

REPUBLIQUE ALGERIENNE DEMOCRATIQUE POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABDELHAMID IBN BADIS DE MOSTAGANEM



UNIVERSITE
Abdelhamid Ibn Badis
MOSTAGANEM

FACULTE DES SCIENCES ET DE LA TECHNOLOGIE
DEPARTEMENT GENIE ELECTRIQUE

N° D'ORDRE : M.../2013

MEMOIRE

Pour l'obtention du

DIPLOME DE MAGISTER

SPECIALITE : ELECTRONIQUE
OPTION : SIGNAUX ET APPLICATIONS

Thème

**Segmentation d'image médicale par
champs de Markov**

Soutenu en **2013** par :

Mr. kharroubi Mohamed Amine

Devant le jury composé de :

Président :	Mr .A. Ahmadouche	Maître de conférences A	université de Mostaganem
Examineur:	Mr. A. Benotsmane	Maître de conférences A	université de Mostaganem
Rapporteur :	Mme. M. MIMI.	Professeur	université de Mostaganem

REMERCIEMENTS

Avant tout je remercie le puissant ,le grand , le roi Allah elhamdo wechokro leh .

Je tiens a remercier mon encadrant Prof. Mimi Malika pour la considération et l'intérêt qu'elle a porté a mon travail. Je lui suis tres reconnaissant de m'avoir fait profiter de son expertise en traitement statistiques d'image. Je remercie également *M^{elle}* Mehidi Aicha pour son énorme aide informatique. Je tiens également a remercier toute l'équipe de l'option signaux & applications .

Je tiens également à remercier tous les membres de jury de m'avoir accepter de juger mon travail.

Sommaire

Sommaire

Résumé	A
Abstract	B
Listes des tableaux	C
Listes des figures	E
Liste des symboles	F
Introduction Générale	1
Chapitre I : Position du problème	6
Chapitre II : Modélisation markovienne	30
Chapitre III : Segmentation d'image	56
Chapitre IV : Simulation et résultats	66
Chapitre V : Conclusion Générale	83
Annexe	85
Bibliographie	90.

Tables de matières

Résumé.....	A
Abstract.....	B
Listes des tableaux	C
Listes des figures	E
Liste des symboles	F
Introduction générale	2
Chapitre I	
I.1 Introduction.....	6
I.2 Vision par ordinateur.....	6
I.2.1. problématique de l'interprétation d'image.....	7
I.2.2 La gestion de l'incertitude et de l'imprécision.....	8
I.3 Segmentation d'images.....	9
I.3.1 Principe de segmentation	9
I.4 Méthodes de segmentation d'images.....	11
I.4.1 Méthodes de détection de discontinuité.....	11
I.4.2 Méthodes de détection de similarité.....	11
I.5 Segmentation par classification.....	12
I.5.2 Quelques méthodes de classification.....	12
I.5.2.1 Classification hiérarchique.....	13
I.5.2.2 Classification non hiérarchique (partitionnelle)	13
I.5.3 La classification par les k-moyennes.....	13
I.5.3.1 Les nuées dynamiques.....	13
I.5.3.2 K-means ou classification binaire.....	15
I.5.4 La classification floue.....	16
I.5.4.1 La logique floue.....	16
I.5.4.2 La théorie des sous ensembles flous.....	17
I.5.4.4 Travaux utilisant la segmentation par FCM.....	18

Tables de matières

I.5.5 Quelques mots sur la segmentation par classification.....	19
I.6 Segmentation contextuelle.....	19
I.6.1 Segmentation par croissance de région.....	20
I.6.1.1 Présentation de la croissance de région.....	20
I.6.1.2 Algorithme général.....	21
I.6.1.3 Propriétés de la croissance de région.....	22
I.6.2 Approche contextuelle par morphologie mathématique.....	23
I.6.2.1 Outils de la morphologie mathématique.....	24
I.7 Coopération en segmentation d'image.....	26
I.7.1 Coopération séquentielle.....	27
I.7.2 Coopération itérative.....	27
I.8 Généralités sur les champs de Markov cachés.....	27
I.9 Conclusion.....	28
Chapitre II	
II.1 Introduction.....	30
II.1 Champ de Markov et distribution de Gibbs.....	31
II.1.1 Système de voisinage et Système de voisinage.....	32
II.1.2 Champs de Markov.....	34
II.1.3 Champs de Gibbs.....	34
II.1.4 Equivalence champ de Markov – champ de Gibbs.....	35
II.1.5 Exemple: Modèle de Potts.....	36
II. 2 Simulation d'un champ de Markov.....	38
II.2.1 Algorithme de Metropolis.....	38
II.2.2 Echantillonneur de Gibbs.....	39
II.2.2.1 Rôle du choix des cliques.....	44
II.2.2.2 Rôle de la constante de couplage	44
II.2.2.3. Conclusion.....	44

Tables de matières

II.2.3 Algorithme de Swendsen – Wang.....	45
II.2.4 Conclusion.....	46
II. 3 Modèle markovien caché pour la segmentation d'images.....	46
II. 4 Apprentissage.....	47
II. 5 Estimation bayésienne.....	48
II.5.1 Estimateur Maximum A Posteriori (MAP).....	50
II.5.2 Algorithme du recuit simulé.....	52
II.5.3 Estimateur du Mode Conditionnel Itéré (ICM).....	53
I.5.4 Estimateur du Mode des Marginales à posteriori (MMP).....	54
II. 6 Conclusion.....	54
Chapitre III	
III .1. Introduction.....	56
III.2. Synoptique de la segmentation.....	56
III.3 Initialisation.....	57
III.4. Estimation des paramètres.....	58
III.4.1. Principe de l'algorithme ICE.....	58
III.4.2. L'algorithme ICE.....	59
III.3.3. L'algorithme MPM pour la segmentation d'image.....	61
III.3.4. Simulation de l'algorithme MPM pour la segmentation d'image.....	61
III.4 Conclusion.....	64
Chapitre IV	
IV.1 Introduction.....	66
IV.2 Critères d'évaluation.....	66
IV.2 .1 Le rapport signal sur bruit crête (PSNR).....	66
IV.2 .2 Résultats sur d'images synthétiques.....	67
IV.2 .3 Analyse et discussion.....	67
IV.2 .4 Conclusion.....	68

Tables de matières

IV.3 Présentation des images par résonance magnétique nucléaire (RMN).....	69
IV.4.1 Présentation des résultats.....	72
IV.4.2 Analyse et discussion.....	73
IV.5 Conclusion.....	81
Conclusion générale.....	82
Annexes.....	83
Bibliographie.....	88

Résumé

La segmentation (détection des contours de textures) 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 est supposé markovien et doit être estimé à partir du champ observé. Un tel traitement est possible par l'analyse de la markovianité de conditionnellement. 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.

Mots Clés : Champs de Markov, détection, contour, Région, segmentation, images.

ABSTRACT

Segmentation (edge detection textures) images is one of the key problems in image processing. Among the various models and approaches developed, some of the statistical methods commonly used are based on the model by Hidden Markov fields (CMC). This success is mainly due to the model's ability to take into account the spatial dependencies of random variables, even when in large numbers, which can exceed one million. In this model the hidden field is assumed Markov and must be estimated from the observed field. Such treatment is possible by analysis of markovianité conditionally. This model was then generalized to Markov fields couples (CMCouples), where it is assumed directly markovianité the pair (X, Y) that offer the same treatment options as CMC and provide a better noise model which allows in particular, to better take into account the existence of textures.

Keywords: Markov random fields, detection, contour, region segmentation, images.

Listes des tableaux

Tab. IV.1 - Caractéristiques des images utilisées	72
Tab. IV.2 – informations sur la segmentation d’image « coupe axiale du cerveau »	74
Tab. IV.3– informations sur la segmentation d’image « coupe sagittale du cerveau ».....	75
Tab. IV.4 – informations sur la segmentation d’image « coupe abdominale ».....	76
Tab. IV.5 informations sur la segmentation d’image « coupe frontale du cerveau».....	77
Tab. IV.6 – informations sur la segmentation d’image « coupe de genou ».....	78
Tab. IV.7 – informations sur la segmentation d’image «Moelle épinière».....	79
Tab. IV.8 – informations sur la segmentation d’image «mammographie».....	80

Listes des figures

FIG. I.1: Principe de la croissance de région sur une image 2D	19
FIG. I.2: Segmentation d'une image synthétique.....	21
FIG. I.3: Dilatation par un carré 3x3, Érosion par un carré 3x3.....	24
FIG. I.4: Fermeture par un carré 3x3, Ouverture par un carré 3x3	25
FIG. I.5: Édition de l'encéphale.....	25
FIG. II.1 – Cliques associées à un système de voisinage en 4-connexité et en 8-connexité	32
FIG II.2 orientation d'un spin à l'état d'équilibre	36
FIG.II. 3 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs pour différentes valeurs de β_V & β_H	39
FIG. II.4– Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs pour différentes valeurs de β	40
FIG.II. 5 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs a posteriori pour différentes valeurs de β_V & β_H	41
FIG.II. 6 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs a posteriori pour différentes valeurs de β	42
FIG. II.7 – Evolution de la fonction d'énergie en fonction de la température	51
FIG III.1 organigramme de la segmentation d'image.....	56
Fig III.2 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm $\beta_h = \beta_v = 0.5$	62
Fig III.3 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm $\beta_h = \beta_v = 0.5$	62
Fig III.3 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm $\beta = 0.8$	63
Fig III.3 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm $\beta = 0.8$	63
FIG.IV.1 - (a) Image synthétique originale; (b) Image bruitée	67
FIG.IV.2 Résultat de segmentation en niveau de gris de l'image synthétique	68

Listes des figures

FIG.IV.3: Les images de simulation	70-71
FIG. IV.4 : Résultat de segmentation de l'image « coupe axiale du cerveau»	74
FIG. IV.5 : Résultat de segmentation de l'image « coupe sagittale du cerveau»	75
Fig. IV.6 : Résultat de segmentation de l'image « coupe abdominale».....	76
Fig. IV.7: Résultat de segmentation de l'image « coupe frontale du cerveau».....	77
Fig. IV.8: Résultat de segmentation de l'image « coupe de genou ».....	78
Fig. IV.9: Résultat de segmentation de l'image «Moelle épinière».....	79
Fig. IV.10: Résultat de segmentation de l'image «mammographie».....	80

$ $	Déterminant
α	Paramètre d'interaction
β	Paramètre associé aux singletons
η	Paramètres associés au bruit
θ	Vecteur Paramètre
C	Ensemble de cliques
c	Clique
C_r	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_s	Système de voisinage
V	Ensemble de voisins de s
X_s, Y_s, U_s	Champ caché
U	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 GSADI Gradient Stochastique Adapté
$ICM Mode$	Conditionnel Itéré
MAP	Maximum a posteriori
MPM	Mode des Marginales à posteriori

Introduction générale

Introduction

L'imagerie médicale a révolutionné la médecine en donnant un accès immédiat et fiable à des informations invisibles au diagnostic clinique. Elle fournit un volume très important de données à analyser. Des méthodes automatiques de traitement et analyse d'images se sont multipliées ces dernières années dans le but d'aider l'expert dans son interprétation et son diagnostic. Ces méthodes doivent prendre en considération d'une part la quantité de données à analyser et d'autre part la complexité structurelle des images médicales.

L'imagerie médicale regroupe l'ensemble des techniques utilisées par la médecine pour le diagnostic mais aussi pour le traitement d'un grand nombre de pathologies. Parmi ces techniques nous pouvons citer : L'image mammographie (1989), l'image à résonance magnétique nucléaire (RMN en 1971), le scanner hélicoïdal (CT en 1990), la tomographie (1927) et d'autres modalités d'imagerie médicale fournissent des moyens efficaces pour la configuration de l'anatomie d'un sujet, la détection d'anomalies, et à faire des mesures sans intervention directe. Ces technologies ont augmenté la connaissance des anatomies normales et malades pour la recherche médicale. Ils sont un composant critique dans le diagnostic et la planification du traitement.

Avec la croissance de la taille et du nombre d'images médicales, l'utilisation d'ordinateurs pour faciliter leur traitement et leur analyse est devenue nécessaire. En particulier les algorithmes de segmentation d'image qui jouent un rôle crucial dans plusieurs applications d'imagerie médicales en automatisant la délimitation de structures anatomiques et d'autres régions d'intérêt.

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 mémoire, 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 $\mathcal{U} = \{u_1, \dots, u_k\}$ 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 $X = (X_s)_{s \in S}$.

Introduction générale

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 réalisations 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

Le plan de le mémoire est le suivant :

Le premier chapitre est introductif et présente la problématique dans sa généralité. Nous commençons par définir la nature et les objectifs de l'opération de segmentation, ainsi que les notations utilisées au cours des chapitres suivants et une présentation de différents modèles de segmentation qui ont suscité de nombreux travaux dans ce domaine au cours de ces dernières

Introduction générale

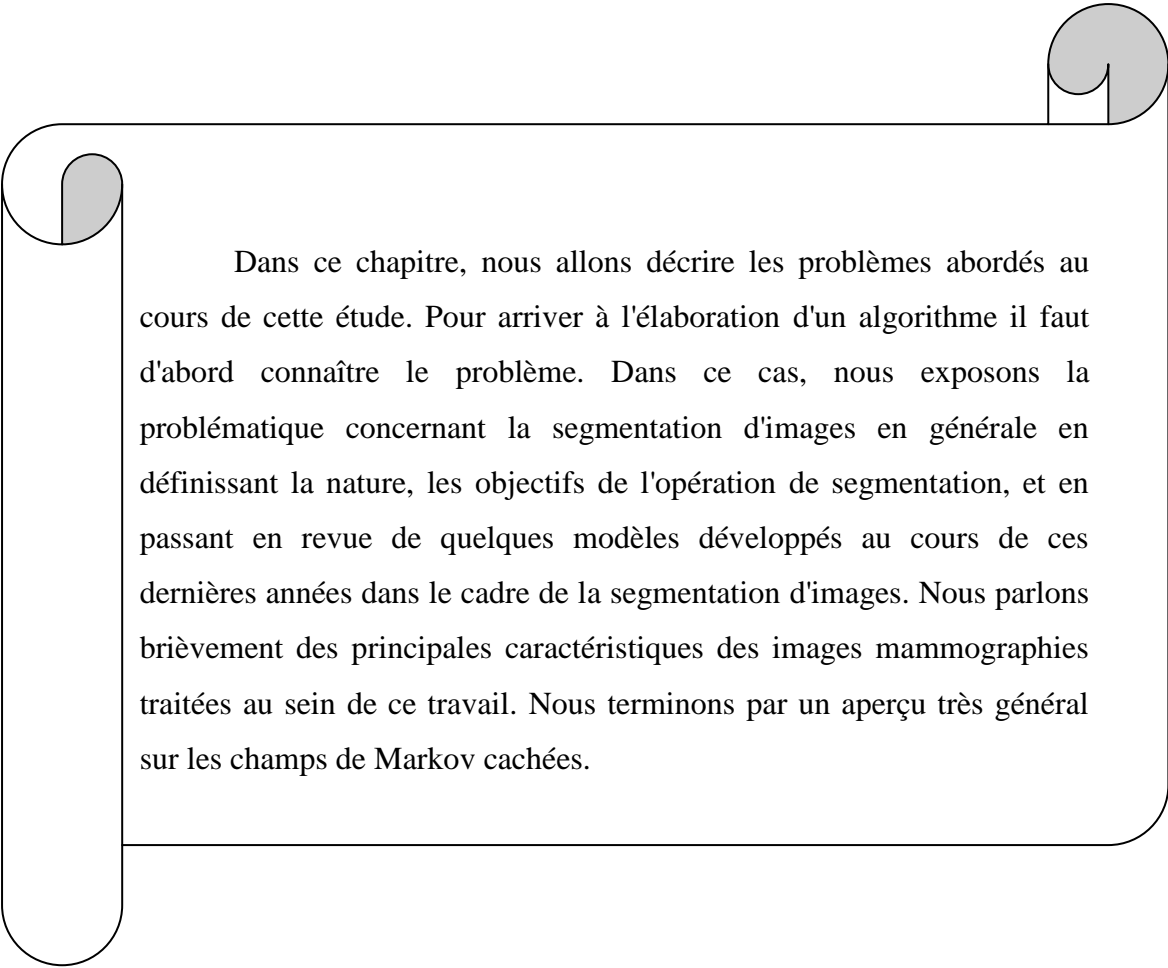
années. Nous introduisons également les images mammographie sur lesquelles nous appliquons le modèle de segmentation markovien. Ce chapitre s'achève par un aperçu très général sur les champs de Markov cachés.

Le deuxième chapitre présente le modèle de segmentation par les champs de Markov cachés. Au début nous précisons les notations qui y sont attachées. Nous développons ensuite le modèle markovien pour la segmentation d'images. Le modèle que nous définissons ici met en jeu le contexte spatial de l'image. Le modèle de champs de Markov cachés permet d'avoir une interprétation des interactions tant locales que globales dans une grille de variables aléatoires.

Le chapitre III est consacré aux problèmes de segmentation d'image et l'apprentissage non supervisé des paramètres pour les champs de Markov cachés. Nous présentons deux algorithmes originaux. Pour le premier sujet, il s'agira du mode des marginales à posteriori (MPM) qui consiste à estimer les informations cachées. Quant à la seconde question, nous adapterons l'algorithme d'espérance conditionnelle itérative (ICE) à l'estimation des paramètres a priori et d'attaches aux données. Ici, c'est une étape supérieure et un niveau de difficulté accru, car il s'agit de réaliser le même objectif que la segmentation supervisée mais en estimant conjointement les paramètres.

Dans le chapitre IV, Nous segmentons d'images médicales.

Ce manuscrit s'achève par un récapitulatif des résultats obtenus avec le modèle de champ et par l'évocation de quelques perspectives et extensions possibles de nos travaux que nous envisageons dans l'avenir.



Dans ce chapitre, nous allons décrire les problèmes abordés au cours de cette étude. Pour arriver à l'élaboration d'un algorithme il faut d'abord connaître le problème. Dans ce cas, nous exposons la problématique concernant la segmentation d'images en générale en définissant la nature, les objectifs de l'opération de segmentation, et en passant en revue de quelques modèles développés au cours de ces dernières années dans le cadre de la segmentation d'images. Nous parlons brièvement des principales caractéristiques des images mammographies traitées au sein de ce travail. Nous terminons par un aperçu très général sur les champs de Markov cachées.

I.1 Introduction

Quelque soit son origine une image constitue une représentation d'un univers composé d'entités : objets dans une scène intérieure, cellules, surfaces sismiques, organes du corps humain... Le but de toute méthode de segmentation est l'extraction d'attributs caractérisant ces entités. Les attributs étudiés correspondent à des points d'intérêt ou à des zones caractéristiques de l'image.

La diversité des images, la difficulté des problèmes, les origines variées des chercheurs, l'évolution de la puissance de calcul des ordinateurs, et un certain empirisme dans l'évaluation des résultats ont conduit à l'introduction d'une multitude d'algorithmes.

La segmentation fiable et précise des volumes anatomiques (normaux ou pathologiques) issus des systèmes d'imagerie reste un objectif important en traitement de l'information médicale car elle constitue le premier maillon de la chaîne d'analyse aboutissant à l'étude de la morphologie des structures internes, et la détection et quantification des lésions. Un grand nombre de méthodes de segmentation ont été proposées depuis de nombreuses années pour ce type d'images.

Des synthèses de méthodes de segmentation d'images médicales sont nombreuses et sont présentées dans la littérature [Har05]. Nous limitons notre présentation aux différentes méthodes utilisées pour notre travail, suivie de travaux antérieurs, utilisant ces méthodes en imagerie médicale.

La segmentation d'images constitue la pierre de base d'un système de vision et nous commençons ce chapitre par une brève présentation de la vision par ordinateur.

I.2 Vision par ordinateur

La vision par ordinateur est un domaine technologique très dynamique et en pleine effervescence, tant du point de vue de la recherche scientifique que des applications grand public.

La vision est un processus de traitement de l'information. L'entrée d'un système de vision est constituée par une séquence d'images. Le système lui-même apporte un certain nombre de connaissances qui interviennent à tous les niveaux. La sortie est une description de l'entrée en termes d'objets et de relation entre ces objets.

L'analyse d'image est généralement vue sous deux angles : les traitements de bas niveau et les traitements de haut niveau (interprétation). Cette distinction est

une fonction du contenu sémantique des entités traitées et extraites de l'image. Les traitements de bas niveau opèrent en général sur des grandeurs calculées à partir des valeurs attachées à chaque point de l'image sans faire nécessairement la liaison avec la réalité qu'elles représentent. Ces traitements manipulent des données de nature numérique, contrairement aux traitements de haut niveau qui opèrent sur des entités de nature symbolique associées à une représentation de la réalité extraite de l'image : ils sont relatifs à l'interprétation et à la compréhension de l'image exprimée avec des mots du vocabulaire et de l'application.

Remarque

La majorité des opérateurs de base peuvent être classés en trois catégories : seuillage, détection de contours, extraction de régions qui sont des modèles de base permettant la conception de méthodes bas niveau de segmentation et de description des images.

Parmi les inconvénients de la vision de bas niveau, on peut citer par exemple la difficulté de segmenter une image en primitives bas niveau sans connaître le domaine d'application de l'image et l'utilisation que l'on veut faire de cette segmentation.

I.2.1. problématique de l'interprétation d'image

L'interprétation des images correspond à l'analyse d'une image ou d'une scène permettant de décrire les objets composant la scène et leurs relations, c'est-à-dire extraire la sémantique de l'image, afin de mieux la comprendre. La problématique de l'interprétation d'image, est un problème de perception de l'environnement et de prise de décision.

Les systèmes de vision doivent œuvrer au sein d'environnements partiellement connus. Pour effectuer des choix, des mécanismes spécifiques doivent être introduits reposant notamment sur l'évaluation de la confiance dans les hypothèses courantes, en fonction de nombreux paramètres tels que la qualité du traitement, le choix des opérateurs, la présence de descripteurs significatifs pour les objets. Des modèles a priori, décrivant les entités présentes dans les images, doivent être définis et combinés avec la mise en place de coopérations entre différents types de traitements [Har05].

La première théorie de la vision numérique a été proposée dans [Mar82] et consiste en une architecture à trois niveaux que tout système de traitement de

l'information doit respecter, pour demeurer cohérent. Ces travaux ont inspiré la plupart des systèmes de traitement de l'information. Les différents niveaux proposés sont les suivants :

- un niveau abstrait : le « quoi » et le « pourquoi », c'est-à-dire que doit-on faire? Ce qui revient à mettre en place une méthodologie ou une théorie en relation avec les données en entrée;
- un niveau de la représentation : le comment? En structurant les données et en concevant les algorithmes;
- un niveau de réalisation : l'implantation des algorithmes, reliée au matériel.

I.2.2 La gestion de l'incertitude et de l'imprécision

Les informations manipulées par le système sont incomplètes, imprécises et potentiellement erronées. Pour pouvoir les utiliser afin de produire de nouvelles informations de façon robuste, c'est-à-dire pour pouvoir prendre des décisions, il est nécessaire d'évaluer et de gérer leur incertitude. Différents formalismes numériques modélisant l'incertain ont été proposés, tels que les approches probabilistes et possibilistes qui permettent d'introduire des mécanismes d'inférence de façon à tirer des conclusions provisoires (dont l'incertitude est quantifiée), nécessaires à la progression de l'interprétation et à l'estimation des modèles.

En effet, le modèle de traitement de l'incertitude le plus connu est celui basé sur la théorie des probabilités. Néanmoins, la notion de probabilité n'est pas très bien adaptée à traiter d'autres aspects du raisonnement humain, comme l'imprécision et la confiance.

En ce qui concerne la modélisation de l'imprécision, la notion d'intervalle d'erreur (utilisée par exemple en physique pour décrire l'erreur de mesure) est la plus ancienne. Par la suite, d'autres notions comme celle du degré de confiance introduite dans la théorie de Dempster-Shafer et celle de degré d'appartenance de la théorie des sous-ensembles flous sont venues compléter la modélisation.

Les innombrables travaux sur la théorie des sous-ensembles flous en ont fait un modèle puissant pour modéliser l'imprécision. Zadeh [Zad65] en introduisant les sous-ensembles flous a rapproché la notion d'intervalle et le raisonnement humain en considérant plutôt des frontières graduelles que des frontières abruptes (nous

aborderons le sujet des sous-ensembles flous dans les prochains paragraphes).

Par exemple, en imagerie par résonance magnétique, la traduction, «chaque individu appartient à une classe» par «chaque pixel appartient à une structure» n'est pas possible. Comme nous l'avons vu dans le premier chapitre, l'effet de volume partiel crée des pixels de mélange (région intermédiaire) qui sont constitués de plusieurs structures. Les données sont alors imprécises [Har05].

I.3 Segmentation d'images

La segmentation d'images est une opération présente à la base de nombreuses chaînes de traitements et d'interprétation de données. On distingue en général deux niveaux dans l'analyse d'images [Coc et Phi, 1995] : les traitements de bas niveau et ceux de haut niveau. Dans les traitements de bas niveau, il s'agit d'extraire des attributs de l'image à partir des grandeurs physiques observées, que ces attributs soient réalisés ou non. Une fois ces attributs de bas niveau extraits, ils peuvent être interprétés par un traitement de haut niveau. Le processus de segmentation, dont il est question au cours de ce travail, appartient aux approches de type bas niveau.

I.3.1 Principe de segmentation

L'information véhiculée par une image est portée par bien d'autres données que les seuls niveaux de gris retournés par chaque pixel. Cette grandeur physique est la plus simple, et la plus souvent utilisée, mais les modèles de segmentation que nous utilisons peuvent s'étendre à des critères différents comme des attributs de texture par exemple, le tout étant que les attributs considérés puissent être caractérisés par des valeurs numériques. Une fois ces attributs caractérisés, le principe de la segmentation consiste à réaliser une partition de l'image I en sous-ensembles R_i , appelés régions. Ainsi, les ensembles qui composent une segmentation doivent satisfaire :

$$\forall i, R_i \neq \phi$$

$$\forall i \neq j, R_i \cap R_j = \phi$$

$$I = \bigcup_{i=1}^k R_i$$

Une région est un ensemble de pixels connexes ayant des propriétés communes qui les

différencient des pixels des régions voisines [Coc et Phi, 1995]. Aujourd'hui, vu la simplicité de la définition de l'objectif de la segmentation d'images et la richesse des travaux effectués dans ce domaine, on pourrait penser que la segmentation d'images est un problème en grande partie résolu. Il n'en est rien, probablement à cause de la très grande diversité des applications et des approches mathématiques mises en œuvre sur des ordinateurs toujours plus puissants permettant l'exploration de nouvelles approches autrefois trop coûteuses en temps de calcul.

De nombreuses méthodes ont été développées pour la segmentation d'images [Fu et Mui, 1981]. Deux familles importantes sont : la segmentation par recherche de frontières et celle par recherches de régions.

Les méthodes par recherche de frontières sont des méthodes de détection de contours correspondant à une variation d'intensité ou à une discontinuité entre deux ensembles connexes de pixels. Ensuite, les contours obtenus doivent être fermés de façon à obtenir une partition de l'ensemble des pixels de l'image.

L'approche pour la segmentation par recherche de régions consiste à regrouper les pixels ayant des propriétés communes et à leur attribuer une même étiquette.

Le processus de segmentation nécessite de la détermination des paramètres de classes caractérisant les différents attributs de l'image. Lorsque les paramètres sont connus au préalable (estimés par un algorithme ou donnés par un opérateur), nous parlerons de segmentation supervisée, au contraire de la segmentation non supervisée où les paramètres sont déterminés au sein même de la segmentation,

Les segmentations supervisées consistent à déterminer des frontières de décision linéaires ou non-linéaires afin de segmenter les données. Les méthodes de segmentation linéaires ne sont généralement pas applicables à des données non linéairement séparables, puisque les frontières de décision linéaires obtenues par ces méthodes ne prennent pas correctement en compte la répartition souvent complexe des données.

Les méthodes non- supervisées sont très intéressantes car elles ne supposent pas d'étapes d'apprentissage ou la mise à disposition d'un ensemble de données préalablement étiquetées. De plus elles ne se heurtent pas au problème de généralisation et/ou de pertinence de l'ensemble d'apprentissage. Cependant leur utilisation est souvent délicate et spécifique à un type de traitement.

I.4 Méthodes de segmentation d'images

La segmentation permet de distinguer les objets d'intérêt d'une image en les isolants du fond ou des autres structures. Après la segmentation d'une image, chaque pixel se voit attribuer une région, et chaque région correspond à une partie sémantique de l'image. Les discontinuités entre les régions correspondent aux contours des objets [KKM].

Les approches de segmentation peuvent se diviser en deux grandes classes :

- Approches de détection de discontinuité (contours ou encore frontières).
- Approches de détection de similarité (régions).

Remarquons qu'il existe aussi d'autres approches moins utilisées telles que les contours actifs, les Template, les modèles pyramidaux et les modèles mixtes.

I.4.1 Méthodes de détection de discontinuité

Les approches contours cherchent à extraire les contours présents dans l'image, en se basant sur l'étude des changements abrupts de la fonction de luminance ou sur la discontinuité des propriétés des ensembles.

I.4.2 Méthodes de détection de similarité

Les approches régions cherchent quand à elles, à détecter les zones de l'image présentant des caractéristiques d'homogénéité et vérifiant un critère d'homogénéité et de similitude. Parmi les approches régions, on distingue les méthodes de classification, la croissance de régions et la division-fusion.

Les deux approches sont complémentaires et aucune n'a prouvé sa supériorité par rapport à l'autre, chacune ayant des avantages et ses domaines d'application.

Cependant au cours de notre travail, nous nous sommes plutôt intéressées aux méthodes basées sur la détection des similarités (approche régions). Nous commençons par présenter les méthodes de classification, en mettant l'accent sur la classification floue puis, nous nous attelons aux méthodes contextuelles, et plus précisément à la croissance de régions.

I.5 Segmentation par classification

Les méthodes de classification permettent de regrouper des objets en groupes ou classes d'objets plus homogènes. Les objets regroupés ont des caractéristiques communes, ils sont similaires mais se distinguent clairement des objets des autres classes.

Généralement les résultats des méthodes de classification sont plus ou moins différents [Ati05] .

I.5.1 Méthodologie de la classification

La mise en œuvre d'une procédure de classification, ayant pour objectif de classer automatiquement des objets, comporte généralement deux phases fondamentales :

- une phase d'apprentissage dont le but est de déterminer un espace de représentation des signaux et de rechercher les paramètres discriminants capables de caractériser chaque classe d'objets;
- une phase de reconnaissance au cours de laquelle on attribue à une classe chacun des objets inconnus dans l'espace de représentation déterminé durant l'apprentissage.

Remarque

La classification peut être supervisée ou non supervisée.

I.5.2 Quelques méthodes de classification

La majorité des algorithmes de classification peuvent être divisés en deux catégories

I.5.2.1 Classification hiérarchique

La classification hiérarchique fournit une hiérarchie de partitions. Ce type de classification regroupe les méthodes ascendantes, descendantes et combinées. Ces algorithmes essaient de créer une hiérarchie de classes. Les objets les plus similaires sont rassemblés dans des groupes aux plus bas niveaux, tandis que les objets moins similaires se retrouvent dans des groupes aux plus hauts niveaux.

I.5.2.2 Classification non hiérarchique (partitionnelle)

Il n'y a pas de hiérarchie. Dans la plupart de ces méthodes le choix a priori du nombre de classes est nécessaire. Les points d'initialisation peuvent être des points du nuage pris au hasard ou les centroides d'une partition préalable. Les classes obtenues n'ont pas plus d'importance les unes que les autres.

Nous distinguons deux types de classification partitionnelle :

classification dure

classification floue

Dans la suite de ce paragraphe, nous présentons la segmentation dure par les k-Means (k moyennes), nous passons ensuite, à la segmentation basée sur la théorie des sous ensembles flous.

I.5.3 La classification par les k-moyennes

Avant de présenter la classification par les k-moyennes, nous parlons des nuées dynamiques de Diday [Did71], qui sont une version généralisée des k-moyennes.

I.5.3.1 Les nuées dynamiques

La méthode des nuées dynamiques consiste à trouver la partition optimale d'un ensemble d'individus en c sous-ensembles (ou classes), chaque sous-ensemble C_i est représenté par un « noyau ». Cette méthode a été proposée par Diday.

On note:

Ω : L'ensemble des partitions π de X en c classes (C_1, \dots, C_c).

$v = (v_1, \dots, v_c)$ un ensemble de c noyaux, où v_i est le noyau de la classe C_i .

$f(x, v_i)$ un critère qui mesure la dissimilarité (ou dissemblance) entre l'individu x et la classe C_i .

La classification cherche à optimiser l'adéquation entre la partition π et la représentation v associée au sens du critère f , c'est-à-dire à avoir la somme des mesures entre les individus et leur classe la plus faible possible.

La fonction objective à minimiser s'écrit alors :

$$g_f(\Omega) = \sum_{v_i \in \Omega} \sum_{x_k \in C_i} f(v_i, x_k)$$

La solution optimale à un algorithme partitionné est de trouver, parmi toutes les partitions possibles contenant c classes, celle qui minimise la fonction de coût.

En pratique, cette option n'est pas envisageable, compte tenu de l'explosion combinatoire, même pour une population d'échantillons de taille raisonnable. Par conséquent, la majorité des approches partitionnais utilisent une technique d'optimisation itérative.

Afin de minimiser la fonction objective, l'algorithme le plus utilisé est le suivant [Gé98] :

1. $it = 0$
2. Choisir une classification initiale de $X : \{ v_1(it), \dots, v_c(it) \}$
3. $it = it + 1$
4. Actualiser chaque classe suivant :

$$C_i(it) = \{x_k \in X, f(x_k, v_i^{(it)}) \leq f(x_k, v_j^{(it)}) \forall j \neq i\}$$

5. si $\exists i, v_i^{(it)} \neq v_i^{(it-1)}$ alors retourner à l'étape

Lors d'une itération, on affecte à chaque individu la classe dont la caractérisation est la plus proche des caractéristiques de cet individu. Ainsi les classes voient leur forme évoluer progressivement.

Discussion

R. Haroun dans [Har05] est arrivée aux remarques suivantes:

1. Le choix des noyaux initiaux V^0 est très important puisque la partition obtenue en dépend.
2. Cet algorithme est un algorithme de descente; donc il converge vers un minimum local. Si l'on souhaite éviter la rencontre d'un minimum local trop éloigné de la solution optimale, il faut alors mettre en œuvre une stratégie de recherche globale. Dans [Gé98], T. Géraud présente quelques méthodes d'optimisation qui permettent de résoudre ce problème.

I.5.3.2 K-means ou classification binaire

Soit $X = \{ x_k / k = 1, \dots, n \}$ un sous-ensemble d'un espace n-dimensionnel. Une partition de X en c classes peut être représentée par des sous-groupes mutuellement disjoints $C_1 \dots C_c$, tel que $C_1 \cup \dots \cup C_c = X$.

Chaque vecteur x_k est assigné à une et seulement une classe qui lui est la plus proche.

La fonction objectif correspondante s'exprime sous la forme [AYFM02] :

$$J = \sum_{i=1}^c \sum_{k=1}^n U_{ik} d^2(x_k, v_i)$$

Où :

- v_i désigne le prototype (centroïde) de cette classe.
- d mesure la similarité entre le prototype et le vecteur de données.
- U_{ik} est un indicateur binaire $\in \{0, 1\}$ désignant l'appartenance ou non du vecteur x_k à la classe C_i ($U_{ik} = 1$ si $x_k \in C_i$ et $U_{ik} = 0$ sinon, $\forall i=1 \dots c$

Et $x_k \in X$)

L'indicateur d'appartenance U_{ik} s'exprime comme suit :

$$u(x_k, v_i) = \{ 1 \text{ si } d(x_k, v_i) = \min_c d(x_k, v_c) \text{ ou } 0 \text{ sinon} \}$$

L'optimisation se fait de manière itérative. Les étapes de cette classification peuvent être résumées comme suit :

1. Initialiser l'ensemble des centroïdes avec des éléments, soit pris au hasard de l'ensemble de données, soit d'après des connaissances a priori.
2. Assigner chaque point au centroïde le plus proche.
3. Calculer les nouveaux centroïdes des classes construites. Les centroïdes étant les centres des classes, ils sont calculés par la moyenne des éléments de la classe. Générer une nouvelle partition de l'ensemble de données en assignant chaque point au centroïde le plus proche.
4. S'il y a changement dans la partition, réitérer depuis l'étape 3.

Discussion

La méthode des k-moyennes est très attrayante pour obtenir rapidement une première classification. De plus, c'est une méthode simple et rapide. Cependant, pour le cas particulier des images médicales, la supposition a priori par ces méthodes que les limites entre les classes sont très bien définies peut poser problème. Pour pallier ces limites, une méthode reposant sur les sous-ensembles flous est née.

I.5.4 La classification floue

Le principe de la classification floue est d'affecter un élément à classer non pas à une classe comme dans les approches classiques (dont fait partie la méthode des k-means), mais à toutes les classes avec un certain degré.

Ce principe découle de ce qu'on appelle la logique floue, qui est venue compléter la logique classique.

I.5.4.1 La logique floue

Dans la vie de tous les jours, nous nous trouvons dans des situations où les informations dont nous disposons ne sont pas toujours précises. Dans le domaine de la science aussi, des imprécisions peuvent être vues; par exemple lorsqu'on veut établir une valeur qui caractérise le degré de similarité de deux formes.

L'être humain est habitué à utiliser des informations entachées d'incertitudes et d'imprécision dans la vie de tous les jours, il utilise ces informations incomplètes, raisonne avec elles et prend des décisions. Dans le domaine scientifique, il a été nécessaire de créer une logique qui admette des valeurs de vérité en dehors de l'ensemble {vrai, faux } pour pouvoir tenir compte et manipuler ce genre d'information incomplète.

Lukasiewicz propose en 1920 une logique ayant les trois valeurs de vérité suivantes : "vrai", "faux" et "doute". Ces valeurs ont été représentées par l'ensemble { 0,1,0.5 }, et ont été ensuite étendues à l'intervalle [0,1].

Néanmoins c'est Zadeh [Zad65], qui à partir de l'idée d'appartenance partielle d'un élément à plusieurs classes, a formellement introduit la logique floue.

I.5.4.2 La théorie des sous ensembles flous

La théorie des sous ensembles flous et les outils de raisonnement qui en

découlent, proposent un cadre formel qui permet de modéliser le langage naturel et de gérer l'imprécis et l'incertain.

Cette théorie est basée sur la notion de degré d'appartenance. Ces degrés sont des valeurs qui expriment l'appartenance incertaine d'un pixel à une région. Le degré d'appartenance se situe dans l'intervalle $[0,1]$ et les classes obtenues ne sont pas forcément disjointes.

Dans ce cas, les données x_j ne sont plus assignées à une classe unique, mais à plusieurs par l'intermédiaire de degrés d'appartenance u_{ij} du vecteur x_j à la classe i

I.5.4.3 L'algorithme Fuzzy C-means (FCM)

L'algorithme des c-moyennes floues ou fuzzy c-means (FCM) repose sur la recherche des valeurs prises par les centroïdes des classes et par les degrés d'appartenance des voxels à ces classes, qui en respectant les contraintes, minimisent la fonction objectif suivante [Har05] :

$$J = \sum_{i=1}^c \sum_{k=1}^n U_{ik}^m d^2(x_k, v_i)$$

avec:

- c : le nombre de classes, connu a priori.
- n : la taille du vecteur de données (nombre de pixels à classer).
- U_{ik} : le degré d'appartenance du pixel x_k à la classe i connue par son centroïde v_i .
- d : le degré de similitude, il peut être la distance Euclidienne.
- m : un réel >1 appelé degré de flou. Plus m tend vers 1, plus la classification devient dure et U_{ik} se rapproche de 0 ou de 1. Inversement quand m devient trop grand il y a moins de tolérance au bruit, et la distribution des degrés d'appartenance tend à se concentrer autour de $1/c$.

Notation

Soit $X = x_k, k = 1, \dots, n$ un sous-ensemble d'un espace E , de dimension n . Nous cherchons à segmenter X en c classes suivant une classification floue. Chaque

classe C est caractérisée par son prototype v_i , et $V = \{v_i, i=1, \dots, c\}$ l'ensemble des prototypes des classes.

L'algorithme de la FCM tente de minimiser la fonction objectif, par rapport à chaque degré d'appartenance flou U_{ik} et chaque prototype v_i .

Le rôle de l'indice de flou m est de contrôler la contribution du bruit présent dans les données. Les détails sur le choix de m se trouvent dans le chapitre 4, dans notre présentation des études sur l'influence de m sur les résultats de la segmentation des images médicales.

I.5.4 Travaux utilisant la segmentation par FCM

L'algorithme FCM a été largement utilisé pour la segmentation des images de cerveau, quels que soient la modalité et le type d'acquisition (mono ou multimodal). De nombreux travaux ont notamment été effectués en imagerie par résonance magnétique. Baumgartner et al. ont ainsi utilisé la FCM pour segmenter les régions activées (simulation et aires motrices) du cerveau. Les pixels étant représentés par leur niveau de gris, les auteurs ont montré non seulement que l'algorithme avait des performances comparables à l'analyse de corrélation standard (avec l'avantage de ne nécessiter aucune connaissance a priori sur le paradigme), mais aussi que la FCM détectait des zones effectivement activées qui restaient silencieuses avec la corrélation.

En IRM traditionnelle, Clark et al. ont utilisé les c -moyennes floues comme étape d'initialisation à un système expert pour segmenter des volumes de tumeur ou d'œdème sur des images IRM cérébrales [STM07].

I.5.5 Quelques mots sur la segmentation par classification

Les limites majeures de la classification viennent des données elles même. Parfois, ces dernières ne contiennent pas suffisamment d'informations sur la région pour réussir une segmentation parfaite. Pour pallier les limites de la classification, des méthodes dites contextuelles sont nées [Har05].

I.6 Segmentation contextuelle

Les méthodes que nous avons présentées ci-dessus sont des méthodes dites non-contextuelles, c'est-à-dire qu'elles ne tiennent pas compte des voisinages des pixels.

Elles considèrent que l'appartenance d'un pixel à une classe ne dépend que de sa radiométrie.

À l'issue de l'application de ces méthodes sur des IRM cérébrales, nous obtenons une information relative aux caractéristiques radiométriques des matières principales présentes dans ces images : le liquide céphalo-rachidien, la substance grise et la substance blanche. La segmentation résultante n'est pourtant pas toujours très satisfaisante

Du fait du bruit et de la présence de volumes partiels, la loi statistique caractérisant la radiométrie d'une classe recouvre presque systématiquement la loi d'une ou de plusieurs autres classes. L'application de seuils radiométriques pour décider de l'appartenance d'un pixel à une classe cause donc des erreurs, qui pourraient être évitées si des connaissances portées par le voisinage de ce pixel étaient prises en compte.

Les paragraphes qui suivent présentent la croissance de région, qui est une méthode très utilisée dans la segmentation d'image médicale. Ensuite l'approche par morphologie mathématique est exposée, méthode généralement utilisée afin d'extraire l'encéphale.

I.6.1 Segmentation par croissance de région

Les méthodes par croissance de région, sont basées sur un processus itératif d'agrégation de pixels. Ces méthodes de par leur relative simplicité de mise en œuvre ont été l'une des premières familles d'outils utilisés pour la segmentation d'image. Dans [Zuc76], Zucker introduit les premiers travaux relatifs à la croissance de région .

I.6.1.1 Présentation de la croissance de région

À partir d'un ensemble de points initiaux, cette méthode agrège les pixels suivant un double critère : l'homogénéité et l'adjacence. Cette agrégation de pixels est contrôlée par un prédicat (fonction booléenne). Le processus s'applique sur une ou plusieurs régions.

Le prédicat permet d'identifier une contrainte que doivent satisfaire les régions, il peut être de nature géométrique (région rectangulaire, région convexe ...), de nature radiométrique (région homogène,..) ou une combinaison des deux.

Exemples de prédicats d'homogénéité pour une région :

- La variance $\sigma^2(R)$ des niveaux de gris de l'image associés aux points de la région R est inférieure à un seuil préfixé.

La proposition $\alpha(R)$ de points dont le niveau de gris se situe hors de l'intervalle $[m(R) - \alpha(R), m(R) + \alpha(R)]$ (où $m(R)$ est la valeur moyenne des niveaux de gris), ne dépasse pas un seuil préfixé.

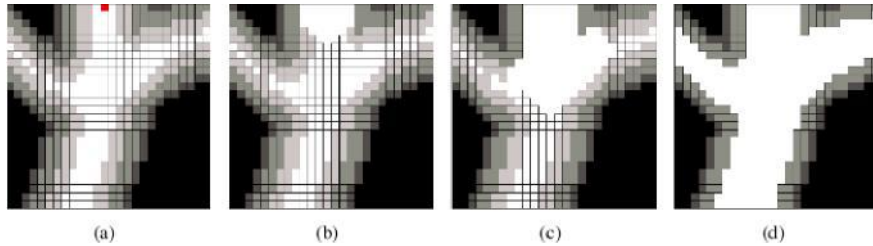


FIG. I.1: Principe de la croissance de région sur une image 2D :(a) image initiale à cinq niveaux de gris (point germe en rouge, en haut de l'image), (b) et (c) sont les étapes successives de la croissance de région jusqu'à obtention du résultat (d).

I.6.1.2 Algorithme général

Pour chaque région, le processus de croissance comprend une phase d'initialisation et une phase itérative.

On suppose que l'ensemble des sites S de l'image a été partiellement partitionné en R_j régions, $j = 1, \dots, i-1$. La phase itérative modifie la région R_i représentant l'objet segmenté. La région $S^{(0)} = S - \cup R_j$ correspond à l'ensemble des éléments non segmentés. Le processus de croissance de région étant itératif, le contenu d'une région à l'itération n est notée $R^{(n)}$.

- **Germes initiaux**

L'initialisation de la croissance de région nécessite le positionnement de germes désignant la ou les premières régions. Bien entendu, le point où la zone germe doit faire partie de la région recherchée, au risque d'obtenir une segmentation partiellement, voire totalement erronée. Ces germes constituent donc la région initiale $R_i^{(0)}$

Le choix d'un germe $R_i^{(0)}$ peut se faire par seuillage sur un attribut (par

exemple le niveau de gris), ou de toute autre manière (extraction de formes géométriques, etc.). Un germe peut éventuellement être réduit à un seul point.

- **Un processus itératif**

Suite à la phase d'initialisation, un processus itératif de déformation ajoute progressivement des points situés à la périphérie de la région en train de croître, s'ils respectent le prédicat. Ces nouveaux points sont regroupés dans l'ensemble que l'on note $S^{(k)}$. Considérons une croissance de région ne faisant croître qu'une seule région. Étant donné une région initiale $R^{(0)}$, la région suivante $R^{(k+1)}$ est obtenue à partir de la région courante $R^{(k)}$ et de la couronne $S^{(k)}$ par la réunion suivante:

$$R^{(k+1)} = R^{(k)} \cup S^{(k)}$$

Dans la croissance de région, la condition d'agglomération implique la définition d'un terme de similarité entre un point candidat et la région segmentée. Ce terme de ressemblance, appelé critère, est utilisé par le prédicat pour décider de l'ajout ou non d'un pixel. En pratique, le critère implique presque toujours une mesure d'homogénéité sur les intensités des points de la région segmentée. La Figure 2.2 présente un résultat de segmentation par croissance de région sur une image synthétique. Le processus a été initialisé à partir de deux germes. Le résultat est donné à différentes itérations de l'évolution.

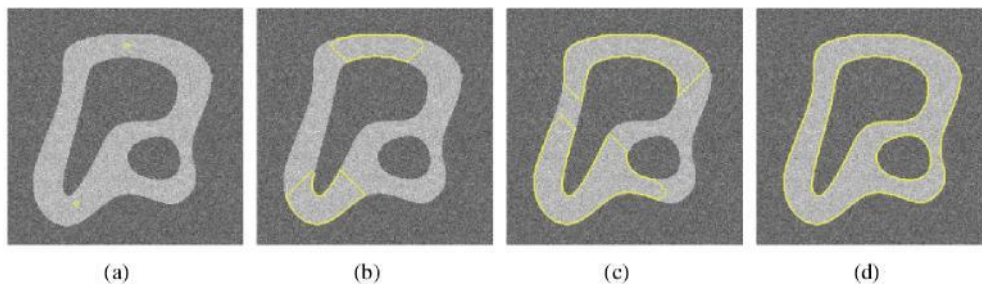


FIG. 1.2: Segmentation d'une image synthétique de 256 niveaux de gris additionnée d'un bruit gaussien (écart type = 10). (b)-(d) évolution de la segmentation à partir de deux germes (a).

Convergence, critère d'arrêt

Comme l'itération porte sur une région croissante et bornée, la convergence est assurée au bout d'un nombre fini d'itérations. Le processus de croissance peut s'arrêter selon deux scénarios :

- toutes les régions satisfont le prédicat et la segmentation comprend n régions.
- il existe $n - 1$ régions qui vérifient le prédicat, la $n^{\text{ième}}$ région comprend les points ne vérifiant pas le prédicat.

I.6.1.3 Propriétés de la croissance de région

La croissance de région a été employée à de multiples reprises pour la segmentation des images médicales. Cela s'explique par le fait que c'est une méthode simple à initialiser et rapide. Parmi les propriétés caractérisant cette méthode, ce qui suit [Ros08] :

1. La croissance de région permet de segmenter un objet d'intérêt par la simple désignation de cet objet. Cette particularité la rend attrayante pour les applications médicales.
2. La croissance de région n'intègre pas par défaut de terme de régularisation. La méthode ne nécessite pas de lisser le contour de la région segmentée contrairement à d'autres méthodes comme les contours actifs. Dans le cadre de la quantification pour l'imagerie médicale, ce terme de régularisation peut provoquer une erreur de mesure.
3. Une propriété consécutive au point précédent et à l'origine de la principale faiblesse de la méthode : sa sensibilité aux points de fuite. Un point de fuite se définit par la connexion de deux régions distinctes aux propriétés statistiques semblables.

I.6.2 Approche contextuelle par morphologie mathématique

La morphologie mathématique est une branche des mathématiques fondée sur l'étude des ensembles permettant d'analyser les structures géométriques. Cette méthode qui a vu le jour à l'Ecole des Mines de Paris, entre 1964 et 1968, sous l'impulsion de G. Matheron puis de J. Serra, a donné lieu depuis, à de nombreux développements.

Le développement de la morphologie mathématique a été inspiré par des problèmes de traitement d'images, domaine qui constitue son principal champ d'application. Elle fournit en particulier des outils de filtrage, segmentation, quantification et modélisation d'images.

Dans ce qui suit, nous présentons les éléments de base de la morphologie mathématique. Nous allons étudier la morphologie mathématique appliquée à des images binaires (deux niveaux de gris) car ceci correspond à notre utilisation de cette méthode. Toutefois, l'approche peut aussi être appliquée à des images en plusieurs niveaux de gris.

I.6.2.1 Outils de la morphologie mathématique

La morphologie mathématique est connue par ses éléments ensemblistes $S \subset X$; $S \cap X = \emptyset$ où X est l'ensemble à analyser et S un élément structurant que l'on choisit en fonction des besoins de l'analyse. Ces relations sont en effet, à la base d'opérateurs morphologiques élémentaires qui sont l'érosion et la dilatation (voir Figure 2.3).

Sur une image binaire, X pourra être l'ensemble des pixels blancs ou tout sous-ensemble de ce dernier, et S un sous-ensemble quelconque de l'ensemble des sites.

On notera:

$X_t = \{ x + t, x \in X \}$, le translaté de X par t .

$\tilde{S} = \{-s, s \in S\}$, le symétrique de S par rapport à l'origine.

- **Soustraction de Minkowski**

La soustraction de Minkowski de l'ensemble X par l'ensemble Y est

définie par: $X \ominus Y = \bigcap X_Y$

- **Addition de Minkowski**

L'addition de Minkowski de l'ensemble X par l'ensemble Y est

définie par: $X \oplus Y = \{ x + y, x \in X, y \in Y \}$

- **Érosion**

L'érosion mathématique par un élément structurant S est une soustraction de Minkowski par \tilde{S} de telle manière que:

$X \ominus \tilde{S} = \{x, S_x \subseteq X\}$

Une opération d'érosion sur une image I binaire contenant des étiquettes 0 et 1 (1 et 0 représentent respectivement le blanc et le noir), par un élément structurant S , consiste à une convulsion de S avec I , centré en i . Si cette convulsion est inférieure à une valeur prédéterminée (généralement la surface de S , le nombre de pixels qui sont à 1 dans l'élément structurant lui-même), on change la valeur du pixel i dans I de 1 à 0.

L'élément structurant (erosion kernel) détermine de quelle façon l'érosion amincit les frontières dans une image.

- **Dilatation**

La dilatation mathématique par un élément structurant S est une addition de Minkowski par \check{S} tel que:

$$X + \check{S} = \{ x, S_x \cap X \neq \emptyset \}$$

Comme pour l'érosion, une dilatation sur une image I contenant les étiquettes 0 et 1, par un élément structurant S change la valeur du pixel i dans I de 1 à 0, si le résultat de la convolution de S avec I , centré en i , est supérieur à une valeur prédéterminée. Cette valeur est généralement nulle. L'élément structurant (dilatation kernel) détermine de quelle façon la dilatation élargit les frontières dans une image.

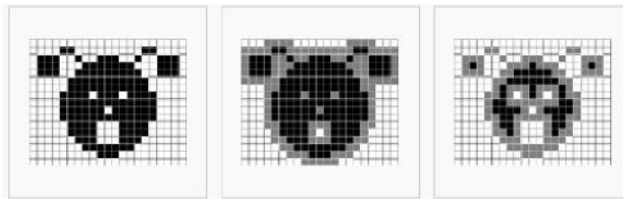


FIG. I.3: De la gauche vers la droite : Image originale (en noir : l'objet; en blanc : le fond), Dilatation par un carré 3x3, Érosion par un carré 3x3.

- **Dilatation conditionnelle**

Une dilatation conditionnelle est une opération de dilatation avec la condition ajoutée que seuls les pixels allumés (qui sont à 1) dans l'image sur laquelle la dilatation est conditionnée, vont être mis à 1 par l'opération de dilatation. Il est équivalent de masquer le résultat de la dilatation par une autre image.

- **Ouverture et fermeture**

Une opération d'ouverture mathématique consiste en une érosion suivie par une

dilatation avec le même élément structurant. La fermeture quant à elle, est une dilatation suivie par une érosion avec toujours le même élément structurant. (Voir Figure I.4)

I.6.2.2 Travaux utilisant la morphologie mathématique

Comme évoqué précédemment, la majorité des travaux utilisant la morphologie mathématique, ont pour but de réaliser l'édition de l'encéphale ou de segmenter les structures sous-corticales comme cela a été fait dans [Gé98] et [Har05].

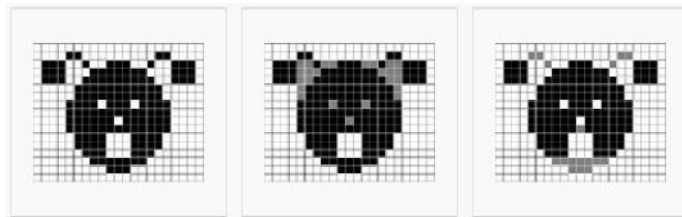


FIG. I.4: De la gauche vers la droite : Image originale, Fermeture par un carré 3x3, Ouverture par un carré 3x3.

- **Édition de l'encéphale**

En plus de la matière blanche, matière grise et du LCR, les images IRM cérébrales initiales contiennent d'autres substances indésirables. Il faut donc extraire l'encéphale du reste de l'image de la tête, cette segmentation est appelée édition de l'encéphale (Figure I.5).

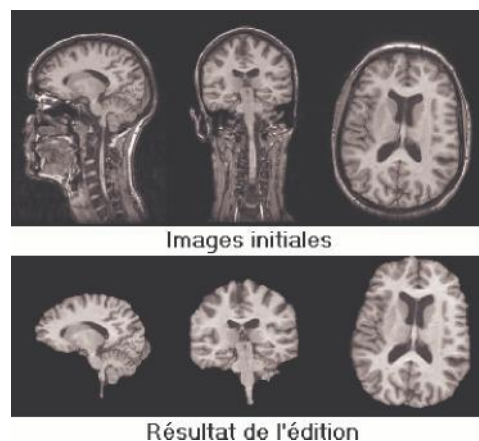


FIG. I.5: Édition de l'encéphale.

Remarque

D'autres méthodes de segmentation contextuelles existent, parmi elles la segmentation par champ de Markov. Cette méthode a été utilisée pendant longtemps et dans beaucoup de travaux se rapportant à la segmentation des images médicale.

I.7 Coopération en segmentation d'image

comme nous l'avons déjà précisé, il existe plusieurs approches de segmentation d'images, chacune agit de manière différente et utilise des attributs différents. De plus, chaque méthode ayant ses avantages et ses limites d'utilisation selon le problème à résoudre. Dans notre travail, nous visons à utiliser les avantages de plusieurs méthodes en même temps.

est réalisé grâce à la coopération. En effet, les méthodes coopératives apportent une meilleure segmentation des images puisque les algorithmes émergents de la coopération prennent en compte les différentes caractéristiques de ces images

Il existe en grande partie deux types de coopération; une coopération dite « séquentielle » dans laquelle il existe une exploitation successive du résultat d'une approche pour guider une autre. Un autre type de coopération est dit « itératif » fondé sur une définition mutuelle de contraintes et qui permet une émergence de la solution [Har05].

I.7.1 Coopération séquentielle

Les approches par coopération séquentielle sont principalement fondées sur des principes de focalisations successives et de corrections de résultats intermédiaires. L'objectif est de réduire progressivement la difficulté du problème en le décomposant en sous-parties de plus en plus fines.

I.7.2 Coopération itérative

Dans les approches par coopération itérative, plusieurs modules de segmentation s'exécutent simultanément. L'objectif est de construire la segmentation progressivement, en optimisant à chaque étape les résultats obtenus par chaque méthode mise en jeu. Il existe relativement peu de travaux proposant un tel type de coopération pour la segmentation.

I.8 Généralités sur les champs de Markov cachés

Les champs de Markov sont des outils probabilistes assez largement utilisés en

traitement d'images. Leur intérêt majeur est de pouvoir modéliser les dépendances spatiales des variables aléatoires, dont les réalisations modélisent les quantités observées ou recherchées, de façon relativement simple et se prêtant bien aux divers traitements. Divers modèles fondés sur les champs markoviens ont été proposés, parmi lesquels les champs de Markov cachés (CMC), très utilisés en segmentation d'images.

Il est important d'abord de préciser que la modélisation par les champs de Markov cachés n'est pas en elle-même une méthode de segmentation, mais un modèle statistique qui peut être employé dans des méthodes de segmentation. On peut noter aussi que les variables cachées n'ont pas, a priori d'existence physique dans le phénomène observé; mais, elles sont avant tout utilisées pour créer des modèles flexibles. Cependant, après analyse des données en regard du modèle, elles trouvent souvent une interprétation concrète a posteriori [Dur, 2003].

Les modèles de champs de Markov cachés permettent, grâce à leur structure, de prendre en compte les dépendances spatiales entre les différents pixels de l'image et d'incorporer une information a priori sur la segmentation [Pey, 2001].

Une difficulté associée aux modèles des champs de Markov cachés est le choix approprié des paramètres qui contrôlent la force d'interactions spatiales. De plus, ces modèles exigent d'habitude des algorithmes de temps de calcul intensifs.

Les champs de Markov cachés sont souvent incorporés dans des algorithmes de segmentation dans un cadre bayésien. La segmentation est alors obtenu en maximisant la probabilité a posteriori sachant les données de l'image en employant des méthodes itératives comme l'ICM (pour «Mode Conditionnel Itéré»), le MAP (pour «Maximum A Posteriori») et le MPM (pour «Mode des Marginales a Posteriori»).

Dans l'imagerie médicale, ils sont typiquement employés pour tenir compte du fait que la plupart des pixels appartiennent à la même classe comme leurs pixels voisins. Ils sont largement utilisés non seulement pour modéliser les classes de segmentation, mais aussi pour modéliser les inhomogénéités d'intensité qui peuvent apparaître par exemple dans des images mammographies

Ainsi, les modèles de Markov cachés peuvent être utilisés pour modéliser l'existence des zones homogènes. Il s'agit donc de modèles appropriés pour étudier la segmentation d'image avec un apprentissage non supervisé.

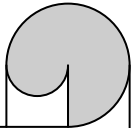
I.9 Conclusion

La présentation du contexte de la vision par ordinateur et de la problématique concernant l'interprétation de l'image nous a paru essentielle à l'introduction de ce chapitre. Cette présentation a permis un élargissement du domaine d'analyse indispensable à la conception d'un module de segmentation.

Après une synthèse des méthodes de segmentation d'images où nous avons exposé les méthodes par classification (les k-moyennes et les fuzzy C-means) et les méthodes contextuelles (croissance de régions et morphologie mathématique), qui sont utilisées dans le cadre de notre travail, le dernier paragraphe nous a permis d'introduire la notion de coopération en segmentation d'image.

Plutôt que d'utiliser des méthodes de segmentation classiques, il peut être intéressant de faire coopérer plusieurs approches; ce courant de pensée est utilisé dans de récents travaux. L'approche que nous proposons dans ce mémoire s'inscrit dans cet axe.

Afin d'utiliser la croissance de régions de manière plus efficace, elle est fondée sur un environnement de « systèmes multi-agents ». Dans le chapitre suivant, nous présenterons les systèmes multi-agents.



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. 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. Enfin, nous présentons ensuite la modélisation par les champs de Markov cachés, ainsi que les différentes techniques d'estimation des paramètres du modèle. Diverses simulations ont été effectuées dont certain feront l'objet d'une comparaison.

Introduction

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.

II. 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).

II.1.1 système de voisinage et clique

Système de voisinage

En analyse d'images, l'image est représentée par une grille $S = \{ S = (i, j) ; 0 \leq i \leq n - 1 \quad 0 \leq j \leq m - 1 \}$ 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.

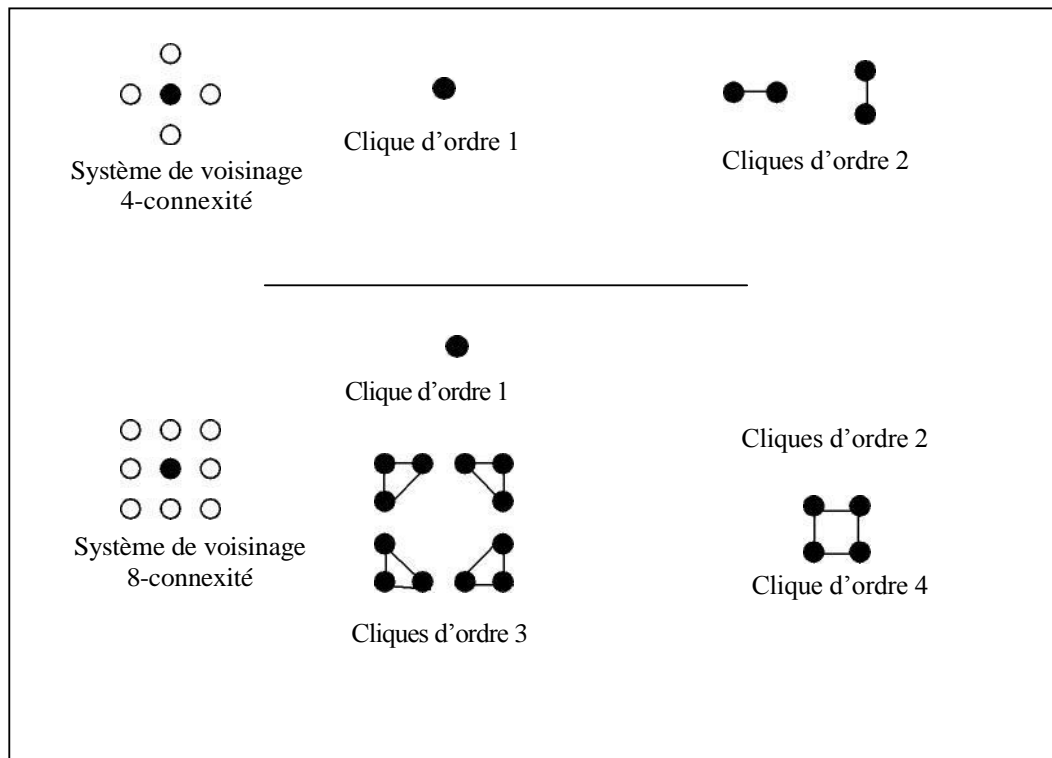


FIG. II.1 – Cliques associées à un système de voisinage en 4-connexité et en 8-connexité

II.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:

$$\forall x \in \Omega^{\text{card}(S)}, P(x) > 0 \text{ (contrainte de positivité)} \quad (\text{II.1})$$

- Markovianité:

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

La propriété (II.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.

II.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:

$$p(X = x) = \frac{1}{Z} \exp^{-U(x)} \quad (\text{II.3})$$

Avec :

$$U(x) = \sum_{c \in C} \varphi_c(x_c) \quad (\text{II.4})$$

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

$$z = \sum_{X \in \Omega^{card}(s)} \exp(-U(x)) \quad (II.5)$$

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 256x256 pixels, le calcul de Z nécessite une somme sur $2^{256 \times 256}$ configurations possibles, ce qui est rédhibitoire.

II.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 :

$$P(\mathbf{x}_S / \mathbf{x}_{V_S}) = \frac{\exp - \sum_{c \in C} \varphi_c(\mathbf{x}_S)}{\sum_{\mathbf{x}_S \in \Omega} \exp - \sum_{c \in C} \varphi_c(\mathbf{x}_S)} \quad (\text{II.6})$$

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

II.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:

$$U(x) = -\beta \sum_{s \in S} x_s - \alpha \sum_{(s,t) \in S} x_s x_t \quad (\text{II.7})$$

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 "*antiferromagnétique*".

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 (II.7) ;
- dans le cas contraire, le modèle est appelé modèle anisotropique. La fonction énergie d'Ising (I.7) devient:

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

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:

$$U(x) = \alpha \sum_{(s,t) \in S} (1 - \delta(x_s, x_t)) \quad (II.9)$$

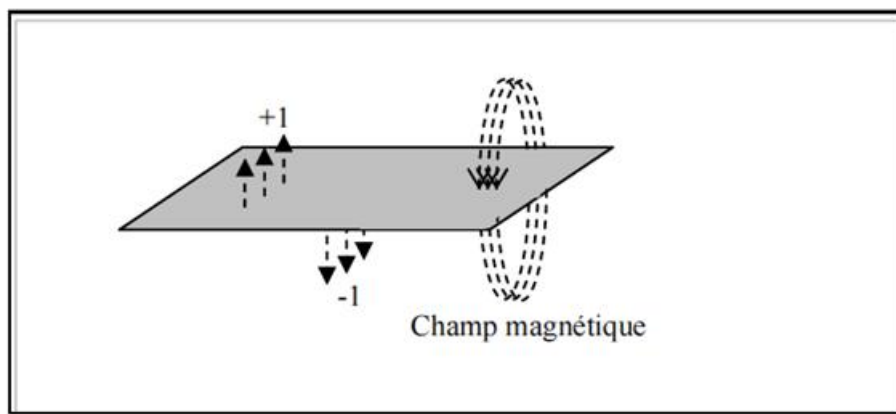


FIG II.2 orientation d'un spin à l'état d'équilibre

pour un champ magnétique nul, où $\delta(x_s, x_t)$ est la fonction de Kronecker, telle que:

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

II. 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* $p(x|y)$.

II.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 k selon la loi uniforme dans Ω ,
 - Calculer la variation d'énergie locale ΔU par:

$$\Delta U = U(x_s = k, x_{v_s}) - U(x_s^n, x_{v_s}) \quad (\text{II.10})$$

- Si $\Delta U \leq 0$ le changement est accepté $x_s^{n+1} = k$
- Sinon, le changement est accepté ($x_s^{n+1} = k$) 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.

II.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 $P(x_s^n | x_{v_s})$
- Effectuer un tirage aléatoire d'une variable k dans Ω , selon ces probabilités conditionnelles et poser $x_s^{n+1} = k$

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

Durant notre travail nous allons utiliser l'échantillonneur de Gibbs ou Il existe deux types de ce dernier : (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 :

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

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. Elles ont été réalisées par l'échantillonneur de Gibbs utilisant 20 balayages déterministes.

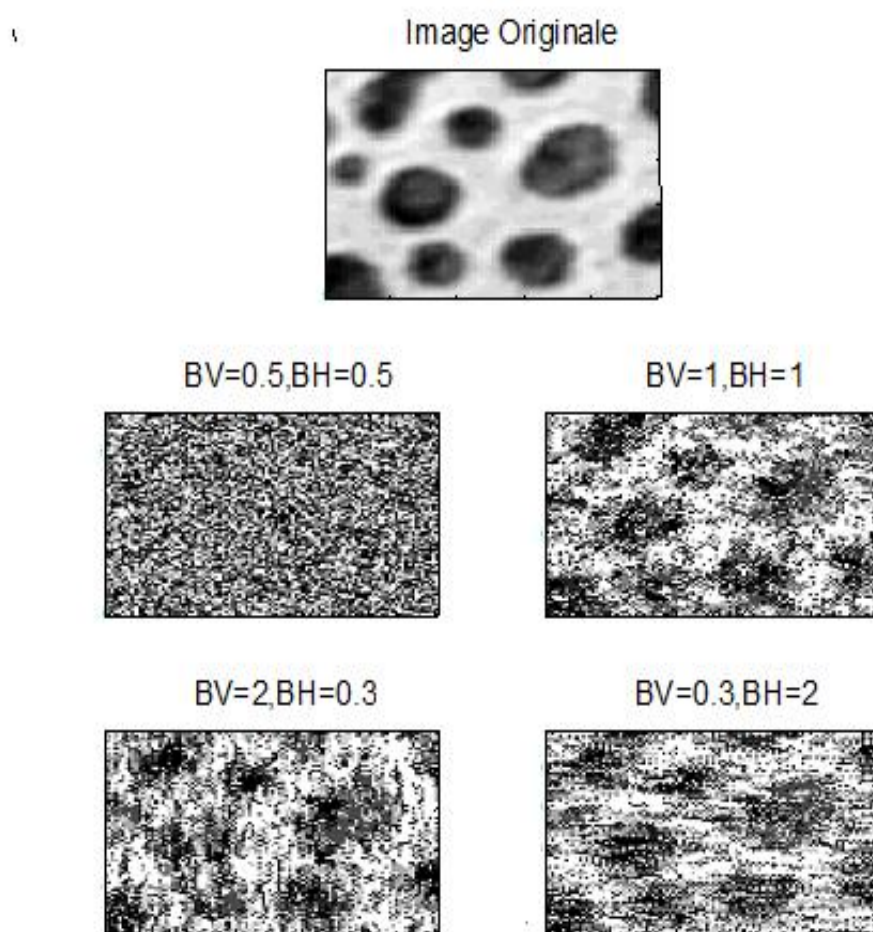


FIG.II. 3 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs pour différentes valeurs de β_V & β_H

La figure précédente montre une réalisation d'échantillonneur de Gibbs avec 20 itérations d'une image à 4 classes pour différentes valeurs de β_V et β_H , ou on a simulé avec la 4

connexité comme type de voisinage , les deux premières images on a choisit des valeurs égaux pour BH et BV , 0.5 pour la première et 1 pour la deuxième

Pour les images qui suivent on a prêt des valeurs grandes pour BV et minimiser BH en suite en inversent l'opération.

On a obtenue des images verticale et horizontal.

Même chose pour la réalisation suivante mais cette fois en utilisant un voisinage de 8 connexité, le résultat de la simulation est montré dans la figure II.4.

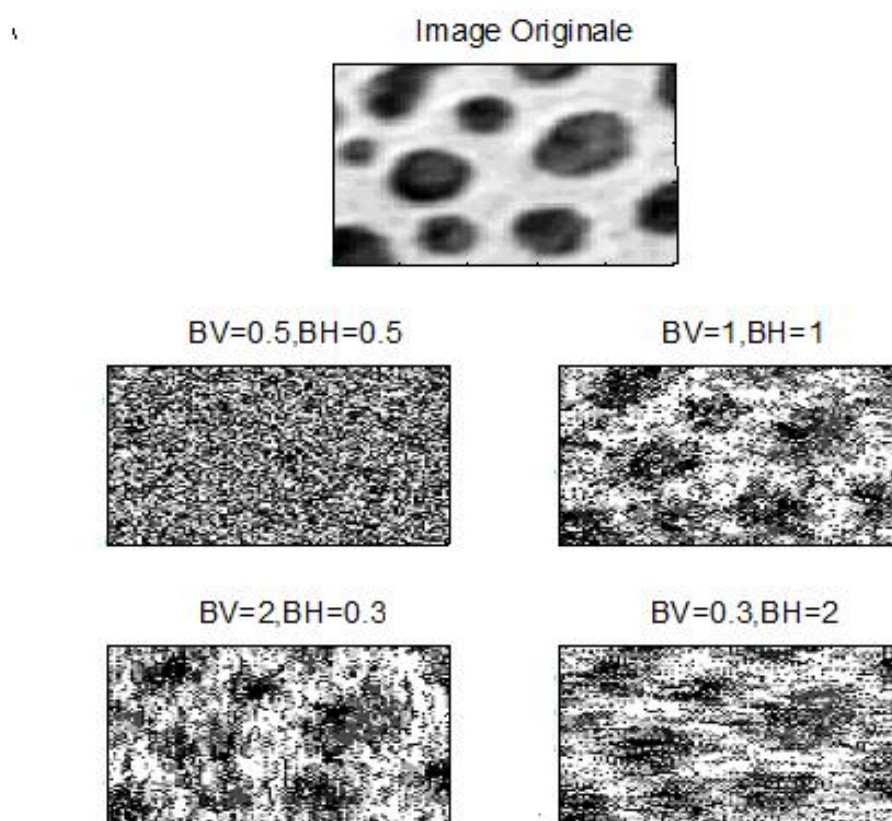


FIG. II.4– Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs pour différentes valeurs de β

Pour la suite on utilise la loi de probabilité a posteriori pour l'échantillonneur de Gibbs .

On a fait plusieurs réalisations (20 itérations) de Gibbs pour déférente valeur de BV et BH en 4

connexité et des valeurs de B pour 8 connexité ou 0 n'a trouvé des images à voisinage de 8 mieux que de 4 connexité ce qu'est montré dans les figures II.5 et II.6 respectivement .

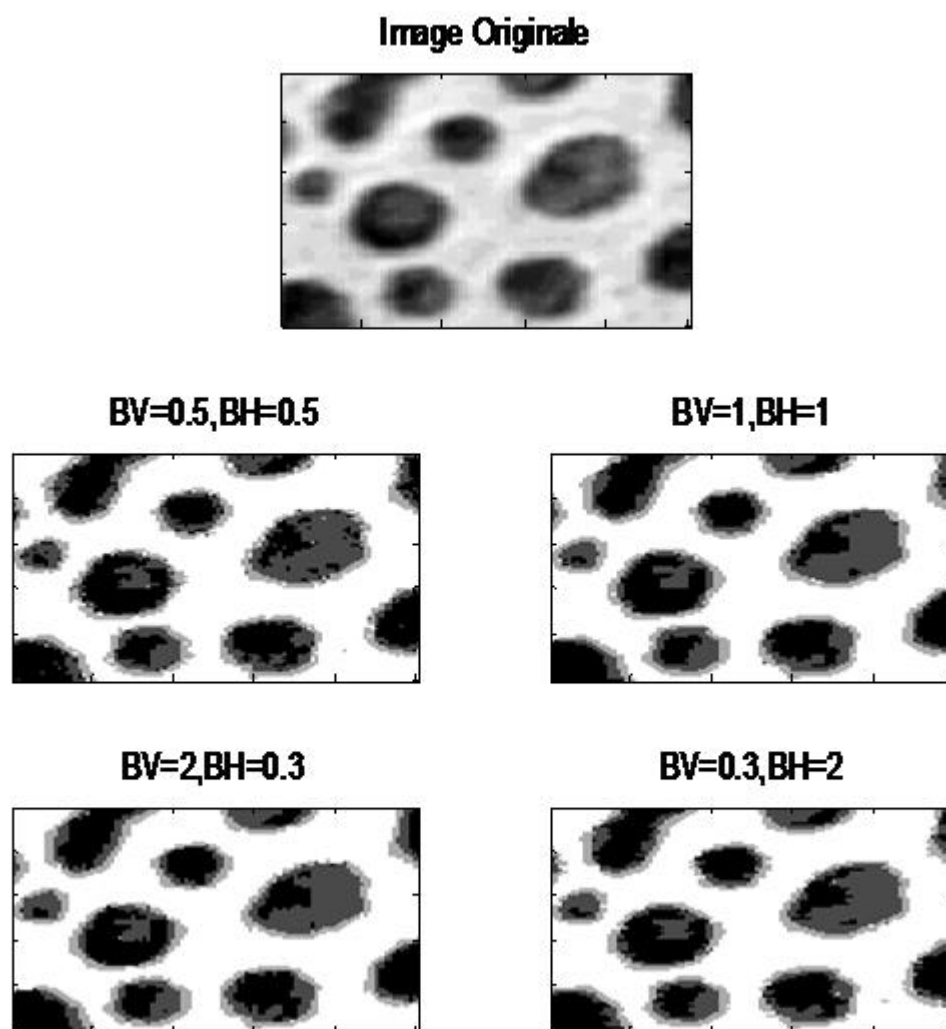


FIG.II. 5 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs a posteriori pour différentes valeurs de β_V & β_H

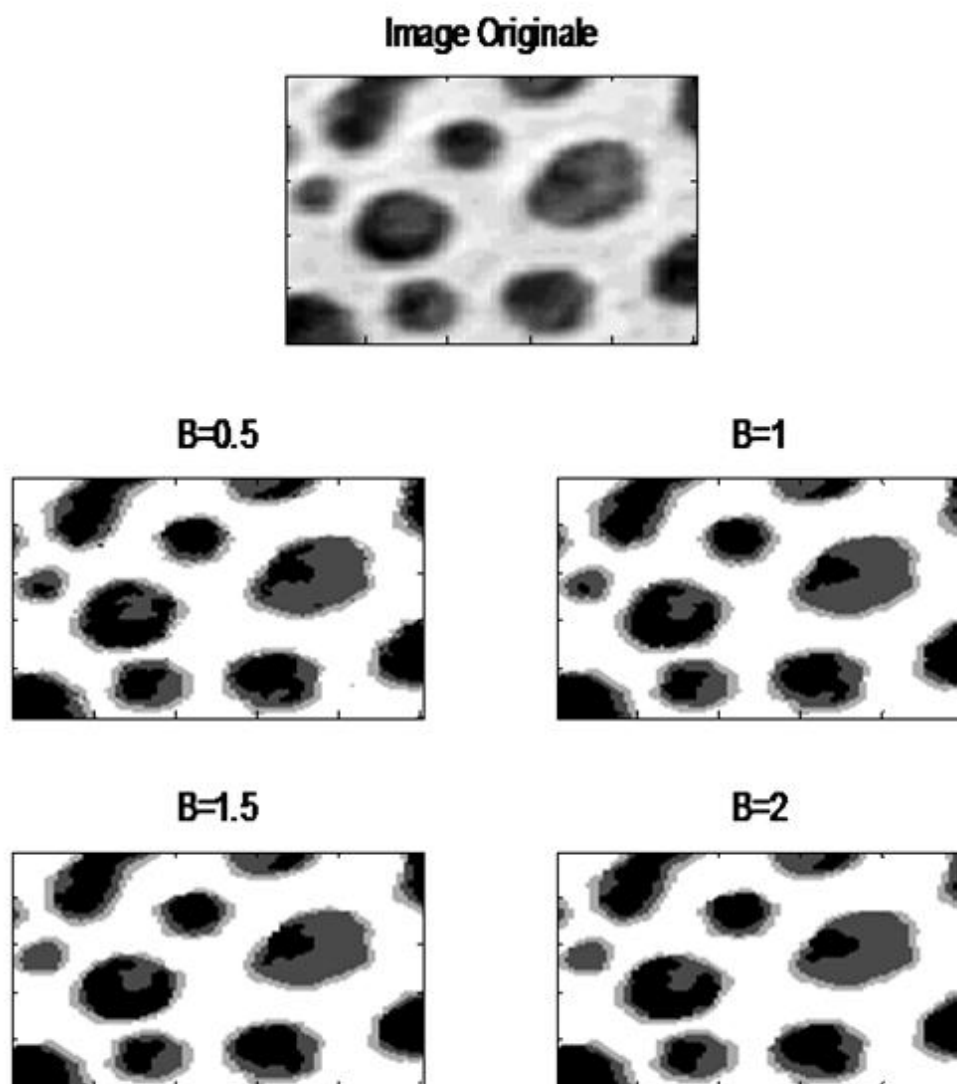


FIG.II. 6 – Exemple de simulation d'un champ de Markov à 4 classes par échantillonneur de Gibbs a posteriori pour différentes valeurs de β

Remarque

La différence avec l'échantillonneur de Gibbs réside dans le tirage au sort du nouveau niveau de gris (ou descripteur), au lieu de considérer la loi définie par tous les descripteurs. Comme on ne considère que la variation énergétique entre les 2 configurations, l'algorithme de Metropolis soit plus rapide à chaque étape que l'échantillonneur de Gibbs, qui lui nécessite le calcul de la fonction de partition locale. Mais la convergence peut être plus lente car le taux d'acceptation est strictement inférieur à 1 (les transitions ne sont pas toujours acceptées, contrairement au cas de l'échantillonneur de Gibbs).

II.2.2.1 Rôle du choix des cliques

D'après les deux cas ci-dessus, on peut conclure qu'il faut prendre des cliques sur toutes les directions pour approcher une certaine isotropie. Nous considérons dans un premier temps à 4 connexité, on trouve un résultat un peu meilleur qu'en 8-connexité mais les contours sont marqués, par contre avec les 8 voisins, on trouve des images mieux régularisées.

Nous pouvons conclure qu'un choix de voisinage plus grand donne une image plus régulière.

II.2.2.2 Rôle de la constante de couplage β

Lors de l'utilisation du modèle de champs de Markov caché nous devons fixer un certain nombre des coefficients du modèle.

Un des coefficients à fixer est le coefficient β qui pondère la partie de l'énergie.

Correspondant au modèle de Potts. Nous illustrons sur les deux cas ci-dessus le rôle de ce coefficient en choisissant différentes valeurs : $\beta = 0.5$, $\beta = 1$, $\beta = 1.5$ et $\beta = 2$.

- Le cas $\beta = 0.5$ donne un résultat un peu meilleur dans le 4- connexités, il présente des nouvelles zones blanches. Mais avec 8 pixels voisins les images sont mieux régularisés.
- Le cas $\beta = 1$ donne un bien meilleur résultat lorsque l'on utilise le 8-connexité. Les petites zones ont disparue. On obtient des zones uniformes. On remarque aussi que le résultat obtenu avec 4- connexité est moins bon.
- Les deux cas $\beta = 1.5$ et $\beta = 2$ donnent des bons résultats dans les deux connexités.

La comparaison des quatre cas présentés permet d'illustrer la raison pour la quelle le coefficient β et parfois appelé facteur d'échelle. Lorsque le coefficient grandit les zones plus petites qu'une certaine échelle n'apparaissent plus dans la modélisation.

II.2.2.3. Conclusion

Du au comparaison des résultats obtenues on trouve que le voisinage de 8 connexité donne des résultats mieux que le 4 connexité.

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

II.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

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

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 :

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

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 \hat{a} , cet algorithme devient plus rapide que l'échantillonneur de Gibbs. A l'inverse, lorsque les valeurs de \hat{a} 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.

II.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 mémoire, pour l'échantillonneur de Gibbs.

II. 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:

$$P(x/y) = \gamma \exp[-\sum_{c \in C} \varphi_{c \in C}(x_c)] \prod_{s \in S} p(y_s/x_s) \quad (\text{II.14})$$

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.

II. 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 carrées 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 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 α_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.

II. 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 $\Omega^{card(s)} \otimes \Omega^{card(s)} \rightarrow R^+$. Celle-ci possède les propriétés suivantes:

$$\forall x, \hat{x} \in \Omega^{card(s)} \otimes \Omega^{card(s)} :$$

$$L(x, \hat{x}) \geq 0 ;$$

$$L(x, \hat{x}) = 0 \Leftrightarrow x = \hat{x} \quad (\text{II.15})$$

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:

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

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

$$\hat{s}_B(y) = \arg \min_{\hat{x} \in \Omega} E \left[L(\hat{x}(Y), X) | Y = y \right] \quad (\text{II.17})$$

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 conditionnal 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.

II.5.1 Estimateur Maximum A Posteriori (MAP)

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

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

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:

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

Par conséquent, l'estimation bayésienne est :

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

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*".

II.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 :

$$P_T(x) = \frac{1}{z(T)} \exp\left\{-\frac{U(x)}{T}\right\} \quad (\text{II.21})$$

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.

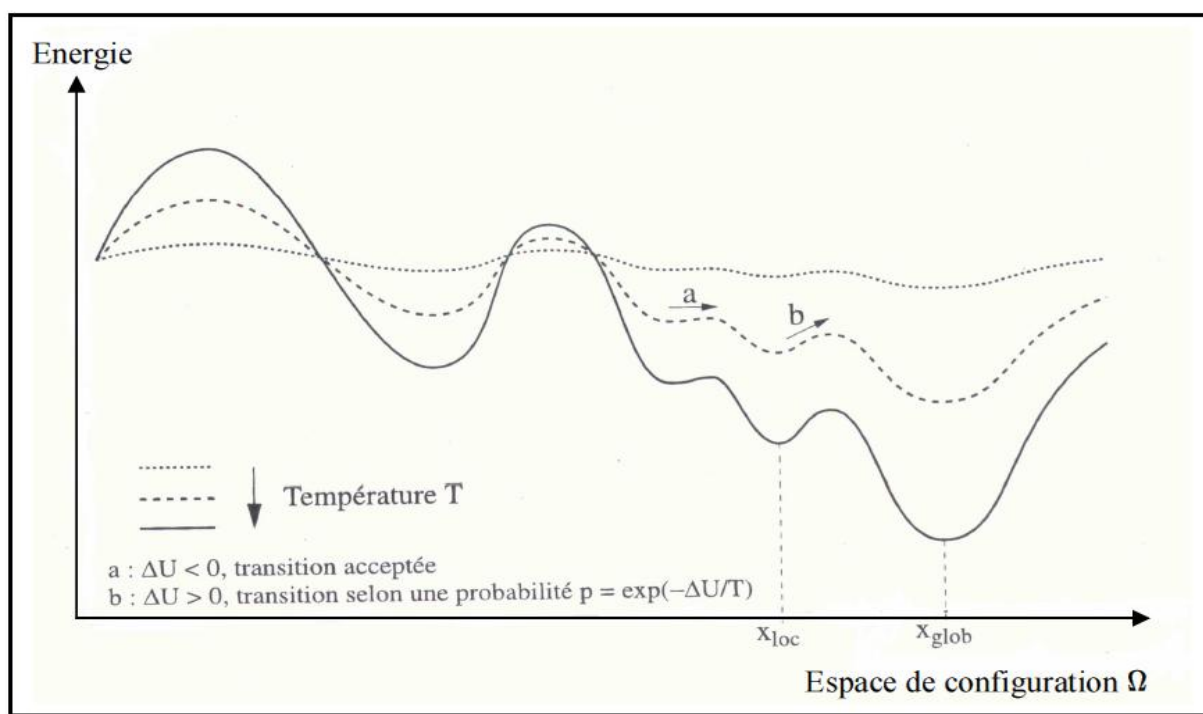


FIG. II.7 – Evolution de la fonction d'énergie en fonction de la température

II.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 :

$$x_s^n = \arg \max_{x_s \in \Omega} P(x_s / x_{V_s}^{n-1}) \quad (II.22)$$

- 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:

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

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 \widehat{x} est ici obtenue suivant:

$$\widehat{x}_s = \arg \max_{x_s} P(x_s | y) \quad (II.24)$$

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, mais le fait de pouvoir simuler des réalisations de X par des approximations de type Monte-Carlo permet leur estimation. En effet, on peut poser :

$$P(x_s = \omega | y) = \frac{1_{[x_s^1 = \omega]} + \dots + 1_{[x_s^N = \omega]}}{N} \quad (II.25)$$

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

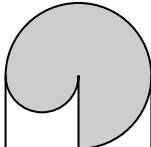
- Simuler N réalisations x^1, x^2, \dots, x^n de X_s selon la probabilité $Y = y$ en utilisant l'échantillonneur de Gibbs (décrit dans § I.2.2) ;
- Estimer à partir des réalisations x^1, x^2, \dots, x^n la loi de chaque X_s par les fréquences (équation (I.25)) ;
- Retenir la classe maximisant la loi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

II. 6 Conclusion

Nous avons exposé dans ce chapitre la problématique générale de l'analyse statistique d'images, abordée par les modèles de Markov. Nous avons présenté les modèles classiques, ainsi que les principaux algorithmes de segmentation bayésienne.

La modélisation par champ de Markov rend compte des propriétés spatiales a priori (et/ou spatio-temporelles dans le cas d'une séquence d'images) des images observées, la modélisation par champ de Markov présente deux intérêts majeurs : (1) tout d'abord, il s'agit d'un modèle facilement paramétrable ; (2) à cette souplesse s'ajoute l'avantage d'une formulation immédiate du terme de probabilité conjointe du champ d'étiquettes grâce au théorème de Hammersley- Clifford qui relie propriétés locales (marginales) et probabilités globales du champs.

Ces modélisations et traitements seront repris et généralisés à des modélisations et traitements originaux dans la suite. Le chapitre suivant présente une méthode de segmentation non supervisée markovienne.



Deux approches permettent de traiter le problème de l'estimation des paramètres : Une première approche dite supervisée suppose connue sur une partie de l'image, une réalisation de X . Une deuxième dite non supervisée. Dans ce cas, il faut déterminer uniquement à partir de l'observation Y les différents paramètres. Cette estimation plus difficile à mettre en œuvre, mais offrant à l'opérateurs une plus grande souplesse d'utilisation. L'apprentissage non supervisé d'un modèle correspond à l'estimation des paramètres le caractérisant. Ces paramètres se divisent en deux types : ceux associés au terme d'attache aux données (les probabilités conditionnelles) et ceux relevant du modèle à priori (les probabilités à priori). Une fois ces paramètres estimés, on peut procéder à l'étape de la segmentation.

Ainsi, l'apprentissage non supervisée des paramètres qui induit à une segmentation non supervisée est le thème central de cette étude.

III.1. Introduction

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 a qu'on appelle "ensemble de p 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.

III.2. Synoptique de la segmentation

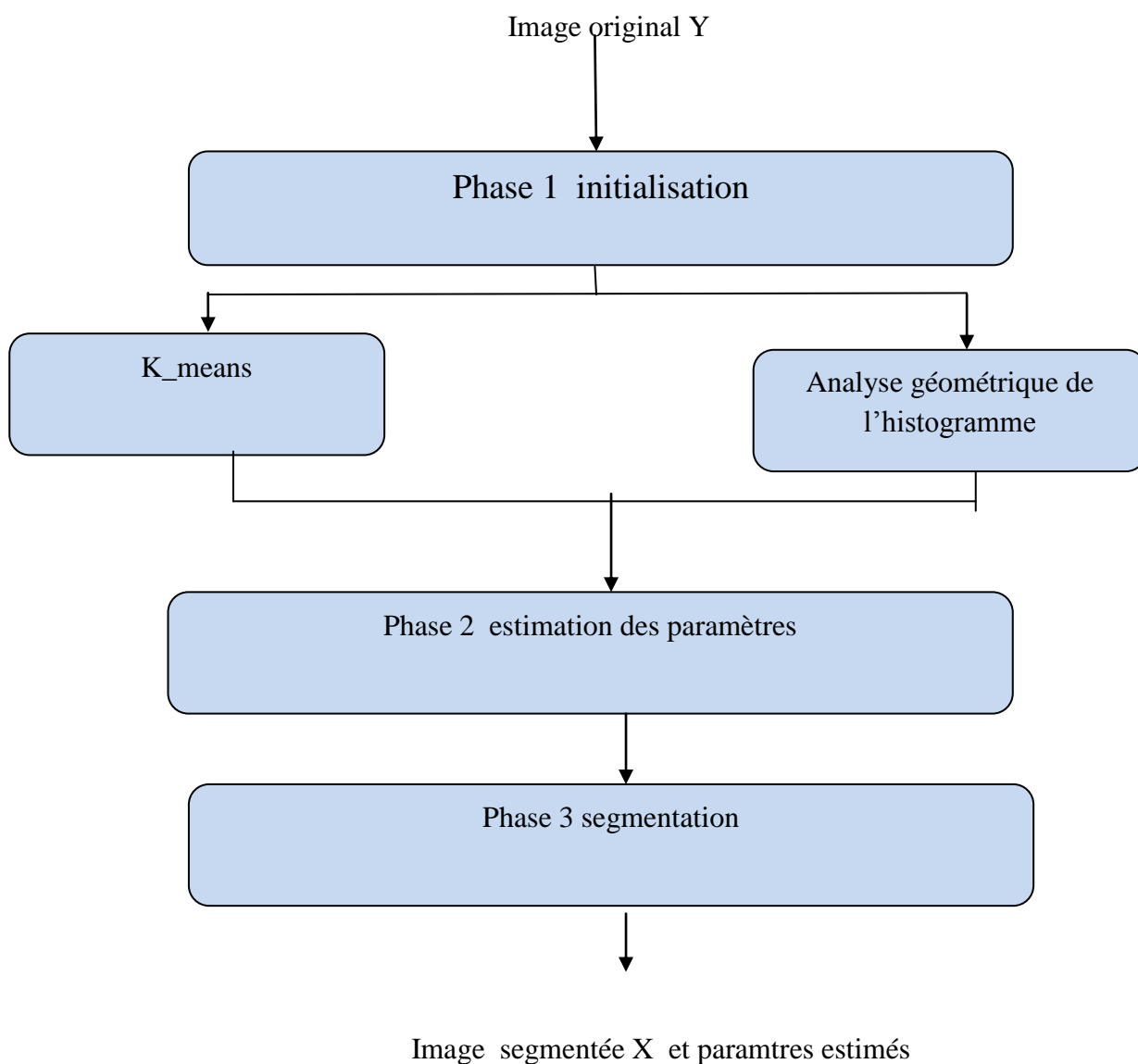


FIG III.1 organigramme de la segmentation d'image

Selon la figure ci-dessus, l'algorithme global de la segmentation non supervisée se présente de la manière suivante :

Phase 1 initialisation

Initialisation de l'image d'entrée peut se faire de deux manières différentes :

- a. K-means.
- b. Analyse géométrique de l'histogramme.

Phase 2 estimation de paramètres

La phase d'estimation des paramètres est réalisée par l'algorithme ICE. Chaque itération fournit une estimation des paramètres. Quand l'estimée finale des paramètres θ_n est obtenue, on procède à la segmentation ordinaire (segmentation supervisée) avec des paramètres connus.

Phase 3 : segmentation

Cette phase permet de segmenter l'image originale à partir des paramètres estimés θ_n (à la sortie de ICE). Dans un contexte bayésien, l'algorithme utilisé est le « Mode Marginales Posteriori » (MPM),

III.3 Initialisation

L'objectif de cette phase est de fournir une première estimation des paramètres des modèles, en vue de la seconde phase. Pour cela, nous avons implémenté un algorithme de classification : l'algorithme des K-means. L'image segmentée résultante de cette initialisation (X°) conduit à une première estimation des paramètres du modèle (θ°). La phase d'initialisation est essentiellement identique dans les deux cas et les étapes qui la composent sont les suivantes :

Lecture de l'image (Y).

Calcul de l'histogramme de l'image.

Segmentation initiale par l'algorithme classique des K-means

Estimation initiale des paramètres θ° à partir de l'image originale (Y) et de l'image segmentée initiale (X°). Cette étape dépend du modèle de Markov. Elle est également utilisée dans la seconde phase (Estimation des paramètres) qui a été présentée précédemment.

III.4. Estimation des paramètres

Dans la pratique, les paramètres de régularité et les paramètres des distributions des classes sont souvent inconnus et doivent être estimés à partir de l'observation $Y = y$. Le problème est alors double : nous ne connaissons pas les caractéristiques des classes et nous ne savons pas quels pixels sont représentatifs pour chaque classe.

Dans ce travail, nous considérons que l'attache aux données est une gaussienne dont ses paramètres sont : la moyenne et la variance des classes. Le modèle a priori est le modèle de Potts de paramètre de régularité λ . Nous supposons que les différentes classes sont définies à partir des niveaux de gris uniquement, c'est à dire que nous ne tenons pas compte de critères de texturation. Nous considérons également que nous avons une seule image de données fixe.

Il y a plusieurs méthodes itératives pour l'estimation de ces paramètres, incluant l'algorithme EM (pour « Expectation-Maximisation ») [Del, 1997], le SEM (pour « Stochastique EM ») [Ban, Gho et Pie, 2000] [Pey, 2001] et l'ICE (pour « Espérance Conditionnelle Itérative ») [Pal et Pal, 1993]. Ici, Nous considérons seulement le dernier.

III.4.1. Principe de l'algorithme ICE

L'algorithme ICE est une méthode très générale d'estimation de paramètres. Elle a été introduite par W.Pieczynski en traitement d'images en 1996 [Char, 2000] [Del, 1997] [Pal et Pal, 1993]. Elle apparaît plus générale car elle accepte toute forme de modélisation.

Son principe s'appuie sur l'hypothèse qu'il existe au moins un estimateur $\hat{\theta}(X, Y)$ de θ défini sur les données (X, Y) . On cherche alors la meilleure approximation de cet estimateur au sens de l'erreur quadratique moyenne qui n'est autre que l'espérance

conditionnelle par rapport à Y : $E[\hat{\theta}(X, Y)/y, \theta^{(q)}]$. Cette espérance dépend de θ et n'est pas calculable explicitement.

On utilise alors la procédure itérative suivante appelée ICE (pour « Espérance Conditionnelle Itérative ») [Mar et Pie, 1991] [Pie, 1992] :

- Initialisation de θ , soit $\theta^{(0)}$.

$$\bullet \quad \theta^{(q+1)} = E[\hat{\theta}(X, Y)/y, \theta^{(q)}] = E[\arg \max_{\theta} \log P(X, Y/\theta)/y, \theta^{(q)}] \quad \text{III. 1}$$

Au sens de l'erreur quadratique moyenne, l'espérance conditionnelle $E[(\hat{\theta}|Y = y^o)]$ est la meilleure approximation. Mais celle-ci dépend elle-même de θ .

L'algorithme adopte donc une démarche itérative pour approcher $E[(\hat{\theta}|Y = y^o)]$ et à chaque étape q , on recherche le paramètre $\theta^{(q+1)} = E[\hat{\theta}(X, Y)/y, \theta^{(q)}]$. Notons que si l'espérance n'est pas explicitement calculable, on peut l'estimer par tirages aléatoires selon la loi conditionnelle à Y .

Alors nous allons dans la suite présenter un ICE pour un champ de Markov caché.

III.4.2. L'algorithme ICE

Dans le cas d'un champ de Markov caché, l'algorithme ICE crée itérativement des réalisations a posteriori et recalcule les paramètres de classes associés à $f_i^0(\mu, \sigma)$ et de régularité spatiale (λ). Dans le cadre d'estimation d'un mélange gaussien, les étapes de calcul suivantes sont effectuées pour chaque itération q et avec $1 \leq i \leq k$:

- Un certain nombre de réalisations a posteriori (avec l'indice r) sont simulées avec l'échantillonneur de Gibbs, en utilisant la formule (II. 11) et les paramètres θ_i^{q-1} (défini par f_i^{q-1}) obtenus dans l'itération précédente.
- Les paramètres de classe $\theta_i^{q,r}$ (qui sont la valeur de la moyenne (μ_i^q) et la variance (σ_i^q) pour une classe de distribution gaussienne) sont estimés pour chaque réalisation, et puis moyennés pour avoir θ_i^q :

$$\mu_i^q = \frac{\sum_{n=1}^N y_n \delta(X_n - \omega_i)}{\sum_{n=1}^N \delta(X_n - \omega_i)} \quad \text{III. 2}$$

$$(\sigma_i^q)^2 = \frac{\sum_{n=1}^N (y_n - \mu_i^q)^2 \delta(X_n - \omega_i)}{\sum_{n=1}^N \delta(X_n - \omega_i)} \quad \text{III. 3}$$

- On a choisi d'estimer les paramètres de régularité λ^q dans chaque itération de ICE avec une approche de gradient stochastique [Del, 1997], réalisée à travers une série de réalisations a priori calculées avec l'échantillonneur de Gibbs en utilisant la formule (II. 11).

Dans le but de réduire le temps de calcul, généralement, on simule seulement une réalisation à posteriori pour chaque itération ICE et seulement une réalisation a priori pour chaque itération de gradient stochastique. Cette simplification ne cause aucune perte de performance significative.

On utilise une approximation de l'équation du gradient stochastique [Del, 1997] [Coc et Phi, 1995]. Soit $U_{aposteriori}^q$ l'énergie de la réalisation a posteriori courante et $U_{apriori}^q$ l'énergie de la réalisation a priori dans l'itération r du gradient stochastique. Posant $\lambda^{q,0}$ la valeur obtenue dans l'itération précédente λ^{q-1} de ICE et on répète le calcul jusqu'à convergence :

$$\lambda^{q,r} = \lambda^{q,r-1} + \frac{1}{r} * \frac{U_{apriori}^{q,r} - U_{aposteriori}^q}{U_{aposteriori}^q} \quad \text{III. 4}$$

III.3.3. L'algorithme MPM pour la segmentation d'image

L'estimateur du MPM, pour les champs de Markov, a été introduit par Marroquin [Char 2000]. L'algorithme MPM consiste à estimer les lois marginales. Pour ce faire, il est nécessaire de générer m échantillons $\{x(1)...x(m)\}$ de la loi a posteriori. La loi des grands nombres devrait alors nous donner accès aux marginales :

$$P(x_s = x_s/y) \approx \frac{1}{m-p} \sum_{q=p+1}^m \delta(x_s - x_s(q))$$

où δ est le symbole de Kroneker. On rejette les p premiers échantillons car ils sont considérés dépendants de la configuration initiale. L'estimateur du MPM sera alors approché par le champ des étiquettes apparaissant le plus fréquemment dans la suite des $(m-p)$ échantillons.

III.3.4. Simulation de l'algorithme MPM pour la segmentation d'image

Comme nous avons vu au cours des articles dans ce chapitre ,nous venons de monter quelque exemple de segmentation d'images simulé par l'algorithme MPM .

Les figures III.2 III.3 illustrent des images segmentées par l MPM a 4 classes utilise un voisinage de 4 connexité pour des valeurs de égaux de BH et BV 0.5 et1 .

Nous avons aussi donner 20 itérations de Gibbs et 10 pour MPM .

En suite nous commençons a utilisé le voisinage du 8 connexité toujours a 4 classes pour le k_means ,20itérations de Gibbs et 0.8 le cardinal de B pour la figure III.4 on a utilisé 10 itérations MPM et 40 pour la figure qui suivre .

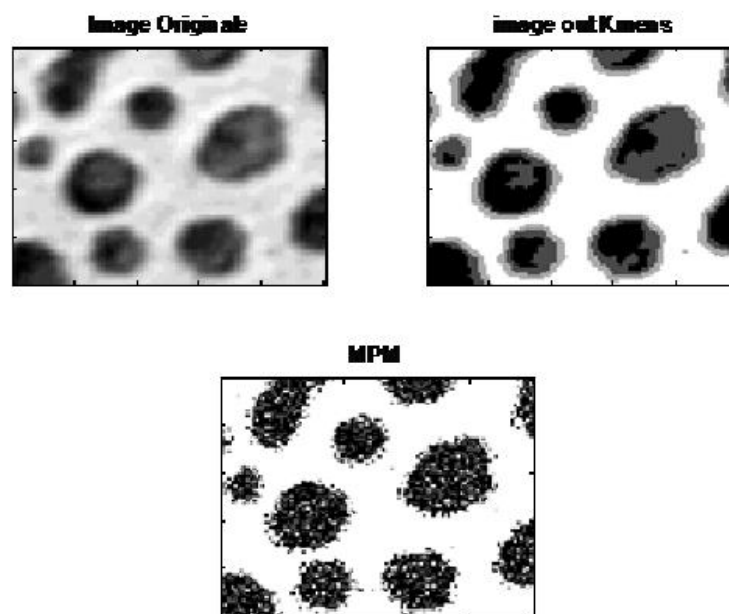


Fig III.2 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm $\beta_h = \beta_v = 0.5$

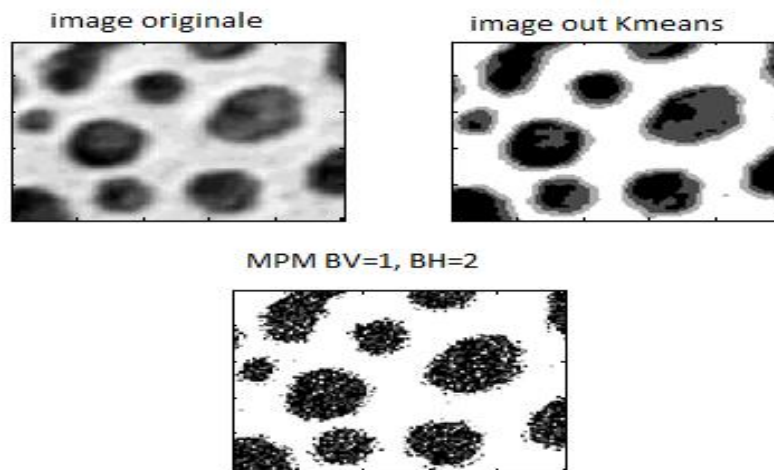


Fig III.3 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm

$$\beta_h = \beta_v = 0.5$$

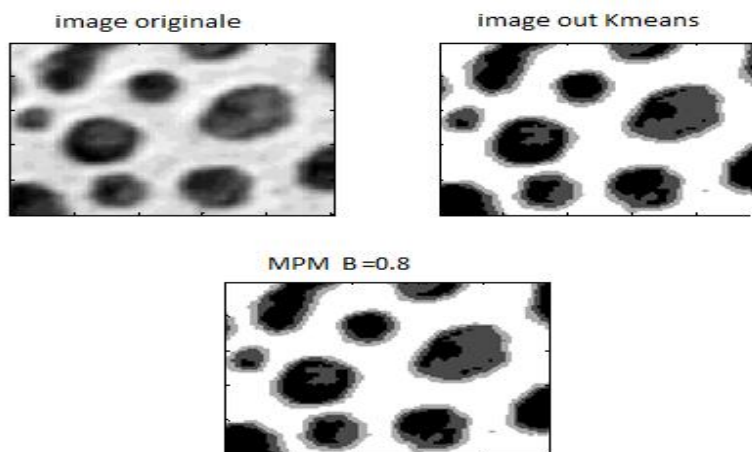


Fig III.4 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm

$$\beta = 0.8$$

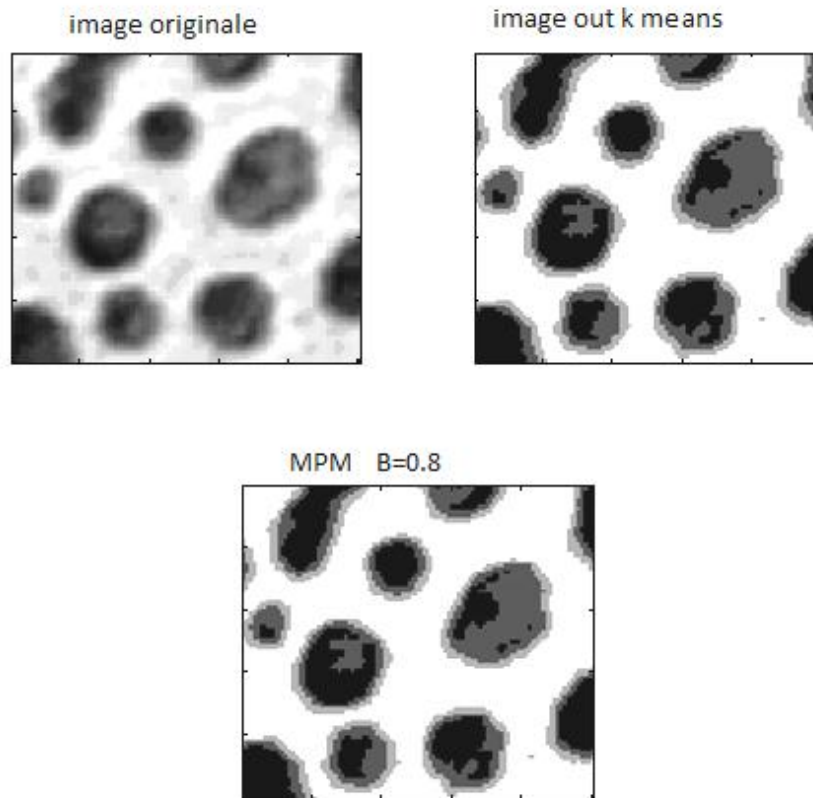


Fig III.5 Résultat de segmentation en niveau de gris de l'image selon algorithme mpm

$$\beta=0.8$$

Remarque

D'après les résultats obtenus, on peut conclure que la simulation de l'algorithme MPM avec le voisinage de 8 connexités est mieux que le 4.

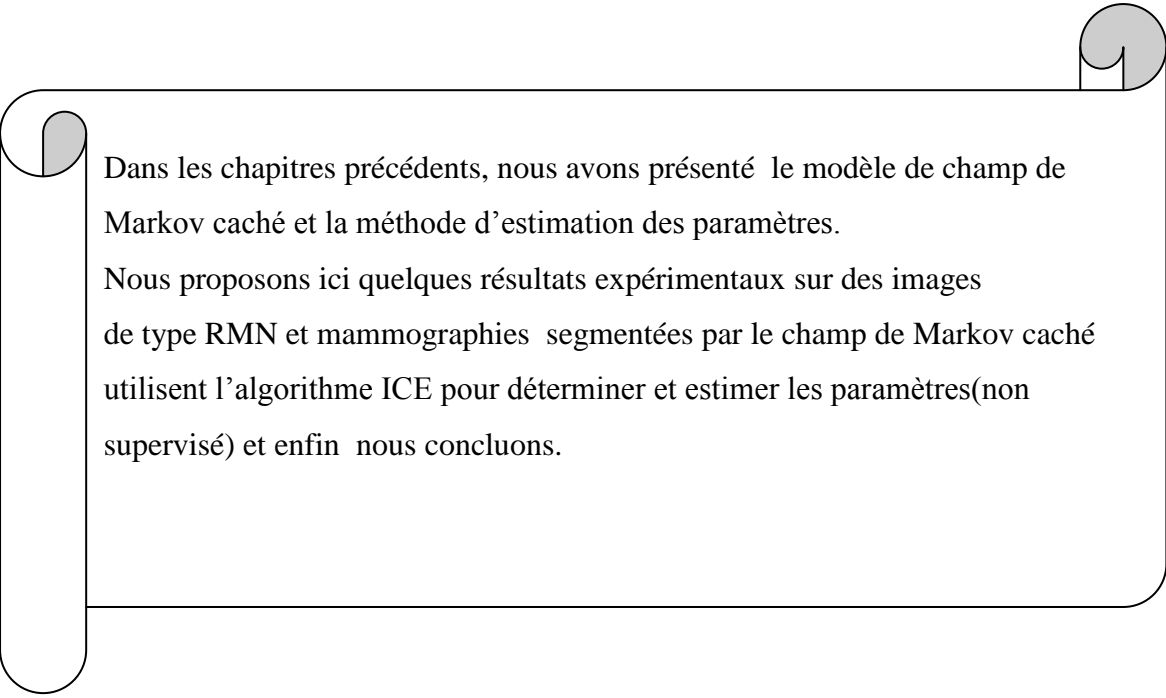
le nombre des itérations de l'algorithme MPM joue un rôle important dans la qualité des images obtenu donc l'augmentation des itérations donnent des meilleurs résultats le seul inconvénient c'est bien le temps de calcul qui serai plus long avec le nombre d'itérations

III.4 Conclusion

Nous avons abordé dans ce 3ème chapitre du mémoire, l'aspect de la segmentation.

En premier lieu nous avons donné une architecture modulaire des étapes de segmentation Supervisée d'images et expliqué le fonctionnement des différents modules des composants. Nous avons présenté une méthode générale récente d'estimation dans le cas des données cachées, dite Estimation du Mode des Marginales à posteriori (MPM).

Plusieurs faits ressortent de cette étude expérimentale. D'abord, l'intérêt de la prise en compte de l'information spatiale pour la classification des pixels d'une image apparaît de manière évidente dans les différents exemples traités. Comme il fallait s'y attendre, le choix d'un modèle gaussien conduit à des segmentations acceptables et l'on observe une nette amélioration au niveau de l'homogénéité de la segmentation avec un modèle spatial. Ce résultat est satisfaisant car cela confirme, que les approximations d'un champ de Markov caché utilisé préservent l'information markovienne.



Dans les chapitres précédents, nous avons présenté le modèle de champ de Markov caché et la méthode d'estimation des paramètres.

Nous proposons ici quelques résultats expérimentaux sur des images de type RMN et mammographies segmentées par le champ de Markov caché utilisant l'algorithme ICE pour déterminer et estimer les paramètres (non supervisé) et enfin nous concluons.

IV.1 Introduction

Dans les chapitres précédents, le modèle de champ de Markov caché et la méthode d'estimation des paramètres ont été présentés. Nous consacrons ce chapitre à l'application de cette modélisation à la segmentation statique non supervisée d'images RMN.

IV.2 Critères d'évaluation

Il est nécessaire d'utiliser un critère numérique permettant une évaluation de la qualité des résultats. Pour évaluer nos résultats de segmentation, nous proposons d'exploiter deux critères utilisés généralement pour l'évaluation d'un résultat de segmentation dans un contexte non supervisé : le premier c'est le rapport signal sur bruit crête (PSNR). C'est un critère global, défini sur l'image complet ou sur une sous-image. L'autre est la matrice de confusion C'est une matrice ou un tableau affichant les statistiques de la précision de classification d'une image, notamment le degré de classification erronée parmi les diverses classes.

Il existe plusieurs critères pour évaluer le résultat comme la matrice de confusion et le PSNR Pour l'évaluation de notre travail on a choisis ce dernier car il est plus simple , très utilisé et il ne prend pas un énorme temps de calcul .

IV.2 .1 Le rapport signal sur bruit crête (PSNR)

L'opérateur PSNR mesure le rapport entre l'information et le bruit dans une image. Il est calculé à partir d'une image initiale img_E qui comporte l'image et le bruit et d'une image img_S qui est la version segmentée de l'image initiale img_E . Les images img_E et img_S doivent avoir la même dimension et le même type. Le PSNR s'agit de quantifier la performance des algorithmes en mesurant la qualité de réinsertion de l'image segmentée par rapport à l'image originale.

Le PSNR est défini par :

$$PSNR = 10 \log_{10} \left(\frac{d^2}{EQM} \right) \quad IV. 1$$

- d est la dynamique du l'image. Dans le cas standard d'une image où les composantes d'un pixel sont codées sur 8 bits, $d = 255$

- EQM est l'erreur quadratique moyenne et est définie pour 2 images image E et image S de taille $M \times N$.

Maximiser le PSNR revient à minimiser l'erreur quadratique. Les valeurs typiques de PSNR pour des images de bonne qualité varient entre 30 et 40 dB.

D'autres distances peuvent être utilisées, faisant par exemple intervenir la norme L_p au lieu de la norme L_2 . On peut ainsi mettre en oeuvre toute une famille d'indicateurs, certains étant plus sensibles aux grands écarts d'intensité, d'autres aux petits écarts. Cependant, le classement des images qui en résulte est tributaire du choix du critère : il est alors difficile de décider, dans l'absolu, quelle est la fonction de distance la plus pertinente. Ainsi, il semble impossible de mettre en oeuvre un critère en-dehors de tout contexte applicatif.

IV.2 .2 Résultats sur d'images synthétiques

Nous avons travaillé sur une images synthétiques de taille 128×128 constituée de 4 classes et a été bruitée en lui additionnant un bruit blanc gaussien de moyenne nul et d'écart type est le même pour toute l'image.

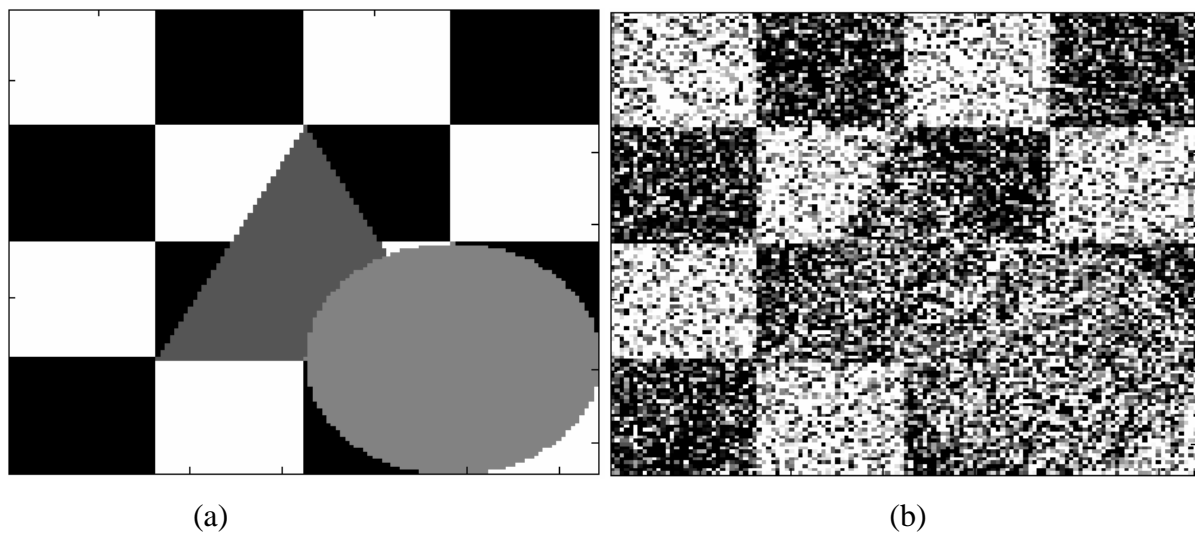


FIG.IV.1 - (a) Image synthétique originale; (b) Image bruitée

IV.2 .3 Analyse et discussion

En générale, les résultats obtenus sont satisfaisants et présentent des performances semblables en termes d'estimation des paramètres des modèles et en terme de segmentation.

Si l'on analyse les variations des segmentations au point de vue visuel, on remarque qu'elles sont cohérentes

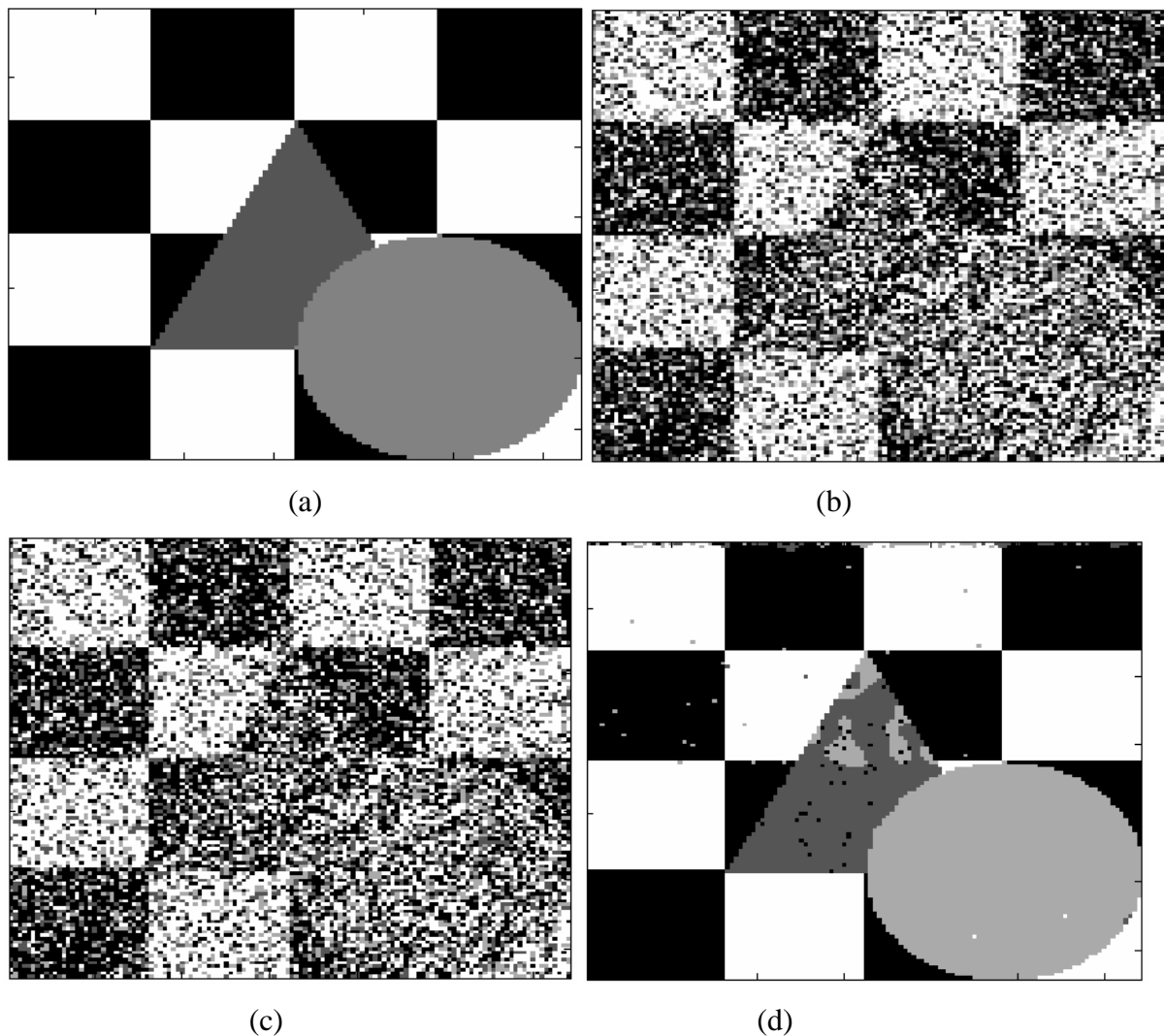


FIG.IV.2 Résultat de segmentation en niveau de gris de l'image synthétique (1) sans flou : (a) image originale, (b) image bruitée, (c) selon algorithme en Kmeans et (d) selon algorithme en champ de Markov caché.

Après le calcul du PSNR qui est égale à 31.4746 dB qui est acceptable ,appartiennent à l'intervalle 30-40 dB donc on peut dire qu'on est sur le bon chemin.

IV.2 .4 Conclusion

Nous avons proposé dans cette partie un algorithme d'estimation des paramètres a priori et conditionnels, ICE, dans un cadre de mélanges de lois gaussiennes, d'un champ de Markov. La méthode de ICE donne de meilleurs résultats d'estimations, surtout lorsqu'il s'agit d'un bruitage important. Aussi l'algorithme ICE est très rapide. Nous avons également présenté l'algorithme MPM qui a été appliqué au problème de segmentation par des champs de

Markov cachés. Nos différents algorithmes seront testés et évalués sur des problèmes de classification non supervisée d'images synthétiques. Ces algorithmes formeront un système de segmentation qui sera par la suite implémenté. Cette implémentation mettra en évidence l'intérêt et la performance de champ de Markov au niveau de la segmentation des images médicales.

IV.3 Présentation des images par résonance magnétique nucléaire (RMN)

Depuis plusieurs années, les connaissances médicales, le diagnostic et les thérapies se transforment de manière durable sous l'influence des progrès technologiques dans le domaine de l'imagerie. Les développements de l'imagerie permettent, entre autres, l'aide au diagnostic, le suivi de l'évolution des pathologies, la détection précoce, la planification et la simulation d'interventions. Quand l'imagerie par résonance magnétique (IRM) a été présentée, la communauté scientifique n'y a accordé que peu d'intérêt, sans pressentir les progrès qu'elle allait apporter dans l'aide au diagnostic ou au confort du malade.

Actuellement, elle a pris une importance considérable dans l'ensemble des modalités d'imagerie diagnostique. En effet, en raison de son caractère non-invasif l'imagerie par IRM est particulièrement bien adaptée au suivi pathologique et thérapeutique, et ceci d'autant plus s'il s'agit de maladies à évolution lente.

Les applications diagnostiques de la RMN fournissent essentiellement des informations morphologiques : des coupes tomographiques du corps humain avec un excellent contraste entre les tissus mous de nature différente. L'imagerie par résonance magnétique est fondée sur la mesure de signaux nucléaires générés par les noyaux d'hydrogène en réponse à des impulsions radiofréquences (RF). La majorité de recherche dans les méthodes de segmentation d'image médicale se concentre sur la RMN, particulièrement l'imagerie cérébrale. La plupart de ces méthodes sont basées sur l'information de l'intensité et leur souci principal est la présence d'inhomogénéités d'intensité des images RMN. Ainsi les images RMNs sont très sensibles aux hétérogénéités résiduelles résultant du mode d'acquisition. Pour corriger ces imperfections et réduire ces hétérogénéités, nous avons exploité une méthode de segmentation automatique basée sur les champs de Markov cachés. Cette méthode permet d'extraire automatiquement les régions de l'image et permet aussi de corriger les hétérogénéités des images grâce à une optimisation de l'entropie (énergie). Notons que le modèle de segmentation que nous présentons n'est pas contraint par la nature des données traitées. En effet, la seule hypothèse formulée sur les données concernent la nature du bruit :

nous supposons que le bruit sur les données est additif, gaussien et de moyenne nulle. Nous ne prenons en compte aucun paramètre physique de la machine à RMN ou des capteurs. Il est cependant évident que cette hypothèse devra être modifiée dans le cas de l'imagerie à résonance magnétique fonctionnelle (IRMf).

Donc ce travail s'articule autour de deux principales thématiques d'application, l'extraction des régions et la réduction des inhomogénéités des images RMN suivantes :

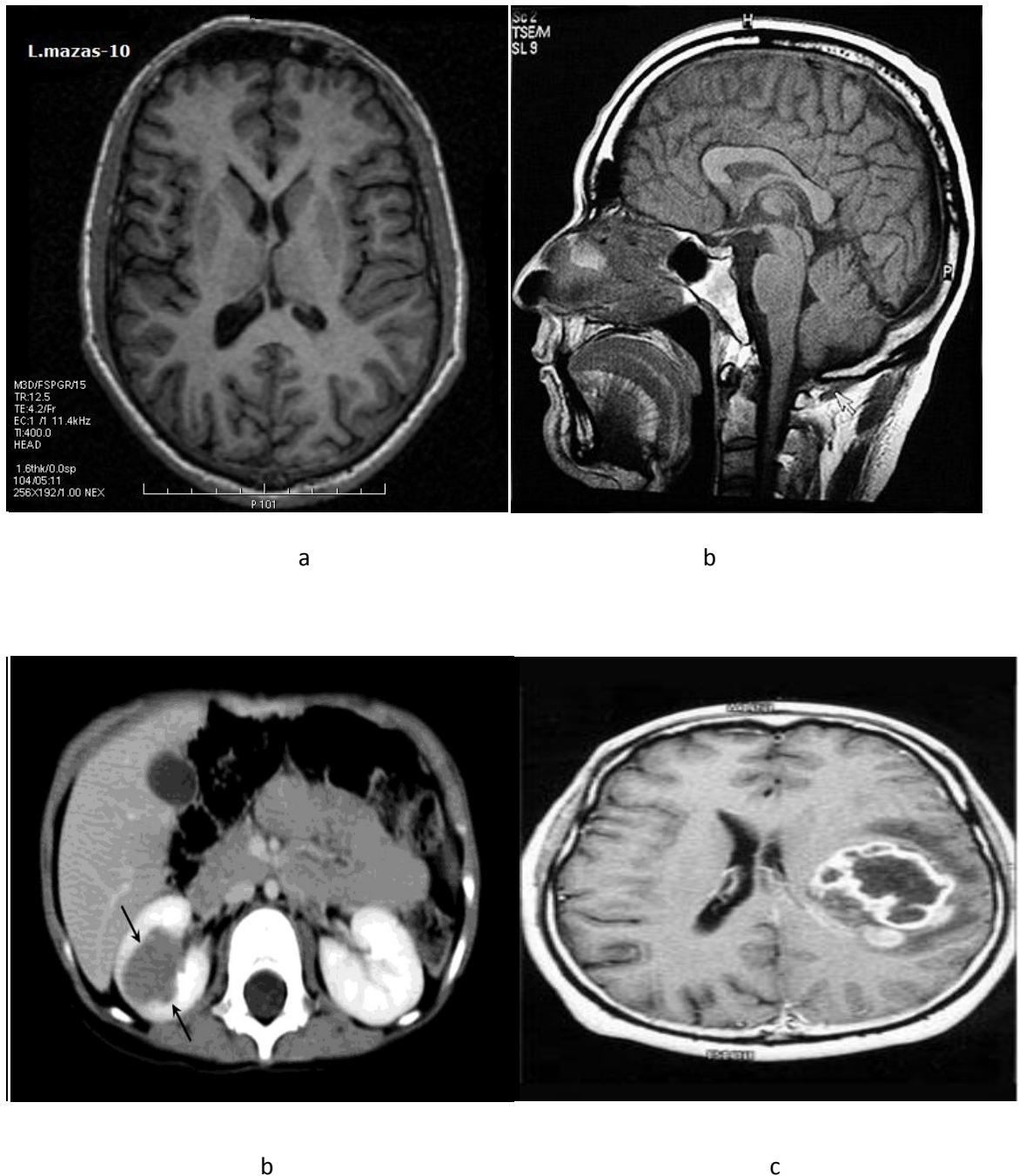
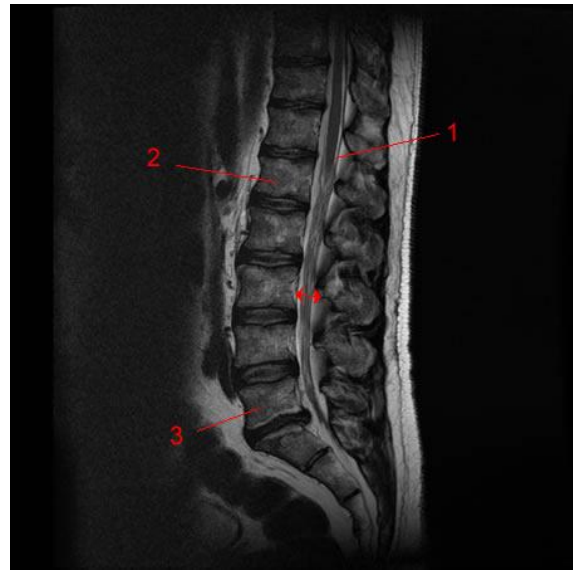


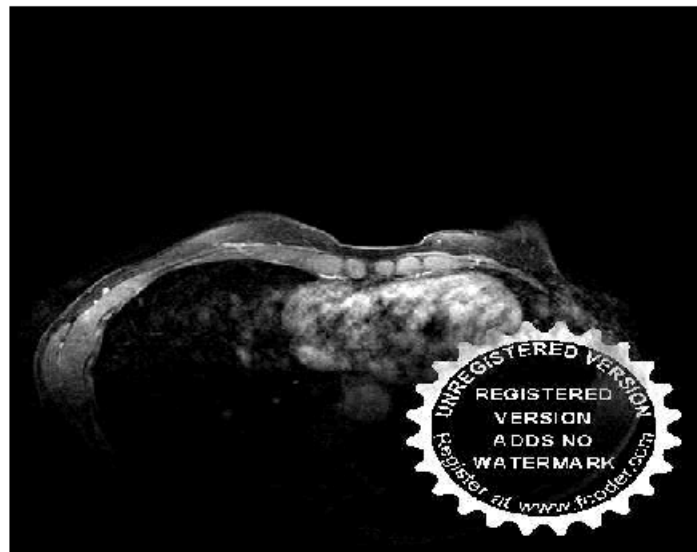
FIG.IV.3: *Les images de simulation partie 1*



e



f



g

FIG.IV.3: *Les images de simulation partie 2*

Figure	Nom des images	Taille	Type	Commentaires
a	Coupe axiale du cerveau	256×282	RMN	cas sain
b	Coupe sagittale du cerveau	256×282	RMN	cas sain
c	Coupe abdominale	256×282	RMN	cas sain
d	Coupe frontale du Cerveau	200×290	RMN	présence d'anomalie (tumeur cérébrale)
e	Coupe de genou	256×256	RMN	cas sain
f	Moelle épinière	142×204	RMN	cas sain
g	mammographie	512×512	RMN	cas sain

Tab. IV.1 - *Caractéristiques des images utilisées.*

IV.4.1 Présentation des résultats

Les exemples traités sont des images en niveau de gris de taille différentes représentés dans le Tab.IV.1. Cette série d'expérimentations concerne des images réelles, pour lesquelles il n'existe pas de valeur exacte du nombre de classes. Cependant il est possible d'obtenir un avis d'expert sur un nombre de classes approprié en fonction du contexte.

L'objectif de ces exemples étant d'obtenir des segmentations en composantes homogènes facilement exploitable et l'évaluation de résultat de segmentation des images synthétiques par des champs markoviens cachés.

Les segmentations de cet image obtenue à partir des champs de Markov cachés . Nous avons utilisé comme modèle a priori un modèle de Potts simple avec un voisinage de huit connexions et un modèle gaussien pour le terme d'attache aux données de moyennes et de variances distinctes.

Le nombre d'itérations de l'algorithme d'ICE est mis à 30 itérations, avec 20 réalisations a posteriori par itération. La valeur initiale du paramètre de régularité du champ de Markov caché β est 0.8. Dans chaque itération d'ICE, le nombre maximal d'itérations pour le gradient stochastique est mis à 10, avec une réalisation *a priori* par itération. L'évaluation de gradient stochastique exige généralement très peu d'itérations. L'échantillonneur de Gibbs qu'on utilise a besoin de 100 itérations pour générer les réalisations a priori et a posteriori. La convergence des énergies globales est en fait tout à fait lente, particulièrement pour des réalisations selon la distribution *a priori*. La classification MPM basée sur les champs de Markov cachés compte sur 100 réalisations a posteriori.

Dans le but de réduire le temps de calcul, on a simulé seulement une réalisation a posteriori pour chaque itération d'ICE de la segmentation par champ de Markov cachés et seulement une réalisation a priori pour chaque itération du gradient stochastique pour l'estimation du paramètre de régularité b de la segmentation par champ de Markov caché.

Cette simplification ne cause aucune perte de performance significative.

IV.4.2 Analyse et discussion

En générale, les résultats obtenus sont satisfaisants qui présente des performances semblables en terme de segmentation. Si l'on analyse les variations des segmentations au point de vue visuel et au sens du critère d'évaluation, on remarque qu'elles sont cohérentes. Mais on peut tirer quelques remarques :

- Afin d'assurer la convergence des algorithmes, nous avons refait les expériences pour toutes les images avec 100 itérations : aucune dégradation des segmentations n'est observée.
- La qualité complète est comparable pour le résultat de classification, mais le temps de calcul est tout à fait différent. En particulier, la procédure d'échantillonneur de Gibbs se montre très coûteuse, puisqu'elle nécessite plus d'une 2 heure pour finir un résultat de segmentation.
- La distribution gaussienne est bien adaptée aux images de type RMN.
- Dans le cas de segmentation par les champs de Markov cachés, la valeur du paramètre de régularité $\beta = 0,8$.

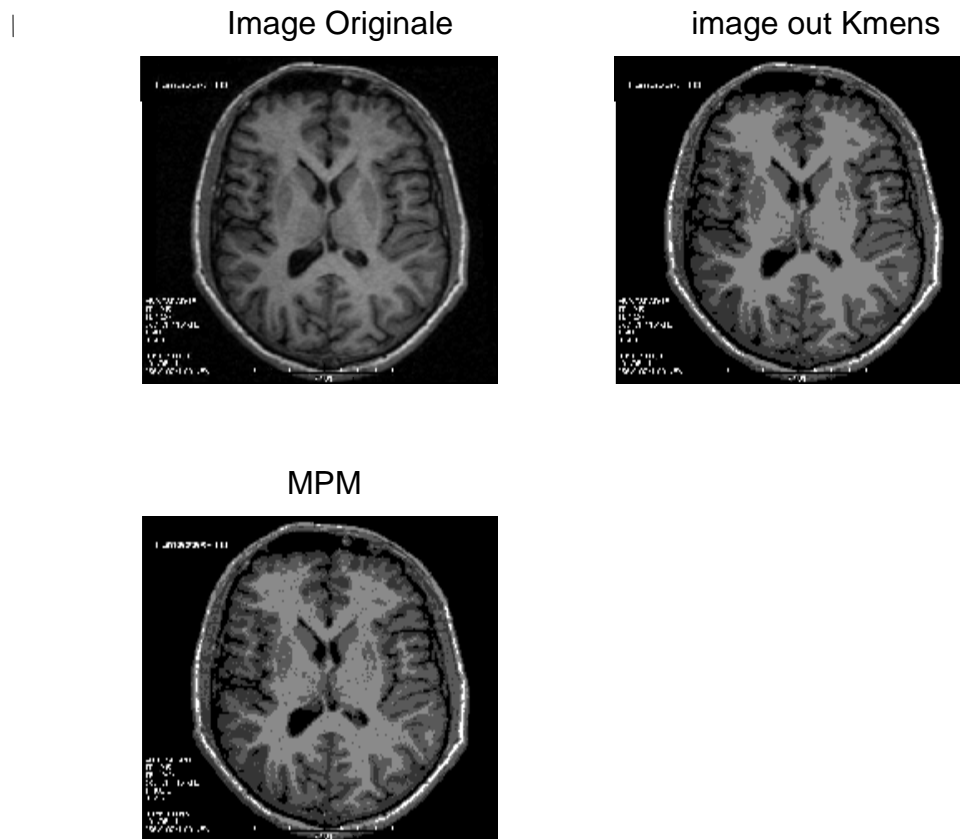


Fig. IV.4 : Résultat de segmentation de l'image « coupe axiale du cerveau » :

Image	Nbr classes	PSNR	Temps calc
<i>coupe axiale du cerveau</i>	5	35.2386	2h.26min

Tab. IV.2 – informations sur la segmentation d'image de « coupe axiale du cerveau »

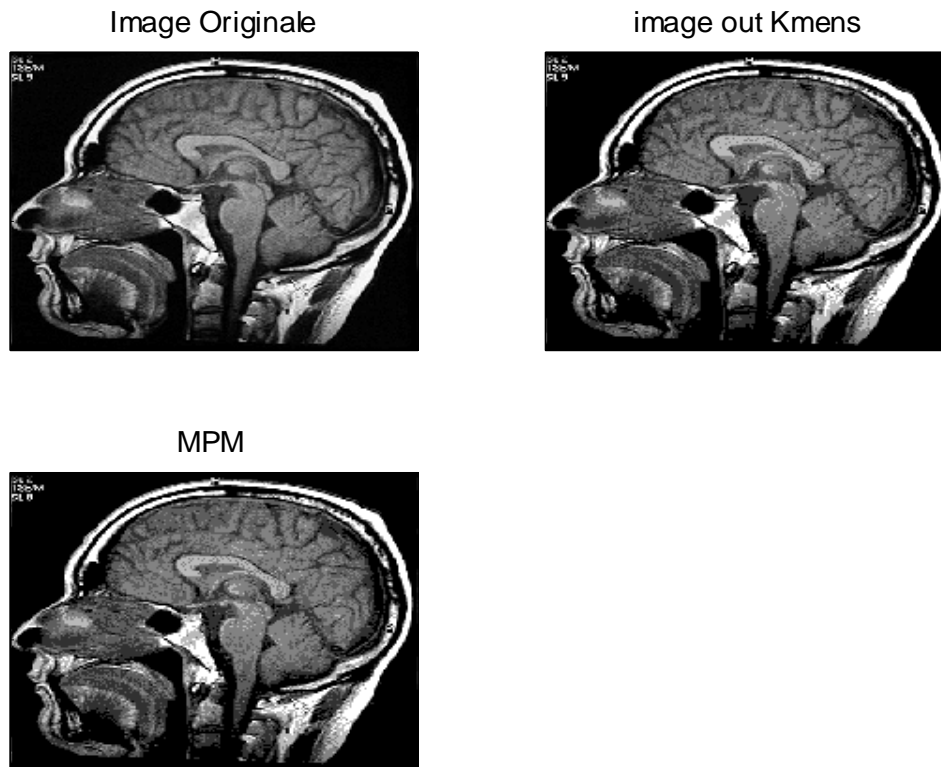


Fig. IV.5 : Résultat de segmentation de l'image « coupe sagittale du cerveau »

Image	Nbr classes	PSNR	Temps calc
<i>coupe sagittale du cerveau</i>	6	33.8160	2h56min

Tab. IV.3– informations sur la segmentation d'image de « coupe sagittale du cerveau »

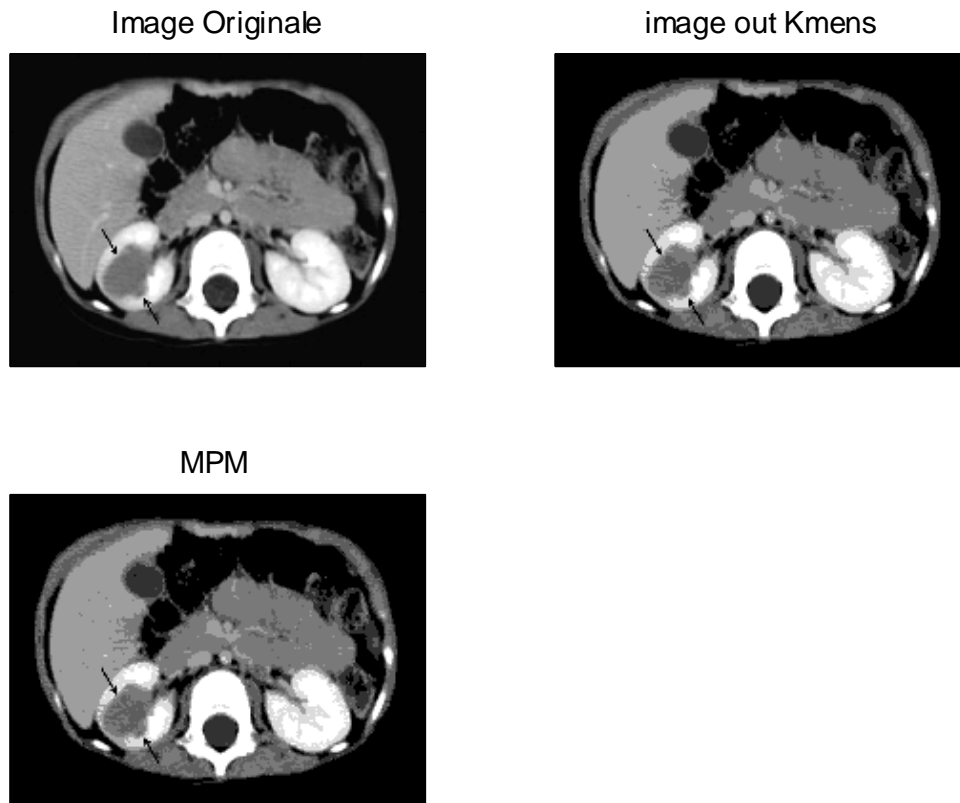


Fig. IV.6 : Résultat de segmentation de l'image « coupe abdominale »

Image	Nbr classes	PSNR	Temps calc
<i>coupe abdominale</i>	7	36.5304	2h33min

Tab. IV.4 – informations sur la segmentation d'image de « coupe abdominale »

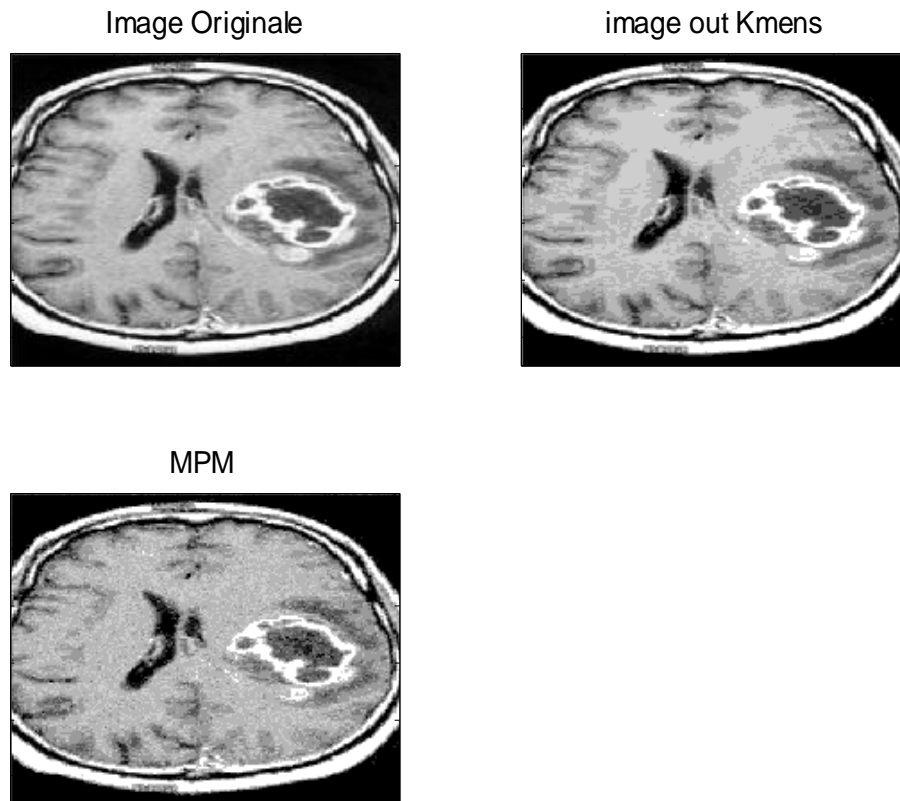


Fig. IV.7: Résultat de segmentation de l'image « coupe frontale du cerveau »

Image	Nbr classes	PSNR	Temps calc
<i>coupe frontale du cerveau</i>	8	25.8450	2h16min

Tab. IV.5 informations sur la segmentation d'image de « coupe frontale du cerveau »

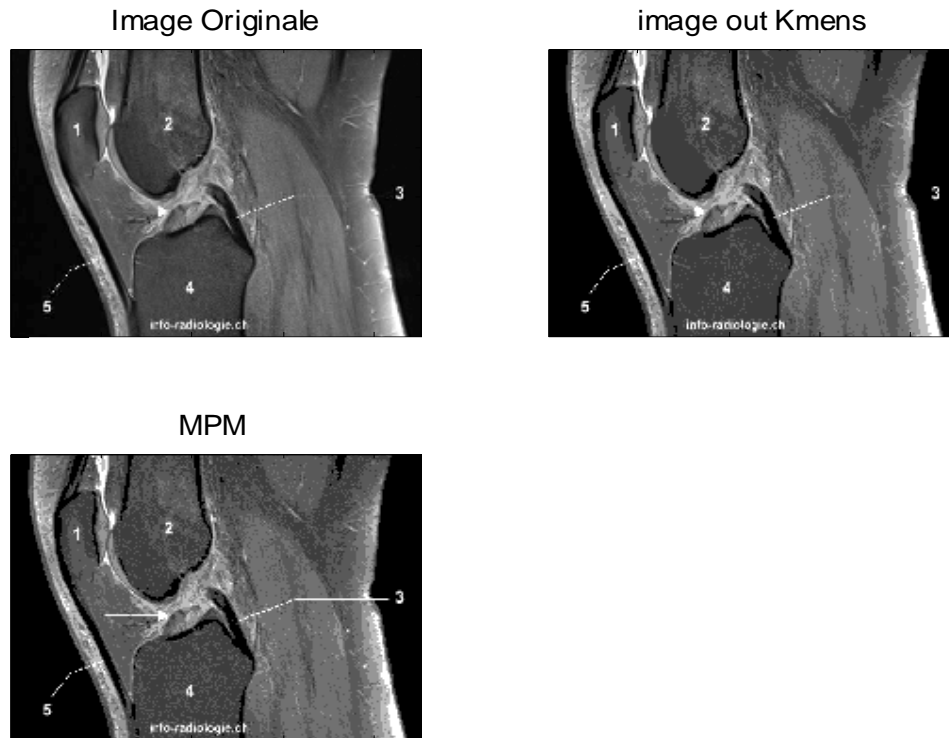


Fig. IV.8: Résultat de segmentation de l'image « coupe de genou »

Image	Nbr classes	PSNR	Temps calc
<i>coupe de genou</i>	7	31.8595	2h32min

Tab. IV.6 – informations sur la segmentation d'image de « coupe de genou »

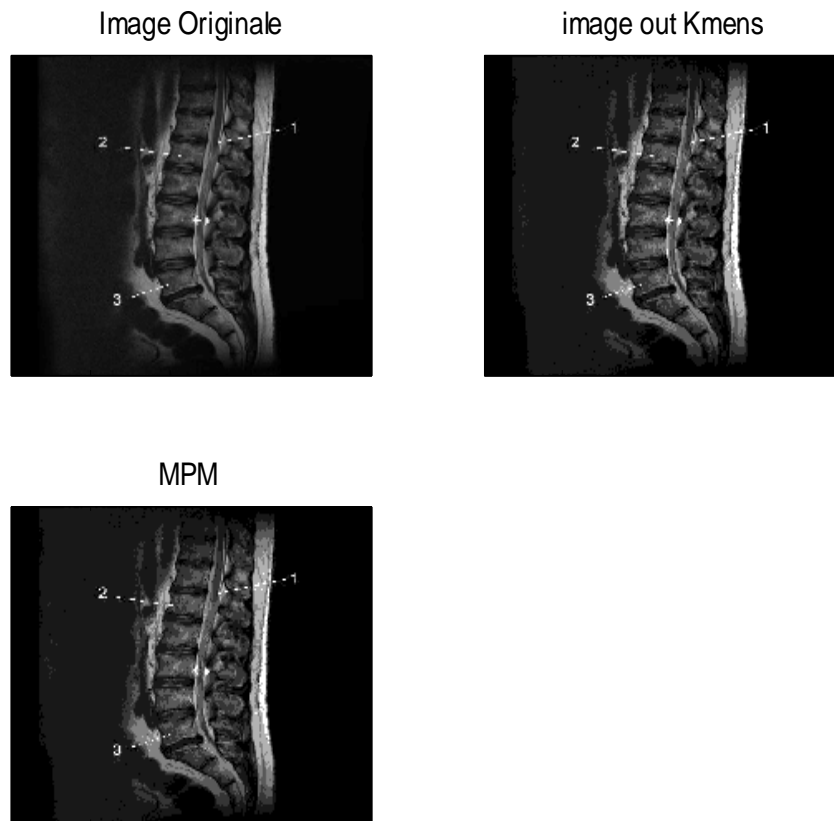


Fig. IV.9: Résultat de segmentation de l'image «Moelle épinière»

Image	Nbr classes	PSNR	Temps calc
Moelle épinière	9	44.66.4	3h05min

Tab. IV.7 – informations sur la segmentation d'image de «Moelle épinière»

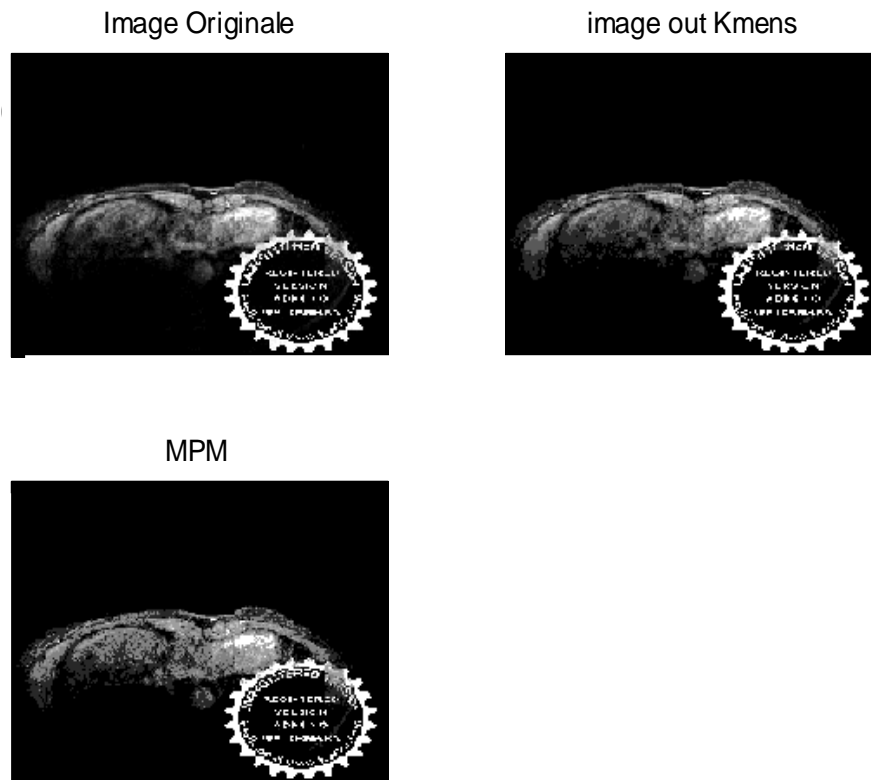


Fig. IV.10: Résultat de segmentation de l'image «mammographie»

Image	Nbr classes	PSNR	Temps calc
mammographie	7	29.9707	2h56min

Tab. IV.8 – informations sur la segmentation d'image de «mammographie»

IV.5 Conclusion

Ce travail décrit la classification non supervisée des images RMN et mammographies dans le cadre de champ de Markov caché et l'estimation gaussien. Les champs de Markov cachés sont fréquemment employés pour imposer des contraintes de régularité spatiales dans l'estimation de paramètre et les étapes de classification. Cette approche produit des résultats excellents dans de nombreux cas, mais nos expériences indiquent que l'estimation du paramètre de régularité β est un problème délicat, particulièrement quand le taux du signal du bruit est bas. Le temps considérable de calcul est un autre inconvénient. Les frontières des régions des structures excellentes sont généralement mieux préservées. En plus de l'estimation robuste des paramètres de régularité

CONCLUSION

Ce travail décrit la classification non supervisée des images dans le cadre des modèles de Markov cachés et l'estimation d'un mélange gaussien. Les champs de Markov cachés sont fréquemment employés pour imposer des contraintes de régularité spatiales dans l'estimation de paramètre et les étapes de classification. Cette approche produit des résultats excellents dans de nombreux cas, mais nos expériences indiquent que l'estimation du paramètre de régularité est un problème délicat. Le temps considérable de calcul est un autre inconvénient. Les frontières de région sont souvent légèrement irrégulières avec cette approche, mais des structures excellentes sont généralement mieux préservées. Plusieurs faits ressortent de cette étude expérimentale. D'abord, l'intérêt de la prise en compte de l'information spatiale pour la classification des pixels d'une image apparaît de manière évidente dans les différents exemples traités. Comme il fallait s'y attendre, le choix d'un modèle de mélange gaussien conduit à des segmentations acceptables et l'on observe une nette amélioration au niveau de l'homogénéité de la segmentation avec un modèle spatial. Ce résultat est satisfaisant car cela confirme, dans le cas du champ, que les approximations d'un modèle de Markov caché utilisé préservent l'information markovienne.

Algorithme d'initialisation k-means

arguments : Taille K (nombre de classes à déterminer).

initialisation : $C_0 = \{y_i, i = 1, 2, \dots, K\}$

tant que $C_m \neq C_{m-1}$ **faire**

- Affectation des éléments de la séquence d'apprentissage

pour tout vecteur z de la séquence d'apprentissage **faire**

si $\forall 1 \in \{1, 2, \dots, K\}, |z - y_i| \leq |z - y_j|$ **alors**

$z \in R_i$;

fin si

fin pour

- Mise à jour des centres de classes

Pour $i = 1, \dots, K$ **faire**

$y_i = \text{centroïde}(\{z / z \in R_i\})$;

fin pour

fin tant que

Echantillonneur de Gibbs

Soit x_0 la configuration initiale

Répéter indéfiniment :

Soit x^q la configuration courante

Pour tous les sites t de S :

- Poser

$$\hat{x}_s = x_s^q, \forall s \neq t \text{ et } x_s^{q+1} = x_s^q, \forall s \neq t$$

- Pour chaque ω_I de Ω :

- Poser $\hat{x}_s = \omega_I$

- Calculer

$$P(\omega_I) = P(x_s = \hat{x}_s | (X_t = x_t)_{t \in V_s}) = \frac{1}{Z_s} \exp(-U_s(\hat{x}_s, (x_t)_{t \in V_s}))$$

- Tirer u selon loi uniforme dans $[0,1]$

- Trouver le plus petit j tel que

$$\sum_{i=1}^j P(\omega_I) > \mu \sum_{i=1}^{|\Omega|} P(\omega_I)$$

- Poser $x_s^{q+1} = j$

Algorithme ICE

Initialisation des paramètres d'a priori $\lambda^{q,0}$ et des classes $f_i^0 (\mu_i^0, \sigma_i^0)$.

Répéter jusqu'à la convergence

Tirage des échantillons $x^q(1), \dots, x^q(t), \dots, x^q(m)$ selon

$$P(X_s = x_s | Y_s = y_s, (X_t = x_t)_{t \in V_s})$$

Estimation des paramètres de classe pour chaque échantillon :

$$\hat{\mu}_i^{q,t} = \frac{\sum_{n=1}^N y_n \delta(x_n - \omega_i)}{\sum_{n=1}^N \delta(x_n - \omega_i)}$$

$$(\hat{\sigma}_i^{q,t})^2 = \frac{\sum_{n=1}^N (y_n - \mu_i^q)^2 \delta(x_n - \omega_i)}{\sum_{n=1}^N \delta(x_n - \omega_i)}$$

Tirage des échantillons $x^q(1), \dots, x^q(t), \dots, x^q(m)$ selon

$$P(X_s = x_s | (X_t = x_t)_{t \in V_s})$$

Estimation des paramètres de régularité pour chaque échantillon :

$$\lambda^{q,r} = \lambda^{q,r-1} + \frac{1}{r} \frac{U_{apriori}^{q,r} - U_{aposteriori}^q}{U_{aposteriori}^q}$$

Remise a jour

$$\mu_i^q = \frac{1}{m} \sum_{t=1}^m \hat{\mu}_i^{q,t}$$

$$(\hat{\sigma}_i^q)^2 = \frac{1}{m} \sum_{t=1}^m (\hat{\sigma}_i^{q,t})^2$$

$$\lambda^q = \frac{1}{m} \sum_{t=1}^m \hat{\lambda}^{q,r}$$

Algorithme MPM

- à partir d'une configuration initiale aléatoire, générer une suite de m réalisations $(x(n), \dots, x(m))$ selon la loi a posteriori $P(X/Y)$.
- éliminer les p premiers échantillons pour limiter l'impact de la configuration initiale
- Procédure de la classification

pour tous les sites s de S :

- approcher la marginale a posteriori .
- chercher l'étiquette qui maximise la marginale approchée, c'est-à-dire celle qui apparaît le plus fréquemment dans la suite $(x_s(p + 1), \dots, x_s(m))$.

Les algorithmes présentés ici sont simulés par MATLAB 2010 version 7.10.0 ,et pour les avoir n'hésitez pas a me contacter sur aminekherroubi@yahoo.fr.

Et pour la partie d'évaluation nous avons utilisé l'algorithme psnr toujours sur le matlab .

Nous avons aussi pris les images a simulées ou bien la base de données a partir de :

www.info-radiologie.ch

www.fooder.com

- [**Ati05**] J-C. Atine. Méthodes d'apprentissage floue : application à la segmentation d'images biologiques. PhD thesis, Institut national des sciences appliquées de Toulouse, 2005.
- [**AYFM02**] M. N. Ahmed, S. M. Yamany, A. A. Farag, and T. Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE Transactions on medical imaging*, 21(3) :193–199, 2002.
- [**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.
- [**Bes 74**] J. Besag. _ Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statist. Soc., Série B*, 36:192_236, 1974.
- [**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.
- [**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.
- [**Coc,phi 1995**] J-P.Cocquerez et S.Philipp. *Analyse d'images : filtrage et segmentation*. Masson. 1995.
- [**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.
- [**Did71**] E Diday. *La méthode des nuées dynamiques*. *Revue de statistique appliquée*, XIX,2 :19–34, 1971.
- [**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.
- [**Dur 2003**] J-B. Durand. *Modèles à structure cachée : inférence, élection de modèles et applications*. Thèse de Doctorat, Université de Joseph Fourier – Janvier 2003.
- [**FU,Mui 1981**] K. FU et J. MUI, A survey on image segmentation, *Pattern Recognition*, Vol. 13, 1981, pp. 3–16.
- [**EAM 95**] R. J. Elliott, L. Aggoun and J.B. Moore, “Hidden Markov Models: Estimation

and Control”, Springer-Verlag, 1995.

[Gé98] T Géraud. Segmentation des structures internes du cerveau en imagerie par résonance magnétique. PhD thesis, ENST Paris, 1998.

[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.

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.

[Guy 95] X. Guyon, “Random fields on network. modeling, statistics, and applications” Springer-Verlag, Probability and its Applications, New York, 1995.

[Har05] R Haroun. Segmentation des tissus cérébraux sur des images par résonance magnétique. Master’s thesis, Université des sciences et de la technologie Houari Boumediène, 2005

[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.

[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.

[KKM] M. Khouadjia, H. Khanfouf, and S. Meshoul. Une approche adaptative pour la segmentation d’images : Implémentation sur la plate-forme multi-agents netlogo.

[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.

[Mar82] D. Marr. Vision. W.H. Freeman, 1982.

[Mai 03] H. Maître, “Le traitement des images”, Hermès, 2003.

[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.

[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.

[Pey 2001] N. Peyrard. Approximation de type champ moyen des modèles de champ de Markov pour la segmentation de données spatiales, Thèse de Doctorat, Université de Joseph Fourier – Octobre 2001.

[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.

[Ros08] J-L. Rose. Croissance de région variationnelle et contraintes géométriques tridimensionnelles pour la segmentation d'image. PhD thesis, Institut National des Sciences Appliquées de Lyon, 2008.

[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.

[STM07] M. Semchedine, L. Toumi, and A. Moussaoui. Classification adaptative robuste pour la segmentation d'images irm cérébrales. JIG'2007 - 3èmes Journées Internationales sur l'Informatique Graphique, 2007.

[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", 3

rd

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.

[**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.

[**Zad65**] L. A. Zadeh. Fuzzy sets. Information Control, 13(8) :338 353, 1965.

[**Zuc76**] S.W. Zucker. Region growing : Childhood and adolescence. Graphics and Image Processing, 5(3) :382–399, 1976. [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. [Zuc76] S.W. Zucker. Region growing : Childhood and adolescence. Graphics and Image Processing, 5(3) :382–399, 1976.

les sites web

<http://www.developpez.net/forums/f148/envIRONNEMENTS-developpement/matlab/>

<http://www.tsi.telecom-paristech.fr/pages/enseignement/ressources/beti/markov/code.txt>

<http://www-public.it-sudparis.eu/~pieczyn/B1.pdf>

<http://www.fresnel.fr/perso/derrode/publi/Taima03.pdf>

.