite, à un problème d'optimisation.

Les enjeux majeurs en analyse d'images, fondée sur un modèle probabiliste, tels que les champs Markoviens considérés dans cette thèse, sont le choix du modèle, l'estimation des paramètres et la segmentation - thèmes qui seront développés de manière plus détaillée dans les chapitres suivants. Le but de ce chapitre est d'introduire la théorie des champs markoviens d'une manière générale et de présenter les algorithmes classiques de simulation, d'estimation et d'optimisation. A cet effet, nous commençons par rappeler quelques concepts généraux

concernant les distributions de champs de Markov, comme le système de voisinage et équivalence champ de Markov – champ de Gibbs. On décrit par la suite les différentes méthodes classiques de simulation, tel que l'échantillonneur de Gibbs ou l'algorithme de Metropolis. On présente ensuite les différentes techniques d'estimation des paramètres du modèle, ainsi que les différents estimateurs existants. Enfin, on évoque les diverses stratégies de segmentation d'images. Pour une présentation plus formelle et plus complète des champs de Markov, nous renvoyons le lecteur à [Jai 89, Win 03, Mai 03, Cha 00, EAM 95, Li 95] dont nous nous sommes largement inspirés.

I. 1 Champ de Markov et distribution de Gibbs

La théorie des champs de Markov est une branche de la théorie des probabilités pour l'analyse des dépendances spatiales de phénomènes physiques.

Dans cette section nous commençons par rappeler quelques notions de base, tels le système de voisinage et clique. Ensuite, nous présentons la définition mathématique des champs de Markov et distribution de Gibbs, ainsi que l'équivalence entre champs de Markov et champs de Gibbs (théorème de Hammersley – Clifford).

I.1.1 Système de voisinage et cliques

Systeme de voisinage

En analyse d'images, l'image est représentée par une grille $S = \{s = (i, j); \quad 0 \leq i \leq n-1; \quad 0 \leq j \leq m-1\}$ rectangulaire finie bidimensionnelle de taille $n \times m$, contenant un ensemble de "sites" $s \in S$, appelés également "pixels". On appelle un système de voisinages toute famille $V = (V_s)_{s \in S}$ vérifiant les propriétés suivantes :

- $s \notin V_s$, i.e. un site n'est pas voisin à lui même ;
- $s \in V_t \Leftrightarrow t \in V_s$, qui signifie que la relation de voisinage est symétrique

A titre d'exemple, l'ensemble des "voisins" d'un site s peut être défini comme un ensemble de sites situés dans un rayon r : $V_s = \{t \in S / d(s,t) \leq r, t \neq s\}$, où $d(s,t)$ est la distance euclidienne entre le pixel s et le pixel t , et r est à valeur dans l'ensemble des entiers [Li 95].

Notons que les systèmes de voisinage les plus usuels, représentés en FIG. 1, sont :

- Système "4 - connexité", également appelé voisinage du premier ordre;
- Système "8 - connexité", également appelé voisinage du second ordre.

A partir d'un système de voisinage, un système de clique est déduit.

Clique

Un sous ensemble c de S est appelé clique relative au système de voisinage V , si c est un singleton ou si tous les sites distincts de c sont voisins. Une clique est dite d'ordre P , si elle contient P éléments, en d'autres termes, l'ordre d'une clique est par définition son cardinal. FIG. 1 illustre la notion de cliques dans le cas de voisinage en 4-connexité et en 8-connexité. On notera C l'ensemble des cliques associées à un système de voisinage.

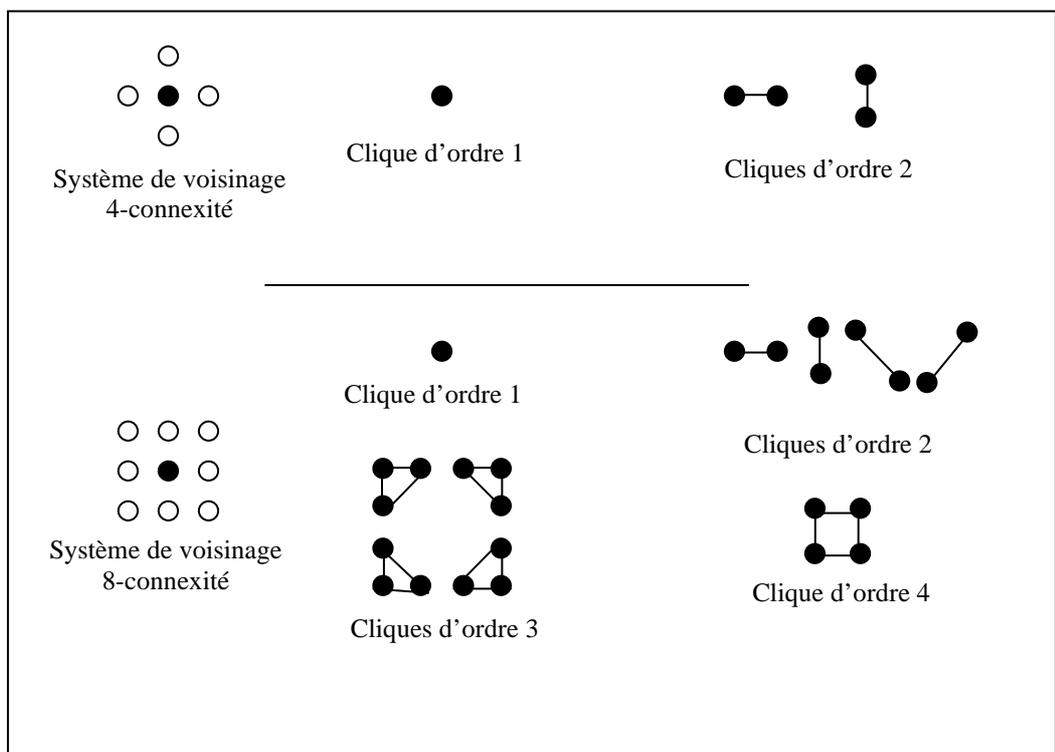


FIG. 1 – Cliques associées à un système de voisinage en 4-connexité et en 8-connexité

I.1.2 Champs de Markov

Soit S un ensemble de pixels et $X = (X_s)_{s \in S}$ une famille de variables aléatoires, définie sur S , où chaque X_s est à valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$. Une telle famille sera dans la suite appelée "*champ aléatoire*". Soit un système de voisinage V . On notera par la même lettre p les diverses lois de probabilité liées à X .

On dit qu'un champ X est un champ de Markov relativement à V , si et seulement si les deux propriétés suivantes sont vérifiées [Li 95]:

- Positivité:

$$\forall x \in \Omega^{\text{Card}(S)}, p(x) > 0 \quad (\text{I.1})$$

- Markovianité:

$$p(x_s | x_t, t \neq s) = p(x_s | x_t, t \in V_s) \quad (\text{I.2})$$

La propriété (I.2) signifie que la probabilité en un site s conditionnelle à tous les autres sites n'est fonction que de la configuration du voisinage du site considéré. En d'autres termes, la probabilité d'apparition d'une classe donnée en un site s conditionnellement au reste du champ, est égale à sa probabilité d'apparition connaissant simplement ses voisins.

Notons que tous les champs positifs aléatoires peuvent être considérés comme markoviens, à condition de choisir un système de voisinage suffisamment grand [Li 95]. Cependant, l'intérêt de restreindre le voisinage est de permettre des temps de calculs exploitables.

I.1.3 Champs de Gibbs

Soit S un ensemble de pixels muni d'un système de voisinage V , et soit C l'ensemble des cliques c associées à V . Le champ X est un champ de Gibbs (CG), si et seulement si sa loi est définie par:

$$p(x) = \frac{1}{Z} \exp(-U(x)) \quad (I.3)$$

$$\text{avec : } U(x) = \sum_{c \in C} \varphi_c(x_c), \quad (I.4)$$

qui est dite "*fonction énergie*", et :

$$Z = \sum_{x \in \Omega^{Card(S)}} \exp(-U(x)) \quad (I.5)$$

Une telle loi est dite une "*distribution de Gibbs*". Notons que φ_c sont des applications de $\Omega^{Card(c)}$ dans R appelées "*fonction potentiel*" avec x_c la restriction de x à c et Z est une constante de normalisation, également appelée "*fonction de partition*" qui fait de $p(x)$ une probabilité. Etant donné que cette constante est une somme sur l'ensemble de toutes les configurations possibles du champ X , son calcul est dans la pratique généralement impossible. Par exemple, pour une image binaire ($Card(\Omega) = 2$) de taille 256×256 pixels, le calcul de Z nécessite une somme sur $2^{256 \times 256}$ configurations possibles, ce qui est rédhibitoire.

I.1.4 Equivalence champ de Markov – champ de Gibbs

Comme nous l'avons évoqué précédemment, le champ de Markov (CM) est caractérisé par sa propriété locale (markovianité) tandis que le champ de Gibbs (CG) est caractérisé par sa propriété globale (distribution de Gibbs). Le théorème de Hammersley – Clifford (1971), dont une démonstration peut être vue dans [Bes 74], établit l'équivalence entre ces deux propriétés.

Théorème

Soit S un ensemble de pixels muni d'un système de voisinage V . Un champ X sur S est un champ de Markov relativement à V , si et seulement si X est un champ de Gibbs de potentiel associé à V .

L'intérêt pratique de ce théorème est qu'il permet d'accéder d'une manière simple à une forme exploitable des probabilités jointes et ce, en spécifiant les fonctions potentiel φ_c

définies sur les restrictions x_c de x aux cliques c . Afin de calculer la probabilité jointe du champ de Markov qui est une distribution de Gibbs, il est nécessaire d'évaluer la fonction de partition (I.5). Comme la somme englobe l'ensemble des configurations, ce calcul est en général impossible. Cependant, il reste possible de simuler des réalisations de ce champ à partir de ses caractéristiques locales, grâce à des méthodes de relaxation que nous allons voir dans la section suivante. C'est cette possibilité de simulations, qui s'avère être un puissant outil, qui est à l'origine des champs de Markov en imagerie.

A titre d'exemple, l'échantillonneur de Gibbs permet cette simulation lorsque les probabilités conditionnelles :

$$p(x_s | x_{V_s}) = \frac{\exp - \sum_{c \in C} \varphi_c(x_s)}{\sum_{x_s \in \Omega} \exp - \sum_{c \in C} \varphi_c(x_s)} \quad (\text{I.6})$$

sont connues. Sachant que la taille des cliques ne dépasse généralement pas quatre éléments, les probabilités conditionnelles (I.6) sont calculables.

I.1.5 Exemple: Modèle de Potts

Nous devons l'introduction du modèle de Potts [Pot 52] à Domb, qui, au début des années 50, proposa à son étudiant R.B. Potts d'étudier une généralisation du célèbre "*modèle d'Ising*". L'intérêt du modèle de Potts (*resp. Ising*), en traitement d'images, est son utilisation en segmentation contextuelle en plus de deux classes (*resp. deux classes*).

Soit S un ensemble d'atomes et $x_s = \pm 1$ l'orientation d'un "*spin*", qui est le moment cinétique intrinsèque des particules quantiques, en un site s à l'état d'équilibre comme le montre FIG. 2. On considère qu'une configuration $x = (x_s)_{s \in S}$ est une réalisation d'un champ de Markov, qui est dit "*modèle d'Ising*". La fonction énergie du modèle d'Ising est définie comme suit:

$$U(x) = -\beta \sum_s x_s - \alpha \sum_{s \sim t} x_s x_t \quad (\text{I.7})$$

Le premier terme représente un champ magnétique externe constant avec une intensité β qui favorise *a priori* par sa valeur, la fréquence d'apparition d'un spin par rapport à un

autre. Le second terme, représente l'énergie d'interaction entre chaque paire de spins voisins ($s \sim t$ signifie que s et t sont voisins). α est une constante: pour $\alpha > 0$, les configurations les plus probables sont celles pour lesquelles les spins sont de même signe. En effet, les spins de même direction (ou de même signe) produisent une faible énergie, ce qui implique une probabilité élevée. Ce cas est appelé "*ferromagnétique*". Par ailleurs, pour $\alpha < 0$ l'alternance des spins de signes opposés est favorisée et nous nous trouvons dans le cas "*anti-ferromagnétique*".

Le modèle d'Ising est parmi les plus simples et il est fréquemment utilisé en analyse d'images. En considérant un système de voisinage d'ordre 2, deux situations peuvent se présenter:

- soit les directions horizontales et verticales ne sont pas distinguées. On a alors un modèle isotropique avec l'énergie U donnée par (I.7) ;
- dans le cas contraire, le modèle est appelé modèle anisotropique. La fonction énergie d'Ising (I.7) devient:

$$U(x) = -\beta \sum_s x_s - \alpha_H \sum_{\dots} x_s x_t - \alpha_V \sum_{\substack{\vdots \\ \cdot}} x_s x_t \quad (\text{I.8})$$

De même que pour le modèle d'Ising, la fonction énergie du modèle de Potts, qui, rappelons le, est une extension du modèle d'Ising à plus de deux états (ou classes), est donnée par:

$$U(x) = \alpha \sum_{s \sim t} (1 - \delta(x_s, x_t)) \quad (\text{I.9})$$

pour un champ magnétique nul, où $\delta(x_s, x_t)$ est la fonction de Dirac, telle que: $\delta(x_s, x_t) = 1$ pour $x_s = x_t$ et $\delta(x_s, x_t) = 0$ pour $x_s \neq x_t$.

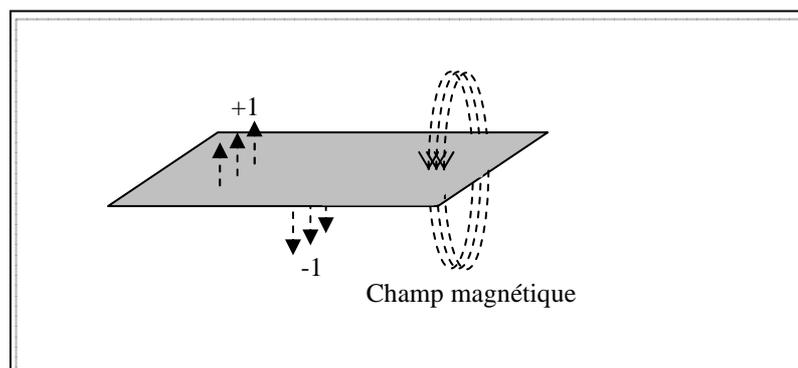


FIG. 2 – orientation d'un spin à l'état d'équilibre

I. 2 Simulation d'un champ de Markov

La simulation des réalisations des champs de Markov est un outil qui peut s'avérer très utile pour résoudre un problème dont on ne peut pas proposer une solution analytique. En effet, dans le cas de la segmentation d'images non supervisée par exemple, où les paramètres du modèle sont à estimer, certaines méthodes d'estimation des paramètres demandent la réalisation du champ X . Les algorithmes d'échantillonnage, tels que l'algorithme de Metropolis [MRR 53], l'échantillonneur de Gibbs [GeG 84], ou l'algorithme de Swendsen – Wang [SwW 87], que nous décrivons ci après, permettent effectivement la simulation des champs des Markov. Bien que la présentation semble concerner la loi *a priori* $p(x)$, les algorithmes sont évidemment valables pour la loi *a posteriori* $p(x|y)$.

I.2.1 Algorithme de Metropolis

L'algorithme de Metropolis est un algorithme itératif qui permet de simuler des réalisations d'un champ markovien X . On considère une suite x^n de réalisations de X , où chaque x^n est obtenu à partir de x^{n-1} en balayant l'ensemble des pixels et en considérant, en chaque pixel s , la variation de l'énergie locale ΔU (on entend par énergie locale, la restriction de l'énergie au site s et ses voisins, donnée par (I.10)). Si ΔU est négative, la classe associée au site s est échangée par celle obtenue par tirage aléatoire. Dans le cas contraire, deux possibilités peuvent se présenter. Les étapes de cet algorithme sont décrites comme suit:

- Initialiser une première carte "image" x^0 d'une façon arbitraire.
- A chaque itération n :
 - Balayer l'ensemble des sites $s \in S$, et en chaque site s :
 - Effectuer un tirage aléatoire d'une variable κ selon la loi uniforme dans Ω ,
 - Calculer la variation d'énergie locale ΔU par:

$$\Delta U = U(x_s = \kappa, x_{V_s}) - U(x_s^n, x_{V_s}) \quad (\text{I.10})$$

- Si $\Delta U \leq 0$: le changement est accepté $x_s^{n+1} = \kappa$
- Sinon, le changement est accepté ($x_s^{n+1} = \kappa$) ou refusé ($x_s^{n+1} = x_s^n$) selon les probabilités $\exp(-\Delta U(x_s^n, x_{V_s}))$ ou $1 - \exp(-\Delta U(x_s^n, x_{V_s}))$

Notons que le voisinage V dans (I.10) contient généralement des sites qui ont déjà été visités et des sites qui ne l'ont pas encore été. Sur les sites t déjà visités x_t dans x_V est x_t^{n+1} , et x_t dans x_V est x_t^n sur les autres sites.

I.2.2 Echantillonneur de Gibbs

Le principe de l'algorithme de l'échantillonneur de Gibbs ressemble à celui de l'algorithme de Metropolis; cependant, la différence réside dans l'étape de mise à jour qui est fondée sur la probabilité conditionnelle au lieu de la variation d'énergie. Cet algorithme se résume à :

- Initialiser une première carte x^0 de façon arbitraire;
- A chaque itération n :
 - Balayer l'ensemble des sites $s \in S$, et en chaque site s :
 - Calculer les probabilités $p(x_s^n | x_{V_s}^n)$;
 - Effectuer un tirage aléatoire d'une variable κ dans Ω , selon ces probabilités conditionnelles et poser $x_s^{n+1} = \kappa$.

On obtient ainsi une suite x^0, x^1, \dots, x^N de réalisations aléatoires du champ X , dont la loi converge vers (I.3) [Win 03].

Remarque

Il existe deux types d'échantillonneurs de Gibbs : (i) à balayage séquentiel : dans ce cas, le balayage s'effectue séquentiellement, par exemple dans l'ordre $1-2-\dots-(n \times m)$, et (ii) à balayage aléatoire : le parcours se fait aléatoirement, à condition de balayer tous les sites de l'image en temps fini avec la probabilité égale à 1.

Nous présentons sur les figures FIG. 3 et FIG. 4 des réalisations de champs de Markov de loi $p(x)$ définie par :

$$p(x) = \gamma \exp \left(\sum_{s \in S} \beta(x_s) + \alpha_H \sum_{s \sim t} (1 - \delta(x_s, x_t)) + \alpha_V \sum_{s \sim t} (1 - \delta(x_s, x_t)) \right) \quad (\text{I.11})$$

avec les paramètres d'interaction α_H correspondant aux cliques horizontales, α_V aux cliques verticales, et $\beta_i = \beta(\omega_i)$ (pour $1 \leq i \leq k$) les paramètres permettant de générer les disproportions de chaque classe dans l'image, respectivement. Ces simulations correspondent à différentes valeurs des paramètres α et β . Elles ont été réalisées par l'échantillonneur de Gibbs utilisant 20 balayages déterministes.

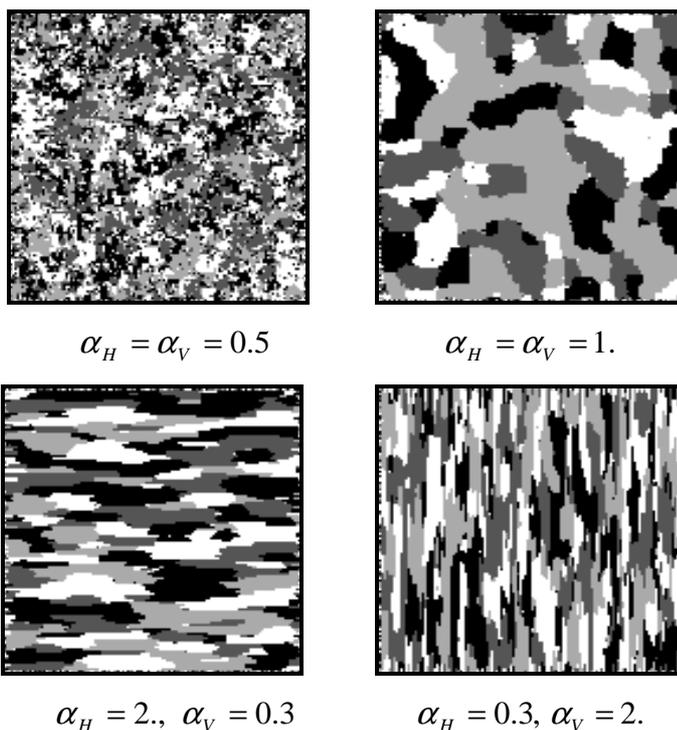


FIG. 3 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs pour différentes valeurs de α ($\beta = 0$. pour tout $1 \leq i \leq 4$)

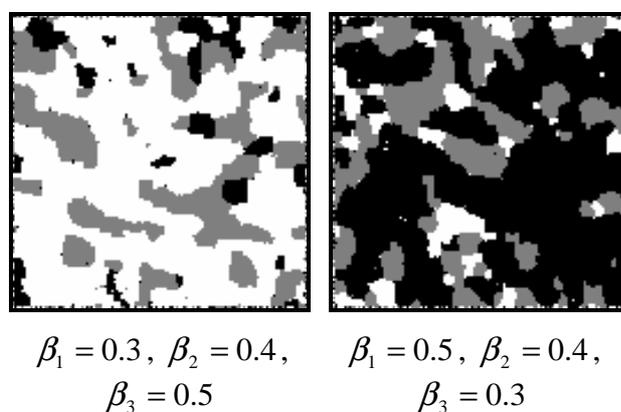


FIG. 4 – Exemple de simulation d'un champ de Markov à 3 classes par échantillonneur de Gibbs pour différentes valeurs de β ($\alpha_H = \alpha_V = 1$.)

I.2.3 Algorithme de Swendsen – Wang

En plus des deux méthodes d'échantillonnage ci-dessus, une troisième technique fondée sur le modèle de Potts, pour la simulation d'un champ de Markov, a été proposée en 1987 par R.H. Swendsen et J.S. Wang [SwW 87] et généralisée par Edwards et Sokal en 1988 [EdS 88]. Aussi, des travaux récents généralisent efficacement l'algorithme de Swendsen et Wang pour la loi a posteriori $p(x|y)$ [BaZ 05].

Cet algorithme consiste en l'introduction d'un ensemble (ou champ) auxiliaire binaire B , appelé "*ensemble de lien*", qui traduit l'existence ou non d'un lien entre deux pixels voisins. Ainsi, la simulation du champ X s'effectue via une transition par ce champ de lien B . On commence par initialiser une première carte x^0 arbitrairement, on simule ensuite $B = b$, selon la loi conditionnelle $p(b|x)$. Ayant le champ B , X est alors obtenu à partir de la loi $p(x|b)$, et ainsi de suite. Les étapes de cet algorithme sont décrites ci-après:

- Initialiser une première carte "*image*" x^0 ;
- A chaque itération n :
 - Balayer l'ensemble des sites $s \in S$ de x^{n-1} , et en chaque pixel s :

Calculer les probabilités $p(b|x_{V_s})$ afin de générer le champ de lien B , tels que :

$$p(b_s^n | x_{V_s}) = \exp\left(-\sum_{s \sim t} \alpha(\delta(x_s, x_t) - 1)\right) \prod_{b(x_s, x_t)=0} \exp(-\alpha) \prod_{b(x_s, x_t)=1} (1 - \exp(-\alpha)) \delta(x_s, x_t) \quad (\text{I.12})$$

Ainsi pour: $x_s \neq x_t$, $b_s^n = 0$ (pas de lien entre les pixels s et t).

$x_s = x_t$, $b_s^n = 0$ (s et t ne sont pas liés) ou $b_s^n = 1$ (s et t sont liés) selon les probabilités $\exp(-\alpha)$ et $1 - \exp(-\alpha)$

- Repérer, à partir de la configuration b^n , les régions connexes déterminées par la présence ou non de lien entre les pixels et leur attribuer une classe choisie uniformément sur l'ensemble des classes. La probabilité $p(x|b)$ est donnée par:

$$p(x|b) = |G|^{-c(b)} \prod_{b(x_s, x_t)=1} \delta(x_s, x_t) \quad (\text{I.13})$$

où $c(b)$ est le nombre de classes dans l'ensemble des liens b

- On obtient ainsi une nouvelle configuration x^n .

Remarque

Comme nous l'avons explicité ci-dessus, la mise à jour des champs B et X s'effectue alternativement, suivant les distributions conditionnelles $p(b|x)$ et $p(x|b)$, respectivement.

Par ailleurs, notons que pour des valeurs élevées du paramètre d'interaction α , cet algorithme devient plus rapide que l'échantillonneur de Gibbs. A l'inverse, lorsque les valeurs de α sont petites, l'algorithme de Swendsen–Wang a tendance à fournir des réalisations formées d'un grand nombre de petites régions, car la probabilité de créer un lien devient faible. Dans ce cas, il devient moins performant que l'échantillonneur de Gibbs.

I.2.4 Conclusion

Les algorithmes décrits dans cette section permettent donc la simulation des réalisations d'un champ de Markov. Nous optons, pour la suite de cette thèse, pour l'échantillonneur de Gibbs.

I.3 Modèle markovien caché pour la segmentation d'images

On considère deux champs aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, avec S l'ensemble des pixels, dans lesquels chaque variable aléatoire X_s est à valeurs dans l'ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque variable aléatoire Y_s est à valeurs dans l'ensemble des nombres réels R . En segmentation d'images, le problème est d'estimer le champ inobservable, ou "caché", $X = x$ à partir du champ observé, ou "bruité", $Y = y$. Dans le modèle "Champ de Markov caché" (CMC) classique, X est de Markov et la forme du bruit (loi de Y conditionnelle à X) est simple, ce qui implique la markovianité de la loi de X *a posteriori* (sa loi conditionnelle à Y). Cette dernière markovianité permet alors de mettre en œuvre les différents traitements bayésiens développés dans la section I.5. Dans les CMC les plus classiques, la loi du couple (X, Y) est donnée par:

$$p(x, y) = \gamma \exp\left(-\sum_{c \in C} \varphi_{c \in C}(x_c)\right) \prod_{s \in S} p(y_s | x_s) \quad (\text{I.14})$$

Cette loi est fonction d'un certain nombre de paramètres, qu'on appellera vecteur paramètre θ , constitué de paramètres d'interaction α liés à la loi *a priori* $p(x)$, et de paramètres η liés

à la loi conditionnelle $p(y|x)$ (ou encore lié au "bruit"). La connaissance de $\theta = (\alpha, \eta)$ est nécessaire aux divers traitements, d'où la nécessité de pouvoir l'estimer. A cet effet, diverses méthodes ont été développées, dont nous mentionnons brièvement dans ce qui suit, les plus connues.

I. 4 Apprentissage

Les méthodes d'estimation des paramètres peuvent être divisées en deux grandes familles, lesquelles peuvent être, à leur tour, classées selon leurs niveaux de complexité :

□ Apprentissage à données complètes

Dans l'apprentissage à données complètes, l'estimation des paramètres du modèle markovien se fait à partir de toutes les données, contenant la vérité terrain $X = x$ et l'observation $Y = y$. Parmi les diverses techniques existantes d'estimation de α à partir de x , citons : la méthode de codage introduite par Besag [Bes 74], la méthode du gradient stochastique proposée par Younes [You 88] et enfin une méthode de complexité inférieure aux deux autres, et qui est la méthode des moindres carrés proposée par Derin et Elliott [DeE 87]. Par ailleurs, dans les modèles simples, comme le modèle (I.14), les paramètres η s'obtiennent généralement aisément à partir de l'estimation des moments.

□ Apprentissage à données incomplètes

De complexité supérieure au cas précédent, l'apprentissage à données incomplètes est très utile, souvent même indispensable dans les applications réelles. Ici, l'estimation est réalisée à partir de données incomplètes $Y = y$. Cette famille regroupe entre autres, l'algorithme "Expectation Maximisation" (EM) développé par [BPS 70], l'algorithme "Stochastic Expectation Maximisation" (SEM) de [BCD 83] et l'algorithme "Itérative Conditional Estimation" (ICE) proposé dans [Pie 92]. Etant donné que ces deux derniers nécessitent des estimateurs de θ à partir de données complètes, les méthodes d'apprentissage à données complètes, citées ci-dessus, peuvent être utilisées.

Sachant que le vecteur paramètre θ est constitué de paramètres α liés à la loi *a priori* $p(x)$, et de paramètres η liés à la loi conditionnelle, il est à noter que α lui-même est généralement constitué de paramètres d'interactions α_i , et de paramètres β_k associés aux singletons dont le rôle est de gérer les proportions de chaque classe de l'image. Par ailleurs, dans le cas de bruit blanc gaussien les paramètres η_k sont les moyennes μ_k et variances σ_k^2 .

Dans cette section nous nous sommes limités à mentionner brièvement les différentes méthodes d'apprentissage. Cette première approche nous permet une description plus rapide des méthodes plus complexes dont certaines originales sont approfondies dans le chapitre III.

I. 5 Estimation bayésienne

Soit $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ deux champs aléatoires comme décrit ci-dessus. Le problème de la segmentation bayésienne consiste en l'estimation de la réalisation invisible X à partir des données observées $Y = y$. Ainsi, le problème est de déterminer une estimation $\hat{x} \in \Omega$ de x à partir de y , obtenue en optimisant un certain critère.

L'estimation bayésienne nécessite la spécification d'une "fonction de coût", L définie dans $\Omega^{Card(S)} \times \Omega^{Card(S)} \rightarrow R^+$. Celle-ci possède les propriétés suivantes:

$$\begin{aligned} \forall x, \hat{x} \in \Omega^{Card(S)} \times \Omega^{Card(S)} : \\ L(x, \hat{x}) \geq 0 ; \\ L(x, \hat{x}) = 0 \Leftrightarrow x = \hat{x} \end{aligned} \tag{I.15}$$

Le risque bayésien associé à la stratégie $\hat{s} : R^{Card(S)} \rightarrow \Omega^{Card(S)}$ est donné par le coût moyen $R = E[L(\hat{s}(Y), X)]$. La stratégie bayésienne \hat{s}_B est une stratégie dont le risque bayésien est minimum:

$$E[L(\hat{s}_B(Y), X)] = \min_{\hat{s}} E[L(\hat{s}(Y), X)] \tag{I.16}$$

L'estimateur bayésien \hat{s}_B est alors obtenu en minimisant l'espérance de cette fonction de coût conditionnellement aux observations:

$$\hat{s}_B(y) = \arg \min_{\hat{x} \in \Omega} \mathbb{E}[L(\hat{x}(y), X) | Y = y] \quad (1.17)$$

A chaque fonction de coût correspond ainsi un estimateur bayésien. Dans ce qui suit nous nous intéressons principalement à deux fonctions de coût, très utilisées, qui définissent les estimateurs les plus répandus dans la littérature, à savoir l'estimateur du "*maximum a posteriori*" (MAP) et l'estimateur du "*mode des marginales a posteriori*" (MMP), très connu sous l'abréviation MPM (en anglais). Aussi, nous nous intéressons à l'estimateur "*iterated conditional mode*" (ICM) qui est une approximation du MAP. Signalons que d'autres estimateurs existent; citons à titre d'exemple, l'estimateur du champ moyen (MF) connu sous le nom TPM "*Threshold Posterior Mean*" utilisé en restauration bayésienne, l'estimateur de la "*moyenne a posteriori*" (MP) appelé également (MMSE) [GRS 96] "*posterior minimum mean squares estimate*" (plus facile d'un point de vue calcul que l'estimateur MAP) et l'estimateur du "*maximum a posteriori séquentiel*" (SMAP) proposé par Bouman et Shapiro [BoS 94], et qui est une nouvelle version du MAP adapté aux problèmes d'estimations dans le cadre des champs multiéchelles "*multiscale random field*" (MSRF).

Le choix d'un estimateur est souvent laissé à l'appréciation de l'utilisateur. Cependant, dans ce travail, notre choix s'est porté sur l'estimateur MPM. Nous détaillons, ci après, les estimateurs les plus répandus, énumérés plus haut.

I.5.1 Estimateur Maximum A Posteriori (MAP)

L'estimateur MAP est associé à la fonction de coût suivante:

$$L(\hat{x}, x) = 1 - \delta(\hat{x}, x) \quad (I.18)$$

où la fonction δ est définie par $\delta(x_s, x_t) = 0$ pour $x_s \neq x_t$ et $\delta(x_s, x_t) = 1$ pour $x_s = x_t$. Cette fonction de coût est sévère, car elle pénalise toute différence entre deux configurations de manière identique. Autrement dit, elle pénalise de façon identique une erreur sur un site et une erreur sur plusieurs sites. Nous pouvons alors écrire:

$$\begin{aligned} E[L(\hat{x}, x) | Y = y] &= \sum_{x \in \Omega} L(\hat{x}, x) p(x|y) \\ &= 1 - p(\hat{x}|y) \end{aligned} \tag{I.19}$$

Par conséquent, l'estimation bayésienne est:

$$\hat{x} = \arg \min_{\hat{x} \in \Omega} (1 - p(\hat{x}|y)) = \arg \max_{\hat{x} \in \Omega} (p(\hat{x}|y)) \tag{I.20}$$

L'estimation au sens du MAP, revient donc à maximiser la probabilité *a posteriori*. Par ailleurs, cet estimateur présente certains inconvénients. Le premier, bien argumenté dans l'article de Bouman et Shapiro [BoS 94], est dû aux méthodes d'approximations, du fait que le calcul de l'estimateur par MAP ne peut pas se faire directement. En effet, le calcul du MAP nécessite la minimisation d'une fonction énergie, souvent non convexe, ce qui a pour effet la présence de minima locaux. Une minimisation exacte est impossible, d'où la nécessité de faire appel à des méthodes d'approximations, tel que le recuit simulé, décrit dans la sous section suivante. Le deuxième, avancé par Marroquin [MVR 01], est lié à la fonction de coût. Celle-ci n'est pas très appropriée du fait qu'elle attribue le coût "0" pour la solution parfaite uniquement et le coût "1" pour toutes les autres solutions, sans tenir compte du nombre des pixels mal classés. Ceci rend l'estimateur MAP très sélectif, fournissant ainsi de mauvais résultats, en particulier lorsque le bruit est fort [MMP 87]. Cependant, Winkler [Win 03] estime que les résultats insatisfaisants obtenus par segmentation MAP ne sont pas dus à l'estimateur lui-même, mais plutôt au choix de la loi *a priori*. Ainsi, selon Winkler, les probabilités *a priori* sont, en pratique, fréquemment choisies de telle sorte que l'estimateur MAP soit facilement calculable, sans correspondre nécessairement aux caractéristiques de l'image. Quant aux temps de calcul, Winkler pense que ce n'est pas un réel problème, étant donné l'avancée rapide des progrès technologiques et méthodologique tel que les méthodes de type "*min-cut/max-flow*".

I.5.2 Algorithme du recuit simulé

L'idée d'introduire un paramètre de température et de simuler un recuit a été initialement proposé par Kirkpatrick [KGV 82], elle a ensuite été reprise par Geman et Geman [GeG 84] dans ce qui est connu de nos jours, sous le nom de "*l'algorithme du recuit simulé*".

Cet algorithme itératif est associé à une décroissance de température. Le but est de rechercher une configuration qui maximise la distribution :

$$p_T(x) = \frac{1}{Z(T)} \exp\left(-\frac{U(x)}{T}\right) \quad (\text{I.21})$$

i.e. rechercher une configuration à énergie minimale. L'idée consiste à simuler une suite de réalisations suivant la distribution (I.21), en faisant décroître la température au cours des itérations. Au départ, une température T élevée, permet d'effectuer des sauts d'énergie afin de ne pas se stabiliser dans des minima locaux. Autrement dit, l'algorithme autorise dans ce cas des passages dans des configurations qui font augmenter la fonction énergie U , ce qui permet d'éviter de rester bloqué dans des minima locaux. Plus tard, lorsque la température décroît, ces sauts d'énergie sont progressivement supprimés. Ainsi, la suite de réalisations évolue progressivement vers le minimum global. Le déroulement de l'algorithme est le suivant:

- Initialiser une première carte "*image*" x^0 d'une façon arbitraire ;
- Initialiser la température $T^{(0)}$ avec une "*valeur assez élevée*" ;
- A chaque étape n :
 - Balayer l'ensemble des sites $s \in S$, et en chaque site s :
 - Calculer les probabilités conditionnelles $p_{T^{(n)}}(x_s | x_{V_s})$.
 - Effectuer un tirage aléatoire, dans l'ensemble Ω , selon ces probabilités conditionnelles, afin de sélectionner l'état (ou la classe) qui sera attribué à ce pixel (qui peut éventuellement être le même).
 - Décroître la température $T^{(n)} = k * T^{(n-1)}$ ($k \approx 0.99$).
- Refaire l'étape n jusqu'à ce que le taux de changement devienne faible.

Remarque

La convergence de l'algorithme du recuit simulé dépend du schéma d'évolution de la température. Cela peut nécessiter une décroissance relativement lente, auquel cas la température peut rester constante sur plusieurs itérations.

Nous montrons dans FIG. 5, extraite de [Mai 03], la version Metropolis de l'évolution de l'énergie en fonction de la décroissance de la température.

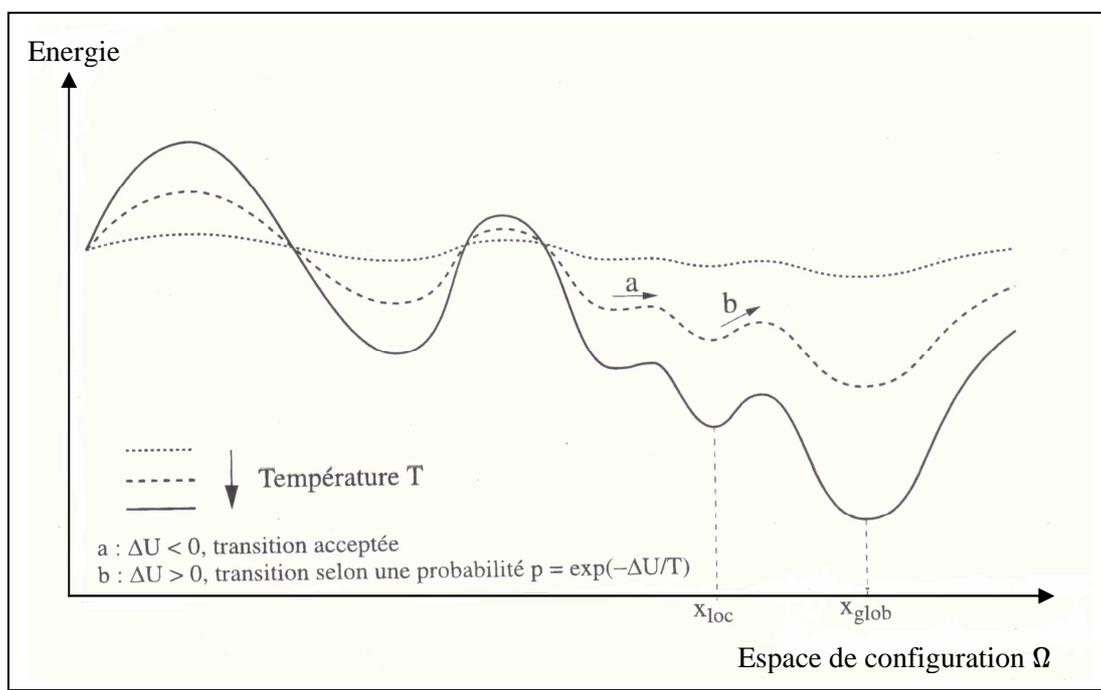


FIG. 5 – Evolution de la fonction d'énergie en fonction de la température

I.5.3 Estimateur du Mode Conditionnel Itéré (ICM)

L'estimateur du mode conditionnel itéré (ICM) a été proposé comme une méthode itérative d'approximation de l'estimateur MAP [Bes 86]. C'est un algorithme déterministe nécessitant une bonne valeur d'initialisation. Son principe est fondé sur la maximisation des probabilités conditionnelles locales en chaque site séquentiellement. En d'autres termes, en

chaque site s on retient la classe qui maximise cette probabilité. L'algorithme ICM se résume comme suit:

- Initialiser une première carte "image" x^0 d'une façon arbitraire.
- A chaque itération n :
 - Balayer l'ensemble des sites $s \in S$ de x^{n-1} et en chaque site on calcule la probabilité conditionnelle, on pose :

$$x_s^n = \arg \max_{x_s \in \Omega} p(x_s | x_{V_s}^{n-1}) \quad (\text{I.22})$$

- Refaire l'étape n jusqu'à réalisation d'un critère d'arrêt (par exemple lorsque le nombre de changements d'une étape à une autre devient suffisamment faible).

Remarque

L'algorithme ICM converge vers le premier minimum local rencontré à partir des conditions initiales.

I.5.4 Estimateur du Mode des Marginales à posteriori (MMP)

Pour éviter la sévérité de la fonction de coût de l'estimateur MAP, une autre fonction de coût moins restrictive, est associée à l'estimateur MPM. Celle ci est donnée comme suit:

$$L(\hat{x}_s, x_s) = \sum_{s \in S} 1 - \delta(\hat{x}_s, x_s) \quad (\text{I.23})$$

Cette fonction consiste à pénaliser l'erreur commise en fonction du nombre de sites mal estimés (ou mal classés).

On montre que l'estimation Bayésienne \hat{x} est ici obtenue suivant:

$$\hat{x}_s = \arg \max_{\hat{x}_s} p(x_s | y) \quad (\text{I.24})$$

Cette estimation ressemble à l'estimation du MAP, mais opérée de façon locale. Autrement dit, on passe de la probabilité conditionnelle globale d'une configuration, à la probabilité conditionnelle en un site. Ainsi, la configuration optimale est atteinte lorsque toutes les lois marginales en chaque site sont maximisées. Cependant, le calcul direct et exact des probabilités *a posteriori* $p(x_s|y)$ est impossible, compte tenu du gigantisme de l'espace des configurations (§ I.1.3), mais le fait de pouvoir simuler des réalisations de X par des approximations de type Monte-Carlo décrites dans (§ I.2), permet leur estimation. En effet, on peut poser :

$$\hat{p}(x_s = \omega|y) = \frac{1_{[x_s^1 = \omega]} + \dots + 1_{[x_s^N = \omega]}}{N} \quad (\text{I.25})$$

Finalement, pour la segmentation MPM, on procède de la manière suivante :

- Simuler N réalisations x^1, x^2, \dots, x^N de X selon la probabilité $Y = y$ en utilisant l'échantillonneur de Gibbs (décrit dans § I.2.2) ;
- Estimer à partir des réalisations x^1, x^2, \dots, x^N la loi de chaque X_s par les fréquences (équation (I.25)) ;
- Retenir la classe maximisant la loi ainsi obtenue. En d'autres termes choisir la classe dont le nombre d'apparitions dans les simulations est le plus grand.

I.6 Conclusion

Nous avons exposé dans ce premier chapitre la problématique générale de la segmentation statistique d'images, abordée par les champs de Markov cachés. Nous avons présenté les modèles classiques, ainsi que les principaux algorithmes de segmentation bayésienne. Les principales méthodes classiques d'estimation des paramètres, permettant la conception des méthodes de segmentation non supervisée, ont été également mentionnées.

Ces modélisations et traitements seront repris et généralisés à des modélisations et traitements originaux dans la suite. Ce premier chapitre permettra ainsi une présentation plus concise, et peut-être plus lisible, des apports originaux de la thèse.

Chapitre II

Modélisation markovienne

Les champs de Markov sont des outils très utilisés en traitement statistique d'images. Ils permettent d'obtenir de remarquables résultats dans un certain nombre de traitements. L'objet de ce chapitre est de présenter différentes modélisations de complexité croissante, dont certaines originales, utilisées au cours de cette thèse. Nous commençons par étudier les champs de Markov cachés (CMC) que l'on appellera également "*modèle classique*", où trois variantes seront explicitées. Nous présentons ensuite les champs de Markov Couples (CMCouples), qui sont une généralisation du modèle classique développé dans le but de surmonter certaines lacunes de ce dernier. En particulier, nous verrons que les CMCouples offrent les mêmes possibilités de traitements que les CMC et permettent de mieux modéliser le bruit et les textures. Ensuite, nous présentons les champs de Markov triplets (CMT) qui sont une généralisation des champs de Markov couples. Plus particulièrement, deux applications originales suivantes des CMT seront explicitées :

1. La modélisation de l'incertitude, grâce à l'utilisation de la théorie de l'évidence dans le contexte des champs de Markov cachés, aboutit à un champ de Markov évidentiel (CME), et peut permettre une amélioration des résultats obtenus par des traitements bayésiens classiques. Nous verrons comment un CME caché est un CMT, ce qui permet son utilisation dans la segmentation d'images ;
2. Le traitement des images non stationnaire pose un problème au niveau de l'estimation des paramètres. Nous verrons comment un modèle classique non stationnaire peut être représenté par un CMT stationnaire, ce qui permet encore la segmentation de l'image non stationnaire de manière non supervisée.

II. 1 Champs de Markov cachés

Soit S un ensemble de pixels de cardinal N ($N = \text{Card}(S)$), et $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$ deux champs aléatoires, l'un caché et l'autre observé. Chaque variable aléatoire X_s est à valeurs dans l'ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque variable aléatoire Y_s est à valeurs dans R . En segmentation d'images, l'objectif est de retrouver le champ inobservé X , à partir des données observées Y qui représentent la version bruitée du champ X . A cet effet, nous devons définir la loi de (X, Y) , ce qui est classiquement obtenu à partir de la loi de X et celle de Y conditionnelle à X . Dans le cas de la modélisation par champs de Markov cachés (CMC), X est supposé markovien. Sa loi est alors donnée par :

$$p(x) = \gamma \exp[-U(x)], \quad (\text{II.1})$$

avec γ une constante de normalisation, appelée également "*fonction de partition*", et U "*la fonction énergie*". Celle-ci se décompose en une somme de fonctions φ_c , appelées "*fonctions de potentiel*", associées à l'ensemble des cliques C (rappelons qu'une clique est soit un singleton, soit un ensemble de pixels mutuellement voisins). Ainsi, $U(x)$ s'écrit :

$$U(x) = \sum_{c \in C} \varphi_c(x_c) \quad (\text{II.2})$$

La fonction de partition γ est généralement impossible à calculer du fait du nombre trop important des réalisations possibles du champ X ; cependant, il reste possible de simuler une réalisation de ce champ à partir de la loi conditionnelle de X_s sachant X_{V_s} tel que X_{V_s} est la restriction du champ X au voisinage V_s . L'échantillonneur de Gibbs que nous allons utiliser au cours de cette thèse et qui a été détaillé au chapitre I (§ I.2.2), permet cette simulation lorsque la loi conditionnelle est connue. Cette loi est donnée par :

$$p(x_s | x_{V_s}) = \frac{\exp\left[-\sum_{c \in C, s \in c} \varphi_c(x_s, x_{c-\{s\}})\right]}{\sum_{\omega \in \Omega} \exp\left[-\sum_{c \in C, s \in c} \varphi_c(x'_s = \omega, x_{c-\{s\}})\right]} \quad (\text{II.3})$$

Ainsi, dans le cas des champs de Markov cachés, le champ inobservé X est supposé markovien, ce qui justifie l'appellation CMC. La loi de X étant définie, il reste à définir la loi conditionnelle $p(y|x)$. Trois CMC, de généralité croissante, peuvent être présentés selon les hypothèses posées sur cette loi conditionnelle et que nous décrivons ci-après.

II.1.1 Champs de Markov cachés à bruit indépendant

Les champs de Markov cachés à bruit indépendant (CMC-BI) sont la modélisation la plus simple des CMC. La probabilité conditionnelle $p(y|x)$ vérifie les deux hypothèses suivantes:

- H1: $p(y_s|x) = p(y_s|x_s)$ pour chaque $s \in S$;
- H2: $p(y|x) = \prod_{s \in S} p(y_s|x)$

La deuxième hypothèse signifie que les Y_s sont indépendants conditionnellement à X . Du fait de H1 et H2, la loi de $p(y|x)$ s'écrit sous la forme :

$$\begin{aligned} p(y|x) &= \prod_{s \in S} p(y_s|x_s) \\ &= \exp \left[\sum_{s \in S} \log(p(y_s|x_s)) \right] \end{aligned} \tag{II.4}$$

La loi du couple (X, Y) résulte des formules (II.1), (II.2) et (II.4). Elle s'écrit comme suit :

$$\begin{aligned} p(x, y) &= p(x)p(y|x) \\ &= \gamma \exp \left[- \sum_c \varphi_{c \in C}(x_c) + \sum_{s \in S} \log(p(y_s|x_s)) \right] \end{aligned} \tag{II.5}$$

La probabilité *a posteriori* $p(x|y)$ du champ X est alors donnée par :

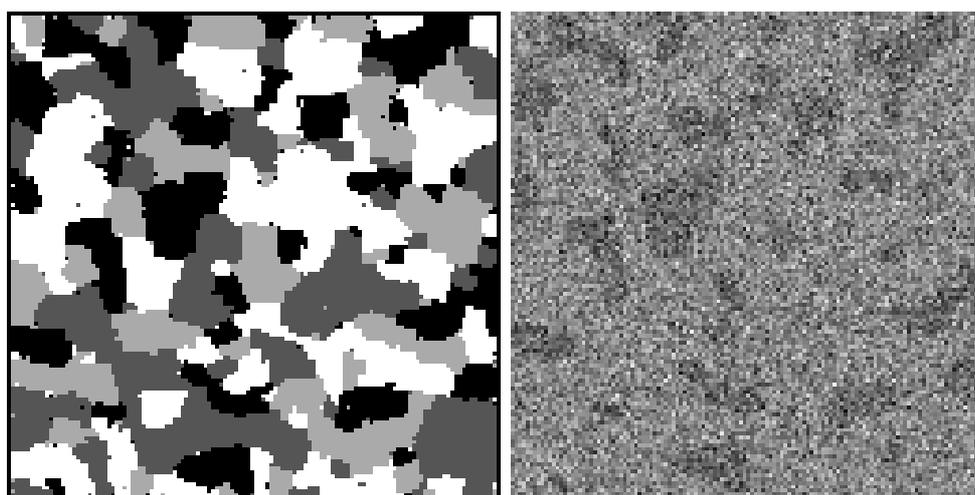
$$p(x|y) = \gamma(y) \exp[-U(x, y)] \quad (\text{II.6})$$

$$= \gamma(y) \exp \left[- \left(\sum_{c \in C} \varphi_c(x_c) - \sum_{s \in S} \log(p(y_s | x_s)) \right) \right]$$

$$\text{avec : } \gamma(y) = \sum_{x \in \Omega^N} \exp \left[- \sum_{c \in C} \varphi_c(x_c) + \sum_{s \in S} \log(p(y_s | x_s)) \right] \quad (\text{II.7})$$

Cette formulation montre que la distribution *a posteriori* du champ X reste une distribution de Gibbs, ce qui signifie que la loi du champ caché conditionnellement à l'observation est également une loi de Markov. Cela rend possible la simulation des réalisations de X selon sa probabilité *a posteriori*, et permet ainsi d'utiliser différentes méthodes de segmentation bayésienne, tels que le "*Maximum a posteriori*" MAP, "*Iterated Conditional Mode*" ICM ou le "*Maximum des Marginales a Posteriori*" MPM.

Un exemple de simulation d'un champ de Markov caché à quatre classes obtenu par échantillonneur de Gibbs, et sa version bruitée avec un bruit blanc gaussien de moyennes $m_0 = 0, m_1 = 1, m_2 = 2, m_3 = 3$ et de variances $\sigma_0^2 = \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, est illustré dans FIG.1.



Champ X

Champ bruité Y

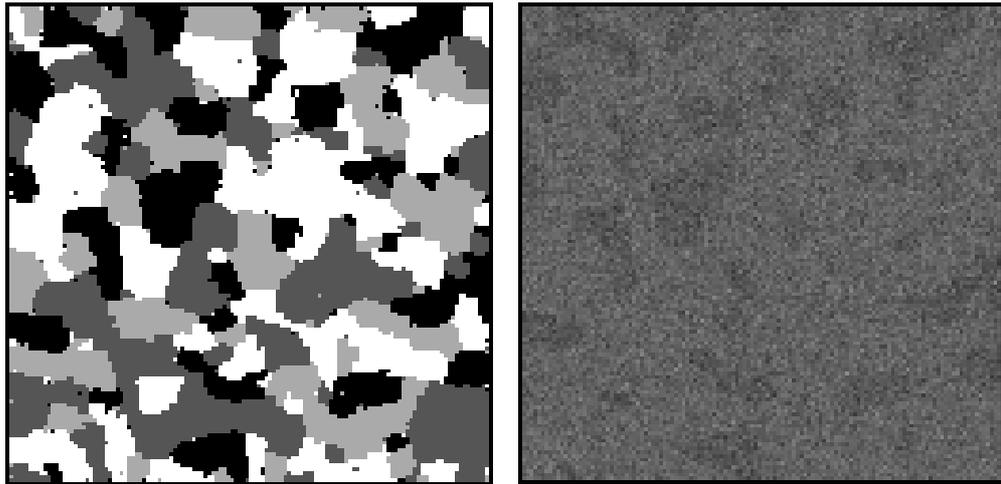
FIG.1 – Exemple d'une réalisation d'un CMC-BI à quatre classes.

II.1.2 Champs de Markov cachés à bruit corrélé

Dans les champs de Markov cachés à bruit corrélé (CMC-BC), l'hypothèse (H2) n'est plus maintenue. La seule hypothèse vérifiée par la loi conditionnelle $p(y|x)$ est $p(y_s|x) = p(y_s|x_s)$. Une manière simple de construire de tels modèles est de considérer deux champs Markoviens X et W définis respectivement sur Ω^N et R^N . On associe à chaque $x_s \in \Omega$ une fonction f_{x_s} bijective et dérivable définie de R dans R . Ensuite, on considère $f(x, w) = (f_{x_s}(w_s))_{s \in S}$ et on définit le champ Y par $Y = f(X, W)$ (voir aussi [Guy 95]). En notant $\sum_{c \in C} \phi_c(x_c)$ l'énergie du champ markovien X et $\sum_{c \in C} \phi_c(w_c)$ l'énergie du champ de Markov W , la distribution $p(x, y)$ est alors donnée par:

$$p(x, y) = \gamma \exp - \left[\sum_{c \in C} \phi_c(x_c) + \sum_{c \in C} \phi_c(f_{x_c}^{-1}(y_c)) - \sum_{s \in S} \log \left| \frac{\partial f_{x_s}^{-1}(y_s)}{\partial y_s} \right| \right] \quad (\text{II.8})$$

avec $f_{x_c}^{-1}(y_c) = (f_{x_s}^{-1}(y_s))_{s \in c}$. Ainsi, le couple $Z = (X, Y)$ est un champ de Markov. Cependant, le champ Z n'est pas un CMC-BI, étant donné que l'hypothèse (H2) d'indépendance conditionnelle des (Y_s) n'est pas vérifiée.



Champ X

Champ Y

FIG. 2 – Exemple d'une réalisation d'un CMC-BC à quatre classes.

FIG. 2 montre un exemple de simulation par échantillonneur de Gibbs d'un champ de Markov à quatre classes à bruit corrélé, et permet de constater une différence visuelle entre le cas présent et celui du bruit blanc gaussien.

II.1.3 Champs de Markov cachés généraux

Dans cette sous-section nous présentons un exemple original de CMC ne vérifiant aucune des deux hypothèses (H1) et (H2). Comme dans le cas des CMC-BC, on considère deux champs aléatoires X et W , définis respectivement sur Ω^N et R^N , et markoviens par rapport au même système de voisinage V . Considérons un ensemble de sous-cliques $C' = \{C_1, \dots, C_q\}$ ($C' \subset C$) de cardinal $a \geq 2$, tels que $C_1 \cup \dots \cup C_q = S$ et $C_i \cap C_j = \emptyset$ pour $i \neq j$ (l'ensemble C' est une partition des S). A titre d'exemple, pour un ensemble S composé de 36 pixels, différentes partitions C_1, \dots, C_q sont possibles. En considérant les 4 plus proches voisins (4-ppv), l'ensemble C_1, \dots, C_q peut s'écrire (1,2), (3,4), (5,6), (7,13), (8,14), (9,15), (10,16), (11,17), (12,18), (19,20), ..., et pour un système à 8-connexités on pourrait, par exemple, opter pour la partition (1,2), (3,4), (5,6), (7,14), (8,13), (9,16), (10,15), (11,18), (12,17), (19,25), (20,26), (21,27), FIG. 3 illustre l'exemple présenté dans le cas des 4-ppv.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

FIG. 3 – Exemple de choix de l'ensemble C_1, \dots, C_{18} pour une image de 36 pixels

Contrairement au cas CMC-BC où nous avons associé une fonction à chaque x_s , dans ce modèle une fonction bijective et dérivable, est associée à chaque clique C_i et $x_{C_i} \in \Omega^a$.

Ensuite, on considère $f(x, w) = (f_{x_{c_i}}(w_{c_i}))_{1 \leq i \leq q}$ et on définit le processus Y par $Y = f(X, W)$.

En notant $\frac{\partial f_{x_{c_i}}^{-1}(y_{c_i})}{\partial y_{c_i}}$ le Jacobien de $f_{x_{c_i}}^{-1}$ (étant une matrice de taille $a \times a$), la distribution $p(x, y)$ s'écrit alors :

$$p(x, y) = \gamma \exp \left[- \left[\sum_{c \in C} \varphi_c(x_c) + \sum_{c \in C} \phi_c(f_{x_c}^{-1}(y_c)) - \sum_{1 \leq i \leq q} \log \left| \det \left(\frac{\partial f_{x_{c_i}}^{-1}(y_{c_i})}{\partial y_{c_i}} \right) \right| \right] \right] \quad (\text{II.9})$$

Comme nous le voyons, l'équation (II.9) est une généralisation de (II.8). Cette dernière étant obtenue pour une fonction $f_{x_{c_i}}$ particulière, vérifiant l'égalité $f_{x_{c_i}}(y_{c_i}) = f_{(x_1, \dots, x_a)}(y_1, \dots, y_a) = (f_{x_1}(y_1), \dots, f_{x_a}(y_a))$. Par ailleurs, la markovianité de la distribution $p(x, y)$ explicitée dans l'équation (II.9) demeure, étant donné que la somme englobe l'ensemble des cliques.

Nous obtenons ainsi un champ de Markov caché dans lequel $p(y_s | x) = p(y_s | x_c)$, où c est l'élément de la partition C' , tel que $s \in c$. Par exemple, dans un CMC par rapport aux 4-ppv, on a $p(y_s | x) = p(y_s | x_s, x_t)$, où t est le voisin de s , tel que $(s, t) \in C'$. Cela signifie que la loi marginale *a posteriori* dépend également de la classe d'un pixel voisin, ce qui peut être utile pour modéliser les bords.

II. 2 Champs de Markov Couples

Proposés dans [PiT 00], les champs de Markov couples (CMCouples) sont une généralisation du modèle précédent. Dans les CMCouples on considère directement la markovianité du couple $Z = (X, Y)$, ce qui implique aussi bien la markovianité de $p(y|x)$ que la markovianité de $p(x|y)$. La markovianité de $p(y|x)$ permet de considérer des bruits markoviens qui sont plus complets que ceux habituellement traités et la markovianité de $p(x|y)$ permet de considérer les mêmes traitements que dans les modèles classiques CMC.

Afin d'illustrer la différence avec le modèle précédant, considérons le cas gaussien. Soit un ensemble de pixels S , avec X et Y deux champs aléatoires. Dans le cas des champs de Markov cachés, le champ X est markovien relativement à un système de voisinage V (dans toutes nos expérimentations nous nous sommes limités aux quatre plus proches voisins), sa loi s'écrit comme suit :

$$p(x) = \gamma \exp \left[- \sum_{c \in C} \phi_c(x_c) \right] \quad (\text{II.10})$$

avec C l'ensemble des cliques c de S . Par ailleurs, en adoptant le modèle classique étudié dans la section précédente (§ II.1), les variables aléatoires Y_s sont supposées indépendantes conditionnellement à X . Cette approche peut engendrer certains problèmes, car en pratique, les Y_s ne sont pas, en général, indépendants conditionnellement à X , cas d'images texturées, par exemple [MeW 02]. Ainsi, pour modéliser des textures par des champs markoviens gaussiens, on est amené à remplacer l'égalité de l'équation (II.4) par la formule suivante :

$$p(y|x) = \gamma(x) \exp \left[- \frac{(\mathbf{y} - \mathbf{m}(x))^T \mathbf{Q}(x) (\mathbf{y} - \mathbf{m}(x))}{2} \right] \quad (\text{II.11})$$

$$\text{où : } \gamma(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \left| \mathbf{Q}^{-\frac{1}{2}}(x) \right|} \quad (\text{II.12})$$

avec $\Gamma(x) = \mathbf{Q}^{-1}(x)$ la matrice de variance-covariance. Pour une image de dimension $N = n \times m = \text{Card}(S)$, celle-ci sera de dimension $N \times N$. En prenant un système des 4-ppv, et en créant un lien fictif entre le dernier pixel d'une ligne et le premier de la ligne suivante, cette matrice s'écrit :

$$\mathbf{Q} = \begin{pmatrix} q_{0,0} & q_{0,1} & 0 & \cdots & 0 & q_{0,p} & 0 & \cdots & 0 & \cdots & \cdots & 0 \\ q_{1,0} & q_{1,1} & q_{1,2} & 0 & \cdots & 0 & q_{1,p+1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & q_{2,1} & q_{2,2} & q_{2,3} & 0 & \cdots & 0 & q_{2,p+2} & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & q_{3,2} & q_{3,3} & q_{3,4} & 0 & \cdots & 0 & q_{3,p+3} & 0 & \cdots & 0 \\ 0 & \ddots & \vdots \\ q_{p,0} & 0 & \cdots & \cdots & 0 & q_{p,p} & q_{p,p+1} & 0 & \cdots & \cdots & \cdots & \vdots \\ 0 & q_{p+1,1} & 0 & \cdots & 0 & q_{p+1,p} & q_{p+1,p+1} & q_{p+1,p+2} & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & q_{N-2,N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & q_{N-1,N-2} & q_{N-1,N-1} \end{pmatrix}$$

La loi $p(y|x)$ est de Markov ; cependant la loi du couple $p(x,y)$ étant le produit de $p(x)$ et de $p(y|x)$ nous avons :

$$p(x,y) = \gamma(x) \exp \left[- \sum_{c \in C} \varphi_c(x_c) + \frac{1}{2} (\mathbf{y} - \mathbf{m}(x))^T \mathbf{Q}(x) (\mathbf{y} - \mathbf{m}(x)) \right] \quad (\text{II.13})$$

Ainsi, X est de Markov, Y est markovien conditionnellement à X , mais ni le couple (X,Y) , ni X conditionnellement à Y ne sont nécessairement markoviens, puisque $\gamma(x)$ ne peut pas être écrit, dans le cas général, comme une distribution markovienne en x . Ceci pose un réel problème aussi bien dans les procédures d'estimation des paramètres qu'en segmentation d'images, du fait que la markovianité de la distribution $p(x|y)$ est indispensable pour de tels traitements. Pour y remédier, il est proposé [PiT 00] de considérer directement la markovianité du couple (X,Y) . Ainsi, la loi du couple devient :

$$p(x,y) = \gamma \exp[-U(x,y)] \quad (\text{II.14})$$

$$\text{avec : } U(x,y) = \sum_{c \in C} \varphi_c(x_c) + \frac{1}{2} (\mathbf{y} - \mathbf{m}(x))^T \mathbf{Q}(x) (\mathbf{y} - \mathbf{m}(x)) \quad (\text{II.15})$$

Le fait de poser directement la markovianité du couple (X, Y) implique la markovianité de Y conditionnellement à X et celle de X conditionnellement à Y , lesquelles rendent possibles la modélisation de textures et les traitements bayesiens, respectivement. Par ailleurs, le champ X n'est plus nécessairement de Markov. En effet, en multipliant et en divisant (II.14) par $\gamma(x)$ et en intégrant par rapport à y , la distribution $p(x) = \frac{\lambda}{\gamma(x)} \exp\left(-\sum_{c \in C} \varphi_c(x_c)\right)$ ne conserve plus nécessairement son écriture markovienne [Pie 03]. L'approche par CMCouple rompt ainsi avec les modélisations classiques dans lesquelles on considère systématiquement que le champ caché est de Markov.

Pour avoir une expression plus pratique de la distribution $p(x, y)$, développons le produit $(\mathbf{y} - \mathbf{m}(x))^T \mathbf{Q}(x)(\mathbf{y} - \mathbf{m}(x))$, on obtient :

$$(\mathbf{y} - \mathbf{m}(x))^T \mathbf{Q}(x)(\mathbf{y} - \mathbf{m}(x)) = \sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C} [q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t})] \quad (\text{II.16})$$

Ainsi, la fonction énergie U aura la forme suivante :

$$U(x, y) = 0.5 \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{c \in C} \varphi_c(x_c) + \sum_{s \sim t} [q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t})] \right] \quad (\text{II.17})$$

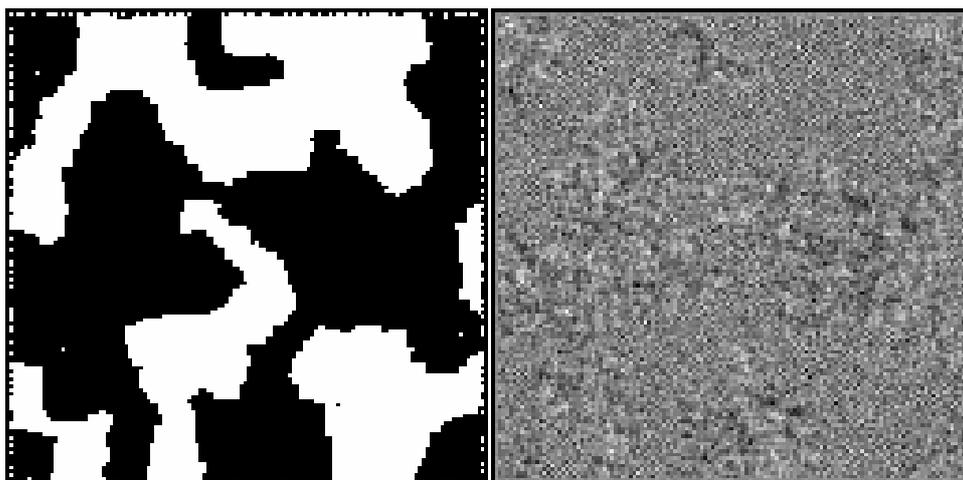
La probabilité *a posteriori* devient donc :

$$p(x|y) \propto p(x, y) \propto \exp[-U(x, y)] \quad (\text{II.18})$$

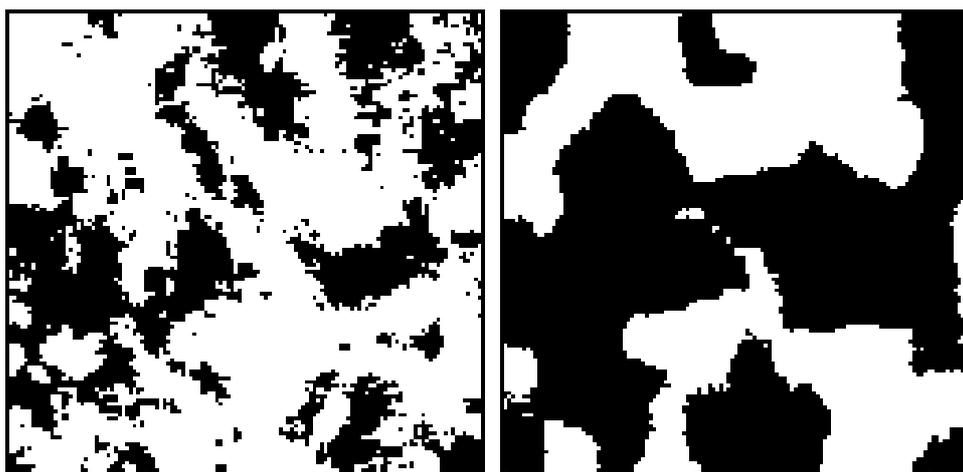
Par ailleurs, pour que la distribution du CMCouple soit définie, les paramètres a_{x_s} et $q_{x_s x_t}$ doivent respecter certaines conditions. En effet, l'existence de la distribution du couple est assurée par l'existence de la distribution $p(y|x)$, ce qui nécessite que la matrice de

variance-covariance $\Gamma(x)$ soit définie positive. En considérant temporairement $x = (x_s)_{s \in S}$, on pose pour tous les x_s égaux $q_{i,i} = a_{x_s}$, et $q_{i,j} = q_{x_s, x_t}$ pour tous $x_s \sim x_t$ (voisins). La matrice ainsi obtenue est une matrice de Toeplitz symétrique dont les valeurs propres sont $\lambda_i = a_{x_s} + 2q_{x_s, x_t} \cos(2\pi Ni)$ [Gra 83]. Ces dernières sont strictement positives si et seulement si $a_{x_s} > 0$ et $-\frac{1}{2} \leq q_{x_s, x_t} \leq +\frac{1}{2}$, ce qui représentent les conditions imposées sur les paramètres de corrélation.

Le modèle de Markov couple permet donc la prise en compte de la corrélation du bruit, fournissant ainsi une information supplémentaire. A titre d'exemple, considérons une image à deux classes, dans laquelle les distributions gaussiennes de chaque Y_s conditionnellement à $X_s = \omega_1, \omega_2$ sont identiques. Cependant, l'unique information dont on dispose permettant la distinction des deux classes est contenue dans les corrélations spatiales du bruit. Le CMCouple permet la segmentation de ce type d'images. FIG. 4 contient un exemple de ce type. Notons que cette figure a été obtenue en bruitant les deux classes par la même loi normale $N(0,1)$, autrement dit $p(y_s | x_s = \omega_1) = p(y_s | x_s = \omega_2)$; seuls les paramètres de corrélation sont distincts $q_{00} = 0.4$ et $q_{11} = -0.4$. Aussi, une comparaison entre les résultats de segmentation non supervisée, fondée sur les modèles CMC-BI et CMCouple est précisée dans FIG. 4 (les paramètres ont été estimés par la nouvelle méthode (NM) exposée dans le chapitre III (§ III.3.2)).



Réalisation d'un champ couple $Z = (X, Y)$



Segmentation fondée sur CMC-BI

Segmentation fondée sur
CMCouples

FIG. 4 – Exemple de simulation d'un CM couple à deux classes dont les distributions Y_s Conditionnelles à $X_s = \omega_1, \omega_2$ sont égales et comparaison entre la segmentation non supervisée fondée sur le modèle CMC-BI et CM couple.

II.2.1 Simulation des CMCouples gaussiens

Considérons la distribution $p(x, y)$ dont la fonction énergie $U(x, y)$ est donnée par (II.17). La simulation de champs de Markov couple se fait par l'échantillonneur de Gibbs. Elle nécessite la connaissance des lois des $Z_s = (X_s, Y_s)$ conditionnellement au voisinage, données classiquement par:

$$\begin{aligned} p(z_s | z_{V_s}) &= p(x_s, y_s | x_{V_s}, y_{V_s}) \\ &= p(x_s | x_{V_s}, y_{V_s}) p(y_s | x_s, x_{V_s}, y_{V_s}) \end{aligned} \quad (\text{II.19})$$

Afin de simplifier l'écriture, nous omettons (x_{V_s}, y_{V_s}) dans les équations (II.20-II.22) ci-après. L'équation (II.19) s'écrit alors $p(z_s) = p(x_s)p(y_s|x_s)$. Ainsi pour simuler le couple $z_s = (x_s, y_s)$ conditionnellement à son voisinage, nous procédons de la manière suivante :

- Simuler d'abord x_s selon la loi :

$$p(x_s) \propto \frac{1}{\sqrt{a_{x_s}}} \exp \left[-0.5 \left(\sum_{c \in C} \varphi_c(x_c) - \frac{1}{4a_{x_s}} \left(\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t})^2 \right) \right) \right] \quad (\text{II.20})$$

- Simuler ensuite y_s selon la densité gaussienne $p(y_s|x_s)$ de moyenne M_{x_s} et de variance Σ_{x_s} , données par :

$$M_{x_s} = m_{x_s} - \frac{\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t})}{2a_{x_s}} \quad (\text{II.21})$$

$$\text{et : } \Sigma_{x_s}^2 = \frac{1}{a_{x_s}} \quad (\text{II.22})$$

Nous présentons dans FIG.5 un exemple de réalisation d'un champ de Markov couple à quatre classes, simulé par échantillonneur de Gibbs en prenant $\alpha = 2$, $a_{x_s} = 1$, $m_{x_s} = 0, 1, 2, 3$ et $q_{x_s, x_r} = -0.4$.

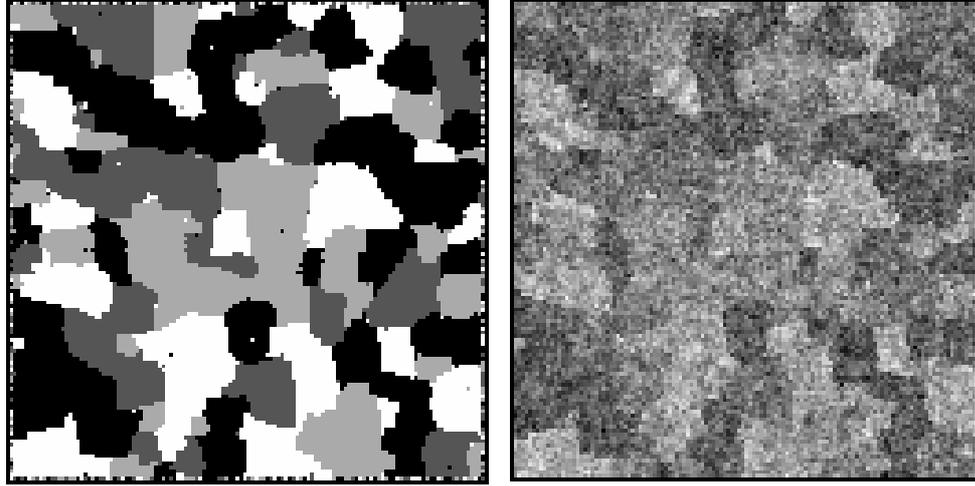


FIG.5 – Exemple de simulation d'un champ de Markov couple à 4 classes.

On constate que contrairement aux deux cas précédents, à savoir : les bruitages corrélé (FIG. 2) et blanc gaussien (FIG. 1), le bruit présent est fortement corrélé.

II. 3 Champs de Markov Triplets

Soit $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ deux champs aléatoires où chaque X_s , Y_s est à valeurs, respectivement, dans $\Omega = \{\omega_1, \dots, \omega_k\}$ et R . La modélisation par champs de Markov triplets (CMT) est une extension des champs de Markov couples, dans laquelle on introduit un troisième processus $U = (U_s)_{s \in S}$, où chaque variable aléatoire U_s est à valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \dots, \lambda_m\}$. Ainsi, le triplet $T = (X, U, Y)$ est un CMT, si sa loi est définie par :

$$p(t) = \gamma \exp - \left(\sum_{c \in C} \varphi_c(t_c) \right) \quad (\text{II.23})$$

où C est l'ensemble des cliques donné par le système de voisinage V considéré.

Afin de simplifier les notations, nous posons $Z = (X, Y)$ et $W = (X, U)$. La loi de W conditionnelle aux observations, notée $p(w|y)$ avec $w = (\omega, \lambda)$, est donnée par :

$$\begin{aligned} p(w|y) &= \gamma(y) \exp[-U(w, y)] \\ &= \gamma(y) \exp\left[-\sum_{c \in C} \varphi_c(w_c, y_c)\right] \end{aligned} \quad (\text{II.24})$$

$$\text{avec : } \gamma(y) = \frac{1}{\sum_{w \in \Omega^N \times \Lambda^N} \exp\left[-\sum_{c \in C} \varphi_c(w_c, y_c)\right]} \quad (\text{II.25})$$

est une loi de Markov. Il en résulte que les lois marginales $p(w_s|y)$ peuvent être estimées à partir des simulations par l'échantillonneur de Gibbs et donc la segmentation MPM est possible. En effet, une telle segmentation s'effectue à partir de la loi $p(x_s|y)$ qui est donnée à partir de la loi $p(w_s = (\omega_i, \lambda_j)|y)$ par :

$$p(x_s = \omega_i|y) = \sum_{\lambda \in \Lambda} p(x_s = \omega_i, u_s = \lambda|y) \quad (\text{II.26})$$

Exemples

Considérons les deux exemples, de complexité croissante, des champs de Markov triplets. Ces exemples seront utilisés ultérieurement dans des expérimentations.

1. On suppose que le couple $W = (X, U)$ est markovien, Y et U sont indépendants conditionnellement à X , et les Y_s sont indépendants conditionnellement à X et $p(y_s|x) = p(y_s|x_s)$ pour tout $s \in S$. Notons que X n'est pas nécessairement de Markov. Sous ces hypothèses la loi de T est donnée par :

$$p(t) = p(w)p(y|x) \tag{II.27}$$

$$= \gamma \exp \left[- \sum_{c \in C} \varphi_c(w_c) \right] \prod_{s \in S} p(y_s | x_s)$$

Cet exemple simple n'est ni nécessairement un CMC (X n'est pas nécessairement de Markov), ni nécessairement un CM couple ((X, Y) n'est pas nécessairement de Markov).

2. On considère directement la markovianité du triplet $T = (X, U, Y)$, avec Y et U indépendants conditionnellement à X et $p(y|x)$ gaussiennes. La loi de T s'écrit :

$$p(t) \propto \exp[-U(w, y)] \tag{II.28}$$

avec :

$$U(w, y) = 0.5 \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{c \in C} \varphi_{c \in C}(w_s) + \sum_{(s,t) \in C} q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right] \tag{II.29}$$

Notons que le champ U peut jouer un rôle important dans la segmentation. En particulier, il peut avoir différentes interprétations. Il peut, par exemple, représenter l'ensemble de sous classes [PBL 02] ou l'ensemble des parties de Ω noté $P(\Omega)$, dans le cas évidentiel [PiB 05, BeP 05], ou encore modéliser les diverses stationnarités contenues dans l'image [BeP 05]. Ces différentes interprétations sont brièvement décrites ci-après :

1. Soit $Z = (X, Y)$ un CM couple à deux classes, telles que les distributions de Y_s conditionnellement aux $X_s = x_s$ possèdent des densités complexes f_1, f_2 . On peut approcher ces densités par un mélange $f_1 \approx \sum_{i=1}^m \alpha_1^i f_1^i$, $f_2 \approx \sum_{i=1}^m \alpha_2^i f_2^i$ où $f_1^m, \dots, f_1^1, f_2^1, \dots, f_2^m$ qui sont des densités données. Une telle situation peut être considérée comme un cas particulier des CMT. En effet, on peut considérer le processus U tel que le couple $Z = (X, Y)$ ait la même structure

et $p(u, y|x) = \prod_{s \in S} p(u_s, y_s | x_s) = \prod_{s \in S} p(u_s | x_s) p(y_s | x_s, u_s)$, avec $p(u_s = \lambda_i | x_s = \omega_j) = \alpha_j^i$ et $p(y_s | x_s = \omega_j, u_s = \lambda_i) = f_j^i(y_s)$. En particulier, on retrouve une démarche classique si le couple $Z = (X, Y)$ est un CMC-BI avec X un champ markovien et la distribution $p(y|x) = \prod_{s \in S} p(y_s | x_s)$.

2. La deuxième interprétation est semblable à la première, à l'exception que les densités

$$f_1 = \sum_{i=1}^m \alpha_1^i f_1^i, \quad f_2 = \sum_{i=1}^m \alpha_2^i f_2^i$$

sont connues. Un tel cas se produit lorsque chacune des deux classes est une union de sous-classes, admettant des densités connues. La segmentation permet alors de retrouver, soit les classes $X = x$, soit les classes et sous-classes $(X, U) = (x, u)$.

3. La troisième interprétation a trait à *la théorie de l'évidence* [Sha 76]. En considérant un CMC-BI $Z = (X, Y)$, et en supposant que la distribution de X n'est pas bien connue et ne peut être estimée à cause, par exemple, d'une non stationnarité élevée. Dans de tels cas, la distribution de X peut être remplacée par "une fonction de masse" et la distribution *a posteriori* de X est obtenue par la *règle de combinaison de Dempster-Shaffer* [Sha 76] qui fusionne la fonction de masse avec la probabilité obtenue à partir de l'observation $Y = y$. En général, une telle fusion détruit la markovianité ; cependant, ces résultats restent exploitables, car le résultat de la fusion est un CMT. Une étude plus approfondie de ce cas est présentée dans la section II.4.

4. La dernière interprétation concerne la modélisation des différentes stationnarités présentes dans l'image. Celle-ci est détaillée dans la section II.5.

II. 4 Champs de Markov évidentiels (CME)

Parfois décrite comme une généralisation de la théorie probabiliste classique, la théorie de l'évidence fut introduite en 1967 par Dempster [Dem 67, Dem 68] puis reprise par Shafer en 1976 [Sha 76]. Elle a l'avantage de pouvoir représenter de manière parfois plus complète l'incertitude sur un événement. A ce titre, Smets [Sme 90] la qualifie de science "*cognitive*" par rapport à la théorie bayésienne, laquelle est qualifiée de science décisionnelle.

Elle peut également permettre une amélioration des résultats obtenus par des traitements bayésiens dans un certain nombre de situations. Cela est particulièrement vrai dans un CMC-BI lorsque les paramètres de $p(x)$ sont mal connus. Auquel cas, cette loi *a priori* de X est remplacée par des "*fonctions de masses*" évidentielles. Une des possibilités de définition d'une fonction de masse M , à partir d'une loi de probabilité, est l'utilisation d'un "*affaiblissement*". La loi *a posteriori* de X est alors obtenue par la fusion de Dempster-Shafer de la masse évidentielle obtenue par affaiblissement avec une loi de probabilité définie par les observations $Y = y$.

Nous commençons cette section par un bref rappel sur les notions de base de la théorie de l'évidence, à savoir les fonctions de masse M , les plausibilités Pl et les crédibilités Cr . Nous donnons ensuite la règle de fusion de Dempster-Shafer et décrivons la possibilité de son utilisation dans le cadre des champs de Markov. Finalement nous montrons le lien existant entre les CMT et les CME.

II.4.1 Notions de base de la théorie de l'évidence

Soit $\Omega = \{\omega_1, \dots, \omega_k\}$ l'ensemble des classes, appelé "*cadre de discernement*" ou "*cadre d'intérêt*" en théorie de l'évidence. Ce dernier est composé de k hypothèses exhaustives (la solution des problèmes est obligatoirement l'une des hypothèses ω_i) et exclusives (la solution est unique). A partir de ce cadre de discernement est déduit l'ensemble des parties de Ω noté $P(\Omega) = \{\emptyset, \{\omega_1\}, \dots, \{\omega_k\}, \dots, \{\omega_1, \omega_k\}, \dots, \Omega\}$ comprenant les 2^k sous-ensembles de Ω . Une fonction de masse est une fonction de $P(\Omega)$ dans $[0,1]$ vérifiant:

$$M(\emptyset) = 0 \tag{II.30}$$

$$\sum_{B \in \mathcal{P}(\Omega)} M(B) = 1 \quad (\text{II.31})$$

La valeur $M(B)$ représente la partie du degré de croyance placée exactement sur la proposition B [Jan 97]. A partir de la fonction de masse M , sont définies les fonctions de crédibilité Cr et de plausibilité Pl par:

$$Cr(B) = \sum_{D \subseteq B} M(D) \quad (\text{II.32})$$

$$Pl(B) = \sum_{B \cap D \neq \emptyset} M(D) \quad (\text{II.33})$$

Aussi, la fonction de crédibilité $Cr(B)$ prend en compte toute la croyance sur la proposition B , elle mesure à quel point les informations fournies par une source soutiennent B . Par ailleurs, la fonction de plausibilité $Pl(B)$ mesure à quel point les informations données par une source ne contredisent pas la proposition B . Classiquement, on montre le lien suivant entre Pl et Cr :

$$Pl(B) + Cr(B^c) = 1 \quad (\text{II.34})$$

où B^c est le complémentaire de B dans Ω . Ainsi, la connaissance de l'une des fonctions (Cr ou Pl) permet d'en déduire l'autre. On montre que la crédibilité vérifie les propriétés suivantes:

$$Cr(\emptyset) = 0$$

$$Cr(\Omega) = 1$$

$$Cr(B_1 \cup \dots \cup B_m) \geq \sum_{I \subseteq \{1, \dots, m\}} (-1)^{\text{Card}(I)+1} Cr\left(\bigcap_{i \in I} B_i\right) \quad (\text{II.35})$$

où B_1, \dots, B_m sont les m éléments de $\mathcal{P}(\Omega)$ et $\text{Card}(I)$ le cardinal de l'ensemble I .

En outre, et par analogie les propriétés concernant la fonction de plausibilité Pl s'obtiennent en remplaçant l'inégalité \geq par \leq . Enfin, soit Cr vérifiant (II.35), il est possible d'établir que M définie par :

$$M(B) = \sum_{D \subset B} (-1)^{\text{Card}(B-D)} Cr(D) \quad (\text{II.36})$$

est une fonction de masse redonnant la même Cr définie par (II.32) [GuB 91].

Ainsi, la connaissance de l'une des fonctions M , Cr ou Pl permet de déduire les deux autres. Par ailleurs, la fonction de probabilité peut être vue comme étant un cas particulier lorsque ces trois fonctions sont égales. En d'autres termes, lorsque M est nulle en dehors des singletons, les notions de masses, de crédibilités et de plausibilités sont confondues et équivalentes à la notion usuelle de probabilité.

II.4.2 Fusion de Dempster-Shafer

Soit $m+1$ fonctions de masse M_0, \dots, M_m définies sur $P(\Omega)$. La règle de fusion de Dempster-Shafer, dont le résultat sera noté par $M = M_0 \oplus M_1 \dots \oplus M_m$, est la suivante:

$$M(B) = \frac{1}{1-H} \sum_{B_0 \cap B_1 \cap \dots \cap B_m = B} \left[\prod_{j=0}^m M_j(B_j) \right] \quad (\text{II.37})$$

avec:

$$H = \sum_{B_0 \cap B_1 \cap \dots \cap B_m = B} \left[\prod_{j=0}^m M_j(B_j) \right] \quad (\text{II.38})$$

Précisons que dans les sommes (II.37) et (II.38) B est fixe et elles sont prises sur tous les $(m+1)$ -uples (B_0, \dots, B_m) , tels que $B_0 \cap \dots \cap B_m = B$. Par ailleurs, le coefficient H représente le degré de conflit entre les différentes sources d'information.

Remarque

On montre classiquement que, si parmi les fonctions de masse M_0, \dots, M_m une au moins est une probabilité, alors $M = M_0 \oplus \dots \oplus M_m$ est également une probabilité.

II.4.3 Fusion DS dans les champs de Markov

Soit S un ensemble de pixels, $(X)_{s \in S}$, $(Y)_{s \in S}$ et $(U)_{s \in S}$ trois champs aléatoires. Chacune des variables X_s , Y_s et U_s est définie dans $\Omega = \{\omega_1, \dots, \omega_k\}$ l'ensemble de classes, R l'ensemble des réels et $\Lambda = P(\Omega)$ l'ensemble des parties de Ω , respectivement. La fonction de masse M_0 est un champ de Markov "évidentiel" (CME) défini sur $(P(\Omega))^N$ si:

$$M_0(B) = \gamma \exp\left(-\sum_{c \in C} \varphi_c(B_c)\right) \quad (\text{II.39})$$

où $B = (B_s)_{s \in S}$ et $B_c = (B_s)_{s \in c}$, l'ensemble des cliques C correspondant à un voisinage donné.

En posant $M_1^y(x) = \frac{\prod_{s \in S} p(y_s | x_s)}{\sum_x \left(\prod_{s \in S} p(y_s | x_s) \right)}$ la probabilité sur Ω^N définie à partir du processus

observé $Y = y$, considérons le résultat de sa fusion de Dempster-Shafer $M^y = M_0 \oplus M_1^y$ avec un CME M_0 . Ce résultat peut être vu comme une généralisation de la probabilité conditionnelle classique $p(x|y) = p(x, y) / p(y)$. En effet, lorsque M_0 est un champ de Markov classique, $p(x, y)$ est un CMC-BI classique et on a $(M_0 \oplus M_1^y)(x) = p(x|y)$.

La probabilité $(M_0 \oplus M_1^y)(x)$ n'est pas nécessairement markovienne ; cependant, elle est formellement la loi marginale d'un CMT, ce qui la rend exploitable. Plus précisément, l'énergie du CMT $T = (X, U, Y)$ correspondant est :

$$\varphi_c(t_c) = \begin{cases} \varphi_c(u_c) & \text{pour } \text{Card}(c) > 1 \\ \varphi_c(u_c) + \log(p(y_s | x_s)) & \text{pour } c = \{s\} \end{cases} \quad (\text{II.40})$$

En effet, la probabilité *a posteriori* $M^y = M_0 \oplus M_1^y$ s'écrit :

$$\begin{aligned} M^y(x) &= (M_0 \oplus M_1^y)(x) \propto \sum_{u: x \in u} \left[\exp - \left(\sum_{c \in C} \varphi_c(u_c) \prod_{s \in S} p(y_s | x_s) \right) \right] \\ &= \sum_{u: x \in u} \left[\exp - \left(\sum_{c \in C} \varphi_c(u_c) - \sum_{s \in S} \log(p(y_s | x_s)) \right) \right] \end{aligned} \quad (\text{II.41})$$

Notons que dans la somme figurant dans l'équation (II.41), $x = (x_s)_{s \in S}$ est fixé et $u = (u_s)_{s \in S}$ varie dans l'ensemble $[P(\Omega)]^N$. Plus précisément, l'énergie dans (II.41) est définie pour $(x, u) \in \Delta^N$ avec $\Delta \subset \Omega \times \Lambda$ tel que $[(\omega, \lambda) \in \Delta] \Leftrightarrow [\omega \in \lambda]$.

A titre d'exemple, pour un champ de Markov à deux classes, on a $\Omega = \{\omega_1, \omega_2\}$, $\Lambda = P(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$, et le couple (X_s, U_s) prend ses valeurs dans $\{(\omega_1, \{\omega_1\}), (\omega_1, \{\omega_1, \omega_2\}), (\omega_2, \{\omega_2\}), (\omega_2, \{\omega_1, \omega_2\})\}$. La fonction de masse M^y devient:

$$\begin{aligned} M^y(x) &= \sum_{u | (x, u, y) \in \Delta^N \times R^N} \left[\exp - \left(\sum_{c \in C} \varphi_c(u_c) - \sum_{s \in S} \log(p(y_s | x_s)) \right) \right] \\ &= \sum_{u | (x, u, y) \in \Delta^N \times R^N} \exp - \varphi(x_c, u_c, y_c) \end{aligned} \quad (\text{II.42})$$

Ainsi, on retrouve le fait que M^y est la marginale du champ triplet $T = (X, U, Y)$ défini par :

$$p(x, u, y) = \gamma \exp[-U(x, u) + \sum_{s \in S} \log(p(y_s | x_s))] \quad (\text{II.43})$$

Finalement, le résultat de la fusion de Dempster-Shafer est la loi conditionnelle $p(x|y)$ induite par un champ triplet $T = (X, U, Y)$, avec le processus auxiliaire U , à valeurs dans l'ensemble des parties $P(\Omega)$ de Ω . De plus, pour $x = u$ l'affaiblissement est réduit à néant et on retrouve l'énergie $U(x)$ classique du champ de Markov caché (équation (II.2)).

Remarques

1. Notons que la théorie de l'évidence peut également être utilisée dans le cadre des champs de Markov au niveau des capteurs, comme proposé dans [BDF 01]. Supposons que l'on dispose d'une image à deux classes et que l'on souhaite utiliser le modèle classique donné par (II.5), en présence des nuages. Le champ de Markov $p(x) \propto \exp[-U(x)]$ n'est pas affecté par les nuages ; par contre, la probabilité $M_1^y(x) \propto \prod_{s \in S} p(y_s | x_s)$ change de nature et tout se passe comme s'il y avait trois classes : ω_1 , ω_2 , et "nuages". La troisième classe "nuages" peut alors être interprétée comme une classe n'apportant aucune information d'intérêt et assimilée à Ω . Finalement, la probabilité $M_1^y(x) \propto \prod_{s \in S} p(y_s | x_s)$ fournie par les capteurs peut être interprétée comme une masse définie sur $\{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}^N$ et l'on montre que sa fusion avec le champ de Markov $p(x) \propto \exp[-U(x)]$ reste un champ de Markov. Ce dernier apparaît alors immédiatement comme une extension du champ de Markov *a posteriori* classique et peut être classiquement utilisé à des fins de segmentation.
2. Nous avons exposé les champs de Markov évidentiels dans le cas caché, par analogie la même étude peut être effectuée dans le cas des champs de Markov couples. Le modèle ainsi obtenu est appelé champ de Markov évidentiel couple (CMEC).

II.4.4 Critique de la règle de la fusion de Dempster-Shafer

Une faiblesse de la fusion de Dempster-Shafer fréquemment soulevée concerne le coefficient de normalisation et la validité de la règle de fusion dans le cas de masses fortement conflictuelles. L'exemple le plus cité dans la littérature est celui de Zadeh [Zad 79], qui peut être formulé comme suit:

On est amené à résoudre une affaire criminelle, on dispose de trois suspects " H_1 : Henri, H_2 : Hector, H_3 : Hervé" et de deux témoins. Soient M_1 le degré de croyance fourni par le premier témoin quant à la culpabilité de chacun des suspects et M_2 le degré de croyance fourni par le second.

	H_1	H_2	H_3
M_1	0.9	0.1	0.0
M_2	0.0	0.1	0.9

En considérant que les valeurs des degrés de croyance M_1 et M_2 sont données par le tableau ci-dessus, nous constatons que le premier témoin est presque sûr que *Henri* est le meurtrier, en revanche le second témoin le disculpe en accusant *Hervé*. Par ailleurs, l'utilisation de la règle de fusion de Dempster-Shafer donne $M(H_1) = M(H_3) = 0.$ et $M(H_2) = 1.$, ce qui apparaît comme peu compatible avec le fait que H_2 soit presque innocenté par les deux témoins. Ainsi, dans le cas de conflit entre les sources, la loi de combinaison de DS a ses limites. Les critiques faites dans ce sens concernent, d'une part, une perte d'information due à la normalisation qui a pour effet de masquer le conflit entre les sources. D'autre part, l'obtention de résultats contre-intuitifs dans le cas d'un conflit élevé, due au facteur de normalisation qui, au voisinage de 1 pour H de l'équation (II.38), est instable [Jan 97]. Pour remédier à ce problème, différentes solutions ont été proposées, dont les principales sont:

- Solutions monde ouvert (*open-world*): Correspond au cas où les sources combinées sont totalement fiables. Ainsi, le conflit entre les sources ne peut provenir que de la non prise en compte d'une ou de plusieurs hypothèses dans l'ensemble Ω . La solution proposée par Smets [Sme 88, Sme 90] est de supprimer la normalisation et d'affecter toute la masse conflictuelle H à l'ensemble vide \emptyset ;
- Solutions monde fermé (*closed-world*): Correspond au cas où une ou plusieurs sources sont non fiables. Différentes solutions peuvent être utilisées, tels que l'affaiblissement des masses de toutes les sources à fusionner [Sha 76], ou l'utilisation des lois de fusion de Yager [YKF 94], ou de Dubois et Prade [DuP 86],...

II. 5 La non stationnarité

Bien que la modélisation classique (CMC-BI) soit souvent bien adaptée au cas stationnaire, fournissant de bons résultats de segmentation, son utilisation en traitement d'images non stationnaires présente des faiblesses et ce, à cause de la non prise en compte des différentes stationnarités présentes dans une image donnée. Ces lacunes peuvent être surmontées par l'utilisation d'un CMT particulier [BeP 05] qu'on appelle "*champ de Markov triplet adapté*" (CMTA-BI). Nous verrons que ce champ triplet (*respectivement "champ de Markov triplet adapté"* (CMTA) est un champ stationnaire qui modélise un CMC-BI (*respectivement CMCouple*) non stationnaire.

II.5.1 Champs de Markov triplet adaptés à bruit indépendant (CMTA-BI)

Soit S un ensemble de pixels, X , Y et U trois champs aléatoires. Pour chaque pixel $s \in S$, les variables aléatoires X_s , Y_s et U_s sont à valeurs dans $\Omega = \{\omega_1, \dots, \omega_k\}$, R et $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, respectivement. Nous avons donc un modèle triplet, dans lequel les m valeurs de U_s modélisent les m différentes stationnarités du champ X . Comme dans le cas stationnaire, le problème est de retrouver le champ inobservé $X = x$, à partir de sa version bruitée $Y = y$. Afin de simplifier les notations, nous posons $Z = (X, Y)$ et $W = (X, U)$. Nous considérons un cas particulier en supposant que le couple (X, U) est de Markov. Nous souhaitons donc définir la loi de (X, U) de manière à ce que $p(x|u)$ modélise de manière satisfaisante la non stationnarité du champs aléatoire X . Supposons que le champ non stationnaire X suit le modèle de Potts :

$$p(x) = \gamma \exp - \left[\sum_{(s,t) \in C} \alpha_{(s,t)} (1 - 2\delta'(x_s, x_t)) \right] \quad (\text{II.44})$$

avec $\delta'(x_s, x_t) = 0$ pour $x_s \neq x_t$, et $\delta'(x_s, x_t) = 1$ pour $x_s = x_t$.

Nous proposons de modéliser sa non stationnarité, due à la dépendance de $\alpha_{(s,t)}$ de (s, t) , par le champ de Markov stationnaire (X, U) dont la fonction énergie $U(x, u)$ est choisie de la forme suivante:

$$U(x, u) = \sum_{(s,t) \in C} \alpha^1 (1 - 2\delta'(x_s, x_t)) - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots + \alpha_{\lambda_M}^2 \delta(u_s, u_t, \lambda_M)) (1 - \delta'(x_s, x_t)) \quad (\text{II.45})$$

avec C l'ensemble des cliques et:

$$\delta(u_s, u_t, \lambda_i) = \begin{cases} 1 & \text{pour } u_s = u_t = \lambda_i \\ 0 & \text{sinon} \end{cases} \quad (\text{II.46})$$

Par ailleurs, en supposant que les variables aléatoires Y_s sont indépendantes conditionnellement à X , et que $p(y_s | x, u) = p(y_s | x_s)$, nous avons:

$$p(y | x, u) = \prod_{s \in S} p(y_s | x_s) \quad (\text{II.47})$$

La distribution $p(x, u, y)$ est alors obtenue à partir des équations (II.44) - (II.47) comme suit:

$$\begin{aligned} p(x, u, y) &= p(x, u) p(y | x, u) \\ &= \gamma \exp \left[\sum_{(s,t) \in C} \alpha^1 (1 - 2\delta'(x_s, x_t)) - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots \right. \\ &\quad \left. + \alpha_{\lambda_M}^2 \delta(u_s, u_t, \lambda_M)) (1 - \delta'(x_s, x_t)) - \sum_{s \in S} \log(p(y_s | x_s)) \right] \end{aligned} \quad (\text{II.48})$$

nous avons donc une distribution markovienne. De plus, $p(x, y | u)$ est un CMC-BI classique non stationnaire : ainsi la non stationnarité de CMC-BI est ici modélisée par un CMT stationnaire particulier.

Ainsi que mentionné précédemment, les probabilités $p((x_s, u_s) = (\omega_i, \lambda_j) | y)$ peuvent alors être estimées, et on peut ainsi calculer la distribution $p(x_s = \omega_i | y) = \sum_{u \in \Lambda} p(x_s = \omega_i, u_s = \lambda_j | y)$ et l'utiliser pour l'appliquer dans la méthode de segmentation MPM. Notons également que $p(u_s | y)$ peut être calculée par :

$$p(u_s|y) = \sum_{\omega_i \in \Omega} p(x_s = \omega_i, u_s = \lambda_j | y) \quad (\text{II.49})$$

permettant ainsi d'estimer également le champ U représentant les différentes stationnarités présentes dans l'image. De plus, le champ CMTA étant stationnaire, il est envisageable d'estimer ses paramètres et considérer des segmentations non supervisées. Des exemples de segmentations non supervisées des images non stationnaires seront présentées dans le chapitre IV.

Exemple

Considérons une image constituée de deux parties dont l'une représente une "zone urbaine" et l'autre une "zone rurale", où chacune d'elles contient deux classes ("espaces verts" et "maisons") à densité disproportionnée. Ainsi, on trouve une prédominance de la classe "maisons" dans la "zone urbaine" alors qu'il y'a prédominance de la classe "espaces vert" dans la "zone rurale". Dans ce type d'images on est confronté à deux problèmes: le premier concerne la modélisation de ces deux zones, le second à trait à la modélisation des différentes classes, étant donné que la loi de X diffère dans chacune des deux zones. Les CMTA-BI permettent cette modélisation. En effet, pour $s \in S$, x_s appartient soit à la classe ω_1 : "espace vert" ou ω_2 : "maison", et u_s appartient soit à λ_1 : la "zone urbaine" ou λ_2 : "zone rurale" représentant les différentes stationnarités de l'image associées au champ $U = u$. Ainsi, lorsqu'on s'intéresse aux différentes stationnarités contenues dans l'image, on se focalise sur le champ U qu'on estime en utilisant l'équation (II.49), et lorsqu'on veut retrouver le champ des classes, on estime le champ caché X par (II.26).

Remarque

La transition du modèle CMTA-BI au modèle classique CMC-BI peut se faire en posant $x = u$. En effet, lorsque $x = u$, la fonction énergie (II.45) sera égale à l'énergie correspondante au CMC-BI (II.2).

II.5.2 Champs de Markov triplet adapté (CMTA)

Dans la sous-section précédente nous avons traité le problème de la non stationnarité dans le cas simple du modèle de Potts avec du bruit indépendant du type "BI", et nous avons proposé une solution par l'utilisation du nouveau modèle CMTA-BI. Cependant, cette modélisation présente les mêmes faiblesses que celles évoquées dans la modélisation par champs de Markov cachés à bruit indépendant, à cause de la simplicité de la distribution $p(y|x)$. En effet, en pratique les variables aléatoires Y_s ne sont pas, en général, indépendantes conditionnellement à X . Pour y remédier, une extension du CMTA-BI peut être envisagée. L'idée est d'enrichir les champs de Markov couples [PiT 00] de telle sorte que la non stationnarité soit prise en compte.

Comme précédemment, considérons un ensemble de pixels S et X , U et Y trois champs aléatoires définis sur S , tel que X est un champ caché et doit être retrouvé ou estimé à partir du champ observé $Y = y$. Pour chaque $s \in S$, les variables X_s , U_s et Y_s sont à valeurs dans $\Omega = \{\omega_1, \dots, \omega_k\}$, $\Lambda = \{\lambda_1, \dots, \lambda_m\}$ et R , respectivement.

Considérons la loi de $T = (X, U, Y)$ donnée par:

$$\begin{aligned}
 p(x, u, y) &= p(x, u)p(y|x, u) = p(x, u)p(y|x) \\
 &\propto \exp - 0.5 \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C} \alpha^1 (1 - 2\delta(x_s, x_t)) \right. \\
 &\quad \left. - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots + \alpha_{\lambda_M}^2 \delta(u_s, u_t, \lambda_M)) (1 - \delta(x_s, x_t)) \right. \\
 &\quad \left. + \sum_{(s,t) \in C} [q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t})] \right]
 \end{aligned} \tag{II.50}$$

Soit $(x_{V_s}, u_{V_s}, y_{V_s})$ la restriction de (x, u, y) au voisinage de $s \in S$, on a alors:

$$\begin{aligned}
 p(x_s, u_s, y_s | x_{V_s}, u_{V_s}, y_{V_s}) &= p(x_s, u_s | x_{V_s}, u_{V_s}, y_{V_s}) p(y_s | x_s, u_s, x_{V_s}, u_{V_s}, y_{V_s}) \\
 &= p(x_s, u_s | x_{V_s}, u_{V_s}, y_{V_s}) p(y_s | x_s, x_{V_s}, y_{V_s})
 \end{aligned} \tag{II.51}$$

où:

$$\begin{aligned}
 p(x_s, u_s | x_{V_s}, u_{V_s}, y_{V_s}) &\propto \frac{1}{\sqrt{a_{x_s}}} \exp - 0.5 \left[\sum_{(s,t) \in C} \alpha^1 (1 - 2\delta(x_s, x_t)) \right. \\
 &\quad \left. - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots + \alpha_{\lambda_M}^2 \delta(u_s, u_t, \lambda_M)) (1 - \delta(x_s, x_t)) \right] \\
 &\quad \left. - \frac{1}{4a_{x_s}} \left(\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t}) \right)^2 \right]
 \end{aligned} \tag{II.52}$$

et $p(y_s | x_s, x_{V_s}, y_{V_s})$ est gaussien dont la moyenne et la variance sont données par:

$$M_{x_s} = m_{x_s} - \frac{\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t})}{2a_{x_s}} \tag{II.53}$$

$$\Sigma_{x_s}^2 = \frac{1}{a_{x_s}} \tag{II.54}$$

Ainsi, la simulation du champ de Markov triplet adapté (CMTA) s'effectue en deux étapes et de la même façon que la simulation des champs de Markov couples étudiés dans la section (II. 2). On simule d'abord le couple (x_s, u_s) en se basant sur l'équation (II.52), ensuite y_s est simulé à partir de la densité gaussienne définie par les équations (II.53) et (II.54).

Comme précédemment, les probabilités $p((x_s, u_s) = (\omega_i, \lambda_j) | y)$ étant estimées, les distributions $p(x_s = \omega_i | y)$ et $p(u_s = \lambda_j | y)$ permettant d'obtenir le champs X et U sont calculées par (II.26) et (II.49), respectivement.

II. 6 Conclusion

Dans ce chapitre, nous avons présenté les différents modèles, dont certains originaux, étudiés au cours de cette thèse. Ainsi les champs de Markov cachés à bruit indépendant (CMC-BI), les champs de Markov cachés (CMC), les champs de Markov couples (CMCouples) et les champs de Markov triplet (CMT) ont été décrits. Les CMT généralisent les CMCouples lesquels généralisent les CMC. En effet, dans $T = (X, U, Y)$ la distribution marginale $Z = (X, Y)$ n'est pas nécessairement de Markov et par conséquent n'est pas nécessairement un CMCouple. Cette généralisation fait la richesse des CMT par rapport aux CMCouples, offrant ainsi une flexibilité et une souplesse de traitement et ce, grâce à l'introduction du champ supplémentaire U à différentes interprétations.

Bien adaptés au cas stationnaire, les CMC et CMCouples fournissent de bons résultats de segmentation. Cependant, dans le cas d'images non stationnaires, l'utilisation de ces modèles présente des limites. Ceci nous a amené à introduire deux variantes originales des CMT, dites "*champs de Markov évidentiels*" (CME) et "*champs de Markov triplet adaptés*" (CMTA). Dans les CME, des fonctions de masses au niveau de la loi *a priori*, ont été introduites. L'introduction d'un certain type d'affaiblissement au niveau de la loi *a priori* aboutit, formellement, à un champ de Markov triplet, ce qui autorise les traitements d'intérêt [PiB 05], en particulier, lorsque la loi *a priori* est non stationnaire et inconnue. D'autre part, en plus des avantages fournis par les CME, les CMTA ont la possibilité de modéliser les différentes homogénéités. Ceux-ci étant des CMT stationnaires particuliers, il est possible, ainsi que nous le verrons dans la suite, d'estimer leurs paramètres et les utiliser dans les segmentations non supervisées d'images.

Chapitre III

Estimation des paramètres

Dans le chapitre précédent, nous avons étudié divers modèles markoviens à base d'énergie dépendant d'un certain nombre de paramètres. Cet ensemble de paramètres de la loi du couple (X, Y) est constitué d'un ensemble α qu'on appelle "*ensemble de paramètres d'interaction*" et qui définit, en général, la loi du champ caché X , et d'un ensemble η définissant la loi de Y conditionnelle à X (ou encore, la loi du "*bruit*"). On notera $\theta = (\alpha, \eta)$ l'ensemble de paramètres nécessaires pour mettre en œuvre une méthode de segmentation choisie. Cependant, dans les applications réelles, ces paramètres sont souvent inconnus, d'où la nécessité de développer des méthodes de leur estimation.

Deux niveaux de difficultés peuvent être distingués. Le premier concerne le cas des données complètes ; dans ce cas on dispose du champ de classes, que l'on appelle aussi "*vérité terrain*". Ainsi, l'estimation s'effectue à partir du champ observé Y et du champ de classes X pour les modèles cachés et couple, et des champs (X, U) et Y pour le modèle triplet. Le deuxième niveau, de complexité beaucoup plus grande, est le cas des données incomplètes. Dans ce cas, l'estimation s'effectue à partir du champ observé Y uniquement. A cet effet, plusieurs méthodes ont été proposées, dont la grande majorité est fondée sur le principe général du maximum de vraisemblance.

Le but de ce chapitre est de décrire les différentes méthodes d'apprentissage des champs de Markov classiques, ainsi que de présenter deux nouvelles techniques récemment

introduites pour l'estimation des paramètres du champ de Markov triplet [BeP 04, BeP 05], et qui font partie de contributions originales de cette thèse.

III.1 Apprentissage à données complètes

Considérons un champ de Markov caché classique (X, Y) à bruit indépendant. Dans l'estimation "à données complètes", on dispose des réalisations $(X, Y) = (x, y)$. L'estimation s'effectue à partir du couple (X, Y) . Dans cette optique, différentes méthodes telles que la méthode de Derin et Elliott, la méthode de codage et la méthode du gradient stochastique ont été proposées. Celles-ci concernent l'estimation des paramètres d'interaction propres à la loi *a priori*. Par ailleurs, l'estimation des paramètres liés au bruit s'effectue de manière différente. Dans le cas classique considéré, le paramètre d'interaction α définissant $p(x|\alpha)$ est estimé à partir de $X = x$ par $\hat{\alpha}(x)$, et le paramètre η définissant $p(y|x, \eta)$ est estimé à partir de $(X, Y) = (x, y)$ par $\hat{\eta}(x, y)$. Nous proposons ci-après une brève description des trois méthodes classiques sus-citées, très utilisées, pour l'estimation de α , dont deux seront reprises dans le cas des champs de Markov triplets. Concernant le paramètre η , son estimation s'effectue par des techniques classiques. Par exemple, dans le cas gaussien les moyennes et les variances sont estimées par :

$$m_i = \frac{\sum_{s \in S} y_s 1_{[x_s = \omega_i]}}{\sum_{s \in S} 1_{[x_s = \omega_i]}} \quad (\text{III.1})$$

$$\sigma_i^2 = \frac{\sum_{s \in S} (y_s - m_i)^2 1_{[x_s = \omega_i]}}{\sum_{s \in S} 1_{[x_s = \omega_i]}} \quad (\text{III.2})$$

III.1.1 Méthode de codage

Proposée par Besag [Bes 74], la méthode de codage (MC) consiste en la partition de l'ensemble S en plusieurs sous ensembles disjoints $S^{(l)}$, appelés "codage", de sorte que deux sites d'un même sous ensemble $S^{(l)}$ ne soient pas voisins. Deux systèmes de codage

correspondant aux voisinages des 4-ppv et 8-ppv, le plus souvent utilisés, sont présentés à la figure (FIG. 1). Ainsi deux sous-ensembles S^1 et S^2 sont nécessaires pour un système à 4-ppv, et quatre pour un système à 8-ppv.

1	2	1	2	1
2	1	2	1	2
1	2	1	2	1
2	1	2	1	2
1	2	1	2	1

1	2	1	2	1
3	4	3	4	3
1	2	1	2	1
3	4	3	4	3
1	2	1	2	1

FIG. 1 – Codage des systèmes à 4-ppv (gauche) et 8-ppv (droite)

Le fait que les sites à l'intérieur de chaque sous-ensemble $S^{(l)}$ ne sont pas mutuellement voisins est un avantage ; en effet, sous l'hypothèse de markovianité, les variables associées aux sites dans chaque sous-ensemble $S^{(l)}$ sont indépendantes conditionnellement à leurs voisins, ce qui permet d'écrire la relation suivante:

$$p^{(l)}(x|\alpha) = \prod_{s \in S^{(l)}} p(x_s | x_{V_s}, \alpha) \quad (\text{III.3})$$

qui, en la maximisant, donne les vecteurs paramètres $\alpha^{(l)}$.

Cependant, cette démarche présente un inconvénient, étant donné que l'estimation de chacun des vecteurs $\alpha^{(l)}$ s'effectue sur la base d'un seul codage. En d'autres termes, à chaque sous ensemble $S^{(l)}$ correspond un vecteur $\alpha^{(l)}$. Ainsi, le nombre de vecteurs paramètres $\alpha^{(l)}$ dépendra du nombre des sous ensembles $S^{(l)}$ (i.e. du système de voisinage choisi), ce qui rend le choix d'une technique de combinaison de ces vecteurs paramètres $\alpha^{(l)}$ délicat pour fournir un résultat optimal. Toutefois, la méthode de combinaison la plus simple, souvent utilisée, est la moyenne empirique. Celle-ci est donnée comme suit:

$$\hat{\alpha} = \frac{1}{L} \sum_{l=1}^L \arg \max_{\alpha} p^{(l)}(x|\alpha) \quad (\text{III.4})$$

A titre d'exemple, considérons le cas classique d'un champ de Markov dont la distribution est donnée par :

$$p(x) = \exp\left(-\left(\alpha_1 \sum_s x_s + \alpha_2 \sum_{s \sim t} (1 - \delta(x_s, x_t))\right)\right) \quad (\text{III.5})$$

avec $\alpha = (\alpha_1, \alpha_2)$. En appliquant le principe de la méthode de codage nous avons :

$$\begin{aligned} p^{(l)}(x|\alpha) &= \prod_{s \in S^{(l)}} p(x_s | x_{V_s}, \alpha) \\ &= \prod_{s \in S^{(l)}} \exp\left(-\left(\alpha_1 x_s + \alpha_2 \sum_{t \in V_s} (1 - \delta(x_s, x_t))\right)\right) \end{aligned} \quad (\text{III.6})$$

On obtient alors les $\alpha^{(l)}$ en résolvant :

$$\frac{\partial \ln p^{(l)}(x|\alpha_1, \alpha_2)}{\partial \alpha_1} = 0 \quad (\text{III.7})$$

$$\frac{\partial \ln p^{(l)}(x|\alpha_1, \alpha_2)}{\partial \alpha_2} = 0 \quad (\text{III.8})$$

Pour $l=1, 2$ le vecteur paramètre final est $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2) = \left(\frac{\alpha_1^{(1)} + \alpha_1^{(2)}}{2}, \frac{\alpha_2^{(1)} + \alpha_2^{(2)}}{2}\right)$.

Remarque

Bien qu'elle soit communément utilisée, cette méthode présente l'inconvénient de faire appel à un système d'équations non linéaires, parfois difficile à résoudre. D'autres critiques concernant cette méthode d'estimation sont explicitées dans l'article de Derin et Elliott [DeE 87].

III.1.2 Le gradient stochastique

Proposée par Younes [You 88], la méthode du gradient stochastique est une méthode itérative permettant d'approcher l'estimateur du maximum de vraisemblance. Le déroulement de la méthode est le suivant :

- Initialiser le vecteur des paramètres α_0 ;
- A chaque itération n , calculer α_{n+1} à partir de α_n et de $X = x$ par :

$$\alpha_{n+1} = \alpha_n + \frac{K}{n+1} [\nabla_{\alpha_n} U(x_{n+1}) - \nabla_{\alpha_n} U(x)], \quad (\text{III.9})$$

où $\nabla_{\alpha_n} U(x)$ est le gradient de $U(x)$ par rapport à α pris en α_n , et K une constante. Notons qu'en pratique, on prend la constante K inversement proportionnelle au nombre des pixels : $K \approx 1/N$.

III.1.3 Méthode de Derin et Elliott

La méthode de Derin et Elliott [DeE 87] est une méthode non itérative et optimale au sens des moindres carrés. Pour des raisons de simplicité, nous rappelons cette méthode dans le cas simple du champs de Markov à k classes $\Omega = \{\omega_1, \dots, \omega_k\}$, 4-ppv, avec l'énergie de la forme

$$U(x) = \sum_{s \in S} \varphi_s(x_s) + \sum_{(s,t)} \varphi_{(s,t)}(x_s, x_t) \quad (\text{III.10})$$

avec $\varphi_s(\omega_i) = \beta_i$, et deux types de $\varphi_{(s,t)}$, selon que la clique (s,t) est horizontale ou verticale. On pose, pour les cliques horizontales, $\varphi_{(s,t)}(x_s, x_t) = -\alpha_H$ pour $x_s = x_t$, et $\varphi_{(s,t)}(x_s, x_t) = \alpha_H$ pour $x_s \neq x_t$, et, pour les cliques verticales, $\varphi_{(s,t)}(x_s, x_t) = -\alpha_V$ pour $x_s = x_t$, et $\varphi_{(s,t)}(x_s, x_t) = \alpha_V$ pour $x_s \neq x_t$.

Finalement, la loi de X est déterminée par les paramètres

$$\alpha = [\beta_1, \dots, \beta_k, \alpha_H, \alpha_V]^T \quad (\text{III.11})$$

Soit s un site donné, V_s l'ensemble de ses voisins. Dans le cas d'un système de voisinage d'ordre 1 considéré, numérotions les pixels, comme indiqué sur la figure FIG. 2, et écrivons V_s sous forme vectorielle :

$$V_s = [t_1, t_2, t_3, t_4]^T \quad (\text{III.12})$$

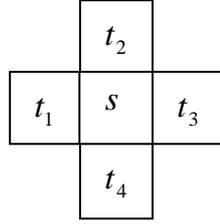


FIG. 2 – Position des t_i dans le système de voisinage d'ordre 1.

Considérons les fonctions indicatrices suivantes :

$$I(x_{s_1}, \dots, x_{s_l}) = \begin{cases} -1 & \text{si } x_{s_1} = x_{s_2} = \dots = x_{s_l} \\ +1 & \text{sinon} \end{cases} \quad (\text{III.13})$$

et

$$J_k(x_s) = \begin{cases} +1 & \text{si } x_s = \omega_k \\ 0 & \text{sinon} \end{cases} \quad (\text{III.14})$$

Les fonctions de potentiel de la distribution de Gibbs du modèle de Markov caché peuvent alors s'exprimer en fonction de ces quantités. En effet, soit $U(x_s, x_{V_s}, \alpha)$ la somme des fonctions de potentiel associées aux cliques contenant le pixel s .

$U(x_s, x_{V_s}, \alpha)$ s'écrit comme suit :

$$U(x_s, x_{V_s}, \alpha) = \Phi^T(x_s, x_{V_s}) \alpha \quad (\text{III.15})$$

avec :

$$\Phi(x_s, x_{V_s}) = [J_1(x_s), \dots, J_k(x_s), (I(x_s, x_{t_1}) + I(x_s, x_{t_3})), (I(x_s, x_{t_2}) + I(x_s, x_{t_4}))] \quad (\text{III.16})$$

La dimension de ce vecteur peut être réduite en mettant les paramètres des singletons à zéro ($\beta_i = 0$), ou élargie en prenant un voisinage plus grand (8-ppv), par exemple.

Nous avons :

$$p(x_s | x_{V_s}) = \frac{p(x_s, x_{V_s})}{p(x_{V_s})} \quad (\text{III.17})$$

sachant que :

$$p(x_s | x_{V_s}) = \gamma \exp[-U(x_s, x_{V_s}, \alpha)] \quad (\text{III.18})$$

$$\text{où: } \gamma = 1 / \sum_{s \in S} \exp[-U(x_s, x_{V_s}, \alpha)] \quad (\text{III.19})$$

Pour deux valeurs distinctes de x_s notées : ω_i , ω_j et pour la même configuration x_{V_s} du voisinage au site s , on a l'expression suivante :

$$[\Phi(\omega_i, x_{V_s}) - \Phi(\omega_j, x_{V_s})]^T \alpha = \ln \left(\frac{p(\omega_j, x_{V_s})}{p(\omega_i, x_{V_s})} \right) \quad (\text{III.20})$$

Ainsi, pour chaque paire distincte de classes (ω_i, ω_j) et pour chacune des configurations possibles de V_s , résulte une équation linéaire. Les probabilités $p(\cdot, x_{V_s})$ peuvent être estimées classiquement, en comptant le nombre de configurations du voisinage du type (\cdot, x_{V_s}) apparaissant dans x . Il est important de noter que les configurations qui n'apparaissent pas dans x et les configurations pour lesquelles $\Phi(\omega_i, x_{V_s}) = \Phi(\omega_j, x_{V_s})$ ne peuvent être utilisées, car elles conduisent aux indéterminations $\ln(0)$ et $0 = C^{st}$, respectivement. Ainsi, le vecteur estimé est la solution d'un système d'équations linéaires. Ce système est surdimensionné, ce qui requiert l'utilisation de la méthode des moindres carrés. Cependant, il arrive parfois que le système ainsi obtenu n'admette pas de solutions (matrice singulière); cette difficulté peut alors être contournée en faisant appel à la méthode SVD explicitée en Annexe.

III.2 Apprentissage à données incomplètes

Dans cette section nous abordons le problème de l'estimation des paramètres dans le cas des données incomplètes. Nous entendons par données incomplètes le cas où l'on ne dispose d'aucun échantillon d'apprentissage, ou encore d'aucune réalisation de X . Auquel cas, l'estimation se fait à partir de la seule observation $Y = y$. Nombreux travaux ont été consacrés au sujet, donnant naissance à divers algorithmes. Parmi les plus utilisés, on peut citer les méthodes générales "Expectation Maximisation" (EM), "Stochastic Expectation Maximisation" (SEM), "Iterative Conditional Estimation" (ICE) ou encore le "Gradient Stochastique" (GS). Cette section est consacrée au rappel de ces méthodes générales et leur utilisation dans les CMC classiques.

III.2.1 Algorithme Expectation Maximisation

Connu sous l'abréviation "EM", cet algorithme a été proposé par Baum *et al* [BPS 70] pour traiter le problème de l'estimation des paramètres dans les chaînes de Markov cachés. Il fut ensuite repris par Dempster *et al.* [DLR 77] pour être appliqué à diverses situations d'observations incomplètes. Son principe consiste à maximiser de manière itérative par rapport à θ la vraisemblance en $Y = y$. Cet algorithme comprend les étapes suivantes:

- Initialiser le vecteur paramètre θ_0 ;
- A chaque itération n ,
 - Etape E: Calcul de l'espérance :

$$E_{\theta_n} [\log(p_{\theta_n}(X, y)|Y = y)] ; \quad (\text{III.21})$$

- Etape M: Maximisation de (III. 21).

$$\theta_{n+1} = \arg \max_{\theta} E_{\theta_n} [\log(p_{\theta}(X, y)|Y = y)]. \quad (\text{III.22})$$

Remarques

1. Dans le cas des champs Markoviens, la résolution de l'équation (III.22) n'est pas envisageable d'un point de vue pratique. D'où la nécessité de la mise en œuvre d'un algorithme stochastique de complexité comparable à celle de l'algorithme EM. A cet effet, Chalmond

[Cha 89] a proposé une modification en remplaçant la vraisemblance donnée en (III.21) par la pseudo-vraisemblance. L'algorithme ainsi obtenu est appelé "*EM Gibbsien*". Une autre approximation dite "*Champ Moyen*" proposée par Chandler [Cha 87] et utilisée par Zhang dans l'algorithme EM a également été proposée [Zha 92] pour contourner cette difficulté.

2. L'algorithme EM peut dépendre fortement de l'initialisation θ_0 . En conséquence, des techniques moins sensibles à l'initialisation, dont certaines sont détaillées, ci-après, ont été recherchées et développées.

III.2.2 Algorithme Stochastic Expectation Maximisation

Introduit en 1983 par Broniatowski *et al.* [BCD 83], cet algorithme itératif est une version stochastique de l'algorithme EM. L'idée principale est d'incorporer une étape stochastique (étape S) précédant l'étape d'estimation (étape E) de l'algorithme EM. Cette étape "S" repose sur un principe d'affectation aléatoire selon un tirage de X suivant les lois *a posteriori*. Les étapes de cet algorithme sont décrites comme suit:

- Initialiser le vecteur paramètre θ_0 ;
- A chaque itération n ,
 - Simulation d'une réalisation x^n de X selon la probabilité *a posteriori* $p(x|y, \theta_n)$;
 - Estimation de θ à partir de (x^n, y) par l'estimateur du maximum de vraisemblance $\hat{\theta}_{MV} : \theta_{n+1} = \hat{\theta}_{MV}(x^n, y)$.

Remarques

1. D'un point de vue pratique, l'algorithme SEM répond aux limitations de l'algorithme EM. En particulier, ses résultats sont moins sensibles à son initialisation ;

2. Il est à noter que Celeux et Diebolt ont démontré que les algorithmes EM et SEM sont reliés d'une manière formelle par une relation de récurrence [CeD 90]. Cette relation a permis de définir une version de type recuit simulé de l'algorithme EM, appelée algorithme "*Simulated Annealing EM*" (SAEM).

III.2.3 Algorithme estimation conditionnelle itérative

L'estimation conditionnelle itérative (ECI, ou ICE de l'anglais "*iterative conditional estimation*") est une procédure générale d'estimation des paramètres dans le cas des données cachées [Pie 92]. De façon générale, soient X et Y deux champs aléatoires dont la loi conjointe dépend d'un paramètre θ , et soit $\hat{\theta} = \hat{\theta}(X, Y)$ un estimateur de celui-ci à partir des données complètes. Le champ X étant inobservable, l'idée de l'ECI est d'approcher $\hat{\theta}(X, Y)$ par l'espérance conditionnelle. Or, l'espérance conditionnelle de $\hat{\theta}$ dépend du paramètre θ . Ainsi, pour calculer $E[\hat{\theta}(X, Y)|Y]$, on est amené à considérer la valeur θ_n (valeur courante) du paramètre. La valeur suivante est donnée par $\theta_{n+1} = E_{\theta_n}[\hat{\theta}(X, Y)|Y = y]$. La procédure itérative ECI est donc :

- Initialiser le vecteur paramètre θ_0 ;
- A chaque itération n , on calcule la valeur θ_{n+1} à partir de θ_n et de $Y = y$ par :

$$\theta_{n+1} = E_{\theta_n}[\hat{\theta}(X, Y)|Y = y] \quad (\text{III. 23})$$

Si $E_n[\hat{\theta}(X, Y)|Y = y]$ n'est pas explicitement calculable, mais la simulation des réalisations de X selon la loi *a posteriori* est possible, on utilise l'approximation :

$$\theta_{n+1} = \frac{\hat{\theta}(x^1, y) + \dots + \hat{\theta}(x^m, y)}{m}, \quad (\text{III. 24})$$

où x^1, x^2, \dots, x^m sont les m simulations des réalisations de X selon sa loi conditionnelle à $Y = y$.

Notons qu'il peut arriver, comme dans le cas des chaînes de Markov cachées, qu'on utilise (III.23) pour estimer certaines composantes de θ , et (II.24) pour estimer les composantes restantes.

Notons également que l'utilisation de $\hat{\theta}(X, Y)$ et la possibilité de simulation de X , selon $p_{\theta}(x|y)$, sont les seules conditions d'utilisation de ECI. Ces conditions sont très faibles dans le cadre de segmentation d'images par champs de Markov qui est le nôtre. En effet, si l'on ne peut pas définir un estimateur à partir des données complètes, on ne peut pas en définir un, à partir des données incomplètes. Par ailleurs, la deuxième condition est toujours vérifiée.

Remarques

1. Les trois méthodes EM, SEM, ICE ont été comparées dans le cadre de la segmentation non supervisée d'images par chaînes de Markov cachées [BeP 95] et par les méthodes contextuelles et aveugles [PeP 95]. Dans le cas des bruits gaussiens considérés dans ces articles, les performances des trois méthodes sont comparables ;
2. Pour un nombre de tirage égal à 1 à chaque itération de l'ECI et lorsque l'estimateur $\hat{\theta}(X, Y)$ est l'estimateur du maximum de vraisemblance, les algorithmes ECI et SEM sont identiques.

III.2.4 Algorithme du gradient stochastique

L'algorithme du gradient stochastique (GS) de Younes, dont le principe est de rechercher le maximum de vraisemblance, a d'abord été proposé dans le cas de données complètes [You 88], puis généralisé au cas des données incomplètes [You 89]. Ainsi, pour une donnée observée $Y = y$, le problème est de trouver le vecteur paramètre θ qui maximise $p_{\theta}(y)$. Comme les méthodes précédentes, GS permet d'estimer les paramètres du modèle d'une manière itérative. Le déroulement de la méthode est le suivant :

- Initialiser le vecteur paramètre θ_0 ;
- A chaque itération n , calculer θ_{n+1} à partir de θ_n et de $Y = y$ par :

$$\theta_{n+1} = \theta_n + \frac{K}{n+1} [\nabla_{\theta_n} U(x_{n+1}, y_{n+1}) - \nabla_{\theta_n} U(x_{n+1}^*, y)] \quad (\text{III.25})$$

où $\nabla_{\theta_n} U(x_{n+1}, y_{n+1})$ est le gradient de $U(x_{n+1}, y_{n+1})$ par rapport à θ pris en θ_n , (x_{n+1}, y_{n+1}) est une réalisation de (X, Y) , simulée par l'échantillonneur de Gibbs en utilisant θ_n , et x_{n+1}^* une réalisation de X simulée par échantillonneur de Gibbs, suivant la loi *a posteriori* utilisant θ_n . K une constante prise, dans la pratique, égale à l'inverse du nombre des pixels.

III.3 Estimation des paramètres du champ de Markov Triplet

Dans cette section nous abordons le problème de l'estimation des paramètres des champs de Markov couple $Z = (X, Y)$ et triplet $T = (X, U, Y)$. Pour ce faire, nous présentons deux méthodes originales, récemment décrites dans [BeP 05]. La première, appelée gradient stochastique adapté (GSA) est une adaptation de la méthode de Younes [You 89] aux cas couple et triplet. La deuxième, qu'on appellera par la suite "*Nouvelle Méthode*" (NM), est une méthode originale de type ECI, où l'estimateur $\hat{\theta}(X, U, Y)$ utilise les moindres carrés pour l'estimation des paramètres du champ de Markov (X, U) , et les méthodes classiques utilisant les moments pour l'estimation des paramètres du bruit.

Notons que l'estimation des paramètres du modèle de Markov triplet est exactement la même que celle du modèle de Markov couple. En effet, un triplet $T = (X, U, Y)$ est également un couple $T = (V, Y)$, avec $V = (X, U)$. Donc, estimer les paramètres de la loi de $T = (X, U, Y)$ à partir de $Y = y$ revient à estimer les paramètres de la loi de $T = (V, Y)$, toujours à partir de $Y = y$. Dans la suite de ce paragraphe nous nous intéressons aux champs de Markov couples $Z = (X, Y)$, sachant que les méthodes d'estimation présentées sont également valables pour les modèles triplets.

A titre indicatif, pour un champ à k classes, le vecteur paramètre θ se compose de deux paramètres d'interaction α_i , de k moyennes m_k , k inverse de la variance σ_k^2 et de $k \times k$ paramètres de corrélation q_{ij} , dans le cas gaussien.

III.3.1 Gradient stochastique adapté

L'idée du gradient stochastique adapté est de reprendre le raisonnement du GS décrit en (§ III.2.4) et de l'adapter au modèle de Markov couple (§ II.2 du chapitre II). Nous l'exposons ici dans le cas d'un couple "gaussien" - dans le sens où les lois $p(y|x)$ sont des lois gaussiennes - utilisé par la suite dans les simulations. Considérons l'énergie $U(x, y)$ de la distribution de Gibbs, donnée par l'équation (II.17) :

$$U(x, y) = \frac{1}{2} \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C} \alpha (1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C} q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right] \quad (\text{III.26})$$

Le vecteur paramètre est donc ici $\theta = (\alpha, a_i, m_i, q_{ij})_{i \leq j \leq k}$. Les étapes de l'algorithme restent inchangées, seule la fonction énergie change, modifiant ainsi le gradient. Ce dernier est calculé comme suit:

$$\begin{aligned} \frac{\partial U(x, y)}{\partial a_i} &= 0.5 \sum_{s \in S} (y_s - m_i)^2 1_{[x_s = \omega_i]} \\ \frac{\partial U(x, y)}{\partial \alpha} &= 0.5 \sum_{(s,t) \in C} [1 - 2\delta(x_s, x_t)] \\ \frac{\partial U(x, y)}{\partial q_{ij}} &= \sum_{(s,t) \in C} (y_s - m_i)(y_t - m_j) 1_{[x_s = \omega_i]} 1_{[x_t = \omega_j]} \\ \frac{\partial U(x, y)}{\partial m_i} &= \sum_{s \in S} a_i (m_i - y_i) 1_{[x_s = \omega_i]} + \sum_{(s,t) \in C} q_{ij} (m_j - y_t) 1_{[x_s = i]} 1_{[x_t = \omega_j \neq \omega_i]} \\ &\quad + \sum_{(s,t) \in C} q_{ii} (2m_i - y_s - y_t) 1_{[x_s = \omega_i]} 1_{[x_t = \omega_i]} \end{aligned} \quad (\text{III.27})$$

Les équations ci-dessus, qui représentent donc le gradient de $U(x, y)$ par rapport à θ dans le cas à données complètes, sont utilisées pour le calcul des deux gradients du (III.25).

Remarques

1. L'algorithme GSA peut être également utilisé dans le cas où l'on dispose d'un échantillon d'apprentissage, i.e. lorsque le champ $X = x$ est observé. Dans ce cas, il suffit de remplacer x_{n+1}^* de l'équation (III.25) par le champ observable $X = x$ qui est fixe à chaque itération.

L'équation (III.25) devient:

$$\theta_{n+1} = \theta_n + \frac{K}{n+1} [\nabla_{\theta_n} U(x_{n+1}, y_{n+1}) - \nabla_{\theta_n} U(x, y)] \quad (\text{III.28})$$

2. L'initialisation du GSA peut se faire à partir de différentes méthodes classiques, telles que la méthode des k -means que nous avons utilisée dans nos programmes et que nous décrivons en Annexe, ou la technique de l'histogramme dans le cas de la segmentation d'images. D'autre part, pour le choix du nombre d'itérations, nous avons opté pour le critère de seuil ; i.e., l'arrêt s'effectue lorsque $\theta_{n-1} - \theta_n < \varepsilon$. Par ailleurs, d'après les diverses expérimentations que nous avons effectuées, nous avons constaté que l'algorithme se stabilise, en général, après une quinzaine d'itérations.

III.3.2 Nouvelle méthode

La nouvelle méthode est fondée sur le principe général ECI, dont l'estimation de $\hat{\theta}(X, Y)$ s'effectue par les moindres carrés (estimation des paramètres d'interaction α du champ X), et les lois conditionnelles (estimation des paramètres du bruit à savoir les paramètres de covariance q_{ij} , l'inverse de la variance a_i et les moyennes m_i). Celle-ci est décrite dans ce qui suit:

Estimation des α_i

Pour l'estimation des paramètres d'interaction α_i , nous avons adapté la méthode de Derin et Elliott explicitée plus haut au modèle de Markov couple. Cette méthode se résume en quatre points :

- Trouver la relation entre les probabilités jointes $p(x_s, x_{V_s})$ et les paramètres du modèle ;
- Estimer toutes les probabilités ;
- Construire le système d'équations ;
- Résoudre ce système en utilisant les moindres carrés.

Estimation des $q = [q_{ij}]_{1 \leq i, j \leq k}$, $a = [a_i]_{1 \leq i \leq k}$ **et** $m = [m_i]_{1 \leq i \leq k}$

Considérons une suite de sous ensembles de pixels W_1, \dots, W_l centrés sur les pixels $1, \dots, l$ où chaque sous ensemble W_i est représenté en FIG. 2 (§ III.1.2). X_{W_i} et Y_{W_i} sont les restrictions de X et Y au sous-ensemble W_i , respectivement. Notons $X_{W_i} = [X_i \ X_{V_i}]^T$ et $Y_{W_i} = [Y_i \ Y_{V_i}]^T$, où V_i est un voisinage des 4-ppv. L'idée est de calculer les moyennes m_{x_w} et les matrices de variance-covariance Γ_{x_w} de la distribution de Y_w conditionnellement à $X_w = x_w$ de chacune des configurations x_w (x_w étant la réalisation possible de X_{W_1}, \dots, X_{W_l}). Ainsi, à une configuration x_w donnée, est associée une sous séquence W_1', \dots, W_l' de W_1, \dots, W_l . Les moyennes m_{x_w} et matrices de variance-covariance Γ_{x_w} sont calculées empiriquement comme suit:

$$\hat{m}_{x_w} = \frac{1}{t} \sum_{j=1}^t y_{W_j'} \quad (\text{III.29})$$

$$\hat{\Gamma}_{x_w} = \frac{1}{t} \sum_{j=1}^t (y_{W_j'} - \hat{m}_{x_w})^T (y_{W_j'} - \hat{m}_{x_w}) \quad (\text{III.30})$$

Cette matrice de variance-covariance (Eqs. III.30) est de la forme $\Gamma_{x_w} = \begin{bmatrix} \sigma^2 & A \\ A^T & B \end{bmatrix}$, avec σ^2 la variance de Y_i conditionnelle à x_i .

Par ailleurs, la distribution de Y_i conditionnellement à Y_{V_i} est une distribution Gaussienne de moyennes et de variances données par:

$$M_{x_s} = m_{x_s} + AB^{-1}A^T (y_{V_i} - m_{V_i}) \quad (\text{III.31})$$

$$\Sigma_{x_s}^2 = \sigma^2 - AB^{-1}A^T \quad (\text{III.32})$$

Ainsi, les paramètres a_{x_s} peuvent être calculés à partir de l'équation (III.32), grâce à la relation $a_{x_s} = 1/\Sigma_{x_s}^2$ (voir équation (II.22) du chapitre II).

Cependant, il reste à calculer les paramètres de corrélations $q_{x_s} = (q_{x_s x_{i_1}}, q_{x_s x_{i_2}}, q_{x_s x_{i_3}}, q_{x_s x_{i_4}})$. Ceux-ci sont obtenus par identification des équations (III.31) et (II.21) du chapitre II, ce qui donne:

$$q_{x_s} = -2a_{x_s} AB^{-1} \quad (\text{III.33})$$

Ainsi, à chaque configuration x_w sont associés les paramètres m_i , a_i et q_{ij} . Par conséquent, lorsque les configurations varient de 1 à r , on obtient m_i^1, \dots, m_i^r estimées de m_i , de même pour a_i et q_{ij} . En notant d_1 le nombre de configurations de type 1, ..., d_r le nombre de configurations de type r (on a $d_1 + \dots + d_r = n$), les paramètres m_i , a_i et q_{ij} estimés sont obtenus par les moyennes pondérées suivantes:

$$\hat{m}_i = \frac{1}{n} \sum_{t=1}^r d_t m_i^t, \quad \hat{a}_i = \frac{1}{n} \sum_{t=1}^r d_t a_i^t, \quad \text{et} \quad \hat{q}_{ij} = \frac{1}{n} \sum_{t=1}^r d_t q_{ij}^t \quad (\text{III.34})$$

Finalement, l'estimateur $\hat{\theta}(X, Y)$ à partir des données complètes, utilisé dans l'approche de type ECI que nous avons développée, se résume ainsi:

1. Estimation des α_i par la méthode de Derin et Elliott adaptée ;
2. Pour chaque configuration x_w , estimation de la matrice des moyennes m_{x_w} variance-covariance Γ_{x_w} en utilisant les équations (II.21) et (II.22) ;
3. Pour chaque configuration x_w , calcul de a_{x_s} et de $q_{x_s} = (q_{x_s x_{i_1}}, q_{x_s x_{i_2}}, q_{x_s x_{i_3}}, q_{x_s x_{i_4}})$, à partir des équations (III.32) et (III.33) en leur attribuant un poids d_i , obtenu par les fréquences ;
4. Calcul des a_i , m_i et q_{ij} finaux, en utilisant la moyenne pondérée (équation III.34).

En intégrant l'approche sus développée dans ECI, on obtient l'algorithme suivant:

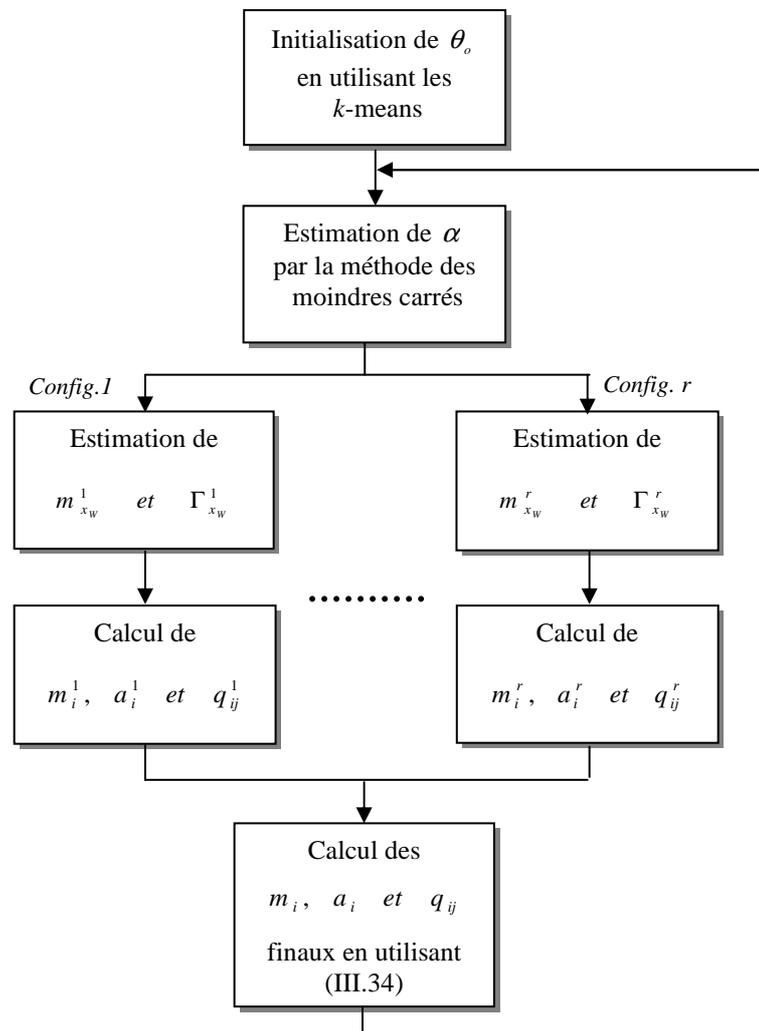


FIG. 3 – Algorithme de la nouvelle méthode (NM)

Remarque

Lorsque le nombre de classes devient important, la prise en compte de toutes les configurations x_w devient impossible. Dans ce cas, il est possible d'envisager de ne prendre en compte que certains types de configurations, comme par exemple celles contenant au plus deux classes.

III. 3. 3 Analyse de performances

Nous proposons dans cette sous-section d'analyser les performances des algorithmes GSA et NM présentés ci-dessus et ce, en testant, d'une part les performances de chacun des algorithmes séparément et d'autre part, en effectuant une comparaison entre eux. En particulier, on teste la résistance des deux méthodes au bruit, en considérant des cas de plus en plus bruités. Cette résistance est mesurée par la différence entre l'estimation à partir des données complètes (DC) et des données incomplètes (DI).

Pour ce faire, nous simulons d'abord un champ de Markov couple à deux classes dont la fonction énergie est donnée comme suit:

$$U(x, y) = \frac{1}{2} \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C^H} \alpha_H (1 - 2\delta(x_s, x_t)) \right. \\ \left. \sum_{(s,t) \in C^V} \alpha_V (1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C} q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right] \quad (\text{III.35})$$

avec C^H l'ensemble des cliques horizontales et C^V l'ensemble des cliques verticales (c'est une expression légèrement plus générale que celle donnée dans l'équation (II.17)). Ensuite, les paramètres sont estimés à partir des deux méthodes décrites plus haut, de deux manières différentes. La première est de considérer le cas de données complètes, i.e. en supposant que le champ de classes X est observable. La deuxième consiste en l'estimation à partir de données incomplètes.

Nous présentons ci-après deux séries d'expérimentations. La première série, qui correspond au couple dont la réalisation est représentée en FIG. 4, est obtenue à partir de paramètres d'interactions $\alpha_H = \alpha_V = 2$, de moyennes $m_1 = 0$ et $m_2 = 2$, d'inverse de variances $a_1 = a_2 = 1$ et de corrélations $q_{11} = q_{22} = -0.2$ ($q_{12} = q_{21} = 0$).

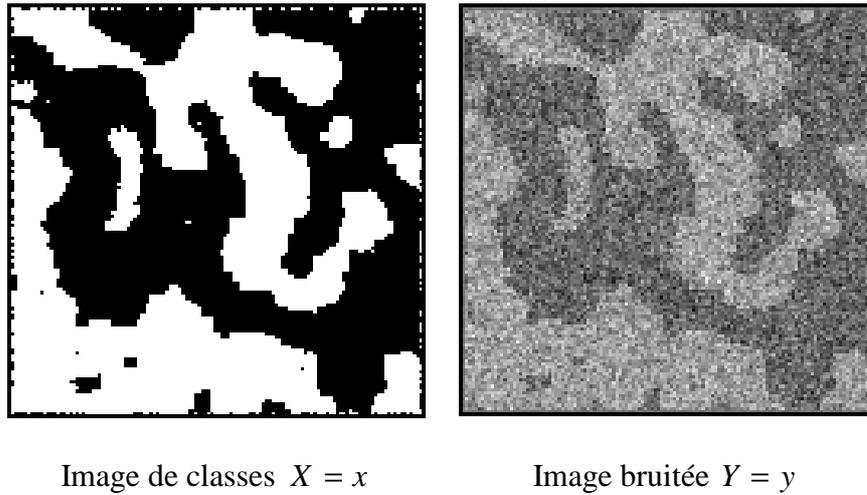


FIG. 4 – Champ de Markov couple

TAB. 1 Résultats correspondant à FIG. 4

Paramètres	Valeurs réelles	Nouvelle méthode		Méthode du GS adaptée	
		NMDC	NMDI	GSADC	GSADI
α_H	2.00	2.10	1.87	2.36	2.24
α_V	2.00	2.01	1.68	2.37	2.24
a_1	1.00	1.00	1.06	0.91	0.58
a_2	1.00	1.03	1.09	1.00	1.03
m_1	0.00	0.00	-0.01	0.10	0.26
m_2	2.00	2.00	2.07	1.86	0.78
q_{11}	-0.20	-0.19	-0.19	-0.20	-0.06
q_{22}	-0.20	-0.20	-0.19	-0.21	-0.40
q_{12}	0.00	0.00	0.00	0.02	0.00
tps (sec.)		<1	2	2	6

NMDC et GSADC: Estimés à partir de données complètes, NMDI et GSADI: estimés à partir de données incomplètes, tps: temps d'exécution.

La deuxième série concerne un bruit plus élevé et plus corrélé. Les paramètres utilisés sont $\alpha_H = \alpha_V = 2$, $m_1 = 0$ et $m_2 = 1$, $a_1 = a_2 = 1$ et $q_{11} = q_{22} = -0.4$ ($q_{12} = q_{21} = 0$), et une réalisation du couple est présentée sur la figure FIG. 5.

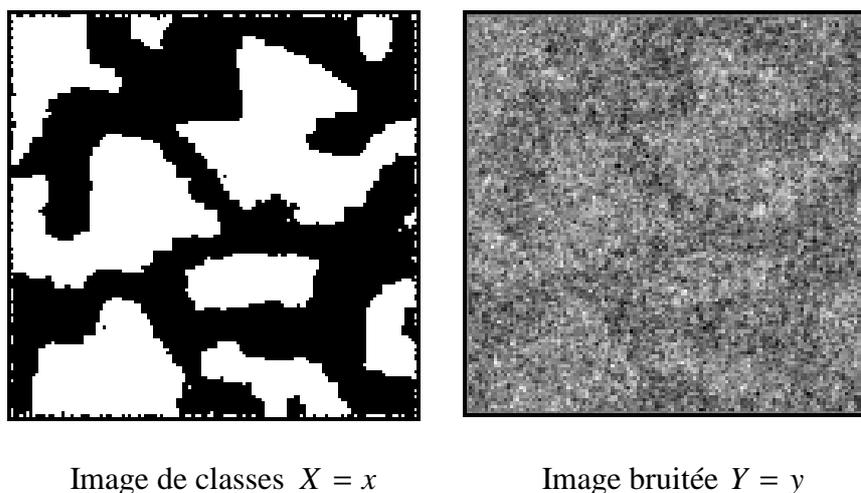


FIG. 5 – *Champ de Markov couple*

TAB. 2 *Résultats correspondant à FIG. 5*

Paramètres	Valeurs réelles	Nouvelle méthode		Méthode du GS adaptée	
		NMDC	NMDI	GSADC	GSADI
α_H	2.00	2.11	1.11	2.36	1.75
α_V	2.00	2.02	1.03	2.35	1.75
a_1	1.00	0.96	1.11	0.91	0.91
a_2	1.00	0.98	1.13	0.86	0.83
m_1	0.00	0.01	0.00	0.01	0.20
m_2	1.00	0.97	1.14	1.08	0.40
q_{11}	-0.40	-0.37	-0.41	-0.32	-0.41
q_{22}	-0.40	-0.38	-0.41	-0.32	-0.30
q_{12}	0.00	0.00	0.00	-0.01	0.00
Tps (sec.)		<1	2	2	6

NMDC et GSADC: Estimés à partir de données complètes, NMDI et GSADI: estimés à partir de données incomplètes, tps: temps d'exécution.

Les résultats des différentes valeurs estimées sont présentés dans le tableau TAB. 1 pour la première simulation, et dans le tableau TAB. 2 pour la deuxième.

Ces résultats, ainsi que les résultats obtenus à partir des simulations analogues dont nous ne présentons pas les résultats ici, nous permettent de tirer les conclusions suivantes. Les estimées obtenues par la nouvelle méthode à données complètes (NMDC) et par la méthode du gradient stochastique adapté à données complètes (GSADC), sont assez proches des paramètres réels, et cela reste vrai pour différents niveaux de bruitage (voir TAB. 2). Par ailleurs, dans le cas des données incomplètes, la NMDI se détache de la méthode GSADI fournissant ainsi de meilleurs résultats. La différence est particulièrement visible dans les paramètres liés au bruit, à savoir les moyennes, variances et paramètres de corrélations. Aussi, nous constatons que dans tous les cas (données complètes et incomplètes) la nouvelle méthode est plus rapide que le gradient stochastique adapté. Le temps d'exécution du programme est inférieur à 1 sec. pour la NMDC, de 2 sec. pour NMDI, 2 sec. pour GSADI et 6 sec. pour GSADI. Ainsi, dans le cas non supervisé, la NM est 3 fois plus rapide que le GS. Enfin, 15 itérations ont été effectuées dans le cas du GSADC et 30 pour le GSADI. Pour la NMDI, le nombre d'itérations est fixé à 30.

En conclusion, la NM semble plus intéressante et sera retenue pour la suite de cette étude.

Remarque

La convergence de la méthode du GSA (*resp.* GS) dépend fortement du choix de la constante K . Celle-ci est souvent choisie égale à $1/ Card(S)$ dans le cas des champs de Markov cachés, à bruit indépendant, pour l'estimation des paramètres d'interaction α . Cependant, en prenant cette valeur pour l'estimation des paramètres du modèle de Markov couple considéré, nous avons constaté que l'algorithme divergeait, surtout lors de la mise à jour des paramètres du bruit (moyennes, variances, paramètres de corrélation). Une des raisons possibles pourrait être due aux dérivées qui sont fonctions des moyennes m_i et des y_i (équation III.27), ce qui n'est pas le cas pour la dérivée de cette même fonction d'énergie par rapport à α . Ceci nous a contraint à choisir une valeur plus petite pour cette constante K . Expérimentalement, nous avons optimisé ce choix, et pour les deux séries d'expériences ci-dessus, $K \approx (1/ Card(S))10^{-1}$.

III.4 Conclusion

Dans ce chapitre nous avons commencé par rappeler les principales méthodes d'estimation des paramètres des champs markoviens cachés classiques, aussi bien dans le cas des données complètes que dans le cas des données incomplètes. Nous avons ensuite exposé deux nouvelles méthodes d'apprentissage des champs de Markov couple et triplet, que nous avons récemment développées [BeP 04, BeP 05]. La première est une adaptation aux modèles couples de la méthode de gradient stochastique de Younes [You 89]. La deuxième, de type "*Estimation Conditionnelle Itérative*", utilise la méthode de Derin *et al.* pour les paramètres relatifs au champ caché, et une démarche originale pour les paramètres du bruit. Une première étude comparative entre ces deux méthodes, par le biais d'images simulées, a été effectuée et deux exemples de résultats d'expériences ont été présentés. Cette étude montre que les deux algorithmes se comportent approximativement de la même façon, fournissant ainsi d'assez bons résultats, dans le cas des données complètes. Cependant, dans le cas des données incomplètes, la méthode ECI se détache de la méthode du gradient stochastique, donnant ainsi de meilleurs résultats d'estimation, surtout lorsqu'il s'agit d'un bruitage important. Aussi, l'algorithme ECI est trois fois plus rapide que l'algorithme du gradient stochastique.

Nous allons étudier dans la suite diverses modélisations et applications à la segmentation non supervisée d'images. Dans toute la suite, les paramètres sont estimés par la nouvelle méthode ECI de ce chapitre.

Chapitre IV

Simulations et segmentation non supervisée d'images

Dans les chapitres précédents, trois modèles markoviens de complexité croissante et diverses méthodes d'estimation des paramètres ont été présentés. Nous consacrons ce chapitre à l'application de ces différentes modélisations et techniques d'estimation des paramètres à la segmentation statistique non supervisée d'images. Afin de valider les nouvelles modélisations et techniques, nous considérerons classiquement des images simulées, des images de synthèse et des images réelles.

IV.1 Segmentation couple et triplet

Le but de cette section est de comparer, d'une part, les différents algorithmes d'estimation des paramètres et d'autre part, la segmentation non supervisée MPM fondée sur le modèle de Markov couple à celle fondée sur le modèle de Markov triplet. Pour ce faire, trois séries d'expérimentations ont été effectuées. Classiquement, la première série concerne des images simulées, la seconde des images de synthèse, quant à la troisième, elle a trait aux images réelles.

Les fonctions énergies associées aux modèles de Markov couple et triplet utilisés au cours de ces expérimentations sont données, respectivement, comme suit:

$$\begin{aligned}
 U(x, y) = 0.5 & \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C^H} \alpha_H (1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C^V} \alpha_V (1 - 2\delta(x_s, x_t)) \right. \\
 & \left. + \sum_{(s,t) \in C} q_{x_s, x_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right]
 \end{aligned} \tag{IV.1}$$

$$\begin{aligned}
 U(w, y) = 0.5 & \left[\sum_{s \in S} a_{w_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C^H} \alpha_H (1 - 2\delta(w_s, w_t)) + \sum_{(s,t) \in C^V} \alpha_V (1 - 2\delta(w_s, w_t)) \right. \\
 & \left. + \sum_{(s,t) \in C} q_{w_s, w_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right]
 \end{aligned} \tag{IV.2}$$

où $w = (x, u)$, C^H l'ensemble des cliques horizontales, C^V l'ensemble des cliques verticales et δ la fonction de Dirac, vérifiant:

$$\delta(x_s, x_t) = \begin{cases} 1 & \text{si } x_s = x_t \\ 0 & \text{si } x_s \neq x_t \end{cases} \tag{IV.3}$$

IV.1.1 Simulations

Les images de cette sous-section ont été simulées en utilisant la fonction énergie donnée par l'équation (IV.1) et segmentées par la méthode MPM décrite dans le chapitre I. Notre premier objectif consiste à comparer les deux méthodes d'estimation des paramètres à savoir: la nouvelle méthode (NM) et la méthode du gradient stochastique adapté (GSA), étudiées dans le chapitre III, dans le cadre de la segmentation non supervisée. Pour une image, à deux classes, considérée, les paramètres sont regroupés dans un vecteur $\theta = (\alpha_H, \alpha_V, a_1, a_2, m_1, m_2, q_{11}, q_{22}, q_{12})$.

Dans la première série d'expérimentations, les simulations correspondent exactement au modèle théorique (CMCouple) étudié. Ainsi les paramètres du modèle utilisé peuvent être contrôlés et comparés aux valeurs estimées obtenues, à partir de la NM et le GSA, dans le cas des DC et DI. Les résultats obtenus sont exposés dans FIG. 1 et FIG. 2. Les paramètres et leurs valeurs estimées à partir des données complètes (DC) et incomplètes (DI), sont présentés

dans TAB. 1 et TAB. 2. Le premier exemple est relativement peu bruité, alors que dans le deuxième le niveau de bruit a été augmenté de manière à départager les deux méthodes.

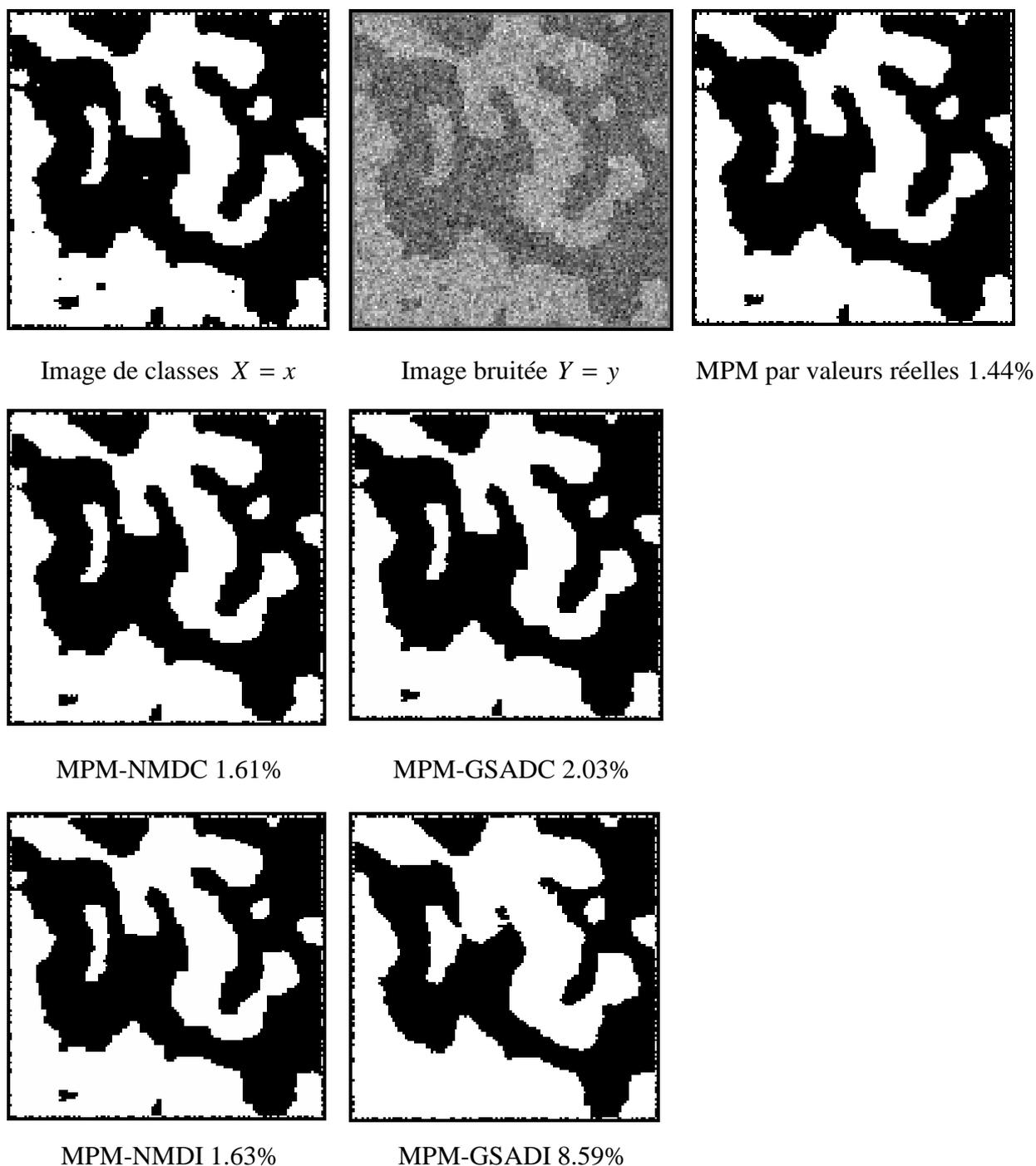


FIG. 1 – Simulation d'un CMCouple à deux classes et comparaison entre la segmentation MPM utilisant le GSA et la NM pour l'estimation des paramètres.

TAB. 1 Résultats correspondant à FIG. 1

Paramètres	Valeurs réelles	Nouvelle méthode		Méthode du GS adaptée	
		NMDC	NMDI	GSADC	GSADI
α_H	2.00	2.10	1.87	2.36	2.24
α_V	2.00	2.01	1.68	2.37	2.24
a_1	1.00	1.00	1.06	0.91	0.58
a_2	1.00	1.03	1.09	1.00	1.03
m_1	0.00	0.00	-0.01	0.10	0.26
m_2	2.00	2.00	2.07	1.86	0.78
q_{11}	-0.20	-0.19	-0.19	-0.20	-0.06
q_{22}	-0.20	-0.20	-0.19	-0.21	-0.40
q_{12}	0.00	0.00	0.00	0.02	0.00
TE	1.44%	1.61%	1.63%	2.03%	8.59%

NMDC et GSADC: Estimés à partir de données complètes, NMDI et GSADI: estimés à partir de données incomplètes, TE: taux d'erreur

TAB. 2 Résultats correspondant à FIG. 2

Paramètres	Valeurs réelles	Nouvelle méthode		Méthode du GS adaptée	
		NMDC	NMDI	GSADC	GSADI
α_H	2.00	2.11	1.11	2.36	1.75
α_V	2.00	2.02	1.03	2.35	1.75
a_1	1.00	0.96	1.11	0.91	0.91
a_2	1.00	0.98	1.13	0.86	0.83
m_1	0.00	0.01	0.00	0.01	0.20
m_2	1.00	0.97	1.14	1.08	0.40
q_{11}	-0.40	-0.37	-0.41	-0.32	-0.41
q_{22}	-0.40	-0.38	-0.41	-0.32	-0.30
q_{12}	0.00	0.00	0.00	-0.01	0.00
TE	11.53%	11.65%	12.23%	12.40%	48.67%

NMDC et GSADC: Estimés à partir de données complètes, NMDI et GSADI: estimés à partir de données incomplètes, TE: taux d'erreur

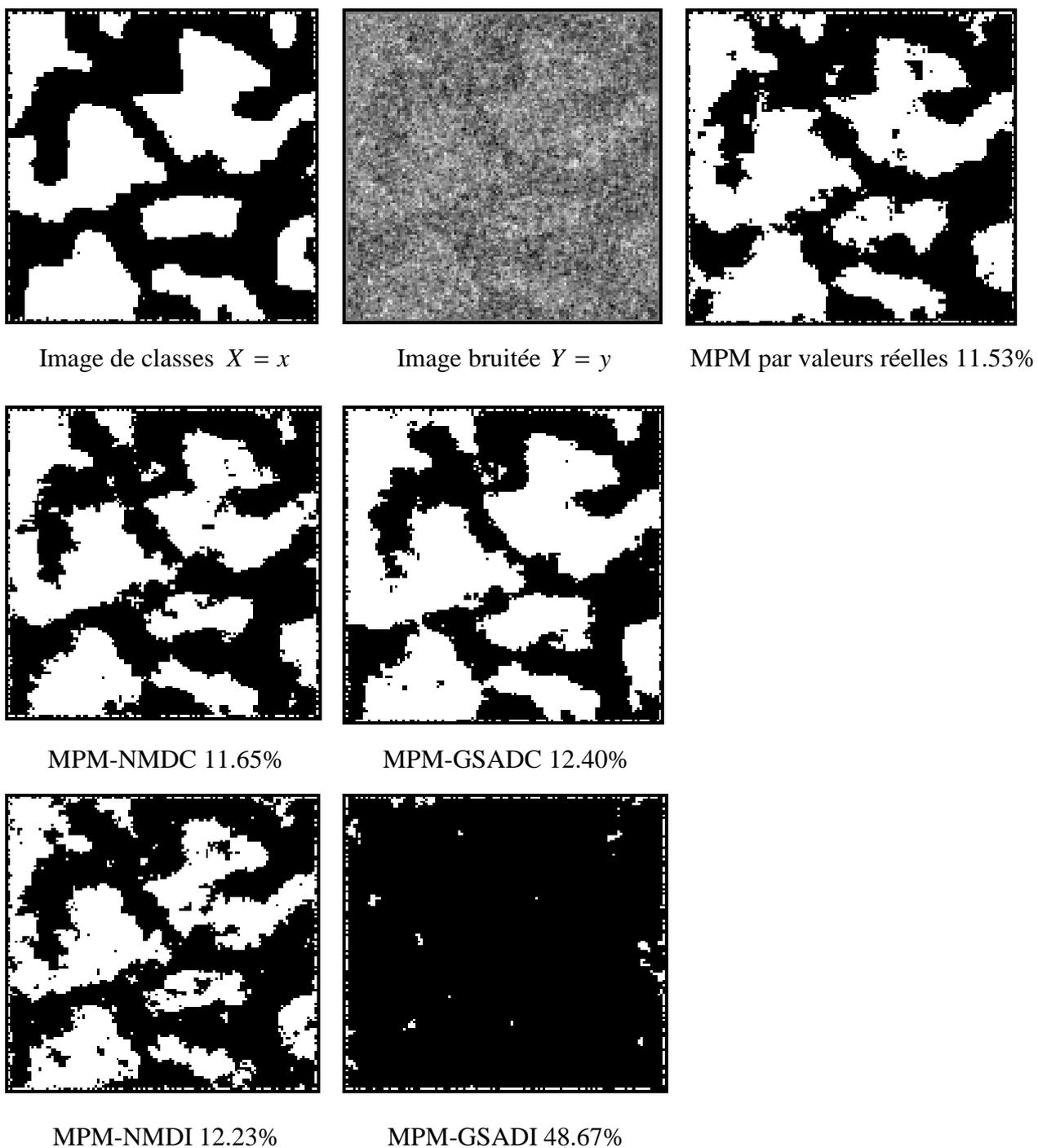


FIG. 2 – Simulation d'un CMCouple à deux classes et comparaison entre la segmentation MPM utilisant le GSA et la NM pour l'estimation des paramètres.

Nous pouvons conclure, à partir des diverses simulations effectuées, dont deux présentées ci-dessus, que dans le cas à données complètes (DC), les taux d'erreur de la segmentation MPM obtenus en utilisant les paramètres estimés par la nouvelle méthode (NM) et le gradient

stochastique adapté (GSA), sont toujours proches du résultat obtenu, à partir des valeurs réelles. Cependant, dans le cas des données incomplètes (DI), le taux d'erreur de la segmentation MPM obtenu à partir des paramètres estimés par la NMDI est systématiquement inférieur à celui obtenu en estimant les paramètres du modèle par le GSADI. La différence augmente avec le niveau de bruitage et le GSADI peut diverger, comme montré dans le deuxième exemple présenté à la figure FIG. 2.

Remarque

Le nombre d'itérations de l'algorithme MPM a été fixé expérimentalement à 100, dont chacune contient 20 itérations dans l'échantillonneur de Gibbs utilisé. Pour les deux méthodes d'estimation (GSADI et NMDI), le nombre d'itérations utilisé est de 30.

IV.1.2 Image de synthèse

Nous souhaitons tester la robustesse des mêmes algorithmes sur un modèle proche du modèle de Markov couple, dans le sens où les (Y_s) sont gaussiennes et corrélées conditionnellement à X ; cependant, le modèle utilisé ne correspond pas exactement au modèle couple théorique. Ce modèle est obtenu en prenant une image de synthèse, "image ring" à deux classes, par exemple, qu'on bruitage par un bruit blanc gaussien de paramètres $N(0,1)$ et $N(0.5,1)$ associés à chacune des deux classes. L'image ainsi obtenue, notée $W = (W_s)_{s \in S}$, est ensuite moyennée sur les 4-ppv comme suit:

$$Y_s = \frac{1}{5} (W_s + \sum_{t \in V_s} W_t) \quad (\text{IV.4})$$

Bien sûr, un tel modèle ne correspond pas aux caractéristiques du modèle de Markov couple étudié précédemment ; en particulier, le bruit obtenu par (IV.4) est à "portée finie", alors que dans un modèle de Markov couple, le bruit est markovien, donc à portée infinie. C'est un modèle "intermédiaire" entre le modèle théorique (CMCouple) et pratique (images réelles). L'étude des cas intermédiaires de ce type présente deux avantages : on maîtrise le niveau du bruit, ce qui permet de départager les différentes méthodes en augmentant le niveau, et on dispose de la vérité terrain $X = x$.



Image de classes $X = x$

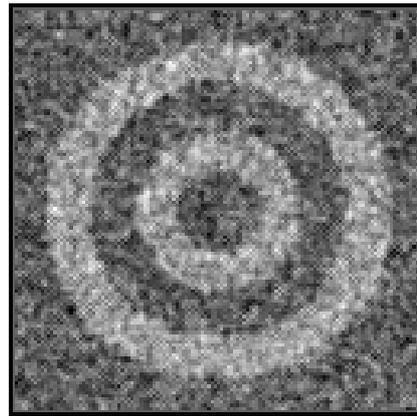
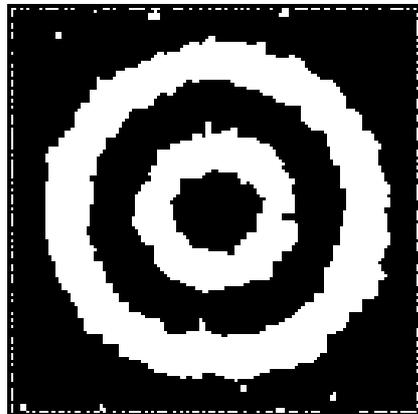
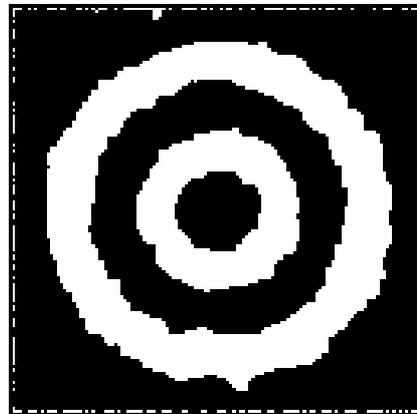


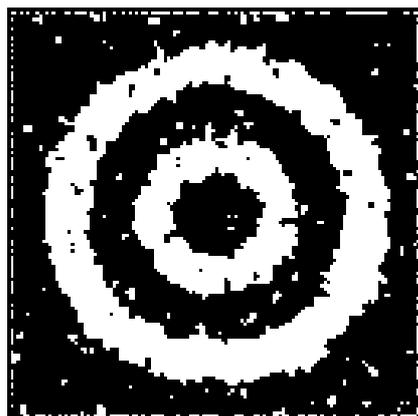
Image bruitée $Y = y$



MPM-NMDC 3.10%



MPM-GSADC 3.70%



MPM-NMDI 5.50%



MPM-GSADI 15.30%

FIG. 3 – Image "ring", sa version bruitée et différentes méthodes de segmentation MPM.

TAB. 3 Résultats correspondant à FIG. 3

Paramètres	Nouvelle méthode		Méthode du GS adaptée	
	NMDC	NMDI	GSADC	GSADI
α_H	2.01	1.17	2.38	2.30
α_V	2.01	1.19	2.38	2.30
a_1	7.37	8.19	4.88	2.04
a_2	7.74	8.23	3.43	2.05
m_1	0.01	-0.05	0.01	0.02
m_2	0.99	1.04	0.98	0.60
q_{11}	-0.50	-0.50	-0.35	0.00
q_{22}	-0.48	-0.45	-0.23	-0.50
q_{12}	0.00	0.00	-0.01	0.00
TE	3.10%	5.50%	3.70%	15.30%

NMDC et GSADC: Estimés à partir de données complètes, NMDI et GSADI: estimés à partir de données incomplètes, TE: taux d'erreur

Plusieurs tests ont été effectués sur différentes images de synthèse et une série de résultats est donnée, ci-dessus. Les conclusions générales sont similaires à celles des simulations de la section précédente. Ainsi, les deux algorithmes se comportent de la même façon dans le cas à données complètes. Cependant, dans le cas à données incomplètes, les résultats de segmentation MPM obtenus en estimant les paramètres par la NMDI, sont meilleurs que ceux fournis en utilisant le GSADI. Pour cette raison, nous utiliserons la nouvelle méthode (NM), dans la suite de notre étude.

IV.1.3 Image réelle

Dans cette sous-section, nous comparons la segmentation MPM non supervisée fondée sur le modèle de Markov couple (CMCouple) à celle fondée sur le modèle de Markov triplet (CMT), sur une image réelle. Nous considérons une image radar, représentant une partie de la ville d'Istres (voir FIG. 4), constituée de 4 classes: ("*eau*", "*foret et culture*", "*champ de maïs*" et "*maisons*") que nous segmentons en deux classes "*eau*" et "*terre*", cette dernière classe étant l'union des trois classes ("*foret et culture*", "*champ de maïs*" et "*maisons*"). Ainsi,

l'exemple considéré fait partie de l'une des interprétations du processus auxiliaire où les différentes valeurs de U_s correspondent aux sous-classes des classes "eau" et "terre" (nous sommes devant un cas où l'existence des sous-classes ne concerne que l'une de deux classes ; à savoir la classe "terre").

Dans nos expérimentations, nous avons fixé le paramètre de corrélation $q_{ij} = 0$ pour $i \neq j$. Par ailleurs, le nombre d'itérations est fixé à 30 pour la NM et à 100 itérations pour la segmentation MPM, chacune contenant 20 itérations dans l'échantillonneur de Gibbs utilisé.

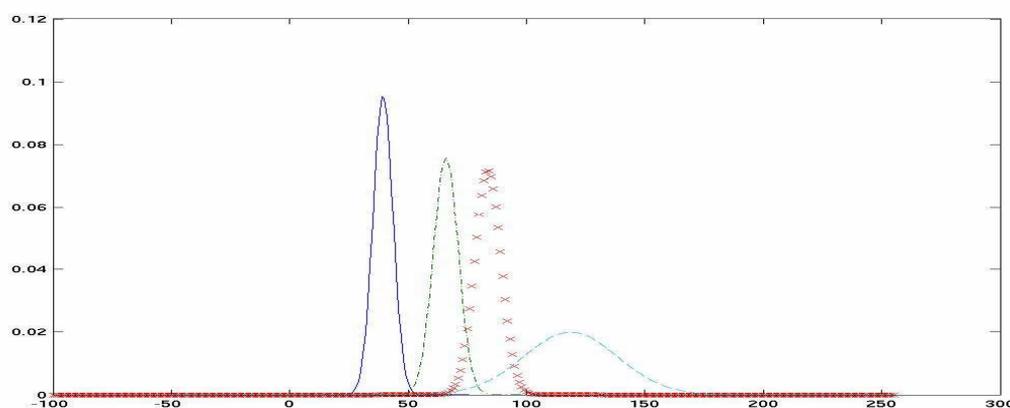


FIG.4 – Représentation des quatre distributions gaussiennes associées aux 4 classes ("eau", "foret et culture", "champ de maïs" et "maisons", respectivement) de l'image de la "ville d'Istres" représentée dans FIG. 5 estimés par la NMDI, où $\alpha = (\alpha_H, \alpha_V) = (0.93, 0.89)$ et q_{x_s, x_t} égales à $-0.12, -0.09, -0.07$ et -0.02 .

Les quatre densités gaussiennes estimées par la méthode d'estimation "NMDI" de la section précédente sont présentées sur la figure FIG.4, qui contient également, dans la légende, les autres estimées. La segmentation MPM non supervisée en deux classes "eau" et "terre", par un CMCouple (IV.1), donne un taux d'erreur de 21.99%, alors que l'utilisation du modèle de Markov triplet (IV.2) fournit un taux d'erreur de 6.66%. Le résultat de segmentation montre ainsi que le modèle de Markov triplet peut permettre, en situations réelles, d'améliorer la segmentation MPM fondée sur les CMCouples.

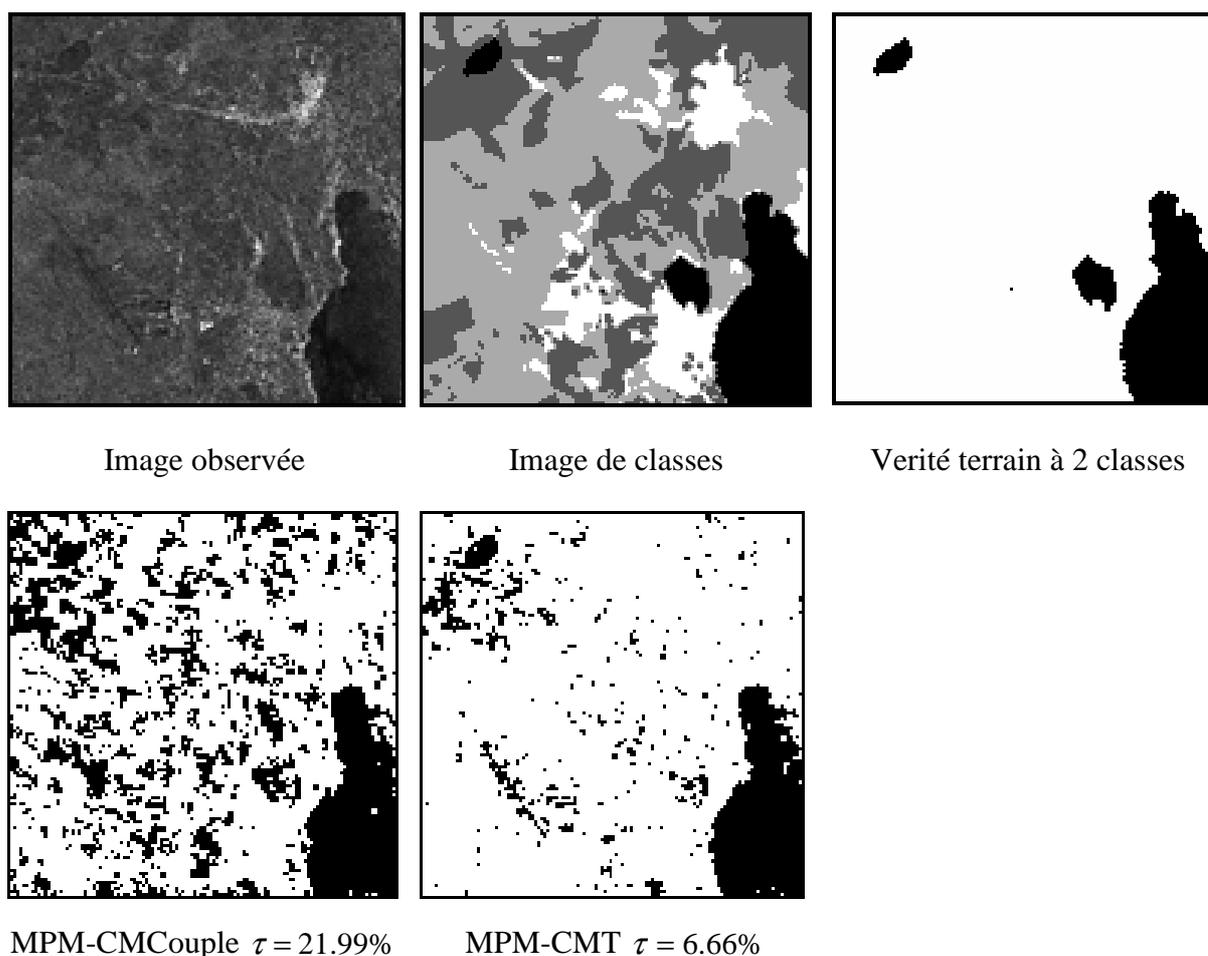


FIG. 5 – Images représentant une partie de la ville "d'Istres", sa vérité terrain à deux classes et les résultats de la segmentation MPM basée sur le CMCouple et CMT. τ : taux d'erreur en prenant comme référence la vérité terrain à 2 classes.

IV.2 Segmentation évidentielle

Nous nous proposons dans cette partie d'étudier l'apport de la fusion de Dempster-Shafer dans certaines situations où les champs de Markov cachés classiques apparaissent comme des modélisations trop simples. L'idée est de modéliser le fait que l'image des classes est très peu homogène, ce qui rend sa modélisation par un champ de Markov stationnaire difficile. A cet effet, nous utilisons le modèle CME, détaillé dans le chapitre II, obtenu par affaiblissement de la loi *a priori* et dont le résultat de la fusion aboutit à un champ de Markov triplet.

La fonction énergie que nous allons utiliser est donnée par:

$$U(w, y) = \sum_{i=1}^3 \left(\sum_{(s,t) \in C^H} \alpha_1^i (1 - 2\delta(u_s, u_t)) + \sum_{(s,t) \in C^V} \alpha_2^i (1 - 2\delta(u_s, u_t)) \right) - \sum_{s \in S} \log(p(y_s | x_s)) \quad (\text{IV.5})$$

La segmentation s'effectue d'une manière non supervisée, dans laquelle les paramètres sont estimés par la méthode NMDI. Les résultats ainsi obtenus sont comparés à ceux obtenus en utilisant la modélisation classique (CMC-BI).

Considérons une image "oiseau" non stationnaire à deux classes, illustrée en FIG. 6, et sa version bruitée, avec un bruit blanc gaussien de variance 1 et de moyennes, respectivement, de 0 et 2. L'utilisation de la méthode bayésienne MPM fondée sur le modèle triplet correspondant au modèle markovien affaibli, dont la fonction énergie est donnée par l'équation (IV.5), fournit 6.70% de taux d'erreur, alors que celle fondée sur le modèle CMC classique (II.5) donne 12.80% de taux d'erreur. Dans les deux cas, la segmentation est non supervisée. Les paramètres ont été initialisés en utilisant la méthode des k -means et estimés par la méthode NMDI. Les paramètres ainsi obtenus sont représentés dans TAB. 4.

TAB. 4 Résultats correspondant à FIG. 6

Paramètres	CMC	CME
μ_1	0.88	0.02
μ_2	2.01	2.02
σ_1	1.36	0.94
σ_2	0.99	0.67
α_H^1, α_V^1		0.25, 0.31
α_H^2, α_V^2	0.65, 0.82	0.38, 0.44
α_H^3, α_V^3		0.01, 0.01

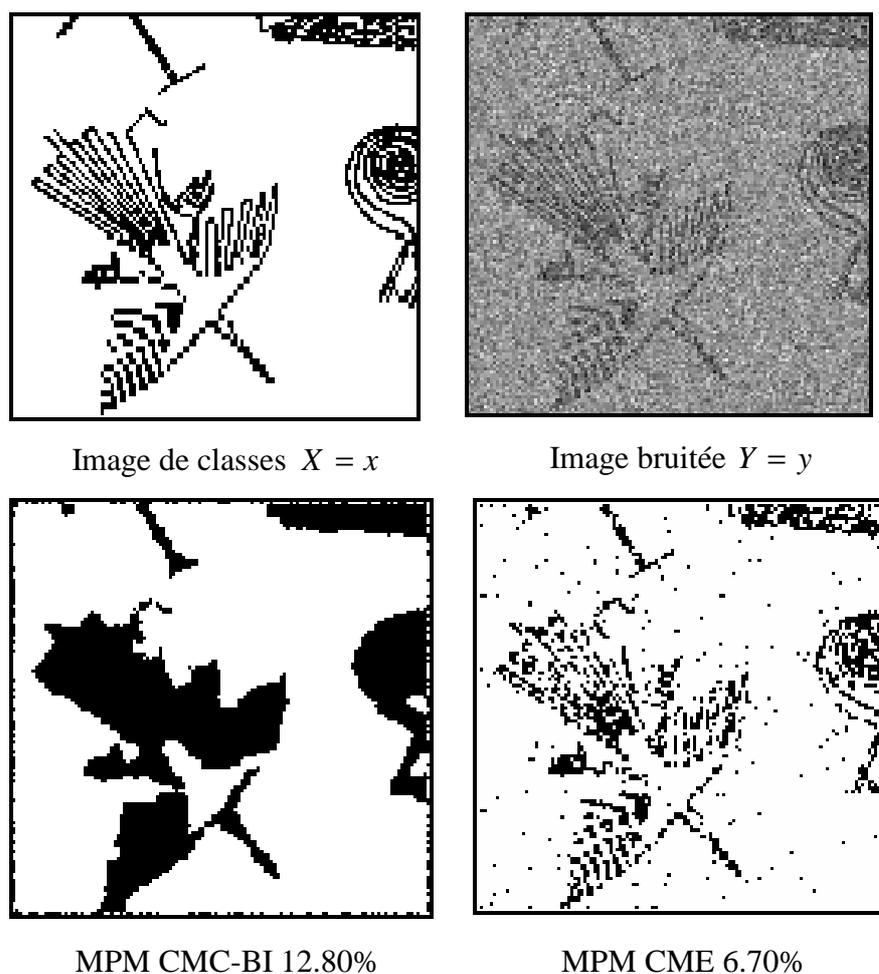


FIG. 6 – Image "oiseau", sa version bruitée et résultats de la segmentation MPM non supervisée.

Les résultats de TAB. 4 montrent que les paramètres du bruit estimés à partir du modèle évidentiel, sont plus proches des paramètres réels que ceux estimés à partir des CMC-BI classiques. Par ailleurs, aucune conclusion ne peut être tirée quant à l'estimation des paramètres d'interaction α_i , étant donné que leurs vraies valeurs sont inconnues. Au niveau de la segmentation, le point important est que le nouveau modèle permet de retrouver certains détails (ailes et queue de l'oiseau), ce que le modèle classique ne permet pas.

La même démarche que celle suivie pour la segmentation de l'image "oiseau" est effectuée pour l'image "mosaïque", qui est une image plus complexe, représentée dans FIG. 7. Ainsi, l'image "mosaïque" est bruitée avec un bruit blanc gaussien de variance 1 et de moyennes, respectivement, de 0 et 2. Ensuite, l'image bruitée est segmentée d'une manière

non supervisée en utilisant, d'une part, la segmentation MPM fondée sur le modèle CMC-BI et d'autre part, la segmentation MPM fondée sur le modèle CME. Les résultats ainsi obtenus sont présentés dans le tableau TAB. 5:

TAB. 5 Résultats correspondant à FIG. 7

Paramètres	CMC	CME
μ_1	-0.3	0.02
μ_2	1.86	2.03
σ_1	0.85	1.00
σ_2	1.05	0.97
α_H^1, α_V^1		0.29, 0.33
α_H^2, α_V^2	0.41, 0.5	0.47, 0.55
α_H^3, α_V^3		0.02, 0.02

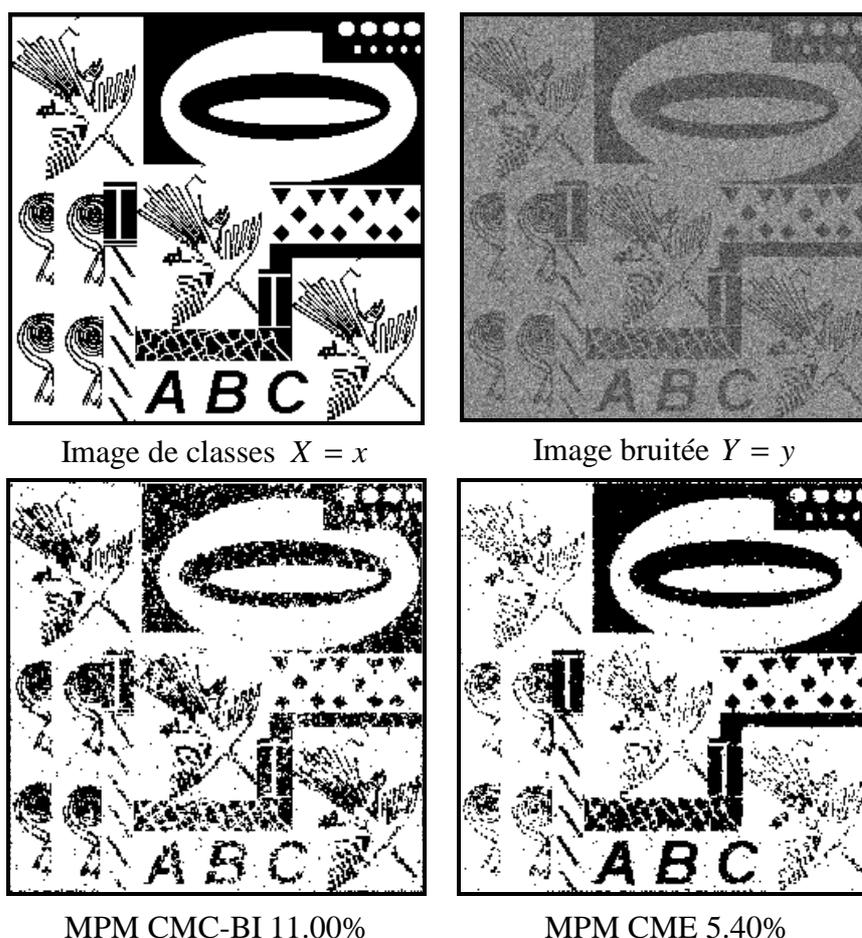


FIG. 7 – Image "mosaïque", sa version bruitée et résultats de la segmentation MPM non supervisée fondée sur les modèles CMC-BI et CME.

TAB. 5 illustre les valeurs des paramètres estimés. Là aussi, les valeurs estimées des paramètres du bruit à partir du modèle évidentiel sont plus proches des valeurs réelles que celles estimées à partir des CMC classiques. On a pu ainsi améliorer le résultat de segmentation non supervisée en utilisant les CME.

IV.3 Segmentation triplet adaptée

Bien que les CME modélisent bien les images non stationnaires, ils présentent cependant la particularité de ne pas pouvoir modéliser les différentes stationnarités présentes dans l'image.

L'objet de cette section est de présenter des résultats d'applications des CMT "*adaptés*" présentés dans § II.5, au chapitre II. L'objectif est de montrer leur intérêt dans la segmentation non supervisée des images non stationnaires et comparer leur efficacité à celle des CMC-BI. Ainsi, plusieurs expérimentations, sur des images de différentes natures, ont été effectuées dont un échantillon est présenté dans ce qui suit.

Dans toutes les expérimentations de cette sous-section 20 itérations ont été effectuées pour les simulations des images (échantillonneur de Gibbs), 20 itérations pour la NMDI (estimation des paramètres), et 100 itérations pour la segmentation MPM, contenant chacune 20 itérations dans l'échantillonneur de Gibbs utilisé.

IV.3.1 Champ de Markov triplet adapté à bruit indépendant

Considérons un champ de Markov X non stationnaire, d'énergie $U(x) = \sum_c \varphi_c(x_c)$ tel que φ_c dépend de la position de c dans S . Supposons qu'il y ait m stationnarités différentes. Nous modélisons cette situation en introduisant $u = (u_s)_{s \in S}$, avec $u_s \in \Lambda$ ($\Lambda = \{\lambda_1, \dots, \lambda_m\}$) et en considérant $\Phi_c(x_c, u_c) = \varphi_c(x_c)$, avec Φ_c ne dépendant pas de la position de c dans S . Considérons, comme application, le CMTA-BI (BI pour "*bruit indépendant*") dont la fonction énergie est donnée par:

$$\begin{aligned}
 U(x, u, y) = & \sum_{(s,t) \in C^H} \alpha_H^1 (1 - 2\delta(x_s, x_t)) + (\alpha_{H\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots \\
 & + \alpha_{H\lambda_M}^2 \delta(u_s, u_t, \lambda_M))(1 - \delta(x_s, x_t)) \\
 & + \sum_{(s,t) \in C^V} \alpha_V^1 (1 - 2\delta(x_s, x_t)) + (\alpha_{V\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots \\
 & + \alpha_{V\lambda_M}^2 \delta(u_s, u_t, \lambda_M))(1 - \delta(x_s, x_t)) - \sum_{s \in S} \log(p(y_s | x_s))
 \end{aligned} \tag{IV.6}$$

Nous voyons que la loi $p(x, y|u)$ est la loi d'un champ de Markov caché classique non stationnaire. Ainsi, la non stationnarité de (X, Y) est modélisée par un TMC "adapté" $T = (X, U, Y)$. On peut alors estimer les paramètres du TMC "adapté" et l'utiliser pour effectuer une segmentation non supervisée.

Afin d'étudier l'efficacité des CMTA-BI, nous proposons une série de tests sur trois types d'images. Le premier type concerne les images simulées où l'on dispose de la vérité terrain et de tous les paramètres. Le second a trait aux images de synthèses où l'on dispose de la connaissance partielle (vérité terrain, paramètres du bruit). Enfin, le dernier concerne les images réelles. Ces trois cas sont étudiés dans les trois sous-sections ci-dessous.

IV.3.1.1 Simulations

Cette première série de tests concerne des images simulées en utilisant les champs de Markov triplet adaptés, à bruit indépendant (CMTA-BI). Pour cela, nous avons simulé par l'échantillonneur de Gibbs, trois CMTA-BI dont la forme générale de l'énergie est donnée par l'équation (IV.6). Ensuite, nous avons estimé les paramètres du modèle par les méthodes NMDC et NMDI,s et enfin nous avons effectué une segmentation MPM basée sur les CMTA-BI. Les résultats ainsi obtenus sont alors comparés à ceux obtenus en utilisant la modélisation classique (CMC-BI).

Les résultats de la première simulation, présentés à la figure FIG. 8, montrent que les taux d'erreurs fournis par la segmentation MPM fondée sur les CMTA-BI, dans le cas supervisé, sont approximativement identiques à ceux obtenus dans le cas non supervisé (DI).

En comparant les deux méthodes de segmentation (MPM CMTA-BI et MPM CMC-BI), nous constatons que les résultats obtenus par la segmentation MPM fondée sur les CMTA-BI sont meilleurs que ceux fournis par la segmentation MPM, fondée sur les CMC-BI. Aussi, les CMTA-BI ont la capacité de détecter les différentes stationnarités présentes dans l'image (champ $U = u$), ce qui n'est pas le cas pour le modèle classique.

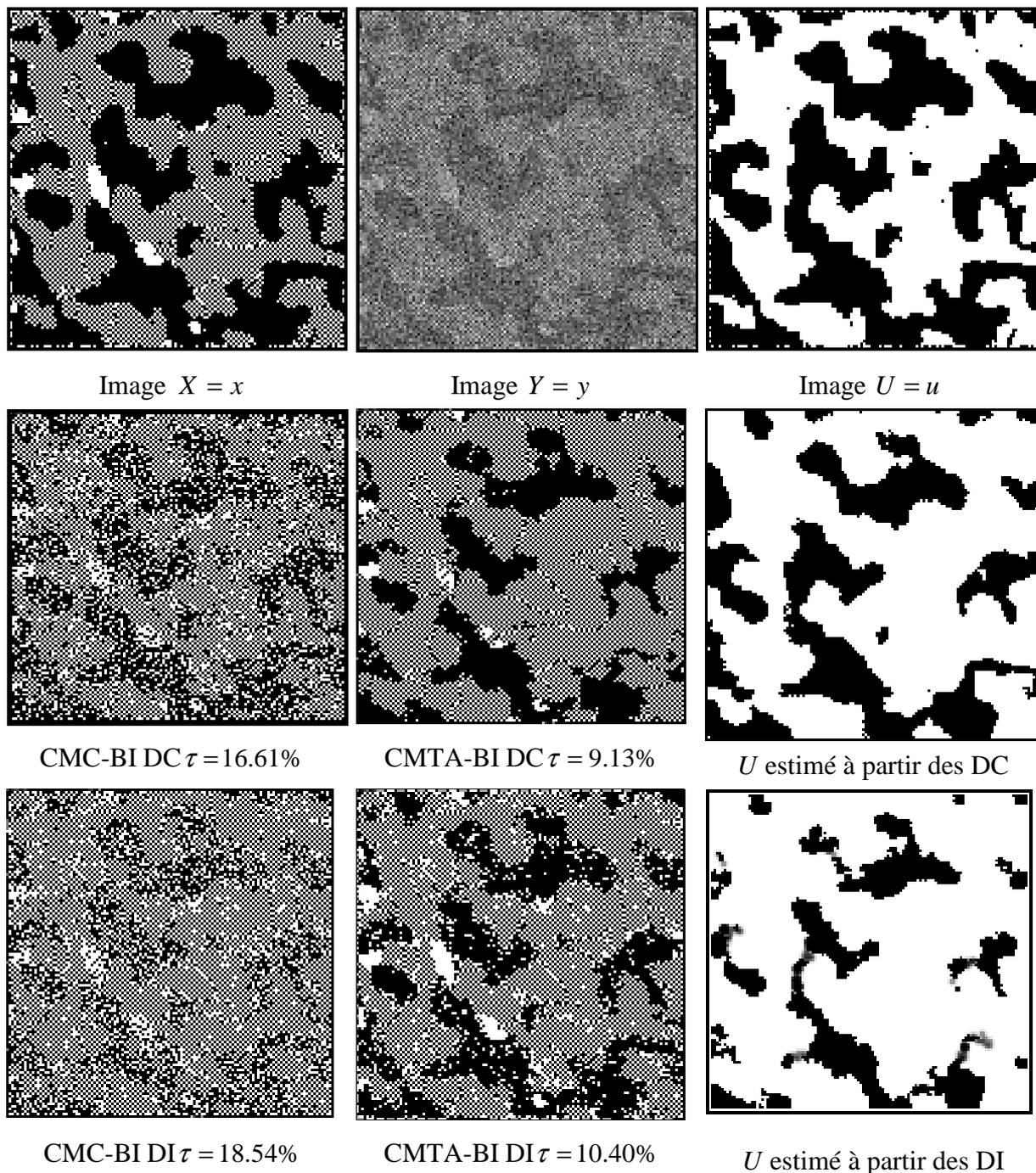


FIG. 8 – Simulation d'un CMTA-BI et segmentation MPM basée sur les CMC-BI et CMTA-BI. DI et DC sont les données incomplètes et complètes, respectivement.

Les valeurs réelles utilisées pour la simulation ainsi que les valeurs estimées par NMDI et NMDC sont explicitées dans TAB. 6.

TAB. 6 Résultats correspondant à FIG. 8

Paramètres	Valeurs réelles	CMTA-BI		CMC-BI	
		NMDC	NMDI	NMDC	NMDI
α_H^1, α_V^1	1.00, 1.00	0.96, 1.03	0.84, 0.81		
$\alpha_{H\lambda_1}^2, \alpha_{V\lambda_1}^2$	1.00, 1.00	0.51, 0.66	0.7, 0.72	-0.37, -0.46	-0.42, -0.44
$\alpha_{H\lambda_2}^2, \alpha_{V\lambda_2}^2$	-1.00, -1.00	-0.95, -0.87	-0.86, -0.46		
m_1	0.00	0.01	-0.10	0.01	0.01
m_2	2.00	2.00	1.55	2.00	2.00
σ_1^2	1.00	1.01	0.95	1.01	1.01
σ_2^2	1.00	1.00	1.10	1.00	1.00

NMDC: Estimés à partir de données complètes, NMDI: estimés à partir de données incomplètes

La simulation représentée en FIG. 8 a été obtenue en fixant le nombre de stationnarités à 2 ; autrement dit, $\Lambda = \{\lambda_1, \lambda_2\}$. Cependant, la simulation d'un champ contenant plus de 2 stationnarités ne pose pas de problème particulier. Aussi, il est possible de gérer les proportions des différentes stationnarités contenues dans l'image en enrichissant l'énergie (IV.6) par la prise en compte du paramètre β associé aux singletons ($\beta_{i=0, \dots, \text{Card}(\Lambda)} \neq 0$) qui a été jusque là, mis à zéro. Les simulations présentées dans ce qui suit ont été obtenues dans un cadre plus général, en prenant $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ et $\beta_i \neq 0$.

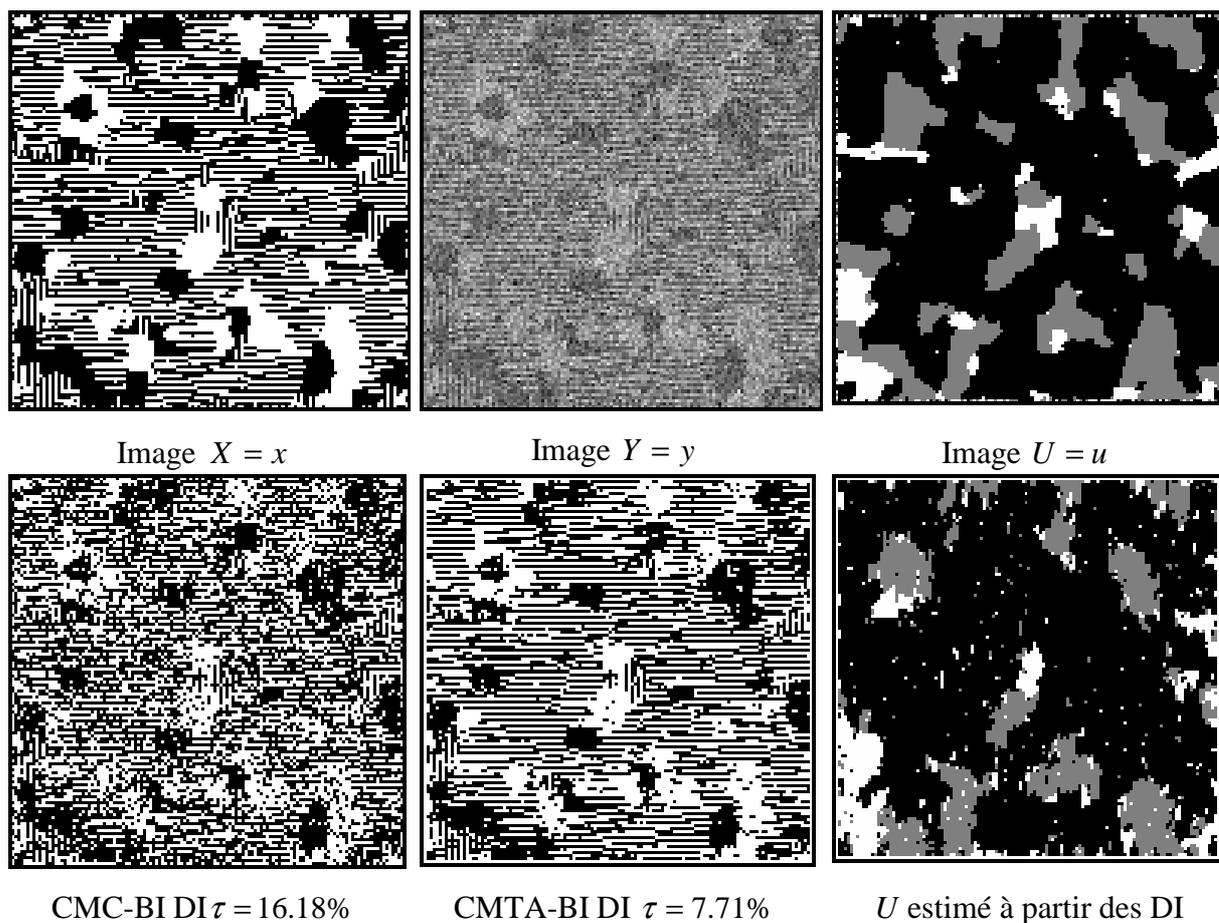


FIG. 9 – Simulation d'un CMTA-BI à trois stationnarités et segmentation MPM non supervisée basée sur les CMC-BI et CMTA-BI. DI: données incomplètes.

TAB. 7 Résultats correspondant à FIG. 9

Paramètres	Valeurs réelles	CMTA-BI DI	CMC-BI DI
$\beta_1, \beta_2, \beta_3$	0.50, 0.40, 0.30	0.49, 0.37, 0.28	
α_H^1, α_V^1	1.00, 1.00	0.59, 0.69	
$\alpha_{H\lambda_1}^2, \alpha_{V\lambda_1}^2$	1.00, -0.30	0.75, -0.34	0.00, 0.00, 0.42, -0.02
$\alpha_{H\lambda_2}^2, \alpha_{V\lambda_2}^2$	1.00, 1.00	0.56, 0.62	
$\alpha_{H\lambda_3}^2, \alpha_{V\lambda_3}^2$	-0.30, 1.00	-0.24, 0.74	
m_1	0.00	-0.05	0.09
m_2	2.00	2.01	2.04
σ_1^2	1.00	0.97	1.05
σ_2^2	1.00	0.97	0.99

CMTA-BI DI: Estimés à partir de données incomplètes en utilisant les CMTA-BI,
 CMC-BI DI: estimés à partir de données incomplètes en utilisant les CMC-BI

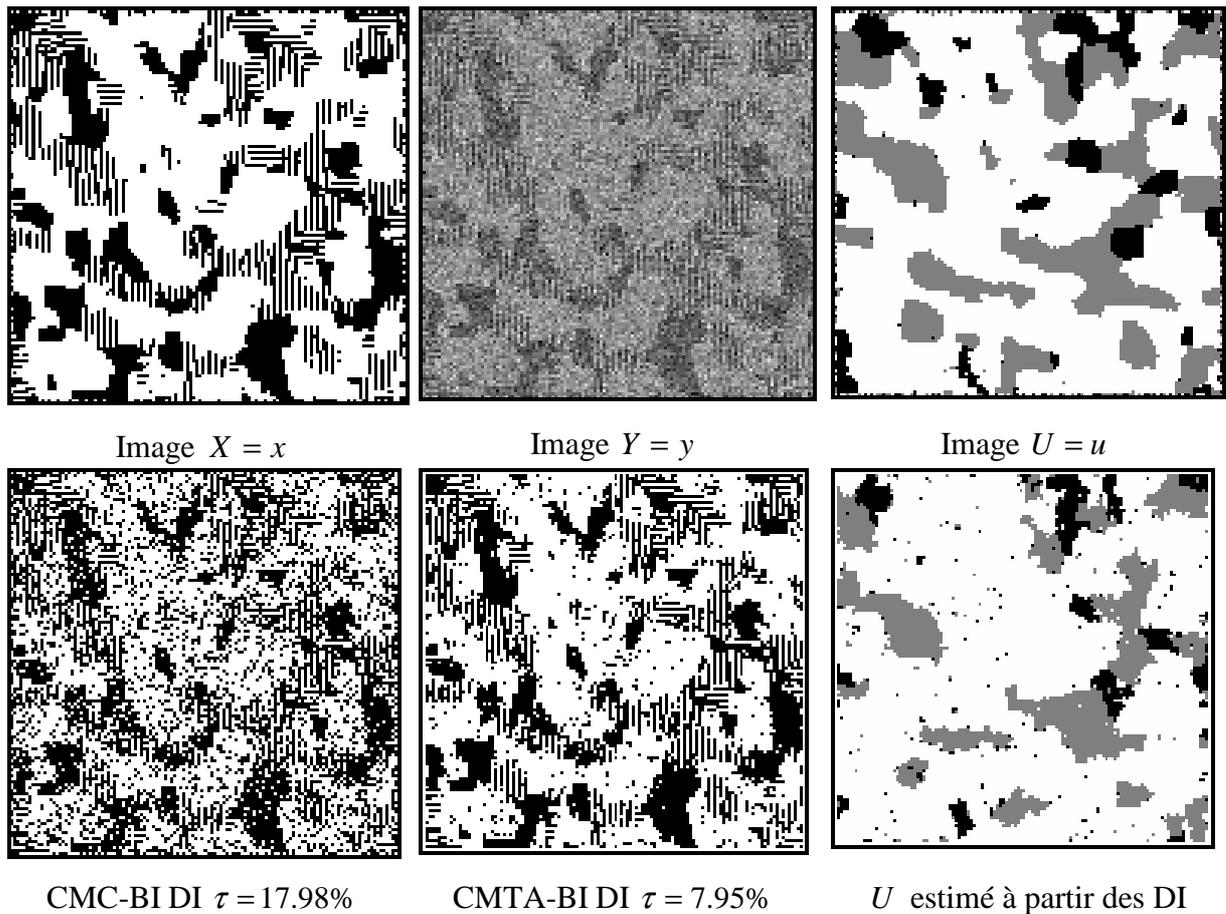


FIG. 10 – Simulation d'un CMTA-BI à trois stationnarités et segmentation MPM non supervisée basée sur les CMC-BI et CMTA-BI. DI: données incomplètes.

TAB. 8 Résultats correspondant à FIG. 10

Paramètres	Valeurs réelles	CMTA-BI DI	CMC-BI DI
$\beta_1, \beta_2, \beta_3$	0.30, 0.40, 0.50	0.27, 0.36, 0.43	
α_H^1, α_V^1	1.00, 1.00	0.60, 0.69	
$\alpha_{H\lambda_1}^2, \alpha_{V\lambda_1}^2$	1.00, -0.30	0.58, -0.24	0.00, 0.00, 0.18, 0.15
$\alpha_{H\lambda_2}^2, \alpha_{V\lambda_2}^2$	1.00, 1.00	-0.30, 0.68	
$\alpha_{H\lambda_3}^2, \alpha_{V\lambda_3}^2$	-0.30, 1.00	0.58, 0.65	
m_1	0.00	0.11	0.41
m_2	2.00	2.13	2.21
σ_1^2	1.00	0.98	1.12
σ_2^2	1.00	0.92	0.92

CMTA-BI DI: Estimés à partir de données incomplètes en utilisant les CMTA-BI,
 CMC-BI DI: estimés à partir de données incomplètes en utilisant les CMC-BI

Les FIG. 9 et FIG. 10 ont été obtenues en augmentant le nombre de stationnarités à trois et en tenant compte du paramètre β , afin de générer des proportions inégales de ces différentes stationnarités. Ainsi, dans FIG. 9, nous avons privilégié la zone de stationnarité λ_1 (trait horizontal de l'image $X = x$) en lui affectant un poids plus important ($\beta_1 = 0.5$) et dans FIG. 10, nous avons privilégié une autre zone de stationnarité. Les différents résultats montrent que la segmentation MPM non supervisée, fondée sur les CMTA-BI, fournit de meilleurs résultats que la segmentation MPM non supervisée, fondée sur les CMC-BI. En effet, nous constatons que les taux d'erreurs sont divisés par deux. Accessoirement, nous présentons également les valeurs estimées des champs $U = u$, qui peuvent avoir un intérêt propre. Aussi, les CMTA-BI permettent de modéliser les différentes stationnarités (Champ $U = u$).

Les valeurs des paramètres réelles et estimées correspondant à FIG. 9 et 10. sont représentées dans TAB. 7. et TAB. 8, respectivement.

Finalement, il apparaît que lorsque les données correspondent au modèle et présentent diverses stationnarités différentes, l'utilisation du nouveau modèle donne des résultats significativement meilleurs que l'utilisation du modèle classique.

IV.3.1.2 Image de synthèse

Comme précédemment, nous souhaitons vérifier si les performances de la segmentation MPM CMTA-BI, qui sont normales d'un point de vue théorique dans le cadre du modèle étudié ci-dessus, sont conservées lorsque les images utilisées ne suivent plus le modèle. Cependant, nous gardons une maîtrise partielle des données traitées dans la mesure où la vérité terrain et le bruitage sont donnés. Pour cela, nous avons effectué plusieurs expérimentations dont une est présentée, ci-après. Nous prenons une image de synthèse non stationnaire à deux classes que nous bruitons avec un bruit blanc gaussien de variance 1 et de moyennes 0 et 2. Ensuite, l'image bruitée est segmentée en utilisant la méthode MPM fondée, d'une part, sur les CMTA-BI et d'autre part, sur les CMC-BI. Les résultats fournis par la segmentation non supervisée, illustrés en FIG. 11, montrent les performances des CMTA-BI sur une l'image de synthèse traitée. Ainsi, la segmentation MPM CMTA-BI donne $\tau = 6.41\%$ tandis que celle fondée sur le CMC-BI donne $\tau = 10.37\%$.

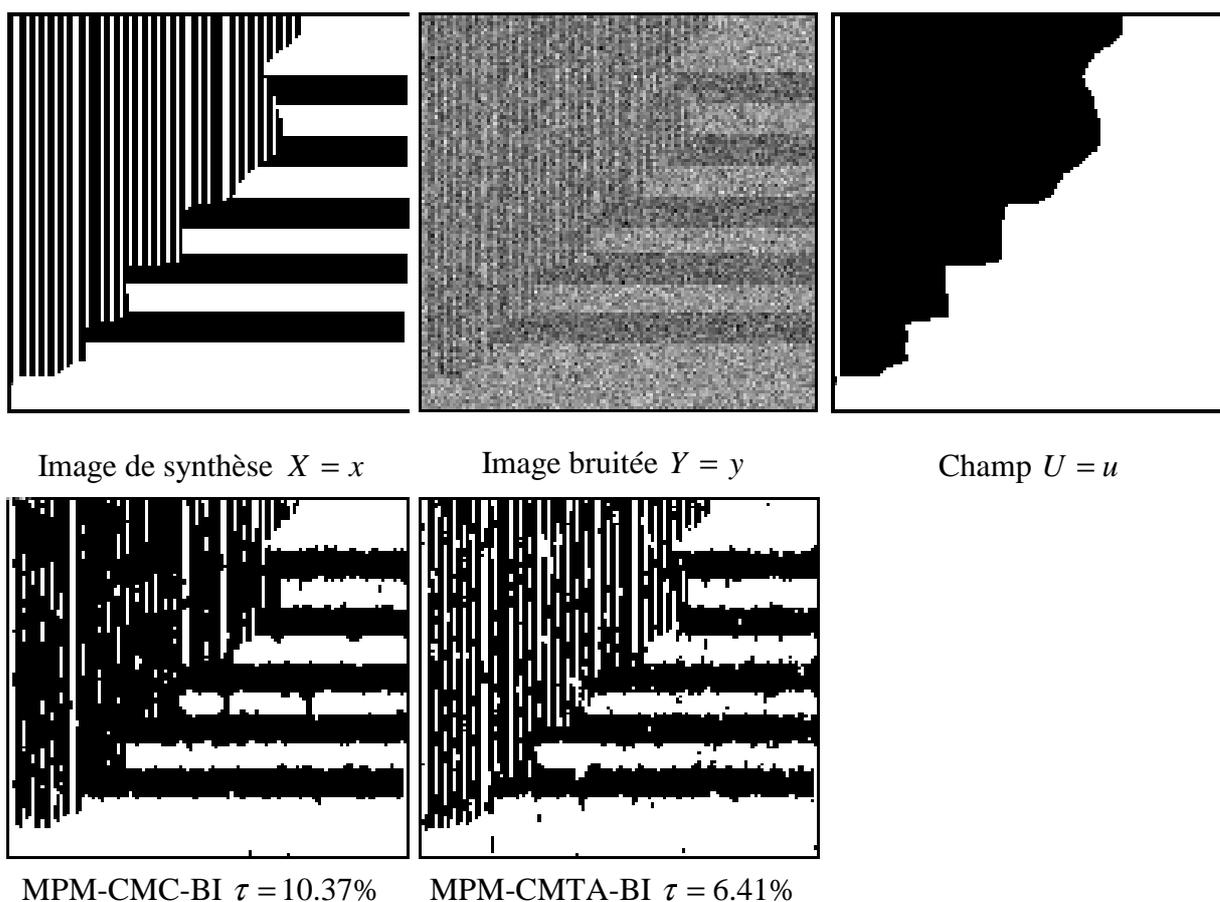


FIG. 11 – Image de synthèse, sa version bruitée et résultats de la segmentation MPM non supervisée

Finalement, l'exemple présenté, ainsi que d'autres simulations analogues, permettent d'affirmer que l'avantage du nouveau modèle par rapport au modèle classique reste significatif.

IV.3.1.3 Images réelles

Afin de valider l'algorithme de segmentation non supervisée MPM, fondé sur CMTA-BI et NMDI, nous l'appliquons sur des images réelles. Nous considérons des images des zones urbaines, généralement réputées pour être difficiles à segmenter. Les résultats de la segmentation non supervisée en 4 classes sont présentés à la figure FIG. 12 ci après.

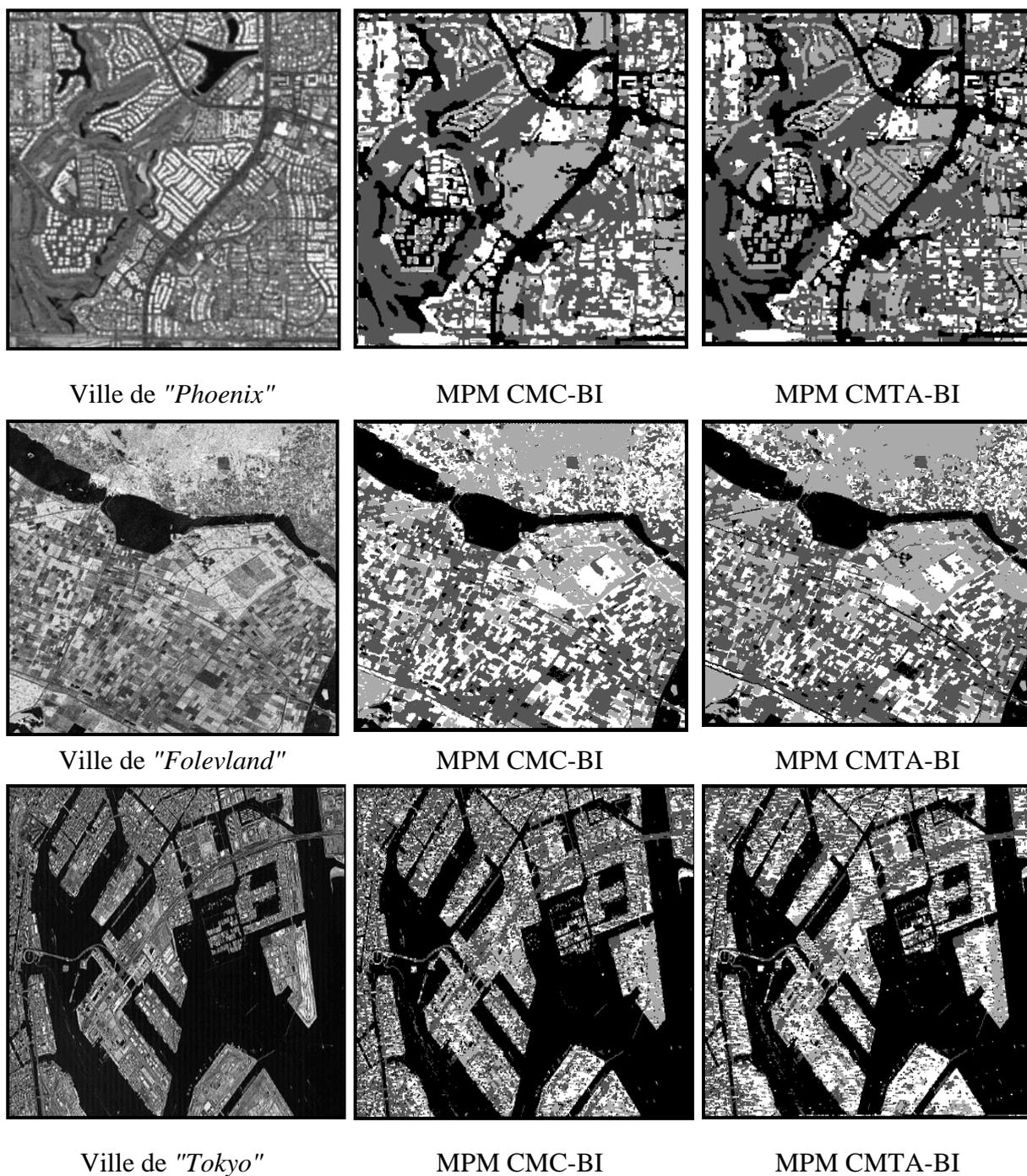


FIG. 12 – Résultats de segmentation MPM non supervisée d'images réelles basée sur les CMC-BI et les CMTA-BI.

Ainsi, FIG. 12 montre les résultats de segmentation MPM non supervisée fondée, d'une part, sur les CMC-BI et d'autre part, sur les CMTA-BI (à 2 stationnarités). Nous ne disposons pas, en absence de vérité terrain, d'un instrument de mesure objective de la qualité des

segmentations. Cependant, il apparaît visuellement que les résultats fournis par la segmentation non supervisée en utilisant ce nouveau modèle sont de meilleure qualité que ceux obtenus en utilisant le modèle classique. Les CMTA-BI ont ainsi permis l'amélioration du résultat de segmentation non supervisée.

IV.3.2 Champ de Markov triplet adapté

Contrairement au cas précédent, où les variables aléatoires Y_s étaient supposées indépendantes conditionnellement à $X = x$, nous considérons ici le cas plus général où cette indépendance n'a plus nécessairement lieu. Le modèle s'appellera simplement "*champ de Markov triplet adapté*" (CMTA), et les corrélations entre pixels peuvent être prises en considération. Dans les CMTA gaussiennes que nous considérons, la fonction énergie s'écrit:

$$\begin{aligned}
 U(x, y) = & \frac{1}{2} \left[\sum_{s \in S} a_{x_s} (y_s - m_{x_s})^2 + \sum_{(s,t) \in C^H} \alpha_H^1 (1 - 2\delta(x_s, x_t)) + (\alpha_{H\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots \right. \\
 & + \alpha_{H\lambda_M}^2 \delta(u_s, u_t, \lambda_M) (1 - \delta(x_s, x_t)) \\
 & + \sum_{(s,t) \in C^V} \alpha_V^1 (1 - 2\delta(x_s, x_t)) + (\alpha_{V\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots \\
 & \left. + \alpha_{V\lambda_M}^2 \delta(u_s, u_t, \lambda_M) (1 - \delta(x_s, x_t)) + \sum_{(s,t) \in C} q_{x_s x_t} (y_s - m_{x_s})(y_t - m_{x_t}) \right]
 \end{aligned} \tag{IV.7}$$

En reprenant le même plan que celui de la sous-section IV.3.1, nous exposons ci-après des résultats de segmentation en utilisant les CMTA, une comparaison avec la segmentation utilisant les CMCouple est effectuée. Les avantages relatifs aux divers modèles étant analogues à ceux de l'étude précédente, nous limitons ainsi la quantité des simulations et traitement présentés.

IV.3.2.1 Simulation

Un exemple de simulation d'un CMTA dont les paramètres sont présentés dans TAB. 9, est illustré dans FIG. 13. Une comparaison, entre la segmentation MPM fondée sur le

CMTAG et celle utilisant les CMCouple, est effectuée. Ainsi, l'utilisation de la méthode MPM non supervisée, fondée sur le modèle triplet dont la fonction énergie représentée par l'équation (IV.7), donne 15.10% de taux d'erreur, alors que celle fondée sur le modèle CMCouple (équation (IV.1) donne 21.51% de taux d'erreur.

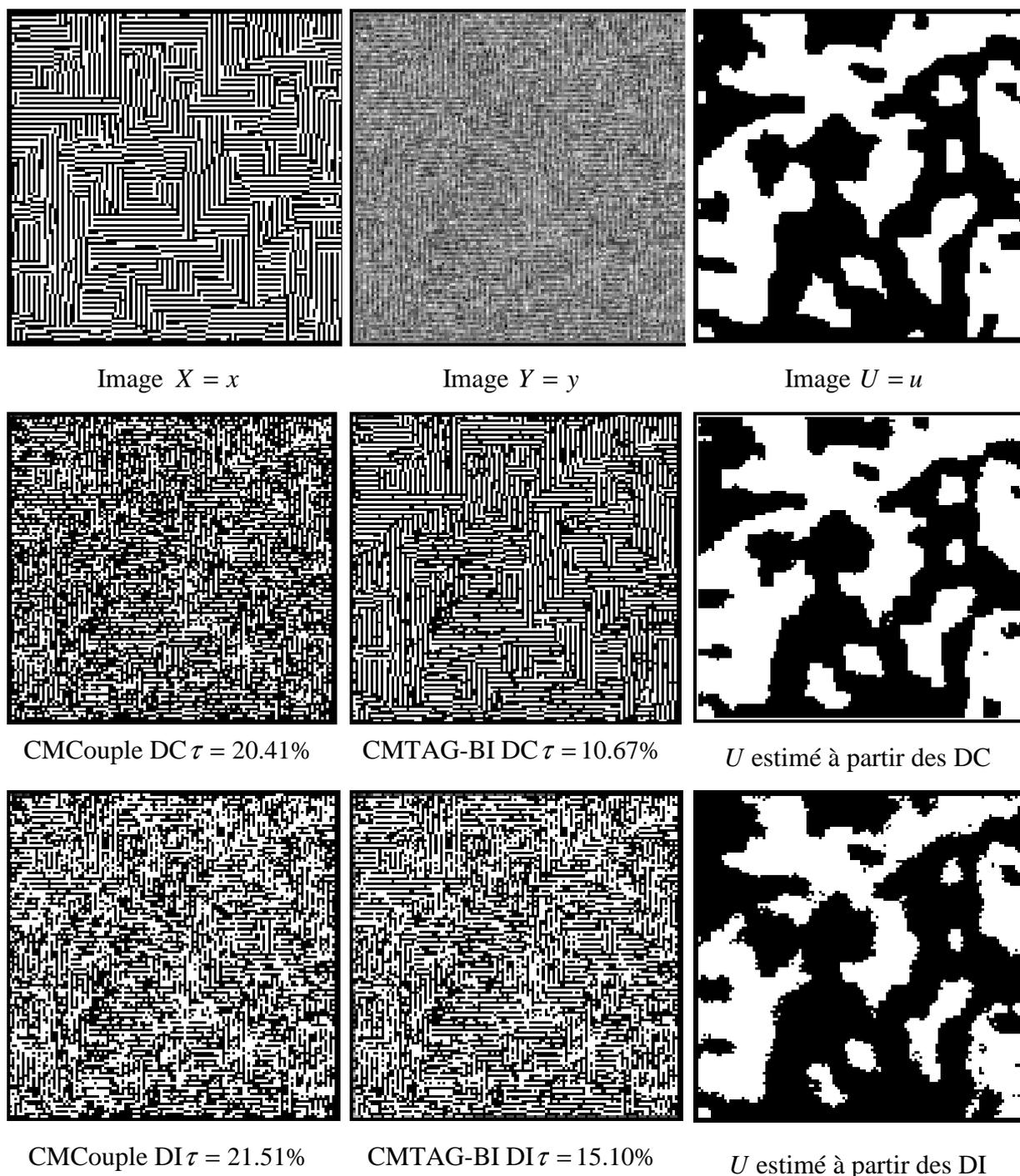


FIG. 13 – Simulation d'un CMTAG et segmentation MPM non supervisée, basée sur les CMCouple et CMTAG. DI et DC sont les données incomplètes et complètes, respectivement.

TAB. 9 Résultats correspondant à FIG. 13

Paramètres	Valeurs réelles	CMTG		CMCouple	
		DC	DI	DC	DI
α_H^1, α_V^1	2.00, 2.00	1.23, 1.26	0.86, 0.78		
$\alpha_{H\lambda_1}^2, \alpha_{V\lambda_1}^2$	2.00, -2.00	1.72, -0.84	0.80, -0.65	0.34, 0.33	0.11, 0.14
$\alpha_{H\lambda_2}^2, \alpha_{V\lambda_2}^2$	-2.00, 2.00	-1.81, 1.01	-1.06, 0.66		
m_1	0.00	-0.04	-0.02	-0.04	0.1
m_2	2.00	1.97	1.94	1.97	1.79
a_1	1.00	1.56	0.85	1.56	0.69
a_2	1.00	1.87	0.87	1.87	0.68
q_{11}	-0.30	-0.26	-0.19	-0.26	-0.21
q_{22}	-0.30	-0.40	-0.20	-0.40	-0.21
q_{12}	0.00	0.00	0.00	0.00	0.00

DC: Estimés à partir de données complètes, DI: estimés à partir de données incomplètes

IV.3.2.2 Images de synthèse

Considérons une image de synthèse "textures", présentée à la figure FIG. 14, que nous bruitons avec un bruit corrélé (nous utilisons la même technique de bruitage par moyenne mobile que celle appliquée à l'image "ring", équation (IV.4)). Comme dans la sous section précédente, une comparaison entre la segmentation MPM non supervisée, utilisant le modèle de Markov triplet adapté, en considérons deux stationnarités ($\Lambda = \{\lambda_1, \lambda_2\}$), et celle basée sur le modèle CMCouple, est effectuée. Là aussi, le modèle proposé permet d'améliorer les résultats de la segmentation non supervisée de ce type d'image. Cependant, en général, le gain en qualité est moins important que dans le cas de bruit indépendant. Les résultats des segmentations sont représentés dans FIG. 14.

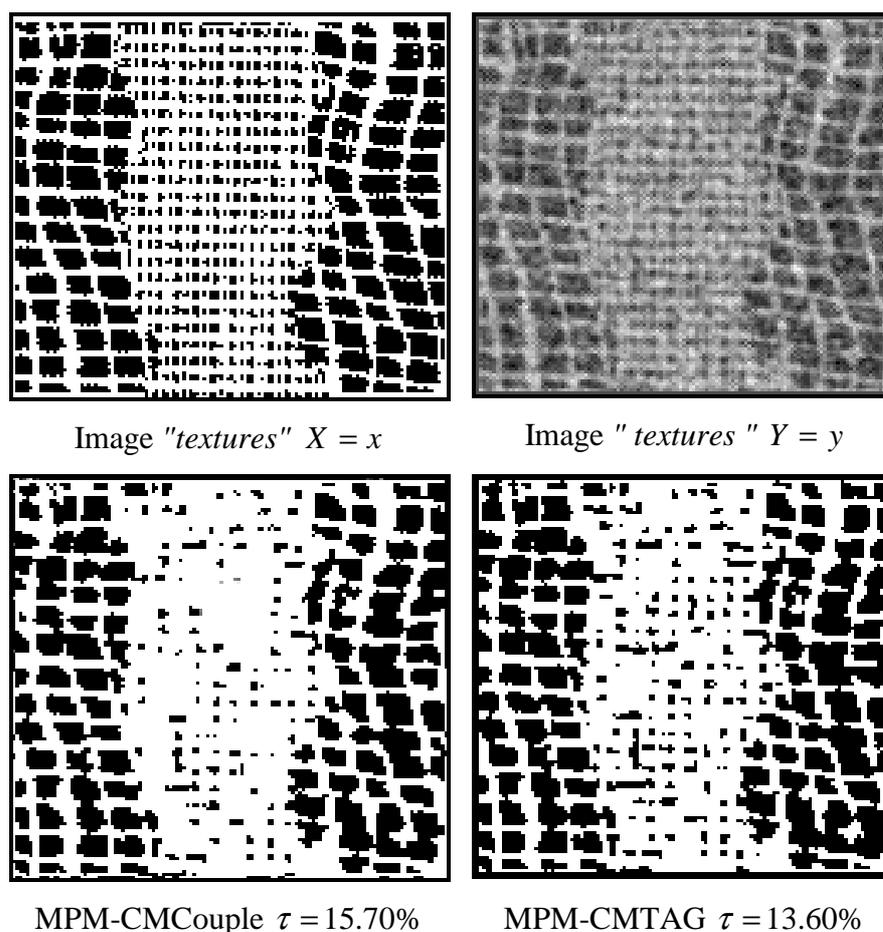


FIG. 14 – Image de synthèse, sa version bruitée et résultats de la segmentation MPM non supervisée

IV.3.2.3 Image réelle

Finalement, nous présentons un exemple d'application des CMTA dans le cas d'image réelle. Les résultats de segmentation MPM non supervisée à deux classes, dans le cas des CMTA - en considérant $\Lambda = \{\lambda_1, \lambda_2\}$ - et CMCouple, sont exposés.

L'image illustrée dans FIG. 15 représente une partie de la ville de "Phoenix". En l'absence de la vérité terrain, il nous est impossible d'évaluer les performances de la segmentation non supervisée, fondée sur le nouveau modèle en terme de taux d'erreur. Cependant, nous remarquons visuellement que l'utilisation de la modélisation CMTA a permis de retrouver certains détails, comme immeubles au milieu de l'image, qui n'apparaissent pas dans le résultat de la segmentation obtenue par le modèle classique.

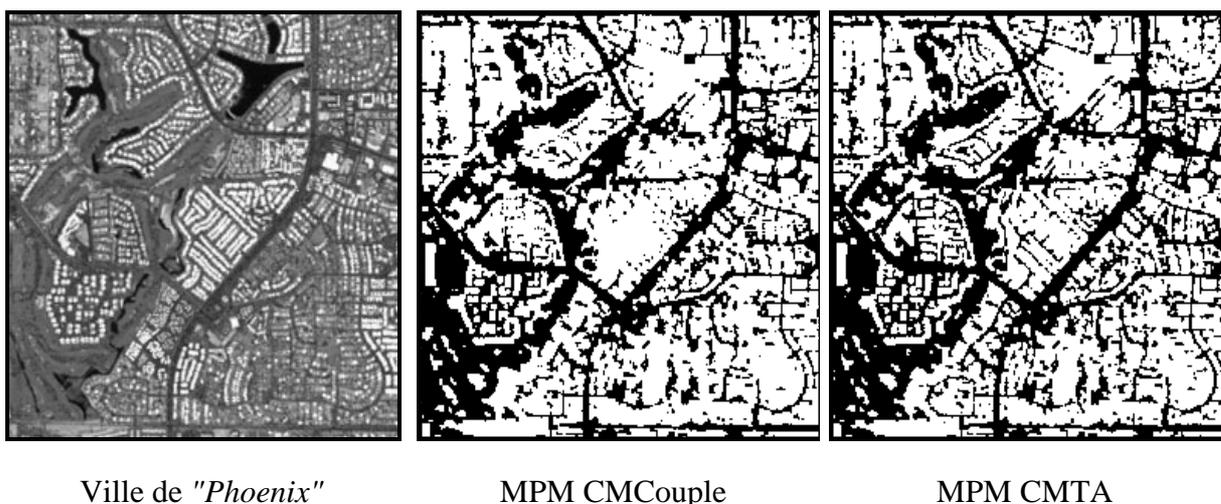


FIG. 15 – Résultats de segmentation MPM non supervisée en deux classes de la ville de "Phoenix" basée sur les CMCouple et CMTAG.

IV.4 Conclusion

Nous avons présenté, dans ce chapitre, des exemples d'application des différents modèles étudiés, en segmentation statistique non supervisée d'images.

Dans un premier temps, nous avons testé les méthodes fondées sur les modèles couples (CMC Couple) et triplets (CMT), sur des images simulées, de synthèse et réelles. Les résultats obtenus montrent que le modèle triplet peut fournir des résultats meilleurs que ceux obtenus en utilisant la modélisation classique (CMC-BI) ou la modélisation (CMC Couple).

Ensuite, nous avons proposé deux utilisations des modèles CMT pour traiter des images non stationnaires : une fondée sur la théorie de l'évidence et une autre où le champ auxiliaire modélise directement les différentes stationnarités présentes dans une image. Nous avons montré, via des simulations et traitements d'un certain nombre d'images réelles, que l'utilisation des CMT permet non seulement l'amélioration des résultats de segmentation non supervisée obtenus par les méthodes classiques, mais présente également l'avantage de pouvoir détecter les différentes stationnarités.

Chapitre V

Extension au cas non gaussien et mélange généralisé

L'étude que nous avons menée jusqu'à présent (modélisation et estimation des paramètres) s'est limitée au cas gaussien. Cette restriction est tout d'abord due au fait que les différentes lois peuvent souvent être approchées ou vues comme un mélange de gaussiennes, et à l'utilisation fréquente de la loi normale qui est généralement justifiée par des motifs asymptotiques (théorème centrale limite). Cependant, une telle modélisation n'est pas toujours valable pour tout type d'images. En effet, le bruit présent dans les images radar, sonar, ultrason, infrarouge et IRM, n'est pas nécessairement gaussien [ATA 95, CAS 05, KaV 94, MaW 91]. De plus, différents types de bruits, variant avec la classe, peuvent être présents dans une image donnée [DeP 02]. En d'autres termes, différentes classes peuvent être modélisées par des bruits de nature différente.

Le but de ce chapitre est de proposer des modèles et des méthodes de traitement valables dans des cas non gaussiens. Ainsi, une approche originale fondée sur la nouvelle méthode d'estimation des paramètres, présentée au Chapitre III, sur le système de Pearson et sur les champs de Markov triplets adaptés (CMTA), sera présentée. Cette approche peut être vue comme une extension au cas non stationnaire avec bruit corrélé de l'approche proposée dans [DMP 97].

V.1 Modélisation statistique

Des différentes expérimentations effectuées, dont une partie a été présentée dans le chapitre IV, nous avons constaté que l'utilisation des champs de Markov triplets adaptés (CMTA) a permis l'amélioration des résultats de segmentation non supervisée des images non stationnaires. Nous proposons, ci-après, une extension du CMTA gaussien au cas où les marginales du bruit conditionnellement aux classes sont quelconques, le bruit restant corrélé.

Soit S un ensemble de pixels, X , Y et U trois champs aléatoires, tels que chaque variable aléatoire X_s , Y_s et U_s est à valeurs dans l'ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$, l'ensemble des réels R et l'ensemble des stationnarités $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, respectivement. Le but est de retrouver ou estimer le champ caché X , à partir du champ observé Y . Pour ce faire, on suppose que le triplet $T = (X, U, Y)$ est markovien et que le couple $W = (X, U)$ l'est également. Considérons la fonction énergie $U(x, u)$ de la loi du couple $W = (X, U)$ donnée par:

$$U(x, u) = \sum_{(s,t) \in C} \alpha^1 (1 - 2\delta'(x_s, x_t)) - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots + \alpha_{\lambda_m}^2 \delta(u_s, u_t, \lambda_m)) (1 - \delta(x_s, x_t)) \quad (V.1)$$

avec C l'ensemble des cliques,

$$\delta'(x_s, x_t) = \begin{cases} 1 & \text{pour } x_s = x_t \\ 0 & \text{sinon} \end{cases}, \text{ et } \delta(u_s, u_t, \lambda_i) = \begin{cases} 1 & \text{pour } u_s = u_t = \lambda_i \\ 0 & \text{sinon} \end{cases}. \quad (V.2)$$

Ainsi, la loi $p(x, u)$ de $W = (X, U)$ s'écrit:

$$p(x, u) = \gamma \exp - \left[\sum_{(s,t) \in C} (\alpha^1 (1 - 2\delta'(x_s, x_t)) - (\alpha_{\lambda_1}^2 \delta(u_s, u_t, \lambda_1) + \dots + \alpha_{\lambda_m}^2 \delta(u_s, u_t, \lambda_m)) (1 - \delta(x_s, x_t))) \right] \quad (V.3)$$

La loi $p(x, u)$ étant définie par (V.3), il reste à définir les lois $p(y|x, u)$. Dans le cas gaussien à bruit indépendant, on peut considérer que le champ $Y = (Y_s)_{s \in Y}$ est obtenu à partir d'un ensemble de variables aléatoires gaussiennes $(Y'_s)_{s \in S}$ indépendantes, de moyennes $E[Y'_s] = 0$ et de variances $VAR[Y'_s] = 1$ par $Y_s = m_{X_s} + \sigma_{x_s} \cdot Y'_s$. Notons que $Y' = (Y'_s)_{s \in S}$ est indépendant de $X = (X_s)_{s \in S}$ et de $U = (U_s)_{s \in S}$. Plaçons nous maintenant dans le cadre non gaussien. L'idée est d'utiliser de manière analogue un ensemble $Y' = (Y'_s)_{s \in S}$ de variables centrées de variance égale à 1, avec la différence que $Y' = (Y'_s)_{s \in S}$ est un champs gaussien de Markov, ce qui permettra d'obtenir un bruit corrélé. Par ailleurs, notons F la fonction de répartition de la loi normale $N(0,1)$, et notons G_{x_s} la fonction de répartition de la loi marginale $p(y_s|x_s)$ (notée aussi G_i pour $p(y_s|x_s = \omega_i)$). En posant $Y_s = G_{x_s}^{-1} \circ F(Y'_s)$ (qui remplace et généralise la relation simple $Y_s = m_{X_s} + \sigma_{x_s} \cdot Y'_s$ ci-dessus), nous savons que la loi de Y_s conditionnelle à $X_s = x_s$ est donnée par la fonction de répartition G_{x_s} , nous avons donc la loi voulue. Par ailleurs, les variables $(Y'_s)_{s \in S}$ étant corrélées, les variables (Y_s) sont également corrélées.

Finalement, le modèle est le suivant. Soit $Y' = (Y'_s)_{s \in S}$ est un champs gaussien d'énergie $\phi(y') = \sum_{c \in C} \phi_c(y'_c)$, F la fonction de répartition de $N(0,1)$, G_{x_s} les fonction de répartition des $p(y_s|x_s)$. On pose

$$Y_s = G_{x_s}^{-1} \circ F(Y'_s), \tag{V.4}$$

l'énergie de la loi markovienne $p(y|x, u) = p(y|x)$ est alors

$$\sum_{c \in C} \phi_c(F^{-1} \circ G_{x_c}(y_c)) - \sum_{s \in S} \log \left| \frac{\partial(F^{-1} \circ G_{x_s}(y_s))}{\partial y_s} \right| \tag{V.5}$$

avec $F^{-1} \circ G_{x_c}(y_c) = (F^{-1} \circ G_{x_s}(y_s))_{s \in S}$ et $\frac{\partial(F^{-1} \circ G_{x_s}(y_s))}{\partial y_s} = \frac{g_{x_s}(y_s)}{f[F^{-1} \circ G_{x_s}(y_s)]}$, tel que $g_{x_s}(y_s) = \frac{\partial G_{x_s}(y_s)}{\partial y_s}$ et f la densité de $N(0,1)$.

La loi du triplet est donnée par :

$$p(x, u, y) = \gamma \exp \left[-U(x, u) - \sum_{c \in C} \varphi_c(F^{-1} \circ G_{x_c}(y_c)) + \sum_{s \in S} \log \left| \frac{\partial(F^{-1} \circ G_{x_s}(y_s))}{\partial y_s} \right| \right] \quad (\text{V.6})$$

On obtient bien la markovianité de $p(x, u|y)$, ce qui permet d'estimer les marginales $p(x_s, u_s|y)$, qui donnent $p(x_s|y) = \sum_{u \in \Lambda} p(x_s, u_s|y)$ et $p(u_s|y) = \sum_{x \in \Omega} p(x_s, u_s|y)$ utilisées dans les segmentations. Comme précédemment, ces deux lois permettent d'obtenir, d'une part, une image segmentée et d'autre part, une image représentant les différentes stationnarités.

V.2 Système de Pearson

La méthode d'estimation exposée dans le paragraphe suivant permet d'estimer des "mélanges généralisés", à savoir : identifier la forme de G_{x_s} pour chaque classe x_s dans $\Omega = \{\omega_1, \dots, \omega_k\}$, et estimer ses paramètres. Ce type de méthodes a déjà été proposé dans [DMP 97] dans le cas des bruits indépendants et images stationnaires ; notre méthode peut ainsi être vue comme une extension de la méthode proposée dans [DMP 97], au cas des images non stationnaires et du bruit corrélé.

Nous supposons que chaque G_{x_s} est dans le système de Pearson. Avant d'exposer la méthode d'estimation des paramètres, nous rappelons brièvement ce système.

Le système de Pearson est constitué des densités de probabilité vérifiant l'équation différentielle suivante :

$$\frac{df(x)}{dx} = -\frac{(x+a)}{c_0 + c_1x + c_2x^2} f(x) \quad (\text{V.7})$$

a, c_0, c_1 et c_2 sont les paramètres de Pearson. En centrant la variable aléatoire X et en notant m_1, m_2, m_3, m_4 les moments d'ordre 1, 2, 3 et 4 définis par:

$$m_1 = E[Y] \quad (\text{V.8})$$

$$m_k = E[(Y - E(Y))^k], \quad k \geq 2 \quad (\text{V.9})$$

Les coefficients du système de Pearson s'obtiennent comme suit:

$$c_0 = \frac{m_2(4\gamma_2 - 3\gamma_1)}{10\gamma_2 - 12\gamma_1 - 18} \quad (\text{V.10})$$

$$c_1 = \frac{\sqrt{\gamma_1 m_2}(\gamma_2 + 3)}{10\gamma_2 - 12\gamma_1 - 18} \quad (\text{V.11})$$

$$c_2 = \frac{2\gamma_2 - 3\gamma_1 - 6}{10\gamma_2 - 12\gamma_1 - 18} \quad (\text{V.12})$$

$$a = c_1 \quad (\text{V.13})$$

avec γ_1 et γ_2 le *kurtosis* et le *skewness*, respectivement, définis par :

$$\gamma_1 = \frac{(m_3)^2}{(m_2)^3} \quad (\text{V.14})$$

$$\gamma_2 = \frac{m_4}{(m_2)^2} \quad (\text{V.15})$$

En posant:

$$\kappa = \frac{\gamma_1(\gamma_2 + 3)^2}{4(4\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1 - 6)}, \quad (\text{V.16})$$

huit familles de distribution peuvent être distinguées, suivant des tests effectués sur la constante κ , les valeurs du *skewness* et du *kurtosis*. Ainsi on a :

Loi de type I : $\kappa < 0$	Loi bêta de 1 ^{er} espèce
Loi de type II : $\gamma_1 = 0 \quad \gamma_2 < 3$	Loi bêta de 1 ^{er} espèce
Loi de type III : $2\gamma_2 - 3\gamma_1 - 6 = 0$	Loi gamma
Loi de types IV : $0 < \kappa < 1$	Loi Bessel modifiée de 2 ^{eme} espèce
Loi de type V : $\kappa = 1$	Loi Bessel modifiée de 2 ^{eme} espèce
Loi de type VI : $\kappa > 1$	Loi bêta de 2 ^{eme} espèce
Loi de type VII : $\gamma_1 = 0 \quad \gamma_2 > 3$	Loi Bessel modifiée de 2 ^{eme} espèce
Loi de type VIII : $\gamma_1 = 0 \quad \gamma_2 = 3$	Loi normale

Le système de Pearson peut alors être utilisé dans la reconnaissance de la forme de G_{x_s} donnée de la manière suivante. Si on dispose d'un échantillon suivant G_{x_s} , on peut estimer les quatre premiers moments qui donnent le *kurtosis* et le *skewness* définis par (V.14) et (V.15). Ces deux quantités permettent de définir la forme de la loi G_{x_s} comme indiqué ci-dessus. Sachant la forme, on utilise les moments pour calculer les paramètres. Nous allons voir comment intégrer cette démarche (bien entendu, en segmentation non supervisée nous ne disposons pas d'échantillon de G_{x_s}) dans la NMDI au paragraphe suivant.

Parmi ces huit types de famille, nous nous sommes limités au cours de cette étude à trois lois usuelles, à savoir : de type I (loi bêta), de type III (loi gamma) et de type VIII (loi normale).

V.3 NMDI généralisée

La méthode que nous décrivons ci-après, est une généralisation de celle étudiée dans § III.3.2. dans le sens où la forme du bruit de chaque classe est identifié, et les paramètres identifiant chaque bruit sont estimés.

Soit α , ν^1 et ν^2 , les paramètres associés à (X, U) , Y et G_i , respectivement. Supposons une image à k classes, chacune d'elle bruitée selon une loi donnée. Dans ce cas, le vecteur paramètre à estimer est composé de $\theta = (\alpha, \nu^1, \nu^2)$, tel que $\nu^2 = (\nu_1^2, \dots, \nu_k^2)$ (chaque ν_i^2 défini G_i). Celui-ci est estimé suivant les étapes :

- Initialiser un θ^0 ;
- A chaque itération n ,
 - Simuler (x^{n+1}, u^{n+1}) selon $p(x, u | y, \theta^n)$
 - Estimer α^{n+1} à partir de (x^{n+1}, u^{n+1}) comme dans la méthode NMDI ;
 - Pour chaque $1 \leq i \leq k$, utiliser $S_i = \{s \in S / x_s = \omega_i\}$ et l'échantillon $y^i = (y_s^i)_{s \in S_i}$ pour rechercher G_i^{n+1} dans le système de Pearson (ce qui donne la forme de G_i et les paramètres correspondants) ;
 - Calculer $(y_s^{n+1})' = F^{-1} \circ G_{x_s}^{n+1}(y_s)$ et estimer $(\nu^1)^{n+1}$ à partir de $(y_s^{n+1})' = ((y_s^{n+1})')_{s \in S}$ par la méthode NMDI;

Nous retrouvons la méthode NMDI lorsque tous les G_i sont gaussiennes.

V.4 Expérimentations

Nous rappelons brièvement deux lois (gamma et bêta) qui seront utilisées dans les simulations ci-après.

La loi Gamma $\Gamma(p, \lambda)$ de paramètres p et λ ($(p, \lambda) > 0$), s'écrit sous la forme:

$$f(x) = \begin{cases} \frac{\lambda^{-p} x^{p-1} e^{-x/\lambda}}{\Gamma(p)} & \text{pour } x > 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{V. 17})$$

où Γ est la fonction Gamma définie par :

$$\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt \text{ pour } p > 0 \quad (\text{V.18})$$

Sa moyenne et sa variance sont données par:

$$m = p\lambda \quad (\text{V.19})$$

$$\sigma^2 = p\lambda^2 \quad (\text{V.20})$$

La loi Bêta $\beta(r, s)$ est définie par les paramètres r et s ($(r, s) > 0$), elle est de densité:

$$f(x) = \begin{cases} \frac{x^{r-1}(1-x)^{s-1}}{B(r, s)} & \text{pour } 0 < x < 1 \\ 0 & \text{sinon} \end{cases} \quad (\text{V.21})$$

où B est la fonction Bêta définie par:

$$B(r, s) = \int_0^1 t^{r-1}(1-t)^{s-1} dt = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} \quad (\text{V.22})$$

La loi bêta est de moyenne :

$$m = \frac{r}{r+s} \quad (\text{V.23})$$

et variance:

$$\sigma^2 = \frac{rs}{(r+s)^2(r+s+1)} \quad (\text{V.24})$$

Remarque

On peut montrer que les coefficients de l'équation différentielle (V.16) correspondant à loi Gamma $\Gamma(p, \lambda)$ sont :

$$p = -\frac{a}{c_1} + 1, \quad \lambda = c_1 \quad (\text{V.25})$$

et ceux correspondant à la loi Bêta $\beta(r, s)$.

$$r = \frac{a}{2c_2 \sqrt{\frac{c_1^2 - 4c_0c_2}{4c_2^2}}} + 1, \quad s = -\frac{a}{2c_2 \sqrt{\frac{c_1^2 - 4c_0c_2}{4c_2^2}}} + 1 \quad (\text{V.26})$$

Nous proposons dans cette partie quelques résultats obtenus à partir des images simulées. Afin d'analyser les performances de cette méthode "généralisée", une comparaison entre celle-ci à celle utilisant la modélisation gaussienne est effectuée. Pour ce faire, des images ont été simulées en utilisant la fonction énergie définie dans l'équation (V.1), ensuite bruitées comme suit:

- Simuler un champ markovien gaussien Y' tel que $Y' \sim N(0,1)$;
- Bruiter chacune des classes $x_s = \omega_i$, de l'image simulée $X = x$, par un bruit selon la loi de probabilité donnée G_i et en utilisant Y' par :

$$y_s = G_i^{-1} \circ F(y_s')$$
(V.30)

Le modèle est ensuite identifié comme décrit dans V.4, et l'image $Y = y$ est segmentée par la méthode MPM.

Nous présentons, ci-après, deux séries d'expériences : une, avec bruit indépendant et une autre avec bruit corrélé. Les deux marginales sont une loi bêta et une loi gamma.

(i) bruitage indépendant

Dans cette série d'expérimentations, nous présentons une image simulée à deux classes ω_1, ω_2 où chacune d'elles a été bruitée selon loi bêta ($B(2,1)$) et la loi gamma ($\Gamma(1,2)$), respectivement. Ainsi le champ $Y' = (y_s')_{s \in S}$ est composé des variables gaussiennes indépendantes. L'image ainsi obtenue est segmentée, d'une part, en utilisant les lois correspondantes et d'autre part, en supposant que le bruitage utilisé pour chacune des classe est un bruit blanc gaussien. Les résultats sont illustrés en FIG. 1.

Nous constatons que l'utilisation des lois gaussiennes au lieu des vraies lois bêta et gamma peut dégrader considérablement la qualité de la segmentation.

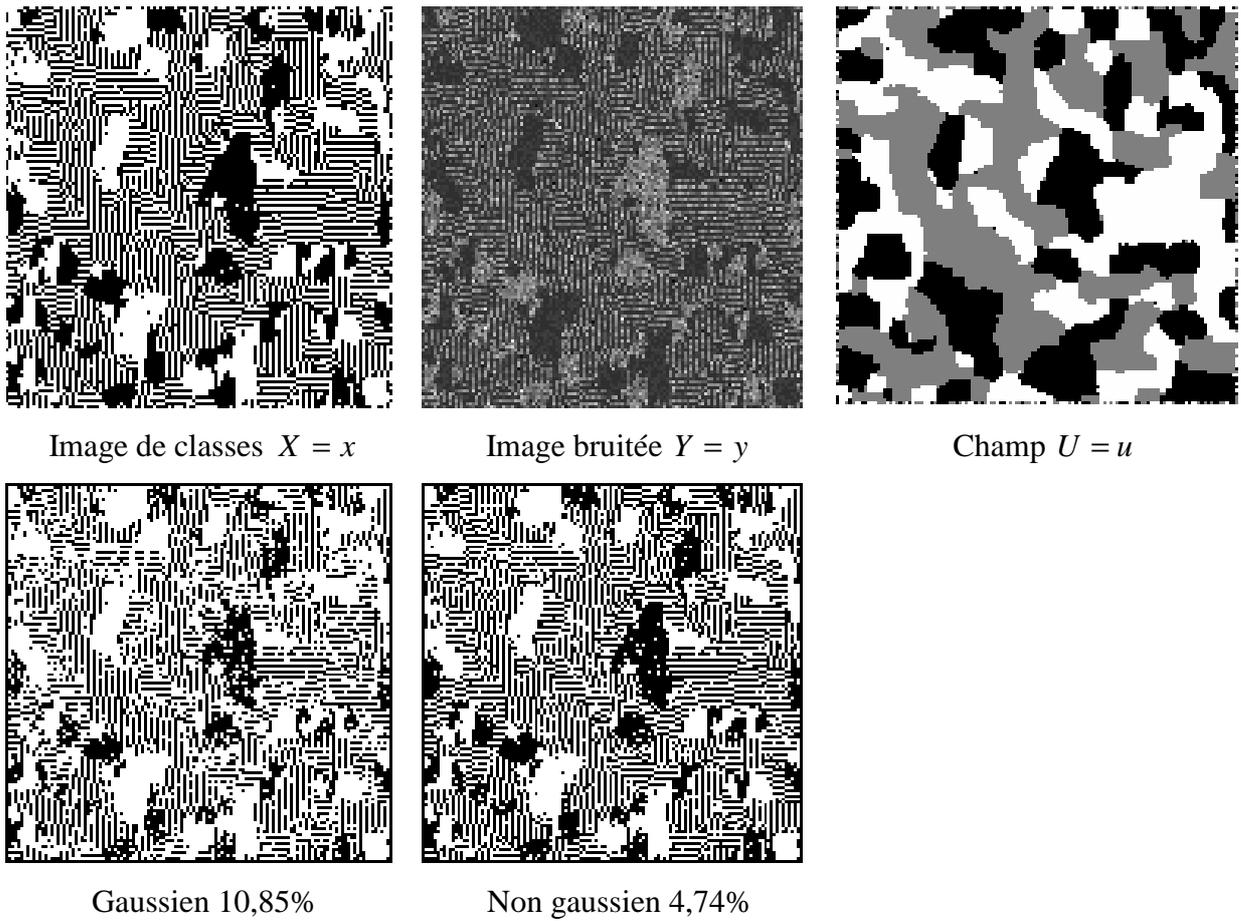


FIG. 1 – Résultats de segmentation non supervisée en utilisant les lois Gamma et Bêta, bruit indépendant

(ii) bruitage dépendant

Dans cette série d'expérimentations, nous considérons toujours une image simulée à deux classes ω_1, ω_2 , dont chacune d'elles a été bruitée selon la loi bêta et la loi gamma, respectivement. Ici le champs $Y' = (y_s')_{s \in S}$ est un champ markovien et nous nous trouvons donc dans le cas le plus général. Plus précisément, les lois marginales utilisées dans cette expérience sont $p(y_s | x_s = \omega_1) = \Gamma(1,2)$, $p(y_s | x_s = \omega_2) = B(2,1)$, et l'énergie du champs de

Markov gaussien Y' est $\varphi(y') = \frac{1}{2} \left[\sum_{s \in S} (y_s')^2 + \sum_{(s,t)} -0.2 y_s' y_t' \right]$. L'image ainsi obtenue est

segmentée, d'une part, en utilisant les estimées des lois dans le système de Pearson et d'autre

part, en supposant que le bruitage est gaussien. Les résultats sont illustrés en FIG. 2 (nous présentons également l'estimée des différentes stationnarités $U = u$).

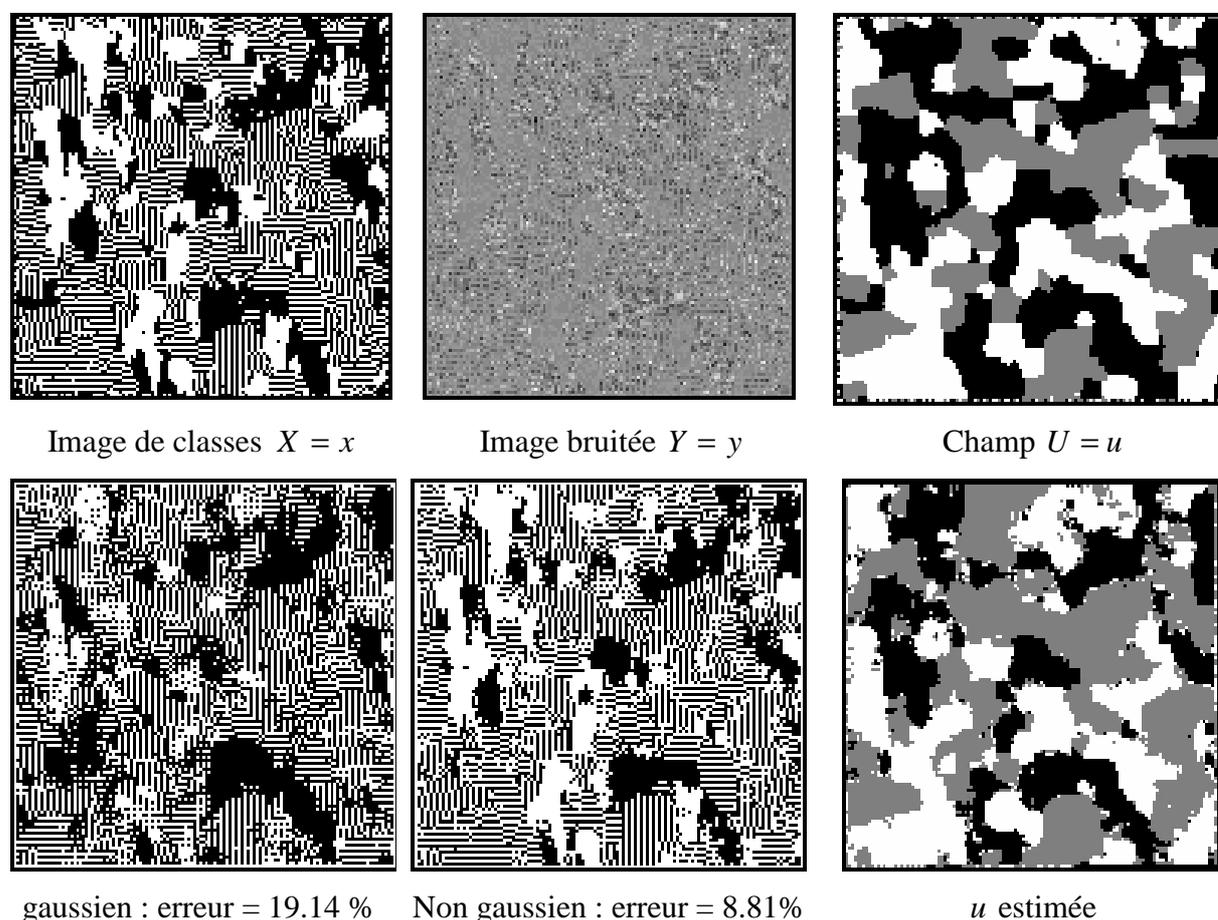


FIG. 2 – Résultats de segmentation non supervisée en utilisant les lois Gamma et Bêta, bruit markovien

Comme dans le cas précédent, nous constatons que l'utilisation des lois gaussiennes au lieu des vraies lois bêta et gamma peut dégrader considérablement la qualité de la segmentation. Par ailleurs, nous constatons visuellement que le cas considéré est très bruité, ce qui montre le bon comportement de la méthode non supervisée présentée.

Remarque

Notons que pour les deux expérimentations présentées, le nombre d'itérations de l'algorithme MPM est de 100, dont chacune contient 20 itérations dans l'échantillonneur de Gibbs utilisé. Pour l'estimation des paramètres par la NM généralisé, le nombre d'itérations utilisé est de 30.

V.5 Conclusion

Dans ce chapitre nous avons étudié la possibilité d'extension des champs de Markov triplet au cas du bruitage non gaussien. Afin de continuer à modéliser les images non stationnaires, notre choix a été porté sur les champs de Markov triplet adaptés (CMTA).

Nous avons proposé un modèle général, permettant de tenir compte de la forme des marginales du bruit quelconque et pouvant varier avec les classes. Nous avons également proposé une méthode d'estimation "*généralisée*", permettant de retrouver les formes des densités dans le système de Pearson. Les premières expériences montrent l'intérêt de la nouvelle modélisation et de la méthode d'estimation des paramètres proposées.

Notons que la méthode d'estimation des paramètres proposée généralise, d'une part, la méthode proposée dans le chapitre II [BeP 05] et d'autre part, celle proposée dans [DMP 97].

Conclusion et perspectives

Nous avons présenté dans cette thèse différentes variantes originales des champs de Markov triplets (CMT), ainsi que des méthodes d'estimation des paramètres correspondantes. Nous avons également étendu ces variantes au cas d'un bruit corrélé non gaussien et dont la nature peut varier selon les classes.

Dans le premier chapitre, nous avons rappelé quelques notions de base des champs markoviens, telles que les différentes méthodes de simulation, à savoir : l'échantillonneur de Gibbs, l'algorithme de Metropolis ou l'algorithme de Swendsen-Wang. Nous avons également exposé les différentes méthodes de segmentation, telle que la méthode " *Maximum des Marginales a Posteriori* " utilisée au cours de cette thèse. Le choix d'une méthode par rapport à une autre a été évoqué.

Dans le chapitre II, il a été traité les différentes modélisations markoviennes en commençant par les champs de Markov cachés (CMC) où trois niveaux de généralité ont été précisés. Le premier modèle communément utilisé dans la littérature est le champ de Markov caché à bruit indépendant (CMC-BI), où les variables aléatoires Y_s (du processus bruité) sont supposées indépendantes conditionnellement au champ caché X , et la loi de Y_s conditionnellement à X est égale à la loi de Y_s conditionnellement à X_s . Le deuxième modèle, qu'on appelle CMC-BC permet de modéliser un bruit corrélé. Le troisième modèle où les deux hypothèses du CMC-BI ne sont pas vérifiées, est original. Il permet encore l'application des méthodes bayésiennes de segmentation.

Ensuite, nous avons présenté un autre modèle plus général, appelé CMCCouple qui offre l'avantage de prendre en compte les corrélations et de modéliser les textures.

Enfin, nous nous sommes attardé sur les champs de Markov triplets (CMT) qui sont une généralisation des CMCCouples et qui permettent de modéliser les non stationnarités. Deux variantes des CMT ont été étudiées. La première a trait à la théorie de l'évidence. La deuxième a été obtenue par enrichissement de la fonction énergie classique, ce qui a permis de modéliser les champs classiques non stationnaires.

Dans le but d'une segmentation non supervisée, nous avons d'abord présenté dans le chapitre III différentes méthodes utilisées dans la littérature que nous avons groupées en deux catégories : le cas à données complètes et le cas à données incomplètes. Nous avons ensuite détaillé deux méthodes originales, appelées gradient stochastique adapté (GSA) et une nouvelle méthode (NM), fondée sur le principe de l'Estimation Conditionnelle Itérative (ECI). Une comparaison entre les deux méthodes a montré que dans le cas à données complètes les deux algorithmes sont comparables ; mais, dans le cas à données incomplètes les résultats fournis par la nouvelle méthode fondée sur ECI présentent des performances plus intéressantes ; ce qui nous a amené à opter pour celle-ci.

Le chapitre IV est une application des champs de Markov triplet en segmentation non supervisée d'images, en utilisant les techniques d'estimation introduites. L'étude de performances de l'approche proposée a montré que l'utilisation des CMT, en particulier les CMTA appliqués aux cas non stationnaires, permet une nette amélioration des résultats de segmentations obtenus avec les méthodes classiques.

Enfin, dans le chapitre V, nous avons étendu les diverses techniques développées au cas non gaussien corrélé et ce, en utilisant le système de Pearson et les mélanges généralisés.

Comme perspective, on peut envisager l'extension des nouveaux modèles proposés dans ce travail dans les diverses directions déjà explorées par les modèles classiques. On peut notamment envisager le traitement de cas multi-capteurs, avec des capteurs éventuellement évidentiels, comme suggéré dans [PiB 05]. Une extension vers l'imagerie 3D, avec applications en imagerie médicale, est une autre perspective possible. La recherche des différents cas nouveaux des champs de Markov triplets, comme par exemple des modèles avec le troisième processus à valeurs continues, offre des possibilités pour la poursuite de ce travail.

Annexe

Méthode SVD

Soit A une matrice réelle de dimension $m \times n$ ($m \geq n$), la décomposition en valeurs singulières "*Singular Value Decomposition*", exprime la matrice A sous la forme :

$$A = U \cdot S \cdot V^T \quad (\text{A.1})$$

U est une matrice orthogonale de dimension $m \times n$

V est une matrice orthogonale de dimension $n \times n$

$S = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix}$ est une matrice diagonale de dimension $n \times n$. Ses termes diagonaux s_i

tous positifs ou nuls, sont les valeurs singulières de A .

* Si la matrice A est *régulière* : tous les $s_i > 0$, on a :

$$A^{-1} = V \cdot S^{-1} \cdot U^T \quad (\text{A.2})$$

* Si la matrice A est *singulière* : on remplace pour chaque valeur singulière nulle, le terme $1/s_i$ par 0. On obtient le pseudo inverse de la matrice A à partir de l'équation (A.2).

Matrice de Toeplitz

Matrice hermitienne définie positive

On dit que la matrice carrée complexe A est hermitienne si :

$$A^H = A \quad (\text{A.3})$$

Dans le cas réel on dit que A est symétrique.

Une matrice hermitienne est définie positive (*resp.* semi- définie positive) si toutes ses valeurs propres sont strictement positives (*resp.* positives ou nulles).

Matrice de Toeplitz

Soit A une matrice réelle de dimension $n \times n$, A est une matrice de Toeplitz si $a_{i,j} = a_{i+1,j+1}$ ($a_{i,j}$ sont les éléments de la matrice). Autrement dit, les matrices de Toeplitz ont des éléments égaux sur leurs diagonales. Celles-ci s'écrivent donc sous la forme :

$$A = \begin{pmatrix} a & b & c & d & e \\ f & a & \ddots & \ddots & d \\ g & \ddots & \ddots & \ddots & c \\ h & \ddots & \ddots & a & b \\ i & h & g & f & a \end{pmatrix} \quad (\text{A.4})$$

***k*-means**

La méthode des k -means est une méthode itérative permettant de classer les pixels d'une image, en k classes ($k \in \Omega$), selon leur niveau de gris. Le pixel est affecté dans la classe pour laquelle la distance du centre de la classe au pixel est minimale. Les étapes de l'algorithme k -means sont décrites comme suit:

-
- Initialiser k noyaux (centres des classes) d'une manière aléatoire ;
 - A chaque itération n balayer l'image,
 - Affecter x_s à la classe ω_i si :

$$\|x_s - m(\omega_i)\| = \min_{l=1}^k \|x_s - m(\omega_l)\| \quad (\text{A.6})$$

- Recalculer le centre de chaque classe par :

$$m(\omega_k) = \frac{1}{|\Omega|} \sum_{x \in \Omega} x \quad (\text{A.7})$$

Liste de notations

$ $	Déterminant
α	Paramètre d'interaction
β	Paramètre associé aux singletons
η	Paramètres associés au bruit
θ	Vecteur Paramètre
C	Ensemble de cliques
c	Clique
Cr	Crédibilité
∇	Gradient
δ	Fonction de Dirac
E	Espérance mathématique
γ	Fonction de partition
Ω	Ensemble de classes
ω_i	Classe
φ	Fonction potentielle
M	Fonction de masse
$P(\Omega)$	Ensemble des parties de Ω
Pl	Plausibilité
R	Ensemble des réels

S	Ensemble des pixels
s	site ou pixel
U	Fonction énergie
U	Champ aléatoire
V	Système de voisinage
V_s	Ensemble de voisins de s
X	Champ caché
X_s, Y_s, U_s	Variables aléatoires
Y	Champ observé
BI	Bruit indépendant
CG	Champ de Gibbs
CM	Champ de Markov
CMC	Champ de Markov Caché
CMCouple	Champ de Markov Couple
CME	Champ de Markov Evidentiel
CMT	Champ de Markov Triplet
CMTA	Champ de Markov Triplet Adapté
DC	Données Complètes
DI	Données Incomplètes
DS	Dempster – Shafer
ECI	Estimation conditionnelle Iterative
EM	Expectation Maximisation
GS	Gradient Stochastique
GSA	Gradient Stochastique Adapté
GSADC	Gradient Stochastique Adapté à Données Complètes
GSADI	Gradient Stochastique Adapté à Données Incomplètes
ICM	Mode Conditionnel Itéré
MAP	Maximum a posteriori
MPM	Mode des Marginales à posteriori

NM	Nouvelle Méthode
NMDI	Nouvelle Méthode à Données Complètes
NMDI	Nouvelle Méthode à Données Incomplètes
ppv	plus proches voisins
SEM	Stochastic Expectation Maximisation

Bibliographie

- [ATA 95] K. Arai, Y. Terayama and T. Arata, “Image Classification Based on Bêta Distribution for SAR Image”, *IEEE International Geoscience and Remote Sensing Symposium*, Vol. 2, pp. 1263-1265, 1995.
- [BaD 01] N. Bartoli et P. Del Moral, “Simulation et algorithmes stochastiques une introduction avec applications”, 2001.
- [BaZ 05] A. Barbu and P. S. Zhun, “Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp. 1239-1253, 2005.
- [BCD 83] M. Broniatowski, G. Celeux et J. Diebolt, “Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabilistes”, *In Data Analysis and Informatics 3, E. Diday (Ed.)*, North Holland, Amsterdam, pp. 359-373, 1983.
- [BDF 01] 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.
- [BeP 04] D. Benboudjema and W. Pieczynski, “Parameter estimation in pairwise Markov fields”, *Advanced Concepts for Intelligent Vision Systems*, Aug. 31-Sept. 3, Brussels, Belgium, 2004.
- [BeP 05] D. Benboudjema and W. Pieczynski, “Unsupervised image segmentation using triplet Markov fields”, *Computer Vision and Image Understanding*, Vol. 99, No. 3, pp. 476-498, 2005.
- [BeP 05] D. Benboudjema and W. Pieczynski, “Segmentation Non supervisée d’images Non stationnaires avec champs de Markov evidentiels”, *20 colloque GRETSI*, Louvain-La-Neuve, Belgique, 6-9 septembre 2005.

- [BeP 05] D. Benboudjema and W. Pieczynski, "Segmenting Non stationary images with triplet Markov fields", *IEEE International Conference on Image Processing*, Genova, Italy, September 11-14, 2005.
- [BeP 95] B. Benmiloud, 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.
- [Bes 74] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *Journal of Royal Statistical Society*, B, 36, pp 192-236, 1974.
- [Bes 83] J. Besag, "Discussion of paper by P. Switzer", *Bulletion of the international Statistical Institute*, 50 (Bk. 3), pp 422-425, 1983.
- [Bes 86] J. Besag, "Statistical analysis of dirty picture", *Journal of Royal Statistical Society*, B, 48, pp 259-302, 1986.
- [BoS 94] C. A. Bouman et M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 3, pp. 162-177, 1994.
- [BPS 70] L. Baum, T. Petrie, G. Soules and N. Weiss, "A maximisation technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Ann. Math. Statist.*, 41, pp. 164-171, 1970.
- [CaS 05] Y. Cao and H. Sun, "An unsupervised Segmentation Method Based on MPM for SAR Images", *IEEE Geoscience and Remote Sensing Letters*, Vol. 2, No. 1, 2005.
- [CeD 90] G. Celeux and J. Diebolt, "Une version de type recuit simulé de l'algorithme EM", *C. R. Acad. Sci. Paris t. 310 Série I*, 119-124, 1990.
- [CFP 03] 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.
- [Cha 00] B. Chalmond, "Eléments de modélisation pour l'analyse d'images", *Springer*, 2000.
- [Cha 89] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images", *Pattern Recognition*, 22, pp. 747-761, 1989.
- [Cha 87] D. Chandler, "Introduction to modern statistical mechanics", *Oxford University Press.*, 1987.

- [DaG 98] M. Dang and G. Govaert, "Spatial fuzzy clustering using EM and Markov random fields", *International Journal of System Research and Information Science*, Vol. 8, No. 4, pp. 183–202, 1998.
- [DeE 87] H. Derin and H. Elliott, "Modeling and Segmentation of Noisy and textured images using Gibbs random fields", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 9 No. 1, pp. 39-55, 1987.
- [Del 97] 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.
- [Dem 67] A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Annals of Mathematical Statistics*, AMS-38, pp 325-339, 1967.
- [Dem 68] A. P. Dempster, "A generalization of Bayesian inference", *Journal of Royal Statistical Society*, 30, pp 205-247, 1968.
- [DeP 02] Y. Delignon and W. Pieczynski, "Modeling Non-Rayleigh Speckle Distribution in SAR Images", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 40, No. 6, pp. 1430-1435, 2002.
- [DLR 77] A. Dempster, N. Laird and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of Royal Statistical Society*, B-39, pp. 1-38, 1977.
- [DMP 97] Y. Delignon, A. Marzouki and W. Pieczynski, "Estimation of generalized mixture and its application in image segmentation", *IEEE Trans. on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.
- [DFM 98] Y. Dong, B. Forster and A. Milne, "Evaluation of Radar Image Segmentation by Markov Random Field Model with Gaussian Distribution and Gamma Distribution", *IEEE Proc. IGRASS*, Vol. 3, pp. 1617-1619, 1998.
- [DuP 86] D. Dubois and H. Prade, "On the unicity of Dempster's rule of combination", *Int. J. Intell. Systems*, Vol. 1, pp. 133-142, 1986.
- [EAM 95] R. J. Elliott, L. Aggoun and J.B. Moore, "Hidden Markov Models: Estimation and Control", *Springer-Verlag*, 1995.
- [EdS 88] R. G. Edwards and A. D. Sokal, "Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm", *Phys. Rev. D*, 38, pp. 2009-2012, 1988.
- [GeG 84] 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, pp. 721-741, 1984.

- [GIP 97] 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, 1997.
- [Gra 83] F. A. Graybill, "Matrices with applications in statistics", 2nd Edition, Wadsworth, Belmont C. A., 1983.
- [GRS 96] W. R. Gilks, S. Richardson and D. J. Spiegelhalter, editors. "Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics", *Chapman & Hall, London, Weinheim, New York, Tokyo, Melbourne, Madras*, 1996.
- [GuB 91] J. Guan and D. A. Bell, "Evidence theory and its applications", *Amsterdam, North-Holland*, 1991.
- [Guy 95] X. Guyon, "Random fields on network. modeling, statistics, and applications" Springer-Verlag, Probability and its Applications, New York, 1995.
- [HeB 93] F. Heitz and P. Bouthemy, "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields", *IEEE Trans. on Pattern Analysis Machine Intelligence*, Vol. 15, No 12, pp. 1217-1232, 1993.
- [HPB 94] F. Heitz, P. Perez 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.
- [Jai 81] A. K. Jain, "Advances in mathematical models for image processing", *Proc. IEEE*, Vol. 69, pp. 502-526, 1981.
- [Jai 89] A. K. Jain, "Fundamental of digital image processing", *PrenticeHall*, 1989.
- [Jan 97] F. Janez, "Rappels sur la théorie de l'évidence", *Note technique*, 1997.
- [KeH 98] 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.
- [KaV 94] G. K. Kaleh and R. Vallet, "Joint Parameter Estimation and Symbol Detection for linear or Nonlinear UnkNown Channels", *IEEE Trans. Commun.*, Vol. 42, No. 7, pp 2,406-2,413, 1994.
- [KGV 82] S. Kirkpatrick, C. D. Gellatt and M. P. Vecchi, "Optimization by simulated annealing", *IBM Thomas J. Watson research Center, Yorktown Heights, New York*, 1982.
- [LaK 91] A. M. Law, W. D. Kelton. "Simulation modeling and analysis", *Mc Graw-Hill book company*, 2nd edition, 1991.

- [LaP 05] P. Lanchantin and W. Pieczynski, "Unsupervised restoration of hidden Non stationary Markov chain using evidential priors", *IEEE Trans. on Signal Processing*, Vol. 53, No. 8, pp. 3091-3098, 2005.
- [Lev 48] P. Lévy, "Chaînes double de Markov et fonction aléatoires de deux variables", *Comptes Rendues de l'Académie des Sciences, Série I* 226, 53-55, 1948.
- [Li 95] S. Z. Li, "Markov Random Field Modelling in Computer Vision", *Springer*, 1995.
- [Mai 03] H. Maître, "Le traitement des images", *Hermès*, 2003.
- [MaW 91] A. Maffet and C. Wackerman, "The modified Beta Density Function as a Model for Synthetic Aperture Radar Clutter Statistics", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 29, No. 2, 1991.
- [McK 97] G. J. McLachlan and T. Krishnan, "EM Algorithm and Extensions", *Wiley, Series in Probability and Statistics*, 1997.
- [MCP 00] M. Mignotte, C. Collet, P. Pérez, 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.
- [MeW 02] 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.
- [MMP 87] J. Marroquin, S. Mitter, T. Poggio, "Probabilistic solution of ill-posed problems in computational vision", *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.
- [MMS 02] M. Mignotte, J. Meunier, J.-P. Soucy, C. Janicki, "Comparison of deconvolution techniques using a distribution mixture parameter estimation: application in SPECT imagery", *Journal of Electronic Imaging*, Vol. 11, No. 1, pp. 11-25, 2002.
- [MRR 53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equations of state calculations by fast computing machines", *J. Chem. Phys.*, 21, pp. 1087-1092, 1953.
- [MVR 01] J. L. Marroquin, F.A. Velasco, M. Rivera and M. Nakamura, "Gauss-Markov Measure Field Models for Low-Level Vision", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 337-348, 2001.

- [PeP 95] A. Peng, W. Pieczynski, "Adaptive Mixture Estimation and Unsupervised Local Bayesian Image Segmentation", *Graphical Models and Image Processing*, Vol. 57, No. 5, pp. 389-399, 1995.
- [Per 98] P. Perez, "Markov random fields and images". « *CWI Quarterly* », Vol. 11, No. 4, pp. 413-437, 1998.
- [PBL 02] 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.
- [PiB 05] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence", *Image and Vision Computing*, accepted, September 2005.
- [Pie 92] W. Pieczynski, "Statistical image segmentation", *Machine Graphics and Vision*, Vol.1, No. 1/2, pp. 261-268, 1992.
- [Pie 03] W. Pieczynski, "Modèles de Markov en traitement d'images", *Traitement du Signal*, Vol.20, No. 3, pp. 255-278, 2003.
- [PiT 00] 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.
- [Pot 52] R. B. Potts, "Some generalized order-disorder transitions", *Proc. Camb. Phil. Soc.*, 48, pp. 106-109, 1952.
- [PSZ 05] G. Poggi, G. Scarpa, and J. Zerubia, "Supervised segmentation of remote sensing images based on a tree-structured MRF model", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 43, No. 8, pp. 1901-1911, 2005.
- [Ric 05] J. A. Richards, "Analysis of remotely sensed data : the formative decades and the future", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 43, No. 3, pp. 422-432, 2005.
- [Rob 91] C. Robert, "L'analyse statistique Bayésienne", Economica, 1991.
- [RPB 03] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar", *IEEE Journal of Oceanic Engineering*, Vol. 28, No. 1, pp. 90-105, 2003.
- [Swi 69] P. Switer, "Mapping a geographically correlated environment (with discussion)", *In International Symposium on statistical Ecology: Spatial Patterns and Statistical Distribution*, Vol 1, pp. 235-269, 1969.
- [Sha 76] G. Shafer, "A mathematical theory of evidence", *Princeton University Press, New Jersey.*, 1976.

- [Sme 90] P. Smets, "The combination of evidence in the transformable belief model", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, pp. 447-458, 1990.
- [SwW 87] R. H. Swendsen and J.S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations", *Physical Review Letters*, 58, pp. 86-88, 1987.
- [Tan 96] M. A Tanner, "Tools for Statistical Inference: Methods for the exploration of Posterior distributions and likelihood functions", 3rd Edition, *Springer*, 1996.
- [TNT 04] C. Tison, J.M. Nicolas, F. Tupin and H. Maître, "A new statistical model of urban areas in high resolution SAR images for Markovian segmentation", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 42, No. 10, pp. 2046-2057, 2004.
- [TuR 05] F. Tupin and M. Roux, "Markov random field on region adjacency graphs for the fusion of SAR and optical data in radar grammetric applications", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 43, No. 8, pp. 1920-1928, 2005.
- [Win 03] G. Winkler, "Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A mathematical introduction", *Springer*, 2003.
- [YKF 94] R. R. Yager, J. Kacprzyk and M. Fedrizzi, "Advances in the Dempster-Shafer theory of evidence", *John Wiley & Sons, New York*, 1994.
- [You 88] L. Younes, "Estimation and annealing for Gibbsian fields", *Annales de l'institut Henri Poincaré*, 2, 1988.
- [You 89] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields", *Probability Theory and Related Fields*, 82, pp. 652-645, 1989.
- [Zad 79] L. A. Zadeh, "On the validity of Dempster's rule of combination of evidence", *Memo UCB/ERL 79/24, University of California*, 1979.
- [Zha 92] J. Zhang, "The mean field theory in EM procedures for Markov random fields", *IEEE Trans. on Image. Processing*, Vol. 40, No.10, pp. 2570-2583, 1992.

Résumé

La segmentation non supervisée d'images est parmi les problèmes clé en traitement d'images. Parmi les différents modèles et approches développées, une partie des méthodes statistiques, couramment utilisées, sont fondées sur le modèle par champs de Markov Cachés (CMC). Ce succès est principalement dû à l'aptitude du modèle de prendre en compte des dépendances spatiales des variables aléatoires, même lorsqu'elles sont en très grand nombre, pouvant dépasser le million. Dans un tel modèle le champ caché X est supposé markovien et doit être estimé à partir du champ observé Y . Un tel traitement est possible du fait de la markovianité de X conditionnellement à Y . Ce modèle a été ensuite généralisé aux champs de Markov couples (CMCouples), où l'on suppose directement la markovianité du couple (X, Y) , qui offrent les mêmes possibilités de traitements que les CMC et permettent de mieux modéliser le bruit ce qui permet, en particulier, de mieux prendre en compte l'existence des textures. Par la suite, les CMCouples ont été généralisés aux champs de Markov triplet (CMT), où la loi du couple (X, Y) est une loi marginale d'un champ de Markov triplet $T = (X, U, Y)$, avec un champ auxiliaire U . L'objet de cette thèse est d'étudier les CMT. Deux variantes originales sont présentées : les champs de Markov évidentiels (CME), permettant la modélisation des incertitudes sur les paramètres ; et les champs de Markov Triplets Adaptés (CMTA), permettant la modélisation des différentes stationnarités de l'image cachée. Pour une segmentation non supervisée, deux méthodes originales d'estimation des paramètres sont proposées. La première est fondée sur le principe du gradient stochastique, et la seconde est fondée sur le principe de l'estimation conditionnelle itérative (ECI) et les moindres carrés. Cette dernière est ensuite généralisée aux cas des images non stationnaires avec du bruit non Gaussien corrélé, et à marginales quelconques. Elle permet de rechercher la forme même des différentes lois marginales en utilisant le système de Pearson, ainsi que d'estimer tous les paramètres du modèle. L'intérêt des différentes méthodes non supervisées obtenues est attesté par des simulations informatiques ainsi que par les premiers traitements des images réelles.

Mots clés : Champs de Markov Cachés (CMC), Champs de Markov Couples (CMCouples), Champs de Markov Triplets (CMT), Champs de Markov Evidentiels (CME), segmentation statistique non supervisée, estimation des paramètres, Estimation Conditionnelle Itérative (ECI), Gradient Stochastique (GS), Système de Pearson, images non stationnaires.

Abstract

Image segmentation is a fundamental and yet difficult task in machine vision. Several models and approaches have been proposed, and the ones which have probably received considerable attention are hidden Markov fields (HMF) models. In such model the hidden field X which is assumed Markovian, must be estimated from the observed –or noisy– field Y . Such processing is possible because the distribution X conditional on the observed process Y remains markovian. This model has been generalized to the Pairwise Markov field (PMF) which offer similar processing and superior modelling capabilities. In this model we assume directly the markovianity of the couple (X, Y) . Afterwards, triplet Markov fields (TMF) which are the generalization of the PMF, have been proposed. In such model the distribution of the couple (X, Y) is the marginal distribution of a Markov field $T = (X, U, Y)$, where U is latent process.

The aim of this thesis is to study the TMF models. Two original models are presented: the Evidential Markov field (EMF) allowing to model the evidential aspects of the prior information and the adapted triplet Markov field (ATMF), allowing to model the simultaneous presence of different stationarities in the class image. For the unsupervised processing, two original approaches of estimation the model's parameters have been proposed. The first one is based on the stochastic gradient and the second one is based on the iterative conditional estimation (ICE) and the least square method, as well. The latter, have then been generalized to the non stationary images with non Gaussian correlated noise, which uses the Pearson system to find the natures of margins of the noise, which can vary with the class. Experiments indicate that the new models and related processing algorithms can improve the results obtained with the classical ones.

Index Terms: Hidden Markov Field (HMF), Pairwise Markov Field (PMF), Triplet Markov Field (TMF), Evidential Markov Field (EMF), unsupervised statistical segmentation, Parameters estimation, Iterative conditional estimation (ICE), Stochastic Gradient (SG) Pearson system, non stationary images.