N° d'ordre : 5790

UNIVERSITÉ LOUIS PASTEUR STRASBOURG

THÈSE

présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur - Strasbourg I

École Doctorale	: Mathématiques, Sciences de l'Information et de l'Ingénieur
Discipline	: Électronique, Électrotechnique, Automatique
Spécialité	: Traitement d'images et vision par ordinateur

Segmentation d'images IRM anatomiques par inférence bayésienne multimodale et détection de lésions

Soutenue publiquement
le 06 novembre 2008
par

Stéphanie BRICQ

Membres du jury :

M.	Christian	BARILLOT	Rapporteur externe	IRISA, INRIA Rennes-Bretagne Atlantique
Mme	Su	RUAN	Rapporteur externe	CReSTIC, Univ. de Reims
M.	Fabrice	HEITZ	Rapporteur interne	LSIIT, Univ. de Strasbourg
M.	Wojciech	PIECZYNSKI	Examinateur	Telecom & Management SudParis, Evry
M.	Christophe	COLLET	Directeur de thèse	LSIIT, Univ. de Strasbourg
M.	Jean-Paul	ARMSPACH	Directeur de thèse	LINC, Univ. de Strasbourg

Équipes d'accueil :

Laboratoire des Sciences de l'Image,
de l'Informatique et de la Télédétection
UMR CNRS 7005

Laboratoire d'Imagerie
et de Neurosciences Cognitives
UMR CNRS 7191

Remerciements

Je remercie Madame Su Ruan, Professeur à l'Université de Reims, Monsieur Christian Barillot, Directeur de recherche à l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA, Rennes), et Monsieur Fabrice Heitz, Professeur à l'Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS, Strasbourg) pour l'intérêt qu'ils ont porté à mes travaux en acceptant la tâche fastidieuse de rapporteurs. Je remercie également Monsieur Wojciech Pieczynski, Professeur à l'INT Evry d'avoir accepté de présider mon jury de thèse.

Je remercie mes directeurs de thèse Monsieur Christophe Collet, Professeur à l'Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS, Strasbourg), et Monsieur Jean-Paul Armspach, Ingénieur de recherche au Laboratoire d'Imagerie et de Neurosciences Cognitives (LINC, Strasbourg) qui ont su me communiquer tout au long de ces trois années leur enthousiasme et leur motivation. Je les remercie également pour les conseils qu'ils m'ont prodigués pendant ces trois années.

Je tiens également à remercier les membre de l'équipe IGG du LSIIT et plus particulièrement Dominique Bechmann, Sylvain Thery et David Cazier pour m'avoir permis d'utiliser leur plateforme de modélisation géométrique et pour leurs conseils pour la réalisation de la dernière partie de cette thèse.

Je remercie mes collègues de l'équipe MIV et du LINC-Médecine : Bessem, Matthieu, Hervé, Felix, Christelle, Steven, Benoît, Slim et Vincent Mazet pour leur gentillesse et leur bonne humeur. Je remercie particulièrement François Rousseau pour ses conseils et son aide lors du challenge MICCAI'08.

Je tiens à remercier mes parents, mon frère et ma famille pour leur soutien pendant toutes ces années.

Enfin merci à Jawad pour son soutien pendant ces trois années et pour son aide précieuse et ses conseils lors de la préparation de la soutenance.

Table des matières

Notations

Notations

Modélisation markovienne

X	Champ des étiquettes
Y	Champ des observations
x	Réalisation du champ des variables cachées
y	Réalisation du champ des observations
G	Graphe
S	Ensemble de sites ou nœuds
L	Ensemble d'arêtes dans un graphe
$\mathbb{E}(x)$	Espérance mathématique de x
$\mathcal{L}(.,.)$	Fonction de coût
$\mathcal{V}_s$	Ensemble des voisins du site s
$\mathcal{N}$	Système de voisinage
c	Clique relative à un système de voisinage
C	Ensemble de cliques
$pa(s)$	Ensemble des parents de s
$ch(s)$	Ensemble des enfants de s
$U(x)$	Energie associée au champ x
$m_{ij}(.)$	Message passé du nœud i au nœud j
t_{ij}	Probabilité de transition de la classe i à la classe j
π_i	Probabilité initiale d'appartenance à la classe i
$f_k(y_n)$	Vraisemblance des observations y_n conditionnellement à $x_n = \omega_k$
Φ_x	Paramètres *a priori*
Φ_y	Paramètres de la loi d'attache aux données
Φ	Ensemble des paramètre Φ_x et Φ_y
$\alpha_n(k)$	Probabilité *Forward*
$\beta_n(k)$	Probabilité *Backward*
$b_n(k)$	Probabilité du voxel n d'appartenir à la classe k donnée par l'atlas

Estimation robuste

Ψ	Ouvert
w_i	Pondération
h	Paramètre de tamisage
ν	Permutation
$\hat{\theta}$	Estimateur de θ
H	Sous-ensemble des observations
Q	Estimateur du MV sur le sous-ensemble H
r_n	Résidu

Modélisation par contours actifs statistiques

s	Image
w	Partition associée à l'image
Ω_a	Région définie par la cible
Ω_b	Région définie par le fond
Θ	Ensemble des vecteurs de paramètres
$\Delta(s, w, \Theta)$	Complexité stochastique
$\Delta_G(w)$	Nombre de nats nécessaires pour coder la partition w
$\Delta_L(s\|\theta, w)$	Terme de codage de l'image connaissant le modèle
N	Nombre de voxels dans l'image
k	Nombre de triangles définissant le maillage
p	Nombre de segments
Δ_{w^k}	Longueur de description du maillage surfacique
Δ_a	Terme de codage des niveaux de gris de la cible
Δ_b	Terme de codage des niveaux de gris du fond
$\mathcal{L}_e[.\|.]$	Log-vraisemblance
P_m	ddp de paramètre m
S	Entropie
P_i	$i^{\text{ème}}$ nœud
m_i	Longueur moyenne des segments reliés au nœud i
a_i	Amplitude de déplacement associée au nœud i

Acronymes

BET	*Brain Extraction Tool*
CAS	Contour Actif Statistique
CHB	*Children's Hospital Boston*
CMC	Chaîne de Markov Cachée
CMC-BI	Chaîne de Markov cachée à bruit indépendant
CMCouple	Chaîne de Markov Couple
CMT	Chaîne de Markov Triplet
CSMC	Chaîne semi-markovienne cachée
ddp	densité de probabilité
EM	*Expectation-Maximisation*
EMS	*Expectation-Maximisation Segmentation*
FCM	*Fuzzy C-Means*
FPF	*False Positive Fraction*
GAS	Grille Active Statistique
HMC	*Hidden Markov Chain*
IBSR	*Internet Brain Segmentation Repository*
IRM	Imagerie par Résonance Magnétique
KI	Kappa Index
LCR	Liquide céphalo-rachidien
MAP	*Maximum a Posteriori*
MB	Matière Blanche
MCMC	*Markov Chain Monte Carlo*
MCS	Minimisation de la complexité stochastique
MDL	*Minimum Description Length*
MG	Matière Grise
MRF	*Markov Random Field*
PDF	*Probability Density Function*
PV	*Partial Volume*
RMS	*Root Mean Square*
SEP	Sclérose en Plaques
SPM	*Statistical Parametric Mapping*
TLE	*Trimmed Likelihood Estimator*
TPF	*True Positive Fraction*
UNC	*University of North Carolian*

Introduction

Contexte du travail

L'imagerie médicale, en constante évolution ces dernières années, fournit un nombre croissant de données. Ce volume important de données doit ensuite être analysé. Les méthodes automatiques de traitement et d'analyse d'images se sont récemment multipliées pour assister l'expert dans l'analyse qualitative et quantitative de ces images et faciliter son interprétation. Ces méthodes doivent prendre en considération d'une part la quantité des données à analyser et d'autre part, la complexité structurelle des images IRM. Parmi ces méthodes, la segmentation fiable et précise des images IRM 3D anatomiques multimodales, normales ou pathologiques, reste un objectif premier en analyse d'images médicales car elle constitue un préalable incontournable pour différentes études telles que l'atrophie/hypotrophie cérébrale en utilisant entre autres les méthodes d'analyse de type Voxel Based Morphometry (VBM), la mesure de la charge lésionnelle dans la pathologie de la Sclérose En Plaques ou encore la visualisation surfacique 3D. Cependant une segmentation manuelle effectuée par un médecin s'avère être une tâche fastidieuse montrant une grande variabilité inter- et intra-expert, d'où l'intérêt de développer des méthodes automatiques de segmentation.

Ce travail de thèse a été mené au sein du Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (LSIIT, UMR 7005, directeur de thèse Ch. Collet) et du Laboratoire d'Imagerie et de Neurosciences Cognitives (LINC, UMR 7191, co-directeur de thèse J.-P. Armspach) dans le cadre du consortium de recherche Imagerie et Robotique Médicale et Chirurgicale (IRMC).

Objectifs

Dans cette thèse nous focalisons notre étude sur des images IRM cérébrales et nous proposons des méthodes automatiques de segmentation. Un grand nombre de méthodes de segmentation ont été proposées depuis plusieurs années et on distingue deux grandes approches :

- ❖ celles dites de « bas niveau » qui s'appuient classiquement sur la réunion de voxels similaires (au sens des « niveaux de gris », de la « couleur » ou de la « texture ») dans les régions 3D ou sur la détection des surfaces frontières entre les différentes structures d'intérêt. Bien qu'elles offrent généralement une bonne précision dans la localisation des frontières, ces méthodes ne donnent des résultats satisfaisants que si les contrastes entre les structures d'intérêt sont suffisamment marqués, si les rapports signal à bruit sont suffisamment élevés et si l'image est exempte d'artefacts. Ces conditions sont rarement réunies en pratique, et les méthodes de bas niveau requièrent donc le plus souvent une correction manuelle des erreurs de segmentation. Il va sans dire que ces corrections peuvent s'avérer délicates dans le cas de structures spatiales

tridimensionnelles ;

- ❖ celles qui s'appuient au contraire sur des modèles a priori, de nature déterministe ou probabiliste, qui de plus en plus prennent la forme de modèles déformables ou d'atlas numériques.

L'objectif de ce travail de thèse est de coupler les méthodes dites de « bas niveaux » aux approches plus modernes qui s'appuient sur des modèles *a priori*. Nous avons donc associé des modèles statistiques markoviens [Pieczynski 03a] à des modèles *a priori* de haut niveau de type atlas statistiques. La difficulté à segmenter une image est due à la complexité structurelle des images IRM et au contraste souvent insuffisant pour extraire la structure d'intérêt, sans aucune connaissance *a priori* ni sur sa forme ni sur sa localisation. Les connaissances *a priori* apportées par les atlas probabilistes permettront donc de mieux différencier les structures anatomiques proches par le contraste et la localisation spatiale.

Contributions

La méthode que nous proposons est basée sur la modélisation par Chaînes de Markov Cachées permettant de prendre en compte partiellement la notion de voisinage 3D avec une réelle efficacité. Elle permet également d'inclure l'information *a priori* apportée par un atlas probabiliste. En outre la méthode proposée prend en considération les hétérogénéités d'intensité présentes sur les images IRM ainsi que les effets de volume partiel en proposant une segmentation en cartes de tissus contenant les proportions de chaque tissu en chaque voxel [$R1$][$C1$][$C7$].

Cette méthode a été appliquée à la segmentation des tissus cérébraux puis étendue en utilisant un estimateur robuste à la localisation et la détection des lésions de Scléroses En Plaques (SEP) observées en IRM. Grâce à cet estimateur robuste et à l'information *a priori* apportée par l'atlas probabiliste, les lésions SEP vont être détectées comme des données atypiques ("outliers") par rapport à un modèle statistique d'images cérébrales non pathologiques [$C2$][$C4$][$C5$][$C6$].

Enfin nous avons développé une deuxième méthode de segmentation d'IRM 3D basée sur les contours actifs statistiques et le principe de la minimisation de la complexité stochastique [Galland 05], pour raffiner la segmentation des lésions en ayant une information plus précise sur le volume et la forme des lésions, ouvrant ainsi la voie à la quantification de leur évolution. Cette méthode permet également de segmenter différentes structures anatomiques cérébrales.

Les différentes méthodes développées ont été validées à la fois sur des bases d'images synthétiques et réelles. Les résultats obtenus ont été comparés avec d'autres méthodes classiques de segmentation, en particulier SPM5[1] (Statistical Parametric Mapping) [Ashburner 05] et EMS[2] (Expectation-Maximization Segmentation) [Van Leemput 99b], ainsi qu'avec des segmentations manuelles réalisées par des experts.

Organisation du manuscrit

Le manuscrit est organisé en six chapitres présentant les contributions essentielles du travail doctoral. Dans le premier chapitre nous décrivons le principe de l'imagerie par résonance magnétique

[1] http ://www.fil.ion.ucl.ac.uk/spm

[2] http ://www.medicalimagecomputing.com/EMS/

et les artefacts associés à l'acquisition des images. Les principales méthodes de segmentation en imagerie cérébrale sont ensuite abordées.

Le deuxième chapitre débute par le rappel de la théorie bayésienne et enchaîne sur la présentation générale des modèles graphiques. Nous présentons ensuite les différents modèles markoviens utilisés en imagerie. Puis nous décrivons la modélisation par chaînes de Markov Triplet et divers cas particuliers, en détaillant plus particulièrement le modèle des chaînes de Markov cachées à bruit indépendant qui sera utilisé pour la segmentation des tissus cérébraux. L'estimation des paramètres de ce modèle est ensuite abordée.

Le chapitre 3 décrit la méthodologie employée pour la segmentation des tissus. Cette méthode prend en compte les principales imperfections pouvant apparaître sur les images IRM (bruit, hétérogénéités d'intensité et effets de volume partiel). L'information sur le voisinage est prise en compte via la modélisation par chaînes de Markov. Nous expliquons comment nous avons intégré de l'information *a priori* apportée par un atlas probabiliste dans le modèle. Nous présentons ensuite les différents prétraitements nécessaires avant d'appliquer la méthode de segmentation proposée. Ces différentes contributions sont ensuite validées à la fois sur la base d'images synthétiques Brainweb[3] et sur la base d'images réelles IBSR[4].

Le chapitre 4 décrit l'extension de la méthode pour la détection de lésions de sclérose en plaques en utilisant un estimateur robuste. Nous commençons par présenter la sclérose en plaques ainsi que les principales méthodes de détection de lésions existantes. Nous détaillons ensuite la méthode que nous proposons pour détecter les lésions. Cette méthode est également basée sur la modélisation par chaînes de Markov. Les lésions sont détectées comme des données atypiques par rapport à un modèle statistique d'images cérébrales non pathologiques en utilisant un estimateur robuste. L'information *a priori* apportée par un atlas probabiliste est également intégrée dans le modèle. Puis nous proposons différents post-traitements afin de garder seulement les outliers correspondant à des lésions SEP. Cette méthode est validée d'abord sur des images synthétiques provenant de la base Brainweb et sur des données cliniques en comparant les résultats obtenus par rapport aux segmentations de deux médecins. Nous exposons ensuite les résultats obtenus lors du challenge MICCAI'08[5] sur la segmentation des lésions SEP.

Enfin le chapitre 5 présente une méthode de raffinement de la segmentation des lésions basée sur les contours actifs statistiques, permettant d'avoir une information plus précise sur le volume des lésions. Cette méthode permet de plus de segmenter différentes structures anatomiques cérébrales. Dans un premier temps, nous décrivons le principe de minimisation de la complexité stochastique sur lequel est basée la méthode de segmentation proposée, puis nous présentons l'algorithme de segmentation basé sur une stratégie multi-résolution. Cette méthode est ensuite validée sur des images synthétiques 3D ainsi que sur des images réelles de lésions et de structures anatomiques cérébrales.

Le dernier chapitre présente les conclusions et les perspectives de ce travail de thèse.

[3] `http://www.bic.mni.mcgill.ca/brainweb/`
[4] `http://www.cma.mgh.harvard.edu/ibsr/`
[5] `http://www.ia.unc.edu/MSseg/index.php`

Chapitre 1

Segmentation d'IRM cérébrales

L'imagerie médicale, en constante évolution ces dernières années, fournit un nombre croissant de données : résolution croissante des imageurs IRM, accès à des modalités variées... Les méthodes automatiques de traitement et d'analyse d'images se sont récemment multipliées pour assister l'expert dans l'analyse qualitative et quantitative de ces images et faciliter son interprétation. La segmentation automatique des tissus du cerveau est devenue une étape fondamentale pour ces analyses quantitatives des images dans de nombreuses pathologies cérébrales telles que la maladie d'Alzheimer, la schizophrénie ou la sclérose en plaques. Elle fait également partie de phases de prétraitements dans de nombreux algorithmes tels que VBM (*Voxel-Based Morphometry*). La délimitation manuelle des principaux tissus du cerveau par un expert est en effet trop longue et fastidieuse pour traiter un volume important de données d'où l'intérêt d'avoir des méthodes automatiques de segmentation. La variabilité inter- et intra-observateur est de plus élevée ce qui nécessite le développement de méthodes reproductibles. De nombreuses approches ont été proposées pour segmenter les IRM cérébrales. On peut les classer en deux groupes : les méthodes supervisées qui requièrent une interaction avec l'utilisateur et les méthodes non supervisées, où l'estimation des paramètres est réalisée automatiquement. Ce chapitre présente un état de l'art des différents problèmes liés à la segmentation d'images IRM et fait le point sur les différentes méthodes de segmentation existantes. Nous commencerons par décrire l'anatomie cérébrale puis nous détaillerons le principe de l'imagerie par résonance magnétique. Nous présenterons ensuite les caractéristiques des IRM cérébrales avec les différents artefacts pouvant apparaître et nous détaillerons les principales méthodes de segmentation existantes prenant en compte ces artefacts. Enfin nous positionnerons l'innovation de l'approche que nous proposons par rapport à ces méthodes.

1.1 Imagerie par Résonance Magnétique (IRM) cérébrale

1.1.1 Anatomie cérébrale

L'encéphale est constitué du cerveau, du cervelet et du tronc cérébral. Le cerveau est composé de trois matières principales : la matière blanche (MB), la matière grise (MG) et le liquide céphalo-rachidien (LCR). La matière blanche est constituée des fibres des cellules nerveuses appelées axones qui permettent la transmission de l'information traitée au niveau de la matière grise. La matière

grise contient le corps des cellules nerveuses et est répartie en deux types de structures : le cortex et les noyaux. Le cortex est caractérisé par de nombreuses fissures appelées sillons. Sur la face intérieure du cortex se trouve la matière blanche et sur la face extérieure circule la matière grise. Les noyaux, constitués essentiellement de matière grise, sont des structures plus compactes au centre du cerveau. Le liquide céphalo-rachidien baigne la surface extérieure du cerveau et du cervelet et remplit le système ventriculaire. Les figures 1.1 et 1.2 illustrent cette description de l'anatomie cérébrale.

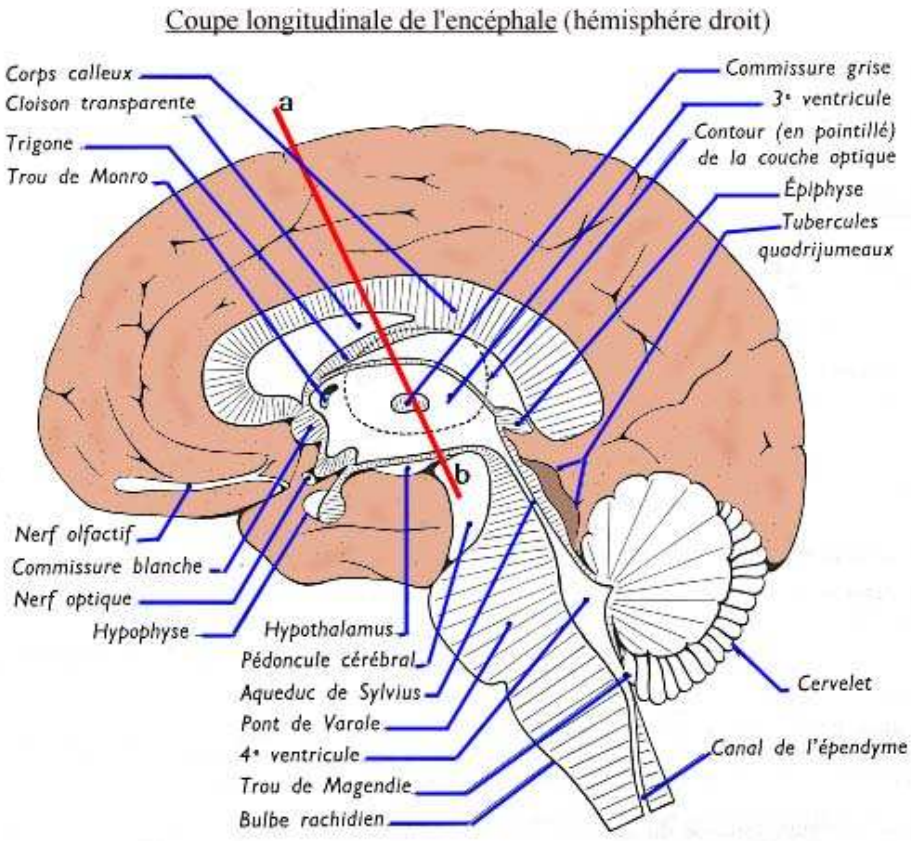


FIG. 1.1 – *Coupe longitudinale de l'hémisphère droit. Illustration tirée de* `www.medecine-et-sante.com`.

1.1.2 Principe de l'Imagerie par Résonance Magnétique (IRM)

L'IRM est une technique de diagnostic médical fournissant des images tridimensionnelles et en coupe de grande précision anatomique. Cette technique récente, non invasive, est basée sur le phénomène physique de la résonance magnétique nucléaire. La résonance magnétique nucléaire est une technique en développement depuis une soixantaine d'années dont le phénomène physique a été conceptualisé en 1946 par Bloch et Purcell. Les premiers développements en Imagerie par Résonance Magnétique datent des années 1973. Les premières images chez l'homme ont été réalisées en 1979. Aujourd'hui, l'IRM est devenue une technique majeure de l'imagerie médicale moderne. Le principe de l'IRM repose sur la propriété de certains atomes à entrer en résonance dans certaines conditions :

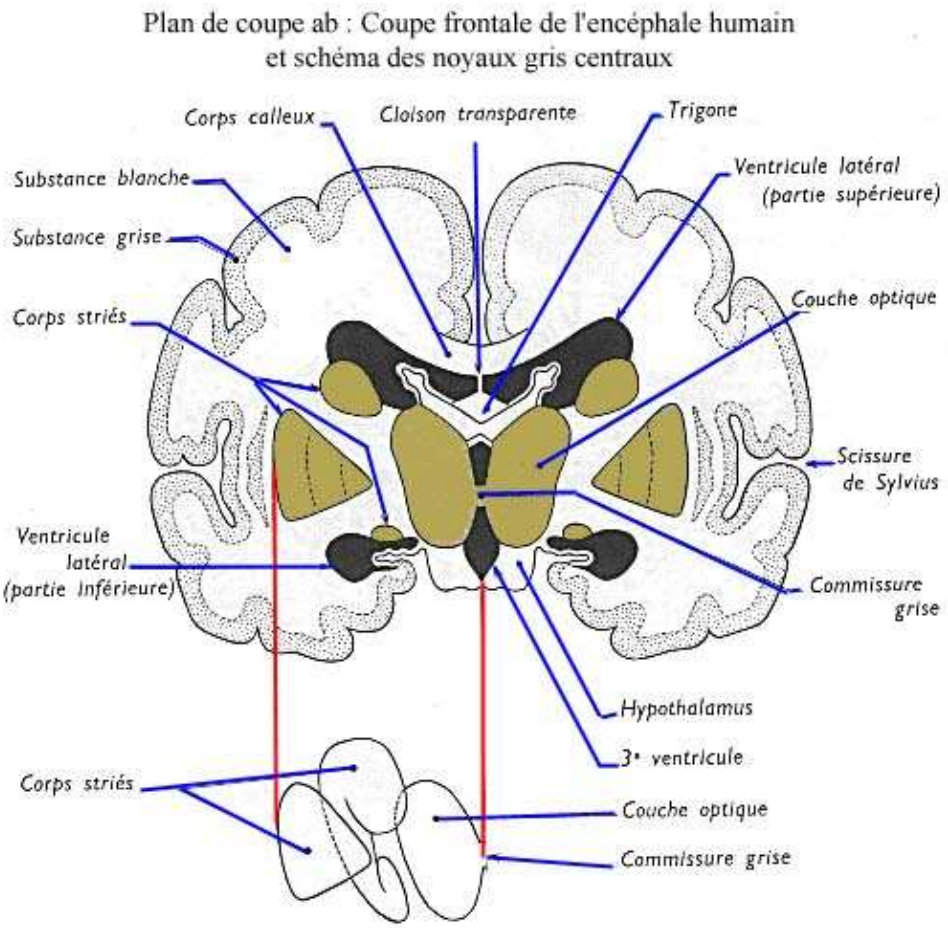


FIG. 1.2 – *Coupe frontale de l'encéphale et schéma des noyaux gris centraux. Illustration tirée de* `www.medecine-et-sante.com`.

c'est le cas de l'atome d'hydrogène (H) qui est un des deux constituants de la molécule d'eau (H_2O) que l'on trouve en grande quantité dans le corps humain (entre 60 et 75%). De plus, la quantité d'eau varie d'un tissu à l'autre, mais aussi à l'intérieur d'un même tissu selon son état physiologique, ce qui est utilisé pour établir une cartographie du corps humain et de ses pathologies.

Lorsque les atomes d'hydrogène entrent en résonance sous l'action d'un champ magnétique, ils absorbent de l'énergie : ainsi, plus la concentration en eau du milieu est élevée, plus il y a absorption d'énergie. A l'arrêt du phénomène de résonance, l'énergie emmagasinée par les atomes est restituée au milieu.

Les atomes possèdent un moment angulaire intrinsèque nommé spin auquel est lié un moment magnétique pouvant être assimilé à un aimant. Sans perturbation extérieure ces aimants élémentaires sont orientés de façon aléatoire dans toutes les directions. Si on applique un champ magnétique $\vec{B}_0$ constant et homogène, ils s'orientent selon des directions privilégiées. Si on applique en plus un champ magnétique tournant $\vec{b}_1$ perpendiculaire au premier, un phénomène de résonance des noyaux de l'échantillon étudié intervient. Lorsqu'on supprime ce second champ, les atomes vont retourner à l'état initial en émettant un signal RMN (Résonance Magnétique Nucléaire). C'est ce

principe qui est utilisé pour former le cube d'images IRM.

1.1.3 Formation des IRM

Le processus se fait en trois étapes : dans un premier temps, le corps est placé dans un champ magnétique qui oriente tous les protons dans la même direction. Puis les protons sont excités par des ondes radio qui modifient leur orientation. Dès l'arrêt de l'impulsion RF (Radio-Fréquence), les protons retournent à l'état d'équilibre. L'acquisition des signaux RMN émis permet de reconstruire l'image.

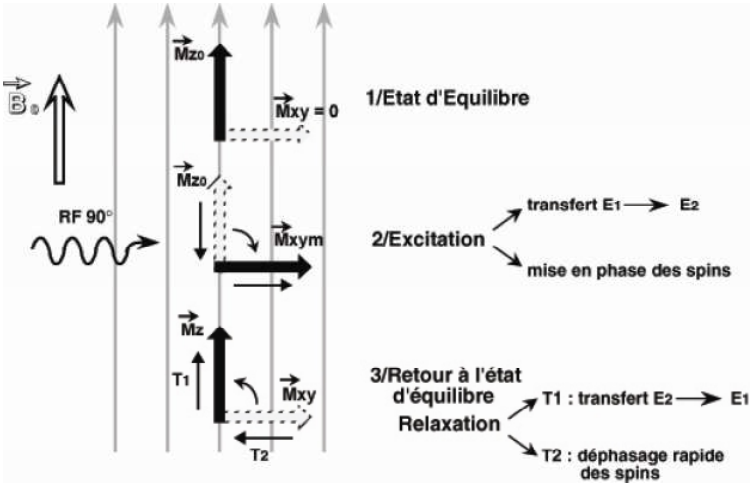


FIG. 1.3 – *L'état d'excitation est instable et dès l'arrêt de l'impulsion RF commence le retour à l'état d'équilibre (stable). La transition progressive $E_2 \rightarrow E_1$ entraîne la réaugmentation de $\vec{M}_z$: phénomène de relaxation T1, et le déphasage rapide des spins entraîne la diminution de $\vec{M}_{xy}$: phénomène de relaxation T2.*

Pour former le volume, il faut pouvoir localiser dans les trois plans de l'espace le signal reçu par l'antenne réceptrice. Pour cela trois gradients de champs magnétiques sont utilisés. Ces gradients sont créés par des bobines qui vont se superposer au champ magnétique $\vec{B}_0$. Le gradient de sélection de coupe (GS) est d'abord appliqué puis le gradient de codage de phase (GP) permet de sélectionner les lignes de la matrice à l'origine de l'image et le gradient de lecture (GL) permet de sélectionner les colonnes (Fig 1.4). Une transformée de Fourier permet ensuite de reconstruire l'image. Lors de l'acquisition du signal, le processus d'échantillonnage engendre un phénomène appelé volume partiel : certains voxels issus de l'échantillonnage du volume observé sont à la frontière entre deux tissus et peuvent donc contenir de l'information de plusieurs classes tissulaires. Ces voxels appartiennent donc à plusieurs classes en même temps, du fait de la résolution spatiale de l'imageur, ce qui pose un problème lors de la labellisation.

Les images IRM sont donc obtenues par application de gradients de champ et de séquences d'impulsions RF.

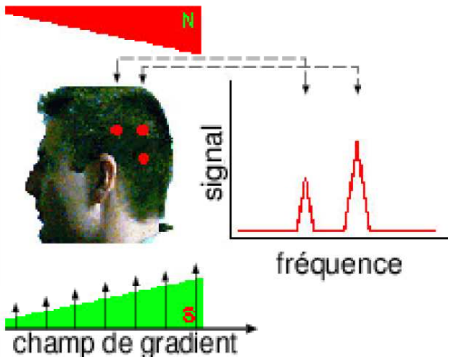


FIG. 1.4 – *Formation de l'image : un champ de gradient permet de sélectionner une position dans une direction spatiale [Bosc 03a].*

1.1.4 Contrastes en IRM

Après avoir vu les notions de base concernant l'imagerie par résonance magnétique, nous allons maintenant décrire les différents types de contrastes qui permettent d'obtenir des images contenant des informations de natures différentes nommées modalités. On peut ainsi pondérer l'image en T1 ou T2, ou en densité de protons. Les différents types de contraste sont obtenus en faisant varier les paramètres d'acquisition : le temps d'écho TE, c'est-à-dire le temps qui sépare l'impulsion RF et l'acquisition du signal, et le temps de répétition TR, c'est-à-dire le temps entre deux impulsions RF consécutives. Les images pondérées en T1 sont obtenues lorsque le TR et le TE sont courts tandis que les images sont pondérées en T2 lorsque le TR et le TE sont longs. Pour les images pondérées en densité de protons, le contraste est obtenu en utilisant un TR long et un TE court.

Chaque modalité contient des informations spécifiques ne se retrouvant pas dans les autres. A partir des images pondérées en T1, on peut par exemple distinguer les différents tissus cérébraux : matière blanche, matière grise et liquide céphalo-rachidien ; tandis que les images pondérées en T2 mettent plus facilement en évidence certaines anomalies comme les lésions dues à la sclérose en plaque (SEP).

1.1.5 Caractéristiques des IRM cérébrales

1.1.5.1 Le bruit

Un bruit aléatoire gaussien apparaît dans le domaine de Fourier de l'image acquise. L'image est ensuite obtenue en calculant le module de la transformée inverse de Fourier. La distribution gaussienne devient donc, dans l'image, une distribution de Rice [Sijbers 98]. Dans les zones où l'intensité n'est pas proche de zéro, cette distribution est approximativement gaussienne, tandis que dans les zones où l'intensité est proche de zéro, elle est proche d'une distribution de Raleigh. Dans la suite du manuscrit, nous ferons l'hypothèse d'une distribution gaussienne.

1.1.5.2 L'effet de volume partiel

Un autre problème de la segmentation d'images IRM est l'effet de volume partiel qui apparaît lorsque plusieurs types de tissus contribuent au même voxel. Ce problème est de plus en plus pris en

compte dans les algorithmes de segmentation. En raison d'une résolution du système d'acquisition limitée, les voxels situés à la frontière entre plusieurs tissus sont composés de deux ou plusieurs tissus (Fig. 1.5). Il est donc nécessaire de prendre en compte ces effets de volume partiel pour obtenir une segmentation fiable des tissus cérébraux. Les méthodes de classification en 3 classes "dures" (matière blanche, matière grise et liquide céphalo-rachidien) ignorent ce problème et perdent ainsi de l'information sur la structure des tissus.

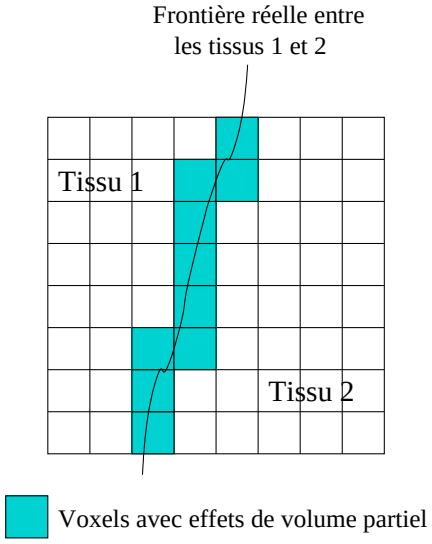


FIG. 1.5 – *Effet de volume partiel d'après [Jaggi 98].*

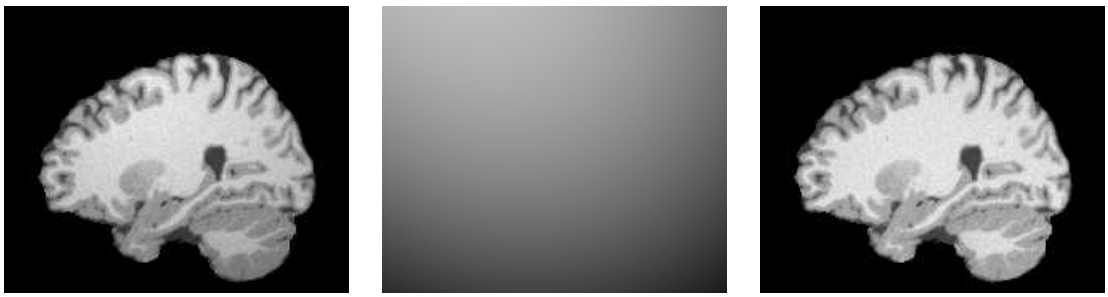
Certaines méthodes considèrent l'effet de volume partiel comme un facteur de dégradation et cherchent à le corriger, alors que d'autres le considèrent comme une propriété de l'image et cherchent à l'estimer pour obtenir une qualité sous-voxel [Ballester 02]. Les différentes méthodes de segmentation prenant en compte ces effets de volume partiel seront présentées au paragraphe 1.4.

1.1.5.3 Les hétérogénéités d'intensité

Une des principales difficultés de la segmentation d'images IRM est la présence d'un artefact d'hétérogénéité d'intensité spatiale pour un même tissu cérébral. Les inhomogénéités du champ RF sont en effet responsables de ces variations spatiales lentes de l'intensité des images (Fig. 1.6). Ce biais peut poser des problèmes de classification pour des techniques de segmentation basées sur l'intensité, si on suppose que l'intensité d'une classe est constante sur toute l'image. La non-uniformité est prise en compte dans la plupart des méthodes de segmentation, soit en la compensant par pré-traitement [Shattuck 01], soit en la modélisant au cours de la segmentation. Belaroussi *et al.* présentent une comparaison des différentes méthodes de correction de non-uniformité existantes [Belaroussi 06]. Nous présenterons différentes modélisations existantes au paragraphe 1.3.

1.1.5.4 Autres artefacts

Les artefacts de mouvement sont dus aux déplacements du patient pendant l'examen ainsi qu'aux mouvements physiologiques (respiration, flux sanguin). L'impact de ces artefacts est



(a) (b) (c)

FIG. 1.6 – *Illustration de l'artefact d'inhomogénéité. (a) correspond à l'image affectée par une hétérogénéité RF. (b) correspond à l'artefact isolé et (c) à l'image corrigée.*

variable selon le moment de l'acquisition mais il se traduit généralement par l'apparition d'images fantômes de la structure en mouvement.

Après s'être intéressés aux différentes caractéristiques et imperfections des images IRM, nous allons maintenant présenter les principales méthodes de segmentation des tissus cérébraux prenant en compte ces artefacts.

L'objectif principal du processus de segmentation est le partitionnement d'une image en régions d'intérêt (ou classes) homogènes du point de vue d'un certain nombre de caractéristiques ou critères. En imagerie médicale, la segmentation est très importante, que ce soit pour l'extraction de paramètres ou de mesures sur les images, ainsi que pour la représentation et la visualisation. Certaines applications nécessitent une classification des images en différentes régions anatomiques : graisse, os, muscles, vaisseaux sanguins, alors que d'autres recherchent une détection des zones pathologiques, comme les tumeurs, ou des lésions cérébrales. Dans ce chapitre, nous nous intéressons principalement à la segmentation des tissus cérébraux (matière blanche, matière grise et liquide céphalo-rachidien). Dans un premier temps, nous présentons les méthodes générales de segmentation existantes puis nous détaillerons les méthodes prenant en compte les hétérogénéités d'intensité ainsi que les modélisations des effets de volume partiel existantes.

1.2 Segmentation des tissus cérébraux

On distingue deux grands types de méthode de segmentation :
- celles basées *voxels* de type seuillage ou classification (*e.g.*, EM, Fuzzy C-Means) ;
- celles basées *modèles* : modèles déformables 2D ou 3D.

Les principales caractéristiques (information sur l'intensité, la forme, approche frontière ou région) des différentes méthodes sont présentées dans le tableau 1.1.

La classification peut être considérée comme un problème de données incomplètes : soit $Y = (Y_s)_{s\in S}$ le champ des observations et $X = (X_s)_{s\in S}$ le champ des étiquettes. L'objectif est de trouver une classification des données manquantes (ou étiquettes cachées) X connaissant $(Y_s)_{s\in S}$. Nous allons détailler différentes modélisations statistiques des données manquantes : dans le cas où aucune corrélation spatiale n'est prise en compte, la densité des données observées est souvent

	Seuillage/Classification	Modèles déformables	MRF (Markov Random Fields)
Forme	Aucune	Importante	Local
Intensité	Essentielle	Importante	Importante
Frontière/Région	Région	Frontière	Région

TAB. 1.1 – *Principales caractéristiques des différentes méthodes de segmentation.*

modélisée par un modèle de mélange fini indépendant. D'autres modélisations font l'hypothèse que les données sont corrélées spatialement.

1.2.1 Méthodes basées voxels

1.2.1.1 Algorithme FCM (Fuzzy C-Means)

L'algorithme des K-moyennes floues est une approche de classification générale non paramétrique développée par Bezdek [Bezdek 81]. Il permet d'estimer la valeur des centroïdes v_k caractérisant les K classes et les degrés flous d'appartenance des voxels aux classes u_{nk}, à partir de l'intensité observée y_n en chaque voxel n. Les degrés flous d'appartenance aux classes vérifient les contraintes suivantes :

$$\forall\, k \in \{1, \ldots, K\},\ 0 \leq u_{nk} \leq 1 \quad \text{et} \quad \sum_{k=1}^{K} u_{nk} = 1 \tag{1.1}$$

L'algorithme consiste à minimiser la fonction objectif :

$$J = \sum_{n} \sum_{k=1}^{K} (u_{nk})^{(q)} ||y_n - v_k||^2 \tag{1.2}$$

avec la contrainte $\sum_{k=1}^{K} u_{nk} = 1 \quad \forall n$ et où q caractérise le degré de flou.

L'algorithme des K-Moyennes floues est présenté dans l'algorithme 1.

Les FCM sont souvent utilisés en segmentation d'images médicales [Pham 99, Shen 05]. Cependant cet algorithme n'est pas robuste face au bruit et est sensible à l'initialisation. Plusieurs extensions ont donc été développées pour surmonter cet inconvénient. Dans [Pham 99], Pham et Prince présentent un algorithme FCM adaptatif (AFCM). La non-uniformité de l'image est modélisée en multipliant les centroïdes par un gain. Deux termes de régularisation sont ajoutés à la fonction à minimiser pour assurer que le gain soit lisse et varie lentement. Pour réduire le temps de calcul, ils utilisent une méthode *coarse to fine* : pendant les premières itérations de l'AFCM, seule une approximation du gain est requise. Une solution sous-échantillonnée est donc utilisée et affinée quand le nombre d'itérations augmente. Une estimation de la solution est affinée en approximant l'erreur de l'estimation sur une grille grossière, puis en mettant à jour l'estimée à partir de l'erreur.

Un autre extension a été proposée par Shen *et al.* : dans [Shen 05], ils prennent en compte un terme d'attraction de voisinage qui dépend de la position relative et des propriétés des pixels voisins. Le degré d'attraction est optimisé par un réseau de neurones.

Algorithme 1 Algorithme des K-Moyennes Floues

1. Initialisation des valeurs des centroïdes $v_k^{[p=0]}$
2. Faire
   - $p \leftarrow p+1$
   - Mise à jour des fonctions d'appartenance

$$u_{nk}^{[p]} = \left(\frac{\|y_n - v_k^{[p-1]}\|^2}{\sum_{l=1}^{K} \|y_n - v_l^{[p-1]}\|^2} \right)^{-1/(q-1)} \tag{1.3}$$

   - Mise à jour des centroïdes

$$v_k^{[p]} = \frac{\sum_n u_{nk}^{[p]} y_n}{\sum_n u_{nk}^{[p]}} \tag{1.4}$$

   Tant que $\|v_k^{[p]} - v_k^{[p-1]}\| < \epsilon$ où ϵ est un seuil empirique fixé au préalable.

1.2.1.2 Modèle de mélange de gaussiennes

Lorsque le modèle ne prend pas en compte la corrélation spatiale, un modèle de mélange de gaussiennes est souvent utilisé pour modéliser la densité des données observées [Ashburner 00, Dugas-Phocion 04, Aït-ali 05]. Considérons m modalités d'IRM. Chaque voxel n est décrit par un vecteur m-dimensionnel $\boldsymbol{y_n} = (y_n^1, \ldots, y_n^m)$. Ces vecteurs sont considérés comme étant les réalisations d'une variable aléatoire suivant un modèle de mélange de gaussiennes défini par :

$$f(\boldsymbol{y_n}; \theta) = \sum_{k=1}^{K} \alpha_k (2\pi)^{m/2} |\Sigma_k|^{-1/2} \exp(-\frac{1}{2}(\boldsymbol{y_n} - \boldsymbol{\mu_k})^t \boldsymbol{\Sigma_k}^{-1} (y_n - \boldsymbol{\mu_k})) \tag{1.5}$$

$$= \sum_{k=1}^{K} \alpha_k f_k(\boldsymbol{y_n}; \theta_k) \tag{1.6}$$

où K est le nombre de composantes du modèle et $\theta = (\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu_k}$ et $\boldsymbol{\Sigma_k}$ sont respectivement le vecteur moyenne et la matrice de covariance de la $k^{\text{ème}}$ distribution gaussienne et α représente les proportions du mélange. $f_k(\boldsymbol{y_n}; \theta_k)$ représente la densité d'une gaussienne multivariée de dimension m et de paramètre $\theta_k = (\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$. Les poids α_k sont tous positifs et vérifient la relation :

$$\sum_{k=1}^{K} \alpha_k = 1.$$

Le but est de donner une étiquette x_n à chaque voxel n avec $x_n \in \{c_1, \ldots, c_K\}$, c_k étant une des trois composantes du mélange. L'estimateur du maximum du vraisemblance est souvent utilisé pour estimer les paramètres inconnus des modèles de mélange de gaussiennes en utilisant l'algorithme EM (Expectation-Maximisation) [Dempster 77] comme schéma d'optimisation. L'algorithme EM sera présenté en détails au paragraphe 2.6 dans le cas de l'estimation des paramètres des chaînes de Markov. A l'itération (q), l'estimation des paramètres fait intervenir le calcul des probabilités suivantes :

$$\hat{z}_{nk}^{(q)} = P(x_n = c_k | y_n; \theta^{(q)}) = \frac{\alpha_k^{(q)} f_k(\boldsymbol{y_n} | \theta_k^{(q)})}{\sum_{l=1}^{K} \alpha_l^{(q)} f_l(\boldsymbol{y_n} | \theta_l^{(q)})} \tag{1.7}$$

La mise à jour des poids est donnée par :

$$\hat{\alpha}_k^{(q)} = \frac{\sum_{n=1}^{N} \hat{z}_{nk}}{N} \tag{1.8}$$

où N représente le nombre total de voxels. L'estimation des paramètres des classes devient :

– estimation du vecteur moyenne de la $k^{\text{ème}}$ classe

$$\boldsymbol{\mu_k}^{(p)} = \frac{\sum_{n=1}^{N} \hat{z}_{nk} \boldsymbol{y_n}}{\sum_{n=1}^{N} \hat{z}_{nk}} \tag{1.9}$$

– estimation de la matrice de covariance de la $k^{\text{ème}}$ classe

$$\boldsymbol{\Sigma_k}^{(p)} = \frac{\sum_{n=1}^{N} \hat{z}_{nk} (\boldsymbol{y_n} - \boldsymbol{\mu_k}^{(p)})(\boldsymbol{y_n} - \boldsymbol{\mu_k}^{(p)})^t}{\sum_{n=1}^{N} \hat{z}_{nk}} \tag{1.10}$$

Certaines méthodes proposent des améliorations de cet algorithme pour prendre en compte les effets de volume partiel [Dugas-Phocion 04] ou les inhomogénéités d'intensité [Ashburner 00]. Ces méthodes seront détaillées dans les paragraphes suivants.

Des adaptations robustes de cet algorithme EM sont présentées dans [Schroeter 98, Aït-ali 05] afin de prendre en compte la présence éventuelle de données atypiques. Aït-ali *et al.* utilisent un estimateur robuste, l'estimateur de vraisemblance tamisée TLE (Trimmed Likelihood Estimator) et remplacent ainsi l'étape de maximisation de l'EM par une étape d'estimation robuste appropriée permettant d'obtenir une estimation robuste des paramètres du mélange. Schroeter *et al* proposent un algorithme EM modifié incluant un schéma de rejet des outliers.

Le principal inconvénient de l'EM est qu'il est très sensible au bruit, d'où la nécessité de prendre en compte une homogénéité spatiale *a priori*, par exemple en utilisant les champs de Markov (Markov Random Field, MRF).

1.2.1.3 Modélisation avec corrélation spatiale

Les champs de Markov permettent de modéliser les interactions entre voxels voisins. Ces modèles sont souvent incorporés dans les algorithmes de segmentation sous forme d'un modèle *a priori* régularisant [Rajapakse 97, Held 97].

Etant donné un ensemble S de sites s et des variables aléatoires X_s qui leur sont associées, le champ $X = (X_s)_{s \in S}$ est un champ de Markov pour un système de voisinage V_s donné si et seulement si :

$$P(X_s | X_t, t \neq s) = P(X_s | X_t, t \in V_s) \tag{1.11}$$

D'après le théorème de Hammersley-Clifford, le champ est alors un champ de Gibbs de distribution [Besag 74] :

$$P(X = x) = \frac{1}{Z} \exp(-U(x)) \tag{1.12}$$

avec $U(x) = \sum_{c \in C} V_c(x_s, s \in c)$, Z est une constante de normalisation appelée fonction de partition. C est l'ensemble des cliques défini par le système de voisinage considéré et V_c est un potentiel dépendant des configurations de la clique c [Geman 84]. X représente le champ des étiquettes, c'est-à-dire des labels recherchés.

Dans [Rajapakse 97], Rajapakse *et al.* présentent un algorithme basé sur un modèle statistique qui consiste en un modèle de mesure (mélange de gaussiennes) et un modèle *a priori* markovien sur le champ des étiquettes X. Le bruit est supposé blanc, gaussien, additif, dépendant du tissu et spatialement variant. Le modèle *a priori* est représenté par les modèles MLL (*Multilevel Logistic*). Si on considère les interactions entre deux sites, la fonction potentiel pour la clique c_l est définie par : $V_{c_l}(x_i) = \sum_{i,j \in c_l}[-\beta_l \delta(x_i = x_j) + \beta_l \delta(x_i \neq x_j)]$, où β est une constante réelle positive. Pour les cliques d'un seul site, le potentiel est : $V_c(x_i = k) = \alpha_k$ pour $k \in \Lambda$ où $i \in c$ et α_k est une constante réelle et Λ est l'espace des labels. Le modèle proposé est basé sur un algorithme ICM (Iterative Conditional Modes). La modélisation MRF permet d'imposer des contraintes de continuité et d'homogénéité aux différentes régions. Le modèle proposé peut être étendu à des images multimodales en utilisant des densités multidimensionnelles pour le modèle de mesure.

Dans [Van Leemput 99b], Van Leemput *et al.* proposent une méthode de segmentation d'images mono- ou multimodales en corrigeant les inhomogénéités (*cf.* section 1.3 pour la correction du biais [Van Leemput 99a]) et en incorporant de l'information contextuelle en utilisant un *a priori* markovien sur la carte de segmentation. L'algorithme prend en compte la classification, l'estimation des paramètres de la distribution normale, l'estimation des paramètres du biais et de ceux du modèle MRF. L'utilisation d'un atlas du cerveau contenant les cartes de probabilités *a priori* de chaque classe de tissus pour l'initialisation de l'algorithme permet d'avoir une méthode entièrement automatique sans intervention de l'utilisateur.

Zhang *et al.* présentent une méthode supervisée basée sur l'algorithme EM [Zhang 01]. Un modèle de Potts dont les paramètres sont fixés par l'utilisateur modélise les interactions spatiales entre étiquettes. Cette méthode alterne une étape de segmentation de l'image à l'aide de l'algorithme ICM et une étape d'estimation des paramètres du modèle d'intensité à l'aide de la classification courante de l'image.

Marroquin *et al.* utilisent un atlas du cerveau qu'ils recalent sur le cerveau à segmenter [Marroquin 02]. Cela permet de supprimer les tissus n'appartenant pas au cerveau, de calculer les probabilités *a priori* pour chaque classe en chaque voxel et d'initialiser la segmentation. Ils proposent une variante de l'algorithme EM prenant en compte les hypothèses de cohérence spatiale sous la forme d'un modèle MRF (modèle de Potts) en utilisant un système de voisinage à six voisins.

Scherrer *et al.* proposent une approche markovienne coopérative fondée sur le raffinement mutuel des segmentations en tissus et en structures [Scherrer 07]. La connaissance *a priori* nécessaire à la segmentation des structures est apportée par une description floue de l'anatomie cérébrale. Elle est intégrée dans un cadre markovien via l'expression d'un champ externe. La segmentation des tissus intègre dynamiquement l'information structure via un autre champ externe, permettant de raffiner l'estimation des modèles d'intensité. Cette approche markovienne dynamique et coopérative est implémentée dans un environnement multiagents : des entités autonomes distribuées dans l'image estiment des modèles markoviens locaux et coopèrent pour assurer leur cohérence.

D'autres méthodes utilisent une modélisation par champs de Markov et prennent en compte les inhomogénéités d'intensité et/ou les effets de volume partiel [Held 97, Ruan 00, Ruan 02]. Elles seront donc détaillées aux paragraphes 1.3 et 1.4.

1.2.2 Modèles déformables

Dans [McInerney 96], McInerney et Terzopoulos présentent une synthèse des méthodes utilisant les modèles déformables. L'avantage de ces modèles est leur capacité à segmenter des structures anatomiques en exploitant à la fois les données de l'image et des connaissances *a priori* sur la position, la forme et la taille des structures. Ces modèles sont basés sur une approche frontière et consistent en la minimisation de la somme de deux énergies : d'une part l'énergie interne mesurant l'adéquation avec la forme et d'autre part l'énergie externe mesurant l'adéquation avec l'apparence. La segmentation par modèles déformables se décompose en deux étapes :

- l'initialisation : il faut tout d'abord détecter si la structure cible se trouve dans l'image puis localiser cette structure cible
- l'optimisation des paramètres du modèle pour délinéer la structure cible dans l'image.

Il existe différents types de modèles déformables continus ou discrets, les plus connus étant les levels-sets. La figure 1.7 présente diverses représentations existantes.

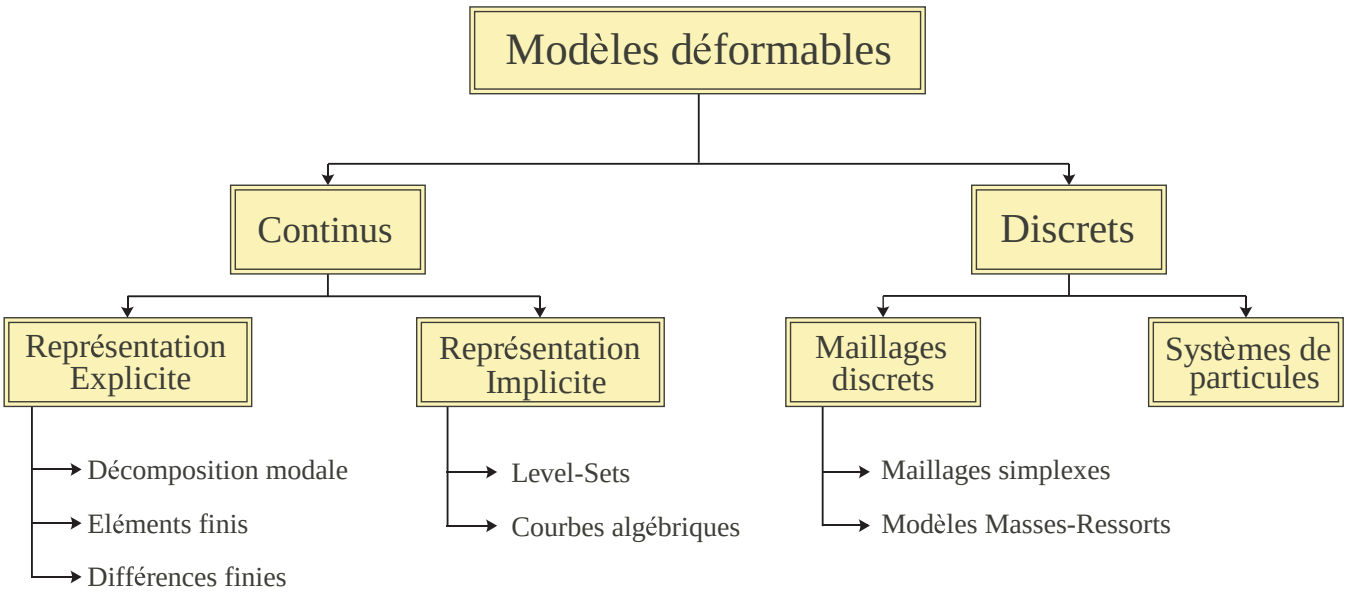


FIG. 1.7 – *Différentes représentations de modèles déformables.*

Dans [Baillard 01], Baillard *et al.* utilisent les level sets pour segmenter le cerveau. Ils proposent d'employer une méthode de recalage pour initialiser la surface et ainsi accélérer la segmentation. Le recalage est réalisé par une méthode de recalage multirésolution en effectuant une minimisation multi-grille basée sur les partitions successives du volume initial, à chaque niveau de résolution.

Ciofolo *et al.* utilisent les level-set en prenant en compte de la logique floue [Ciofolo 04]. La force d'évolution intègre à la fois un terme région et le gradient de l'intensité de l'image à l'intérieur de la même expression. La fusion entre les deux termes est faite par un algorithme de décision floue.

Dans [Yang 04] , Yang *et al.* présentent une méthode de segmentation dans le cadre d'une estimation MAP prenant en compte à la fois l'information *a priori* du voisinage et l'information sur les niveaux de gris de l'image pour segmenter plusieurs objets simultanément.

Dans [Bazin 05], Bazin et Pham présentent une méthode pour segmenter plusieurs objets dans les images en respectant les contraintes topologiques données par un *template*. Ils utilisent

l'algorithme FCM (Fuzzy C-Means), et ajoutent un terme de régularisation pour être moins sensible au bruit. Puis ils emploient une technique de fast marching, où seuls les points "critiques" sont gardés, les points "réguliers" étant supprimés (*thinning process*) de façon à ne garder que le squelette de l'objet original avec la même topologie. Puis ce processus est inversé (*joint marching process*) : à partir du squelette, de nouveaux points sont ajoutés à tous les objets, de façon à retrouver l'image entière.

Après avoir décrit les différentes méthodes de segmentation existantes, nous allons nous intéresser à la modélisation des hétérogénéités d'intensité et des effets de volume partiel.

1.3 Prise en compte des hétérogénéités d'intensité

Le modèle le plus utilisé pour décrire les données est un modèle multiplicatif avec du bruit additif : $v(\vec{r}) = g(\vec{r}).u(\vec{r}) + n(\vec{r})$, avec $v(\vec{r})$ l'intensité du voxel en $\vec{r} = (x, y, z)$, $g(\vec{r})$ un bruit multiplicatif dont les valeurs varient lentement et de façon lisse en fonction de la localisation des voxels dans l'image, $u(\vec{r})$ la "vraie" distribution spatiale de l'intensité et $n(\vec{r})$ le bruit additif. Les variations spatiales de g sont supposées lentes, u est supposée constante par morceaux et n suit une loi de probabilité de Rice qui peut être approximée par une loi gaussienne [Sijbers 98].

Les deux principaux types de méthodes utilisées pour corriger ces hétérogénéités d'intensité sont les méthodes "prospectives" basées sur la connaissance *a priori* des paramètres d'acquisition et les méthodes "rétrospectives" de traitement d'images. On s'intéressera ici aux méthodes appartenant à la deuxième catégorie (*cf.* Fig. 1.8).

Une partie de ces méthodes est basée sur la distribution spatiale en niveaux de gris et l'autre sur des domaines transformés (Fourier, ondelettes).

1.3.1 Méthodes basées sur la distribution spatiale des niveaux de gris

Les méthodes basées sur la distribution en niveaux de gris supposent que la variation de l'artefact est spatialement lisse et varie lentement et que la "véritable" image est constante par morceaux. Les méthodes les plus employées sont les méthodes par ajustement de surface, filtrage spatial et les méthodes statistiques.

1.3.1.1 Ajustement de surfaces

La variation d'inhomogénéité d'intensité est supposée lisse et est approximée par une surface lisse. La correction sera effectuée en divisant voxel par voxel l'image originale par la surface calculée. Différentes fonctions de bases lisses sont utilisées, les deux principales étant les splines et les polynômes [Styner 00].

1.3.1.2 Filtrage spatial

Il peut s'agir soit de filtres passe-bas soit de filtres homomorphiques. Dans [Johnston 96], Johnston *et al.* utilisent un filtrage homomorphique pour corriger la non-uniformité comme prétraitement avant de segmenter. Ces approches supposent que l'information anatomique dans l'image est présente à des fréquences spatiales beaucoup plus élevées que les hétérogénéités

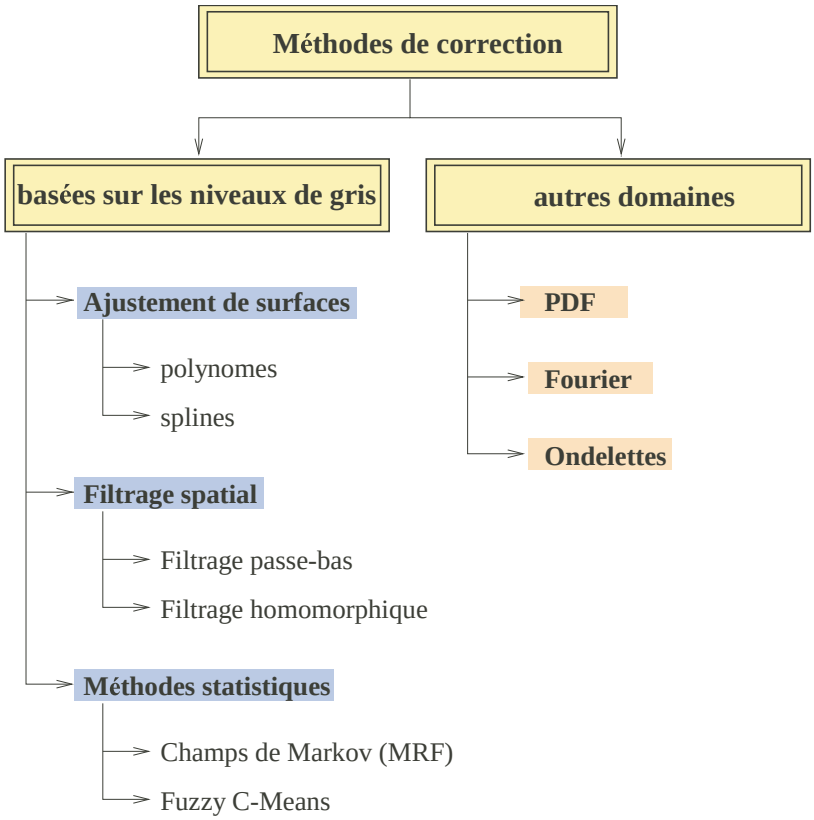


FIG. 1.8 – *Différentes méthodes de correction, d'après [Belaroussi 06].*

d'intensité. L'algorithme HUM (*Homomorphic Unsharp Masking*) permet de corriger les hétérogénéités d'intensité en supprimant les composantes basses fréquences de l'image sans altérer les contours [Brinkmann 98].

1.3.1.3 Méthodes statistiques

Le but initial de ces méthodes est de segmenter les données mais elles prennent aussi en compte l'inhomogénéité dans le modèle. L'estimation est basée sur un des critères suivants : le maximum de vraisemblance ML (Maximum Likelihood) [Van Leemput 99a, Zhang 01] ou le maximum *a posteriori* (MAP) [Wells 96, Guillemaud 97]. La densité de l'intensité peut être modélisée par un modèle de mélange fini FM (*Finite Mixture*) [Van Leemput 99a], ou par un modèle de mélange fini de gaussiennes [Wells 96, Ashburner 00]. La principale limitation de ces différents modèles est qu'ils ne prennent pas en compte l'information spatiale ce qui a motivé l'utilisation des modèles MRF (*Markov Random Field*) [Zhang 01]. Une fois le critère statistique (MAP, MPM) et la modélisation de l'image (mélange de gaussiennes, MRF, ...) choisis, les paramètres du modèle doivent être estimés. L'estimation des paramètres est réalisée dans la plupart des cas par l'algorithme EM [Van Leemput 99a] ou par un algorithme génétique [Schroeter 98] ou encore par l'ICM (*Iterative Conditional Modes*).

Dans [Van Leemput 99a], Van Leemput *et al.* proposent un modèle en trois étapes qui prend en compte l'estimation des paramètres du modèle d'intensité, la classification des tissus et l'estimation

du biais. Leur approche est basée sur la méthode du maximum de vraisemblance en utilisant l'algorithme EM (ou plutôt un algorithme GEM, *Generalized EM*, la vraisemblance n'étant pas maximisée mais seulement augmentée à chaque itération). Ils modélisent l'inhomogénéité comme une combinaison linéaire $\sum_k c_k \phi_k$ de fonctions de base polynômiales ϕ_k. La non-uniformité est modélisée comme un biais multiplicatif, une transformation logarithmique est donc effectuée pour rendre le biais additif. L'utilisation d'un atlas contenant les cartes de probabilité *a priori* pour chaque classe de tissus permet d'initialiser l'algorithme.

Marroquin *et al.* proposent une modélisation variable spatialement de l'intensité des tissus [Marroquin 02]. Ils considèrent le bruit comme indépendant des tissus et de la localisation dans l'image et modélisent les valeurs moyennes d'intensité par une fonction spline. L'algorithme alterne classification des voxels et estimation des paramètres des modèles d'intensité.

Les Fuzzy C-Means ont également été utilisés même si au départ ils n'étaient pas adaptés aux données avec inhomogénéité. Des extensions des Fuzzy C-Means ont donc été proposées pour prendre en compte la non-uniformité soit en adaptant le clustering aux variations locales, soit en modifiant la fonction à minimiser en ajoutant un terme de contrainte pour compenser l'inhomogénéité. Dans [Pham 99], Pham et Prince présentent un FCM adaptatif (AFCM). La non-uniformité de l'image est modelisée en multipliant les centroïdes par un gain. Deux termes de régularisation sont ajoutés à la fonction à minimiser pour assurer que le gain soit lisse et varie lentement.

1.3.2 Méthodes basées sur un domaine transformé

Les autres méthodes sont basées sur un domaine autre que le domaine spatial : domaine PDF (*Probability Density Functions*), de Fourier ou des ondelettes.

1.3.2.1 Domaine de la PDF

Dans ce domaine, l'inhomogénéité est considérée comme un terme de convolution parasite lissant la distribution réelle de l'intensité et augmentant l'entropie. Dans [Sled 98], l'histogramme du biais est supposé gaussien et l'approche de Sled *et al.* est basée sur la déconvolution du signal mesuré. Leur approche consiste à trouver le champ multiplicatif, lisse, variant lentement qui maximise le contenu fréquentiel de U.

1.3.2.2 Domaine de Fourier

Ces méthodes sont une alternative de celles présentées au paragraphe 1.3.1.2, les filtres n'étant pas appliqués dans le même domaine. Seuls quelques filtres passes-bas gaussiens ont été utilisés pour corriger la non-uniformité [Cohen 00].

1.3.2.3 Domaine des ondelettes

La transformée discrète en ondelettes permet de décomposer l'image en une cascade de sous-espaces d'approximation (contenant l'information basse-fréquence) et de détail (contenant l'information haute-fréquence) orthogonaux. L'inhomogénéité étant contenue dans les basses fréquences, elle peut être estimée et corrigée dans le sous-espace des approximations [Lin 03].

1.4 Modélisation des effets de volume partiel

Certaines méthodes considèrent l'effet de volume partiel comme un facteur de corruption et cherchent à le corriger, tandis que d'autres méthodes le considèrent comme une propriété de l'image et cherchent à l'estimer pour obtenir une qualité sous-voxel [Ballester 02]. Au lieu d'attribuer à chaque voxel une seule et unique classe, certains auteurs estiment ainsi le pourcentage de chaque tissu en chaque voxel.

1.4.1 Modélisation sans corrélation spatiale

Dans la littérature, de nombreux articles incorporent ces effets de volume partiel dans la segmentation : des algorithmes EM modifiés ont été proposés pour inclure ces effets de volume partiel. Dans [Dugas-Phocion 04], Dugas-Phocion *et al.* présentent un algorithme EM amélioré prenant en compte les effets de volume partiel. Si on considère deux tissus a et b, d'intensité respective y_a et y_b, et α la proportion du tissu a dans le voxel PV considéré, y_a et y_b suivent une loi normale $\mathcal{N}(\mu_a, \Sigma_a)$ et $\mathcal{N}(\mu_b, \Sigma_b)$ respectivement. L'intensité du voxel PV y_{PV} est une variable aléatoire $y_{PV} = \alpha y_a + (1-\alpha) y_b$. Si α est fixé et connu, y_{PV} suit une loi gaussienne (variables aléatoires indépendantes) :

$$y_{PV} \sim \mathcal{N}(\alpha\mu_a + (1-\alpha)\mu_b, \alpha^2\sigma_a + (1-\alpha)^2\sigma_b) \tag{1.13}$$

Pour α fixé, l'algorithme se décompose en trois étapes itératives :

- labellisation des classes (dont les classes PV) ;
- estimation des paramètres des classes pures ;
- calcul des paramètres des classes PV en utilisant l'équation 1.13.

Cette méthode permet d'améliorer les résultats de l'EM mais elle ne règle pas le problème de la variance de la classe LCR qui reste très haute. Pour résoudre ce problème, Dugas-Phocion *et al.* introduisent une quatrième classe : classe "vaisseaux". Dans le cas où seulement trois classes sont prises en compte, les vaisseaux sont classés comme du LCR. Cette classe LCR a donc été divisée en deux : vaisseaux et LCR. La probabilité *a priori* des vaisseaux est définie comme étant une fraction β (estimée à chaque itération) de la probabilité du LCR donnée par l'atlas. Néanmoins cette méthode ne prend pas en compte l'information spatiale.

Dans [Vincken 97], Vincken *et al.* utilisent une approche multi-échelle (*hyperstack*) avec des liens probabilistes pour segmenter une image. Chaque voxel enfant peut être connecté à plus d'un voxel parent (contrairement aux "hyperstacks" conventionnels où un enfant est lié à un seul parent). $P(V_{i_{l+1}}/V_{i_l})$ représente la probabilité que le voxel i au niveau $l+1$ soit le père du voxel enfant i au niveau l. Les probabilités des "racines" obtenues représentent les fractions des types de tissus des différents voxels (effets de volume partiel).

Bosc *et al.* proposent de résoudre les problèmes liés aux effets de volume partiel en modélisant le processus d'acquisition et en utilisant une image labellisée haute résolution, obtenant ainsi des segmentations haute résolution à partir d'images IRM standards [Bosc 03b].

1.4.2 Modélisation avec corrélation spatiale

Certaines approches modélisent les mixels avec une corrélation spatiale entre les proportions des différents types de tissus. Un *a priori* markovien est introduit : Choi *et al.* utilisent un modèle MRF

pour prendre en compte la corrélation spatiale entre les voxels [Choi 91]. Cependant cette méthode requière des images multimodales qui ne sont pas toujours disponibles. Shattuck *et al.*, Tohka *et al.* commencent par classifier le cerveau en différents labels PV (*Partial Volume*) en utilisant un modèle de Potts [Shattuck 01, Tohka 04]. Ils attribuent à chaque voxel une fraction de tissu pur pour la matière blanche, la matière grise et le liquide céphalo-rachidien, en se basant à la fois sur son label et sur son intensité dans l'image. Cependant cette méthode nécessite qu'un paramètre soit réglé par l'utilisateur : le terme β du modèle de Potts contrôlant la force de l'*a priori*.

D'autres méthodes identifient seulement les voxels contenant des effets de volume partiel mais ne cherchent pas à estimer la proportion de chaque tissu en chaque voxel. Par exemple, Ruan *et al.* prennent en compte les effets de volume partiel en proposant un algorithme en deux étapes [Ruan 00]. Dans un premier temps, ils classent le cerveau en cinq classes : trois types de tissus correspondant aux classes pures (MB, MG, LCR) et deux mélanges de tissus (LCR+MG, MG+MB). La classification se fait en utilisant le modèle MRF (Markov Random Field). Dans un deuxième temps, les classes de mélange sont reclassées en classes pures en ajoutant de la connaissance sur la topologie du cerveau : pour cela, la dimension multifractale qui apporte une information locale sur les variations en niveaux de gris est incorporée au modèle MRF en ajoutant un troisième terme dans la fonction d'énergie (exposant de Hölder α^H). A chaque voxel est cependant attribué un type de tissu pur unique, ce qui fait perdre la notion de volume partiel. Dans un autre article, Ruan *et al.* utilisent un modèle MRF flou pour prendre en compte les effets de volume partiel [Ruan 02]. Cette méthode est malheureusement coûteuse en temps de calcul en raison de la génération du champ de Markov effectuée en utilisant l'échantillonneur de Gibbs. Pour résoudre ce problème, l'estimation est réalisée sur environ 1/6 du volume total, occultant ainsi une partie de l'information.

Van Leemput *et al.* proposent une extension de l'EM prenant en compte les effets du volume partiel en ajoutant une étape de sous-échantillonnage [Van Leemput 03]. Si L est l'image labellisée comprenant J voxels, Y est l'image originale bruitée mais sans effet de volume partiel. $\tilde{Y}$ est l'image sous-échantillonnée comprenant $I = J/M$ voxels, M étant le nombre de sous-voxels par voxel. Le but est donc, connaissant $\tilde{Y}$ de reconstruire l'image labellisée sous-voxel $\mathbf{L}$ ou plutôt d'estimer la fraction des tissus. Pour cela, Van Leemput *et al.* utilisent l'algorithme EM et ils présentent trois différents modèles *a priori* $f(\mathbf{L}|\Phi_L)$ pour l'image labellisée $\mathbf{L}$. Le premier modèle suppose qu'il n'y a pas de corrélation spatiale entre les voxels, l'estimation des paramètres est donc entièrement basée sur l'histogramme. Le deuxième modèle suppose en plus que toutes les combinaisons de mélange entre deux types de tissus sont égales. Le troisième modèle est un modèle MRF. L'avantage de ce modèle est d'imposer des régions homogènes de tissus purs bordés par des voxels PV. Cependant comme ce modèle autorise les voxels PV seulement à l'interface entre deux tissus, il ne permet pas de décrire certaines structures où deux tissus se mélangent vraiment sans être sur une frontière.

Les différentes modélisations des effets de volume partiel sont résumées dans le tableau 1.2.

1.5 Positionnement de l'approche proposée par rapport aux méthodes existantes

L'approche que nous proposons pour la segmentation des tissus cérébraux (chapitre 3) est une approche région prenant en compte les différents artefacts qui ont été présentés : elle prend en considération les inhomogénéités d'intensité présentes sur les images IRM ainsi que les effets

	Sans corrélation spatiale	**Avec corrélation spatiale**
[Choi 91]		☺ modèle MRF ☹ images multimodales
[Vincken 97]	approche multi-échelle	
[Ruan 00]		☺ modèle MRF ☹ classification finale en 3 classes
[Shattuck 01] [Tohka 04]		☺ modèle de Potts ☹ supervisé
[Ruan 02]		☺ modèle MRF flou ☹ estimation sur 1/6 du volume (coûteux en temps de calcul)
[Van Leemput 03]	☺ possibilité d'une modélisation MRF ☹ modélisation PV seulement à la frontière entre 2 tissus	
[Dugas-Phocion 04]	☺ EM modifié classe LCR divisée en 2	

TAB. 1.2 – *Différentes modélisations des effets de volume partiel.*

de volume partiel en proposant une segmentation en cartes de tissus contenant les proportions de chaque tissu en chaque voxel. D'autre part, elle est basée sur la modélisation par Chaînes de Markov Cachées qui sera présentée au chapitre suivant et qui permet de prendre en compte la notion de voisinage. Enfin elle inclut l'information *a priori* apportée par un atlas probabiliste.

Chapitre 2

Segmentation non supervisée par chaînes de Markov cachées

Dans ce chapitre, nous commençons par rappeler brièvement les principes de la théorie bayésienne en présentant quelques estimateurs bayésiens. Puis nous exposons les principes généraux des modèles graphiques. Nous présentons ensuite les principaux modèles markoviens et leur intérêt en analyse d'images, ainsi que le cadre général des chaînes de Markov triplets et couples en détaillant plus particulièrement le modèle des chaînes de Markov cachées que nous utiliserons au chapitre suivant pour segmenter les images IRM. Enfin nous nous intéressons à l'estimation des paramètres de ce modèle.

2.1 Théorie bayésienne

Dans le cadre de la segmentation statistique, une image est représentée par un couple de variables aléatoires $(X, Y) = ((X_s)_{s\in S}, (\boldsymbol{Y_s})_{s\in S})$ où S est l'ensemble des pixels (ou voxels dans le cas de données tridimensionnelles), X le champ des étiquettes et Y celui des observations. Chaque X_s prend ses valeurs dans un ensemble $\Omega = \{\omega_1, \ldots, \omega_K\}$ appelé ensemble des classes et chaque $\boldsymbol{Y_s}$ prend ses valeurs dans $\mathbb{R}^m$, où m est par exemple le nombre de modalités dans le cas de la segmentation d'IRM (ou le nombre de canaux d'observation de manière plus générale). Les observations données par le champ Y sont considérées comme une version dégradée (filtrée, bruitée, ...) de X. Le problème de la segmentation d'images consiste alors à estimer la réalisation de X ayant observé $(\boldsymbol{Y_s})_{s\in S}$. On suppose dans ce cas l'unicité de X associée à des observations à caractère multimodal. La recherche des informations cachées X à partir des observations Y est généralement basée sur l'établissement d'un modèle direct qui intègre toute la connaissance disponible sur la formation des Y à partir des X. Retrouver les informations cachées à partir des observations consiste alors à inverser ce modèle direct. Les problèmes inverses en traitement d'images étant généralement des problèmes mal-posés, cette tâche s'avère complexe. Pour tenter de contourner cette difficulté, on peut favoriser certains types de solution en imposant une connaissance *a priori* sur les étiquettes et leur agencement spatial.

Dans la suite du manuscrit la notation p désignera une loi de probabilité dans le cas où l'ensemble des configurations est discret. Dans le cas où cet ensemble est continu, la même notation p doit être comprise comme une densité de probabilité.

La théorie bayésienne est basée sur la spécification de la distribution *a posteriori* $p(X = x|Y = y)$. La règle de Bayes permet de déterminer la probabilité *a posteriori* des étiquettes connaissant le champ des observations.

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)} = \frac{p(Y = y|X = x)p(X = x)}{p(Y = y)} \tag{2.1}$$

où $p(X = x, Y = y)$ est la probabilité jointe[1], $p(Y = y|X = x)$ est la vraisemblance des observations conditionnellement aux informations cachées, $p(X = x)$ est l'*a priori* sur les étiquettes et $p(Y = y)$ une constante de normalisation.

Nous allons maintenant présenter différents estimateurs permettant d'associer à chaque estimation $\hat{x}$ de x une fonction de coût $\mathcal{L}(x, \hat{x})$ mesurant la "distance" entre le champ des étiquettes et son estimation. L'estimateur optimal au sens de Bayes est celui qui minimise le coût moyen, *i.e.*, le risque de Bayes :

$$\hat{x}_{opt}(y) = \arg\min_{\hat{x}} \mathbb{E}[\mathcal{L}(X, \hat{x})|Y = y)] \tag{2.2}$$

Les principaux estimateurs bayésiens sont présentés dans les trois paragraphes suivants.

L'estimateur du MAP

L'estimateur du *Maximum a Posteriori* (MAP) pénalise de la même façon toutes les estimations différentes de la vraie solution x. La fonction de coût est la suivante [Somersalo 07] :

$$\mathcal{L}(x, \hat{x}) = 1 - \delta(x, \hat{x}) \quad \text{avec } \delta(a, b) = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{sinon} \end{cases} \tag{2.3}$$

le symbole de Kronecker. L'estimateur basé sur cette fonction de coût est :

$$\hat{x}_{MAP}(y) = \arg\max_{x} p_{X|Y}(x|y) \tag{2.4}$$

L'estimateur du MPM

L'estimateur du Mode des Marginales a Posteriori (MPM) pénalise une solution proportionnellement au nombre de sites erronés. La fonction de coût qui lui est associée est la suivante [Somersalo 07] :

$$\mathcal{L}(x, \hat{x}) = \sum_{s \in S} 1 - \delta(x_s, \hat{x}_s) \tag{2.5}$$

L'estimateur du MPM minimise en chaque site la distribution marginale locale *a posteriori* :

$$\forall s \in S, \quad \hat{x}_{sMPM}(y) = \arg\max_{x_s} p_{X_s|Y}(X_s = x_s|Y = y) \tag{2.6}$$

[1]Pour des raisons de simplicité, nous noterons par la suite $p(X = x) = p_X(x)$ par $p(x)$.

L'estimateur du champ moyen

La fonction de coût de l'estimateur du champ moyen (CM) est quadratique [Somersalo 07] :

$$\mathcal{L}(x, \hat{x}) = \sum_{s \in S} (x_s - \hat{x}_s)^2 \tag{2.7}$$

L'estimateur basé sur cette fonction de coût est :

$$\forall s \in S, \quad \hat{x}_{sCM}(y) = \mathbb{E}[X_s | Y = y] \tag{2.8}$$

On peut noter que dans le cas où les étiquettes X prennent leurs valeurs dans l'ensemble discret Ω, l'estimateur du champ moyen ne sera pas nécessairement dans cet ensemble.

2.2 Modèles graphiques

La théorie des modèles graphiques permet de modéliser notre connaissance sur les processus étudiés. Cette modélisation graphique des processus stochastiques a récemment suscité un vif intérêt [Lauritzen 96, Jordan 98, Perez 03]. Un graphe de dépendance permet de représenter de manière visuelle les dépendances existantes entre les variables aléatoires. Ceci permet d'utiliser efficacement les algorithmes existants sur les graphes pour réaliser différentes opérations comme le calcul de la loi jointe.

2.2.1 Notations et terminologie

On appelle *graphe* le couple $G = (S, L)$ où S est un ensemble fini de sites appelés *nœuds* et L un ensemble de liens entre ces sites nommés *arêtes*. Les arêtes $(a, b) \in L$ pour lesquelles à la fois (a, b) et $(b, a) \in L$ sont dites *non-orientées* tandis qu'une arête (a, b) dont l'*opposée* $(b, a) \notin L$ est dite *orientée*.

Ce graphe est souvent représenté par une figure où un *segment* joignant a et b représente une arête non orientée tandis qu'une *flèche* de a vers b représente une arête orientée (a, b) pour laquelle $(b, a) \notin L$.

Si le graphe est seulement constitué d'arêtes non orientées, il s'agit d'un graphe *non orienté*, tandis que si toutes les arêtes sont orientées, le graphe est dit *orienté*. Si il existe une flèche pointant de a vers b, a est un *parent* de b et b un *enfant* de a. L'ensemble des parents de b est noté $pa(b)$ et l'ensemble des enfants de a $ch(a)$. Si il existe un segment entre a et b, a et b sont dits *adjacents* ou *voisins*. L'ensemble des voisins d'un nœud est noté $ne(a)$.

Définition 2.1 (Système de voisinage)
Soit S un ensemble de sites. A tout site $s \in S$ est associé un ensemble $\mathcal{V}_s \subset S$ de voisins tel que
- $s \notin \mathcal{V}_s$,
- $s \in \mathcal{V}_s \Longleftrightarrow t \in \mathcal{V}_s$.

$\mathcal{N} \triangleq \{\mathcal{V}_s, s \in S\}$ est appelé système de voisinage.

Définition 2.2 (Clique)
Une clique c est soit un singleton, soit un sous-ensemble de S tel que tous ses sites soient mutuellement voisins : $\{s, t\} \subset c \Leftrightarrow t \in \mathcal{V}_s$.
On note C l'ensemble des cliques associées au système de voisinage $\mathcal{N}$.

2.2.2 Correspondance entre graphes et processus aléatoires

Une famille de variables aléatoires peut être naturellement indexée par un ensemble de sites constituant les nœuds du graphe représentant le processus aléatoire. Un graphe permet d'interpréter visuellement les relations entre les variables. Ces relations sont en fait des relations de dépendance conditionnelle entre les variables.

Définition 2.3 (Indépendance conditionnelle)
Soient A, B, C des variables aléatoires quelconques. A et B sont indépendantes conditionnellement à C si et seulement si

$$P(A = a, B = b|C = c) = P(A = a|C = c)P(B = b|C = c) \tag{2.9}$$

On note $A \perp B|C$.

Définition 2.4 (Graphe de dépendance)
Un graphe de dépendance $G = (S, L)$, orienté ou non, est composé d'un ensemble de nœuds S représentant des variables aléatoires, et d'un ensemble d'arêtes L tel que l'absence d'une arête entre les nœuds s et t représente une relation d'indépendance conditionnelle entre les variables aléatoires associées à ces deux sites.

La distribution jointe associée à un graphe donné peut être paramétrée par un produit de fonctions potentielles associées à des sous-ensembles de nœuds du graphe.

2.2.2.1 Graphe de dépendance orienté

Considérons d'abord le cas des graphes de dépendance orientés (également appelés *réseaux bayésiens*). Le sous-ensemble sur lequel le potentiel est défini consiste en un nœud et ses parents, et le potentiel correspond à la probabilité conditionnelle du nœud sachant ses parents. La loi jointe d'un processus X décrit par un graphe orienté s'écrit donc sous la forme :

$$p(x) \propto \prod_i p(x_i|x_{pa(i)}) \tag{2.10}$$

avec $p(x_i|x_{pa(i)})$ la probabilité conditionnelle locale associée au nœud i et $pa(i)$ l'ensemble des indices des parents du nœud i.

2.2.2.2 Graphe de dépendance non-orienté

Dans le cas des graphes de dépendance non-orientés, les sous-ensembles correspondent aux cliques du graphe. Pour une clique c donnée, on note $\psi_c(x_c)$ une fonction potentielle assignant un nombre réel positif à chaque configuration x_c. On a :

$$p(x) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c) \tag{2.11}$$

avec C l'ensemble des cliques associées au graphe et Z un facteur de normalisation assurant que $\sum_x p(x) = 1$.

Définition 2.5 (Champ de Markov)
Soit X un champ aléatoire indexé par un ensemble S. X est un champ de Markov relativement à un système de voisinage $\mathcal{N}$ si et seulement si :

$$\forall s \in S, p(X_s|X_t, t \neq s) = p(X_s|X_t, t \in \mathcal{V}_s) \tag{2.12}$$

Un processus aléatoire X défini sur un graphe non orienté est un champ de Markov. La définition des champs de Markov est locale. Cependant le théorème d'Hammersley et Clifford [Besag 74] établit une équivalence entre les champs de Markov et les champs de Gibbs caractérisés par leur distribution jointe. Un des principaux intérêts des champs de Markov provient de la possibilité d'avoir une expression de leur distribution globale $p(X)$.

Définition 2.6 (Champ de Gibbs)
Soit X un champ aléatoire indexé par un ensemble S et à valeur dans Ω. X est un champ de Gibbs relativement à un système de voisinage $\mathcal{N}$ si et seulement si sa distribution jointe est une distribution de Gibbs :

$$p(X = x) \triangleq \frac{1}{Z} \exp\{-U(x)\} \tag{2.13}$$

où Z est la constante de normalisation $Z \triangleq \sum_{x \in \Omega} \exp\{-U(x)\}$ et U se décompose en une somme de potentiels sur les cliques $U(x) \triangleq \sum_{c \in C} V_c(x)$.

Théorème 2.1 (Hammersley-Clifford)
Soit X un champ aléatoire indexé par un ensemble S et à valeur dans Ω. Soit $\mathcal{N}$ un système de voisinage sur S.
X est un champ de Markov relativement à $\mathcal{N}$ avec $\forall x \in \Omega, p(X = x) > 0 \Leftrightarrow X$ est un champ de Gibbs relativement à $\mathcal{N}$.

Remarque : L'équation 2.10 peut être vue comme un cas particulier de l'équation 2.11. $p(x_i|x_{pa(i)})$ est un bon exemple de fonction potentielle excepté le fait que l'ensemble des nœuds sur lesquels elle est définie $(i \cup pa(i))$ n'est en général pas une clique. En effet les parents d'un nœud donné ne sont pas forcément interconnectés. Pour pouvoir traiter de la même manière les équations 2.10 et 2.11, il est donc nécessaire de définir le *graphe moral* G^m associé au graphe orienté G. Le graphe moral est un graphe non-orienté obtenu en transformant toutes les arêtes orientées en non orientées et en ajoutant des arêtes entre les parents initialement non liés et qui ont des enfants en commun.

2.2.3 Lecture markovienne du graphe de dépendance

Nous allons maintenant nous intéresser aux informations probabilistes apportées par le graphe de dépendance $G = (S, L)$ d'un processus aléatoire X. Chaque arête absente capte une propriété d'indépendance conditionnelle de type markovien.

Proposition 2.1 (Markov par paire (MP))
Le processus X vérifie la propriété de Markov par paire (MP) si pour toute paire de sites $\{s, t\} \in S$ non voisins dans G, les variables aléatoires X_s et X_t sont indépendantes conditionnellement à toutes les autres, i.e., *:*

$$\forall \{s, t\} \in S \ : (s, t) \notin L \Rightarrow X_s \perp X_t | Z_{S \setminus \{s,t\}}$$

Si on s'intéresse à un site et son voisinage, on obtient la propriété markovienne la plus utilisée en imagerie [Lauritzen 96] :

Proposition 2.2 (Markov local (ML))
Soit $\mathcal{V}_s$ le voisinage de tout site $s \in S$. Le processus X vérifie la propriété de Markov locale (ML) si conditionellement à son voisinage $\mathcal{V}_s$, la variable aléatoire X_s est indépendante du reste du graphe, i.e., *:*

$$\forall s \in S \ : p(X_s|X_{S\setminus\{s\}}) = p(X_s|X_{\mathcal{V}_s})$$

Ces deux propriétés correspondent respectivement aux cas où deux sites non voisins sont séparés par les autres sites et où un site est séparé du reste par son voisinage. Elles montrent comment des situations de séparation dans le graphe de dépendance traduisent des propriétés d'indépendance conditionnelle. Ceci est également valable dans le cas le plus général d'un sous-ensemble de sites séparant deux autres sous-ensembles.

Proposition 2.3 (Markov global (MG))
Le processus X vérifie la propriété de Markov globale (MG) si pour trois sous-ensembles non vides et disjoints de sites a, b et c tels que a sépare b et c dans G[2]*, X_b et X_c sont indépendants conditionnellement à X_a.*

Ces trois propriétés ne sont pas nécessairement équivalentes. Nous avons [Lauritzen 96] :

$$\textbf{(MG)} \Rightarrow \textbf{(ML)} \Rightarrow \textbf{(MP)}$$

Théorème 2.2 (Pearl et Paz)
Si la propriété suivante est vérifiée :
$\forall$ *a, b, c, d sous-ensembles disjoints de S,*
si $X_a \perp X_b|(X_c \cup X_d)$ et $X_a \perp X_c|(X_b \cup X_d)$, alors $X_a \perp (X_b \cup X_c)|X_d$.
Alors on a l'équivalence :

$$\textbf{\textit{(MG)}} \Leftrightarrow \textbf{\textit{(ML)}} \Leftrightarrow \textbf{\textit{(MP)}}$$

2.2.4 Manipulation des graphes

Le graphe de dépendance permet d'appréhender la structure d'interaction sous-jacente à la loi jointe d'un ensemble de variables aléatoires. La manipulation de cette loi fait appel à trois mécanismes différents : le conditionnement, la marginalisation et le partitionnement. Ces trois mécanismes se traduisent par des transformations simples du graphe de dépendance original.

Conditionnement

Le conditionnement du graphe $G = (S, L)$ par rapport à un sous-ensemble $A \subset S$ revient à figer les variables aléatoires X_A et à considérer la loi conditionnelle résultante sur les variables restantes. Le graphe associé à la loi conditionnelle $p(X_{S\setminus A}|Z_A)$ est le graphe $G' = (S\setminus A, L')$ tel que $L' = \{(s,t) \in L : \{s,t\} \subset \setminus A\}$.

Le conditionnement se traduit graphiquement par la suppression des sites associés aux variables conditionnantes A et de toutes les arêtes ayant au moins une extrémité dans A (Fig 2.1).

[2] tout chemin d'un site de b vers un site de c traverse a.

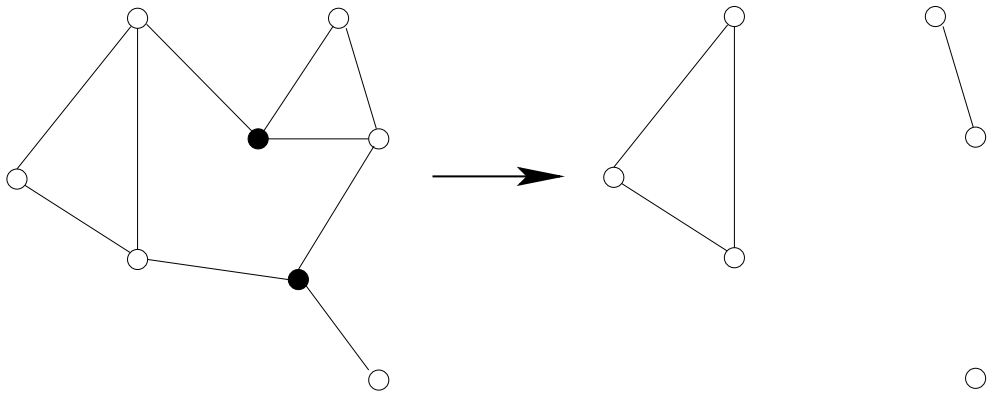


FIG. 2.1 – *Illustration du conditionnement par rapport aux variables sur les sites* •.

Marginalisation

La marginalisation implique la sommation ou l'intégration de la loi jointe par rapport à un sous-ensemble de variables $A \subset S$ de façon à obtenir la loi marginale des variables restantes (Fig 2.2). Le graphe correspondant à la loi marginale est le graphe $G' = (S \backslash A, L')$ tel que $L' = \{(s,t) \in (S \backslash A)^2 : ((s,t) \in L)$ ou $(\exists$ chaine $\subset A$ joignant s et $t)\}$.

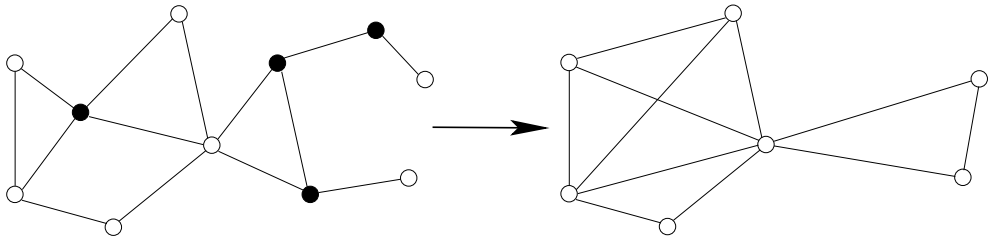


FIG. 2.2 – *Illustration de la marginalisation par rapport aux variables en gris.*

Partitionnement

Le partitionnement du graphe consiste à regrouper certaines variables entre elles pour former des "méta-variables" (Fig 2.3). On définit une partition d'ensembles non vides $(A_k)_{k=1}^p$ de S. Le graphe correspondant au vecteur aléatoire $(X_{A_k})_k$ résultat de la partition de X est le graphe $G' = (S', L')$ tel que $S' = \{1, \ldots, p\}$ et $L' = \{\{k,l\} \in S' : \exists (s,t) \in L, s \in A_k$ et $t \in A_l\}$.

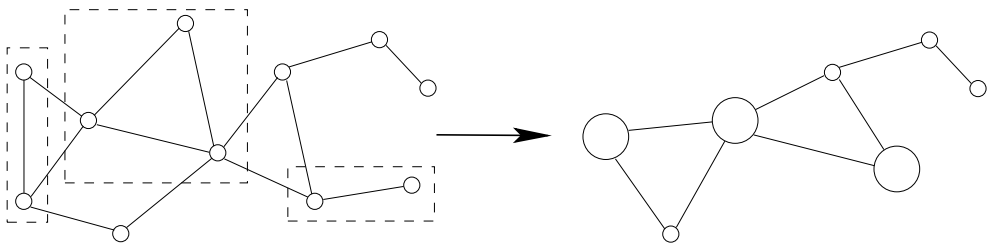


FIG. 2.3 – *Transformation du graphe de dépendance par agrégation.*

Pour tout processus aléatoire Z, constitué d'une partie observée Y et d'une partie cachée X pouvant être décrite par une ou plusieurs classes de lois factorisées, trois tâches génériques peuvent être rencontrées :

- Sélection de modèle : plusieurs classes de lois sont en compétition pour expliquer les données (partiellement) observées. La meilleure doit être déterminée.
- Apprentissage : une seule classe de lois est choisie, $\mathcal{M} = \{p(.;\theta), \theta \in \Theta\}$, paramétrée par un vecteur θ prenant ses valeurs dans un ensemble de valeurs admissibles Θ. Le problème est de trouver la valeur du vecteur θ permettant d'expliquer au mieux une ou plusieurs réalisations (partiellement) observées de x.
- Inférence : une loi jointe $p(x;\theta)$ étant choisie et une partie des composantes de x étant observée, il s'agit d'estimer au mieux celles qui ne le sont pas.

Différentes tâches élémentaires seront à la base de la résolution exacte ou approchée des trois problèmes génériques précédents :

- Marginalisation : dans le cadre d'une approche bayésienne, c'est le principal problème à résoudre, pour retirer du modèle toutes les variables cachées non pertinentes.
- Maximisation : certains estimateurs bayésiens requièrent la maximisation par rapport à tout ou partie de l'ensemble des variables cachées.
- Echantillonnage : afin d'approcher par méthode de Monte Carlo certaines quantités inaccessibles, telles que les marginales locales, il sera nécessaire de tirer des échantillons des lois choisies.

Dans le paragraphe suivant, nous présentons les principales techniques de résolution exactes et approximatives des problèmes d'inférence.

2.2.5 Algorithmes d'inférence probabiliste

2.2.5.1 Méthodes exactes d'inférence

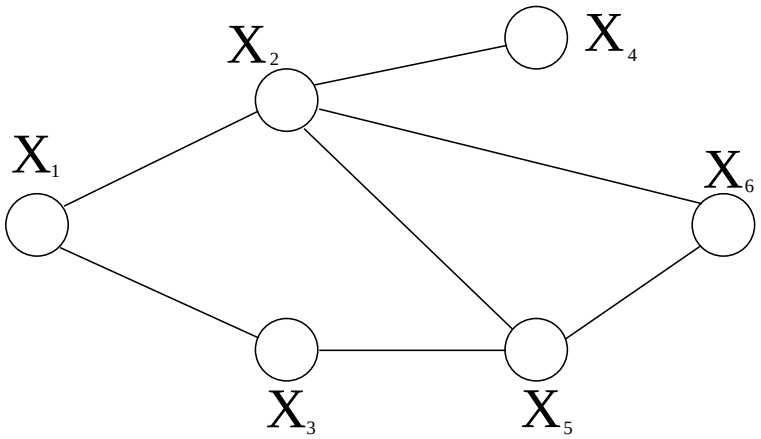


FIG. 2.4 – *Exemple de graphe de dépendance non-orienté.*

Elimination Dans ce paragraphe, nous introduisons un algorithme de base appelé élimination. Commençons par un exemple. Considérons le modèle graphique présenté sur la figure 2.4. Calculons

la marginale $p(x_1)$. Dans le cas de variables aléatoires discrètes, on obtient cette marginale en sommant sur les variables restantes.

$$p(x_1) = \sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5}\sum_{x_6}\frac{1}{Z}\psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6) \tag{2.14}$$

En supposant que le cardinal de chaque variable vaut r, la complexité calculatoire est de r^6. On peut obtenir une complexité plus faible :

$$\begin{aligned}
p(x_1) &= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)\sum_{x_6}\psi(x_2,x_5,x_6)\\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)m_6(x_2,x_5)\\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\sum_{x_4}\psi(x_2,x_4)\\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)m_3(x_1,x_2)\\
&= \frac{1}{Z}m_2(x_1)
\end{aligned}$$

en définissant de facteurs intermédiaires m_i. On obtient la valeur de Z et donc la marginale en sommant l'expression finale suivant x_1. Au plus trois variables apparaissent dans chaque sommation, la complexité calculatoire a donc été réduite à r^3. On peut effectuer la sommation dans un ordre différent. Cependant il est plus judicieux de choisir un ordre tel que la complexité soit minimale.

Une limitation de cette approche par élimination provient du fait qu'on obtient ainsi une seule marginale, alors qu'en pratique, il est souvent nécessaire de connaître plusieurs marginales. Pour éviter de calculer chaque marginale en lançant plusieurs fois l'algorithme d'élimination sans exploiter le fait que certains termes intermédiaires sont communs à plusieurs calculs, d'autres algorithmes peuvent être utilisés.

Algorithme de passage du message Dans ce cas on se restreint au cas particulier des *arbres*. Pour les arbres non-orientés $G = (S, L)$, les cliques correspondent à des singletons ou à des paires de nœuds et la distribution est caractérisée par les potentiels $\psi(x_i) : i \in S$ et $\psi(x_i, x_j) : (i, j) \in L$. Pour calculer la marginale $p(x_f)$, considérons un arbre pour lequel le nœud f est la racine. Dans ce nouvel arbre, il faut choisir un ordre d'élimination tel que tous les enfants de chaque nœud sont éliminés avant leur parent. Les étapes de l'algorithme d'élimination peuvent ainsi s'écrire de la manière générale suivante :

$$m_{ji}(x_i) = \sum_{x_j}\left(\psi(x_j)\psi(x_i,x_j)\prod_{k\in\mathcal{N}(j)\setminus i}m_{kj}(x_j)\right), \tag{2.15}$$

où $\mathcal{N}(j)$ est l'ensemble des voisins du nœud j et $m_{ji}(x_i)$ correspond au terme créé lorsque le nœud j est éliminé. La sommation suivant x_j crée ainsi un *message* $m_{ji}(x_i)$ envoyé au nœud j. Un nœud peut envoyer un message à un nœud voisin une fois qu'il a reçu les messages de tous ses autres voisins. La marginale désirée est obtenue par :

$$p(x_f) \propto \psi(x_f) \prod_{e \in \mathcal{N}(f)} m_{ef}(x_f), \tag{2.16}$$

où la constante de proportionnalité est obtenue en sommant le terme de droite suivant x_f.

Algorithme de l'arbre de jonction L'algorithme de l'arbre de jonction est une généralisation de l'algorithme précédent pour un graphe quelconque. L'idée principale consiste à transformer le graphe en hypergraphe dans lequel les nœuds sont des cliques plutôt que des singletons. En général il n'est pas possible d'utiliser les cliques du graphe original. Les cliques utilisées sont donc celles d'un graphe augmenté obtenu par triangulation du graphe original. Nous ne détaillons pas ici la construction de l'arbre des cliques (ou *arbre de jonction*) mais il est construit en choisissant un ordre d'élimination et en effectuant les opérations graphiques associées.

Considérons un graphe triangulé avec les cliques $c_i \in C$ et les potentiels $\psi_{c_i}(x_{c_i})$ et un arbre de jonction correspondant définissant les liens entre les cliques. Le message passé de la clique c_i à la clique c_j est :

$$m_{ij}(x_{S_{ij}}) = \sum_{c_i \setminus S_{ij}} \left(\psi_{c_i}(x_{c_i}) \prod_{k \in \mathcal{N}(i) \setminus j} m_{ki}(x_{S_{ki}}) \right), \tag{2.17}$$

où $S_{ij} = c_i \cap c_j$ et $\mathcal{N}(j)$ sont les voisins de la clique c_i de l'arbre de jonction. Les marginales sont obtenues en effectuant le produit des messages. Ainsi :

$$p(x_{c_i}) \propto \prod_{k \in \mathcal{N}(i)} m_{ki}(x_{S_{ki}}) \tag{2.18}$$

correspond à la marginale de la clique c_i.

Les méthodes exactes d'inférence telles que l'élimination, l'algorithme de passage du message et l'algorithme de l'arbre de jonction permettent de calculer les marginales en exploitant la structure graphique. Cependant en pratique, ces algorithmes ne sont pas toujours utilisables. Dans ce cas, on peut avoir recours à des méthodes approximatives.

2.2.5.2 Méthodes approximatives d'inférence

Nous présentons deux classes de méthodes approximatives : les algorithmes de Monte Carlo et les méthodes variationnelles.

Algorithmes de Monte Carlo Les algorithmes de Monte Carlo se basent sur le fait que dans le cas où la moyenne suivant une loi $p(x)$ n'est pas calculable, il est possible de tirer des échantillons suivant cette loi ou une loi assez proche de telle sorte que les marginales ou autres quantités d'intérêt puissent être approximées à partir des moyennes empiriques des échantillons

tirés. Différents algorithmes de Monte Carlo sont utilisés, les plus courants étant l'échantillonneur de Gibbs, l'algorithme de Metropolis-Hastings ou l'échantillonnage pondéré (*Importance sampling*).

L'**échantillonneur de Gibbs** est un exemple des algorithmes de Monte Carlo par chaîne de Markov (MCMC[3]). Dans le cas des algorithmes MCMC, les échantillons sont obtenus à partir d'une chaîne de Markov dont l'état stationnaire donnerait la distribution désirée $p(x)$. La chaîne de Markov de l'échantillonneur de Gibbs est construite de la manière suivante :

1. à chaque étape, une des variables X_i est sélectionnée ;
2. la distribution conditionnelle $p(x_i|x_{S\setminus i})$ est calculée (où S est l'ensemble des nœuds du graphe $G = (S, L)$) ;
3. une valeur x_i est choisie suivant cette loi ;
4. l'échantillon x_i remplace la valeur précédente de la $i^{\text{ème}}$ variable.

L'implémentation de l'échantillonneur de Gibbs se réduit donc au calcul de la loi conditionnelle des variables individuelles sachant toutes les autres variables. Dans le cas des modèles graphiques, ces lois conditionnelles sont de la forme :

$$p(x_i|x_{S\setminus i}) = \frac{\prod_{c\in C_i} \psi_c(x_c)}{\sum_{x_i} \prod_{c\in C_i} \psi_c(x_c)} \tag{2.19}$$

où c_i représente l'ensemble des cliques contenant l'index i. Cet ensemble est généralement plus petit que l'ensemble C de toutes les cliques, et dans ce cas chaque étape de l'échantillonneur peut être implémentée facilement.

Dans le cas où le calcul de l'équation 2.19 est trop complexe, l'**algorithme de Metropolis-Hastings** fournit une alternative. L'algorithme de Metropolis-Hastings est un algorithme MCMC non basé sur les probabilités conditionnelles. Par conséquent il ne requière pas de normalisation. Etant donné l'état courant x de l'algorithme, Metropolis-Hastings choisit le nouvel état $\tilde{x}$ suivant une loi $q(\tilde{x}|x)$, qui consiste le plus souvent à tirer une variable X_i de façon aléatoire et à choisir une nouvelle valeur pour cette variable de manière aléatoire. L'algorithme calcule ensuite la probabilité d'acceptation :

$$\alpha = \min\left(1, \frac{q(x|\tilde{x})}{q(\tilde{x}|x)} \frac{\prod_{c\in C_i} \psi_c(\tilde{x}_c)}{\prod_{c\in C_i} \psi_c(x_c)}\right) \tag{2.20}$$

L'algorithme accepte $\tilde{x}$ avec la probabilité α.

Alors que l'échantillonneur de Gibbs et l'algorithme de Metropolis-Hastings utilisent des échantillons tirés selon $p(x)$, l'**échantillonnage pondéré** (*importance sampling*) est une technique de Monte Carlo approximant la moyenne de la fonction $f(x)$ suivant la loi $p(x)$ en utilisant M

[3] *Markov chain Monte Carlo*

échantillons $x^m, m = 1, \ldots, M$ tirés suivant une loi $q(x)$:

$$\begin{aligned} E[f(x)] &= \sum_x p(x) f(x) \\ &= \sum_x q(x) \left(\frac{p(x)}{q(x)} f(x) \right) \\ &\approx \frac{1}{M} \sum_{m=1} M \frac{p(x^{(m)})}{q(x^{(m)})} f(x^{(m)}) \end{aligned}$$

Les principaux avantages des algorithmes de Monte Carlo sont leur simplicité d'implémentation et leur généralité. Cependant le temps de convergence peut être long.

Méthodes variationnelles L'idée de base des méthodes variationnelles [Jordan 99] consiste à convertir le problème d'inférence probabiliste en problème d'optimisation de façon à exploiter les outils standards de l'optimisation sous contrainte. Cette approche peut ressembler à l'échantillonnage pondéré, mais au lieu de choisir une seule distribution $q(x)$ a priori, une famille de distributions $\{q(x)\}$ est utilisée et le système d'optimisation choisit une distribution particulière de cette famille.

2.3 Modèles markoviens utilisés en traitement d'images

Les modèles de Markov cachés suscitent l'intérêt de la communauté du traitement de l'information depuis leurs premiers succès en reconnaissance automatique de la parole au début des années 70 [Rabiner 89]. De nos jours, leurs domaines d'application couvrent le traitement des images [Giordana 97], la génomique, la météorologie [Bellone 00], les finances [Thomas 02], *etc.* Les modèles markoviens fournissent en effet de puissants outils statistiques pour des applications en vision par ordinateur. Ces approches permettent de modéliser notre connaissance sur les processus étudiés, étape nécessaire dans le cadre de la théorie de décision basée sur l'inférence bayésienne. Les champs de Markov cachés sont ainsi souvent utilisés en imagerie depuis plus de vingt ans et se sont avérés être d'une grande efficacité dans certains cas. Leur intérêt est de pouvoir prendre en compte l'information contextuelle dans une image, de manière mathématiquement rigoureuse. 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. Nous les avons introduit dans le cadre de la segmentation d'IRM au paragraphe 1.2.1.3. Nous allons rapidement rappeler les principaux algorithmes de segmentation bayésienne fondés sur les champs de Markov. Les solutions des méthodes bayésiennes MPM et MAP ne sont pas calculables directement et des procédures d'approximation doivent donc être utilisées. La solution au sens du critère MPM est approchée par l'algorithme de Marroquin [Marroquin 87], celle du MAP par l'algorithme du recuit simulé de Geman *et al.* [Geman 84]. L'algorithme ICM de Besag [Besag 86] peut être interprété comme une approximation plus grossière et rapide de la solution du MAP.

Les Chaînes de Markov Cachées que nous définirons plus précisément au paragraphe 2.4.2, constituent une alternative, que nous jugeons pertinente, à l'approche par champs de Markov cachés. Le modèle est certes moins intuitif, mais les algorithmes obtenus sont plus exacts et beaucoup

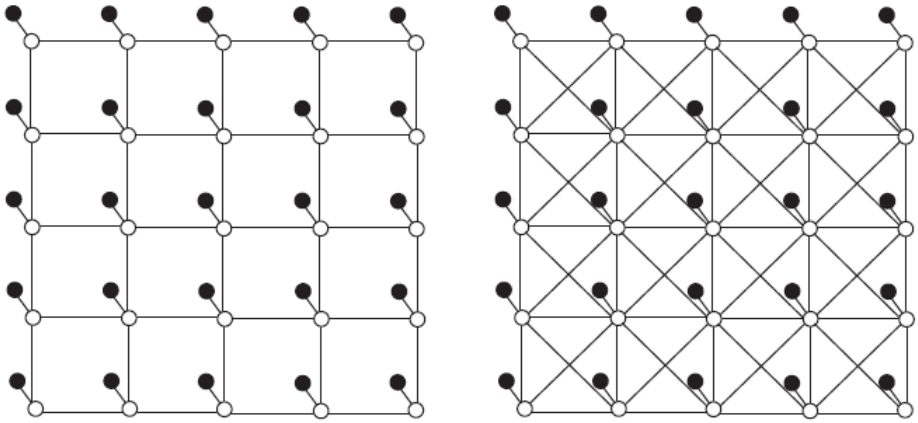


FIG. 2.5 – *Graphe de dépendance d'un champ de Markov : voisinage 4-connexité à gauche et voisinage 8-connexité à droite. Les cercles blancs représentent le états cachés et les cercles noirs les observations.*

plus rapides. En effet les lois de probabilité *a posteriori* sont calculables analytiquement ce qui constitue le principal avantage des méthodes par chaînes par rapport aux modèles par champs où ces probabilités doivent êtres estimées par des procédures itératives fort coûteuses en temps de calcul. Les algorithmes d'optimisation et d'estimation des paramètres sont alors considérablement simplifiés par rapport au cas des champs.

Les arbres de Markov [Laferté 00] constituent une généralisation des chaînes de Markov que l'on retrouve dans le cas où les générations sont réduites au singleton. Ils apparaissent comme concurrents des champs de Markov cachés avec les avantages suivants : une plus grande vitesse d'exécution à la fois des méthodes d'estimation des paramètres et des méthodes de segmentation, ainsi qu'une excellente adéquation au traitement des images multirésolutions [Chardin 00, Provost 01, Flitti 05]. Les relations spatiales sont cependant moins bien modélisées que dans le cas des champs car deux sites voisins n'ont pas forcément le même parent dans l'arbre. L'intérêt des quadarbres réside dans la possibilité d'utiliser des algorithmes itératifs rapides pour calculer les solutions au sens du MAP ou du MPM. Par rapport au cas des champs, les algorithmes d'estimation des paramètres sont simplifiés et des versions exactes de l'algorithme EM peuvent être mises en place dans certains cas.

Des modèles "mixtes" ont récemment été développés. Ces modèles utilisent un champ de Markov à une résolution grossière et un arbre aux résolutions plus fines pour remédier aux effets de blocs pouvant apparaître dans les méthodes par arbres [Pérez 00].

Récemment diverses généralisations des modèles de Markov cachés ont été proposées. La première généralisation concerne les modèles de Markov couples, pour lesquels on considère directement la markovianité du couple constitué du processus caché et du processus observé. La deuxième généralisation concerne l'introduction d'un processus auxiliaire et la considération d'un processus triplet constitué du processus caché, du processus auxiliaire et du processus observé. Ces modèles de Markov triplets permettent en particulier de traiter des données non stationnaires, ou encore des données à mémoire longue modélisées par la semi-markovianité. Ces modèles seront présentés plus en détail dans le cas des chaînes.

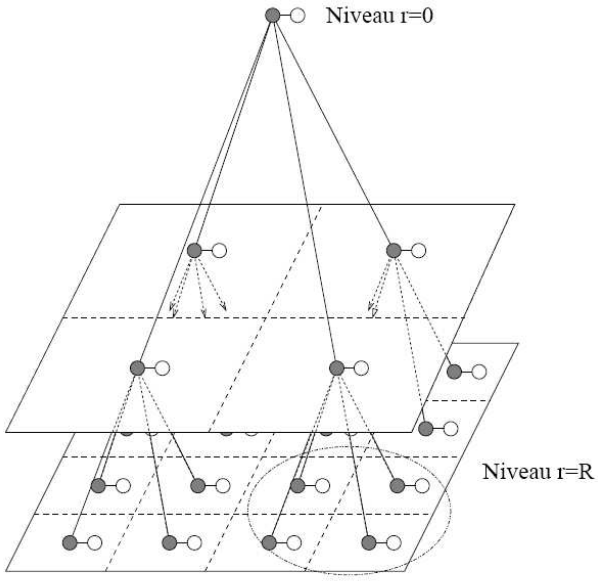


FIG. 2.6 – *Graphe de dépendance d'un quadarbre. Les cercles noirs représentent les étiquettes et les disques blancs les observations qui peuvent être multimodales. Ce modèle permet de gérer des observations à différentes résolutions.*

2.4 Chaînes de Markov

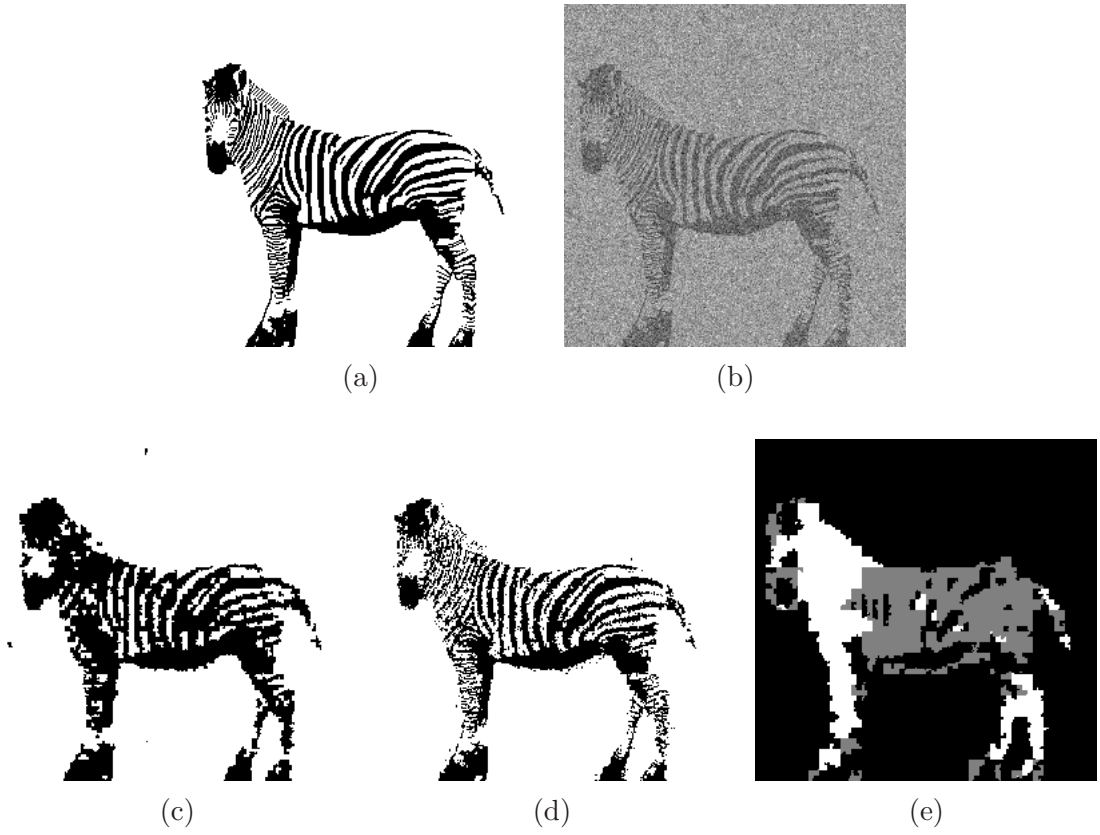
2.4.1 Présentation générale

Nous allons commencer par présenter le modèle général des Chaînes de Markov Triplets (CMT) puis nous détaillerons quelques cas particuliers de ce modèle.

Chaînes de Markov triplets (CMT)

Soient $X = (X_1, \ldots, X_N)$, $Y = (Y_1, \ldots, Y_N)$ et $U = (U_1, \ldots, U_N)$ trois processus aléatoires, X_n prenant ses valeurs dans un ensemble fini $\Omega = \{\omega_1, \ldots, \omega_K\}$, Y_n prenant ses valeurs dans l'ensemble des réels $\mathbb{R}$ et U_n à valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$. On pose $T = (T_1, \ldots, T_N)$, avec $T_n = (X_n, U_n, Y_n)$. On aura une chaîne de Markov triplet (CMT) si le triplet $T = (X, U, Y)$ est de Markov [Pieczynski 02a, Pieczynski 02b]. La chaîne U n'a pas nécessairement une signification physique. Le modèle CMT est ainsi un modèle très riche. Une application possible du modèle CMT est leur utilisation pour traiter le problème de la non stationnarité de la chaîne $Z = (X, Y)$ [Lanchantin 04]. Dans ce cas, la chaîne U a une signification physique, chaque état dans $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$ modélisant une certaine stationnarité. La figure 2.7 illustre la modélisation de la non-stationnarité sur une image de zèbre [Lanchantin 06]. L'image binaire du zèbre (Fig. 2.7.a) est bruitée à l'aide d'un bruit gaussien, spatialement indépendant de moyennes respectives 0 et 2 et de variance égale à 1 pour les deux classes. L'image observée (Fig. 2.7.b) est segmentée de deux

manières différentes : en utilisant le modèle CMC-BI (Chaîne de Markov cachée à bruit indépendant classique qui sera présenté au paragraphe 2.4.2) et le modèle CMT. Dans le cas où on considère trois stationnarités différentes (CMT avec trois classes auxiliaires), la recherche de $U = u$ donne $\hat{U} = \hat{u}$ présenté sur la figure 2.7(e) : les labels noirs sont attribués au fond, les labels gris aux bandes larges (sur le corps du zèbre) et les labels blancs aux bandes étroites (sur le cou et les pattes du zèbre). D'autre part on peut constater que la qualité de la segmentation est nettement améliorée en utilisant le modèle CMT (Fig. 2.7.d) par rapport au modèle CMC-BI (Fig. 2.7.c).



(a) (b)

(c) (d) (e)

FIG. 2.7 – *(a) Image $X = x$ d'un zèbre, (b) sa version bruitée $Y = y$ par du bruit gaussien indépendant, (c) segmentation non supervisée utilisant les CMC-BI, (d) segmentation non supervisée utilisant les CMT, (e) segmentation $\hat{U} = \hat{u}$ en trois stationnarités différentes utilisant les CMT. Illustration tirée de [Lanchantin 06].*

Les chaînes de Markov triplets permettent aussi de généraliser les chaînes semi-markoviennes cachées (CSMC). Considérons un processus $X = (X_1, \ldots, X_n, \ldots)$ où chaque X_i est à valeurs dans $\Omega = \{\omega_1, \ldots, \omega_k\}$. X est une chaîne *semi-markovienne* si sa loi est donnée par la loi de X_1 notée $p(x_1)$, la suite des matrices de transition $(p(x_i|x_{i-1}))_{i\geq 2}$ vérifiant $p(x_i|x_{i-1}) = 0$ pour $x_i = x_{i-1}$, et k suites de lois sur N^*. Pour chaque $i = 1, \ldots, k$, on note $p^*(.|x_n = \omega_i)$ la suite des lois correspondantes. Une réalisation de la chaîne semi-markovienne $X = (X_1, \ldots, X_n, \ldots)$ est obtenue par :

1. $X_1 = x_1$ est simulé selon $p(x_1)$
2. on simule un entier naturel $N_1 = n_1$ selon $p^*(.|x_1)$

3. pour $1 \leq i \leq n_1$, on pose $x_i = x_1$
4. on simule $X_{n_1+1} = x_{n_1+1}$ selon $p(x_{n_1+1}|x_{n_1})$
5. on simule un entier naturel $N_2 = n_2$ selon $p^*(.|x_{n_1})$
6. ...

On peut montrer que les CSMC sont des CMT particulières [Pieczynski 05]. La chaîne U est telle que pour chaque n, U_n modélise le temps restant de séjour de X_n dans x_n. On peut montrer la proposition suivante :

Proposition 2.4
Une chaîne semi-markovienne cachée est une chaîne de Markov triplet $T = (X, U, Y)$ avec U_n à valeurs dans N^*, définie par $p(x_1, u_1, y_1) = p(x_1)p(u_1|x_1)p(y_1|x_1)$ et les transitions

$$\begin{aligned} p(x_{n+1}, u_{n+1}, y_{n+1}|x_n, u_n, y_n) &= p(x_{n+1}|x_n, u_n, y_n)p(u_{n+1}|x_n, u_n, y_n, x_{n+1}) \times \\ & \quad p(y_{n+1}|x_n, u_n, y_n, x_{n+1}, u_{n+1}) \end{aligned} \tag{2.21}$$

données par (δ désignant la masse de Dirac) :

$$p(x_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n) = \begin{cases} \delta(x_n) & si\ u_n > 1 \\ p(x_{n+1}|x_n) & si\ u_n = 1 \end{cases} \tag{2.22}$$

$$p(u_{n+1}|x_n, u_n, y_n, x_{n+1}) = p(u_{n+1}|x_{n+1}, u_n) = \begin{cases} \delta(u_n - 1) & si\ u_n > 1 \\ p(u_{n+1}|x_{n+1}) & si\ u_n = 1 \end{cases} \tag{2.23}$$

$$p(y_{n+1}|x_n, u_n, y_n, x_{n+1}, u_{n+1}) = p(y_{n+1}|x_{n+1}) \tag{2.24}$$

Les chaînes de Markov triplets peuvent aussi modéliser différentes propriétés simultanément en considérant la chaîne auxiliaire U sous forme multi-variée $U = (U^1, \ldots, U^d)$, où chaque U^i modélise une certaine propriété : par exemple U^1 peut modéliser la non-stationnarité et U^2 la semi-markovianité. De plus nous avons supposé que U était à valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$ mais on peut aussi supposer que chaque U_n prend ses valeurs dans l'ensemble des réels $\mathbb{R}$.

A partir de ce modèle général des chaînes de Markov triplets, on peut définir un certain nombre de modèles particuliers dont les graphes de dépendance sont présentés 2.8.

Chaînes de Markov couples (CMCouple)

Soit $Z = (Z_1, \ldots, Z_N)$, avec $Z_n = (X_n, Y_n)$. Le processus Z sera une chaîne de Markov couple (CMCouple) s'il est de Markov [Pieczynski 03b, Derrode 04]. Z est une CMCouple si et seulement si :

$$p(z) = p(z_1)p(z_2|z_1)\ldots p(z_N|z_{N-1}) \tag{2.25}$$

Dans ce modèle $p(y_n|x_{n-1}, x_n)$ dépend de x_n et de x_{n-1} et les observations sont corrélées. Nous obtenons :

$$p(y_{n-1}, y_n|x_{n-1}, x_n) = p(y_n|x_n, z_{n-1})p(y_{n-1}|x_{n-1}, x_n) \tag{2.26}$$

On peut en déduire que :

$$p(z_n|z_{n-1}) = p(x_n|z_{n-1})p(y_n|x_n, z_{n-1}) \tag{2.27}$$

Pieczynski a démontré une condition nécessaire et suffisante pour qu'une CMT soit une CMCouple [Pieczynski 02a].

Lien entre une CMT et une CMCouple Soit $T = (X, U, Y)$ une CMT vérifiant (H) :
(H) Pour tout $n \in N$, $z_n = z_{n+2}$ implique $p(u_{n+1}|z_n, z_{n+1}) = p(u_{n+1}|z_{n+1}, z_{n+2})$.
Alors $Z = (X, Y)$ est une CMCouple si et seulement si pour tout $n \in N$, $p(u_{n+1}|z_n, z_{n+1}) = p(u_{n+1}|z_{n+1})$.

Chaînes de Markov couples à bruit indépendant (CMCouple-BI) [Lanchantin 06]

Dans ce cas, on suppose que les observations sont indépendantes conditionnellement à X. Nous avons donc :

$$p(y_{n-1}, y_n|x_{n-1}, x_n) = p(y_n|x_{n-1}, x_n)p(y_{n-1}|x_{n-1}, x_n) \tag{2.28}$$

Dans le cas d'une CMCouple-BI, les probabilités de transition ont la forme suivante :

$$p(z_n|z_{n-1}) = p(x_n|z_{n-1})p(y_n|x_n, x_{n-1}) \tag{2.29}$$

Chaînes de Markov cachées (CMC)

Dans ce modèle les corrélations entre les observations sont prises en compte tout en faisant l'hypothèse que $p(y_n|x_{n-1}, x_n)$ ne dépend que de x_n. Nous avons donc :

$$p(y_{n-1}, y_n|x_{n-1}, x_n) = p(y_n|y_{n-1}, x_n)p(y_{n-1}|x_{n-1}) \tag{2.30}$$

On peut en déduire que :

$$p(z_n|z_{n-1}) = p(x_n|x_{n-1})p(y_n|y_{n-1}, x_n) \tag{2.31}$$

Chaînes de Markov cachées à bruit indépendant (CMC-BI) [Lanchantin 06]

On suppose que $p(y_n|x_{n-1}, x_n)$ ne dépend que de x_n et que les observations ne sont par corrélées. Nous avons donc :

$$p(y_{n-1}, y_n|x_{n-1}, x_n) = p(y_n|x_n)p(y_{n-1}|x_{n-1}) \tag{2.32}$$

Dans le cas d'une CMC-BI, les probabilités de transition ont la forme suivante :

$$p(z_n|z_{n-1}) = p(x_n|x_{n-1})p(y_n|x_n) \tag{2.33}$$

Intérêt des différentes modélisations

L'intérêt du modèle CMT provient du fait que c'est un modèle très général permettant de multiples modélisations (non-stationnarité, généralisation des CSMC) grâce à l'introduction d'un processus auxiliaire $U = (U_1, \ldots, U_n)$ n'ayant pas forcément une signification physique. Ce modèle suppose la markovianité du triplet $T = (X, U, Y)$. Les modèles par chaînes de Markov couples supposent la markovianité du couple $Z = (X, Y)$, ce qui permet d'autoriser des modélisations complètes de $p(y|x)$ [Pieczynski 03b, Derrode 04]. Considérer que $p(y_n|x_{n-1}, x_n)$ dépend de x_n et également de x_{n-1} peut s'avérer utile dans une application de segmentation d'image pour modéliser le fait que la loi d'observation d'une classe donnée peut être différente près d'une frontière entre classes différentes et à l'intérieur d'un ensemble de sites d'une même classe. De plus, la prise en compte de la corrélation entre les observations peut permettre d'accroître la qualité de la classification [Derrode 04]. Ce modèle général peut être simplifié soit en supposant

que les observations ne sont pas corrélées (modèles de bruit indépendant) soit en considérant que $p(y_n|x_{n-1}, x_n)$ dépend seulement de x_n (modèles de Chaînes de Markov cachées). Le modèle CMC-BI considère à la fois que $p(y_n|x_{n-1}, x_n)$ dépend seulement de x_n et que les observations ne sont pas corrélées. L'intérêt de ce modèle réside dans le fait qu'un certain nombre de quantités d'intérêts sont calculables, en particulier $p(x_n|y)$.

Différents graphes représentant quelques modèles de chaînes de Markov sont présentés sur la figure 2.8. Ces modélisations couples et triplets ainsi que la plupart de leurs propriétés peuvent être facilement étendues aux arbres et champs de Markov.

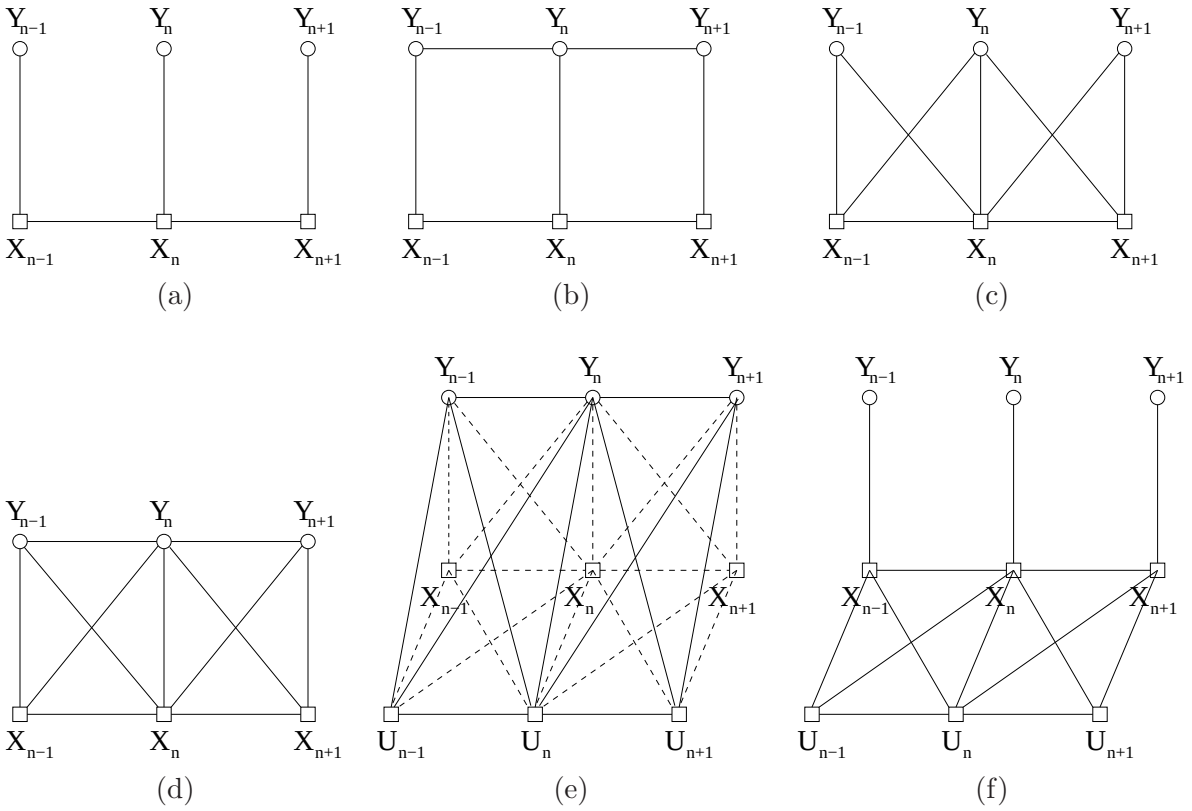


FIG. 2.8 – *Différents graphes : (a) Chaîne de Markov à bruit indépendant. (b) Chaîne de Markov. (c) Chaîne de Markov couple à bruit indépendant. (d) Chaîne de Markov couple. (e) Chaîne de Markov triplet. (f) Chaîne de Markov cachée non stationnaire.*

2.4.2 Chaînes de Markov Cachées

Définition

Dans ce paragraphe nous allons détailler plus particulièrement le modèle des chaînes de Markov cachées. Les chaînes de Markov cachées sont utilisées dans de nombreux domaines du traitement du signal monodimensionnel. Les principales applications concernent la reconnaissance de la parole, la reconnaissance de texte.

Une chaîne de Markov [Fjortoft 03] est une suite de variables aléatoires notées $(X_n)_{n\in S}$, où n est l'indice de l'élément étudié du vecteur. Cette suite est définie dans l'espace des étiquettes

$\Omega = \{\omega_1, ..., \omega_K\}$. La suite $(X_n)_{n \in S}$ est une chaîne de Markov d'ordre 1 si :

$$p(x_{n+1}|x_n, ..., x_1) = p(x_{n+1}|x_n) \tag{2.34}$$

c'est-à-dire si l'influence sur un voxel conditionnellement à son passé se réduit à l'influence conditionnée par son prédécesseur.

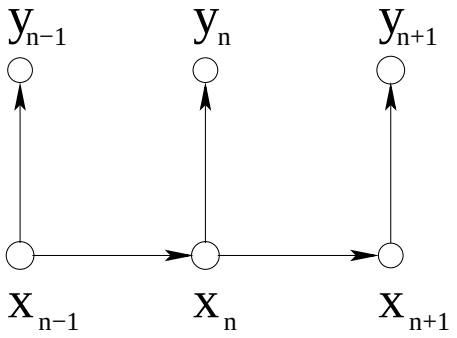


FIG. 2.9 – *Graphe de dépendance d'une chaîne de Markov Cachée d'ordre 1.*

La chaîne sera supposée homogène, c'est-à-dire que la matrice de transition ne dépend pas de n. En revanche les chaînes ne seront pas supposées stationnaires, ce qui implique $\pi_i \neq P(X_n = \omega_i)$.

La loi de X est donc entièrement déterminée par la connaissance, d'une part de la probabilité initiale d'appartenance à une classe π_i, et d'autre part des probabilités de transition d'une classe à l'autre t_{ij} :

$$\pi_i = p(x_1 = \omega_i) \tag{2.35}$$

$$t_{ij} = p(x_{n+1} = \omega_j | x_n = \omega_i) \quad \forall n \neq N \tag{2.36}$$

La figure 2.9 présente le graphe d'une chaîne de Markov d'ordre 1 à bruit indépendant : Y représente les observations à partir desquelles on recherche les étiquettes cachées X.

Application à la segmentation d'images et intérêt par rapport aux champs

La segmentation d'IRM par chaînes de Markov nécessite la manipulation d'un vecteur, c'est-à-dire la transformation d'une image tridimensionnelle en un vecteur monodimensionnel. Lorsque ce dernier a été segmenté, la transformation inverse est effectuée afin d'obtenir l'image segmentée. La vectorisation peut être réalisée grâce au parcours fractal d'Hilbert-Peano en 3D (Fig. 2.10)[Bandoh 99]. Cet algorithme permet de parcourir de manière continue une image cubique de taille $2^n \times 2^n \times 2^n$, $n \in N$ et donc de générer un vecteur monodimensionnel. L'intérêt des méthodes par Chaînes de Markov Cachées par rapport aux méthodes par Champs de Markov 3D pour la segmentation d'images est que, basées sur un modèle unidimensionnel, leur coût calculatoire est plus faible. Néanmoins dans le modèle des Chaînes de Markov, le voisinage est partiellement pris en compte à travers le parcours d'Hilbert-Peano : deux voisins dans la chaîne sont voisins dans la grille mais deux voisins dans l'image 3D peuvent être fréquemment très éloignés dans la chaîne. Cependant, vraisemblablement en raison d'une forte corrélation à l'intérieur du cube de données, le parcours a une faible influence sur les résultats de segmentation. L'influence de différents parcours d'Hilbert-Peano sur la segmentation finale sera étudiée dans le chapitre 3 dans le cas de

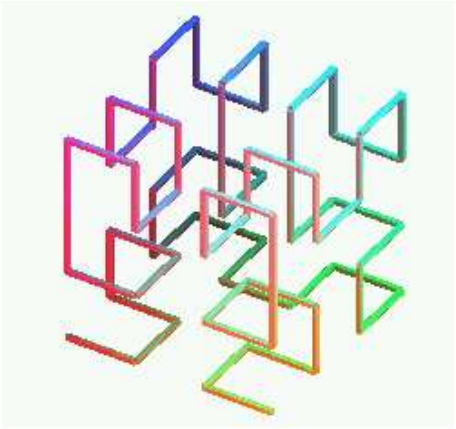


FIG. 2.10 – *Parcours d'Hilbert-Peano en 3D.*

la segmentation d'une image IRM 3D. Un autre avantage des chaînes par rapport aux champs est que dans le cas des chaînes, la loi *a posteriori* peut être estimée exactement, sans nécessiter d'approximation contrairement aux méthodes par champs.

Les méthodes par chaînes de Markov ont déjà été utilisées pour segmenter des images 2D, en particulier des images SAR. Dans [Fjortoft 03], Fjortoft *et al.* comparent la classification obtenue à la fois sur des images synthétiques et des images radar réelles en utilisant les chaînes et les champs de Markov. La classification obtenue avec les chaînes est parfois moins régulière que celle obtenue par les champs mais les structures fines sont généralement mieux préservées.

2.5 Algorithmes d'optimisation

2.5.1 Algorithme d'optimisation sur une chaîne de Markov

La mise en oeuvre des algorithmes d'estimation et de segmentation nécessitent un certain nombre d'hypothèses sur l'indépendance conditionnelle.

– Les observations $\boldsymbol{Y}_n$ (où $\boldsymbol{Y}_n$ prend ses valeurs dans $\mathbb{R}^m$, $\boldsymbol{Y}_n = [Y_n^1, \ldots, Y_n^m]$) sont supposées indépendantes conditionnellement à X :

$$p(\boldsymbol{Y}|X) = \prod_{n=1}^{N} p(\boldsymbol{Y}_n = \boldsymbol{y}_n|X) \tag{2.37}$$

– La loi de $\boldsymbol{Y}_n$ conditionnellement à X est égale à la loi de $\boldsymbol{Y}_n$ sachant X_n :

$$p(\boldsymbol{Y}_n = \boldsymbol{y}_n|X) = p(\boldsymbol{Y}_n = \boldsymbol{y}_n|X_n = \omega_k) \tag{2.38}$$

Les équations 2.37 et 2.38 aboutissent à :

$$p(\boldsymbol{Y}|X) = \prod_{n=1}^{N} p(\boldsymbol{Y}_n = \boldsymbol{y}_n|X_n = \omega_k) \tag{2.39}$$

Les paramètres *a priori* sont regroupés dans $\Phi_x = \{\pi_k, t_{kl}\}$ où π_k et t_{kl} correspondent respectivement à la probabilité initiale d'appartenance à la classe k et à la matrice de transition

d'une classe à une autre. Les paramètres de la loi d'attache aux données sont regroupés dans $\Phi_y = \{\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}\}$ où $\boldsymbol{\mu_k}$ et $\boldsymbol{\Sigma_k}$ correspondent respectivement à la moyenne et l'écart-type de la classe k. Dans la suite nous noterons l'ensemble des paramètres $\Phi = \{\Phi_x, \Phi_y\}$ et $f_k(\boldsymbol{y}_n)$ la vraisemblance $p(\boldsymbol{Y}_n = \boldsymbol{y}_n | X_n = \omega_k)$ des observations $\boldsymbol{y}_n$ conditionnellement à $X_n = \omega_k$. L'objectif consiste à trouver les étiquettes cachées connaissant les observations. X peut être retrouvé à partir du processus observé Y en utilisant différentes techniques de classification bayésienne comme le MPM [Gelman 05] avec l'algorithme de Baum et Welch [Devijver 85], ou le MAP avec l'algorithme de Viterbi [Fornay 73]. Dans le cas de la segmentation non-supervisée les paramètres Φ_x et Φ_y sont inconnus. Nous allons nous baser sur l'algorithme de Baum et Welch [Baum 72] pour estimer ces paramètres et segmenter la chaîne de Markov. Définissons pour cela les probabilités *forward* et *backward* [Baum 72]. Les probabilités *forward* sont définies par :

$$\alpha_n(k) = p(x_n = \omega_k, \boldsymbol{y}_{\leq n}) \tag{2.40}$$

avec $\boldsymbol{y}_{\leq n} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

Les probabilités *backward* sont définies par :

$$\beta_n(k) = p(\boldsymbol{y}_{>n} | \boldsymbol{y}_n, x_n = \omega_k) \tag{2.41}$$

où $\boldsymbol{y}_{>n} = (\boldsymbol{y}_{n+1}, \ldots, \boldsymbol{y}_N)$.

Néanmoins ces définitions posent un problème d'implémentation car on peut être confronté au phénomène de dépassement inférieur de capacité. Pour résoudre ce problème Devijver *et al.* proposent d'utiliser les probabilités conditionnelles à la place des probabilités conjointes [Devijver 85]. Les probabilités *forward* deviennent :

$$\alpha_n(k) = p(x_n = \omega_k | \boldsymbol{y}_{\leq n}) \tag{2.42}$$

et les probabilités *backward* :

$$\beta_n(k) = \frac{p(\boldsymbol{y}_{>n} | x_n = \omega_k)}{p(\boldsymbol{y}_{>n} | \boldsymbol{y}_{\leq n})} \tag{2.43}$$

L'intérêt de la segmentation d'images par chaînes de Markov vient du fait que ces probabilités peuvent être calculées récursivement. Les probabilités *forward* sont obtenues par un calcul récursif en parcourant la chaîne de $n = 1$ à $n = N$. Les probabilités *backward* sont obtenues par un parcours inverse de la chaîne.

Calcul de α

– pour $n = 1$

$$\alpha_1(k) = \pi_k f_k(y_1), \quad \forall k = 1, ..., K \tag{2.44}$$

– Pour $n > 1$

$$\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) t_{lk} f_k(y_n), \quad \forall k = 1, ...K \tag{2.45}$$

Calcul de β

– Pour $n = N$

$$\beta_N(k) = 1, \quad \forall k = 1, ..., K \tag{2.46}$$

– Pour $n < N$

$$\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) t_{kl} f_l(y_{n+1}), \quad \forall k = 1, ..., K \tag{2.47}$$

Les marginales *a posteriori* peuvent être directement calculées :

$$p(x_n = \omega_k | y) = \alpha_n(k)\beta_n(k), \quad \forall n \in [1, ..., N] \tag{2.48}$$

La maximisation de ces marginales permet de segmenter la chaîne au sens du critère du MPM :

$$\hat{x}_n \quad = \quad \arg_{\omega_k} \max p(x_n = \omega_k | y) = \arg_{\omega_k} \max \alpha_n(k)\beta_n(k) \tag{2.49}$$

2.5.2 Cas des chaînes de Markov couples et triplets

Soit $Z = (Z_1, \ldots, Z_N)$, avec $Z_n = (X_n, Y_n)$. Dans le cas où Z est une CMCouple, nous avons :

$$\begin{aligned} p(z_{n+1}|z_n) &= p((x_{n+1}, y_{n+1})|(x_n, y_n)) && (2.50)\\ &= p(x_{n+1}|x_n, y_n)p(y_{n+1}|x_{n+1}, x_n, y_n) && (2.51) \end{aligned}$$

Si les transitions d'une CMCouple vérifient $p(x_{n+1}|x_n, y_n) = p(x_{n+1}|x_n)$ et $p(y_{n+1}|x_{n+1}, x_n, y_n) = p(y_{n+1}|x_{n+1})$, on retrouve le modèle des CMC-BI classiques présenté au paragraphe 2.4.2. Dans les cas des CMCouples, on définit les probabilités *forward* et *backward* de la manière suivante :

$$\alpha_n(x_n) = p(y_1, \ldots, y_{n-1}|z_n) \tag{2.52}$$

$$\beta_n(x_n) = p(y_{n+1}, \ldots, y_N|z_n) \tag{2.53}$$

Elles sont calculables de manière récursive [Pieczynski 03b] :

$$\begin{aligned} \alpha_1(x_1) &= p(z_1) && (2.54)\\ \alpha_{n+1}(x_{n+1}) &= \sum_{x_n \in \Omega} \alpha_n(x_n) p(z_{n+1}|z_n) \quad \text{pour} \quad 2 \leq n \leq N-1 && (2.55)\\ \beta_N(x_N) &= 1 && (2.56)\\ \beta_n(x_n) &= \sum_{x_{n+1} \in \Omega} \beta_{n+1}(x_{n+1}) p(z_{n+1}|z_n) \quad \text{pour} \quad 1 \leq n \leq N-1 && (2.57) \end{aligned}$$

A partir de ces probabilités, on peut en déduire $p(x_n, y) = \alpha_n(x_n)\beta_n(x_n)$ ce qui donne $p(x_n|y)$.

Considérons maintenant une CMT $T = (X, U, Y)$, la chaîne $Z = (X, Y)$ n'est pas nécessairement markovienne. Soit V la chaîne constituée du couple (X, U), $V = (X, U)$, la chaîne $T = (V, Y)$ est une CMCouple. On peut ainsi appliquer les résultats précédents. Les lois $p(v_n|y) = p(x_n, u_n|y)$ sont ainsi calculables. Dans le cas où le cardinal de Λ n'est pas trop élevé, on peut en déduire facilement les marginales *a posteriori* $p(x_n|y)$ [Pieczynski 05] :

$$p(x_n|y) = \sum_{u_n \in \Lambda} p(x_n, u_n|y) \tag{2.58}$$

2.6 Estimation des paramètres du modèle CMC-BI

Dans ce paragraphe nous allons présenter l'estimation des paramètres du modèle CMC-BI que nous utiliserons dans la suite de ce manuscrit. L'estimation des paramètres peut se faire de manière stochastique (SEM (*Stochastic Expectation-Maximisation*)[Masson 93], ICE (*Iterative Conditional Estimation*) [Pieczynski 92]) ou déterministe (algorithme EM (*Expectation-Maximisation*) [Tanner 93]). Dans notre cas, les paramètres du modèle ont été estimés avec l'algorithme EM. Nous détaillerons donc seulement cette méthode d'estimation. L'EM est une méthode d'optimisation itérative pour estimer les paramètres au sens du maximum de vraisemblance. Au lieu d'utiliser seulement les données observées Y, les observations sont augmentées avec les variables cachées X pour simplifier les calculs. Les données observées sont ainsi considérées comme des données incomplètes, auxquelles on ajoute les données manquantes X pour aboutir aux données complètes (X, Y). Dempster *et al.* [Dempster 77] proposent un algorithme en deux étapes :

– Etape E (Estimation)

$$Q(\Theta, \Theta^{[q]}) = E[\log P(X, Y|\Theta)|Y, \Theta^{[q]}] \tag{2.59}$$

– Etape M (Maximisation)

$$\Theta^{[q+1]} = \arg\max_{\Theta} Q(\Theta, \Theta^{[q]}) \tag{2.60}$$

Dans le cas des Chaînes de Markov cachées, nous avons la relation suivante :

$$p(x, \boldsymbol{y}|\Phi) = p(x_1|\Phi_x) \prod_{n=2}^{N} p(x_n|x_{n-1}, \Phi_x) \prod_{n=1}^{N} p(\boldsymbol{y}_n|x_n, \Phi_y) \tag{2.61}$$

Nous pouvons donc écrire :

$$\begin{aligned} Q(\Phi, \Phi^{[q]}) &= E[\log p(x, \boldsymbol{y}|\Phi)|\boldsymbol{y}, \Phi^{[q]}] && (2.62)\\ &= \sum_{n=1}^{N} E[\log p(\boldsymbol{y}_n|x_n, \Phi_y)|\boldsymbol{y}, \Phi^{[q]}] + \sum_{n=2}^{N} E[\log p(x_n|x_{n-1}, \Phi_x)|\boldsymbol{y}, \Phi^{[q]}] \\ &\quad + E[\log p(x_1|\Phi_x)|\boldsymbol{y}, \Phi^{[q]}] && (2.63)\\ &= \sum_i \gamma_1^{[q]}(i) \log \pi_i + \sum_{n \neq 1} \sum_i \sum_j \xi_n^{[q]}(i, j) \log t_{ij} \\ &\quad + \sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(y_n) && (2.64) \end{aligned}$$

avec :

– les probabilités marginales *a posteriori*

$$\begin{aligned} \gamma_n^{[q]}(i) &= p(x_n = \omega_i|y, \Phi^{[q]}) && (2.65)\\ &= \frac{\alpha_n(i)\beta_n(i)}{\sum_j \alpha_n(j)\beta_n(j)} \end{aligned}$$

– les probabilités conjointes *a posteriori*

$$\begin{aligned} \xi_n^{[q]}(i, j) &= p(x_{n-1} = \omega_j, x_n = \omega_i|y, \Phi^{[q]}) && (2.66)\\ &= \alpha_{n-1}(j) t_{ji} f_i(y_n) \beta_n(i) \end{aligned}$$

– les probabilités initiales

$$\pi_i = p(x_1 = \omega_i) \tag{2.67}$$

– la matrice de transition

$$t_{ij}^n = p(x_{n+1} = \omega_j | x_n = \omega_i) \tag{2.68}$$

L'algorithme EM sur une chaîne de Markov a été développé par Baum [Baum 72]. On maximise Q sous les contraintes suivantes : $\sum_i \pi_i = 1$ et $\sum_j t_{ij} = 1$. En utilisant les multiplicateurs de Lagrange, on obtient [Provost 01] :

$$\pi_i^{[q+1]} = \gamma_1^{[q]}(i) \tag{2.69}$$

$$t_{ij}^{[q+1]} = \frac{\sum_{n=2}^{N} \xi_n^{[q]}}{\sum_{n=1}^{N-1} \gamma_n^{[q]}(i)} \tag{2.70}$$

$$\sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(y_n) = \sum_n \sum_i \gamma_n^{[q]}(i)[-\log \sigma_i - \frac{1}{2\sigma_i^2}(y_n - \mu_i)^2] \tag{2.71}$$

En dérivant $\sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(y_n)$ par rapport à μ_i (resp. σ_i) et en égalant le résultat à 0, on obtient :

$$\mu_i^{[q+1]} = \frac{\sum_n \gamma_n^{[q]}(i) y_n}{\sum_n \gamma_n^{[q]}(i)} \tag{2.72}$$

$$\sigma_i^{2[q+1]} = \frac{\sum_n \gamma_n^{[q]}(i)[y_n - \mu_i]^2}{\sum_n \gamma_n^{[q]}(i)} \tag{2.73}$$

L'algorithme 2 présente la segmentation MPM basée sur une chaîne de Markov cachée.

2.7 Conclusion

Dans ce chapitre, nous avons rappelé quelques éléments de l'inférence bayésienne et introduit les principes généraux des modèles graphiques. Nous avons ensuite présenté diverses modélisations markoviennes et leur intérêt en analyse d'images. Nous nous sommes ensuite intéressés aux récentes modélisations par chaînes de Markov triplet et au cas particulier des chaînes de Markov Couples. Nous avons plus particulièrement détaillé le modèle des chaînes de Markov cachées et l'estimation des paramètres par l'algorithme EM sur lesquels est basée notre méthode de segmentation d'IRM cérébrales présentée au chapitre suivant.

Algorithme 2 Segmentation MPM basée sur une chaîne de Markov cachée CMC-BI

1. Vectorisation de l'image en utilisant le parcours d'Hilbert-Peano 3D
2. Initialisation
   (a) Pour les paramètres *a priori* Φ_x

$$\pi_k = \frac{1}{K}, \forall k \tag{2.74}$$

$$t_{ij} = \begin{cases} \frac{3}{4} \text{ si } i = j \\ \frac{1}{4(K-1)} \text{ sinon} \end{cases} \tag{2.75}$$

   (b) Les paramètres de l'attache aux données Φ_y sont initialisés en utilisant les K-moyennes (Algo. 3 p.60).
3. Pour chaque itération
   (a) Calcul de $\alpha_n(k)$ et $\beta_n(k)$ en utilisant les équations 2.42 et 2.43.
   (b) Mise à jour des paramètres Φ en utilisant les équations 2.69, 2.70, 2.72 et 2.73.
4. Segmentation MPM en utilisant l'équation 2.49

$$\hat{x}_n = \arg\max_{\omega_k} P(X_n = \omega_k | Y = y, \Phi) = \arg\max_{\omega_k} \alpha_n(k)\beta_n(k)$$

5. Transformation inverse d'Hilbert-Peano pour obtenir l'image 3D segmentée

Chapitre 3

Modélisation proposée pour la segmentation en tissus

Dans ce chapitre, nous présentons l'approche proposée pour segmenter des IRM cérébrales multimodales. L'algorithme des chaînes de Markov cachées exposé au chapitre précédent permettant d'inclure l'information sur le voisinage dans le modèle a été modifié pour prendre en compte la correction du biais, les effets de volume partiel ainsi que de l'information *a priori* apportée par un atlas probabiliste. Nous présentons tout d'abord l'insertion de l'information *a priori* apportée par un atlas probabiliste (Section 3.1), puis la modélisation des hétérogénéités de l'intensité des images (Section 3.2), et la prise en compte des effets de volume partiel (Section 3.3). La chaîne de traitement complète est ensuite détaillée dans la section 3.4 avec les différents prétraitements à effectuer (extraction du cerveau, recalage de l'atlas et initialisation des paramètres). Cette chaîne de traitement qui présente l'originalité de combiner la prise en compte du voisinage, de la multimodalité, de l'atlas et la modélisation des inhomogénéités d'intensité et des effets de volume partiel sera validée à la fois sur des données synthétiques et réelles. Les résultats obtenus seront comparés avec d'autres méthodes de segmentation existantes.

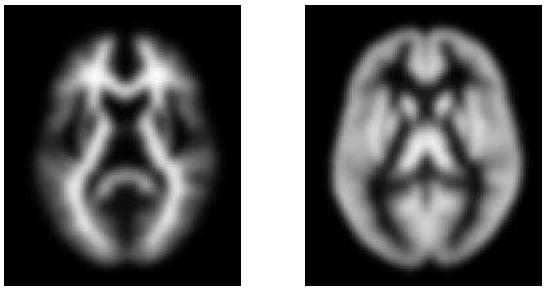
Nous avons choisi d'utiliser la modélisation par Chaînes de Markov Cachées (Section 2.4.2) pour prendre en compte la notion de voisinage. L'avantage de ce modèle comparé à celui des champs de Markov est qu'il requière des coûts calculatoires plus faibles et ne nécessite aucune approximation. En effet les méthodes par champs nécessitent des temps de calcul importants, en particulier pour traiter des données 3D. Les réalisations *a posteriori* de X ne peuvent être créées directement et doivent donc être approximées en utilisant l'algorithme de Metropolis [Metropolis 53] ou l'échantillonneur de Gibbs [Geman 84]. Pour surmonter ce problème, l'estimation des paramètres est souvent basée sur un nombre limité de voxels [Van Leemput 99b, Ruan 02] perdant ainsi une partie de l'information sur les données observées.

3.1 Prise en compte d'information *a priori* apportée par un atlas probabiliste

L'information apportée par un atlas probabiliste permet d'améliorer la précision de la segmentation. Pour cette raison, de l'information *a priori* a déjà été introduite dans certaines

méthodes de segmentation [Al-Zubi 02, Ashburner 05]. Ashburner et Friston proposent par exemple un algorithme basé sur un modèle de mélange de gaussiennes, modifié permettant de prendre en compte des cartes de probabilités [Ashburner 97, Ashburner 00]. Cependant leur méthode ne prend pas en compte l'information sur le voisinage, ce qui dégrade la qualité de la segmentation en cas de bruit important sur les images. D'autres méthodes utilisent un atlas cérébral pour initialiser un processus itératif de segmentation [Van Leemput 99b, Marroquin 02]. Dans chaque cas l'idée principale reste la même : l'atlas est un outil intéressant pour guider le processus de segmentation. Marroquin *et al.* utilisent un atlas pour séparer le cerveau des tissus non-cérébraux et pour calculer les probabilités *a priori* de chaque classe en chaque voxel [Marroquin 02]. Dans certains cas, l'atlas peut aussi permettre de corriger les voxels mal-classés [Al-Zubi 02].

L'approche que nous proposons utilise un atlas apportant de l'information *a priori* comme capteur supplémentaire. L'atlas que nous avons utilisé est l'atlas de SPM d'Ashburner et Friston. Cet atlas est obtenu à partir d'IRM cérébrales de 152 sujets segmentées en images binaires de matière blanche MB, matière grise MG et liquide céphalo-rachidien LCR et normalisées dans le même espace grâce à une transformation affine [Ashburner 97]. Les cartes de probabilités correspondent aux moyennes de ces images binaires, et contiennent ainsi des valeurs comprises entre 0 et 1. Ainsi l'atlas fournit des cartes de probabilités des trois principaux tissus du cerveau (MB, MG et LCR). Un exemple d'atlas utilisé est présenté sur la figure 3.1. Nous allons noter $P(B_n|X_n = \omega_k)$ la probabilité du voxel n d'appartenir à la classe k donnée par l'atlas.



(a) (b)

FIG. 3.1 – *Exemple d'atlas utilisé : (a) Atlas de la matière blanche. (b) Atlas de la matière grise.*

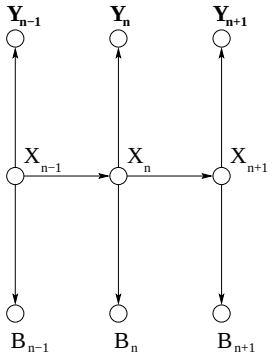


FIG. 3.2 – *Chaîne de Markov Cachée X avec l'information apportée par l'atlas probabiliste et les observations $\boldsymbol{Y}$.*

Le modèle que nous proposons est une chaîne de Markov cachée avec deux capteurs indépendants. Dans ce cas, nous avons comme information à la fois l'intensité des données observées et les cartes de probabilités. Ainsi, nous cherchons :

$$\hat{x}_n = \arg\max_{\omega_k} p(x_n = \omega_k | \boldsymbol{y}, b) \tag{3.1}$$

où $p(x_n = \omega_k|\boldsymbol{y}, b) = \alpha_n(k)\beta_n(k)$ avec $\alpha_n(k)$ probabilité *forward* :

$$\alpha_n(k) = p(x_n = \omega_k | \boldsymbol{y_{\leq n}}, b_{\leq n}) \tag{3.2}$$

et $\beta_n(k)$ probabilité *backward* :

$$\beta_n(k) = \frac{p(\boldsymbol{y_{>n}}, b_{>n} | x_n = \omega_k)}{p(\boldsymbol{y_{>n}}, b_{>n} | \boldsymbol{y_{\leq n}}, b_{\leq n})} \tag{3.3}$$

Le calcul des probabilités *forward* et *backward* est détaillé dans l'annexe A.

Avec ce lien à l'atlas, nous prenons en compte à la fois la connaissance *a priori* apportée par les cartes de probabilités de chaque tissu et l'organisation spatiale dans le modèle de segmentation par chaînes de Markov cachées. Les paramètres de la loi d'attache aux données vont ainsi dépendre d'une part des intensités observées et d'autre part des cartes de probabilité des différents tissus cérébraux. L'utilisation de l'atlas va donc permettre de rendre le processus de segmentation plus robuste, l'information probabiliste apportée par l'atlas étant modélisée comme un capteur supplémentaire et gérée comme tel durant le processus de segmentation.

3.2 Prise en compte des hétérogénéités de l'intensité dans les images

Les hétérogénéités d'intensité dues aux inhomogénéités du champ RF sont modélisées par un biais multiplicatif. Grâce à une transformation logarithmique des intensités, l'artefact peut être modélisé par un biais additif [Van Leemput 99a] : $Z_n = \log Y_n$, où Y_n est l'intensité observée au voxel n.

D'une part, nous modélisons le biais par une combinaison linéaire $\sum_k c_k \phi_k(x)$ de fonctions de base lisses $\phi_k(x)$, comme Van Leemput *et al.* dans [Van Leemput 99a]. D'autre part, chaque classe est modélisée par une distribution normale. Nous avons donc :

$$p(z_n | x_n = \omega_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \frac{(z_n - \mu_i - \sum_k c_k \phi_k(x_n))^2}{2\sigma_i^2} \tag{3.4}$$

Une telle approximation est réalisée dans [Wells 96, Guillemaud 97, Van Leemput 99a]. L'algorithme présenté dans la section précédente est donc légèrement modifié pour introduire cette étape de correction de biais. Nous avons choisi des fonctions de base polynômiales pour $\phi_k(x)$. Etant donné que la vraisemblance n'est plus maximisée, mais seulement augmentée à chaque itération, l'algorithme est un algorithme EM généralisé (GEM). L'estimation des paramètres du biais est détaillée dans l'annexe B.

3.3 Modélisation des effets de volume partiel

Au lieu d'attribuer une seule et unique classe à chaque voxel, nous allons maintenant estimer le pourcentage de chaque tissu en chaque voxel. En effet en raison de la résolution du système d'acquisition, les voxels situés à la frontière entre différentes classes sont composés de plusieurs tissus, créant ainsi des effets de volume partiel (Section 1.4). Les méthodes de classification en "dur" attribuant à chaque voxel la classe la plus représentative perdent ainsi de l'information sur le contenu des tissus. La méthode que nous proposons estime la fraction de chaque tissu pur en chaque voxel en adaptant le modèle par chaînes de Markov cachées présenté au chapitre précédent pour prendre en compte les effets de volume partiel. Cette méthode intervient en post-traitement de la méthode présentée au paragraphe précédent. Durant cette étape, nous n'utilisons pas l'atlas et les paramètres du biais sont fixes.

Notons $\boldsymbol{A_n}$ le vecteur contenant la proportion de chaque tissu en chaque voxel n.

$$\boldsymbol{A_n} = \begin{bmatrix} a_n^1 \\ ... \\ a_n^{K-1} \end{bmatrix} \tag{3.5}$$

où $\forall n, \forall k, a_n^k$ représente la proportion du tissu k dans le voxel n et prend ses valeurs dans $[0,1]$. On peut en déduire les proportions de la classe K : $a_n^K = 1 - \sum_{k=1}^{K-1} a_n^k$. Ainsi a_n^k peut prendre une infinité de valeurs dans $[0,1]$, mais en pratique a_n^k est discretisé en utilisant un pas de 0.1 ou 0.2, ce qui signifie que a_n^k prend ses valeurs dans $\{0, 0.1, 0.2, 0.3, ..., 0.9, 1\}$ (respectivement $\{0, 0.2, ..., 0.8, 1\}$).

Nous représentons l'intensité y d'un voxel par une somme pondérée de K classes pures :

$$y = \sum_{i=1}^{K} a^i y_i^p \tag{3.6}$$

où $\sum_{i=1}^{K} a^i = 1$ et $a^i > 0$. a^i représente la proportion de la classe pure i et y_i^p est une variable aléatoire représentant la classe pure i. Nous supposons que chaque classe pure suit une loi normale $y_i^p \sim \mathcal{N}(\mu_i, \sigma_i^2)$ et que les Y_i^p sont indépendants. D'où, pour un $\boldsymbol{A}$ donné [Stark 86] :

$$\boldsymbol{Y}|\boldsymbol{A} \sim \mathcal{N}(\sum_{i=1}^{K} a^i \mu_i, \sum_{i=1}^{K} (a^i)^2 \sigma_i^2) \tag{3.7}$$

$\boldsymbol{A}$ étant une chaîne de Markov, on peut appliquer les propriétés des chaînes de Markov présentées au chapitre 2.6. La loi de $\boldsymbol{A}$ est donc entièrement déterminée par la connaissance des probabilités initiales et de la matrice de transition d'une classe à une autre. La figure 3.3 présente le graphe de dépendance d'une chaîne de Markov avec effet de volume partiel.

Pour obtenir une segmentation de l'image, on utilise le MPM [Gelman 05] :

$$\hat{\boldsymbol{A}}_{\boldsymbol{n}} = \arg\max_{\boldsymbol{A_n}} p(\boldsymbol{A_n}|\boldsymbol{Y}) = \alpha_n \beta_n \tag{3.8}$$

avec α_n probabilité *forward* :

$$\alpha_n = p(\boldsymbol{A_n}|\boldsymbol{Y_{\leq n}}) \tag{3.9}$$

et β_n probabilité *backward* :

$$\beta_n = \frac{p(\boldsymbol{Y_{>n}}|\boldsymbol{A_n})}{p(\boldsymbol{Y_{>n}}|\boldsymbol{Y_{\leq n}})} \tag{3.10}$$

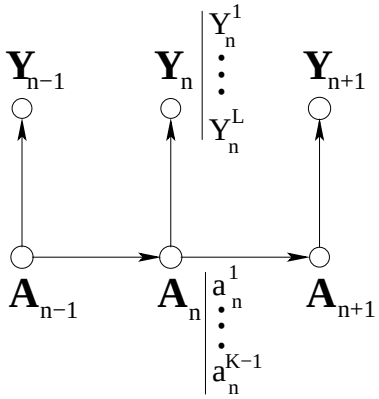


FIG. 3.3 – *Chaîne de Markov Cachée multimodale avec prise en compte de l'effet de volume partiel.*

Pour chaque voxel, on calcule la probabilité de chaque label et on garde le label ayant la plus grande probabilité.

Les probabilités *forward* et *backward* se calculent de manière récursive :

– probabilité *forward*

$$\alpha_1 = p(\boldsymbol{A_1})p(\boldsymbol{Y_1}|\boldsymbol{A_1}) \tag{3.11}$$

$$\alpha_n = p(\boldsymbol{A_n}, \boldsymbol{Y_{\leq n}}), \quad \forall n > 1 \tag{3.12}$$

$$= \sum_{\boldsymbol{A_{n-1}}} \alpha_{n-1} p(\boldsymbol{A_n}|\boldsymbol{A_{n-1}})p(\boldsymbol{Y_n}|\boldsymbol{A_n}) \tag{3.13}$$

– probabilité *backward*

$$\beta_N = 1 \tag{3.14}$$

$$\beta_n = \sum_{\boldsymbol{A_{n+1}}} \beta_{n+1} p(\boldsymbol{A_{n+1}}|\boldsymbol{A_{n+1}})p(\boldsymbol{Y_{n+1}}|\boldsymbol{A_{n+1}}), \quad \forall n < N \tag{3.15}$$

On obtient ainsi K cartes de segmentation contenant la proportion de chaque tissu en chaque voxel.

Remarque : Cette modélisation des effets de volume partiel se rapproche de celle proposée par Ruan dans le cadre des champs de Markov [Ruan 02].

3.4 Chaîne de traitement complète

La méthode de segmentation proposée nécessite un certain nombre de prétraitements avant d'être appliquée. En effet, les différentes modalités que nous étudions doivent être préalablement recalées entre elles, ainsi qu'avec l'atlas. D'autre part, nous focalisons notre étude sur les tissus cérébraux et nous devons donc extraire le cerveau avant d'effectuer la segmentation. Ces différents prétraitements sont détaillés dans les paragraphes suivants.

3.4.1 Prétraitements

3.4.1.1 Recalage

Avant d'étudier deux volumes correspondant au même objet physique, il faut que les deux volumes soient dans le même espace. Le recalage est donc une étape indispensable. Notre algorithme étant applicable à des données multimodales (T1, T2, Flair ...), il faut donc que ces différentes modalités soient placées dans le même espace. D'autre part, l'algorithme développé au paragraphe précédent nécessite l'utilisation d'un atlas probabiliste. Pour utiliser correctement les informations fournies par cet atlas, il est nécessaire de le placer dans le même repère que les images du patient. Le recalage que nous avons utilisé est un recalage non rigide préservant la topologie des structures anatomiques, issu des travaux présentés dans [Noblet 05, Noblet 06].

3.4.1.2 Extraction du cerveau

Les IRM cérébrales contiennent l'image de la tête entière, alors que nous allons nous focaliser sur l'étude des tissus cérébraux seulement. Il faut donc préalablement extraire le cerveau. L'extraction du parenchyme cérébral (matière blanche, matière grise et liquide céphalo-rachidien) est un problème difficile sur les IRM cérébrales, en particulier pour les volumes pondérés en T1. Cette étape est cependant nécessaire avant de pouvoir effectuer la segmentation des différents tissus cérébraux. Les difficultés de l'extraction du cerveau sont principalement dues aux intensités qui peuvent être similaires entre les structures corticales et non corticales et parce que ces régions (par exemple les yeux) apparaissent souvent connectées sur l'IRM.

L'algorithme Brain Extraction Tool (BET) [Smith 02] effectue une estimation basée sur l'intensité du seuil entre les tissus cérébraux et non-cérébraux. Puis l'algorithme détermine le centre de gravité de la tête et définit une sphère initiale basée sur ce centre de gravité. Enfin, la sphère est déformée vers l'extérieur jusqu'à atteindre la surface du cerveau. Le principe de l'algorithme BET est illustré sur la figure 3.4.

3.4.1.3 Initialisation des paramètres

Les algorithmes d'estimation utilisés doivent être initialisés. Cette étape est importante car elle peut influer sur la convergence de l'algorithme EM et conditionne la vitesse de convergence. Nous avons utilisé l'algorithme des K-Moyennes [Bovik 00] basé sur une méthode de clustering. Il s'agit d'un algorithme de quantification vectorielle générant son dictionnaire automatiquement à partir des observations. Le principe consiste à découper l'image en M imagettes sur lesquelles on calcule les paramètres statistiques moyennes et écarts-types de chacune des C modalités. La $m^{\text{ième}}$ imagette sera associée au vecteur $x_m = \{\mu_m^{(c)}, \sigma_m^{(c)}, \forall c = 1, \ldots, C\}$. La quantification vectorielle consiste à associer à chaque vecteur x_m un élément choisi parmi un ensemble de K vecteurs appelé *dictionnaire* ou *alphabet*. Elle réalise une partition de l'espace des observations en raison de la taille réduite de son dictionnaire. L'algorithme des K-Moyennes permet d'établir automatiquement ce dictionnaire. Chaque élément k de l'alphabet sera représenté par le centroïde C_k associé aux paramètres $c_k = \{\mu_m^{(c)}, \sigma_m^{(c)}, \forall c = 1, \ldots, C\}$. Le rattachement d'une imagette m à un centroïde C_k se fait à l'aide d'une mesure de distorsion. Dans le cas de la distance euclidienne, l'algorithme des

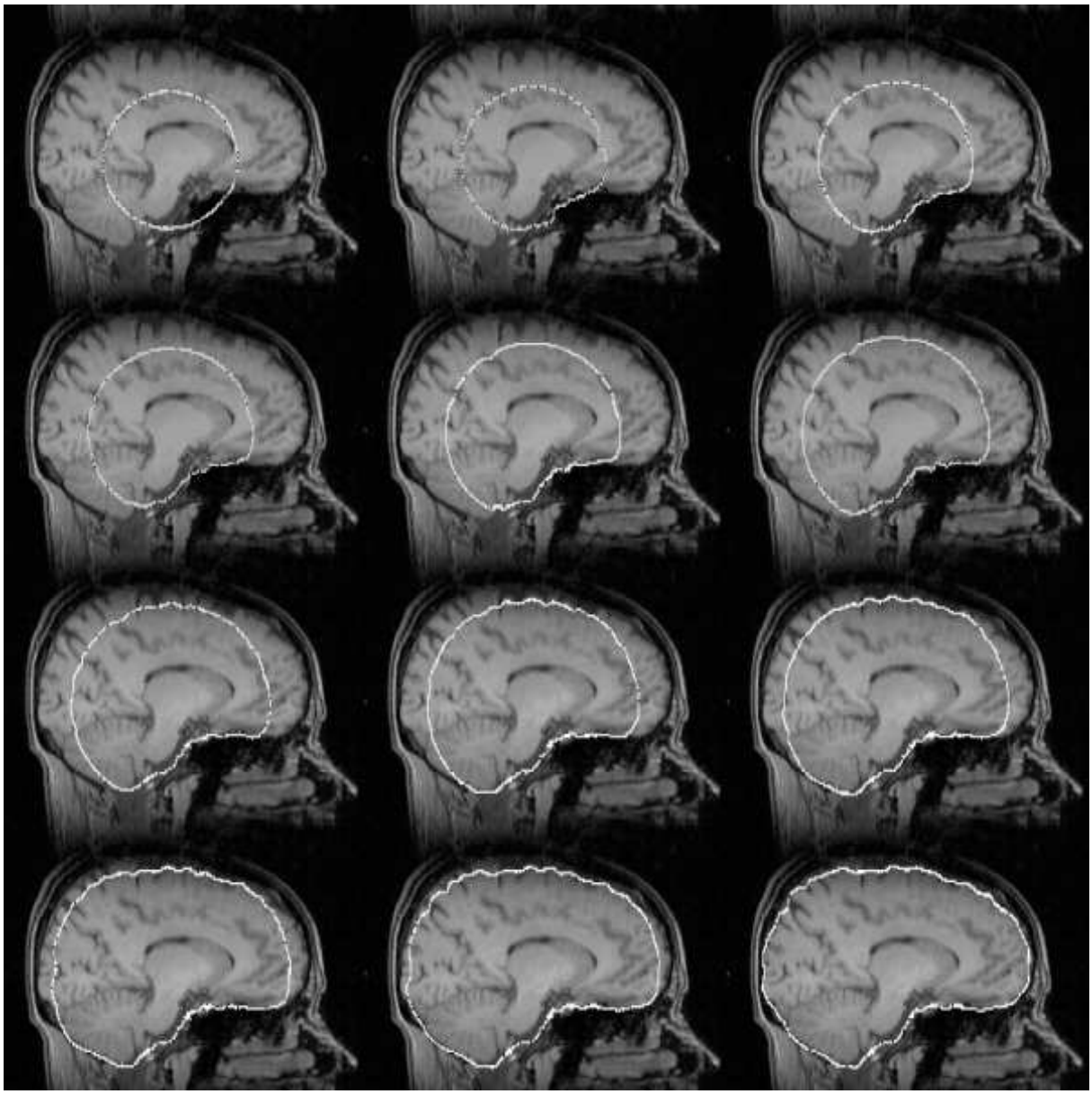


FIG. 3.4 – *Algorithme Brain Extraction Tool (BET). Illustration tirée de [Smith 02].*

K-Moyennes consiste à minimiser le critère :

$$E = \sum_{k=1}^{K} \sum_{x_m \in C_k} \|x_m - c_k\|^2 \tag{3.16}$$

Pour une partition $\{C_k, k \in \{1, \ldots, K\}\}$ donnée, l'indice de dispersion E est minimal si le centre de chaque groupement C_k est donné par la relation suivante :

$$c_k = \frac{1}{card(C_k)} \sum_{m \in C_k} x_m \tag{3.17}$$

L'algorithme des K-Moyennes est détaillé dans l'algorithme 3.

Algorithme 3 Algorithme des K-Moyennes

1. Initialisation des K centroïdes à partir des K premières imagettes

$$c_k^{[0]} = x_k, \quad k \in \{1, \ldots, K\} \tag{3.18}$$

2. Faire
   - $q \leftarrow q + 1$
   - Pour $m \in \{1, \ldots, M\}$, on associe à l'imagette m le centroïde l le plus proche au sens de la distance euclidienne.

$$C_k = \{m, \|x_m - c_k\| < \|x_m - c_l\|, \forall l \neq k\} \tag{3.19}$$

   - Mise à jour des paramètres des centroïdes

$$c_k^{[q]} = \frac{1}{card(C_k)} \sum_{m \in C_k} x_m, \quad \forall k \tag{3.20}$$

   Tant que $\|c^{[q]} - c^{[q-1]}\| < \epsilon$ où ϵ est un seuil empirique.

Remarque : L'initialisation des paramètres de la loi d'attache aux données pourrait aussi être effectuée en utilisant l'information apportée par l'atlas probabiliste.

3.4.2 Chaîne de traitement

Les effets de volume partiel n'affectent pas la correction du biais. Ainsi pour accélérer l'algorithme, nous allons d'abord estimer le biais avant de prendre en compte les effets de volume partiel (Fig. 3.5). Nous commençons par appliquer l'algorithme HMC avec la correction du biais en utilisant l'information *a priori* contenue dans les cartes de probabilité. On obtient ainsi une estimation du biais ainsi que de la moyenne μ et de la variance σ^2 de chaque classe pure. A cette étape, l'algorithme peut aussi fournir une segmentation en trois classes dures.

Pour réduire la complexité numérique, nous considérons qu'un voxel ne peut être composé que de deux différents types de tissu en même temps. La probabilité d'avoir un mélange de plus de deux tissus est en effet très faible. De plus, la probabilité d'avoir un mélange de matière blanche et de

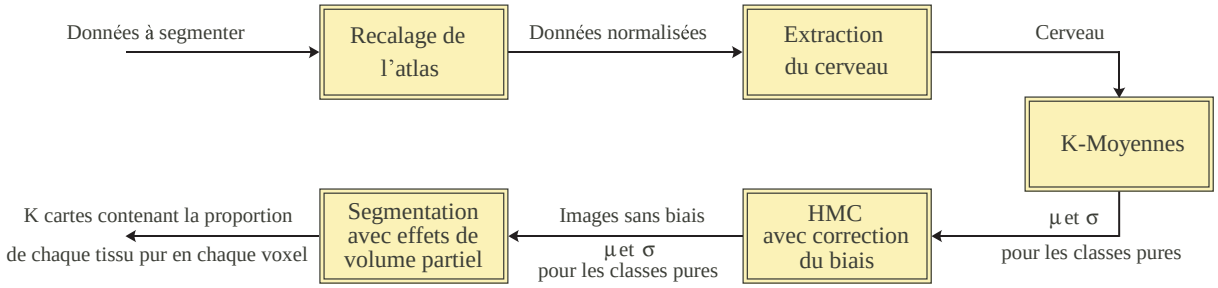


FIG. 3.5 – *Chaîne de traitement proposée pour la segmentation d'IRM cérébrales multimodales.*

liquide céphalo-rachidien est faible [Ruan 00, Shattuck 01, Ruan 02], nous n'autorisons donc que les mélanges MB/MG et MG/LCR. La dernière étape de l'algorithme consiste à appliquer l'algorithme HMC avec prise en compte des effets de volume partiel pour obtenir trois cartes contenant les proportions de chaque tissu pur en chaque voxel.

L'algorithme 4 et la figure 3.5 présentent la chaîne de traitement complète proposée pour segmenter des images IRM multimodales.

3.5 Conclusions et Limites

Nous avons proposé une méthode de segmentation d'IRM cérébrales multimodales basée sur les chaînes de Markov cachées prenant en compte l'information sur le voisinage [Bricq 08d]. Ce modèle prend en considération les hétérogénéités de l'intensité ainsi que les effets de volume partiel et permet d'obtenir les proportions de chaque tissu en chaque voxel. De plus il inclut de l'information *a priori* apportée par un atlas probabiliste. Par conséquent, le résultat de la segmentation pourra dépendre de la qualité de l'atlas utilisé. D'autre part cette méthode ne prend pas en compte la présence éventuelle de données atypiques liées à une pathologie, comme par exemple des lésions de sclérose en plaques ou des tumeurs. L'extension de ce modèle à des données pathologiques (en particulier pour la détection de lésions de sclérose en plaques) sera développée au chapitre 4. La validation de cette méthode de segmentation sur des données non pathologiques est présentée dans les paragraphes suivants à la fois sur une base d'images synthétiques et sur une base d'images réelles.

Nous commençons par présenter les bases de données et les outils utilisés pour la validation. Puis nous comparons notre méthode avec d'autres méthodes de segmentation classiques.

3.6 Données

3.6.1 Base Brainweb

La vérité terrain n'étant pas connue pour des images réelles, nous avons d'abord validé notre algorithme sur des images simulées en utilisant le fantôme Brainweb. Il s'agit d'un outil de validation très utilisé dans le cadre de la segmentation cérébrale [Kwan 96, Cocosco 97].
Le site Web de Brainweb [1] permet de simuler des IRM cérébrales avec différents niveaux de bruit et

[1] `http://www.bic.mni.mcgill.ca/brainweb/`

Algorithme 4 Segmentation d'IRM multimodales par chaîne de Markov cachée

1. Recalage non-rigide des images avec l'atlas [Noblet 05]
2. Extraction du cerveau en utilisant BET [Smith 02]
3. Vectorisation de l'image en utilisant le parcours d'Hilbert-Peano 3D [Bandoh 99]
4. Initialisation
   (a) Pour les paramètres *a priori* Φ_x

$$\begin{aligned} \pi_k &= \frac{1}{K}, \forall k \\ t_{ij} &= \begin{cases} \frac{3}{4} \text{ si } i = j \\ \frac{1}{4(K-1)} \text{ sinon} \end{cases} \end{aligned}$$

   (b) Initialisation des paramètres de la loi d'attache aux données Φ_y par les K-Moyennes [Bovik 00]
5. Algorithme HMC avec correction du biais (sur le logarithme des données), avec l'information apportée par l'atlas probabiliste. Pour chaque itération,
   (a) Calcul de $\alpha_n(k)$ et $\beta_n(k)$ avec Eq. A.6 et A.8.
   (b) Mise à jour des paramètres Φ avec Eq. B.7, B.8, B.10 et B.11.
6. Algorithme HMC avec estimation de l'effet de volume partiel sans atlas et en gardant les paramètres du biais fixés (estimés pendant l'étape précédente).
7. Segmentation MPM en utilisant Eq. 2.49

$$\hat{x}_n = \arg\max_{\omega_k} P(X_n = \omega_k | Y = y, \Phi) = \arg\max_{\omega_k} \alpha_n(k)\beta_n(k)$$

8. Transformation d'Hilbert-Peano inverse pour obtenir l'image segmentée

TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
GT	Vérité terrain
SEG	Segmentation obtenue

TAB. 3.1 – *Abréviations utilisées pour la validation.*

d'inhomogénéités. La réalité terrain étant connue, ce fantôme permet de tester différents algorithmes de segmentation. Les images de synthèse ont été créées à partir de 27 acquisitions réalisées sur un même individu. Ces acquisitions sont ensuite moyennées pour créer le volume initial utilisé pour la construction du fantôme. Un certain nombre de traitements sont appliqués à ce volume pour créer le fantôme : correction des inhomogénéités, classification, corrections manuelles et simulation des niveaux de bruit pour chaque voxel. La construction de ce fantôme est détaillée dans [Collins 98]. Les images Brainweb utilisées sont de taille $181 \times 181 \times 217$ avec des voxels isotropes de 1 mm d'épaisseur.

3.6.2 Base IBSR

Pour valider notre algorithme sur des images réelles, nous avons utilisé la base IBSR 2 (Internet Brain Segmentation Repository) disponible sur le web [2] et fournie par le Center for Morphometric Analysis du Massachusetts General Hospital. Cette base est composée de 18 volumes IRM pondérés en $T1$ acquis sur des patients sains. Ces images sont constituées de 128 coupes de taille 256×256. Cette base dispose d'une segmentation manuelle de la matière blanche, matière grise et du liquide céphalo-rachidien interne, réalisée par des experts.

3.6.3 Outils de validation

Pour tester notre algorithme de façon pertinente, nous avons jugé la qualité de la segmentation obtenue par rapport à plusieurs estimateurs souvent utilisés dans la littérature [Richard 04] :

– Sensibilité :
Elle correspond à la True Positive Fraction (TPF).

$$TPF = \frac{TP}{TP + FN}$$

– Spécificité :
Elle correspond à 1-False Positive Fraction (FPF).

$$specificity = \frac{TN}{TN + FP}$$

[2] `http://www.cma.mgh.harvard.edu/ibsr/`

– Overlap (ou recouvrement) :
Il correspond aussi au coefficient de Jaccard.

$$overlap = OL = \frac{TP}{TP + FP + FN}$$

– Similarité :
Elle correspond aussi au Kappa Index (ou coefficient de Dice) et mesure le recouvrement entre la segmentation obtenue et la vérité terrain.

$$KappaIndex = KI = 2\frac{SEG \cap GT}{SEG + GT} = 2\frac{TP}{2TP + FP + FN}$$

Elle vaut 1 lorsque les deux segmentations sont identiques et 0 lorsqu'il n'y a aucun recouvrement. On peut constater que le coefficient de Jaccard et l'index Kappa sont liés par la relation suivante :

$$KI = 2\frac{OL}{OL + 1}$$

Dans la suite du manuscrit seul l'index Kappa sera utilisé pour la validation.

Ces estimateurs sont utilisés dans le cas d'une segmentation en classes "dures", c'est-à-dire dans le cas où une seule classe est attribuée à chaque voxel.

Pour évaluer les résultats de la segmentation "floue", on mesure l'erreur quadratique moyenne RMS (Root Mean Square)[Shattuck 01, Tohka 04] :

$$e_{RMS}(k) = \sqrt{\frac{1}{\Omega}\sum_{n \in S} |a_n(k) - a_n^f(k)|^2} \tag{3.21}$$

où Ω correspond à la zone du cerveau, $a_n(k)$ représente la proportion du tissu k au voxel n pour le volume segmenté, et $a_n^f(k)$ dénote la même proportion dans le fantôme. $Card(S) = N$ représente le nombre de voxels dans le volume. Plus ce critère est proche de 0, meilleurs sont les résultats. Ce critère est utilisé pour une segmentation floue.

Dans la suite du manuscrit, les différents estimateurs seront donnés en pourcentage. Dans un premier temps, nous testerons l'impact du parcours d'Hilbert-Peano sur le résultat de la segmentation (section 3.7), puis nous validerons notre approche à la fois sur des données mono- et multimodales et nous comparerons nos résultats avec d'autres méthodes utilisées (sections 3.8 et 3.9).

3.6.4 Comparaison avec des méthodes existantes

Pour évaluer les performances de notre algorithme, nous l'avons comparé à trois autres méthodes classiques de segmentation : SPM5[3] (Statistical Parametric Mapping) [Ashburner 05], EMS[4] (Expectation-Maximization Segmentation) [Van Leemput 99a] et avec la méthode proposée par Shattuck *et al.* [Shattuck 01]. Les différentes méthodes sont comparées dans le tableau 3.2. La méthode d'Ashburner et Friston [Ashburner 05] (SPM5) inclue la classification des tissus, la correction du biais et le recalage des images avec les cartes de probabilités dans le même modèle.

[3] `http://www.fil.ion.ucl.ac.uk/spm`
[4] `http://www.medicalimagecomputing.com/EMS/`

TAB. 3.2 – *Comparaison des différentes méthodes de segmentation.*

	SPM5	EMS	Shattuck	HMC
Biais	Oui	Oui	Oui	Oui
Effets de volume partiel	Oui[1]	Non[2]	Oui	Oui
Voisinage	Non	Oui[3] (MRF)	Oui (MRF)	Oui (HMC)
Non supervisé	Oui	Oui	Partiellement[4]	Oui
Atlas	Oui	Oui	Non	Oui
Multimodal	Non	Oui	Non	Oui
Temps de calcul	20 à 35'	20 à 35'	5'	20 à 40'

[1] : les proportions de chaque tissu en chaque voxel ne sont pas calculées.
[2] : pas implémenté dans EMS mais présenté dans [Van Leemput 03].
[3] : l'estimation des paramètres est basée seulement sur un nombre limité de voxels.
[4] : le terme β est fixé.

Van Leemput *et al.* proposent une méthode incluant la correction du biais et la classification des tissus dans un modèle MRF (Markov Random Field) [Van Leemput 99a]. Néanmoins la méthode présentée dans [Van Leemput 03] pour prendre en compte le volume partiel n'est pas implémentée dans EMS. Shattuck *et al.* présentent une méthode prenant en compte l'information sur le voisinage en utilisant un *a priori* markovien [Shattuck 01]. Leur méthode compense les inhomogénéités des images et combine l'estimation des fractions des tissus purs avec un modèle de Gibbs.

3.7 Impact du parcours d'Hilbert-Peano sur les résultats de segmentation

La vectorisation de l'image dépend du parcours de l'image effectué. Le parcours d'Hilbert-Peano est le parcours le mieux adapté pour la segmentation d'images : c'est un parcours fractal qui préserve au mieux la notion de voisinage. Cependant, suivant le point de départ du parcours, le vecteur obtenu sera différent. Nous avons donc testé l'impact de ce parcours sur la segmentation finale. Pour cela, nous avons utilisé le parcours d'Hilbert-Peano pour vectoriser l'image en partant de 3 sommets différents du cube (correspondant respectivement aux parcours 1, 2 et 3). Nous avons ensuite effectué une segmentation HMC du vecteur obtenu (sans correction de biais, ni prise en compte des effets de volume partiel), puis réalisé la transformation inverse d'Hilbert-Peano pour obtenir l'image segmentée. Les résultats obtenus sur une image T1 avec 0% de biais et 3% (resp. 5%) de bruit sont présentés dans le tableau 3.3 (resp. 3.4). Pour chaque parcours, le nombre d'itérations nécessaires avant convergence, le taux de voxels mal classés τ_{err} et l'estimation de la moyenne μ et de l'écart-type σ de chaque classe sont donnés.

On constate que quel que soit le parcours effectué, les résultats obtenus sont très proches : le nombre d'itérations nécessaires avant convergence est le même, les moyennes et écart-types pour chaque classe ainsi que le taux de voxels mal classés sont du même ordre de grandeur.

Méthode	Nb iter	MB		MG		LCR		τ_{err}
		μ	σ	μ	σ	μ	σ	
HMC, Parcours 1	18	135.68	6.37	104.16	9.73	55.48	16.87	7.76 %
HMC, Parcours 2	18	135.70	6.35	104.21	9.71	55.62	16.97	7.87 %
HMC, Parcours 3	18	135.68	6.37	104.21	9.68	55.66	16.99	7.86 %

TAB. 3.3 – *Résultats obtenus sur une image T1 avec 3% de bruit et 0% d'inhomogénéité.*

Méthode	Nb iter	MB		MG		LCR		τ_{err}
		μ	σ	μ	σ	μ	σ	
HMC, Parcours 1	16	129.87	8.54	99.25	10.56	52.16	15.85	7.75 %
HMC, Parcours 2	16	129.89	8.52	99.32	10.53	52.37	16.01	7.86 %
HMC, Parcours 3	16	129.86	8.53	99.29	10.54	52.30	15.95	7.81 %

TAB. 3.4 – *Résultats obtenus sur une image T1 avec 5% de bruit et 0% d'inhomogénéité.*

3.8 Validation sur la base Brainweb

3.8.1 Images monomodales

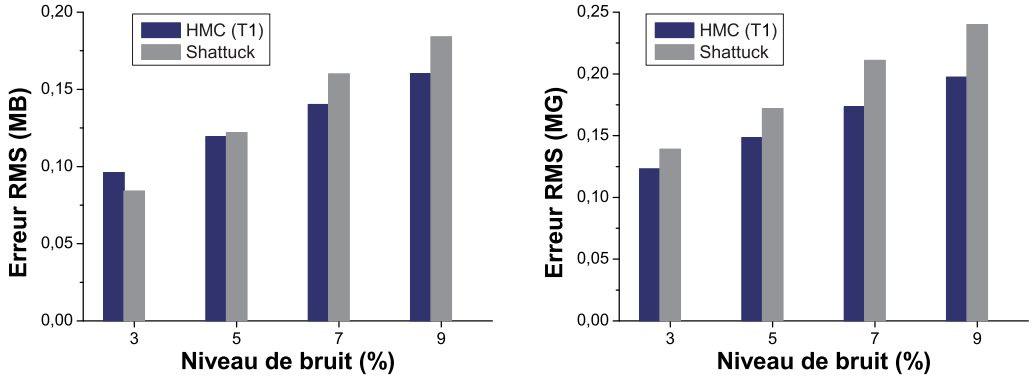
3.8.1.1 Segmentation "floue"

Nous testons notre algorithme sur les images T1 de la base Brainweb avec 20% d'inhomogénéité et différents niveaux de bruit. Un exemple des cartes de matière blanche, matière grise et liquide céphalo-rachidien obtenues est présenté sur la figure 3.6. Chaque carte contient les proportions de chaque tissu.

L'erreur RMS obtenue par notre méthode à la fois pour la matière blanche et la matière grise sur des images dont le niveau de bruit varie de 3% à 9% a été comparée à celle présentée par Shattuck [Shattuck 01] (Fig. 3.7). Excepté pour la matière blanche avec 3% de bruit, la méthode que nous proposons obtient les meilleurs résultats à la fois pour la matière blanche et la matière grise. On peut remarquer que l'erreur RMS augmente avec le niveau de bruit. Dans le meilleur des cas, la méthode basée sur les chaînes de Markov cachées fournit une erreur RMS 12.9% meilleure

(a) (b) (c)

FIG. 3.6 – *Cartes de segmentation obtenues sur une image Brainweb avec 5% de bruit et 20% d'inhomogénéité. (a), (b) et (c) correspondent respectivement aux cartes de matière blanche, matière grise et liquide céphalo-rachidien. L'erreur RMS obtenue pour cette image est présentée sur la figure 3.7.*



(a) (b)

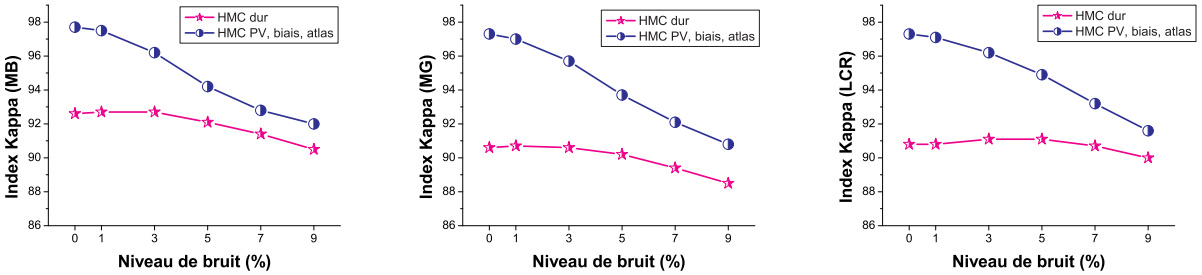
FIG. 3.7 – *Erreur RMS obtenue sur une image T1 de la base Brainweb avec 20% d'inhomogénéité et différents niveaux de bruit avec notre méthode et la méthode développée par Shattuck [Shattuck 01]. (a), (b) correspondent respectivement aux résultats obtenus pour la matière blanche (MB) et la matière grise (MG).*

pour la matière blanche et 17.8% meilleure pour la matière grise par rapport aux résultats donnés par Shattuck [Shattuck 01].

3.8.1.2 Segmentation "dure"

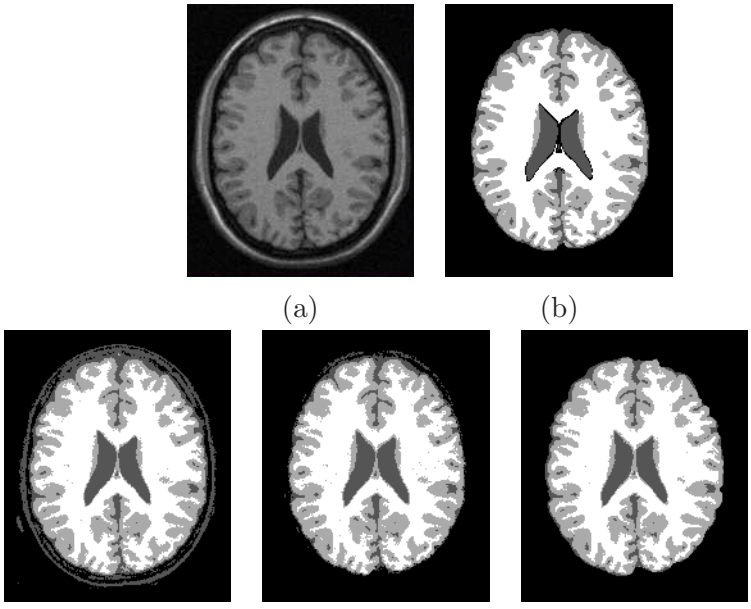
Nous allons maintenant attribuer à chaque voxel la classe la plus représentative afin d'obtenir une segmentation en trois classes "dures". Nous comparons d'abord notre méthode avec une segmentation par chaînes de Markov cachées sans prise en compte de l'atlas, des hétérogénéités d'intensité et des effets de volume partiel. Nous pourrons ainsi valider l'apport de ces différents éléments. Les résultats obtenus sur des images avec 6 différents niveaux de bruit et 20% d'inhomogénéité sont présentés sur la figure 3.8. On constate que l'inclusion dans le modèle du biais, de l'information apportée par l'atlas et des effets de volume partiel permet d'améliorer nettement la qualité de la segmentation.

Nous comparons ensuite la méthode proposée (HMC) avec les méthodes EMS et SPM5. La segmentation "dure" obtenue sur une image Brainweb T1 avec 5% de bruit et 20% d'inhomogénéité avec les trois différentes méthodes testées est illustrée sur la figure 3.9. La figure 3.10 présente les index Kappa obtenus avec différentes méthodes de segmentation sur des images de la base Brainweb avec 20% d'inhomogénéité et un niveau de bruit variant de 0 à 9%. La méthode que nous proposons obtient les meilleurs résultats pour chaque niveau de bruit à la fois sur la matière blanche et la matière grise. On peut remarquer que l'index Kappa de la méthode HMC proposée décroît lorsque le niveau de bruit augmente. Pour les méthodes EMS et SPM5, l'index Kappa est maximal pour un niveau de bruit situé entre 3 et 5% ce qui correspond en pratique au niveau de bruit présent sur les images réelles. Cependant, même pour ces niveaux de bruit la méthode proposée prenant en compte l'effet de volume partiel obtient la segmentation la plus précise.



(a) (b) (c)

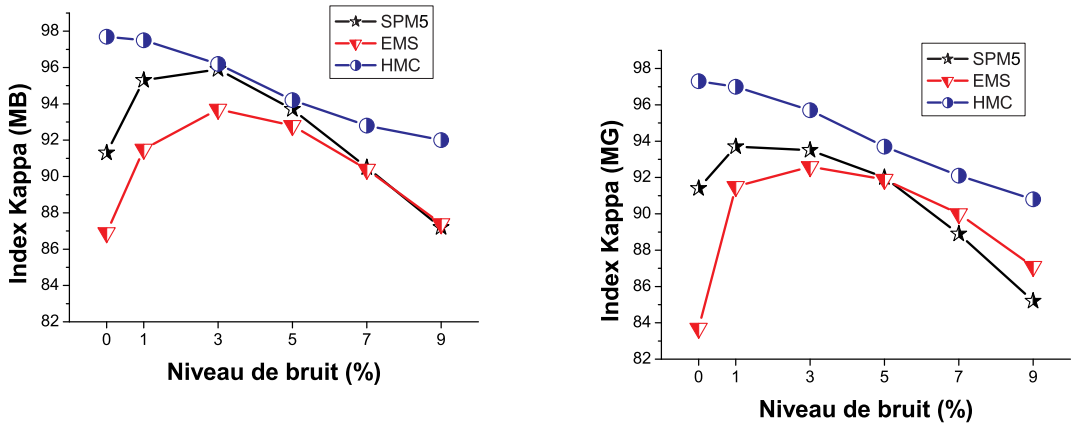
FIG. 3.8 – *Comparaison des résultats obtenus sur des images T1 de la base Brainweb avec 20% d'inhomogénéité avec d'une part la méthode proposée (HMC PV, biais, atlas) et d'autre part la segmentation par chaînes de Markov cachées sans prise en compte de l'atlas, des hétérogénéités d'intensité et des effets de volume partiel (HMC "dur"). (a), (b) et (c) correspondent respectivement aux index Kappa obtenus pour la matière blanche (MB) et la matière grise (MG) et le liquide céphalo-rachidien (LCR).*



(a) (b)

(c) (d) (e)

FIG. 3.9 – *Résultats obtenus sur une image Brainweb avec 5% de bruit et 20% d'inhomogénéité. (a) représente les données à segmenter et (b) la vérité terrain. (c), (d) et (e) correspondent respectivement aux résultats obtenus avec SPM5, EMS et la méthode proposée. L'index Kappa obtenu pour cette image avec les différentes méthodes est présenté sur la figure 3.10.*



(a) (b)

FIG. 3.10 – *Comparaison des résultats obtenus sur des images T1 de la base Brainweb avec 20% d'inhomogénéité avec les trois différentes méthodes : SPM5, EMS et la méthode HMC proposée HMC. (a), (b) correspondent respectivement aux index Kappa obtenus pour la matière blanche (MB) et la matière grise (MG).*

3.8.2 Images multimodales

3.8.2.1 Segmentation "floue"

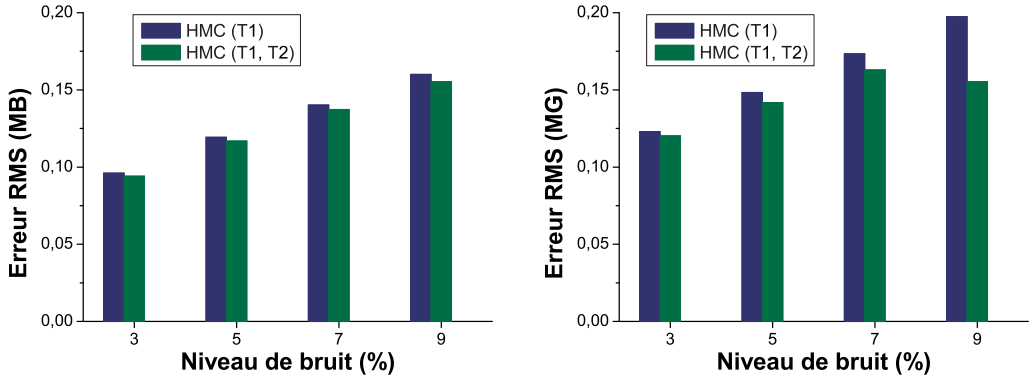
La base Brainweb fournit des images multimodales avec différents niveaux de bruit et d'inhomogénéité. L'approche proposée par Shattuck [Shattuck 01] ayant été seulement appliquée sur des images monomodales, nous allons seulement comparer les résultats sur des images multimodales T1/T2 avec ceux obtenus dans le cas monomodal avec notre méthode HMC. La figure 3.11 présente l'erreur RMS obtenue sur des données mono- et multimodales. L'utilisation d'images multimodales apporte plus d'information et permet de diminuer l'erreur RMS à la fois pour la matière blanche et la matière grise. La multimodalité améliore les résultats de 2% à 3% pour la matière blanche et de 2.7% à 7.3% pour la matière grise.

3.8.2.2 Segmentation "dure"

Après l'étape de segmentation "floue", nous attribuons à chaque voxel la classe la plus représentative pour obtenir une segmentation "dure" en trois classes. Nous avons ensuite comparé les résultats obtenus sur la base Brainweb T1/T2 avec ceux de la méthode EMS. Les graphiques de la figure 3.12 soulignent le fait que la multimodalité améliore les résultats. Les index Kappa obtenus par la méthode HMC sur les images multimodales T1 et T2 sont supérieurs à ceux obtenus sur les images monomodales T1 et ceux obtenus par la méthode EMS sur les images multimodales.

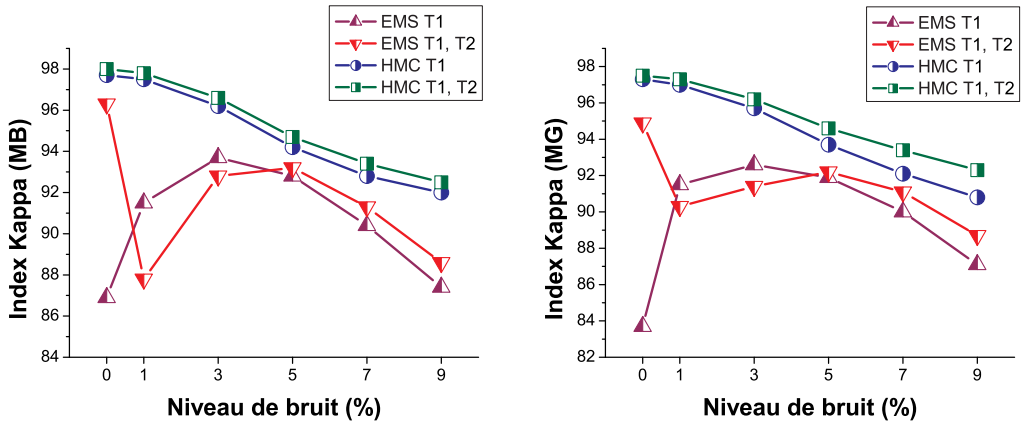
3.9 Validation sur la base IBSR

Après avoir validé notre algorithme sur des images synthétiques, nous l'avons testé sur des images réelles. Les 18 images et leurs segmentations manuelles sont fournies par le Center for



(a) (b)

FIG. 3.11 – *Erreur RMS obtenue sur la base Brainweb avec 20% d'inhomogénéité à la fois pour des images mono- et multimodales. (a) et (b) correspondent respectivement aux résultats obtenus pour la matière blanche (MB) et la matière grise (MG).*



(a) (b)

FIG. 3.12 – *Comparaison des résultats obtenus sur des images T1 de la base Brainweb avec 20% d'inhomogénéité avec les différentes méthodes, EMS et la méthode HMC proposée, à la fois sur des images monomodales (T1) et multimodales (T1, T2). (a) et (b) correspondent respectivement aux index Kappa obtenus pour la matière blanche (MB) et la matière grise (MG).*

Morphometric Analysis du Massachusetts General Hospital[5]. Seule la segmentation en 3 classes "dures" par des experts est disponible. On utilisera donc les critères d'évaluation présentés dans la section 3.6.3. Pour utiliser ces critères, une segmentation en trois classes est requise. Nous réalisons donc une segmentation "dure" en assignant à chaque voxel la classe la plus représentative. Pour les données réelles, les images doivent préalablement être recalées sur l'atlas. Le cerveau est ensuite extrait en utilisant l'algorithme BET [Smith 02]. Néanmoins après cette étape, certains tissus n'appartenant pas au cerveau apparaissent encore sur les images et biaisent l'estimation des paramètres et par conséquent le résultat de la segmentation. Pour régler ce problème, l'atlas est utilisé pour enlever les tissus restants n'appartenant pas au cerveau : les voxels pour lesquels l'atlas indique une probabilité nulle d'appartenir à chacun des tissus sont ainsi enlevés.

Comme pour les images synthétiques de la base Brainweb, nous avons comparé notre méthode avec SPM5 et EMS pour chaque cas de la base IBSR. Un exemple de résultats obtenus sur la base IBSR est présenté sur la figure 3.13. Les différents coefficients ont été calculés à la fois pour la matière blanche et la matière grise. La moyenne des résultats obtenus sur les 18 images IBSR est reportée dans le tableau 3.5. La figure 3.14 compare les performances les différentes méthodes sur chaque image de la base IBSR. Etant donné que la segmentation de l'expert inclut seulement le liquide interne (ventricules) et que les différentes méthodes segmentent aussi le liquide externe, nous n'avons pas reporté les résultats pour le liquide céphalo-rachidien.

Algorithme	Critère	Matière Blanche (%)		Matière Grise (%)	
		Moyenne	Ecart-type	Moyenne	Ecart-type
SPM5	Sensibilité	84.72	5.52	79.97	8.81
	Spécificité	99.38	0.18	97.56	4.48
	Kappa Index	85.27	4.13	78.7	13.98
	Overlap	74.52	5.88	66.6	15.88
EMS	Sensibilité	81.09	4.92	72.86	7.62
	Spécificité	99.67	0.15	99.03	0.26
	Kappa Index	85.87	2.27	78.94	5.68
	Overlap	75.31	3.32	65.56	7.48
Méthode proposée	Sensibilité	90.64	2.53	71.78	6.85
	Spécificité	99.16	0.24	99.35	0.26
	Kappa Index	86.53	1.73	79.94	5.57
	Overlap	76.29	2.67	66.93	7.6

TAB. 3.5 – *Résultats obtenus sur 18 cerveaux de la base IBSR avec trois différentes méthodes. La moyenne sur les 18 images est présentée.*

La méthode proposée obtient une sensibilité élevée pour la matière blanche, ce qui signifie que la matière blanche est bien détectée comme étant de la matière blanche. Pour la matière grise, la sensibilité est plus faible avec notre méthode, mais la spécificité est plus élevée. Ainsi notre méthode détecte un peu moins de matière grise par rapport aux autres méthodes. D'autre part la méthode proposée obtient un index Kappa et un overlap plus élevé que ceux de SPM5 et EMS, à la fois

[5] `http://www.cma.mgh.harvard.edu/ibsr/`

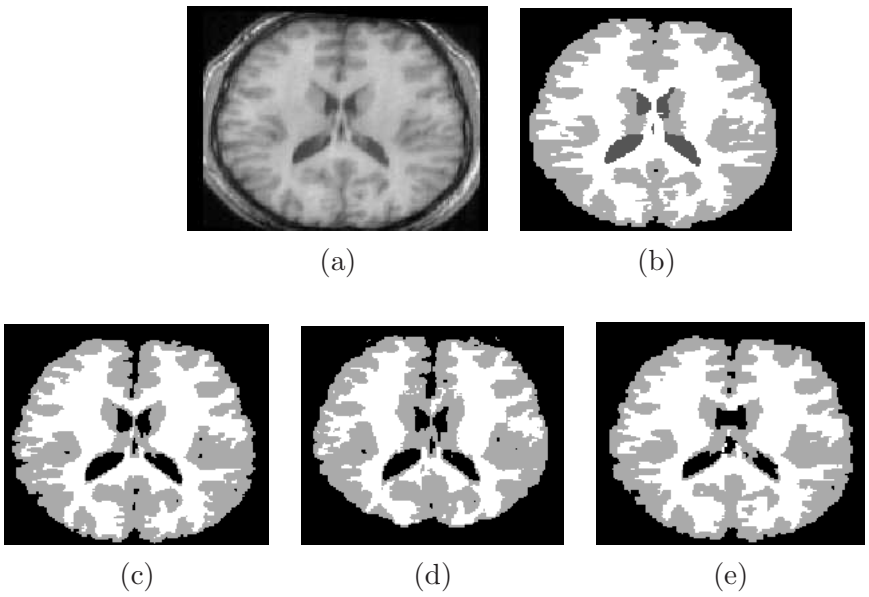
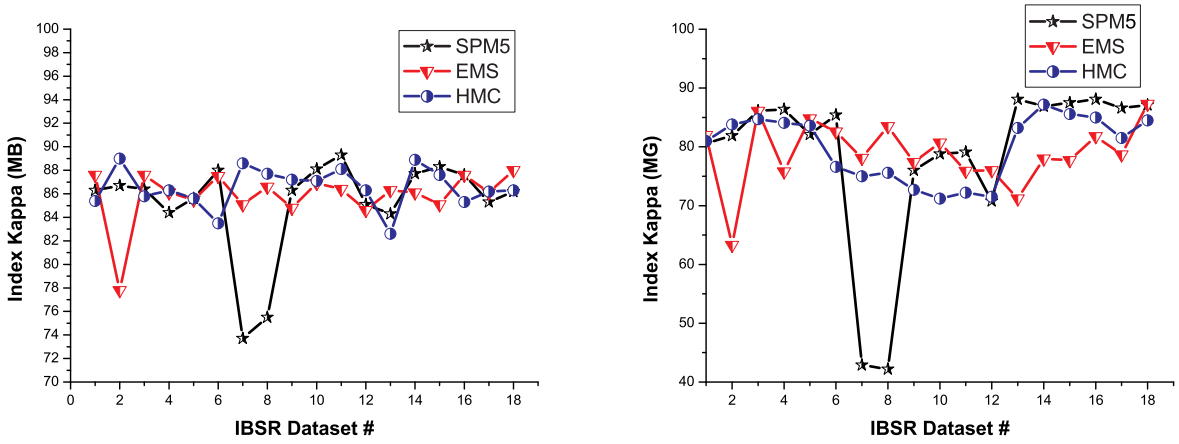


FIG. 3.13 – *Résultats obtenus sur l'image IBSR n°2. (a) correspond aux données que l'on souhaite segmenter et (b) à la segmentation de l'expert. (c), (d) et (e) correspondent respectivement aux résultats obtenus avec SPM5, EMS et la méthode proposée.*

pour la matière blanche et la matière grise. Néanmoins, pour deux images de la base IBSR, la segmentation effectuée par SPM5 est très éloignée de la segmentation de l'expert (Fig. 3.14), en particulier au niveau de la matière grise, ce qui explique l'écart-type élevé. Lorsque ces deux images ne sont pas incluses dans la moyenne, les résultats obtenus avec SPM5 sont proches de ceux de HMC pour la matière blanche et meilleurs pour la matière grise. La figure 3.14 présente les index Kappa obtenus pour chaque image de la base IBSR avec les trois différentes méthodes. Dans certains cas, la méthode présentée obtient la segmentation de la matière blanche la plus précise. L'index Kappa de la matière blanche se situe entre 82 et 90%, avec une moyenne de 86.5 et un faible écart-type. Au niveau de la matière grise, l'index Kappa se situe entre 70 et 90%, avec une moyenne de 79.9%, supérieure à la moyenne de SPM5 et EMS, et un faible écart-type.



(a) (b)

FIG. 3.14 – *Comparaison des résultats obtenus sur la base IBSR avec les trois différentes méthodes : SPM5, EMS et la méthode proposée. L'index Kappa est reporté pour : (a) la matière blanche (b) la matière grise.*

Chapitre 4

Extension à des données pathologiques : Cas de la Sclérose en Plaques

La méthode que nous avons présentée et validée dans les chapitres précédents était dédiée à la segmentation des tissus cérébraux mais ne prenait pas en compte l'existence potentielle de données atypiques comme des tumeurs ou des lésions de Sclérose En Plaques (SEP). L'objet de ce chapitre est donc d'étendre ce modèle à la détection de lésions SEP. Nous commençons par une présentation de cette maladie et par un bref état de l'art de la segmentation de lésions SEP (Section 4.1). Puis nous développons la méthode proposée qui considère ces lésions comme des données atypiques par rapport à un modèle statistique d'images non pathologiques (Section 4.2). Enfin cette méthode sera validée à la fois sur des bases d'images synthétiques et réelles (Section 4.4) où nous comparerons les résultats obtenus avec les segmentations manuelles réalisées par des experts.

4.1 Sclérose en plaques

4.1.1 Connaissances actuelles de cette maladie

Nous allons d'abord présenter la sclérose en plaques en évoquant les caractéristiques et les causes potentielles de cette maladie[1][2].

4.1.1.1 Présentation de la sclérose en plaques

La sclérose en plaques est une "affection du système nerveux central caractérisée par un processus de démyélinisation localisé dans la substance blanche aboutissant à la constitution de plaques de sclérose et évoluant par poussées successives, plus ou moins régressives, survenant à intervalles irréguliers dont la durée est imprévisible."[3] Il s'agit d'une maladie inflammatoire du système nerveux central provoquant la démyélinisation des nerfs. La myéline est une substance lipidique entourant

[1] Source : `http://www.msif.org/`
[2] Source : `http://www.lfsep.asso.fr/`
[3] Source : Dictionnaire de Médecine Flammarion, 7ème édition, 2001.

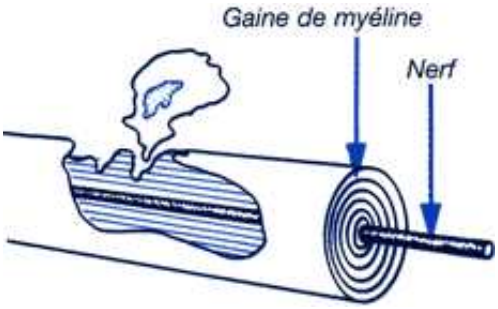


FIG. 4.1 – *Gaines de myéline. Source :* `http://www.lfsep.asso.fr/`.

les fibres nerveuses de la substance blanche permettant la transmission rapide de l'influx nerveux. La perte de la myéline (démyélinisation) s'accompagne d'une perturbation dans la faculté des nerfs à transmettre l'influx nerveux à partir du cerveau vers les membres et à partir des membres vers le cerveau ce qui occasionne divers symptômes de la SEP. Les endroits où la myéline disparaît se présentent comme des zones cicatricielles durcies (plaques ou lésions).

4.1.1.2 Fréquence

Environ 2,5 millions de personnes dans le monde sont atteintes de sclérose en plaques. En Europe occidentale presque un habitant sur mille est atteint de SEP. En France, plus de 80 000 personnes sont touchées. La SEP est une maladie qui frappe en général de jeunes adultes, l'âge moyen d'apparition se situant entre 20 et 40 ans dans 70% des cas avec une prédominance féminine : environ 3 femmes pour 2 hommes. Il existe une répartition géographique inégale de la maladie caractérisée par un gradient Nord-Sud : à savoir la maladie diminue en fréquence entre le Nord et le Sud. Dans l'Europe du Nord et l'Amérique du Nord, notamment en Scandinavie, Ecosse et Canada, on note un taux élevé de SEP atteignant 1 pour 1000 habitants.

4.1.1.3 Causes

La cause de la SEP reste encore partiellement inconnue. Il s'agirait plutôt d'une maladie multifactorielle associant principalement des facteurs environnementaux et génétiques :

- auto-immunité : les dommages causés à la myéline dans la SEP peuvent être dus à une réponse anormale du système immunitaire du corps (qui est normalement appelé à défendre le corps contre les organismes envahissants : bactéries et virus). Le corps s'attaquerait à ses propres cellules et tissus, soit à la myéline dans le cadre de la SEP. Un virus probablement dormant dans le corps pourrait perturber le système immunitaire ou indirectement déclencher le processus auto-immun.
- génétique : en faveur d'un facteur génétique, outre la faible prévalence de la SEP chez les japonais, on relève la rareté de la maladie chez les Noirs américains au Nord comme au Sud des Etats-Unis. L'intervention du patrimoine génétique est confirmée par le dénombrement des familles multi cas plus fréquentes que ne donnerait le hasard : on estime le risque à 2% pour les frères et soeurs d'un patient. Une étude canadienne de jumeaux dont l'un est porteur de la maladie montre à la fois l'importance du facteur génétique dans le déterminisme de

la maladie et le fait qu'il ne s'agit pas d'une maladie héréditaire. Il existe une susceptibilité d'origine génétique mais qui ne suffit pas pour que se produise la maladie.
- facteur d'environnement : l'hypothèse d'un facteur d'environnement est appuyée en particulier sur l'étude des migrations de populations entre des zones de prévalence inégale. Par exemple ceux qui migrent après l'âge de 15 ans ont le risque de la région d'origine, alors que ceux qui migrent avant cet âge ont le risque de la région d'arrivée.

4.1.1.4 Symptômes

La SEP commence le plus souvent par une poussée au cours de laquelle les signes et symptômes apparaissent rapidement puis disparaissent en totalité ou en partie. Elle peut se reproduire à plus ou moins brève échéance une ou plusieurs fois. Les symptômes dépendent des zones du système nerveux central qui ont été affectées. La SEP ne présente pas d'évolution type et chaque patient présente un ensemble distinct de symptômes qui peuvent varier d'une période à l'autre et dont la gravité et la durée peuvent également changer. Les principaux symptômes consistent en :
- troubles de la vue :
  - névrite optique : perte brutale de l'acuité visuelle précédée de douleurs à la mobilisation du globe oculaire,
  - vision floue ;
- troubles de l'équilibre et de la coordination :
  - perte d'équilibre, vertiges,
  - tremblements ;
- troubles moteurs : fatigabilité à la marche ;
- troubles sensitifs :
  - picotements, engourdissements, sensations de brûlures,
  - douleurs faciales, musculaires ou articulaires ;
- troubles de la parole avec ralentissement et modification du rythme de la parole ;
- troubles génito-sphinctériens (soit incontinence, soit rétention) ;
- troubles de la mémoire et de la concentration.

L'échelle EDSS, mise au point par Kurtzke [Kurtzke 83] et présentée dans le tableau C.1 est le principal outil de cotation clinique commun à tous les neurologues pour juger l'évolution des patients.

4.1.1.5 Sclérose en plaques et IRM

Durant ces dernières années, l'imagerie par résonance magnétique a eu un impact majeur sur le diagnostic et la connaissance de la maladie. Elle permet notamment de suivre l'évolution de certaines structures ou lésions au cours du temps sans avoir recours à des explorations invasives. Son rôle est de fournir un critère diagnostique de dissémination dans l'espace et dans le temps. Elle permet également d'exclure d'autres pathologies qui pourraient être à l'origine des signes et des symptômes du patient. McDonald *et al.* ont choisi d'utiliser les critères de Barkhof [Barkhof 97] comme critères diagnostiques lorsqu'on fait appel à l'IRM [McDonald 01]. L'IRM est considérée comme positive pour le diagnostic de la SEP, selon les critères de McDonald (Tab. 4.1), lorsque 3 des 4 critères IRM de Barkhof sont remplis :

Présentation clinique	Données complémentaires requises pour le diagnostic
au moins 2 poussées et au moins 2 sites affectés	aucunes
au moins deux poussées et un seul site affecté	dissémination spatiale des lésions à l'IRM ou poussée clinique suivante dans un site différent
1 poussée et au moins 2 sites affectés	dissémination spatiale des lésions à l'IRM ou 2ème poussée clinique
1 seule poussée	dissémination spatiale des lésions à l'IRM (ou au moins 2 lésions évocatrices à l'IRM et LCR positif) ET dissémination temporelle sur des IRM successives
progression insidieuse évocatrice de SEP	LCR positif ET dissémination spatiale démontrée par l'IRM ET dissémination temporelle démontrée par l'IRM

TAB. 4.1 – *Critères de McDonald [McDonald 01].*

1. une lésion prenant le contraste ou à défaut au moins 9 lésions sur la séquence pondérée en T2 ;
2. au moins une lésion sous-tentorielle ;
3. au moins une lésion située à la jonction cortico-souscorticale ;
4. au moins trois lésions péri-ventriculaires.

L'IRM peut aussi démontrer une dissémination temporelle « douteuse » ou absente sur les seuls critères cliniques, et permet ainsi de poser un diagnostic selon les critères de McDonald, si l'on démontre :

1. Une lésion prenant le contraste ou une nouvelle lésion T2, sur une IRM réalisée 3 mois après un événement inaugural cliniquement compatible avec une poussée ;
2. Si l 'IRM initiale est réalisée moins de 3 mois après la poussée inaugurale, il suffit d'une prise de contraste sur une nouvelle IRM de contrôle réalisée au moins trois mois plus tard. Si cette dernière est négative, un deuxième contrôle distant d'au moins 3 mois devra faire la preuve de l'apparition d'une nouvelle lésion (T2 ou prise de contraste).

4.1.2 Méthodes existantes pour la segmentation de lésions SEP

La plupart des méthodes de segmentation de lésions SEP décrites dans la littérature sont des méthodes semi-automatiques destinées à faciliter la tâche fastidieuse de délimitation manuelle des lésions par des médecins et à réduire la variabilité inter- et intra-experts associée à ces segmentations manuelles. Dans ces approches, l'interaction de l'utilisateur peut consister à accepter ou rejeter des lésions segmentées automatiquement [Udupa 97] ou à délimiter manuellement des régions de tissus purs pour entraîner un classifieur automatique [Kamber 95, Johnston 96, Udupa 97]. Udupa *et al.* ont utilisé le concept de connectivité pour détecter les objets flous représentant la matière blanche, la matière grise et le liquide céphalo-rachidien. Les zones de trous restantes entre les différents

tissus cérébraux représentent les localisations des lésions potentielles [Udupa 97]. L'inconvénient de cette méthode est qu'elle requière l'intervention manuelle d'un opérateur pour éliminer les faux positifs. La connectivité floue est basée sur l'interaction entre les différentes structures de l'image. La force de la connexion entre voxels est déterminée à la fois par leur distance dans l'image et leurs niveaux de gris. Kamber *et al.* utilisent un modèle probabiliste des tissus cérébraux [Kamber 95]. Ce modèle a été construit à partir de cerveaux segmentés de douze sujets sains recalés dans l'espace de Talairach puis moyennés. Ce modèle est utilisé à la fois comme un masque pour confiner la recherche des lésions à l'intérieur de la matière blanche et pour fournir des propriétés géométriques permettant de classifier les lésions : par exemple la probabilité du LCR ventriculaire peut s'avérer utile pour détecter des lésions périventriculaires. La détection des lésions est effectuée en utilisant un arbre de décision. Johnston *et al.* fusionnent les cartes de segmentation de la matière blanche et des lésions obtenues à partir d'un classifieur statistique basé sur un modèle MRF, produisant ainsi un masque couvrant la totalité de la matière blanche [Johnston 96]. Les voxels à l'intérieur de ce masque sont ensuite reclassifiés soit en matière blanche soit en lésion. Dans le système INSECT, Zijdenbos *et al.* présentent une chaîne de traitement basée sur un réseau de neurones [Zijdenbos 94, Zijdenbos 98, Zijdenbos 02].

Les lésions SEP sont souvent détectées comme des voxels atypiques (”outliers”) par rapport à un modèle statistique d'images cérébrales ”normales”. Schroeter *et al.* utilisent des mélanges de gaussiennes pour modéliser la présence de différentes composantes en chaque voxel [Schroeter 98] . A chaque itération, les voxels excédant une distance de Mahalanobis prédéfinie par rapport à chaque distribution gaussienne sont rejetés et la mise à jour des paramètres du modèle est uniquement basée sur les voxels qui n'ont pas été rejetés. La méthode de Van Leemput *et al.* [Van Leemput 01] est proche d'un point de vue algorithmique de celle proposée par Schroeter *et al.* mais prend aussi en compte l'information sur le voisinage en utilisant un modèle de Potts. Ils introduisent également des poids reflétant le caractère plus ou moins atypique de chaque voxel et l'utilisation d'un atlas probabiliste du cerveau permet de rendre la méthode totalement automatique.

L'approche proposée par Dugas-Phocion utilise directement la distance de Mahalanobis au sein des itérations de l'EM pour le rendre plus robuste aux données atypiques et emploie les différentes séquences d'IRM de façon hiérarchique [Dugas-Phocion 04, Dugas-Phocion 06]. Aït-ali *et al.* proposent un EM robuste spatio-temporel fournissant une segmentation longitudinale d'IRM multi-séquences [Aït-ali 05, Ait-Ali 06]. L'estimation robuste des classes est basée sur l'estimateur de vraisemblance tamisée. Les lésions sont détectées en utilisant la distance de Mahalanobis puis classifiées en utilisant des contraintes *a priori* sur l'intensité des manifestations de la SEP en IRM. L'inconvénient de ces deux méthodes provient du fait qu'elles ne prennent pas en compte l'information sur le voisinage spatial.

Les méthodes de détection de lésions que nous venons de présenter sont basées sur la segmentation des lésions. D'autres approches permettent de suivre l'évolution des lésions au cours du temps à la fois au niveau de la taille, de la forme et de la localisation. Ces approches que nous ne détaillerons pas ici sont basées sur l'étude de l'évolution de la maladie entre plusieurs instants distincts (étude *longitudinale*). Cette approche longitudinale peut évidemment être effectuée à l'aide des méthodes de segmentation présentées, en appliquant au préalable un algorithme de recalage entre les segmentations des différents instants. Les autres méthodes sont plutôt des méthodes de

comparaison entre différentes images, soit par analyse d'un champ de déformation [Rey 02], soit par comparaison directe des intensités. Cette dernière approche peut être effectuée par un simple seuillage des images de soustraction [Lemieux 98] ou par des méthodes de détection de changements statistiques [Bosc 03c, Prima 03].

4.2 Modélisation proposée

Nous avons considéré les lésions SEP comme des outliers par rapport à un modèle statistique d'images cérébrales non pathologiques. Nous avons donc remplacé l'étape d'estimation des paramètres du modèle HMC basée sur l'algorithme EM par une estimation robuste basée sur l'estimateur de vraisemblance tamisée. L'utilisation d'un atlas probabiliste va permettre de détecter plus facilement ces outliers. De plus un certain nombre de contraintes sur ces outliers et une étape de post-traitement ont été ajoutés pour garder seulement les outliers présents dans la substance blanche correspondant à des lésions SEP.

4.2.1 Estimateur robuste

L'estimateur de vraisemblance tamisée *Trimmed Likelihood Estimator* (TLE) a été introduit par Neykov et Neytchev [Neykov 90] puis développé par Vandev et Neykov en 1993 [Vandev 93], et Müller et Neykov en 2003 [Müller 03] pour estimer un mélange de lois normales multivariées et des modèles linéaires généralisés de façon robuste. Le schéma d'optimisation utilisé pour calculer cet estimateur dérive du schéma d'optimisation du critère des moindres carrés tamisés de Rousseeuw et Leroy [Rousseeuw 87]. Cet algorithme a été utilisé pour la segmentation d'IRM cérébrales par Aït-Ali [Aït-ali 05] dans le cadre d'une modélisation par mélange de gaussiennes.

Le principe est le suivant : on recherche h observations parmi un échantillon de taille n pour lesquelles la vraisemblance est maximale.

4.2.1.1 Estimateur pondéré tamisé

Soit un ouvert $\Psi = \{\psi_{(i)} : \Theta \to \mathbb{R}^+, \forall i \in [1, \ldots, n]\}, \quad \Theta \subset \mathbb{R}^r$. L'estimateur tamisé pondéré est défini par :

$$W_h = \arg_{\theta \in \Theta} \min \sum_{i=1}^{h} w_{(i)} \psi_{(i)}(\theta) \tag{4.1}$$

avec

- $\psi_i(\theta) \leq \psi_{i+1}(\theta)$
- h paramètre de tamisage
- $w_{(i)} \geq 0, \forall i = 1, \ldots, n$ pondérations, et $w_{(h)} > 0$.

4.2.1.2 Estimateur de vraisemblance tamisée pondéré

L'estimateur de vraisemblance tamisée pondéré peut être trouvé à partir de l'estimateur pondéré tamisé en prenant la fonction ψ telle que : $\psi_{(i)}(\theta) = -\log f(y_i, \theta)$ où $y_i \in \mathbb{R}^m \ \forall i = 1, \ldots, n$ observations indépendantes et identiquement distribuées ayant pour fonction de densité $f(y, \theta)$.

L'estimateur de vraisemblance tamisée pondéré est défini par :

$$W_h = \hat{\theta}_{WTLE} = \arg_{\theta\in\Theta} \min \sum_{i=1}^{h} w_{\nu(i)} \psi_{\nu(i)}(\theta) \tag{4.2}$$

avec

- $\psi_{\nu(i)}(\theta) \leq \psi_{\nu(i+1)}(\theta)$, $\nu(i)$ permutation des indices $\forall i = 1, \ldots, n$
- $w_{\nu(i)} \geq 0, \forall i = 1, \ldots, n$

4.2.1.3 Estimateur de vraisemblance tamisée

On prend $w_{\nu(i)} = 1, \forall i = 1, \ldots, h$ et $w_{\nu(i)} = 0$ sinon. L'estimateur de vraisemblance tamisée est défini par :

$$\hat{\theta}_{TLE} = \arg_{\theta\in\Theta} \max \prod_{i=1}^{h} f(y_{\nu(i)}, \theta) \tag{4.3}$$

avec $\nu = (\nu(1), \ldots, \nu(n))$ permutations des indices telles que $f(y_{\nu(i)}, \theta) \geq f(y_{\nu(i+1)}, \theta)$.

Les conditions générales pour l'existence d'une solution de l'équation 4.2 sont prouvées dans [Dimova 04]. La convergence et les propriétés asymptotiques sont étudiées dans [Čížek 02].

4.2.1.4 Algorithme rapide

On note $\psi_{(i)}(\theta) = -\log f(y_i, \theta)$ où $y_i \in \mathbb{R}^m$, $\forall i = 1, \ldots, n$ observations indépendantes et identiquement distribuées ayant pour fonction de densité $f(y, \theta)$.

- soit un sous-ensemble $H^{old} = \{y_{j_1}, \ldots, y_{j_h}\} \subset \{y_1, \ldots, y_n\}$
- on calcule l'estimateur de θ, $\hat{\theta}^{old}$ maximisant la vraisemblance du sous-ensemble H^{old} ;
- on définit $Q^{old} = \sum_{i=1}^{h} \psi_{j_i}(\hat{\theta}^{old})$;
- on trie $\psi(y_i, \hat{\theta}^{old})$ pour $i = 1, \ldots, n$ en ordre ascendant : $\psi_{\nu(i)}(\hat{\theta}^{old}) \leq \psi_{\nu(i+1)}(\hat{\theta}^{old})$ où $\nu = (\nu(1), \ldots, \nu(n))$ représente la permutation ;
- on définit le nouveau sous-ensemble : $H^{new} = \{y_{\nu(1)}, \ldots, y_{\nu(h)}\}$;
- on calcule l'estimateur de θ, $\hat{\theta}^{new}$ maximisant la vraisemblance du sous-ensemble H^{new} ;
- on définit $Q^{new} = \sum_{i=1}^{h} \psi_{\nu(i)}(\hat{\theta}^{new})$.

4.2.2 Modélisation proposée

4.2.2.1 Estimation robuste des paramètres des chaînes de Markov cachées

Dans ce paragraphe, nous utilisons cet estimateur de vraisemblance tamisée pour estimer de façon robuste les paramètres de notre modèle des Chaînes de Markov Cachées (HMC). Dans le cas des Chaînes de Markov cachées, l'algorithme est le suivant :

1. maximisation de la vraisemblance de toutes les données en utilisant l'algorithme EM, donnant une première estimation $\hat{\theta}^{(p-1)}$;

2. tri des résidus $r_n = -\log f(y_n, \hat{\theta}^{(p-1)})$ pour $n = 1, \ldots, N$

$$p(Y_n = l, \theta) = \sum_{\omega_k} p(Y_n = l, X_n = \omega_k, \theta) \quad (4.4)$$

$$= \sum_{\omega_k} P(X_n = \omega_k) p(Y_n = l | X_n = \omega_k, \theta) \quad (4.5)$$

3. maximisation de la vraisemblance pour les vecteurs ayant les h plus petits résidus donnant une nouvelle estimation $\hat{\theta}^{(p)}$ (pour les données considérées comme des outliers, la vraisemblance est mise à 1 : $f_k(y_n) = p(Y_n = l | X_n = \omega_k, \theta) = 1$). On définit $H^{(p)} = \{y_{\nu(1)}, ..., y_{\nu(h)}\}$ le sous-ensemble contenant les h vecteurs ayant les plus "grandes vraisemblances" pour $\hat{\theta}^{(p-1)}$.
   Le calcul des différentes probabilités devient :
   - probabilités *Forward* :
     - $\alpha_1(k) = \pi_k f_k(y_1)$
     - $\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) a_{lk} f_k(y_n)$ avec $f_k(y_n) = 1$ si y_n est considéré comme un outlier.
   - probabilités *Backward* :
     - $\beta_N(k) = 1$
     - $\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) a_{kl} f_l(y_{n+1})$ avec $f_l(y_{n+1}) = 1$ si y_{n+1} est considéré comme un outlier.
   - probabilités conjointes *a posteriori* :

$$\xi_n(i,j)^{(p)} = \frac{\alpha_{n-1}(j) a_{ji} f_i(y_n) \beta_n(i)}{\sum_k \alpha_n(k)} \quad (4.6)$$

   - probabilités marginales *a posteriori* :

$$\gamma_n(i)^{(p)} = \frac{\alpha_n(i) \beta_n(i)}{\sum_j \alpha_N(j)} \quad (4.7)$$

   Les paramètres du modèle sont obtenus par :
   - $\mu_i^{(p)} = \frac{\sum_{n_1} \gamma_{n_1}(i) y_{n_1}}{\sum_{n_1} \gamma_{n_1}(i)}$
   - $\Sigma_i^{(p)} = \frac{\sum_{n_1} \gamma_{n_1}(i)(y_{n_1} - \mu_i)(y_{n_1} - \mu_i)^t}{\sum_{n_1} \gamma_{n_1}(i)}$

     avec y_{n_1} appartenant au sous-ensemble $H^{(p)}$.

4. retour à l'étape 2 jusqu'à convergence. La convergence est considérée comme atteinte lorsque les paramètres n'évoluent plus à ϵ près.

4.2.2.2 Illustration sur des images synthétiques

Nous avons réalisé des tests sur une image synthétique de taille 64×64, composée de 3 classes bruitées avec un bruit gaussien dont les paramètres sont donnés dans le tableau 4.2. De plus 30 voxels de la classe 2 ont été mis à 200, formant ainsi 0.73% d'outliers. Nous avons ensuite testé l'approche robuste pour différentes valeurs de h (correspondant au pourcentage de données gardées pour l'estimation des paramètres) [Bricq 08a].

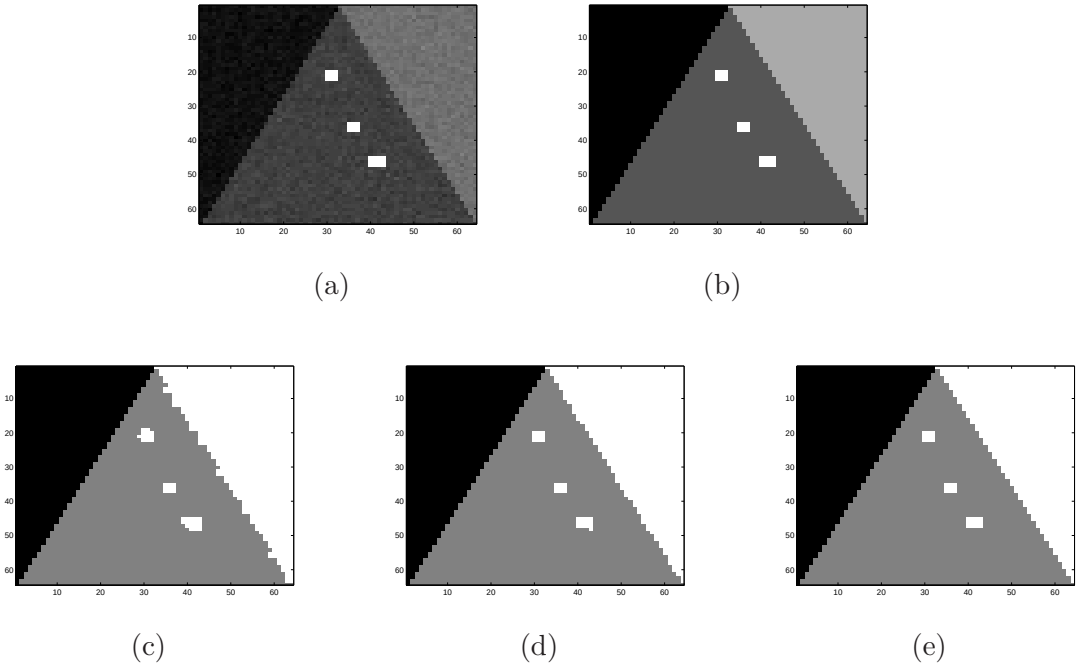
La figure 4.2 présente les résultats de la segmentation en trois classes obtenus en estimant les paramètres de façon robuste avec différentes valeurs du paramètre de tamisage h.

Classe	Moyenne μ	Variance σ^2
Classe 1	75	9
Classe 2	100	9
Classe 3	125	9

TAB. 4.2 – *Paramètres de l'image synthétique.*

h	μ_1	σ_1^2	μ_2	σ_2^2	μ_3	σ_3^2
100	75	8.47	100.08	8.95	126.21	182.37
99.5	75	8.47	100.08	9.02	125.41	69.10
99.27	75	8.47	100.08	9.08	124.95	8.89
99	75.08	7.96	100.08	9.08	124.95	8.89
98	75.30	7.03	100.08	9.08	124.95	8.89
95	75.86	5.40	100.08	9.08	124.95	8.89

TAB. 4.3 – *Résultats obtenus sur l'image synthétique. Estimation des paramètres pour différentes valeurs de h.*

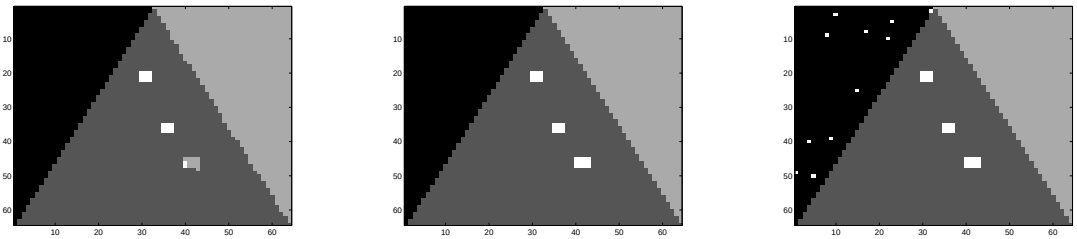


(a) (b)

(c) (d) (e)

FIG. 4.2 – *Résultats obtenus sur une image synthétique. (a) correspond à l'image bruitée et (b) à la vérité. Les segmentations en trois classes obtenues pour différentes valeurs du paramètre de tamisage h sont présentées. (c), (d) et (e) correspondent respectivement aux résultats obtenus en utilisant la segmentation HMC robuste avec h de valeur respective 100, 99.5 et 99.*

Lorsque l'estimation des paramètres ne se fait pas de façon robuste ($h = 100$), la variance de la classe 3 est très élevée : les pixels "outliers" sont en effet considérés comme faisant partie de la classe 3. La frontière entre la classe 2 et la classe 3 est par conséquent mal segmentée (Fig. 4.2 c). Lorsque h diminue, c'est-à-dire lorsque le nombre de pixels considérés comme des outliers et donc non pris en



(a) (b) (c)

FIG. 4.3 – *Résultats obtenus sur l'image synthétique bruitée de la figure 4.2(a). Pour chaque image, les zones en blanc correspondent aux pixels considérés comme des outliers. (a), (b) et (c) correspondent respectivement aux résultats obtenus en utilisant la segmentation HMC robuste avec h de valeur respective 99.5, 99.27 et 99.*

compte dans l'estimation des paramètres augmente, la variance de la classe 3 diminue. Si h diminue trop, la variance de la classe 1 diminue et la moyenne de la classe 1 augmente légèrement : les pixels ayant les intensités les plus faibles sont considérés comme étant les plus éloignés du modèle et donc rejetés.

La figure 4.3 présente les résultats obtenus sur l'image synthétique bruitée avec différentes valeurs de h. Les zones en blanc correspondent aux pixels considérés comme des outliers à la dernière itération. Le nombre d'outliers augmente lorsque h diminue. Pour $h = 99.27$ (ce qui correspond à 0.73% d'outliers), tous les pixels correspondant à des outliers sont bien rejetés, ce qui est normal car ce cas synthétique est assez simple.

Le principal inconvénient de cette méthode est que le paramètre h correspondant au pourcentage de voxels gardés pour l'estimation des paramètres des classes doit être fixé à une valeur ni trop élevée pour pouvoir rejeter tous les outliers, ni trop basse pour ne pas perdre trop d'information. La méthode ainsi implémentée est donc sensible à la valeur retenue pour h. Pour pallier à ce problème, nous proposons une méthode utilisant un paramètre de tamisage adaptatif qui sera présentée au paragraphe 4.2.2.4, permettant d'être davantage robuste.

4.2.2.3 Information *a priori* apportée par un atlas probabiliste

De l'information *a priori* sous forme d'atlas probabiliste a ensuite été injectée dans le modèle présenté précédemment. Cette information supplémentaire va permettre de mieux détecter les lésions. En effet sur les images T1, certaines lésions de Sclérose en Plaques situées dans la matière blanche ont une intensité voisine de l'intensité de la matière grise et seront donc segmentées comme étant de la matière grise au lieu d'être identifiées comme étant des outliers. L'utilisation d'un atlas permet de prendre en compte une information complémentaire sur la localisation spatiale. Ainsi des lésions ayant une intensité proche de la matière grise mais situées dans la matière blanche seront considérées comme des outliers. L'approche que nous proposons utilise un atlas apportant de l'information *a priori*. L'atlas noté B fournit des cartes de probabilités des trois principaux tissus du cerveau (matière blanche MB, matière grise MG et liquide céphalo-rachidien LCR). $b_n(k) = P(B_n = b_n | X_n = \omega_k)$ représente la probabilité du voxel n d'appartenir à la classe k.

Dans ce cas, nous avons comme information à la fois l'intensité des données observées et les cartes de probabilités. Ainsi, nous cherchons :

$$\hat{X}_n = \arg\max_X P(X_n|\boldsymbol{Y}, B) \tag{4.8}$$

où $P(X_n|\boldsymbol{Y}, B) = \alpha_n\beta_n$ avec $\alpha_n = P(X_n|\boldsymbol{Y}_{\leq n}, B_{\leq n})$ probabilité *forward* et $\beta_n = \frac{P(\boldsymbol{Y}_{>n}, B_{>n}|X_n)}{P(\boldsymbol{Y}_{>n}, B_{>n}|\boldsymbol{Y}_{\leq n}, B_{\leq n})}$ probabilité *backward*.

En prenant en compte l'information apportée par l'atlas probabiliste et en estimant les paramètres de façon robuste, l'algorithme de segmentation devient :

1. maximisation de la vraisemblance de toutes les données en utilisant l'EM, donnant une première estimation $\hat{\theta}^{(p-1)}$;

2. tri des résidus $r_n = -\log f(y_n, b_n, \hat{\theta}^{(p-1)})$ pour $n = 1, \ldots, N$

$$P(Y_n = l, B_n, \theta) = \sum_{\omega_k} P(Y_n = l, B_n, X_n = \omega_k, \theta) \tag{4.9}$$

$$= \sum_{\omega_k} P(X_n = \omega_k)P(Y_n = l|X_n = \omega_k, \theta)P(B_n|X_n = \omega_k) \tag{4.10}$$

3. maximisation de la vraisemblance pour les vecteurs ayant les h plus petits résidus donnant une nouvelle estimation $\hat{\theta}^{(p)}$ (pour les données considérées comme des outliers, la vraisemblance est mise à 1 : $f_k(y_n) = 1$). On définit $H^{(p)} = \{y_{\nu(1)}, ..., y_{\nu(h)}\}$ le sous-ensemble contenant les h vecteurs ayant les plus "grandes vraisemblances" pour $\hat{\theta}^{(p-1)}$.
   Le calcul des différentes probabilités devient :
   - probabilités *Forward* :
     - $\alpha_1(k) = \pi_k f_k(y_1)b_1(k)$
     - $\alpha_n(k) = \sum_{l=1}^K \alpha_{n-1}(l)a_{lk}f_k(y_n)b_n(k)$ avec $f_k(y_n) = 1$ si y_n est considéré comme un outlier.
   - probabilités *Backward* :
     - $\beta_N(k) = 1$
     - $\beta_n(k) = \sum_{l=1}^K \beta_{n+1}(l)a_{kl}f_l(y_{n+1})b_{n+1}(k)$ avec $f_l(y_{n+1}) = 1$ si y_{n+1} est considéré comme un outlier.
   - probabilités conjointes *a posteriori* :

$$\xi_n(i,j)^{(p)} = \frac{\alpha_{n-1}(j)a_{ji}f_i(y_n)b_n(i)\beta_n(i)}{\sum_k \alpha_n(k)} \tag{4.11}$$

   - probabilités marginales *a posteriori* :

$$\gamma_n(i)^{(p)} = \frac{\alpha_n(i)\beta_n(i)}{\sum_j \alpha_N(j)} \tag{4.12}$$

   Les paramètres du modèle sont obtenus par :
   - $\mu_i^{(p)} = \frac{\sum_{n_1} \gamma_{n_1}(i)y_{n_1}}{\sum_{n_1} \gamma_{n_1}(i)}$

– $\Sigma_i^{(p)} = \frac{\sum_{n_1} \gamma_{n_1}(i)(y_{n_1} - \mu_i)(y_{n_1} - \mu_i)^t}{\sum_{n_1} \gamma_{n_1}(i)}$
avec y_{n_1} appartenant au sous-ensemble $H^{(p)}$.

4. retour à l'étape 2 jusqu'à convergence. La convergence est considérée comme atteinte lorsque les paramètres n'évoluent plus à ϵ près.

4.2.2.4 Choix du seuil de détection

Le principal inconvénient de cette approche est que le paramètre de tamisage h représentant le pourcentage de voxels utilisés pour l'estimation des paramètres doit être fixé par l'utilisateur. Pour pallier à ce problème, nous proposons d'utiliser un paramètre de tamisage adaptatif à l'aide d'un seuil sur la probabilité $P(Y_n = y_n, B_n, \theta)$. A chaque itération, les voxels pour lesquels la probabilité $P(Y_n = y_n, B_n, \theta)$ est inférieure à un certain seuil s fixé sont considérés comme des outliers et ne sont pas pris en compte dans l'estimation des paramètres du modèle HMC. Ainsi la valeur du paramètre h, c'est-à-dire le pourcentage de voxels utilisés lors de l'estimation des paramètres pourra changer à chaque itération.

4.3 Contraintes sur les lésions et post-traitements

4.3.1 Contraintes sur les lésions

Les voxels considérés comme des outliers ne correspondent pas forcément à des lésions SEP. En effet certains outliers apparaissent dans la classe LCR. L'interprétation de ce phénomène est la suivante : l'approximation gaussienne du bruit Ricien n'est pas valide dans le liquide céphalo-rachidien. En effet la distribution de Rice est approxivement gaussienne seulement dans les zones où l'intensité n'est pas proche de zéro, et le LCR correspond à la classe où l'intensité est la plus faible. Une grande partie des voxels de cette classe vont donc être considérés comme des outliers. Par conséquent, ils ne seront pas pris en compte dans l'estimation des paramètres de la classe LCR. L'estimation des paramètres de moyenne et de variance du LCR risque donc d'être biaisée car ces paramètres ne seront estimés qu'à partir d'un faible nombre d'échantillons. Pour surmonter ce problème, nous avons ajouté la contrainte suivante sur les outliers dans notre algorithme : à chaque itération les voxels pour lesquels l'intensité y vérifie $\mu_{Flair} - 3\sigma_{Flair} < y < \mu_{Flair} + 3\sigma_{Flair}$ ne seront pas considérés comme des outliers.

Remarque Dans le cas où la modalité Flair n'est pas disponible, nous utilisons l'information *a priori* apportée par l'atlas. Ainsi, à chaque itération, les voxels pour lesquels la probabilité de LCR est supérieure à 0.5 ne sont pas considérés comme des outliers.

4.3.2 Post-traitements

Afin de ne garder que les outliers correspondant à des lésions SEP, nous avons appliqué une étape de post-traitement. Lors de cette étape, seuls les outliers pour lesquels l'atlas indique une probabilité d'être dans la matière blanche supérieure à 0.5 et dont le volume est supérieur à $3mm^3$ sont considérés comme des lésions SEP.

4.3.3 Algorithme

Le synoptique utilisé pour la détection des lésions SEP est présenté dans l'algorithme 5.

Algorithme 5 Détection de lésions de sclérose en plaques

1. Recalage non-rigide des images avec l'atlas [Noblet 05]
2. Extraction du cerveau en utilisant BET [Smith 02]
3. Vectorisation de l'image en utilisant le parcours d'Hilbert-Peano 3D [Bandoh 99]
4. Initialisation
   (a) Pour les paramètres *a priori* Φ_x

$$\begin{aligned} \pi_k &= \frac{1}{K}, \forall k \\ t_{ij} &= \begin{cases} \frac{3}{4} \text{ si } i = j \\ \frac{1}{4(K-1)} \text{ sinon} \end{cases} \end{aligned}$$

   (b) Initialisation des paramètres de la loi d'attache aux données Φ_y par les K-Moyennes [Bovik 00]
5. Algorithme HMC robuste avec détection des outliers
6. Segmentation MPM
7. Transformation d'Hilbert-Peano inverse pour obtenir l'image segmentée
8. Post-traitements sur les voxels considérés comme outliers
   (a) seuls les outliers pour lesquels l'atlas indique une probabilité de la matière blanche supérieure à 0.5 sont gardés ;
   (b) les lésions de volume inférieur à $3mm^3$ sont exclues.

4.4 Validation

4.4.1 Matériel

Nous disposons de plusieurs bases de données afin d'évaluer qualitativement et quantitativement les résultats de la détection des lésions. Nous avons d'abord testé notre méthode sur la base synthétique Brainweb [Bricq 08b], puis sur une base d'images réelles [Bricq 08c]. Enfin nous avons évalué les résultats de notre méthode sur les bases fournies lors de notre participation au challenge MICCAI'08[4] sur la segmentation des lésions SEP. Ces résultats ont été évalués par rapport à la segmentation manuelle de deux experts.

Pour chaque base, l'atlas probabiliste utilisé pour la détection des outliers provient d'images de 31 patients sains qui ont été recalées en utilisant un recalage non rigide [Noblet 05], segmentées en utilisant la méthode décrite au chapitre 3 puis moyennées afin d'obtenir des cartes de probabilité des différents tissus sains.

[4] `http://www.ia.unc.edu/MSseg/index.php`

4.4.2 Base Brainweb

Le site Brainweb [5] permet également de simuler des IRM avec des lésions de sclérose en plaques. Nous avons choisi d'utiliser les images avec 3 ou 5% de bruit et 20% d'inhomogénéité qui sont les plus représentatives de données réelles. Différents tests ont été effectués en faisant varier le seuil de détection avec ou sans prise en compte de l'information apportée par l'atlas probabiliste et en utilisant différentes modalités (T1, T2 et densité de protons DP). Les performances des différentes méthodes sont décrites par des courbes de type ROC (Receiver Operating Characteristic) : la sensibilité volumique en fonction de 1−spécificité volumique (Figures 4.4 et 4.5). Le meilleur compromis entre la sensibilité volumique et la spécificité volumique est dans chaque cas obtenu pour une valeur du seuil s de l'ordre de 0.02. Dans la suite du manuscrit, tous les résultats présentés sont obtenus avec une valeur de seuil s égale à 0.02.

Nous avons aussi utilisé l'index Kappa présenté au paragraphe 3.6.3 pour évaluer les résultats obtenus par les différentes méthodes (Tableau 4.4). Ces différents résultats montrent d'une part

Modalités	3% de bruit		5% de bruit	
	Sans atlas	Avec atlas	Sans atlas	Avec atlas
T1/DP	33.8	56.2	29	41.5
T1/T2	75.6	79	74	79
T1/T2/DP	75.5	79	74.1	79

TAB. 4.4 – *Index Kappa obtenus sur les images Brainweb avec 3 et 5% de bruit. Les index Kappa présentés sont les meilleurs obtenus pour chaque méthode en faisant varier le seuil de détection. Les valeurs du seuil s optimales sont de 0.02.*

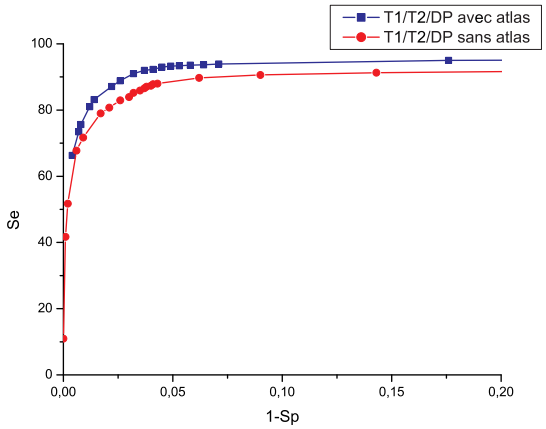
l'apport de la modalité T2 par rapport à la modalité DP et d'autre part l'apport de la prise en compte de l'information apportée par l'atlas. En effet on constate que les index Kappa obtenus à partir des trois modalités T1/T2/DP sont semblables à ceux obtenus avec les deux modalités T1/T2 (75% sans utilisation de l'atlas et 79% avec l'atlas). On peut également remarquer que les index Kappa obtenus avec les deux modalités T1/DP sont très faibles. Dans chaque cas, on observe que l'information apportée par l'atlas améliore nettement les résultats.

Des exemples de résultats obtenus sur les images Brainweb sont présentés sur les figures 4.6 et 4.7.

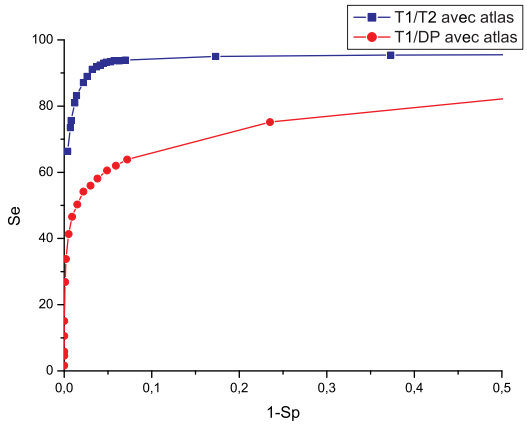
4.4.3 Base réelle

La validation sur des images réelles est effectuée dans le cadre du protocole "Quick". L'objectif de ce protocole consiste d'une part à développer des outils informatiques d'analyse des images pour le diagnostic de la SEP (basés sur les critères de McDonald) et d'autre part à créer une base d'images multi-site (en particulier Strasbourg et Colmar), multi-IRM pour le suivi longitudinal. Ce protocole se base sur le comptage et la localisation des lésions nécessitant leur segmentation et le suivi longitudinal. Dans le cadre de ce protocole, plusieurs bases d'images ont été segmentées manuellement par des experts. Nous avons donc pu comparer les résultats de segmentation des lésions obtenus par notre algorithme par rapport à ces segmentations manuelles.

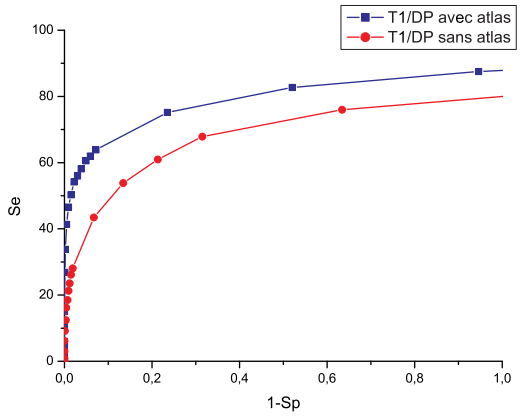
[5] `http://www.bic.mni.mcgill.ca/brainweb/`



(a)



(b)



(c)

FIG. 4.4 – *Résultats obtenus sur les données Brainweb avec 3% de bruit et 20% d'inhomogénéité en faisant varier le seuil s. (a) présente la comparaison de la méthode avec ou sans atlas en prenant en compte les modalités T1/T2/DP. (b) présente les résultats obtenus avec l'atlas en prenant en compte soit les modalités T1/T2 soit T1/DP. (c) présente la comparaison de la méthode avec ou sans atlas en prenant seulement en compte les modalités T1/DP.*

Les résultats que nous présentons ici sont issus de la base provenant de Strasbourg. Nous disposons d'une base contenant les acquisitions de patients à différents instants. Pour chaque instant d'acquisition, le jeu de données contient une séquence 3D d'IRM pondérée en T1 et une Flair. Les données ont été acquises sur une IRM Siemens Allegra $1,5T$. Nous avons appliqué notre méthode sur 8 cas de la base pour le premier examen en prenant en compte les deux modalités T1 et Flair. Ces résultats sont ensuite comparés aux segmentations manuelles de deux différents experts. Nous avons aussi comparé les segmentations manuelles des 2 experts entre elles. Comme outil de comparaison, nous avons utilisé :

– l'index Kappa KI ;

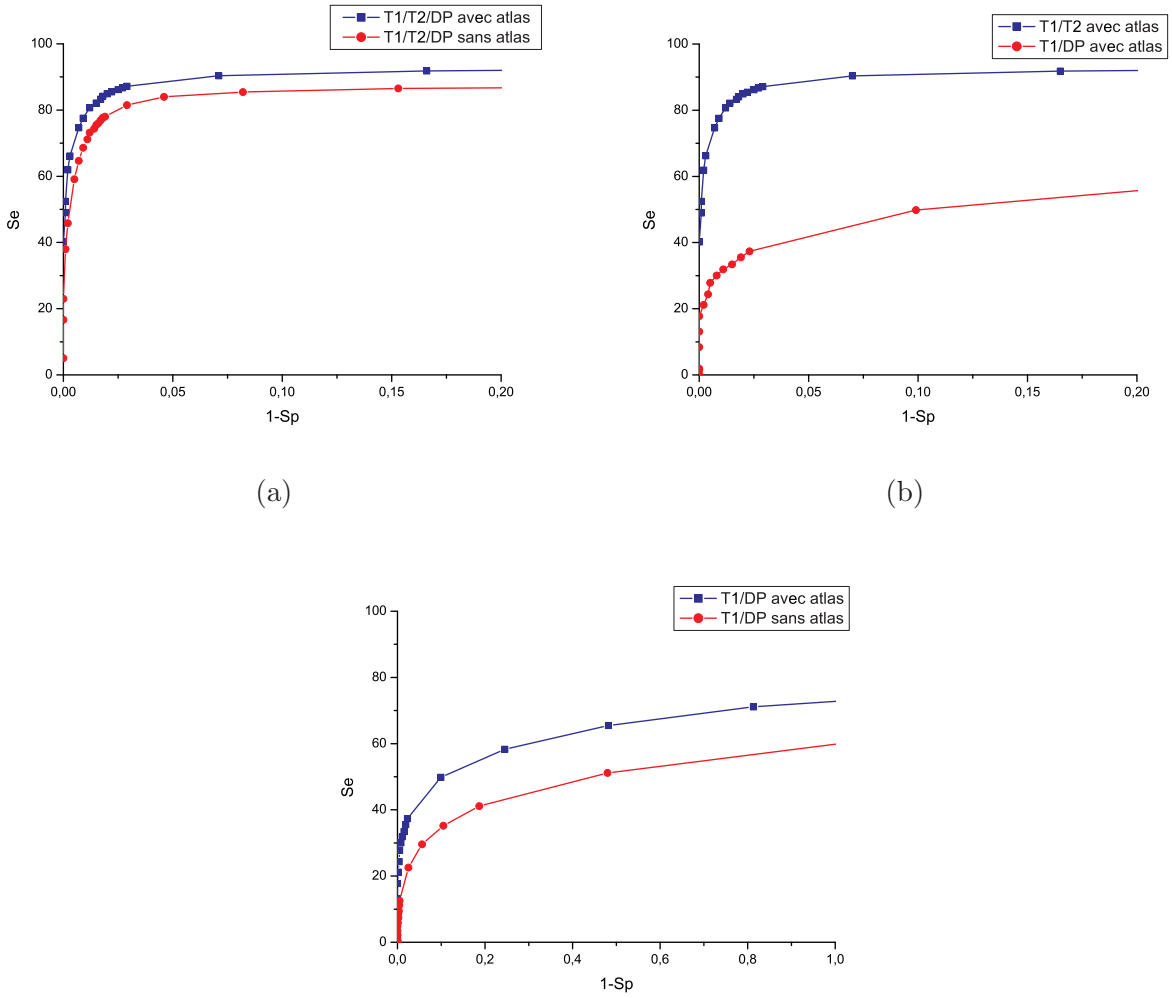


FIG. 4.5 – *Résultats obtenus sur les données Brainweb avec 5% de bruit et 20% d'inhomogénéité en faisant varier le seuil s. (a) présente la comparaison de la méthode avec ou sans atlas en prenant en compte les modalités T1/T2/DP. (b) présente les résultats obtenus avec l'atlas en prenant en compte soit les modalités T1/T2 soit T1/DP. (c) présente la comparaison de la méthode avec ou sans atlas en prenant seulement en compte les modalités T1/DP.*

- le taux de vrais positifs TPF, c'est-à-dire le nombre de lésions dans la segmentation ayant une intersection non nulle avec une lésion segmentée par l'expert divisé par le nombre total de lésions segmentées par l'expert ;
- le taux de faux positifs FPF, c'est-à-dire le nombre de lésions dans la segmentation n'ayant d'intersection avec aucune des lésions segmentées par l'expert divisé par le nombre total de lésions de la segmentation automatique.

Les index Kappa obtenus sur chacun des 8 cas de la base sont présentés sur la figure 4.8 et les taux de vrais et faux positifs sur la figure 4.9. Concernant l'index Kappa, on constate que les résultats obtenus par la méthode proposée sont plus proches de ceux obtenus par l'expert 2.

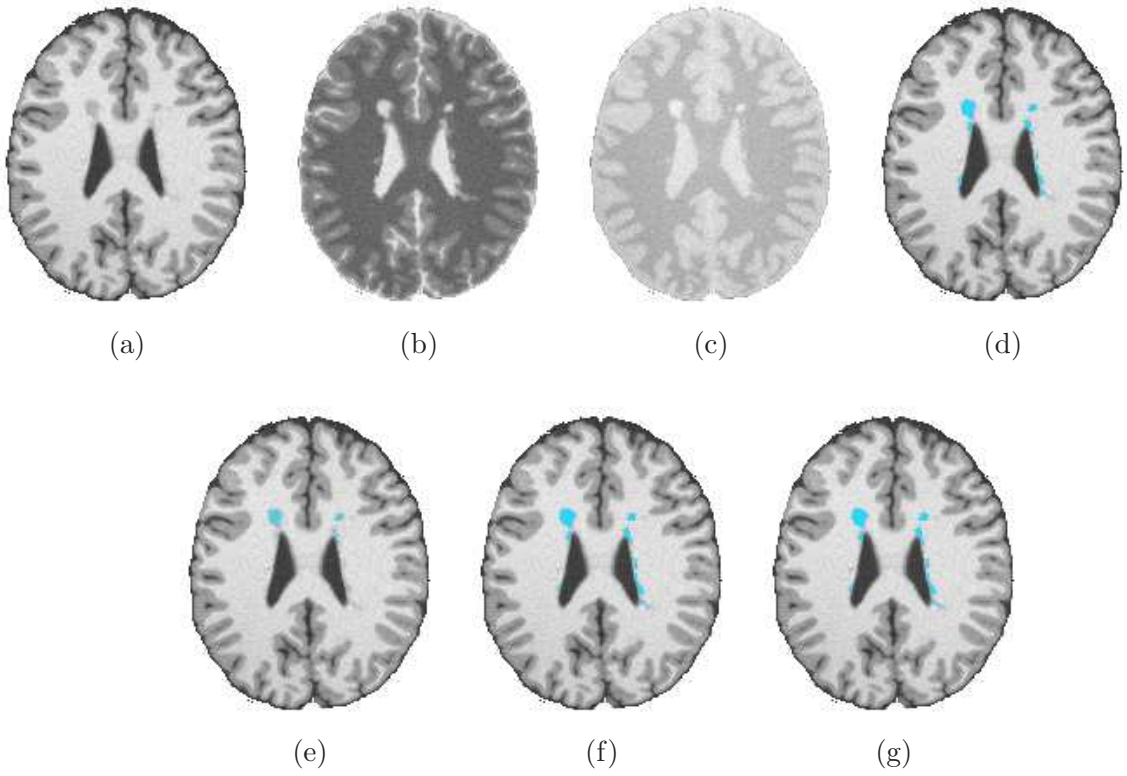


FIG. 4.6 – *Résultats obtenus sur les données Brainweb avec 3% de bruit et 20% d'inhomogénéité. (a), (b) et (c) correspondent respectivement aux images T1, T2 et DP. (d) correspond à la vérité terrain. (e) et (f) correspondent respectivement aux résultats obtenus sans prise en compte de l'atlas à partir des images T1/DP, respectivement T1/T2/DP. (g) correspond aux résultats obtenus à partir des images T1/T2/DP en utilisant la segmentation HMC robuste avec prise en compte de l'atlas.*

L'index Kappa moyen sur les 8 cas est en effet de 0.66 par rapport à l'expert 2 contre 0.62 par rapport à l'expert 1. Ces résultats sont légèrement inférieurs à l'index Kappa moyen comparant les segmentations des deux experts dont la valeur est de 0.74. Concernant les taux de vrais et faux positifs (respectivement TPF et FPF), les résultats obtenus par la méthode proposée sont dans l'ensemble meilleurs par rapport à l'expert 1 que par rapport à l'expert 2. L'expert 2 obtient un TPF moyen de 0.79 et un FPF moyen de 0.24 par rapport à l'expert 1. La méthode HMC proposée obtient un TPF moyen de 0.77 et un FPF moyen de 0.28 par rapport à l'expert 1, alors qu'elle obtient un TPF moyen de 0.71 et un FPF moyen de 0.3 par rapport à l'expert 2.

Pour le cas 03, le taux de faux positifs est élevé. Ceci est du au faible nombre de lésions pour ce patient : l'expert 1 ne détecte que 2 lésions tandis que l'expert 2 en détecte 5 et la méthode proposée 9.

Pour chaque cas, le nombre d'outliers détectés représente moins de 20% du volume total, par conséquent la statistique porte bien sur un nombre suffisant d'échantillons.

Des exemples de segmentations obtenues sur différents cas de la base de Strasbourg sont présentés sur les figures 4.10 et 4.11.

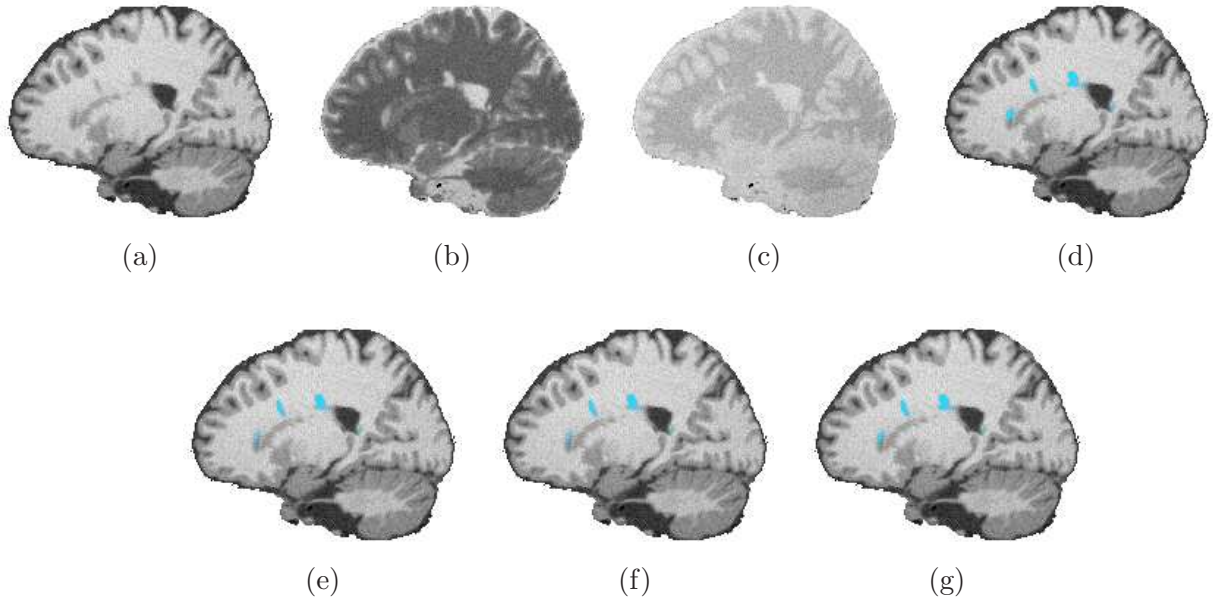


FIG. 4.7 – *Résultats obtenus sur les données Brainweb avec 5% de bruit et 20% d'inhomogénéité. (a), (b) et (c) correspondent respectivement aux images T1, T2 et DP. (d) correspond à la vérité terrain. (e) et (f) correspondent respectivement aux résultats obtenus sans prise en compte de l'atlas à partir des images T1/DP, respectivement T1/T2/DP. (g) correspond aux résultats obtenus à partir des images T1/T2/DP en utilisant la segmentation HMC robuste avec prise en compte de l'atlas.*

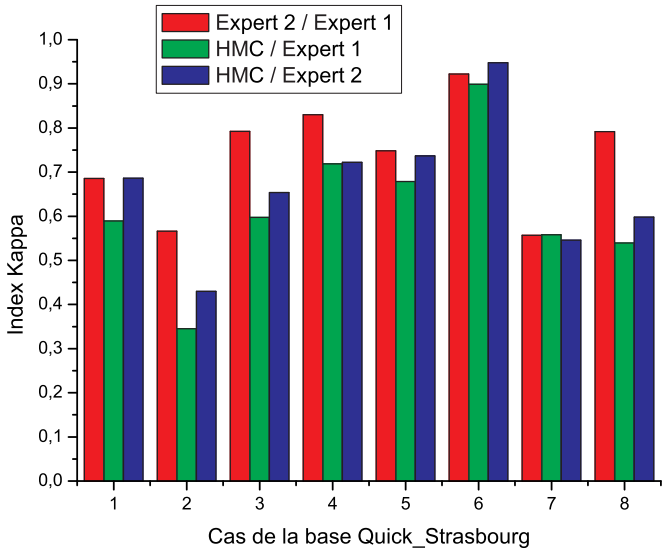
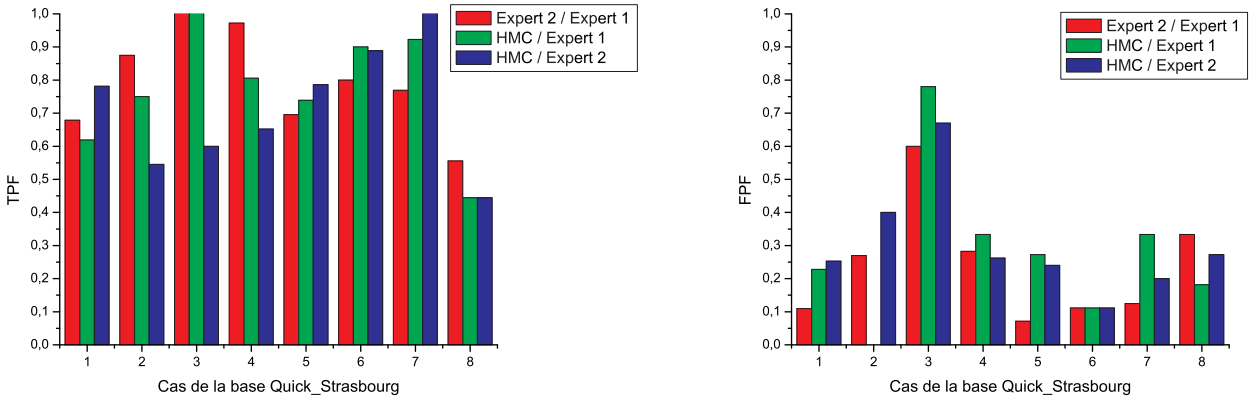
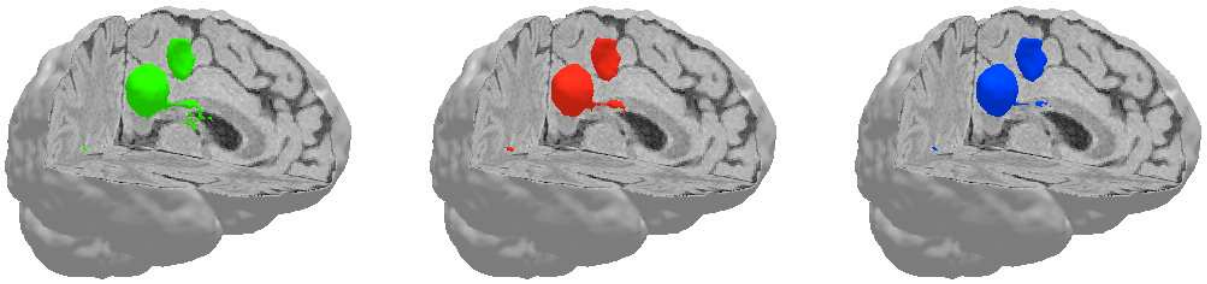


FIG. 4.8 – *Index Kappa obtenus sur chacun des 8 cas de la base de Strasbourg. Nous avons comparé nos résultats par rapport à chaque expert et les résultats des experts entre eux.*



(a) (b)

FIG. 4.9 – *Résultats obtenus sur chacun des 8 cas de la base de Strasbourg. (a), (b) correspondent respectivement au taux de vrais positifs TPF et au taux de faux positifs FPF.*



(a) (b) (c)

FIG. 4.10 – *Représentation 3D de résultats obtenus sur le cas 06 de la base de Strasbourg. (a), (b) et (c) correspondent respectivement aux segmentations de l'expert 1, de l'expert 2 et de la méthode HMC proposée.*

4.4.4 Base du challenge MICCAI'08

Les tests ont été effectués sur les images fournies lors du challenge MICCAI 2008[6]. Les images proviennent du Children's Hospital Boston (CHB) et de l'Université de Caroline du Nord (UNC). Les données UNC ont été acquises sur une IRM Siemens 3T Allegra avec une épaisseur de 1mm et une résolution planaire de 0.5mm. La base contient des images T1, T2 et Flair. Pour chaque cas, les données ont été recalées dans un référentiel commun en utilisant un recalage rigide et les voxels sont rendus isotropiques par une interpolation B-spline. La base est ainsi constituée d'images de taille $512 \times 512 \times 512$. Les données sont composées de 20 cas (10 UNC et 10 CHB) pour la base d'entraînement et de 25 cas (10 UNC et 15 CHB) pour la base test. Les différents cas ont été segmentés par un expert de chaque institution. Ces segmentations manuelles sont disponibles pour la base d'entraînement.

Les segmentations obtenues par notre méthode ont été effectuées à partir des deux modalités

[6] `http://www.ia.unc.edu/MSseg/index.php`

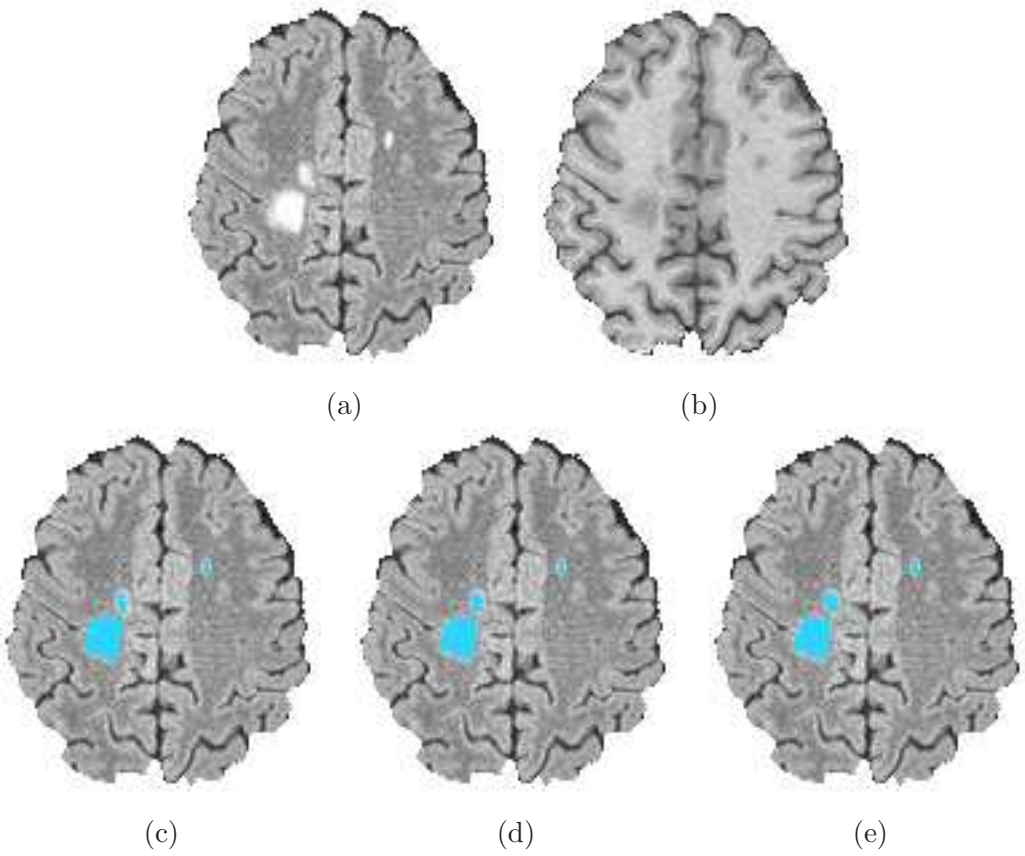


FIG. 4.11 – *Exemple de résultats obtenus sur le cas 06 de la base de Strasbourg. (a) et (b) correspondent respectivement aux images Flair et T1. (c), (d) et (e) correspondent respectivement aux segmentations de l'expert 1, de l'expert 2 et de la méthode proposée.*

T2 et Flair.

4.4.4.1 Base d'entraînement

La base d'entraînement est constituée de 10 cas UNC et de 10 cas CHB. Les segmentations des deux experts étant seulement disponibles pour les cas UNC, nous ne présentons dans ce paragraphe que les résultats obtenus sur la base UNC. Nous avons comparé les segmentations obtenues avec l'algorithme développé par rapport à la segmentation manuelle des différents experts (l'un provenant de UNC et l'autre de CHB), ainsi que les segmentations des deux experts l'une par rapport à l'autre. A cette fin, nous avons utilisé les mêmes métriques que pour la base de Strasbourg, à savoir l'index Kappa, le taux de vrais positifs TPF et le taux de faux positifs FPF.

Les index Kappa obtenus sur chacun des 10 cas de la base sont présentés sur la figure 4.12 et les taux de vrais et faux positifs sur la figure 4.13. Les résultats obtenus sur cette base sont nettement moins bons que ceux obtenus sur la base de Strasbourg présentés au paragraphe précédent. Mais on peut remarquer que la variabilité inter-expert est aussi nettement plus importante sur les images du challenge MICCAI. En effet, l'index Kappa moyen comparant les segmentations des deux experts est seulement de 0.28, alors que notre méthode obtient un index Kappa moyen de 0.3 par rapport à l'expert UNC et de 0.37 par rapport à l'expert CHB. Concernant les taux de vrais et faux positifs,

l'expert CHB obtient un TPF moyen de 0.68 et un FPF moyen de 0.66 par rapport à l'expert UNC. La méthode proposée obtient d'une part un TPF moyen de 0.69 et un FPF moyen de 0.54 par rapport à l'expert UNC, et d'autre part un TPF moyen de 0.55 et un FPF moyen de 0.24 par rapport à l'expert CHB. Dans l'ensemble, les segmentations obtenues par la méthode proposée semblent plus proches des segmentations manuelles effectuées par l'expert CHB. On peut en effet observer sur la figure 4.13.(b) que les taux de faux positifs de la méthode HMC par rapport à l'expert CHB sont très faibles. De plus, même si les résultats obtenus sur cette base sont nettement moins bons que ceux obtenus sur la base de Strasbourg, on peut cependant constater que les résultats obtenus par la méthode proposée sont moins éloignés des résultats des deux experts que les segmentations des deux experts entre eux. La variabilité des résultats entre les experts des deux institutions est en effet très importante ce qui constitue un des points faibles du challenge proposé dans le cadre de MICCAI'08 qui se tiendra à New York en septembre prochain.

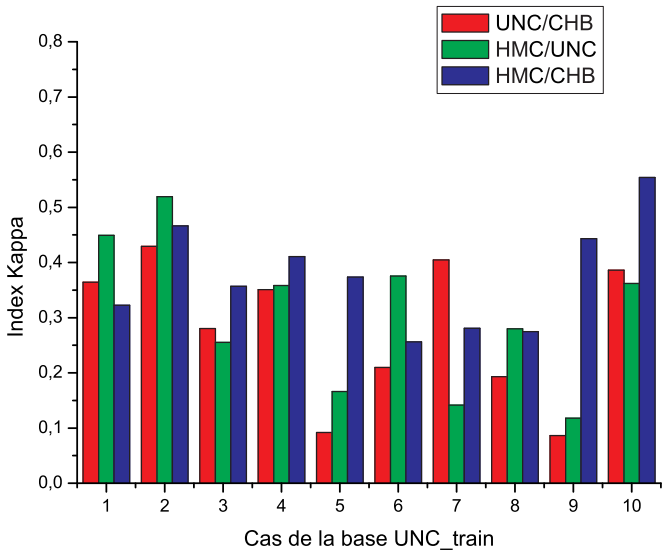


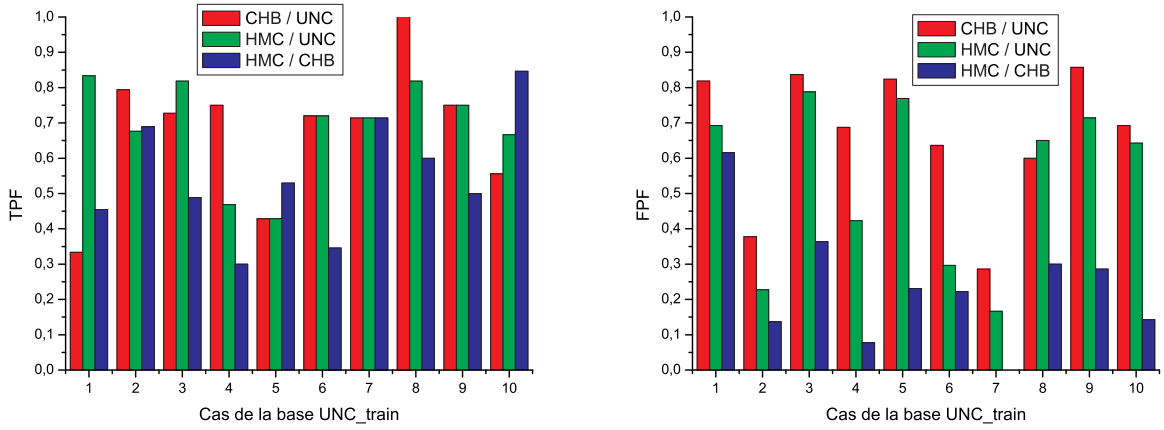
FIG. 4.12 – *Index Kappa obtenus sur chacun des 10 cas de la base d'entraînement UNC du challenge MICCAI. Nous avons comparé nos résultats par rapport à chaque expert (UNC et CHB) et les résultats des experts entre eux.*

Des exemples de résultats obtenus sur la base d'entraînement sont présentés sur les figures 4.14 et 4.15.

4.4.4.2 Base test

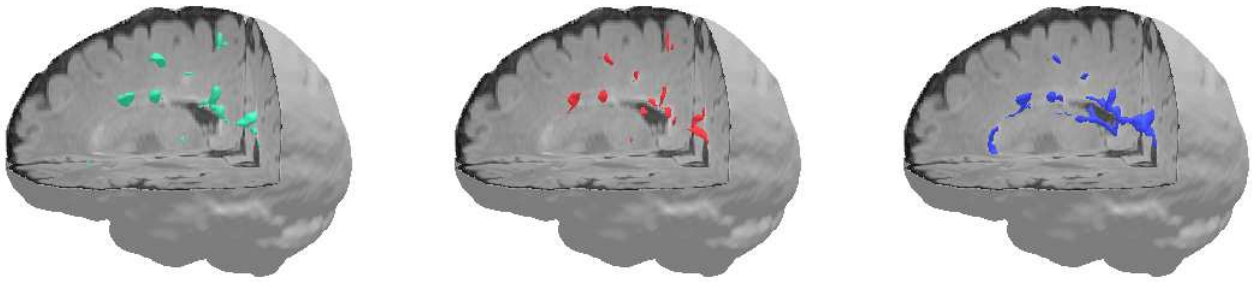
La base de test est constituée de 10 cas UNC et de 15 cas CHB. Différentes métriques ont été utilisées par les organisateurs du challenge pour évaluer le résultat de la segmentation des lésions :

- la différence de volumes (*Volume Diff.*) correspondant au pourcentage de la différence de volume par rapport à la segmentation de l'expert ;
- la distance moyenne (*Avg. Dist.*). Pour déterminer cette métrique, les voxels à la frontière de chaque lésion sont déterminés à la fois pour la segmentation automatique et pour la référence.



(a) (b)

FIG. 4.13 – *Résultats obtenus sur chacun des 10 cas de la base d'entraînement UNC du challenge MICCAI. (a), (b) correspondent respectivement au taux de vrais positifs TPF et au taux de faux positifs FPF.*



(a) (b) (c)

FIG. 4.14 – *Représentation 3D de résultats obtenus sur le cas 02 de la base d'entraînement UNC. (a), (b) et (c) correspondent respectivement aux segmentations de l'expert UNC, CHB et de la méthode HMC proposée.*

Ces voxels sont définis comme étant les voxels de l'objet (*i.e.* la lésion) ayant au moins un voisin (dans la 26-connexité) n'appartenant pas à cet objet. Pour chacun de ces voxels dans la segmentation automatique, le voxel le plus proche dans la référence est déterminé en utilisant la distance Euclidienne. La moyenne de ces différentes distances donne la distance moyenne. Cette valeur est nulle pour une segmentation parfaite ;

- le taux de vrais positifs TPF (*True Pos.*) ;
- le taux de faux positifs FPF (*False Pos.*).

Un score est attribué aux différentes métriques en fonction de la relation entre les segmentations des deux experts. Pour chaque métrique, un score de 100 est attribué à une segmentation parfaite. Un score de 90 pour chaque métrique égalera la performance entre les experts.

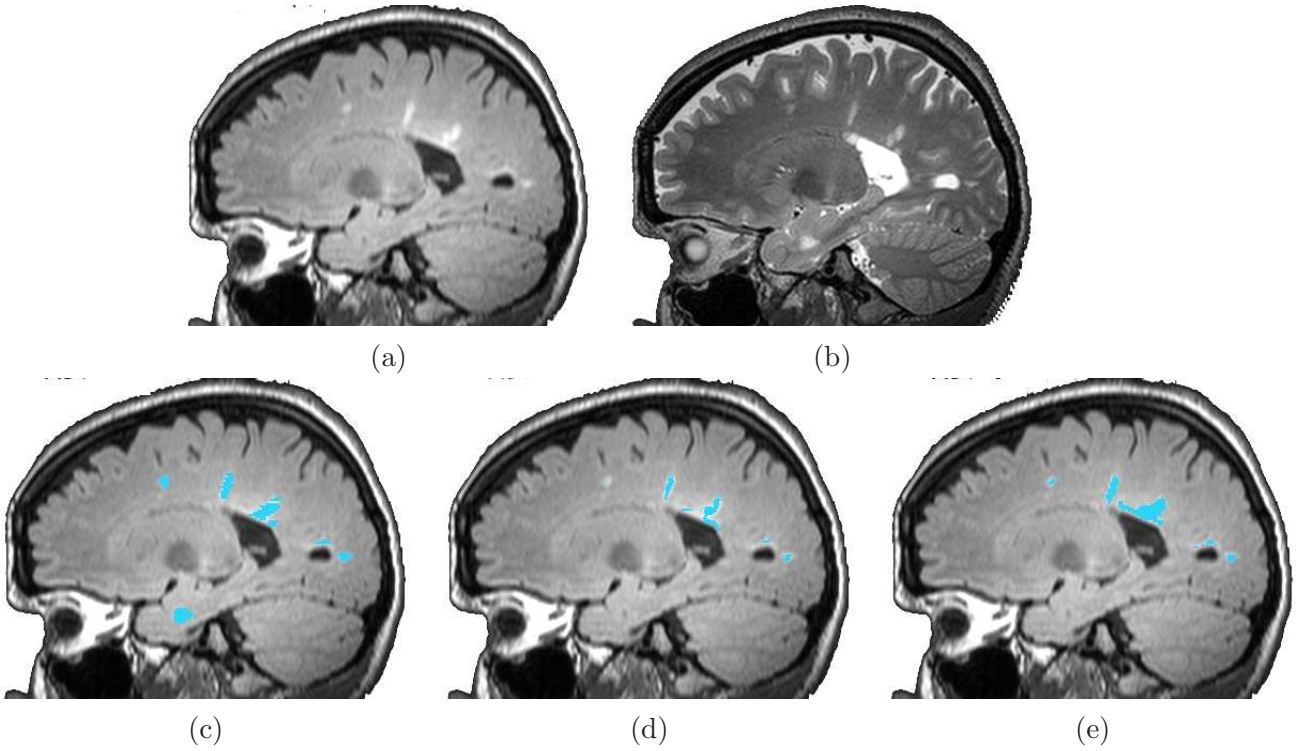


FIG. 4.15 – *Exemple de résultats obtenus sur le cas 02 de la base d'entraînement UNC. (a) et (b) correspondent respectivement aux images Flair et T2. (c), (d) et (e) correspondent respectivement aux segmentations de l'expert UNC, de l'expert CHB et de la méthode proposée.*

Les résultats obtenus sur les deux bases tests sont présentés dans le tableau 4.16. Ces résultats sont fournis par les organisateurs du challenge. Les segmentations manuelles effectuées par les experts ne sont pas disponibles pour les données des deux bases tests ce qui limite actuellement nos possibilités d'investigation.

On peut constater que les scores obtenus pour la différence de volume et la distance moyenne sont proches de 90 par rapport aux deux experts. Les scores obtenus pour les taux de vrais et faux positifs sont de l'ordre de 75. On peut cependant remarquer la forte variabilité entre les experts. Par exemple pour le cas 04 de la base UNC, la différence de volume de la segmentation proposée est de 7.1% par rapport à l'expert UNC et de 142% par rapport à l'expert CHB. De plus pour le cas 09 de la base UNC, le taux de vrais positifs est de seulement 33.3% par rapport à l'expert UNC alors qu'il est de 100% par rapport à l'expert CHB. Ces différents résultats soulignent la difficulté de la validation.

Des exemples de résultats obtenus sur les bases tests sont présentés sur les figures 4.17 et 4.18.

Remarque : Le workshop du challenge MICCAI sur la segmentation des lésions SEP ayant lieu le 6 septembre 2008, les résultats obtenus par les autres méthodes ayant participé au challenge ne sont pour l'instant pas connus. Par conséquent, nous n'avons pas pu comparer nos résultats par rapport aux autres méthodes.

Ground Truth	UNC Rater								CHB Rater								
All Dataset	Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Total
	[%]	Score	[mm]	Score	[%]	Score	[%]	Score	[%]	Score	[mm]	Score	[%]	Score	[%]	Score	
UNC test1 Case01	62.9	91	13.5	72	25.6	66	73.8	65	45.7	93	14.5	70	28.1	67	66.7	69	74
UNC test1 Case02	105.2	85	5.9	88	33.8	71	60.5	73	72.9	89	3.2	93	18.2	62	14.0	100	83
UNC test1 Case03	58.4	91	2.9	94	24.6	65	29.7	92	46.3	93	2.2	95	27.2	67	20.3	97	87
UNC test1 Case04	30.2	96	3.2	93	47.4	78	33.3	89	8.3	99	1.7	96	63.0	87	45.5	82	90
UNC test1 Case05	7.1	99	2.9	94	52.4	81	58.1	74	142.0	79	3.6	92	73.9	93	75.8	63	85
UNC test1 Case06	64.3	91	4.8	90	48.3	79	54.7	76	59.4	91	16.7	66	62.5	87	79.2	61	80
UNC test1 Case07	38.4	94	2.3	95	47.5	78	32.7	90	42.8	94	2.9	94	70.0	91	55.8	76	89
UNC test1 Case08	51.5	92	2.9	94	48.9	79	45.2	82	20.7	97	2.3	95	83.3	99	57.1	75	89
UNC test1 Case09	0.9	100	25.0	49	33.3	70	93.8	53	39.8	94	28.4	42	100.0	100	93.8	53	70
UNC test1 Case10	43.9	94	16.3	66	20.0	63	84.0	58	420.3	38	18.0	63	50.0	80	88.0	56	65
CHB test1 Case01	78.6	88	6.6	86	18.7	62	16.0	100	69.4	90	4.3	91	41.9	75	28.0	93	86
CHB test1 Case02	27.5	96	6.4	87	31.8	70	76.5	63	69.1	90	4.7	90	42.1	75	38.2	86	82
CHB test1 Case03	41.8	94	5.5	89	57.1	84	65.2	70	71.9	89	7.1	85	46.7	78	69.6	67	82
CHB test1 Case04	56.3	92	5.2	89	63.6	88	47.8	80	79.0	88	11.3	77	50.0	80	13.0	100	87
CHB test1 Case05	60.3	91	10.5	78	29.6	68	94.9	52	69.6	90	3.7	92	47.8	79	57.3	75	78
CHB test1 Case06	51.1	93	4.5	91	55.6	83	96.7	51	57.9	92	4.6	91	36.4	72	96.9	51	78
CHB test1 Case07	62.1	91	9.5	80	30.0	69	77.0	63	76.9	89	3.8	92	34.2	71	36.3	88	80
CHB test1 Case08	40.5	94	2.8	94	66.7	89	34.5	89	60.2	91	4.9	90	47.1	78	20.7	97	90
CHB test1 Case09	16.2	98	3.2	93	32.9	70	58.3	74	29.3	96	2.6	95	24.1	65	24.0	95	86
CHB test1 Case10	55.5	92	4.8	90	52.6	81	63.8	71	78.3	89	6.5	87	34.5	71	48.3	80	83
CHB test1 Case11	119.2	83	6.0	88	34.1	71	95.2	52	29.1	96	2.0	96	37.9	73	67.8	68	78
CHB test1 Case12	75.4	89	4.3	91	20.5	63	51.6	78	75.5	89	4.4	91	28.2	68	49.8	79	81
CHB test1 Case13	42.4	94	5.8	88	40.0	74	66.7	69	64.7	91	4.0	92	33.3	70	26.7	93	84
CHB test1 Case15	42.1	94	3.8	92	32.9	70	49.1	80	23.7	97	2.3	95	40.4	74	52.6	78	85
All Average	51.3	92	6.6	86	39.5	74	60.8	73	73.0	89	6.7	86	46.7	78	51.1	78	82
All UNC	46.3	93	8.0	84	38.2	73	56.6	75	89.8	87	9.4	81	57.6	83	59.6	73	81
All CHB	54.9	92	5.6	88	40.4	74	63.8	71	61.0	91	4.7	90	38.9	74	44.9	82	83

FIG. 4.16 – *Résultats obtenus sur les deux bases tests du challenge MICCAI'08. Ce tableau est fourni par les organisateurs de challenge MICCAI'08. Pour chaque cas les résultats obtenus par notre méthode sont comparés aux segmentations manuelles de l'expert UNC et de l'expert CHB. Les critères de comparaison sont les suivants : différence de volumes (Volume Diff.), la distance moyenne (Avg. Dist.), le taux de vrais positifs (True Pos.) et le taux de faux positifs (True Pos.). Pour chaque critère un score est attribué. La moyenne de ce score est présentée pour chaque cas ainsi que pour chaque critère.*

4.5 Conclusion

Dans ce chapitre nous avons étendu la méthode proposée au chapitre 3 à la détection de lésions SEP. L'utilisation d'un estimateur robuste a permis de détecter les voxels atypiques. Différentes contraintes et post-traitements ont ensuite été appliqués sur ces outliers afin d'obtenir seulement les lésions SEP. L'intérêt de la méthode proposée consiste à prendre en compte l'information sur le voisinage. De plus, l'utilisation de l'atlas a permis d'améliorer nettement les résultats. Cette méthode nécessite cependant que le seuil de détection soit préalablement fixé. Des tests sur des images synthétiques ont permis de choisir une valeur permettant de bien détecter les lésions tout en limitant le nombre de faux positifs. Des tests sur différentes bases réelles ont été effectués pour valider cette méthode en utilisant différentes métriques pour quantifier les résultats obtenus

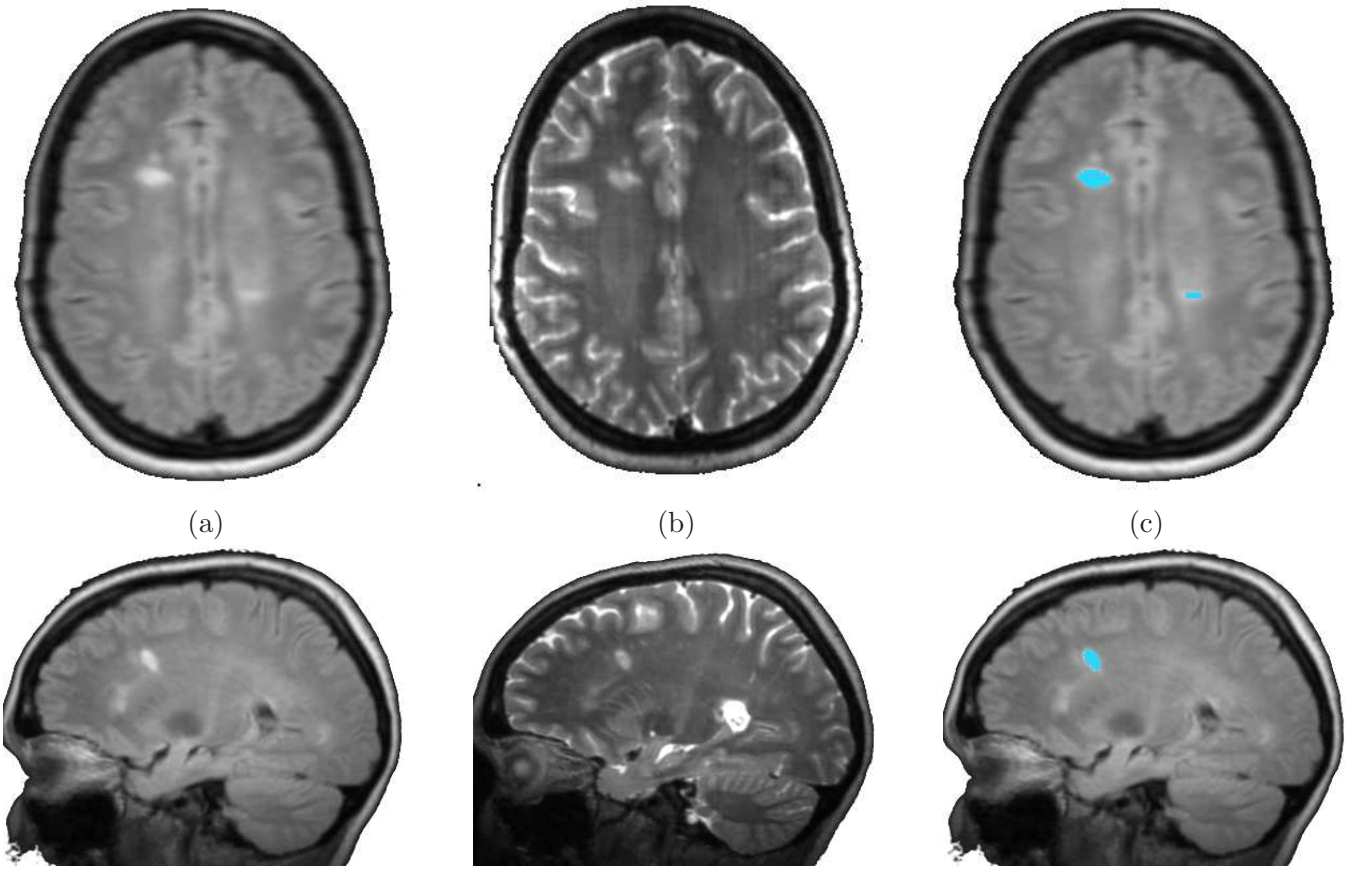


FIG. 4.17 – *Exemple de résultats obtenus sur le cas 02 de la base test CHB. (a) et (d) correspondent à l'image Flair, (b) et (e) à l'image T2, (c) et (f) aux lésions segmentées par la méthode HMC proposée.*

par rapport aux segmentations manuelles de deux experts. Cependant les tests effectués sur les différentes bases réelles, en particulier celles fournies lors du challenge MICCAI'08, ont montré la variabilité inter-expert. Avoir un plus grand nombre de segmentations de différents experts pourrait nous aider à établir un étalon or à partir d'un algorithme de type STAPLE [Warfield 04].

Une perspective d'amélioration de cette méthode consisterait à tester d'autres modélisations du bruit. En effet l'approximation gaussienne du bruit n'étant pas valide dans le LCR, il a été nécessaire de mettre en œuvre des contraintes d'intensité sur les outliers pour ne pas biaiser l'estimation des paramètres. L'utilisation de gaussiennes généralisées pourrait par exemple être envisagée.

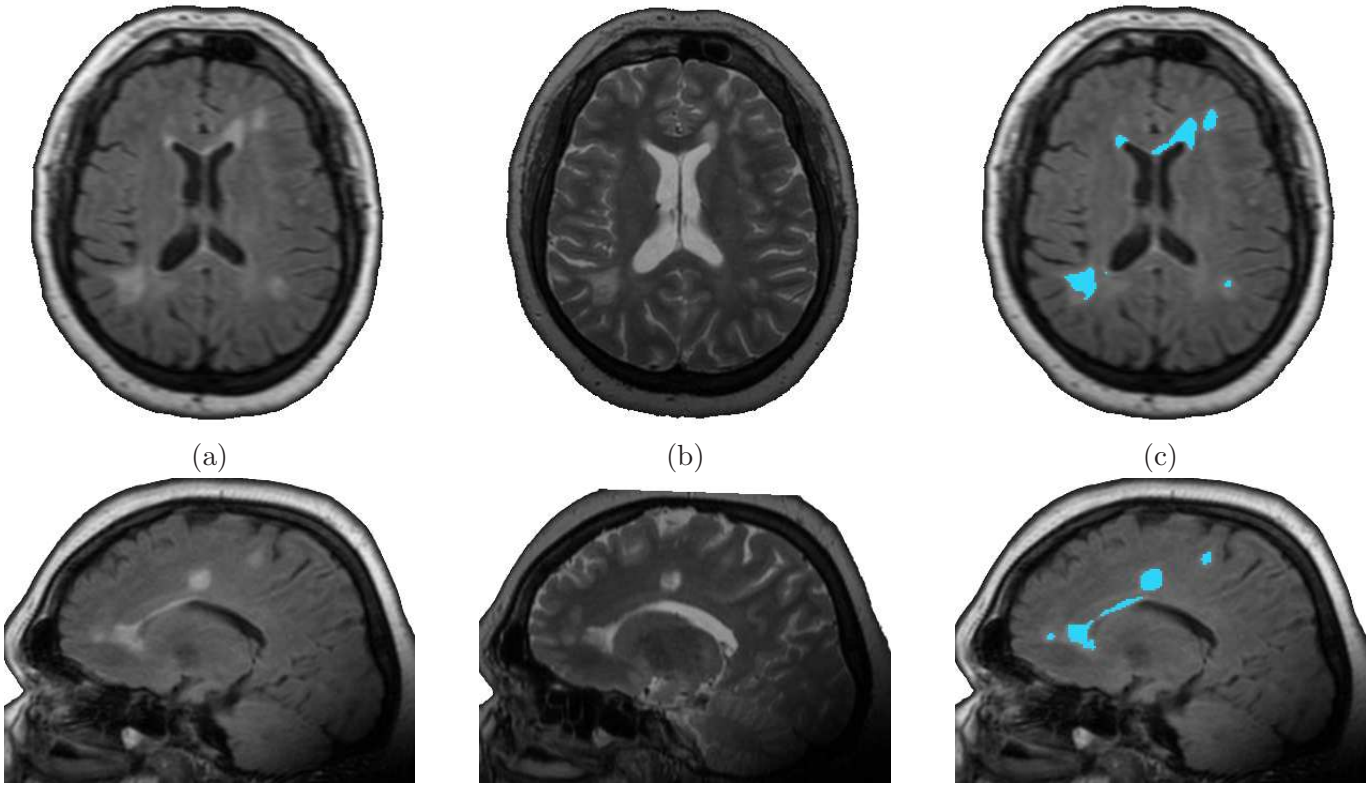


FIG. 4.18 – *Exemple de résultats obtenus sur le cas 04 de la base test UNC.(a) et (d) correspondent à l'image Flair, (b) et (e) à l'image T2, (c) et (f) aux lésions segmentées par la méthode HMC proposée.*

Chapitre 5

Contours actifs statistiques 3D

Dans le chapitre précédent, nous avons présenté une méthode automatique de détection des lésions SEP sur des IRM multimodales. Après la détection de ces lésions, l'objectif était d'obtenir une segmentation plus précise des lésions pour quantifier plus facilement leur évolution. De plus le deuxième objectif consistait à segmenter des structures anatomiques comme par exemple les ventricules ou le corps calleux et non plus les tissus cérébraux (MB, MG, LCR). A cette fin nous proposons une méthode de segmentation basée sur les contours actifs statistiques 3D. La section 5.1 présente brièvement les méthodes de segmentation par contours actifs. La section 5.2 introduit la notion de complexité stochastique en explicitant le principe de minimisation de la complexité stochastique et les méthodes de segmentations existantes basées sur ce principe. Nous détaillons ensuite la méthode de segmentation proposée basée sur les contours actifs statistiques 3D (Section 5.3). Enfin la section 5.4 présente quelques résultats obtenus sur des images synthétiques et réelles.

5.1 Contours actifs

Les contours actifs ont été introduits en 1988 par Kass [Kass 88]. Le principe est le suivant, il s'agit de faire évoluer une courbe paramétrée $C(v(s) = [x(s), y(s)]^t, s \in [0, 1])$ en fonction de contraintes internes ou externes.

$$E_{totale} = \int_0^1 [E_{interne}(v(s)) + E_{image}(v(s)) + E_{ext}(v(s))]ds \tag{5.1}$$

avec

$$E_{interne} = \alpha(s) \left(\frac{dv}{ds} \right)^2 + \beta(s) \left(\frac{d^2v}{ds^2} \right)^2 \tag{5.2}$$

De nombreuses extensions à ce modèle de base ont été proposées comme par exemple l'introduction d'une force ballon [Cohen 92]. Il s'agit d'une force normale au contour en chaque point. Cette force va tendre à gonfler le contour actif ou à accélérer sa rétraction selon le signe de la force introduite. D'autres extensions ont été développées comme les "gradient vector flow" introduits par Xu [Xu 98] ou l'utilisation d'ensemble de niveaux [Malladi 95, Vese 02]. Storvik a proposé un cadre totalement bayésien avec un modèle markovien pour la régularisation [Storvik 94]. Il introduit également un critère de vraisemblance à l'intérieur et à l'extérieur de ce contour (et non

plus sur le contour). Cette approche "région" des contours actifs a été développée par l'équipe de P. Réfrégier à l'institut Fresnel à Marseille dans le cas du contour actif statistique polygonal (CASP) puis des grilles actives statistiques [Galland 03, Galland 04].

5.2 Complexité stochastique et segmentation

5.2.1 Principe de minimisation de la complexité stochastique

Rissanen a introduit en 1978 le principe de minimisation de la complexité stochastique (MCS) basé sur la théorie de l'information et de l'estimation statistique [Rissanen 78]. Ce principe est dénommé principe de la longueur de description minimale (MDL[1]). La longueur de description correspond en fait à la complexité de Kolmogorov [Cover 91]. La complexité de Kolmogorov $C(\chi)$ d'une suite de nombres $\chi = x_1, x_2, \ldots, x_n$ de longueur n est égale à la longueur du plus petit programme (mesurée en bits) permettant de l'engendrer. La complexité stochastique consiste à remplacer dans la complexité de Kolmogorov le nombre de bits nécessaires pour coder la suite par un code entropique auquel il faut ajouter le nombre de bits nécessaires pour décrire le modèle probabiliste permettant de déterminer ce code. La complexité stochastique est dénommée ainsi par opposition à la complexité *algorithmique* de Kolmogorov.

La complexité stochastique permet de définir une mesure de la complexité intégrant un terme relatif au modèle sous-jacent aux données. Le principe de la MCS propose un critère pour choisir le meilleur modèle parmi un ensemble de modèles sans aucune connaissance *a priori* sur le véritable modèle sous-jacent [Rissanen 89]. Dans la suite du manuscrit nous emploierons plutôt le terme de principe de minimisation de la complexité stochastique. Un des intérêts de ce principe est de permettre la résolution des problèmes d'estimation lorsque l'ordre du modèle est inconnu. Il peut être appliqué à la segmentation d'images.

5.2.2 Méthodes de segmentation existantes basées sur le principe de minimisation de la complexité stochastique

Différentes méthodes de segmentation d'images 2D basées sur le principe de minimisation de la complexité stochastique (MCS) ont été développées. La première application a été proposée par Leclerc en 1989 pour segmenter des images optiques, la description du contour actif étant faite pixel par pixel [Leclerc 89]. Kamungo *et al.* ont proposé une méthode de segmentation d'images multi-bandes entièrement non supervisée basée sur la MCS et sur des fusions de régions [Kanungo 94]. Zhu et Yuille proposent une interprétation bayésienne du principe de la MCS avec un algorithme de segmentation combinant dilatations de régions, fusions de régions et contours actifs [Zhu 96]. Cependant cette méthode nécessite le réglage de plusieurs paramètres de pondération.

Pour ces différentes méthodes, la minimisation de la complexité stochastique permet d'estimer le nombre de régions présentes dans l'image. Elle peut aussi conduire à l'estimation de l'ordre du modèle utilisé pour décrire le contour d'un objet. Ainsi Figueiredo *et al.* ont proposé un algorithme non supervisé pour segmenter un objet unique [Figueiredo 00]. Le principe de la MCS permet d'estimer l'ordre du modèle utilisé pour décrire le contour, c'est-à-dire le nombre de points de contrôle des splines du contour. Cette méthode a ensuite été généralisée par Ruch *et*

[1] *Minimum Description Length* en anglais

al. aux contours polygonaux [Ruch 01]. Ce modèle a ensuite été étendu par Galland aux grilles actives statistiques [Galland 04]. La méthode de partition d'images qu'il propose combine à la fois l'estimation du nombre de régions présentes dans l'image et l'ordre du modèle paramétrique nécessaire pour décrire les contours des différentes régions.

Ces différentes méthodes ont seulement été appliquées à la segmentation d'images 2D. Dans les paragraphes suivants nous allons proposer une méthode de segmentation d'objets 3D à partir d'images mono- ou multimodales basée sur le principe de la MCS en étendant au cas 3D les méthodes par contours actifs et grilles actives statistiques proposées par Ruch et Galland dans le cas 2D.

5.3 Contours actifs statistiques 3D

5.3.1 Calcul de la complexité stochastique

5.3.1.1 Modèle d'image

Considérons une image $\mathbf{s} = \{s(x,y,z)|(x,y,z) \in N_x \times N_y \times N_z\}$ contenant $N = N_x \times N_y \times N_z$ voxels et où $s(x,y,z)$ représente le niveau de gris du voxel de coordonnées (x,y,z). Nous supposons que cette image est composée de deux régions Ω_a et Ω_b correspondant respectivement à l'objet cible et au fond. Chaque région $\Omega_{l,l\in\{a,b\}}$ est supposée homogène, *i.e.,* tous les voxels de la même région Ω_l sont distribués selon la même densité de probabilité (ddp) P_{θ_l} où θ_l est le vecteur de paramètres de la ddp dans la région Ω_l. Dans notre cas, nous supposerons que les niveaux de gris des voxels dans chacune des régions Ω_l sont distribués selon une loi gaussienne. $\theta_l = (\mu_l, \sigma_l)$ où μ_l et σ_l sont respectivement la moyenne et l'écart-type dans la région Ω_l. Dans la suite du manuscrit, nous noterons $\boldsymbol{\theta} = (\theta_a, \theta_b)$ l'ensemble des deux vecteurs de paramètres.

La partition de l'image peut être décrite à l'aide de la fonction de partition binaire w : $w = \{w(x,y,z)|(x,y,z) \in N_x \times N_y \times N_z\}$ avec $w(x,y,z) = 1$ à l'intérieur de l'objet cible et $w(x,y,z) = 0$ ailleurs.

5.3.1.2 Détermination de la complexité stochastique

L'objectif de la segmentation consiste à estimer la forme $\mathbf{w}$ de la cible dans la scène. Nous avons choisi de définir la surface de la forme par un maillage triangulaire composé de k triangles modélisant le contour entre l'objet cible et le fond. Pour estimer le nombre de triangles k définissant la forme, nous utilisons le principe de minimisation de la complexité stochastique introduit par Rissanen [Rissanen 89]. La complexité stochastique est une approximation de la longueur minimale de code Δ nécessaire pour décrire l'image et s'exprime en bits ou nats (1 nat = $\log 2$ bits). Dans la suite du manuscrit, nous exprimerons cette complexité en nats pour pouvoir utiliser le logarithme népérien (log) à la place du logarithme en base 2 ($\log_2$). L'image considérée est composée de trois parties : le fond, la cible et le contour. Par conséquent Δ sera la somme des trois termes suivants :

- la longueur Δ_a nécessaire pour décrire la distribution des niveaux de gris de la cible ;
- la longueur Δ_b nécessaire pour décrire la distribution des niveaux de gris du fond ;
- la longueur Δ_{w^k} requise pour la description du maillage surfacique composé de k triangles

La complexité stochastique se décompose donc en deux termes : l'un correspondant à la longueur de code associée au modèle de l'image et l'autre correspondant à la longueur de code associée au codage des données connaissant le modèle. Soit $\Delta(\mathbf{s}, w, \boldsymbol{\theta})$ la complexité stochastique nécessaire

pour décrire l'image $\mathbf{s}$ associée à la partition ω et aux deux densités de probabilités définies par $\boldsymbol{\theta}$. Nous avons :

$$\Delta(\mathbf{s}, w, \boldsymbol{\theta}) = \Delta_G(w) + \Delta_L(\mathbf{s}|\boldsymbol{\theta}, w) \tag{5.3}$$

avec

- $\Delta_G(w)$ le nombre de nats nécessaires pour coder la partition w

$$\Delta_G(w) = \Delta_{w^k} \tag{5.4}$$

- $\Delta_L(\mathbf{s}|\boldsymbol{\theta}, w)$ le terme de codage de l'image connaissant le modèle, *i.e.* le terme de codage des niveaux de gris des voxels dans chacune des deux régions Ω_a et Ω_b.

$$\Delta_L(\mathbf{s}|\boldsymbol{\theta}, w) = \Delta_a + \Delta_b \tag{5.5}$$

Nous allons maintenant détailler le calcul de ces différents termes.

Terme de codage du maillage surfacique Il faut déterminer une approximation de la plus petite longueur de code nécessaire pour décrire une partition w quelconque. La partition w est décrite à l'aide d'un maillage triangulaire composé de k triangles. Pour coder un triangle, il suffit d'encoder le sommet de départ et les segments composant le triangle. Comme l'image contient N voxels, la localisation d'un sommet peut être codée par $\log N$ nats. Pour coder le maillage triangulaire, il faut donc coder les k sommets de départ et les p vecteurs associés aux segments, où p représente le nombre total de segments. Les k sommets de départ peuvent être codés avec $k \log N$ nats.

Nous allons maintenant détailler le codage des p vecteurs associés aux segments. Une première approximation de ce codage peut être effectuée en $p \log N$ nats. En effet pour définir tous les segments, il suffit de donner la liste des coordonnées des nœuds dans l'ordre où on les rencontre en parcourant le maillage. Si on suppose que toutes les positions de nœuds sont équiprobables et indépendantes de la position des nœuds auxquels ils sont reliés, chaque coordonnée de nœuds peut être codée en $\log N$ nats, d'où les $p \log N$ nats pour coder les p segments. Cependant cette approximation ne constitue pas une bonne estimation du nombre minimal de nats nécessaires pour le codage des p vecteurs car elle suppose la répartition des nœuds uniforme dans l'image indépendamment de la position des nœuds auxquels ils sont reliés, ce qui n'est pas le cas en pratique.

Une meilleure approximation du codage des coordonnées des vecteurs peut être obtenue en utilisant un code entropique de Shannon [Shannon 48]. Pour déterminer ce code, il est nécessaire de connaître la distribution associée aux vecteurs. Soit $d_u(i)$ les coordonnées selon u du segment numéro i : $d_u(i) = u_2(i) - u_1(i)$, où $u = x$ ou y ou z, et $(x_1(i), y_1(i), z_1(i))$, $(x_2(i), y_2(i), z_2(i))$ sont les coordonnées des extrémités du segment numéro i. Supposons que $d_u(i)$ est distribué avec une ddp $P_{m_u}(d)$ de paramètre m_u. Si le paramètre m_u de la ddp $P_{m_u}(d)$ est connu, la longueur de code moyenne requise pour coder les composantes (selon x, y, z) des p vecteurs peut être approximée par l'opposée de la log-vraisemblance $\mathcal{L}_e[\chi_u|m_u]$ où $\chi_u = (d_u(i))_{i\in[1,p]}$. La longueur de code minimale nécessaire pour décrire le maillage est donc :

$$\Delta_G(w) = k \log N - \mathcal{L}_e[\chi_x|m_x] - \mathcal{L}_e[\chi_y|m_y] - \mathcal{L}_e[\chi_z|m_z] + \Delta(m_x) + \Delta(m_y) + \Delta(m_z) \tag{5.6}$$

où $\chi_x = (d_x(i))_{i\in[1,p]}$, $\chi_y = (d_y(i))_{i\in[1,p]}$, $\chi_z = (d_z(i))_{i\in[1,p]}$, et où $\Delta(m_x)$, $\Delta(m_y)$ et $\Delta(m_z)$ correspondent aux nombres de nats nécessaires pour coder les paramètres m_x, m_y et m_z.

Les paramètres $m_{u,u\in\{x,y,z\}}$ sont inconnus, il faut donc les remplacer par leurs estimées $\hat{m}_{u,u\in\{x,y,z\}}$ au sens du maximum de vraisemblance (MV). D'après [Rissanen 89], $m_{u,u\in\{x,y,z\}}$ étant estimé sur un échantillon de taille p, il peut être encodé avec $\log\sqrt{p}$ nats. On obtient donc :

$$\Delta_G(w) = k\log N - \mathcal{L}_e[\chi_x|\hat{m}_x] - \mathcal{L}_e[\chi_y|\hat{m}_y] - \mathcal{L}_e[\chi_z|\hat{m}_z] + (3/2) * \log p \tag{5.7}$$

Pour déterminer la ddp $P_{m_{u,u\in\{x,y,z\}}}(d_u)$ pour les différentes composantes des segments, nous appliquons le principe du maximum d'entropie [Galland 03]. Nous supposons de plus que les longueurs moyennes des projections des vecteurs selon x, y, z sont connues, c'est-à-dire les moyennes m_x, m_y et m_z de $|d_x|$, $|d_y|$ et $|d_z|$. Il faut donc déterminer la ddp P maximisant l'entropie $S = -\int_{-\infty}^{+\infty} P(x)\log P(x)dx$ sous la contrainte que la valeur absolue de la moyenne est égale à $m : \int_{-\infty}^{+\infty} |x|P(x)dx = m$. La ddp correspondante est une loi de Laplace, nous obtenons donc :

$$P_{m_u}(d_u) = \frac{1}{2m_u} e^{\frac{-|d_u|}{m_u}} \tag{5.8}$$

D'où :

$$\mathcal{L}_e[\chi|m] = \sum_{i=1}^{p} \log[P_m(d(i))] \tag{5.9}$$

En remplaçant m par son estimée au sens du maximum de vraisemblance $\hat{m} = (1/p)\sum_{i=1}^{p} d(i)$, nous obtenons :

$$\mathcal{L}_e[\chi|m] = -p\log(2\hat{m}) - p \tag{5.10}$$

Le terme de codage du maillage peut être déduit de l'équation 5.7 :

$$\Delta_G(w) = k\log N + (3/2) * \log p + p(3 + \log(2\hat{m}_x) + \log(2\hat{m}_y) + \log(2\hat{m}_z)) \tag{5.11}$$

Terme de codage des niveaux de gris $\Delta_L(\mathbf{s}|\boldsymbol{\theta}, w)$ Nous devons ensuite déterminer la longueur de code nécessaire pour décrire les niveaux de gris des voxels de l'image avec un codage entropique dans chacune des régions Ω_a et Ω_b. Considérons dans un premier temps la longueur de code $\Delta_l(\theta_l)$ nécessaire pour coder les niveaux de gris des voxels de la région $\Omega_{l,l\in\{a,b\}}$ connaissant le vecteur de paramètres θ_l.

D'après Shannon, le nombre moyen de bits nécessaires pour décrire une suite d'événements aléatoires correspond à l'entropie de la ddp [Shannon 48]. Ainsi le nombre moyen de bits $\delta_{l,l\in\{a,b\}}$ requis pour décrire N_l variables aléatoires distribuées avec une ddp $P_{\Theta_l}(x)$ est égale à $\Delta_l = N_l * S_l$, l'entropie S_l étant donnée par $S_l = -\int P_{\Theta_l}(x)\log_2[P_{\Theta_l}(x)]dx - \log_2(q)$ où q correspond au pas de quantification. Comme la contribution de q consiste seulement à ajouter un terme constant à la longueur de description, nous ne le prendrons pas en compte par la suite. La longueur de code moyenne (en nats) pour un codage entropique des niveaux de gris dans la région Ω_l est donc donnée par [Shannon 48] :

$$\Delta_l = -\sum_{(x,y,z)\in\Omega_l} \log P_{\Theta_l}(s(x,y,z)) \tag{5.12}$$

avec $l = a$ ou b. Ainsi $\Delta_a + \Delta_b$ est égale à l'opposée de la log-vraisemblance de l'image $\mathcal{L}_e[\mathbf{s}|w, \theta_a, \theta_b]$.

$$\Delta_L(\mathbf{s}|\boldsymbol{\theta}, w) = -\mathcal{L}_e[\mathbf{s}|w, \theta_a, \theta_b] = -\mathcal{L}_e[\Omega_a|\theta_a] - \mathcal{L}_e[\Omega_b|\theta_b] \tag{5.13}$$

avec l'opposée de la log-vraisemblance $\mathcal{L}_e[\Omega_l|\theta_l]$ dans la région Ω_l (où $l \in \{a, b\}$) :

$$\mathcal{L}_e[\Omega_l|\theta_l] = \sum_{(x,y,z)\in\Omega_l} \log P_{\theta_l}(s(x,y,z)) \tag{5.14}$$

$\mathcal{L}_e[\Omega_l|\theta_l]$ dépend du vecteur de paramètre θ_l de la ddp. θ_l étant inconnu, nous devons l'estimer. Soit $\hat{\theta}_l$ l'estimée de θ_l au sens du maximum de vraisemblance. En remplaçant θ_l par $\hat{\theta}_l$, on obtient la log-vraisemblance généralisée $\mathcal{L}_e[\Omega_l|\hat{\theta}_l]$ et l'équation 5.13 devient :

$$\Delta_L(\mathbf{s}|\boldsymbol{\theta}, w) = -\mathcal{L}_e[\Omega_a|\hat{\theta}_a] - \mathcal{L}_e[\Omega_b|\hat{\theta}_b] \tag{5.15}$$

Nous supposons que les niveaux de gris des voxels dans chacune des régions Ω_l sont distribués selon une loi gaussienne. La ddp de cette loi s'écrit :

$$P_{\theta_l}(x) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp^{-\frac{1}{2}\left(\frac{x_l-\mu_l}{\sigma_l}\right)^2} \tag{5.16}$$

où μ_l et σ_l sont respectivement la moyenne et l'écart-type dans la région Ω_l. La log-vraisemblance dans la région Ω_l s'écrit donc :

$$\begin{aligned} \mathcal{L}_e[\Omega_l|\theta_l] &= \sum_{(x,y,z)\in\Omega_l} \log P_{\theta_l}(s(x,y,z)) & (5.17)\\ &= -\frac{N_l}{2}\log(2\pi) - N_l \log\sigma_l - \frac{1}{2\sigma_l^2} \sum_{(x,y,z)\in\Omega_l} \left(s(x,y,z) - \mu_l\right)^2 & (5.18) \end{aligned}$$

avec N_l le nombre de voxels dans la région Ω_l.

En remplaçant θ_l par $\hat{\theta}_l$, on obtient la log-vraisemblance généralisée $\mathcal{L}_e[\Omega_l|\hat{\theta}_l]$:

$$\mathcal{L}_e[\Omega_l|\hat{\theta}_l] = -\frac{N_l}{2}\log(2\pi) - N_l \log\hat{\sigma}_l - \frac{1}{2}N_l \tag{5.19}$$

$\hat{\mu}_l$ et $\hat{\sigma}_l^2$ sont la moyenne et la variance empiriques calculées sur la région Ω_l :

$$\begin{aligned} \hat{\mu}_l &= \frac{1}{N_l} \sum_{(x,y,z)\in\Omega_l} s(x,y,z) & (5.20)\\ \hat{\sigma}_l^2 &= \frac{1}{N_l} \sum_{(x,y,z)\in\Omega_l} \left(s(x,y,z) - \hat{\mu}_l\right)^2 & (5.21) \end{aligned}$$

Le terme de codage des niveaux de gris s'écrit donc :

$$\Delta_L(\mathbf{s}|\hat{\boldsymbol{\theta}}, w) = \frac{1}{2}N_a \log\hat{\sigma}_a^2 + \frac{1}{2}N_b \log\hat{\sigma}_b^2 + K \tag{5.22}$$

avec $N_{l,l\in\{a,b\}}$ le nombre de voxels dans la région $\Omega_{l,l\in\{a,b\}}$ et $\hat{\sigma}^2_{l,l\in\{a,b\}}$ la variance empirique dans la région $\Omega_{l,l\in\{a,b\}}$ et K une constante indépendante de la partition.

5.3.1.3 Minimisation de la complexité stochastique

A partir des deux termes $\Delta_G(w)$ et $\Delta_L(\mathbf{s}|\hat{\boldsymbol{\theta}}, w)$ issus des équations 5.11 et 5.22, nous pouvons en déduire la complexité stochastique $\Delta(\mathbf{s}, w, \hat{\boldsymbol{\theta}})$ associée à une partition w donnée :

$$\begin{aligned}\Delta(\mathbf{s}, w, \hat{\boldsymbol{\theta}}) &= k \log N + (3/2) * \log p + p(3 + \log(2\hat{m}_x) + \log(2\hat{m}_y) + \log(2\hat{m}_z)) \\ &\quad + \frac{1}{2} N_a \log \hat{\sigma}_a^2 + \frac{1}{2} N_b \log \hat{\sigma}_b^2\end{aligned} \tag{5.23}$$

Dans la suite du manuscrit, nous noterons la complexité stochastique $\Delta(w)$ pour simplifier les notations. Le but de la segmentation est de trouver la fonction w^{MCS} minimisant la complexité stochastique $\Delta(w)$:

$$w^{MCS} = \arg\min_{w} \Delta(w) \tag{5.24}$$

5.3.2 Algorithme de segmentation

5.3.2.1 Schéma d'optimisation

La segmentation consiste à trouver la fonction $\mathbf{w}$ minimisant le critère de complexité stochastique $\Delta(w)$. Pour résoudre un tel problème d'optimisation, nous utilisons une stratégie multi-résolution basée sur deux étapes. La première consiste à segmenter l'objet en augmentant le nombre de triangles et de sommets afin d'augmenter la résolution 3D et d'obtenir un maillage avec un nombre sur-estimé de triangles k_0 ; la seconde étape consiste à réduire la complexité du maillage obtenu à la fin de la première étape. A cette fin, chaque arête du maillage est considérée successivement et la valeur du critère Δ' dans le cas où cette arête est supprimée est calculée. L'arête dont la suppression permet d'obtenir la valeur minimale de Δ' est supprimée.

Le maillage est initialisé avec un faible nombre de triangles (par exemple comme le cube présenté sur la figure 5.5 de la section 5.4). Pour être moins sensible à l'initialisation, nous avons adopté une procédure en deux étapes. La première étape alterne des phases de déplacement de nœuds et de subdivision de triangles afin d'obtenir un maillage dont le nombre de triangles a été surestimé. Pendant la phase de subdivision de triangles la valeur du critère Δ augmente. La seconde étape alterne des phases de déplacement de nœuds et de suppression d'arêtes pour diminuer la valeur du critère Δ. Les phases de subdivision de triangles, de déplacement de nœuds et de suppression d'arêtes sont détaillées dans les paragraphes suivants.

5.3.2.2 Subdivision de triangles

Au début de l'algorithme de segmentation, le maillage est constitué d'un faible nombre de triangles. Le nombre de triangles est donc augmenté pendant la phase de subdivision. Chaque triangle est divisé en quatre triangles en utilisant le schéma de Butterfly modifié [Zorin 96] pour lisser le maillage. La phase de subdivision de triangles est présentée sur la figure 5.1.

5.3.2.3 Déplacement de nœuds

La phase de déplacement consiste à déplacer les nœuds de la grille pour minimiser la complexité stochastique. On considère chaque nœud successivement et on lui affecte un certain déplacement. Si ce déplacement entraîne une diminution de la complexité stochastique, il est accepté, sinon le nœud

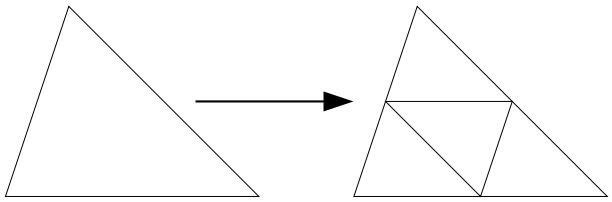


FIG. 5.1 – *Illustration de l'étape de subdivision de triangles.*

est replacé à sa position initiale. En pratique, le nombre de déplacements autorisé est restreint pour limiter le temps de calcul. Pour chaque nœud examiné successivement, on considère 26 déplacements d'amplitude a fixée (Fig. 5.2). L'objectif est de déterminer la position correspondant à la plus faible valeur de la complexité stochastique parmi les 27 possibles (les 26 déplacements testés et la position initiale). Il est donc nécessaire de régler au préalable l'amplitude de déplacement a autorisée. Cependant Galland a proposé, dans le cas 2D, une stratégie multi-résolution pour définir de paramètre de manière à ce que l'utilisateur n'ait pas besoin de le régler [Galland 04]. Nous avons choisi d'utiliser cette même méthode dans le cas 3D. Dans un premier temps, pour chaque nœud P_i de la grille, l'amplitude a_i choisie est égale à $m_i/2$ où m_i est la longueur moyenne des segments reliés au nœud P_i considéré. Ainsi pour chaque nœud l'amplitude sera différente. Lorsque plus aucun nœud ne peut être déplacé, une nouvelle étape de déplacements de nœuds est mise en oeuvre avec toutes les amplitudes a_i divisées par 2. Ce processus de division des amplitudes est effectué jusqu'à ce que toutes les amplitudes soient inférieures à un certain seuil fixé a_{min}, par exemple 1 voxel. L'intérêt de cette stratégie consiste d'une part à utiliser une grande amplitude au début de la phase de déplacement pour permettre une convergence rapide vers la bonne position du contour sans rester piégé dans un minimum local de la complexité stochastique ; et d'autre part à diminuer cette amplitude en fin de convergence pour estimer la position du contour avec une meilleur précision.

L'étape de déplacement de nœuds est présentée dans l'algorithme 6.

Remarque Il faut évidemment vérifier que le déplacement du nœud n'engendre pas de croisement entre les segments du contour.

5.3.2.4 Suppression d'arêtes

Pour simplifier le maillage, nous utilisons une phase de suppression d'arêtes [Hoppe 96]. Chaque arête est considérée successivement et celle dont la suppression permet d'obtenir la plus faible valeur de la complexité stochastique Δ est supprimée. Cette étape entraîne une déformation du contour. Par conséquent, après chaque suppression d'arête, nous appliquons une phase de déplacement de nœuds pour ajuster le maillage. Ce processus est effectué jusqu'à ce que la valeur du critère Δ ne diminue plus. Cette phase de suppression d'arêtes est illustrée sur la figure 5.3.

5.3.2.5 Algorithme de segmentation

La première étape consiste à alterner des phases de subdivision de triangles et de déplacement de nœuds jusqu'à ce que la distance entre deux nœuds soit inférieure à une distance minimale d_{min}

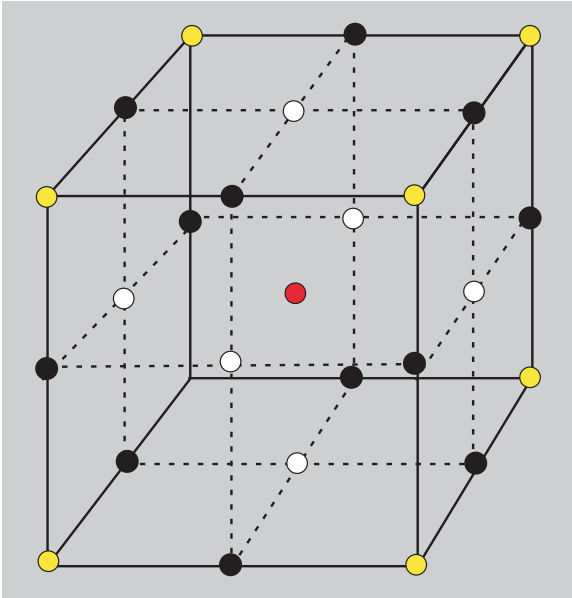


FIG. 5.2 – *Les 26 déplacements possibles et la position initiale.*

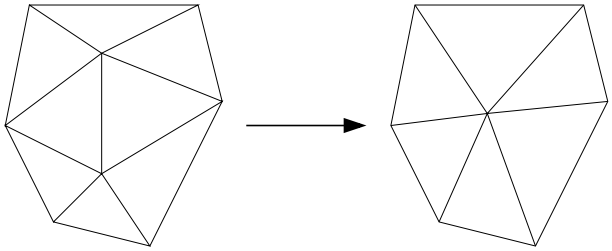


FIG. 5.3 – *Illustration de l'étape de suppression d'arêtes.*

fixée. En pratique on prendra une distance d_{min} égale à un voxel. On obtient ainsi un contour dont le nombre de nœuds et de triangles a été surestimé. Ensuite une étape alternant suppression d'arêtes et déplacements de nœuds est effectuée en minimisant la complexité stochastique afin d'estimer le nombre de nœuds du contour. L'algorithme de segmentation par contours actifs statistiques est résumé dans l'algorithme 7.

5.3.2.6 Représentation graphique

Pour représenter l'objet que l'on cherche à segmenter, un maillage triangulaire nous a paru le plus simple à mettre en oeuvre. Pour implémenter l'algorithme de segmentation, nous avons utilisé la plate-forme de modélisation géométrique de l'équipe d'Informatique Géométrique et Graphique (IGG) du LSIIT[2]. L'objectif de cette plate-forme est de proposer un cadre général pour la définition

[2] `http://lsiit.u-strasbg.fr/igg-fr/`

Algorithme 6 Etape de déplacement de nœuds

(1) Pour chaque nœud P_i, $a_i = m_i/2$
(2) Tant qu'un nœud a été déplacé, Faire
- Pour chaque nœud P_i
  - Pour chacun des 26 déplacements d'amplitude a_i
    - Appliquer le déplacement au nœud considéré $\rightarrow$ **w'**
    - Tester si aucun croisement n'a été créé
    - Déterminer la valeur du critère $\Delta(\mathbf{w'}, \mathbf{s})$ pour ce nouveau maillage
  - Parmi ces déplacements, garder celui qui produit la plus faible valeur du critère $\Delta(\mathbf{w'}_{min}, \mathbf{s})$
  - Comparer $\Delta(\mathbf{w'}_{min}, \mathbf{s})$ et $\Delta(\mathbf{w}, \mathbf{s})$
    - Si $\Delta(\mathbf{w'}_{min}, \mathbf{s}) < \Delta(\mathbf{w}, \mathbf{s})$, appliquer la déformation : $\mathbf{w} \leftarrow \mathbf{w'}_{min}$
    - Sinon annuler la transformation

(3) S'il existe des nœuds associés à une amplitude $a_i \geq 2a_{min}$,
- Pour ces nœuds P_i, $a_i \leftarrow a_i/2$
- Retourner en (2)

  Sinon la convergence est terminée

Algorithme 7 Algorithme de segmentation par contour actif statistique 3D

1. Initialisation du contour
2. Tant que la distance entre 2 nœuds $d > d_{min}$, Faire
   - Subdivision de triangles (Fig. 5.1)
   - Déplacements de nœuds (Algo. 6)
3. Tant que la CS diminue, Faire
   - Suppression d'arêtes (Fig. 5.3)
   - Déplacements de nœuds (Algo. 6)

de modèles topologiques à base de structures combinatoires permettant de représenter la topologie de subdivision d'espaces géométriques. Le modèle que nous avons utilisé est celui des 2-cartes duales. Les modèles des 2-cartes ou cartes combinatoires de dimension 2 reposent sur la notion d'hypercarte [Kraemer 07] dont quelques notions sont explicitées en annexe D. La figure 5.4 présente

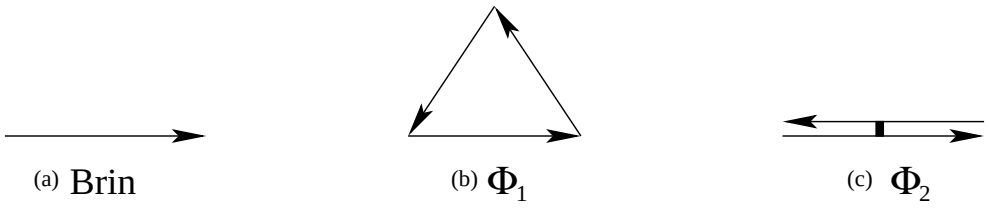


FIG. 5.4 – *Conventions graphiques utilisées pour représenter les 2-cartes duales.*

les conventions graphiques utilisées pour représenter les 2-cartes duales. Un brin est représenté par une flèche. ϕ_1 est une permutation. Une séquence de brins liés par ϕ_1 forme ainsi une face orientée. ϕ_2 permet de lier deux faces orientées le long d'arêtes d'orientations opposées. Ce modèle permet seulement de décrire la topologie de subdivision des objets. Afin de représenter complètement un objet, il va donc falloir définir un modèle de plongement décrivant les données géométriques que l'on va associer à chaque cellule du maillage. Le plongement que nous utilisons est le 0-plongement qui associe un point 3D à chaque sommet du maillage. Ce modèle des 2-cartes duales nous a permis de gérer plus facilement les différentes phases de subdivision de triangles, déplacement de nœuds et de suppression d'arêtes.

5.4 Validation

5.4.1 Résultats sur des images synthétiques

Pour valider la méthode proposée, des tests ont d'abord été effectués sur des images synthétiques. L'image synthétique 3D est composée d'un cône sur un fond altéré par un bruit gaussien dont les paramètres sont décrits dans le tableau 5.1. Le contour initial utilisé est un cube composé de 24 triangles et 14 sommets (Fig.5.5(b)). La figure 5.5 illustre la méthode proposée pour la segmentation du cône.

Classe	Moyenne μ	Variance σ^2
Cône	1	0.1
Fond	0	0.1

TAB. 5.1 – *Paramètres de l'image synthétique.*

Nous avons testé l'algorithme avec 3 positions différentes du contour initial. Nous avons calculé les taux d'erreur obtenus par rapport à la vérité terrain en considérant les voxels situés à l'intérieur du contour. Les taux d'erreur obtenus sont respectivement de 3.9, 4 et 4.1%.

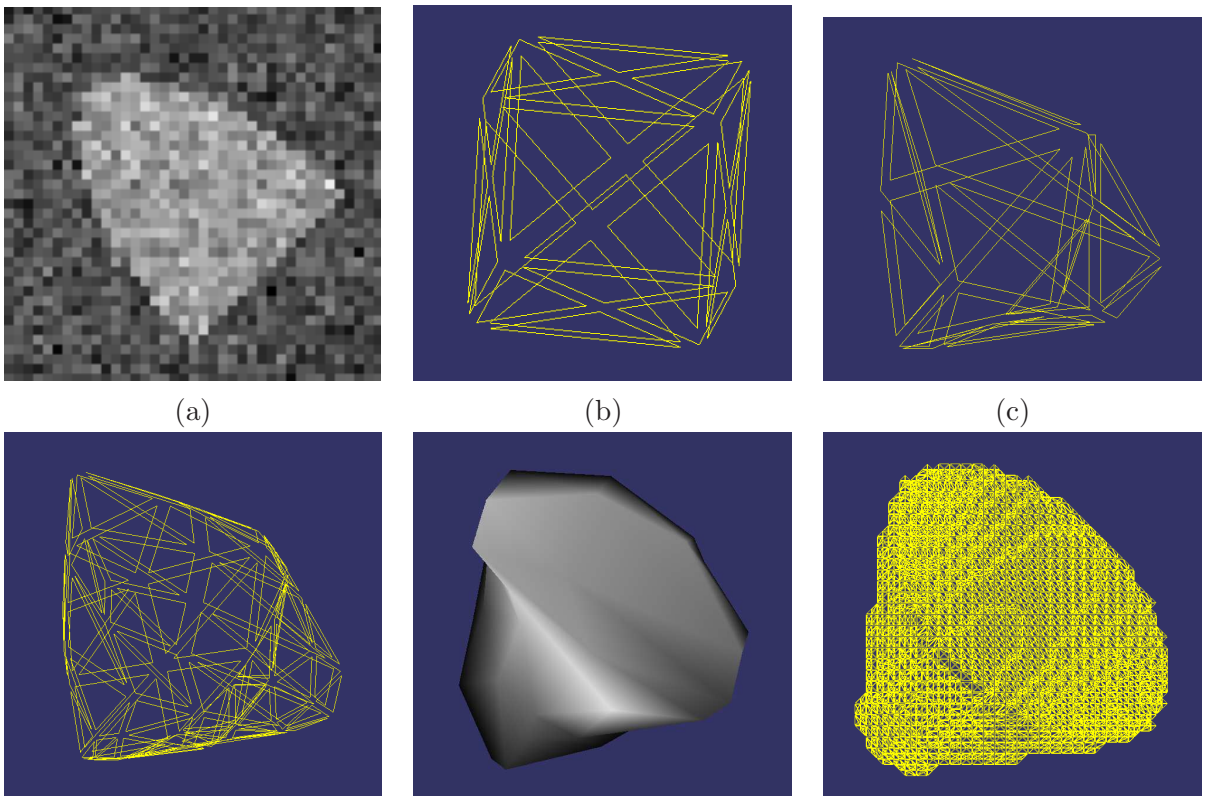


FIG. 5.5 – *Illustration de la méthode proposée pour la segmentation du cône. (a) correspond à l'image à segmenter. (b) représente le contour initial. (c) et (d)correspondent respectivement aux résultats après 1 et 2 phases de déplacement de nœuds et de subdivision de triangles. (e) représente le résultat final. (f) correspond à la vérité terrain.*

5.4.2 Résultats sur des images réelles

Nous avons ensuite effectué des tests sur des images réelles de lésions SEP et de différentes structures anatomiques (putamen, thalamus, ventricule, noyaux caudés). Nous commençons par présenter les résultats obtenus sur les lésions puis ceux obtenus sur quelques structures anatomiques cérébrales.

5.4.2.1 Segmentation de lésions SEP

Nous présentons ici quelques résultats obtenus sur des lésions SEP, d'abord sur une image provenant de la base de Strasbourg puis sur une image issue de la base d'entraînement fournie lors du challenge MICCAI'08. Les résultats obtenus seront comparés par rapport au marching cube des segmentations manuelles de deux experts [Lorensen 87]. Pour chaque cas, l'algorithme est initialisé par un cube comportant 24 triangles et 14 sommets.

Les résultats obtenus sont présentés sur les figures 5.6 et 5.7. Nous avons calculé les taux d'erreur obtenus par rapport aux segmentations manuelles des experts. Pour l'image provenant de la base de Strasbourg, les taux d'erreur par rapport aux deux experts sont respectivement de 5.8 et 5.9%.

Pour celle de la base d'entraînement fournie lors du challenge MICCAI'08 les taux d'erreur par rapport aux deux experts sont respectivement de 12.8 et 15.9%. Ces derniers taux d'erreur sont relativement importants. Cependant nous avons pu voir au chapitre 4 la grande variabilité entre les segmentations des deux experts. Cette variabilité est également visible sur la figure 5.7.

5.4.2.2 Segmentation de structures anatomiques cérébrales

Afin de valider notre méthode sur des structures anatomiques cérébrales, nous avons utilisé les images de la base IBSR[3]. En plus des segmentations en tissus utilisées au chapitre 3, cette base contient des images réelles avec les segmentations en structures correspondantes.

Nous avons segmenté différentes structures : putamen, thalamus, ventricule et noyaux caudés. Dans chaque cas, le contour initial correspond à un cube placé sur la structure. Nous avons comparé les segmentations obtenues avec le marching cube des segmentations manuelles de la base IBSR. Les figures 5.9, 5.10, 5.11 ,5.12 montrent les segmentations obtenues pour chaque structure ainsi que les segmentations manuelles correspondantes. Les taux d'erreur obtenus sont présentés dans le tableau 5.2.

Structure	Ventricule latéral	Thalamus	Noyau caudé	Putamen
Taux d'erreur	7.9	8.1	6.8	7.2

TAB. 5.2 – *Taux d'erreur obtenus par la méthode des CAS 3D sur les différentes structures anatomiques cérébrales.*

5.5 Discussion

La méthode que nous avons proposée semble prometteuse. Basée sur le principe de minimisation de la complexité stochastique, elle permet de segmenter de manière non supervisée des lésions ou des structures anatomiques 3D. Cependant une validation sur un plus grand nombre de données et une comparaison avec des méthodes de segmentation par contours actifs permettrait de mieux quantifier l'intérêt de cette modélisation. D'autre part, de nombreuses extensions sont envisageables :

- ❖ Ce modèle peut être étendu à d'autres types de ddp. En effet l'hypothèse de bruit gaussien n'est pas nécessairement valide pour les structures à segmenter. D'autre part, nous avons supposé que le fond était homogène. Cette hypothèse n'est cependant pas valide étant donnée la complexité des images cérébrales.

- ❖ Cette méthode peut facilement être étendue à la segmentation de données multimodales. Soit une image vectorielle $\mathbf{s}$ constituée de M images $s_{i,i\in\{1,\ldots,M\}}$.

$$\mathbf{s} = \{\mathbf{s}(x,y,z) = (s_1(x,y,z),\ldots,s_M(x,y,z))|(x,y,z) \in [1,N_x]\times[1,N_y]\times[1,N_z]\} \quad (5.25)$$

  Pour chaque région $\Omega_{l,l\in\{a,b\}}$, on suppose que les vecteurs de niveaux de gris de pixels $\mathbf{s}(x,y,z)$, $(x,y,z) \in \Omega_l$ sont distribués selon une ddp P_{Θ_l} de vecteur paramètre Θ_l. Pour le calcul de la complexité stochastique, la différence entre le cas mono- et multimodal se situe

[3] `http://www.cma.mgh.harvard.edu/ibsr/`

dans l'expression de la log-vraisemblance. Pour certains types de ddp, cette expression n'est cependant pas triviale à déterminer.

❖ Dans le cas où les paramètres (moyenne et variance dans notre cas) de l'objet cible (lésion ou structure) sont connus, ces paramètres peuvent être pris en compte dans le calcul de la complexité stochastique.

Le principal inconvénient de cette méthode provient du temps de calcul important : plusieurs heures pour segmenter une structure anatomique cérébrale. En effet les phases de déplacement de nœuds sont coûteuses en temps de calcul. Pour chaque nœud, il faut tester 26 déplacements possibles et effectuer cette opération jusqu'à ce qu'il n'y ait plus aucun déplacement accepté. Ce temps de calcul peut cependant être facilement amélioré. Nous présentons ici quelques améliorations envisageables, la plupart d'entre elles ayant été mises en oeuvre par Galland dans le cas des contours actifs 2D [Galland 04] :

❖ Lors de la phase de suppression d'arêtes, nous appliquons une étape de déplacement de nœuds après chaque suppression d'arête. Cependant la suppression d'une arête n'entraîne qu'une modification locale du contour, il est donc seulement nécessaire de déplacer les nœuds reliés à l'arête venant d'être supprimée. Il suffirait donc d'appliquer l'algorithme de déplacement de nœuds aux sommets reliés à cette arête et non plus à l'ensemble des nœuds.

❖ Une autre solution pourrait consister à ne plus utiliser de phase de déplacement de nœuds entre chaque suppression d'arêtes mais à l'appliquer seulement une fois que toutes les arêtes inutiles ont été supprimées.

❖ L'amplitude des déplacements de nœuds diminue au cours d'une même phase. La faire décroître seulement phase après phase permettrait d'obtenir un gain en temps de calcul. Ainsi lors de la première phase de déplacements de nœuds, l'amplitude de déplacement du nœud P_i sera fixée à la valeur $a_i = m_i/2$ (avec m_i la longueur moyenne des segments reliés au nœud P_i), puis lors de la phase de déplacement suivante (après une phase de subdivision de triangles), la valeur a_i sera $a_i = m_i/4$.

De plus, une initialisation du contour plus proche de la segmentation finale permettrait de diminuer le temps de convergence. Dans le cas des lésions SEP, on peut ainsi utiliser le résultat de la méthode de détection de lésions présentée au chapitre précédent. Le contour initial des structures peut être obtenu en recalant un atlas des structures sur l'image à segmenter. Cette initialisation permettrait de plus d'avoir une méthode totalement automatique sans interaction de l'utilisateur pour initialiser l'algorithme.

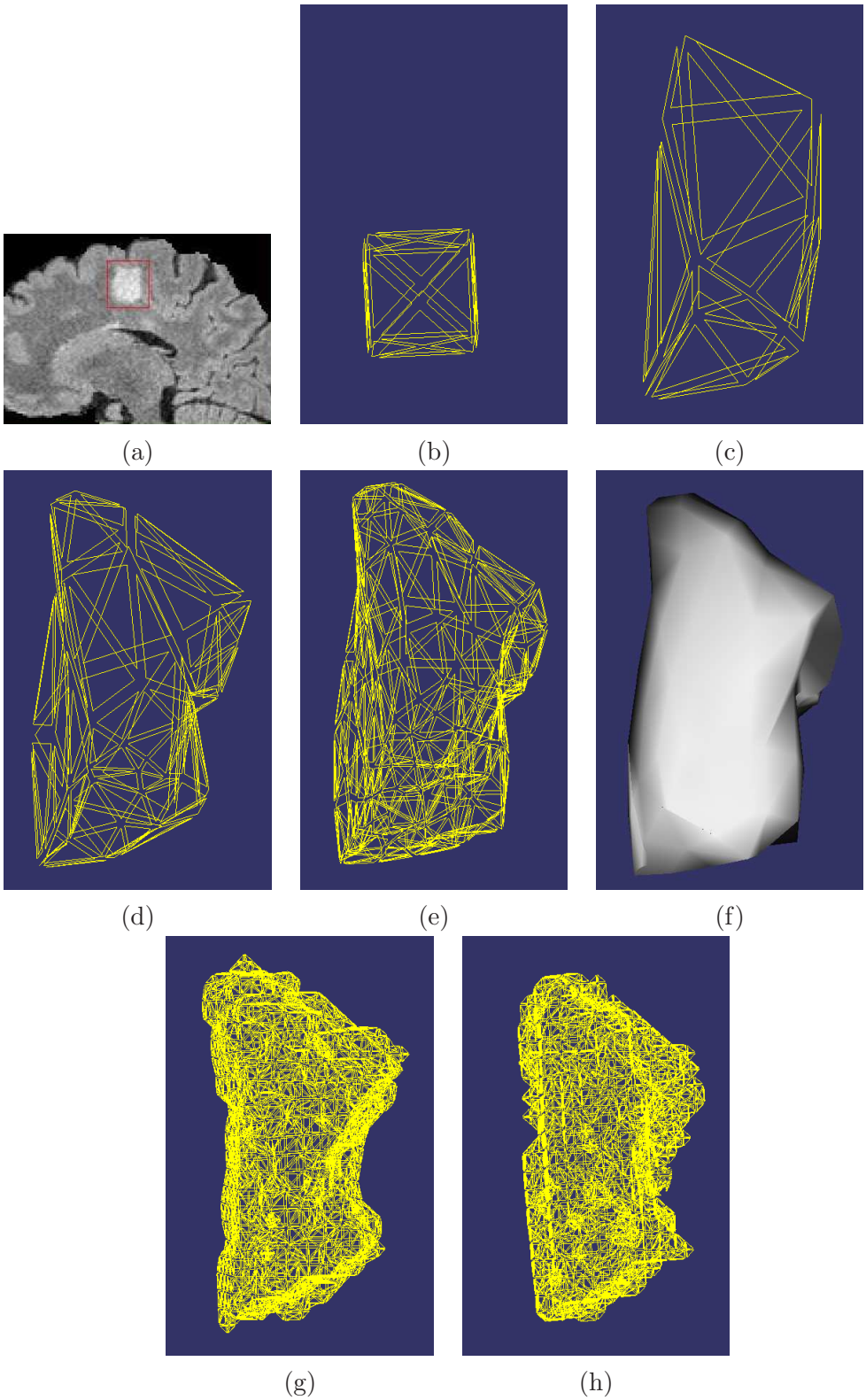


FIG. 5.6 – *Résultats obtenus sur une lésion de la base de Strasbourg. (a) correspond à l'image à segmenter. (b) représente le contour initial. (c), (d) et (e) correspondent respectivement aux résultats après 1, 2 et 3 phases de déplacement de nœuds et de subdivision de triangles. (f) représente le résultat final. (g) et (h) correspondent aux segmentations manuelles des experts.*

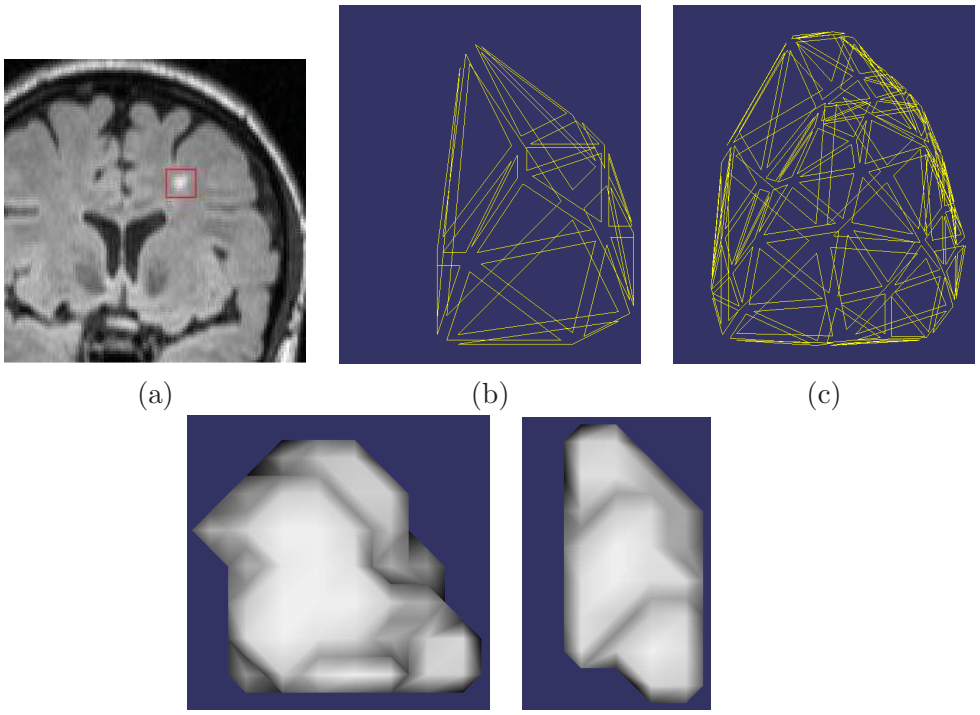


FIG. 5.7 – *Résultats obtenus sur une lésion de la base d'entraînement du challenge MICCAI'08. (a) correspond à l'image à segmenter. (b) représente le contour après une phase de déplacement de nœuds. (c) représente le résultat final. (d) et (e) correspondent aux segmentations manuelles des experts. On peut constater la grande variabilité entre les deux segmentations d'experts.*

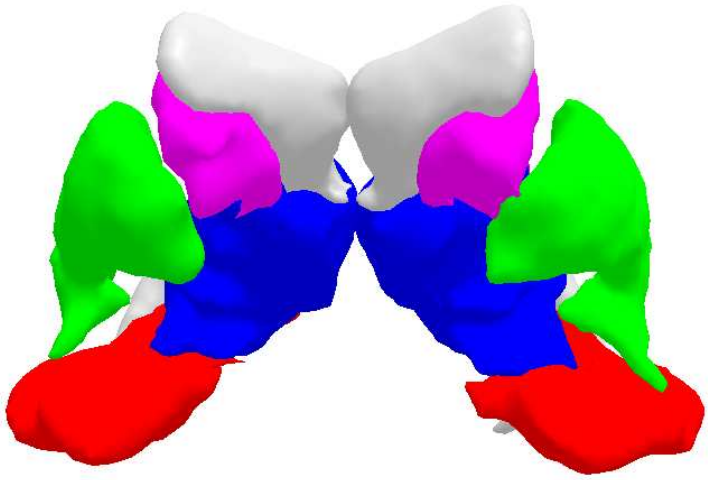


FIG. 5.8 – *Exemples de structures disponibles sur la base IBSR : en gris les ventricules latéraux ; en violet les noyaux caudés ; en vert le putamen ; en bleu le thalamus et en rouge l'hippocampe.*

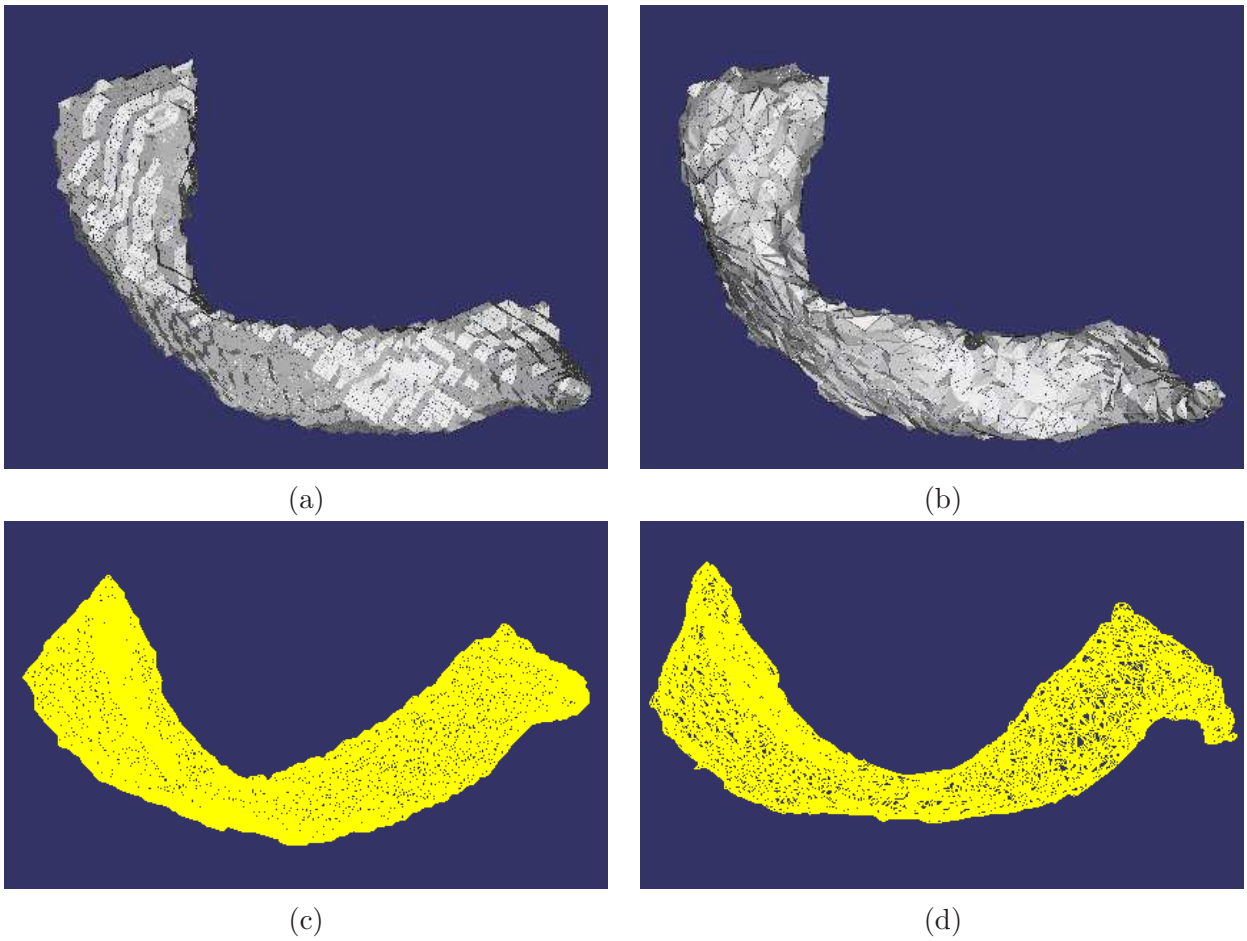


FIG. 5.9 – *Résultats obtenus sur le ventricule latéral gauche de l'image* $IBSR_{01}$. *(a) et (c) correspondent aux segmentations manuelles. (b) et (d) correspondent aux résultats obtenus par notre méthode basée sur les CAS 3D.*

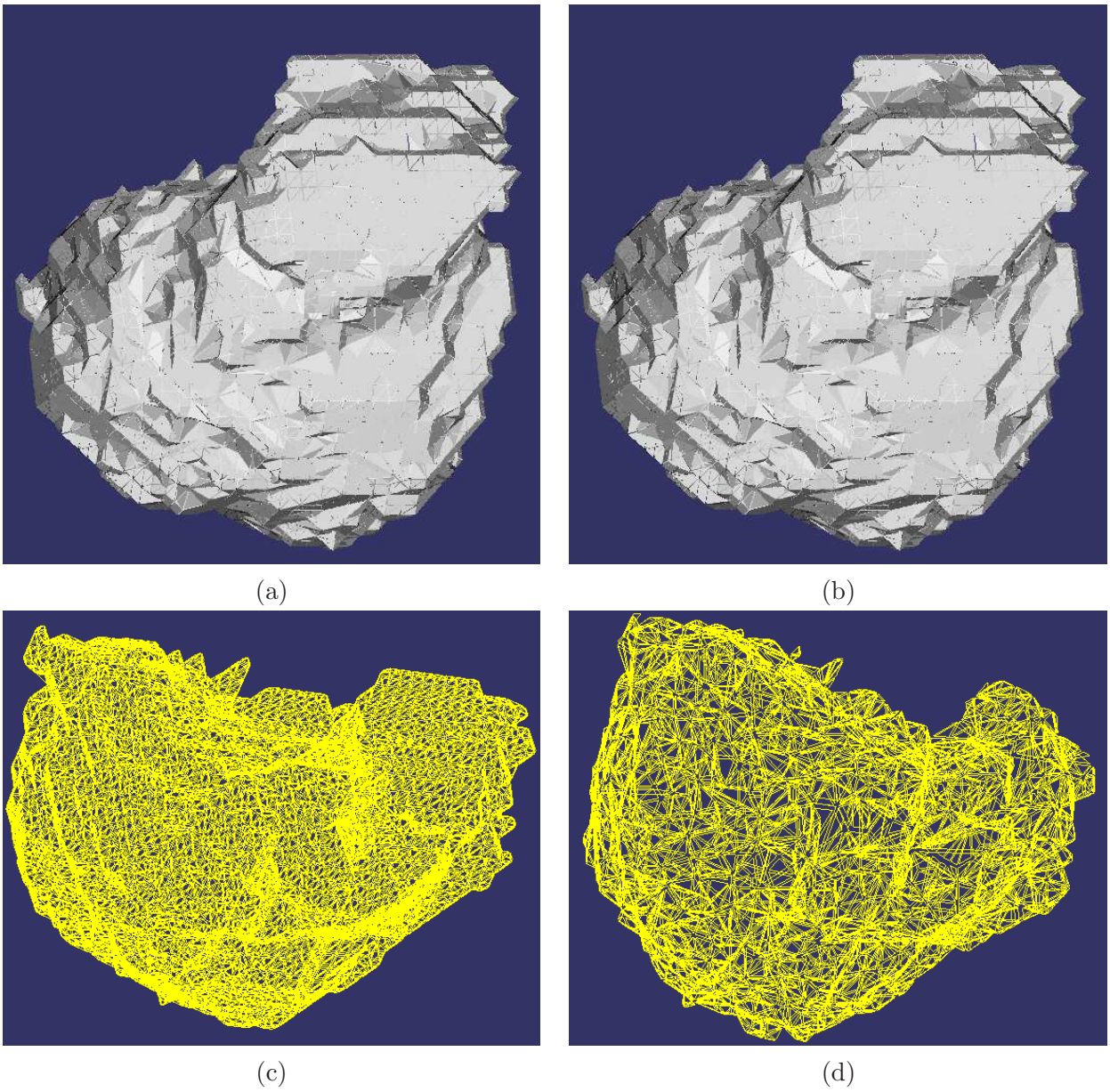
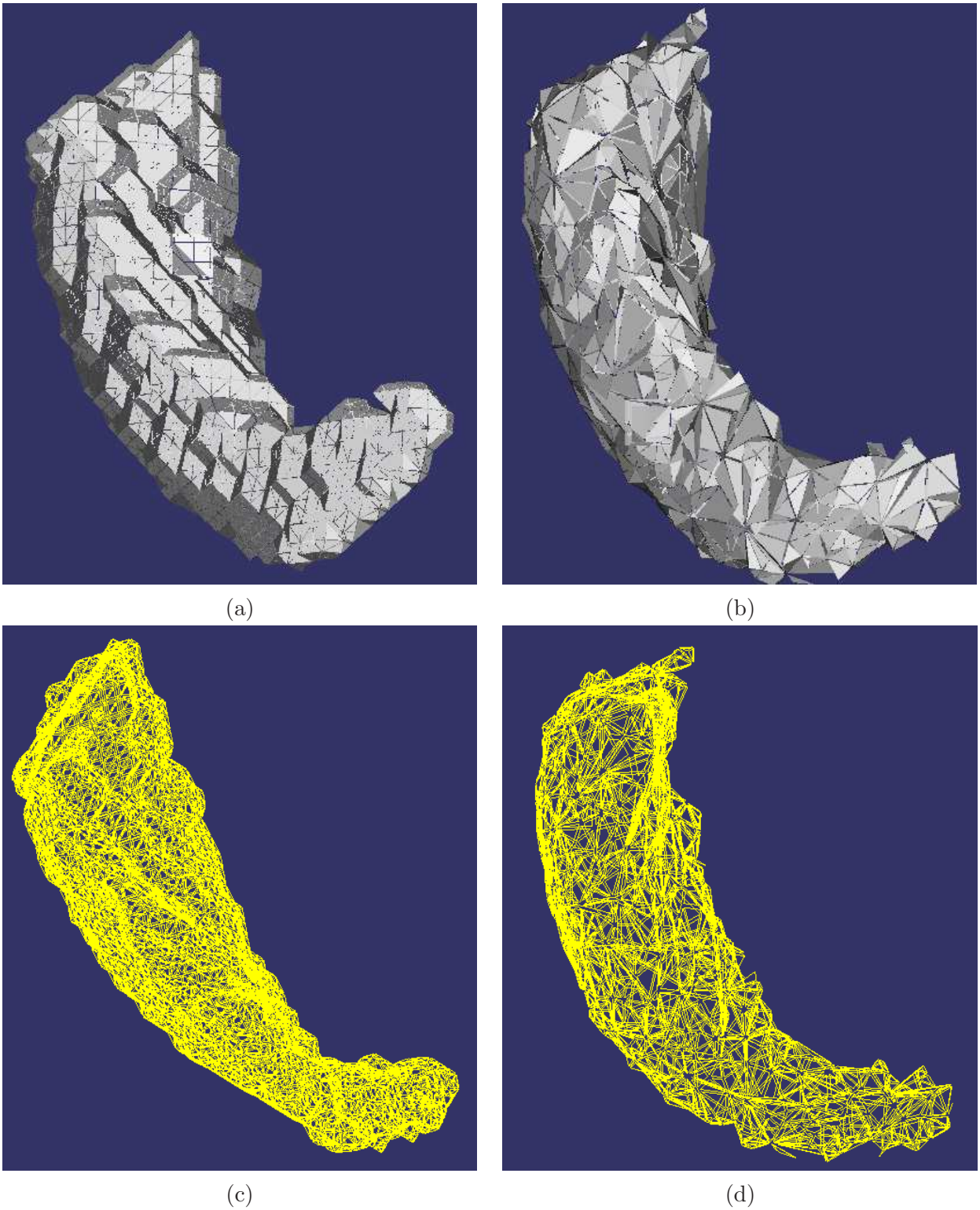


FIG. 5.10 – *Résultats obtenus sur le thalamus gauche de l'image $IBSR_{01}$. (a) et (c) correspondent aux segmentations manuelles. (b) et (d) correspondent aux résultats obtenus par notre méthode basée sur les CAS 3D.*



(a) (b)

(c) (d)

FIG. 5.11 – *Résultats obtenus sur le noyau caudé gauche de l'image $IBSR_{01}$. (a) et (c) correspondent aux segmentations manuelles. (b) et (d) correspondent aux résultats obtenus par notre méthode basée sur les CAS 3D.*

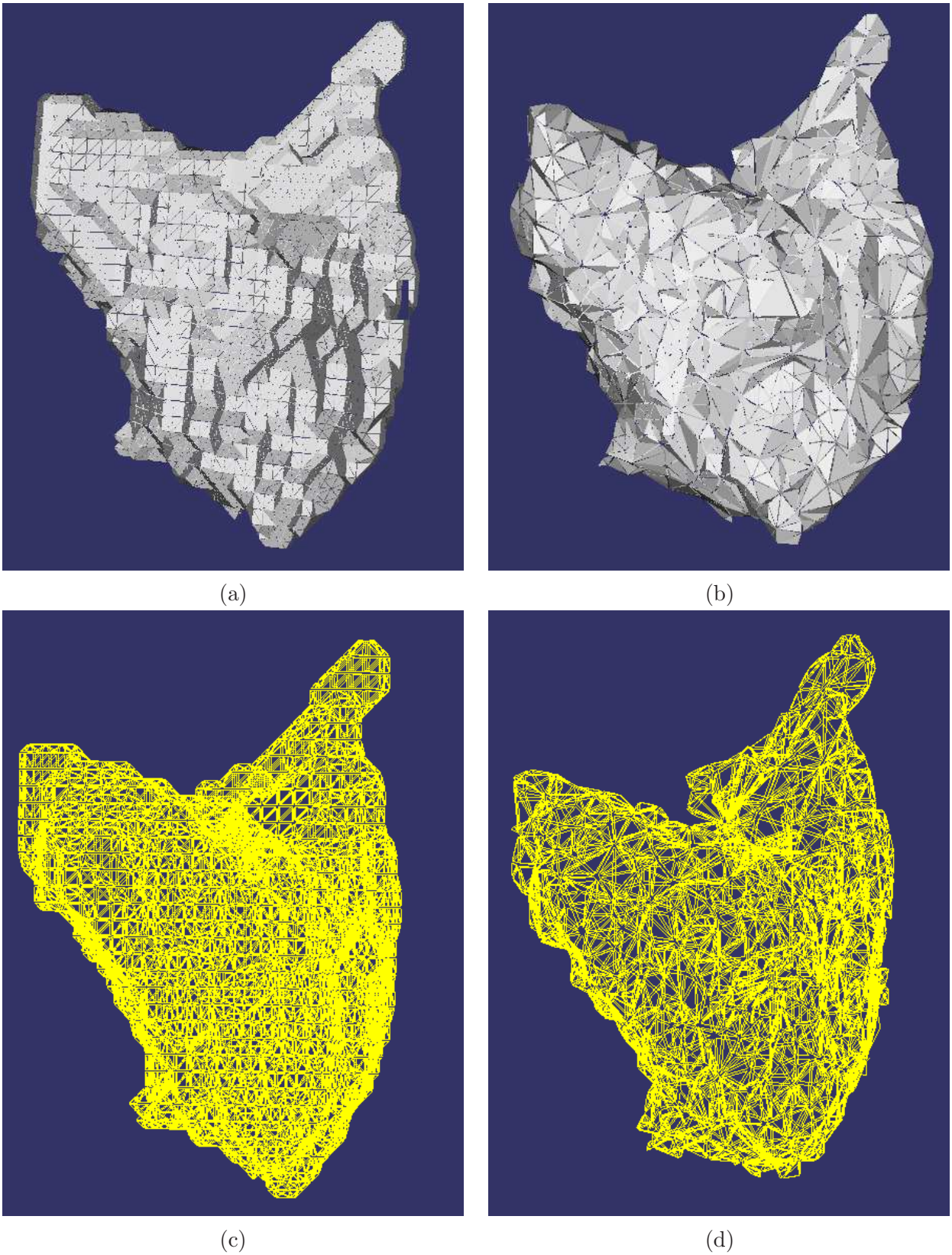


FIG. 5.12 – *Résultats obtenus sur le putamen gauche de l'image $IBSR_{01}$. (a) et (c) correspondent aux segmentations manuelles. (b) et (d) correspondent aux résultats obtenus par notre méthode basée sur les CAS 3D.*

Conclusions et Perspectives

Bilan des contributions

Ce travail de recherche a permis d'aborder différentes problématiques liées à la segmentation d'IRM anatomiques multimodales. Nous avons dans un premier temps proposé une nouvelle méthode de segmentation des tissus cérébraux basée sur la modélisation par chaînes de Markov cachées pour prendre en compte l'information sur le voisinage. Cette méthode prend en compte à la fois les hétérogénéités d'intensité présentes sur les images IRM et les effets de volume partiel en calculant la proportion de chaque tissu en chaque voxel. Cette méthode inclue également l'information *a priori* apportée par un atlas probabiliste et permet de segmenter des images mono- ou multimodales. Nous avons validé cette méthode sur la base d'images synthétiques Brainweb et sur la base d'images réelles IBSR en la comparant à différentes méthodes de segmentation existantes, en particulier SPM5 et EMS. Ce travail a abouti à une publication [$R1$] et plusieurs communications [$C1$][$C7$].

Dans un second temps nous avons cherché à étendre cette méthode de segmentation à la détection de lésions SEP. Nous avons donc utilisé l'estimateur TLE pour estimer de façon robuste les paramètres du modèle HMC et détecter les outliers. Par ailleurs, l'utilisation d'un atlas probabiliste dérivant de patients sains a permis d'apporter une information sur la localisation spatiale des outliers. Afin de ne garder que les lésions SEP, nous avons ensuite appliqué différentes contraintes sur ces outliers (volume minimal, localisation dans la matière blanche). Nous avons d'abord testé cette méthode sur la base synthétique Brainweb avec lésions pour évaluer l'apport de la prise en compte de l'information apportée par l'atlas probabiliste. Cette méthode a ensuite été testée sur des images réelles et comparée par rapport à la segmentation de deux experts dans le cadre du protocole Quick. Ce travail a fait l'objet de plusieurs communications [$C2$][$C4$][$C5$]. Par ailleurs, nous avons participé au challenge MICCAI'08 sur la segmentation des lésions SEP et notre méthode a fait l'objet d'une communication orale lors de ce challenge [$C6$].

Nous avons également cherché à obtenir une segmentation plus précise de ces lésions en proposant une méthode de segmentation basée sur les contours actifs statistiques 3D. Cette méthode basée sur le principe de minimisation de la complexité stochastique est une extension au cas tridimensionnel des méthodes proposées dans le cadre de la segmentation d'images SAR 2D par Chesnaud *et al.*. Cette méthode de segmentation consiste à trouver la fonction de partition w minimisant le critère de complexité stochastique Δ. Une stratégie multi-résolution alternant dans un premier temps des phases de subdivision de triangles et déplacements de nœuds puis des phases

de suppression d'arêtes et déplacements de nœuds a été implémentée pour obtenir le maillage triangulaire surfacique de l'objet cible à partir d'un contour initial comportant peu de nœuds. Cette méthode permet également de segmenter différentes structures cérébrales.

Perspectives

Diverses perspectives de ce travail de thèse sont envisageables. On peut d'abord envisager pour chacune des différentes méthodes proposées de tester d'autres modélisations du bruit. En effet l'approximation gaussienne du bruit n'étant pas valide dans le LCR, il a été nécessaire de mettre en oeuvre des contraintes d'intensité sur les outliers pour ne pas biaiser l'estimation des paramètres. L'utilisation de gaussiennes généralisées pourrait donc dans un premier temps être envisagée.

Dans un second temps, on pourrait envisager d'améliorer les post-traitements effectués après la détection des outliers afin de ne garder que les lésions SEP. Actuellement seuls les outliers pour lesquels la probabilité de la matière blanche donnée par l'atlas est supérieure à 0.5 sont gardés. L'inconvénient de cette méthode provient du fait qu'elle dépend de la méthode de recalage employée. Une autre solution pourrait consister à utiliser la segmentation en tissus obtenue. On pourrait ainsi garder les outliers situés dans le masque de la segmentation en matière blanche obtenue.

D'autre part nous avons vu la difficulté de la validation sur des images réelles. En effet la variabilité inter-expert peut parfois être importante, comme nous avons pu le constater sur les données réelles du challenge MICCAI'08. Cette variabilité est cependant moins importante entre les segmentations manuelles réalisées par les deux experts sur la base d'images de lésions SEP provenant de Strasbourg. Néanmoins, on pourrait envisager de créer un étalon or à partir de segmentations effectuées par plusieurs experts en utilisant par exemple un algorithme tel que STAPLE [Warfield 04].

Les méthodes de segmentation en tissus et de détection des lésions SEP que nous avons proposées prennent en compte l'information *a priori* apportée par un atlas probabiliste. Dans le cas de la segmentation en tissus, nous avons utilisé l'atlas de SPM afin de pouvoir comparer nos résultats à ceux de SPM5 et EMS, tandis que pour la détection des lésions, l'atlas utilisé a été créé à partir de notre base d'images. Il serait donc judicieux de tester l'influence de l'atlas ainsi que celle de la méthode de recalage utilisée sur les résultats.

Concernant la méthode de segmentation 3D basée sur les contours actifs statistiques développée, différentes perspectives sont envisageables pour la validation du modèle, l'initialisation du contour et l'amélioration du temps de calcul. En effet cette méthode semble prometteuse mais il serait cependant nécessaire d'effectuer un plus grand nombre de tests et de comparer les résultats obtenus avec d'autre méthodes basées sur les contours actifs afin de mieux évaluer l'apport de cette méthode par rapport aux méthodes existantes.

Concernant l'initialisation de l'algorithme, les résultats de la détection des lésions SEP basée sur les chaînes de Markov peuvent être utilisés pour créer le contour initial. Pour la segmentation des

structures, le contour initial de la structure à segmenter peut être obtenu en recalant un atlas des structures sur l'image à segmenter. Cette initialisation permettra d'obtenir un contour initial plus proche de l'image à segmenter et par conséquent de diminuer le temps de convergence nécessaire.

Le principal inconvénient actuel de la méthode par contours actifs statistiques 3D provient du temps de calcul important. Cependant plusieurs modifications de l'algorithme peuvent être effectuées pour le diminuer. En plus de l'initialisation du contour initial que nous venons d'aborder, des simplifications déjà mises en œuvre par Galland *et al.* [Galland 04] dans le cas des contours actifs statistiques 2D peuvent être réalisées pour diminuer ce temps de calcul. Les phases de déplacement de nœuds sont les principales phases concernées par ces simplifications. En effet il n'est pas nécessaire de diminuer l'amplitude de déplacement de nœuds au cours d'une même phase mais seulement phase après phase. De plus après une phase de suppression d'arêtes, il n'est pas nécessaire d'appliquer l'algorithme de déplacements de nœuds à l'ensemble des sommets mais seulement à ceux reliés à l'arête venant d'être supprimée car cette suppression a seulement entraîné une modification locale du contour.

Enfin, différentes extensions de cette modélisation par contours actifs 3D peuvent être envisagées. Dans un premier temps, ce modèle peut facilement être étendu à des données multimodales, la seule différence avec le modèle actuel se situant dans le calcul du terme de log-vraisemblance. La modélisation que nous avons proposée permet actuellement de segmenter un seul objet-cible. Ce modèle des contours actifs mono-région peut ainsi être étendu à des contours actifs multi-régions dans le cas où on souhaite segmenter plusieurs structures en même temps. Cette extension suppose cependant que le nombre de régions à segmenter est connu. Dans le cas où ce nombre est inconnu, on pourrait envisager de développer une méthode de segmentation par grilles actives statistiques (GAS) 3D sur le principe des GAS 2D développées par Galland [Galland 04].

Annexe A

EM sur chaîne de Markov avec atlas

Dans cette annexe, nous détaillons le calcul des probabilités *forward* et *backward* avec prise en compte de l'information apportée par l'atlas. Nous avons pour les probabilités *forward* :

$$\alpha_n \propto p_{X_n, \boldsymbol{Y}_{\leq \boldsymbol{n}}, B_{\leq n}}(x_n, \boldsymbol{y}_{\leq \boldsymbol{n}}, b_{\leq n}) \tag{A.1}$$

Ce qui mène à :

$$\begin{aligned} p(x_n = \omega_k, \boldsymbol{y}_{\leq \boldsymbol{n}}, b_{\leq n}) &= \sum_{\omega_l} p(x_{n-1=\omega_l}, x_n = \omega_k, \boldsymbol{y}_{\leq \boldsymbol{n-1}}, \boldsymbol{y_n}, b_{\leq n-1}, b_n) \\ &= \sum_{\omega_l} p(x_{n-1} = \omega_l, \boldsymbol{y}_{\leq \boldsymbol{n-1}}, b_{\leq n-1}) p(x_n = \omega_k | x_{n-1} = \omega_l) \\ &\quad p(b_n | x_n = \omega_k) p(\boldsymbol{y_n} | x_n = \omega_k) \\ &= \sum_{\omega_l} \alpha_{n-1}(l) p(x_n = \omega_k | x_{n-1} = \omega_l) p(b_n | x_n = \omega_k) p(\boldsymbol{y_n} | x_n = \omega_k) \end{aligned} \tag{A.2}$$

Nous avons pour les probabilités *backward* :

$$\beta_n \propto p(\boldsymbol{y}_{\boldsymbol{>n}}, b_{>n} | x_n) \tag{A.3}$$

D'où

$$\begin{aligned} p(\boldsymbol{y}_{\boldsymbol{>n}}, b_{>n} | x_n = \omega_k) &= \sum_{\omega_l} p(\boldsymbol{y}_{\boldsymbol{>n}}, b_{>n}, x_{n+1} = \omega_l | x_n = \omega_k) \\ &= \sum_{\omega_l} [p(\boldsymbol{y}_{\boldsymbol{>n+1}}, b_{>n+1} | x_{n+1} = \omega_l) p(b_{n+1} | x_{n+1} = \omega_l) \\ &\quad p(\boldsymbol{y_{n+1}} | x_{n+1} = \omega_l) p(x_{n+1} = \omega_l | x_n = \omega_k)] \\ &= \sum_{\omega_l} [\beta_{n+1}(l) p(b_{n+1} | x_{n+1} = \omega_l) p(\boldsymbol{y_{n+1}} | x_{n+1} = \omega_l) \\ &\quad p(x_{n+1} = \omega_l | x_n = \omega_k)] \end{aligned} \tag{A.4}$$

Les probabilités *forward* et *backward* peuvent être calculées de la manière suivante :

A.1 Calcul des probabilités *forward* α

– Pour $n = 1$

$$\alpha_1(k) = \pi_k f_k(\boldsymbol{y_1}) b_1(k), \quad \forall k = 1, ..., K \tag{A.5}$$

– Pour $n > 1$

$$\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) t_{lk} f_k(\boldsymbol{y_n}) b_n(k), \quad \forall k = 1, ..., K \tag{A.6}$$

où $b_n(k)$ représente la probabilité du voxel n d'appartenir au tissu k donnée par l'atlas.

A.2 Calcul des probabilités *backward* β

– Pour $n = N$

$$\beta_N(k) = 1, \quad \forall k = 1, ..., K \tag{A.7}$$

– Pour $n < N$

$$\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) t_{kl} f_l(\boldsymbol{y_{n+1}}) b_{n+1}(l), \quad \forall k = 1, ..., K \tag{A.8}$$

Annexe B

EM généralisé avec prise en compte du biais

Dans cette annexe, nous détaillons l'algorithme EM généralisé du chapitre 1.3 avec l'étape d'estimation des paramètres du biais dans le cas monomodal. z_n représente le logarithme des données $z_n = \log y_n$ et $\Phi = \{\Phi_x, \Phi_y\}$ contient les paramètres *a priori* $\Phi_x = \{\pi_k, t_{kl}\}$, et ceux de la loi d'attache aux données Φ_y.

$$\begin{aligned} Q(\Phi, \Phi^{[q]}) &= E[\log p(x, z|\Phi)|z, \Phi^{[q]}] \quad \text{(B.1)} \\ &= \sum_i \gamma_1^{[q]}(i) \log \pi_i + \sum_{n\neq 1} \sum_i \sum_j \xi_n^{[q]}(i,j) \log t_{ij} \\ &\quad + \sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(z_n) \quad \text{(B.2)} \end{aligned}$$

avec :

– les probabilités marginales *a posteriori* :

$$\begin{aligned} \gamma_n^{[q]}(i) &= p(x_n = \omega_i|z, \Phi^{[q]}) \quad \text{(B.3)} \\ &= \frac{\alpha_n(i)\beta_n(i)}{\sum_j \alpha_n(j)\beta_n(j)} \end{aligned}$$

– les probabilités conjointes *a posteriori* :

$$\begin{aligned} \xi_n^{[q]}(i,j) &= p(x_{n-1} = \omega_j, x_n = \omega_i|z, \Phi^{[q]}) \quad \text{(B.4)} \\ &= \alpha_{n-1}(j) t_{ji} f_i(z_n) \beta_n(i) \end{aligned}$$

– les probabilités initiales :

$$\pi_i = p(x_1 = \omega_i) \quad \text{(B.5)}$$

– la matrice de transition :

$$t_{ij}^n = p(x_{n+1} = \omega_j|x_n = \omega_i) \quad \text{(B.6)}$$

On maximise Q sous les contraintes suivantes : $\sum_i \pi_i = 1$ et $\sum_j t_{ij} = 1$. En utilisant les multiplicateurs de Lagrange, on obtient :

$$\pi_i^{[q+1]} = \gamma_1^{[q]}(i) \tag{B.7}$$

$$t_{ij}^{[q+1]} = \frac{\sum_{n=2}^{N} \xi_n^{[q]}}{\sum_{n=1}^{N-1} \gamma_n^{[q]}(i)} \tag{B.8}$$

$$\begin{aligned} \sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(z_n) &= \sum_n \sum_i \gamma_n^{[q]}(i) \\ &\quad [-\log \sigma_i - \frac{1}{2\sigma_i^2}(z_n - \mu_i - \sum_k c_k \psi_k(t_n))^2] \end{aligned} \tag{B.9}$$

En dérivant $\sum_n \sum_i \gamma_n^{[q]}(i) \log f_i(z_n)$ par rapport à μ_i (resp. σ_j) et en égalant le résultat à 0, on obtient :

$$\mu_i^{[q+1]} = \frac{\sum_n \gamma_n^{[q]}(i)[z_n - \sum_k c_k \psi_k(t_n)]}{\sum_n \gamma_n^{[q]}(i)} \tag{B.10}$$

$$\sigma_i^{2[q+1]} = \frac{\sum_n \gamma_n^{[q]}(i)[z_n - \mu_i - \sum_k c_k \psi_k(t_n)]^2}{\sum_n \gamma_n^{[q]}(i)} \tag{B.11}$$

En dérivant par rapport à c_k et en égalant le résultat à 0, on obtient [Van Leemput 99a] :

$$\sum_n \psi_k(x_n) \sum_i \frac{\gamma_n^{[q]}(i)}{\sigma_i^2}[z_n - \mu_i - \sum_l c_l \psi_l(t_n)] = 0, \quad \forall k \tag{B.12}$$

On obtient ainsi [Van Leemput 99a] :

$$C = (D^T W D)^{-1} D^T W R \tag{B.13}$$

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix}, \quad R = \begin{bmatrix} y_1 - \tilde{y}_1 \\ y_2 - \tilde{y}_2 \\ \vdots \end{bmatrix}, \quad D = \begin{bmatrix} \psi_1(t_1) & \psi_2(t_1) & \dots \\ \psi_1(t_2) & \psi_2(t_2) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

avec $w_{ni} = \frac{\gamma_n(i)}{\sigma_i^2}$, $w_n = \sum_i w_{ni}$, $W = diag(w_n)$, et les données corrigées avec l'estimation du biais $\tilde{y}_n = \frac{\sum_i w_{ni}\mu_i}{\sum_i w_{ni}}$.

Annexe C

Sclérose en plaques

L'Echelle de Cotation du Handicap EDSS (Expanded Disability Status Scale) présentée dans le tableau C.1 est le principal outil de cotation clinique commun à tous les neurologues pour juger l'évolution des patients. Un score chiffré de sévérité croissante (0 à 6 ou 7) est donné à chaque paramètre fonctionnel (PF). Le score global de l'échelle se mesure sur une échelle de 20 niveaux (0 à 10 par demi-points). Jusqu'au niveau 3,5, le score obtenu dans chaque PF et le nombre de PF atteints déterminent automatiquement le score EDSS. De 4 à 7, la définition de chaque niveau est aussi donnée par l'incapacité de marche .

Score	Symptômes
0	Examen neurologique normal
1.0	Pas de handicap, signes minimes d'un des paramètres fonctionnels (PF)
1.5	Pas de handicap, signes minimes dans plus d'un des PF
2.0	Handicap minime d'un des PF
2.5	Handicap minime dans deux PF
3.0	Handicap modéré d'un PF sans problème de déambulation
3.5	Handicap modéré dans un PF sans problème de déambulation
4.0	Indépendant, debout 12 heures par jour en dépit d'un handicap relativement sévère Capable de marcher 500 mètres sans aide et sans repos
4.5	Déambulation sans aide, debout la plupart du temps durant la journée capable de travailler une journée entière, peut cependant avoir une limitation dans une activité complète ou réclamer une assistance minimale handicap relativement sévère Capable de marcher 300 mètres sans aide et sans repos
5.0	Déambulation sans aide ou repos sur une distance d'environ 200 mètres handicap suffisamment sévère pour altérer les activités de tous les jours.
5.5	Déambulation sans aide ou repos sur une distance d'environ 100 mètres handicap suffisant pour exclure toute activité complète au cours de la journée
6.0	Aide unilatérale (canne, canne anglaise, béquille), constante ou intermittente, nécessaire pour parcourir environ 100 mètres avec ou sans repos intermédiaire
6.5	Aide permanente et bilatérale (cannes, cannes anglaises, béquilles) nécessaire pour marcher 20 m sans s'arrêter
7.0	Ne peut marcher plus de 5 m avec aide ; essentiellement confiné au fauteuil roulant fait avancer lui-même son fauteuil et effectue le transfert est au fauteuil roulant au moins 12 h par jour
7.5	Incapable de faire quelques pas ; strictement confiné au fauteuil roulant a parfois besoin d'une aide pour le transfert ; peut faire avancer lui-même son fauteuil ne peut y rester toute la journée ; peut avoir besoin d'un fauteuil électrique
8.0	Essentiellement confiné au lit ou au fauteuil, ou promené en fauteuil par une autre personne peut rester hors du lit la majeure partie de la journée conserve la plupart des fonctions élémentaires conserve en général l'usage effectif des bras
8.5	Confiné au lit la majeure partie de la journée, garde un usage partiel des bras conserve quelques fonctions élémentaires
9.0	Patient grabataire ; peut communiquer et manger
9.5	Patient totalement impotent, ne peut plus manger ou avaler ni communiquer
10.0	Décès lié à la SEP

Tab. C.1 – *L'échelle EDSS mise au point par Kurtzke [Kurtzke 83].*

Annexe D

Hypercartes

Dans cette annexe nous présentons quelques notions sur les hypercartes en détaillant plus particulièrement les 2-cartes et les 2-cartes duales [Kraemer 07].

D.1 Les hypercartes

Une hypercarte est constituée d'un ensemble de brins, reliés entre eux par des relations. Formellement, étant donnés un ensemble fini de brins B et des permutations α_0 et α_1 sur cet ensemble B, une hypercarte est un triplet :

$$H = (B, \alpha_0, \alpha_1) \tag{D.1}$$

Les cellules de la subdivision sont définies implicitement par des sous-ensembles de brins. Ces sous-ensembles sont obtenus grâce à la notion d'orbite. Pour une permutation σ sur B, l'orbite de $x \in B$ est le sous-ensemble de B noté $\langle\sigma(x)\rangle$ égal à $\{x, \sigma(x), ..., \sigma^k(x)\}$, où k est le plus petit entier positif tel que $\sigma^{k+1}(x) = x$. On voit clairement que tous les éléments de $\langle\sigma(x)\rangle$ ont la même orbite. En pratique, appliquée à un brin x de B, l'orbite représente l'ensemble des brins atteignables depuis x par les applications successives de la permutation σ.

D.2 Les 2-cartes

Une 2-carte est une hypercarte à laquelle on ajoute la contrainte suivante : α_0 est une involution sans point fixe, c'est-à-dire une permutation telle que $\forall x \in B, \quad \alpha_0(\alpha_0(x)) = x$ et $\alpha_0(x) \neq x$. Ce modèle permet de représenter la topologie de la subdivision de 2-variétés orientables fermées, i.e. de surfaces englobant des volumes. La figure D.1 présente les conventions graphiques adoptées pour la représentation des brins et des relations.

Les sommets sont définis par l'orbite α_1, les arêtes par l'orbite α_0, et les faces par l'orbite $\alpha_0 \circ \alpha_1$.

D.3 Les 2-cartes duales

Soit $M = (B, \alpha_0, \alpha_1)$ une 2-carte, le triplet $M' = (B, \phi_1, \phi_2)$ avec $\phi_1 = \alpha_0 \circ \alpha_1$ et $\phi_2 = \alpha_0$ est aussi une 2-carte appelée carte duale de M. La relation ϕ_1 est une permutation et ϕ_2 est

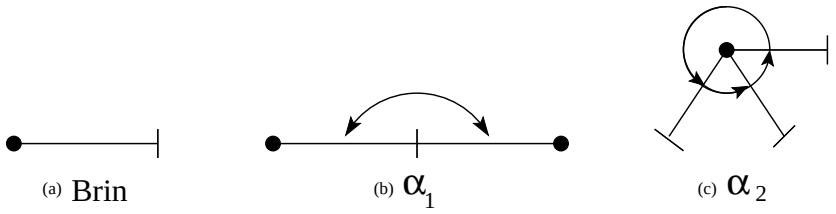


FIG. D.1 – *Conventions graphiques utilisées pour représenter les 2-cartes.*

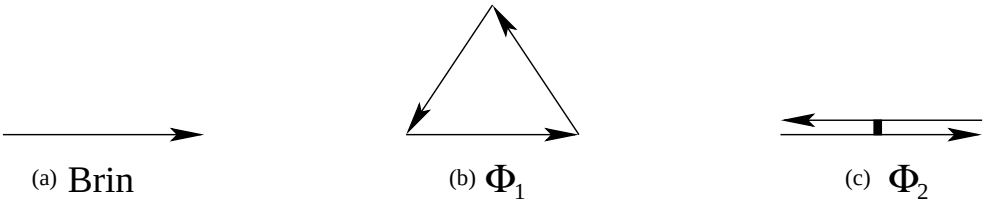


FIG. D.2 – *Conventions graphiques utilisées pour représenter les 2-cartes duales.*

une involution sans point fixe. La figure D.2 présente les conventions graphiques utilisées pour représenter les 2-cartes duales. Un brin est représenté par une flèche. ϕ_1 est une permutation. Une séquence de brins liés par ϕ_1 forme ainsi une face orientée. ϕ_2 permet de lier deux faces orientées le long d'arêtes d'orientations opposées. Les sommets sont définis par l'orbite $\phi_1 \circ \phi_2$, les arêtes par l'orbite ϕ_2, et les faces par l'orbite ϕ_1 .

Liste des Publications

Chapitres de livres

[$B1$] Ch. Collet, A. Jalobeanu, **S. Bricq**, F. Flitti, "Fusion et imagerie multimodale", Chapitre 14 du livre "Problèmes Inverses en Imagerie et en Vision", ouvrage collectif coordonné par Ali Mohammad-Djafari, Editor, Ed Hermes, Traité IC2, *A paraître.*

Revues internationales

[$R1$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Unifying framework for Multimodal Brain MRI Segmentation based on Hidden Markov Chains", Medical Image Analysis, Ed. Elsevier, vol. 12(6), pp. 639–652, Décembre 2008.

Conférences internationales

[$C1$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Triplet Markov Chain for 3D MRI Brain Segmentation Using a Probabilistic Atlas", *IEEE 2006 International Symposium on Biomedical Imaging*, April 6-9, ISBI'06, 2006.
[$C2$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "3D brain MRI segmentation based on robust Hidden Markov Chain", *IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP'08, March 30 - April 4, 2008, Las Vegas, Nevada, USA.
[$C3$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Lesions detection on 3D brain MRI based on Robust Hidden Markov Chain", *ISMRM 2008 Annual Meeting*, Toronto, Canada.
[$C4$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Lesions detection on 3D brain MRI using trimmed likelihood estimator and probabilistic atlas", *IEEE 2008 International Symposium on Biomedical Imaging*, ISBI'08, Paris, France.
[$C5$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Markovian Segmentation of 3D Brain MRI to detect Multiple Sclerosis Lesions", *IEEE 2008 International Conference on Image Processing*, ICIP'08, October 1215, 2008, San Diego, California, U.S.A.
[$C6$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "MS Lesion Segmentation based on Hidden Markov Chains", *MICCAI 2008 Workshop Proceedings : 3D Segmentation in the Clinic - A Grand Challenge.* , New-York, 6 septembre 2008.

Conférences nationales

[$C7$] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Segmentation multimodale d'IRM cérébrales avec effet de volume partiel par Chaînes de Markov Cachées", *Onzième congrès francophone des jeunes*

chercheurs en vision par ordinateur, Obernai, ORASIS'O7, France, 4-8 juin 2007.
[*C*8] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Segmentation de lésions de Sclérose en Plaques en IRM cérébrale multimodale par Chaînes de Markov Cachées", *12ème congrès du GRAMM*, Lyon, 26-28 Mars 2008.

Autres communications

[*C*9] **S. Bricq**, Ch. Collet, J.-P. Armspach, "Segmentation d'IRM anatomiques multimodales couplant inférence markovienne et atlas probabiliste", *journée du GDR–ISIS*, Paris, 23 novembre 2006.

Bibliographie

[Aït-ali 05] L.S. Aït-ali, S. Prima, P. Hellier, B. Carsin, G. Edan & C. Barillot. *STREM : a robust multidimensional parametric method to segment MS lesions in MRI.* In J. Duncan & G. Gerig, editeurs, 8th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2005, volume 3749 of *Lecture Notes in Computer Science*, pages 409–416, Palm Springs, USA, October 2005. Springer.

[Ait-Ali 06] L. Ait-Ali. *Analyse spatio-temporelle pour le suivi de structures évolutives en imagerie cérébrale multi-séquences.* PhD thesis, Université de Rennes 1, 2006.

[Al-Zubi 02] S. Al-Zubi, K. Toennies, N. Bodammer & H. Hinrichs. *Fusing Markov Random Fields with Anatomical Knowledge and Shape Based Analysis to Segment Multiple Sclerosis White Matter Lesions in Magnetic Resonance Images of the Brain.* Bildverarbeitung für die Medizin, pages 185–188, 2002.

[Ashburner 97] J. Ashburner & K.J. Friston. *Multimodal Image Coregistration and Partitioning -A- Unified Framework.* NeuroImage, vol. 6, no. 3, pages 209–217, 1997.

[Ashburner 00] J. Ashburner & K.J. Friston. *Voxel-Based Morphometry–The Methods.* NeuroImage, vol. 11, pages 805–821, 2000.

[Ashburner 05] J. Ashburner & K.J. Friston. *Unified Segmentation.* NeuroImage, vol. 26, pages 839–857, 2005.

[Baillard 01] C. Baillard, P. Hellier & C. Barillot. *Segmentation of brain 3D MR images using level sets and dense registration.* Medical Image Analysis, vol. 5, pages 185–194, 2001.

[Ballester 02] M. A. Gonzalez Ballester, A. P. Zisserman & M. Brady. *Estimation of the partial volume effect.* Medical Image Analysis, vol. 6, pages 389–405, 2002.

[Bandoh 99] Y. Bandoh & A. Kamata. *An address generator for a 3-dimensional pseudo-Hilbert scan in a cuboid region.* pages I :496–500, 1999.

[Barkhof 97] F. Barkhof, M. Filippi, D.H. Miller, P. Scheltens, A. Campi, C.H. Polman, G. Comi, H.J. Ader, N. Loseff & J. Valk. *Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis.* Brain, vol. 120, no. 11, pages 2059–2069, 1997.

[Baum 72] L.E. Baum. *An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes.* Inequalities, vol. 3, pages 1–8, 1972.

[Bazin 05] P.-L. Bazin & D. L. Pham. *Topology Preserving Tissue Classification with Fast Marching and Topology Templates.* In Proc. of IPMI'05, pages 234–245, 2005.

[Belaroussi 06] B. Belaroussi, J. Milles, S. Carme, Y. M. Zhu & H. Benoit-Cattin. *Intensity non-uniformity correction in MRI : Existing method and their validation.* Medical Image Analysis, vol. 10, no. 2, pages 234–246, April 2006.

[Bellone 00] E. Bellone, J. Hugues & P. Guttorp. *A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts.* Climate Research, vol. 15, no. 1, pages 1–15, 2000.

[Besag 74] J. Besag. *Spatial Interaction and the Statistical Analysis of Lattice Systems.* Journal of the Royal Statistical Society, vol. 36, pages 192–236, 1974.

[Besag 86] J. Besag. *On the Statistical Analysis of Dirty Pictures.* Journal of the Royal Statistical Society, vol. B-48, pages 259–302, 1986.

[Bezdek 81] J. C. Bezdek. Pattern recognition with fuzzy objective function algorithms. Plenum, New York, 1981.

[Bosc 03a] M. Bosc. *Contribution à la détection de changements dans les séquences IRM 3D multimodales.* PhD thesis, Université Louis-Pasteur-Strasbourg I., 2003.

[Bosc 03b] M. Bosc, F. Heitz & J.-P. Armspach. *Statistical Atlas-Based Sub-Voxel Segmentation of 3D Brain MRI.* In IEEE International Conference on Image Processing (ICIP), pages 1077–1080, Barcelone, Espagne, September 2003.

[Bosc 03c] M. Bosc, F. Heitz, J.P. Armspach, I. Namer, D. Gounot & L. Rumbach. *Automatic change detection in multimodal serial MRI : application to multiple sclerosis lesion evolution.* NeuroImage, vol. 20, pages 643–656, 2003.

[Bovik 00] A. Bovik. Handbook of image and video processing. Academic Press, San Diego, 2000.

[Bricq 08a] S. Bricq, C. Collet & J.-P. Armspach. *3D brain MRI segmentation based on robust Hidden Markov Chain.* In IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, USA, april 2008.

[Bricq 08b] S. Bricq, C. Collet & J.-P. Armspach. *Lesions detection on 3D brain MRI using trimmed likelihood estimator and probabilistic atlas.* In Fifth IEEE International Symposium on Biomedical Imaging ISBI'08, Paris, France, may 2008.

[Bricq 08c] S. Bricq, C. Collet & J.-P. Armspach. *Markovian Segmentation of 3D brain MRI to detect multiple sclerosis lesions.* In IEEE International Conference on Image Processing ICIP'08, San Diego, CA, USA, october 2008.

[Bricq 08d] S. Bricq, C. Collet & J.-P. Armspach. *Unifying framework for Multimodal Brain MRI Segmentation based on Hidden Markov Chains.* Medical Image Analysis, vol. 12, no. 6, pages 639–652, December 2008.

[Brinkmann 98] B. H. Brinkmann, A. Manduca & R. A. Robb. *Optimized homorphic unsharp masking for MR grayscale inhomogeneity correction.* IEEE Transactions on Medical Imaging, vol. 17, no. 2, pages 161–171, 1998.

[Chardin 00] A. Chardin. *Modèles énergétiques hiérarchiques pour la résolution des problèmes inverses en analyse d'images Application à la télédétection.* PhD thesis, Université de Rennes 1, 2000.

[Choi 91] H. S. Choi, D. R. Haynor & Y. M. Kim. *Partial volume tissue classification of multichannel magnetic resonance images – a mixel model*. IEEE Transactions on Medical Images, vol. 10, pages 395–407, 1991.

[Ciofolo 04] C. Ciofolo, C. Barillot & P. Hellier. *Combining fuzzy logic and level set methods for 3D MRI brain segmentation*. In IEEE International Symposium on Biomedical Imaging : From Nano to Macro, ISBI'2004, pages 161–164, Arlington, USA, April 2004.

[Cocosco 97] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan & A.C. Evans. *BrainWeb : Online Interface to a 3D MRI Simulated Brain Database*. Proceedings of 3-rd International Conference on Functional Mapping of the Human Brain, Copenhagen, vol. 5, no. 4, 1997.

[Cohen 92] I. Cohen. *Modèles Déformables 2D et 3D : Application à la Segmentation d'Images Médicales*. PhD thesis, Université Paris IX - Dauphine, 1992.

[Cohen 00] M. Cohen, R. DuBois & M. Zeineh. *Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging*. Human Brain Mapping, vol. 10, no. 4, pages 204–211, 2000.

[Collins 98] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes & A.C. Evan. *Design and Construction of a Realistic Digital Brain Phantom*. IEEE Transactions on Medical Imaging, vol. 17, no. 3, pages 463–468, June 1998.

[Cover 91] T. M. Cover & J. A. Thomas. Elements of Information Theory. Wiley, 1991.

[Dempster 77] A.P. Dempster, N.M. Laird & D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society B, vol. 39, pages 1–38, 1977.

[Derrode 04] S. Derrode & W. Pieczynski. *Signal and Image Segmentation Using Pairwise Markov Chains*. IEEE Transactions on Signal Processing, vol. 52, no. 9, pages 2477–2489, September 2004.

[Devijver 85] P. A. Devijver. *Baum's forward-backward algorithm revisited*. Pattern Recognition Letters, vol. 3, no. 6, pages 369–373, December 1985.

[Dimova 04] R. Dimova & N.M. Neykov. *Generalized d-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications*. Theory and applications of recent robust methods, 2004.

[Dugas-Phocion 04] G. Dugas-Phocion, M. Á. González Ballester, G. Malandain, C. Lebrun & N. Ayache. *Improved EM-Based Tissue Segmentation and Partial Volume Effect Quantification in Multi-Sequence Brain MRI*. In Proc. of MICCAI'04, Lecture Notes in Computer Science, Saint-Malo, France, September 2004. Springer.

[Dugas-Phocion 06] G. Dugas-Phocion. *Segmentation d'IRM Cérébrales Multi-Séquences et Application à la Sclérose en Plaques*. PhD thesis, Ecole des Mines de Paris, 2006.

[Figueiredo 00] M. Figueiredo, J. Leitao & A. K. Jain. *"Unsupervised contour representation and estimation using B-splines and a minimum description length criterion"*. IEEE Transactions on Image Processing, vol. 9, pages 1075–1087, 2000.

[Fjortoft 03] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle & F. Tupin. *Unsupervised Classification of Radar Images Using Hidden Markov Chains and Hidden Markov*

Random Fields. IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 3, pages 675–686, March 2003.

[Flitti 05] F. Flitti. *Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov*. PhD thesis, Université Louis-Pasteur, Strasbourg I, 2005.

[Fornay 73] G. D. Fornay. *The Viterbi algorithm*. Proceedings of the IEEE, vol. 61, no. 3, pages 268–278, March 1973.

[Galland 03] F. Galland, N. Bertaux & Ph. Réfrégier. *Minimum Description Length Synthetic Aperture Radar image segmentation*. IEEE Transactions on Image Processing, vol. 12, no. 9, pages 995–1006, 2003.

[Galland 04] F. Galland. *Partition d'images par minimisation de la complexité stochastique et grille active : application à la segmentation d'images de radar à ouverture synthétique*. PhD thesis, Université d'Aix-Marseille III, 2004.

[Galland 05] F. Galland & Ph. Réfrégier. *Minimal stochastic complexity snake-based technique adapted to unknown noise model*. Optics Letters, vol. 30, no. 17, pages 2239–2241, 2005.

[Gelman 05] A. Gelman, J. Carlin, H. Stern & D. Rubin. Bayesian data analysis. Chapman and Hall - New York, 2005.

[Geman 84] S. Geman & D. Geman. *Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pages 721–741, November 1984.

[Giordana 97] N. Giordana & W. Pieczynski. *Estimation of generalized multisensor hidden markov chains and unsupervised image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 5, pages 465–475, 1997.

[Guillemaud 97] R. Guillemaud & M. Brady. *Estimating the Bias Field of MR Images*. IEEE Transactions on Medical Imaging, vol. 16, no. 3, pages 238–251, June 1997.

[Held 97] K. Held, E. Rota Kops, B. J. Krause, R. Kikinis W. M. Wells III & H.-W. Müller Gärtner. *Markov Random Field Segmentation of Brain MR Images*. IEEE Transactions on Medical Imaging, vol. 16, no. 6, pages 878–886, December 1997.

[Hoppe 96] H. Hoppe. *Progressive meshes*. In ACM SIGGRAPH '96, pages 99–108, 1996.

[Jaggi 98] C. Jaggi. *Segmentation par méthode markovienne de l'encéphale humain en imagerie par résonance magnétique : théorie, mise en oeuvre et évaluation*. PhD thesis, Université de Caen/Basse-Normandie, 1998.

[Johnston 96] B. Johnston, M. S. Atkins & B. Mackiewich. *Segmentation of Multiple Sclerosis Lesions in Intensity Corrected Multispectral MRI*. IEEE Transactions on Medical Imaging, vol. 15, no. 2, pages 154–169, April 1996.

[Jordan 98] M. I. Jordan. Learning in graphical models. The MIT Press, 1998.

[Jordan 99] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola & L. K. Saul. *An Introduction to Variational Methods for Graphical Models*. Machine Learning, vol. 37, no. 2, pages 183–233, 1999.

[Kamber 95] M. Kamber, R. Shinghal, D.L. Collins, G.S. Francis & A.C. Evans. *Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images.* IEEE Transactions on Medical Imaging, vol. 14, no. 3, pages 442–453, 1995.

[Kanungo 94] T. Kanungo, B. Dom, W. Niblack & D. Steele. *A fast algorithm for MDL-based multi-band image segmentation.* In Proc. IEEE Conf. CVPR, pages 609–616, June 1994.

[Kass 88] M. Kass, A. Witkin & D. Terzopoulos. *Snakes : Active contours models.* International Journal of Computer Vision, vol. 1, pages 321–331, 1988.

[Kraemer 07] P. Kraemer, D. Cazier & D. Bechmann. *Extension multirésolution des cartes combinatoires : application à la représentation de surfaces de subdivision multirésolution.* Journal REFIG (Revue Electronique Francophone d'Informatique Graphique), vol. 1, no. 1, pages 21–32, march 2007.

[Kurtzke 83] J.F. Kurtzke. *Rating neurological impairment in multiple sclerosis : an expanded disability status scale (EDSS).* Neurology, vol. 33, pages 1444–1452, 1983.

[Kwan 96] R.K.-S. Kwan, A.C. Evans & G.B. Pike. *An Extensible MRI Simulator for Post-Processing Evaluation.* Visualization in Biomedical Computing (VBC'96), vol. 1131, pages 135–140, 1996.

[Laferté 00] J.-M. Laferté, P. Pérez & F. Heitz. *Discrete Markov Image Modeling and Inference on the Quad-tree.* IEEE Transaction on Image Processing, vol. 9, no. 3, pages 390–404, March 2000.

[Lanchantin 04] P. Lanchantin & W. Pieczynski. *Unsupervised non stationary image segmentation using Triplet Markov Chains.* In Advanced Concepts for intelligent Vision Systems (ACVIS 04), Brussels, Belgium, Aug. 31-Sept. 3, 2004.

[Lanchantin 06] P. Lanchantin. *Chaînes de Markov Triplets et Segmentation non supervisée de signaux.* PhD thesis, Institut National des Télécommunications, 2006.

[Lauritzen 96] S. L. Lauritzen. Graphical models. Clarendon Press, Oxford, 1996.

[Leclerc 89] Y.G. Leclerc. *Constructing simple stable descriptions for image partitionning.* Computer Vision, vol. 3, pages 73–102, 1989.

[Lemieux 98] L. Lemieux, U.C. Wieshmann, N.F. Moran, D.R. Fish & S.D. Shorvon. *The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization.* Medical Image Analysis, vol. 2, pages 227–242, September 1998.

[Lin 03] F.-H. Lin, Y.-J. Chen, J. Belliveau & L. Wald. *A wavelet-based approximation of surface coil sensitivity profiles for correction of image intensity inhomogeneity and parallel imaging reconstruction.* Human Brain Mapping, vol. 19, no. 2, pages 96–111, 2003.

[Lorensen 87] W. E. Lorensen & H. E Cline. *Marching Cubes : A High Resolution 3D Surface Construction Algorithm.* Computer Graphics, vol. 21, no. 3, pages 163–169, 1987.

[Malladi 95] R. Malladi, J. A. Sethian & B. C. Vemuri. *Shape Modeling with Front Propagation : A Level Set Approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pages 158–175, 1995.

[Marroquin 87] J. Marroquin, S. Mitter & T. Poggio. *Probabilistic Solution of ill-Posed problems in Computation Vision.* Journal of the American statistical Association, vol. 82, no. 397, pages 76–89, March 1987.

[Marroquin 02] J.L. Marroquin, B.C. Vemuri, S. Botelo & al. *An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI.* IEEE Transactions on Medical Imaging, vol. 21, no. 8, pages 934–945, August 2002.

[Masson 93] P. Masson & W. Pieczynski. *SEM Algorithm and Unsupervised Statistical Segmentation of satellite Images.* IEEE Trans. on Geoscience and Remote Sensing, vol. 31, no. 3, pages 618–633, 1993.

[McDonald 01] W.I. McDonald, A. Compston, G. Edan, D. Goodkin, H.P. Hartung, F.D. Lublin, H.F. McFarland, C.H. Polman D.W. Paty, S.C. Reingold, M. Sandberg-Wollheim, W. Sibley, A. Thompson, S. van den Noortand B.Y. Weinshenker & J.S. Wolinsky. *Recommended diagnostic criteria for multiple sclerosis : guidelines from the International Panel on the diagnosis of multiple sclerosis.* Annals of Neurology, vol. 50, no. 1, pages 121–127, 2001.

[McInerney 96] T. McInerney & D. Terzopoulos. *Deformable models in medical image analysis : a survey.* Medical Image Analysis, vol. 1, no. 2, pages 91–108, 1996.

[Metropolis 53] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller & E. Teller. *Equation of State Calculations by Fast Computing Machines.* Journal of Chemical Phys., vol. 21, pages 1087–1091, 1953.

[Müller 03] C.H. Müller & N.M. Neykov. *Breakdown Points of the Trimmed Likelihood and Related Estimators in GLMs.* J. Statist. Plann. Inference, vol. 116, pages 503–519, 2003.

[Neykov 90] N.M. Neykov & P.N. Neytchev. *A robust alternative of the MLE.* Compstat'90, pages 99–100, 1990.

[Noblet 05] V. Noblet, C. Heinrich, F. Heitz & J.-P. Armspach. *3-D Deformable Image Registration : A Topology Preservation Scheme Based on Hierarchical Deformation Models and Interval Analysis Optimization.* IEEE Transactions on Image Processing, vol. 14, no. 5, pages 553–566, 2005.

[Noblet 06] V. Noblet, C. Heinrich, F. Heitz & J.-P. Armspach. *Retrospective evaluation of a topology preserving non-rigid registration method.* Medical Image Analysis, vol. 10, no. 3, pages 366–384, june 2006.

[Pérez 00] P. Pérez, A. Chardin & J.-M. Laferté. *Noniterative manipulation of discrete energy-based models for image analysis.* Pattern Recognition, vol. 33, no. 4, pages 573–586, April 2000.

[Perez 03] P. Perez. *Modèles et algorithmes pour l'analyse probabiliste des images.* Habilitation à diriger des recherches, Université de Rennes 1, 2003.

[Pham 99] D. Pham & J. L. Prince. *Adaptative Fuzzy Segmentation of Magnetic Resonance Images.* IEEE Transactions on Medical Imaging, vol. 18, no. 9, pages 737–752, September 1999.

[Pieczynski 92] W. Pieczynski. *Statistical image segmentation.* Machine Graphics and Vision, vol. 1, no. 2, pages 261–268, 1992.

[Pieczynski 02a] W. Pieczynski. *Chaînes de Markov Triplet*. Comptes Rendus de l'Académie des Sciences, pages 275–278, 2002.

[Pieczynski 02b] W. Pieczynski, C. Hulard & T. Veit. *Triplet Markov Chains in hidden signal restoration*. In SPIE's International Symposium on Remote Sensing, Crete, Greece, September 2002.

[Pieczynski 03a] W. Pieczynski. *Modèles de Markov en traitement d'images*. Traitement du Signal, vol. 20, no. 3, pages 255–277, 2003.

[Pieczynski 03b] W. Pieczynski. *Pairwise Markov Chains*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pages 634–639, May 2003.

[Pieczynski 05] W. Pieczynski & F. Desbouvries. *On Triplet Markov Chains*. In International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005), Brest, France, May 2005.

[Prima 03] S. Prima, D.L. Arnold & D.L. Collins. *Multivariate statistics for detection of MS activity in serial multimodal MR images*. In R.E. Ellis & T.M. Peters, editeurs, 6th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2003, volume 2878 of *Lecture Notes in Computer Science*, pages 663–670, Montreal, Canada, November 2003. Springer.

[Provost 01] J.N. Provost. *Classification Bathymétrique en Imagerie Multispectrale Spot*. PhD thesis, Université de Bretagne Occidentale, 2001.

[Rabiner 89] L.R. Rabiner. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of IEEE, vol. 77, no. 2, pages 257–286, 1989.

[Rajapakse 97] J. C. Rajapakse, J. N. Giedd & J. L. Rapoport. *Statistical Approach to segmentation of Single-Channel Cerebral MR Images*. IEEE Transactions on Medical Imaging, vol. 16, no. 2, pages 176–186, April 1997.

[Rey 02] D. Rey, G. Subsol, H. Delingette & N. Ayache. *Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images : Application to Multiple Sclerosis*. Medical Image Analysis, vol. 6, no. 2, pages 163–179, 2002.

[Richard 04] N. Richard, M. Dojat & C. Garbay. *Automated segmentation of human brain MR images using a multi-agent approach*. Artificial Intelligence in Medecine, vol. 30, pages 153–175, 2004.

[Rissanen 78] J. Rissanen. *Modeling by shortest data description*. Automatica, vol. 14, pages 465–471, 1978.

[Rissanen 89] J. Rissanen. Stochastic Complexity in Statistical Inquiry. Worl Scientific, Singapore, 1989.

[Rousseeuw 87] P.J. Rousseeuw & A.M. Leroy. Robust Regression and Outlier Detection. Wiley, 1987.

[Ruan 00] S. Ruan, C. Jaggi, J. Fadili & D. Bloyet. *Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling*. IEEE Transactions on Medical Imaging, vol. 19, no. 12, pages 1179–1187, December 2000.

[Ruan 02] S. Ruan, B. Moretti, J. Fadili & D. Bloyet. *Fuzzy Markovian Segmentation in Application of Magnetic Resonance Images*. Computer Vision and Image Understanding, vol. 85, pages 54–69, 2002.

[Ruch 01] O. Ruch & P. Réfrégier. *Minimal-complexity segmentation with a polygonal snake adapted to different optical noise models*. Optics Letters, vol. 26, no. 13, pages 977–979, 2001.

[Scherrer 07] B. Scherrer, M. Dojat, F. Forbes & C. Garbay. *LOCUS : LOcal Cooperative Unified Segmentation of MRI Brain Scans*. In in the Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), pages 219–227, Berlin, 2007. Springer-Verlag.

[Schroeter 98] P. Schroeter, J.-M. Vesin, T. Langenberger & R. Meuli. *Robust Parameter Estimation of Intensity Distributions for Brain Magnetic Resonance Images*. IEEE Transactions on Medical Imaging, vol. 17, no. 2, pages 172–186, April 1998.

[Shannon 48] C. E. Shannon. *A mathematical theory of communication*. The Bell System Technical Journal, vol. 27, pages 379–423, 1948.

[Shattuck 01] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg & R. M. Leahy. *Magnetic Resonance Image Tissue Classification Using a Partial Volume Model*. NeuroImage, vol. 13, pages 856–876, 2001.

[Shen 05] S. Shen, W. Sandham, M. Granat & A. Sterr. *MRI Fuzzy Segmentation of Brain Tissue Using Neighborhood Attraction With Neural-Network Optimization*. IEEE Transactions on Information Technology In Biomedicine, vol. 9, no. 3, pages 459–467, September 2005.

[Sijbers 98] J. Sijbers, A. J. den Dekker, P. Scheunders & D. Van Dyck. *Maximum Likelihood estimation of Rician distribution parameters*. IEEE Transactions on Medical Imaging, vol. 17, no. 3, pages 357–361, 1998.

[Sled 98] J. G. Sled & A. P. Zijdenbos. *A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data*. IEEE Transactions on Medical Imaging, vol. 17, no. 1, pages 87–97, February 1998.

[Smith 02] S. Smith. *Fast Robust Automated Brain Extraction*. Human Brain Mapping, vol. 17, pages 143–155, 2002.

[Somersalo 07] E. Somersalo & D. Calvetti. Introduction to bayesian scientific computing. Springer-Verlag, 2007.

[Stark 86] H. Stark & J. W. Woods. Probability, random processes, and estimation theory for engineers. Prentice-Hall, Englewood Cliffs, NJ, 1986.

[Storvik 94] G. Storvik. *A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, pages 976–986, 1994.

[Styner 00] M. Styner, C. Brechbuhler, G. Szekely & G. Gerig. *Parametric estimate of intensity inhomogeneities applied to MRI*. IEEE Transactions on Medical Imaging, vol. 19, no. 3, pages 153–165, March 2000.

[Tanner 93] M.A. Tanner. Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions. Springer Verlag, 1993.

[Thomas 02] L. C. Thomas, D. E. Allen & N. Morkel-Kingsbury. *A hidden Markov chain model for the term structure of bond credit risk spreads.* International Review of Financial Analysis, vol. 11, no. 3, pages 311–329, 2002.

[Tohka 04] J. Tohka, A. Zijdenbos & A. Evans. *Fast and robust parameter estimation for statistical partial volume models in brain MRI.* NeuroImage, vol. 23, no. 1, pages 84–97, 2004.

[Čížek 02] P. Čížek. *Robust estimation in nonlinear regression and limited dependent variable models.* EconPapers, 2002.

[Udupa 97] J.K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M.A. van Buchem & R.I. Grossman. *Multiple sclerosis lesion quantification using fuzzy-connectedness principles.* IEEE Transactions on Medical Imaging, vol. 16, no. 5, pages 598–609, 1997.

[Van Leemput 99a] K. Van Leemput, F. Maes, D. Vandermeulen & P. Suetens. *Automated Model-Based Bias Field Correction of MR Images of the Brain.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 885–896, Oct. 1999.

[Van Leemput 99b] K. Van Leemput, F. Maes, D. Vandermeulen & P. Suetens. *Automated Model-Based Tissue Classification of MR Images of the Brain.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 897–908, Oct. 1999.

[Van Leemput 01] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester & P. Suetens. *Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection.* IEEE Transactions on Medical Imaging, vol. 20, no. 8, pages 677–688, August 2001.

[Van Leemput 03] K. Van Leemput, F. Maes, D. Vandermeulen & P. Suetens. *A Unifying Framework for Partial Volume Segmentation of Brain MR Images.* IEEE Transactions On Medical Imaging, vol. 22, no. 1, pages 10–113, January 2003.

[Vandev 93] D.L. Vandev & N.M. Neykov. *Robust Maximum Likelihood in the Gaussian Case.* In S. Morgenthaler et al., editeur, New Directions in Data Analysis and Robustness, pages 259–264, Birkhäuser Verlag Basel, Switzerland, 1993.

[Vese 02] L. Vese & T. Chan. *"A multiphase level set framework for image segmentation using the Mumford and Shah model".* International Journal of Computer Vision, vol. 50, no. 3, pages 271–293, 2002.

[Vincken 97] K. L. Vincken, A. S. E. Koster & M. A. Viergever. *Probabilistic Multiscale Image Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, pages 109–120, February 1997.

[Warfield 04] Simon K. Warfield, Kelly H. Zou & William M. Wells III. *Simultaneous truth and performance level estimation (STAPLE) : an algorithm for the validation of image segmentation.* IEEE Trans. Med. Imaging, vol. 23, no. 7, pages 903–921, 2004.

[Wells 96] W. M. Wells, W. E. L. Grimson, R. Kikinis & F. A. Jolesz. *Adaptative Segmentation of MRI Data.* IEEE Transactions on Medical Imaging, vol. 15, no. 4, pages 429–442, August 1996.

[Xu 98] C. Xu & J.L. Prince. *Snakes, shapes, and gradient vector flow.* IEEE Transactions on Image Processing, vol. 7, no. 3, pages 359–369, 1998.

[Yang 04] J. Yang, L. H. Staib & J. S. Duncan. *Neighbor-Constrained Segmentation With Level Set Based 3-D Deformable Models.* IEEE Transactions on Medical Imaging, vol. 23, no. 8, pages 940–948, August 2004.

[Zhang 01] Y. Zhang, M. Brady & S. Smith. *Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm.* IEEE Transactions On Medical Imaging, vol. 20, no. 1, pages 45–57, January 2001.

[Zhu 96] S. C. Zhu & A. L. Yuille. *Region Competition : Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pages 884–900, 1996.

[Zijdenbos 94] A. Zijdenbos, B. M. Dawant, R. A. Margolin & A. C. Palmer. *Morphometric analysis of white matter lesions in MR images : Method and validation.* IEEE Transactions on Medical Imaging, vol. 13, no. 4, pages 716–724, 1994.

[Zijdenbos 98] A. Zijdenbos, R. Forghani & A. Evans. *Automatic quantification of MS lesions in 3D MRI brain data sets : Validation of INSECT.* In W.M. Wells, A.C.F. Colchester & S.L. Delp, editeurs, 1st International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'98, volume 1496 of *Lecture Notes in Computer Science*, pages 439–448, Boston, USA, October 1998. Springer.

[Zijdenbos 02] A. Zijdenbos, R. Forghani & A.C. Evans. *Automatic "pipeline" analysis of 3-D MRI data for clinical trials : application to multiple sclerosis.* IEEE Transactions on Medical Imaging, vol. 21, no. 10, pages 1280–1291, 2002.

[Zorin 96] D. Zorin, P. Schröder & W. Sweldens. *Interpolating Subdivision for meshes with arbitrary topology.* In SIGGRAPH '96 : Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pages 189–192, New York, NY, USA, 1996. ACM.

Résumé

L'imagerie médicale, en constante évolution ces dernières années, fournit un nombre croissant de données. Les méthodes de traitement et d'analyse d'images se sont récemment multipliées pour assister le médecin dans l'analyse qualitative et quantitative de ces images et faciliter son interprétation. La segmentation automatique est devenue une étape fondamentale pour l'analyse quantitative de ces images dans de nombreuses pathologies cérébrales telles que la sclérose en plaques (SEP). Dans le cadre de cette thèse nous avons focalisé notre étude sur la segmentation d'IRM cérébrales et la détection de lésions SEP. Nous avons dans un premier temps proposé une méthode de segmentation des tissus cérébraux basée sur la modélisation par chaînes de Markov cachées prenant en compte la notion de voisinage. Cette méthode permet également d'inclure l'information *a priori* apportée par un atlas probabiliste et prend en considération les hétérogénéités d'intensité présentes sur les images IRM ainsi que les effets de volume partiel en calculant les proportions de chaque tissu en chaque voxel. Nous avons ensuite étendu cette méthode à la détection de lésions SEP en utilisant un estimateur robuste. Grâce à cet estimateur robuste et à l'information *a priori* apportée par un atlas probabiliste, les lésions sont détectées comme des données atypiques par rapport à un modèle statistique d'images cérébrales non pathologiques. Nous avons également développé une méthode de segmentation d'IRM 3D basée sur les contours actifs statistiques et sur le principe de minimisation de la complexité stochastique, d'une part pour raffiner la segmentation des lésions et d'autre part pour segmenter des structures anatomiques cérébrales. L'ensemble des algorithmes ont été validés sur des bases d'images synthétiques et réelles. Les résultats obtenus ont été comparés avec d'autres méthodes de segmentation existantes, ainsi qu'avec des segmentations manuelles réalisées par des médecins.

Mots-clés : Segmentation, IRM, détection de lésions, chaîne de Markov, estimateur robuste, atlas probabiliste, contour actif statistique, complexité stochastique.

Abstract

Medical imaging provides a growing number of data. Methods processing and analysis of images have recently been developed to assist experts in the qualitative and quantitative analysis of these images and to facilitate their interpretation. Automatic segmentation has become a fundamental step for quantitative analysis of these images in many brain diseases such as multiple sclerosis (MS). In this thesis we focused our study on the segmentation of brain MRI and MS lesion detection. At first we proposed a method of brain tissue segmentation based on hidden Markov chains modeling taking into account the neighborhood information. This method can also include prior information provided by a probabilistic atlas and takes into account the heterogeneity of the intensity on the MRI images and partial volume effects by calculating the amount of each tissue in each voxel. Then we extended this method to detect MS lesions by using a robust estimator. Thanks to this robust estimator and prior information provided by a probabilistic atlas, lesions are detected as outliers to a statistical model of normal brain images. We have also developed a 3D MRI segmentation method based on statistical active contours and on the principle of minimisation of the stochastic complexity, on the one hand to refine the lesion segmentation and on the other hand to segment brain anatomical structures. All algorithms have been validated on synthetic and real images. The results were compared with other existing methods of segmentation, and with manual segmentations carried out by experts.

Keywords : Segmentation, MRI, lesion detection, Markov chain, robust estimator, probabilistic atlas, statistical active contour, stochastic complexity.