THÈSE

présentée

devant l'Université de Rennes I

pour obtenir

le grade de : Docteur de l'Université de Rennes I

Mention : Traitement du Signal et Télécommunications

par

Jean-Michel ROUET

Équipe d'accueil :	Département Image et Traitement de l'Information, ENST Bretagne
École doctorale :	Informatique, Traitement du Signal et Télécommunications
Composante universitaire :	Structure et Propriétés de la Matière

Recalage 3D élastique multimodalité par optimisation génétique

soutenue le 8 décembre 1999 devant la commission d'examen :

Composition du Jury :

Président :	М.	René Collorec, Professeur (Univ. de Rennes 1)
Rapporteurs :	М. М.	Jean-Yves Boire, Maître de conf. (Univ. Clermont-Ferrand) Philippe Cinquin, Professeur (Univ. de Grenoble 1)
Examinateurs :	Mme M. M.	Mireille Garreau, Maître de conf. (Univ. de Rennes 1) Christian Roux, Professeur (ENST Bretagne) Jean-José Jacq, Ingénieur de recherche (ENST Bretagne)

Résumé

CETTE thèse traite du problème du recalage 3D élastique multimodalité en imagerie médicale. L'originalité du travail repose, d'une part, sur l'utilisation d'un algorithme d'optimisation stochastique (algorithme génétique), et d'autre part, sur une approche hiérarchique du recalage et l'utilisation d'informations structurelles.

En premier lieu, un large aperçu de l'état de l'art couvrant les domaines du recalage en imagerie médicale et de l'optimisation, permet de dégager les principales problématiques de ce domaine. Par la suite, un modèle de déformation globale élastique 3D, ainsi qu'une mesure de distance entre images sont proposés. Une présentation générale des algorithmes génétiques, puis une description plus précise de ceux que nous avons utilisés, nous permet de décrire la chaîne de traitement hiérarchique globale mise en œuvre.

La structure hiérarchique du recalage est basée sur l'utilisation d'une segmentation des images multimodales permettant de réduire le calcul d'une distance entre deux images de natures différentes, à celui d'une distance entre deux surfaces numériques.

Par ailleurs, la structure hiérarchique, et notamment la robustesse de l'algorithme, reposent sur la décomposition du problème en trois étapes : une première étape d'approximation par un recalage rigide rapide guidant une deuxième étape d'appariement de points (*point-matching*), et enfin une étape d'affinement local sur les diverses solutions semi-optimales trouvées.

Cette thèse utilise conjointement, à des fins de validation, un algorithme de rendu direct multi-volumiques (rendus surfaciques flous ne faisant pas appel à la segmentation utilisée lors du recalage) permettant notamment la visualisation d'intersections floues entre deux volumes médicaux.

Ce mémoire présente ensuite l'application de l'algorithme que nous avons développé à une base partagée d'images cérébrales (université de Vanderbilt) originellement constituée pour valider des méthodes de recalage rigide. Les modalités considérées sont l'IRM et le scanner, et l'information structurelle extraite est l'interface air-chair. Cette application nous permet, finalement, de conclure sur les bonnes performances en terme de rapidité et de précision de notre algorithme de recalage élastique 3D.

Abstract

THIS Ph.D. thesis is devoted to 3D elastic registration of multi-modality medical images. This work presents, on the first hand, an original use of a stochastic optimization method (genetic algorithm), and on the other hand, a hierarchical approach for the registration using structural information.

At first, a broad overview of medical image registration and optimization techniques points out the main questions intrinsic to this field. Therefore, a global 3D elastic warping model, and an image to image distance computation are proposed. A general presentation of genetic algorithms, and a more concise definition of those we used, lead to the description of the global hierarchical registration procedure achieved.

The hierarchical structure of the registration lies on a segmentation of multi-modal images reducing the burden of the distance computation between images of different kinds, to the mere inter-numerical-surfaces distance computation.

Moreover, the hierarchical structure, and especially the algorithm's robustness come from the segmentation into three sequential main steps : a leading step finding quickly out a rigid registration approximation driving a point-matching second step; a local refining of the suboptimal solutions found by the point-matching algorithm constitutes the third and last step.

This thesis takes profit from a direct multi-volume rendering algorithm (fuzzy surface renderings without using an explicit segmentation) in order to assess the quality of the registrations (fuzzy intersection between medical volumes).

Finally, this manuscript presents the application of the registration algorithm on a shared cerebral medical images database (from Vanderbilt University) used to validate different rigid registration methods. The used modalities are MRI and CT-scans, while the structural information extracted is the air to skin interface. This application lead us to conclude on the good performances obtained by our 3D elastic registration algorithm in terms of fastness and accuracy.

Remerciements

L'ensemble de ce travail a été effectué au sein du département *Image et Traitement de l'information* (ITI), dirigé par le professeur Christian Roux, de l'École Nationale Supérieure des Télécommunications de Bretagne (ENST Bretagne).

Je tiens tout d'abord à remercier Monsieur René Collorec, professeur à l'université de Rennes 1 pour avoir bien voulu présider le jury de cette thèse.

Je remercie également MM. Jean-Yves Boire, maître de conférence à l'université de Clermont-Ferrand, et Philippe Cinquin, médecin et professeur à l'université de Grenoble 1 pour avoir accepté la lourde tâche d'être rapporteurs. Leur lecture critique et leur rapport de thèse m'ont été d'un grand secours pour la préparation de la soutenance. Par ailleurs, je tiens également à les remercier pour m'avoir fourni des paires d'images médicales qui ont pu servir de support à cette thèse.

Mes pensées vont également à Mireille Garreau, Maître de conférence à l'université de Rennes 1 et au Laboratoire de Traitement du Signal et des Images, pour ses questions intéressantes sur les algorithmes génétiques lors de la soutenance.

Bien entendu, j'adresse également mes remerciements au professeur Christian Roux pour m'avoir accueilli dans son laboratoire à l'ENST Bretagne et pour m'avoir encouragé tout au long de ces 3 années de thèse. Mes plus sincères remerciements s'adressent également à mon encadrant Jean-José Jacq, Ingénieur de recherche à l'ENST Bretagne, avec qui les nombreuses discussions ont toujours été riches d'enseignements et m'ont permis de cheminer efficacement dans les méandres de la recherche.

Je veux, par ailleurs, remercier tous les membres du Laboratoire d'Analyse et de Traitement de l'Information Médicale (LATIM) de Brest ainsi que les membres permanents du département ITI de l'ENST Bretagne qui ont su me faire profiter de leur expérience en imagerie médicale et en traitement d'images et su critiquer efficacement mes travaux.

Outre ces remercerciements que je qualifierais d'officiels, des remerciements plus spécifiques s'adressent à la bande de joyeux thésards du département. Mathieu, maintenant papa, ton rire et ton éternelle bonne humeur résonnent encore dans ma tête; sans toi et surtout sans Béa, je n'aurais pas eu une aussi bonne vue sur mes travaux. Grégoire ou plutôt greg, ton efficacité au boulot et aussi ta connaissance quasi-abyssale de linux m'impressionnent toujours. Mais comment fais-tu pour avoir encore le temps de kitiwecquer? Un grand merci également à Stéph, dit grincheux pour son accueil au Conquet, les bonnes bouffes, les p'tits whiskys, et les randos avec Pat (grand amateur de biâèrre)... et j'allais oublier le goût prononcé du suicide retardant d'autant l'avancement de sa rédaction. Comment ne pas citer également le petit Dr. Ir. J.A. Rodenás puisqu'il a tellement participé à l'animation du bureau... ton départ pour la Hollande a fait comme un vide ici : plus personne à qui répondre RTFM... :-). Enfin,

je ne peux oublier le *tilolo* et son légendaire coup de perceuse dans les fils électriques à 3 heures du mat', sa descente de bières et de cigarettes phénoménale et son rythme de vie aussi incompréhensible que la météo bretonne. Il ne me reste qu'à te souhaiter également une bonne fin de rédaction et une bonne thèse.

Ces remerciements ne sauraient être complets si je ne citais tous les amis de télécom avec qui j'ai partagé quasiment 6 ans de ma vie. OA (grand clown en chef), Luigi (scout avant tout), Branou (non, pas la tête !), Gloguy (mate in two), binôme de binôme (lutin mangeur de gâââteaux danois, oublions le reste...), qu'est-ce qu'on a pu en faire des conneries quand même...! Martin (merci pour les nombreux cours de guitare, de ski et de vols en falaise), Thierry (merci pour les pâtées aux go, les tours de moto et bien sûr les amies), vous aussi, bon courage pour la fin de thèse...! Je remercie également le néo-canadien-ex-réunionnais pilote Joël pour son infinie gentillesse et tous les tours en avions au dessus des côtes bretonnes... Que la GtkAdaForce soit avec toi. Et puis je pense aussi à arno, roch et sale chat pour leurs bouffonades au tarot (spéciale ! spéciale !). Enfin, un grand merci à Gérald et son vieux totor, non seulement pour m'avoir accueilli chez eux en fin de thèse alors que je devenais SDF, mais aussi pour m'avoir fait découvrir les charmes des jeudis du port en bateau. Puisse les haies de ton jardin ne pas pousser trop rapidement :-).

Je suis sûr que j'ai encore oublié plein de monde, non seulement parmi les thésards mes aussi parmi les élèves, les anciens élèves, les profs de l'école, et les amis des amis... bon quand on se reverra, je vous paierai un pot...

En bref, une thèse effectuée avec un entourage aussi agréable et aussi compétent ne peut qu'avoir été une bonne expérience.

Table des matières

Ré	ésum	é		i
A۱	Abstract iii			
Re	emer	ciemer	ıts	v
A١	/ant-j	propos	í	1
1	État	t de l'a	art	5
	1.1	Le but	du recalage en imagerie médicale	6
		1.1.1	Le nombre de modalités	6
		1.1.2	La réalisation du recalage	7
		1.1.3	Les applications cliniques	7
	1.2	Prétra	itements et segmentation des données brutes	8
	1.3	Choix	du modèle de déformation géométrique	8
		1.3.1	Les grandes classes de transformations	9
		1.3.2	Transformations utilisées en imagerie médicale	10
	1.4	Choix	d'une fonctionnelle	10
		1.4.1	Quelques algorithmes déterministes	10
		1.4.2	Quelques algorithmes itératifs	13
	1.5	Straté	gie d'exploration de l'espace de recherche	31
		1.5.1	Recherche locale	31
		1.5.2	Recherche globale	33
		1.5.3	Algorithmes hybrides	37
	1.6	Valida	tion des résultats	37
		1.6.1	Validation visuelle	38
		1.6.2	Validation à l'aide de marqueurs	42
		1.6.3	Test de robustesse statistique	42
	17	Synthè	èse du chapitre	43

2	Rec	calage élastique 3D : Modélisation	45
	2.1	Modèles de transformation géométrique 3D	46
		2.1.1 Transformation 3D rigide et globale	47
		2.1.2 Transformations 3D élastiques	48
		2.1.3 Choix du modèle de transformation élastique $\ldots \ldots \ldots \ldots \ldots$	52
		2.1.4 Description de la transformation par amers	52
		2.1.5 Application de la transformation par interpolation	53
	2.2	Application des modèles de transformation 3D	54
		2.2.1 Déformation 2D	55
		2.2.2 Déformation 3D	55
	2.3	Distance géométrique et fonction de coût	57
		2.3.1 Fonction distance et information structurelle	58
		2.3.2 Utilisation d'une carte de distances	59
		2.3.3 Utilisation d'une distance stochastique	59
	2.4	Synthèse du chapitre \ldots	62
3	Rec	calage élastique 3D : Optimisation génétique	63
	3.1	Pourquoi un algorithme génétique	64
	3.2	Origines et Principes généraux	64
		3.2.1 Fonctionnement	65
		3.2.2 Les principaux paramètres des AG	71
	3.3	Exemple d'algorithme génétique sur un problème concret	71
	3.4	Les AG utilisés	74
		3.4.1 Algorithmes génétiques et paramètres réels	74
		3.4.2 Algorithmes génétiques et espace de recherche indexé	76
	3.5	Les limites des algorithmes génétiques	78
	3.6	Algorithmes génétiques évolutionnistes	79
	3.7	Couplage avec un algorithme d'optimisation locale	79
	3.8	Synthèse du chapitre	80
4	App	proche hiérarchique robuste	81
	4.1	Structure hiérarchique	82
		4.1.1 Pourquoi la structure hiérarchique	82
		4.1.2 Les différentes étapes	84
	4.2	Étape 1 : Approximation globale (recalage rigide)	86
		4.2.1 Description	86
		4.2.2 Mise en œuvre	87
	4.3	Étape 2 : Recalage élastique (appariement de points)	88

		4.3.1	Description	88
		4.3.2	$Mise en œuvre \dots \dots$	93
	4.4	Étape	3 : Affinement local (post-analyse)	94
		4.4.1	Description	94
		4.4.2	Mise en œuvre	97
	4.5	Influe	nce comparée des différentes étapes	98
		4.5.1	Protocole de l'analyse	99
		4.5.2	$Mise en œuvre \dots \dots$	100
		4.5.3	Analyse des résultats	101
		4.5.4	Discussion et interprétation des Résultats	105
	4.6	Synth	èse du chapitre	105
5	Арр	olicatio	on et évaluation du recalage élastique	109
	5.1	La bas	se de données RREP	110
		5.1.1	Historique	110
		5.1.2	Principe d'utilisation et de comparaison	110
		5.1.3	Les images de la base de données	110
	5.2	Métho	dologie d'utilisation de la base	113
	5.3	Princi	paux résultats	114
		5.3.1	Recherche du recalage rigide optimal	114
		5.3.2	Recalage élastique par optimisation génétique	121
		5.3.3	Résumé des résultats	126
	5.4	Comp	m araison	126
	5.5	Synthe	èse du chapitre	128
С	onclu	sion		131
\mathbf{A}	Car	actéris	ation des images médicales	133
	A.1	Tomog	graphie X \ldots	133
		A.1.1	Acquisition des images scanner	133
		A.1.2	Caractéristiques	134
	A.2	Image	rie par Résonance Magnétique (IRM)	136
		A.2.1	Acquisition des images IRM	137
		A.2.2	Reconstruction de l'image (écho de gradient)	138
		A.2.3	Caractéristiques	140
в	Seg	menta	tion morphologique 3D	143
	B.1	Préser	ntation générale	143
	B.2	Défini	tions de base - graphe morphologique	144

	B.2.1	Image numérique	144
	B.2.2	Voisinage et connexité	144
	B.2.3	Distance et transformation géodésique	145
	B.2.4	Graphe morphologique	145
B.3	Distar	nces	146
	B.3.1	Distance et métrique	146
	B.3.2	Transformation distance	146
B.4	Squele	ette par zone d'influence (SKIZ)	146
	B.4.1	Définitions	146
	B.4.2	Étiquetage des composantes connexes	146
B.5	Ligne	de partage des eaux	147
	B.5.1	Définitions	147
	B.5.2	Mise en œuvre sur des images médicales	149
	B.5.3	Autres exemples	151
U VIS	uansat	ion . Rendu Direct Multi-volume	100
C_{1}	Carac	tárisation dos donnéos	153
C.1	Carac	térisation des données	153
C.1 C.2	Carac Fonde	térisation des données	153 154 155
C.1 C.2	Carac Fonde C.2.1	térisation des données	153 154 155 157
C.1 C.2	Carac Fonde C.2.1 C.2.2	térisation des données	153 154 155 157
C.1 C.2	Carac Fonde C.2.1 C.2.2 C.2.3	térisation des données	153 154 155 157 158
C.1 C.2 C.3	Carac Fonde C.2.1 C.2.2 C.2.3 Exemp	térisation des données	153 154 155 157 158 161
C.1 C.2 C.3 C.4	Carac Fonde C.2.1 C.2.2 C.2.3 Exemp Exten	térisation des données	153 154 155 157 158 161 161
C.1 C.2 C.3 C.4 Liste d	Carac Fonde C.2.1 C.2.2 C.2.3 Exem Exten	térisation des données	 153 154 155 157 158 161 161 165
C.1 C.2 C.3 C.4 Liste d Liste d	Carac Fonde C.2.1 C.2.2 C.2.3 Exemj Exten les figu	térisation des données	 153 154 155 157 158 161 161 165 169

Avant-propos

UN week-end de Novembre, en 1995, alors que je profitais des charmes de la côte bretonne du côté de Pen-Hir, après avoir escaladé quelques belles falaises de granite accompagné d'un enseignant-chercheur de l'ENST BRETAGNE, je me rappelle que la discussion a dérivé vers *les algorithmes génétiques*, tous nouveaux pour moi. La fascination pour ces algorithmes étonnamment simples et efficaces, de même que ma curiosité, m'ont poussé à chercher un stage de DEA tournant autour de ce sujet. Cette recherche, non point stochastique, a abouti sur *l'imagerie médicale*, domaine largement traité au département Image et Traitement de l'Information (ITI) de l'ENST BRETAGNE en collaboration avec le Laboratoire de Traitement de l'Information Médicale (LATIM); pour être plus précis, c'est le thème du recalage d'images médicales 3D qui a été retenu pour servir de champ d'applications, d'expérimentation et de validation aux algorithmes génétiques.

Le DEA ayant précédé cette thèse, début 1996, a en effet eu pour objectif le recalage, à l'aide d'un algorithme génétique simple d'images médicales 3D provenant de la même modalité (ou appareil imageur). Ce DEA m'a permis de découvrir ces deux domaines passionnants que sont l'optimisation stochastique d'une part, et l'imagerie médicale de l'autre. En tant qu'extension directe du sujet de stage, dont les résultats encourageants nous ont incités à étendre et affiner la méthode, la présente thèse adresse le problème, plus générique, du recalage élastique d'images médicales 3D issues de modalités différentes.

Dans le domaine de l'imagerie médicale, le thème du recalage a reçu un intérêt croissant de la part des chercheurs durant ces dernières années, et les moyens de scruter les volumes cérébraux ont considérablement progressé. Notons par exemple le développement des techniques d'imagerie anatomique (IRM, ou Imagerie par Résonance Magnétique, scanner, angiographie), mais aussi celui de l'imagerie fonctionnelle (TEP, ou Tomographie par Émission de Positrons, TEMP ou Tomographie par Émission Mono-Photonique, IRM*f*, MEG, EEG). Tous ces capteurs mesurent des signaux dont les origines physiologiques diffèrent. Par conséquent, les informations qu'ils apportent sont complémentaires et, dès lors, la fusion, ou plutôt la confrontation de ces informations devient un objectif important. Étant données les origines différentes de ces images, la fusion des informations impose une étape préalable de recalage, c'est-à-dire de mise en correspondance géométrique des points ou des structures des images.

Parmi les différents types de recalage, notons l'existence de quatre principales catégories :

- Le recalage monomodalité intrapatient est surtout utilisé à des fins de validation clinique. Des images IRM ou Scanner sont, par exemple, acquises avant et après un acte chirurgical, permettant ainsi d'apprécier la réussite de l'opération. Le recalage monomodalité intrapatient peut également être utilisé pour suivre, au cours du temps, l'évolution d'une tumeur, cancéreuse par exemple.
- Le recalage multimodalité intrapatient est utile pour fusionner des informations complémentaires d'un même patient, provenant de différents examens (IRM, Scanner, PET, etc.). Un tel cumul d'informations permet notamment d'aider le neurochirurgien à préparer une intervention.

- Le recalage interpatient monomodalité est surtout utilisé pour de la mesure de forme (morphométrie), la comparaison et la détection d'anomalies. Le recalage s'effectue alors à partir d'images de la même modalité mais provenant de patients différents (une base de donnée). Une fois toutes les images recalées sur une référence, des comparaisons plus fines entre les images sont possibles.
- Enfin, notons le recalage multimodalité intrapatient qui consiste souvent à utiliser un atlas comme cible. Une image réelle est alors recalée sur une image synthétique (atlas), ce qui permet de rapidement étiqueter les zones de l'image réelle en fonction des données prédéfinies de l'atlas.

La plupart des recalages qui sont effectués utilisent une transformation rigide (translation, rotation et éventuellement facteur d'échelle). Un tel modèle de transformation est largement justifié lorsqu'il s'agit d'un recalage monomodalité intrapatient. Par contre, dès que l'on considère des images, soit issues de patients différents, soit d'un même patient mais de modalités différentes, une transformation rigide avoue souvent ses limites (cas du recalage IRM sur scanner où les images IRM peuvent présenter des distorsions d'acquisition non-linéaires), où alors est complètement inefficace (cas du recalage interpatient). Lorsque l'on veut mettre en œuvre un algorithme de recalage élastique, la plus grosse difficulté, outre le choix du modèle de déformation, est la croissance de complexité par rapport à un recalage rigide. Cette accroissement de complexité se traduit aussitôt par une augmentation considérable du temps de calcul. Il devient alors absolument nécessaire de mettre en place une stratégie d'optimisation différente de celles que l'on emploie couramment pour les recalages rigides.

Comme nous l'avons mentionné précédemment, cette thèse dresse le problème du recalage 3D élastique multimodalité, en utilisant de manière originale les bonnes propriétés d'optimisation des algorithmes génétiques. Le travail traité dans ce mémoire s'articule autour de quatre grands thèmes qui sont :

- 1. La définition d'une transformation géométrique élastique 3D permettant de mettre en correspondance deux volumes médicaux, ainsi que la définition d'une fonction d'évaluation d'une transformation donnée (ou fonction distance entre deux volumes médicaux) rapide à calculer.
- 2. La présentation générale des algorithmes génétiques dans un premier temps, puis des algorithmes génétiques personnalisés que nous avons adaptés à la recherche d'information pour le recalage d'images médicales.
- 3. La description globale d'une chaîne de traitement hiérarchique robuste et efficace, adaptée au recalage d'images médicales 3D.
- 4. La mise en œuvre de l'algorithme sur une base de données d'images médicales IRM et scanner d'une série de patients.

Étant donnés les quatre thèmes principaux constituant le travail de ce mémoire, ce dernier s'organise naturellement de la manière suivante :

- Le chapitre 1 reprend les principales étapes nécessaires à l'élaboration d'un algorithme de recalage. Les problématiques qui seront soulevées, ainsi que les principales méthodes de la littérature, y seront classées de manière à dresser un état de l'art aussi vaste que possible. Ce premier chapitre permettra en outre au lecteur de situer notre travail parmi les diverses méthodes de recalage en imagerie médicale.
- Le chapitre 2 est consacré à la définition de la transformation géométrique élastique 3D que nous avons retenue, ainsi qu'à la description de la fonction d'évaluation rapide, ou fonction distance, permettant de qualifier l'optimalité d'une transformation donnée.

- Le chapitre 3 présente et étudie un algorithme génétique standard, puis aboutit au choix de deux algorithmes génétiques particuliers que nous utiliserons dans notre chaîne de traitement globale.
- Dans le **chapitre 4** nous définirons l'approche hiérarchique robuste constituant notre chaîne de traitement globale. On commencera d'abord par exprimer pour quelles raisons cette approche est hiérarchique, puis nous détaillerons les trois principales étapes de cette chaîne de traitement. Une analyse de l'influence des paramètres des diverses étapes sur le résultat global sera également discutée.
- Enfin, dans le **chapitre 5** nous présenterons une base de données d'images cérébrales, et discuterons les résultats de recalages obtenus par notre algorithme sur cette base. Ceci nous permettra de conclure sur la robustesse et la validation de la méthode d'optimisation stochastique que nous avons mise en place.

CHAPITRE **1** État de l'art

Ce premier chapitre introductif est consacré à la présentation des problèmes généraux, ainsi que des solutions que l'on rencontre dans la littérature, en recalage d'images médicales. Après avoir assis les principes généraux du recalage, nous détaillerons plus particulièrement les différentes étapes usuelles des algorithmes de recalage. Des exemples d'algorithmes originaux seront également présentés.

Le but de ce chapitre est également, une fois le tour d'horizon des techniques effectué, de positionner l'approche que nous avons privilégiée pour résoudre notre problème de recalage par rapport aux autres travaux de la littérature.

1.1 Le but du recalage en imagerie médicale

De manière générale, le recalage en imagerie médicale consiste à mettre en correspondance deux (ou plus) images différentes de manière à ce que les coordonnées d'un point physique soient les mêmes dans les différentes images.

Un recalage peut se faire entre des images de même modalité (de même source) ou entre images multimodales.

1.1.1 Le nombre de modalités

Ce que l'on appelle modalité en imagerie médicale est la source d'une image [40]. Les plus courantes sont le scanner-X (ou Computed Tommography, CT) également appelé TomoDensitoMétrie (TDM), l'imagerie par résonance magnétique IRM (ou Magnetic Resonance Imaging, MRI), la tomographie par émission de positrons TEP (ou Positron Emission Tomography, PET), la tomographie par émission mono-photonique TEMP (ou Single Photon Emission Tomography). Notons également l'existence des modalités EEG (ou électroencéphalogramme) et MEG (ou magnétoencéphalogramme). Toutes ces techniques d'imageries sont différentes puisqu'elles permettent, à partir d'un objet réel (le patient) d'obtenir des informations différentes. Par exemple, le scanner-X et l'IRM rendent compte d'information structurelles (les os, les tissus), alors que les modalités TEP et TEMP donnent des informations fonctionnelles (activation, débit sanguin). L'annexe A page 133 reprend plus en détails les principes de formation des images TDM et IRM, images que nous avons utilisées comme support pour notre algorithme de recalage. En plus de ces types d'images, nous pouvons rajouter toutes les images synthétiques et plus particulièrement les atlas qui peuvent, d'un point de vue algorithme de recalage, être considérées comme des modalités à part entière.

1.1.1.1 Recalage monomodal

Le cas le plus simple de recalage est le recalage monomodal. Il est souvent utilisé pour mettre en correspondance des images d'un même patient acquises à des instants différents pour, par exemple, évaluer l'évolution d'un organe comme le coeur (à une échelle de temps faible), ou d'une tumeur cancéreuse (à une échelle de temps plus longue). Par ailleurs, le recalage monomodal peut aussi être utilisé pour recaler des images de patients différents afin d'effectuer des statistiques, ou de simples comparaisons de formes.

Le recalage monomodal est considéré comme le plus simple puisque les images rendent compte d'une information très similaire. Pour un même patient, tout ce qui se voit dans une image doit aussi se voir dans l'autre image, ce qui n'est pas le cas avec le recalage multimodal. De fait, il est souvent possible de travailler directement sur les images monomodales (sur les niveaux de gris), ce qui n'est pas le forcément le cas par exemple avec des images multimodales où les niveaux de gris ne sont pas forcément corrélés (exemple entre une image IRM et un atlas).

1.1.1.2 Recalage multimodal

Le recalage multimodal englobe d'abord toutes les utilisations du recalage monomodal car la monomodalité est, en effet, un cas particulier de multimodalité. Cependant, l'essence même d'un recalage multimodal est non plus l'analyse d'une évolution entre deux images, mais plutot le cumul d'informations complémentaires issues de modalités différentes. Par exemple il est intéressant de recaler des images fonctionnelles dont la mauvaise résolution rend difficile l'identification des bords des objets, avec des images scanner-X ou IRM ayant une meilleure résolution spatiale.

1.1.2 La réalisation du recalage

La réalisation d'un recalage en imagerie médicale passe par plusieurs étapes (cf. figure 1.1). En premier lieu, on effectue au besoin un prétraitement des données brutes (voir section 1.2), ensuite on détermine une façon de quantifier la similarité entre deux images (voir section 1.4). La mesure s'effectue entre une image appelée image de *référence* et une image créée synthétiquement par déformation (voir section 1.3) d'une image à *recaler*. La maximisation de la similarité, ou la recherche de la transformation modélisant au mieux la déformation à corriger est le cœur de l'algorithme de recalage (voir section 1.5). Enfin, la dernière étape d'un processus de recalage est l'évaluation, outre la mesure de similarité, que l'on peut faire sur la qualité du recalage (ce qui est abordé à la section 1.6).



Figure 1.1 — Synoptique général d'un algorithme de recalage.

1.1.3 Les applications cliniques

Bien entendu le recalage en imagerie médicale correspond à un réel besoin clinique. Néanmoins, nous préférons nous concentrer ici sur les aspects techniques et sur la réalisation du recalage en mettant de côté l'utilisation qui en est faite. Notons simplement quels sont les besoins et contraintes imposés par le praticien.

Le recalage est avant tout un compromis entre la précision du recalage et sa rapidité d'exécution. L'idéal est bien entendu de minimiser le temps de calcul tout en maximisant la précision obtenue.

Les images sur lesquelles les médecins travaillent sont soit des coupes 2D soit des volumes 3D. Le recalage rigide (cf. section 1.3) des images 2D est un travail qui doit pouvoir se faire quasiment en temps réel alors que ce n'est encore absolument pas envisageable pour des images 3D, et encore moins pour ce qui concerne du recalage élastique (cf. section1.3).

Pour de plus amples informations sur les aspects cliniques du recalage on pourra se reporter aux articles [54, 79], ainsi qu'aux articles de synthèse sur le recalage en imagerie médicale [60, 93], et sur le recalage en général [15].

1.2 Prétraitements et segmentation des données brutes

La plupart du temps il est nécessaire de prétraiter les données car leur forme brute est souvent trop complexe à utiliser. Il existe principalement deux types de prétraitements complémentaires.

- La première catégorie vise à améliorer le rapport signal à bruit des données (qui sont pour ce qui nous concerne des images). Différents types de filtrages élémentaires sont envisageables comme les filtres gaussiens (lissage des données), ou le filtre médian permettant, au prix également d'un lissage, d'éliminer les valeurs aberrantes d'une image (du type *poivre et sel*).
- La deuxième catégorie de prétraitements que nous rencontrons est celle regroupant les opérations visant à extraire l'information pertinente contenue dans les différents jeux de données. Ces opérateurs sont principalement les opérateurs de segmentation, ou des opérateurs de filtrage (filtre passe-haut, détection de contours, gradient). À titre indicatif, les articles [33, 58, 59, 94] présentent des méthodes d'extraction de lignes de crêtes (notamment l'opérateur L_{vv} , sur lequel nous reviendrons plus tard). D'autres articles, plutôt orientés vers le *pattern-matching* (ou association de formes) nécessitent également l'extraction explicite ou implicite de formes comme les lignes, surfaces, objets [5, 7, 16, 24, 39, 48, 63, 75].

On notera également dans les prétraitements, et au sein des algorithmes de segmentation, les méthodes de détermination de régions (comme la *croissance de régions*) appliquées ensuite au calcul des moments principaux [26].

Enfin, des informations plus générales comme les courbures locales (géométrie différentielle) peuvent faire l'objet de prétraitements importants [38, 91].

1.3 Choix du modèle de déformation géométrique

Une fois effectués les prétraitements adéquats, un deuxième aspect du recalage en imagerie médicale consiste à choisir le type de transformation qui modélisera, c'est-à-dire qui corrigera, au mieux les défauts d'alignements des deux images.

Principalement, on distingue d'une part les transformations globales et les transformations locales, et d'autre part les transformations rigides et les transformations élastiques.



Figure 1.2 — Exemples de transformations 2D.

La première référence qu'il faut noter en ce qui concerne la modélisation de la transformation géométrique est le livre de WOLBERG intitulé « Digital Image Warping » [99].

Le but de la transformation est de modéliser une fonction faisant correspondre les coordonnées de chaque point d'une image à recaler avec les coordonnées de leur homologue sur une image de référence (ou vice versa). Pour cela, différentes approches existent dépendant de divers critères sur la transformation (voir l'article [93]).

1.3.1 Les grandes classes de transformations

Un algorithme de recalage peut s'appliquer soit localement, soit globalement. Une transformation locale présente l'avantage d'optimiser le recalage pour une région d'intérêt, en temps et en qualité. Cependant une approche locale présente l'inconvénient d'être très sensible au bruit, et très dépendante de la qualité de la segmentation (dans le cas où l'on utilise des images segmentées). À ce moment-là, un recalage global serait plus adapté. De plus, un recalage global peut parfois être utile pour traiter des images où l'on sait pertinemment par exemple que la seule différence entre elles réside dans une translation, une rotation ou un simple facteur d'échelle. Van den Elsen [93] rappelle plus longuement les avantages et inconvénients de ces différents types d'approches. Cet article rappelle également comment on peut obtenir des transformations plus générales, pouvant tenir compte d'une éventuelle élasticité (voir également les articles [4, 89]).

La figure 1.2, extraite de [93], montre les différents types de transformations envisageables dans le cas 2D. On peut noter que toutes les transformations locales peuvent introduire des trous ou des recouvrements dans l'image (ce qui rend la fonction non bijective). Cependant il est possible de restreindre les transformations affines, les projections et les transformations gauches de manière à éviter cette non-bijectivité. Une fois la transformation modélisée, cette dernière se résume en un ensemble de paramètres caractérisant une fonction, en général bijective, de l'image à recaler vers l'image de référence. Ces paramètres sont par exemple des vecteurs de translation selon divers axes, des angles de rotation, des facteurs d'échelle, des coefficients de torsion, etc.

1.3.2 Transformations utilisées en imagerie médicale

Le deuxième chapitre de ce mémoire étant consacré, entre autre, au choix du modèle de transformation géométrique, nous avons préféré ne pas entrer ici plus avant dans le détail des divers modèles existants. Notons simplement que nous pouvons distinguer deux grandes catégories de transformations 3D : d'une part les transformations rigides, et d'autre part les transformations élastiques. Parmi les transformations élastiques, nous aborderons au chapitre 2 les transformations affines, bilinéaires, trilinéaires (celles-là même que nous avons choisi), polynomiales, mais encore les transformations modélisées par des plaques minces ou les transformations fluides.

1.4 Choix d'une fonctionnelle

Lorsque l'on veut créer un algorithme de recalage nous avons deux possibilités. Soit faire un algorithme déterministe qui, après analyse des images à recaler calcule directement les paramètres de la transformation qui recalera les images. Soit faire un algorithme itératif qui recherchera au fur et à mesure les paramètres du recalage en minimisant une fonction estimant la différence entre deux images (on parle également de distance entre images), ou maximisant une fonction relative à la similarité des images. C'est précisément cette notion de mesure de similarité que nous appelons la fonctionnelle.

Les articles de synthèse sur le recalage en imagerie médicale [15, 93] et [60] présentent un tour d'horizon des mesures de similarité rencontrées en imagerie médicale.

De même, Liao et Zhang [56] présentent quelques mesures de similarité pour les systèmes flous, l'article d'Armand [2] cite quelques mesures spécifiques à la reconnaissance d'objets dans une image, et la thèse de Chiron [18] propose également quelques critères de similitude en imagerie médicale. Un aperçu de ces mesures, classées en deux groupes (appliquées aux algorithmes déterministes, ou aux algorithmes itératifs) est présenté ci-dessous.

1.4.1 Quelques algorithmes déterministes

On trouve dans la littérature quelques méthodes de recalage utilisant des algorithmes déterministes. Ces algorithmes sont néanmoins assez rares. Nous présentons ici trois exemples d'algorithmes déterministes : un premier utilisant les descripteurs de Fourier [17], un autre utilisant une transformation de Fourier discrète et procédant au recalage dans le domaine fréquentiel [25], et un troisième reposant sur la donnée d'un ensemble de marqueurs sur les images à recaler [41]. Dans le cadre de ces algorithmes déterministes, la fonctionnelle ne dirige pas une exploration d'un espace de recherche ; elle constitue seulement une fonction mesurant la qualité intrinsèque du recalage final.

1.4.1.1 Descripteurs de Fourier

LIN et CHELLAPPA présentent dans leur article [17] une méthode de classification de formes 2D partielles et tronquées à l'aide de descripteurs de Fourier. Cette méthode s'appuie sur le fait que les descripteurs de Fourier de deux formes 2D sont corrélés lorsque les 2 formes sont images l'une de l'autre par une transformation rigide (translation + rotation + facteur d'échelle). Cette méthode pourrait donc permettre de recaler (retrouver les paramètres de rotation, translation et facteur d'échelle) deux formes dont les frontières sont connues. Dans l'article [17], les descripteurs de Fourier sont utilisés non pas pour recaler des formes mais pour faire de la classification entre des formes partiellement connues et une base de données de formes.

La théorie sous-jacente est la suivante :

soit x(n) et y(n), n variant de 0 à N-1 un ensemble de N points définissant une forme. On pose r(n) = x(n) + iy(n). On a alors, pour tout n compris entre 0 et N-1,

$$r(n) = \sum_{k=-NC}^{NC} C(k) \exp\left(i\frac{2\pi nk}{N}\right),$$

où $\{C(k), -NC \leq k \leq NC\}$ sont les descripteurs de Fourier. Ces 2NC + 1 descripteurs peuvent être obtenus par la relation $\mathbf{C} = \mathbf{W}^{-1}\mathbf{R}$ où \mathbf{C} est un vecteur de dimension 2NC + 1 représentant l'ensemble des descripteurs de Fourier, la *j*-ième composante du vecteur \mathbf{R} de dimension 2NC + 1 est

$$\sum_{n=0}^{N-1} r(n) \exp\left(-i\frac{2\pi n j}{N}\right),\,$$

et W est la matrice carrée de taille 2NC + 1 dont l'élément de position (p, q) est

$$\sum_{n=0}^{N-1} r(n) \exp\left(i\frac{2\pi n(p-q)}{N}\right).$$

Alors si C'(k) sont les descripteurs de Fourier de la forme obtenue par rotation, translation et mise à l'échelle de la première forme, nous avons la relation suivante :

$$\forall k \neq 0, \quad C'(k) = S.C(k). \exp(i\phi).$$

Dans cette relation S est le facteur d'échelle et ϕ l'angle de la rotation. La valeur de la translation est retrouvée en comparant C'(0) et C(0), les descripteurs de Fourier d'ordre 0.

1.4.1.2 Recalage par la phase

E. DE CASTRO et C. MORANDINI présentent dans leur article [25] une méthode de recalage rigide d'images 2D utilisant des TFD¹.

Considérons deux images qui diffèrent d'une translation (x_0, y_0) et d'une rotation d'angle θ_0 . $S_t(x, y) = S_0(u, v)$ avec :

$$\begin{cases} x - x_0 = u \cos \theta_0 - v \sin \theta_0 \\ y - y_0 = u \sin \theta_0 + v \cos \theta_0 \end{cases}$$

Il est possible de retrouver les paramètres de recalage rigide en considérant les images dans le domaine fréquentiel. Soit $S_0(\xi,\eta)$ la TFD de $s_0(x,y)$ et $S_t(\xi,\eta)$ la TFD de $s_t(x,y)$.

- Lorsque la transformation entre les deux images est purement une translation (i.e. pas de rotation), nous avons la relation suivante :

$$S_t(\xi,\eta) = e^{-i2\pi(\xi x_0 + \eta y_0)} S_0(\xi,\eta).$$

Considérant le rapport $S_t S_0^*$, $|S_t S_0^*| = \exp(-i2\pi(\xi x_0 + \eta y_0))$, sa TFD inverse est une distribution de Dirac centrée en (x_0, y_0) : δ_{x_0, y_0} .

¹TFD signifie Transformée de Fourier Discrète.

- Lorsque nous sommes en présence d'une rotation d'angle θ_0 , la relation devient

$$S_t(\xi,\eta) = e^{-i2\pi(\xi x_0 + \eta y_0)} S_0(\xi\cos\theta_0 + \eta\sin\theta_0, -\xi\sin\theta_0 + \eta\cos\theta_0).$$

On pose alors la fonction

$$G(\xi,\eta;\theta) = \frac{S_t(\xi,\eta)}{S_0(\xi\cos\theta + \eta\sin\theta, -\xi\sin\theta + \eta\cos\theta)}$$

où θ est pris comme variable. On obtient de manière évidente pour $\theta = \theta_0$, $G(\xi, \eta; \theta_0) = \exp(-i2\pi(\xi x_0 + \eta y_0))$; lorsque la différence $\theta - \theta_0$ augmente, $G(\xi, \eta; \theta)$ s'éloigne de cette forme exponentielle.

Pour revenir à notre problème de recalage rigide 2D, il suffit de retrouver l'angle θ tel que $g = \text{TFD}^{-1}(G)$ soit le plus proche possible d'une distribution de Dirac, dont le centre nous donne en plus la translation recherchée.

Notons que cette méthode n'est que semi-déterministe car il faut faire appel à un algorithme itératif d'exploration pour trouver θ_0 mais la détermination de la translation (x_0, y_0) , elle, est déterministe.

Cette technique a été mise en œuvre par Sébastien DURBEC pour effectuer une mise en correspondance d'images multimodales. Son rapport de stage [29] présente entre autre l'application de cette méthode au recalage d'images cérébrales de modalité IRM et PET, et montre la robustesse de l'algorithme sur des images bruitées et sur des images avec des parties manquantes.

1.4.1.3 Recalage à l'aide de marqueurs

Un dernier exemple simple d'algorithme déterministe est le cas des algorithmes utilisant des marqueurs sur les images. Les marqueurs sont deux ensembles de points désignés interactivement ou sélectionnés automatiquement par la machine. Les points d'une image sont appariés avec les points d'une autre image.

Considérons le cas d'un recalage rigide en deux dimensions. La transformation (une similitude) à retrouver est parfaitement définie par la donnée de deux points images et de deux points antécédents². Donc, connaissant deux paires de points se correspondant mutuellement sur les deux images, on est capable de retrouver analytiquement les paramètres de la transformation rigide à appliquer pour recaler nos images.

D.L.G. HILL expose dans l'article [41] une méthode de recalage multimodal rigide à l'aide de marqueurs. En trois dimensions, trois points linéairement indépendants par image suffisent pour définir de manière unique la transformation recherchée³. Cependant il propose d'utiliser plus de points (typiquement de 6 à 12) et d'effectuer une minimisation de l'erreur au sens des moindres carrés pour s'affranchir d'éventuelles erreurs de positionnement des marqueurs.

Supposons deux ensembles de n points chacun $\{p_i\}$ et $\{p'_i\}$ supposés image l'un de l'autre par une transformation rigide (rotation et translation) avec une certaine marge d'erreur sur le positionnement. Les deux ensembles de points sont alors reliés par la relation :

$$\forall i \in [1, n], \quad p'_i = \mathbf{R}p_i + \mathbf{T} + \mathbf{N}_i$$

où \mathbf{R} est une matrice (3×3) de rotation, \mathbf{T} est un vecteur de translation et p_i et p'_i sont les coordonnées des points se correspondant mutuellement. \mathbf{N}_i est un vecteur de bruit permettant de tenir compte de l'erreur de positionnement des points.

Trouver le recalage mettant en correspondance nos deux ensembles de points revient à trouver \mathbf{R} et \mathbf{T} minimisant l'expression :

$$f(\mathbf{R},\mathbf{T}) = \sum_{i=1}^{n} \left\| p'_i - (\mathbf{R}p_i + \mathbf{T}) \right\|^2.$$

²À condition que les points ne soient pas confondus.

 $^{^{3}}$ À condition que la taille des voxels dans les deux modalités soit connue pour s'affranchir du problème d'échelle.

Ce problème de minimisation de deux variables peut se découpler en utilisant la remarque de ARUN et al. dans l'article [3] qui stipule que, si nous déterminons les centroïdes p et p' des deux ensembles et que nous construisons les ensembles $\{q_i\}$ et $\{q'_i\}$ avec $\forall i \in [1, n], q_i = p_i - p$ et $q'_i = p'_i - p'$, alors la somme à minimiser devient :

$$f(\mathbf{R}) = \sum_{i=1}^{n} \|q'_i - \mathbf{R}q_i\|^2.$$

Dès lors, pour déterminer \mathbf{R} , on considère la matrice $\mathbf{M} = \sum_{i=1}^{n} q_i q'_i^i$ que l'on décompose selon les valeurs singulières en $\mathbf{M} = \mathbf{UWV}^t$ (ce développement est également repris au chapitre 4, paragraphe 4.4.1.1 page 95). On en déduit, après quelques calculs que $\hat{\mathbf{R}} = \mathbf{VU}^t$, est une matrice orthonormale de déterminant⁴ +1, donc une matrice de rotation, minimisant $f(\mathbf{R})$.

Une fois que nous avons déterminé $\hat{\mathbf{R}}$ nous retrouvons le vecteur de translation avec la relation $\hat{\mathbf{T}} = p' - \hat{\mathbf{R}}p$.

La fonctionnelle dans cet exemple, qui n'est autre que f(), n'est explicitement utilisée que pour donner *in fine* une idée de la qualité du recalage.

Notons enfin l'existence d'algorithmes de recalage élastiques déterministes utilisant le *point-matching* (association de points). Par exemple ROHR *et al.* [79] se servent de la théorie des plaques minces pour faire correspondre au mieux deux ensembles, préalablement identifiés, de points supposés appariés. La transformation associant au mieux ces points est une transformation, soit rigide, soit élastique, constituant la transformation globale à appliquer à l'ensemble de l'image à recaler.

1.4.2 Quelques algorithmes itératifs

Dans le cadre d'un algorithme itératif, la philosophie est de mettre en œuvre une fonctionnelle permettant d'orienter la recherche au fil des itérations. Le choix de la fonctionnelle est très important : de celui-ci dépend, d'une part, la qualité du recalage et d'autre part, la vitesse de convergence de l'algorithme vers une solution optimale globale ou locale.

Ces méthodes sont les plus courantes dans la littérature, aussi nous ne présenterons d'abord que les idées générales, puis ne détaillerons que quelques unes parmi les plus originales.

1.4.2.1 Fonctionnelles à base de corrélation

Étant donné que le but de la fonctionnelle est en général de mesurer la similarité entre deux images (une image de référence et la transformée d'une image à recaler) la première chose à laquelle on pense est de calculer la « cross-correlation » directement sur les niveaux de gris de l'image (par exemple [87]). Comme la corrélation explicite à partir des niveaux de gris présente une combinatoire élevée (surtout pour des images 3D), son emploi est très souvent restreint à quelques cas particuliers. Comme nous le verrons par la suite, on trouve également dans la littérature quelques algorithmes utilisant non pas la corrélation des niveaux de gris mais la corrélation de caractéristiques géométriques (exemple [94]).

La corrélation sur les niveaux de gris

Une mesure de similitude entre deux images peut être la mesure de la corrélation entre ces images [80] selon la formule suivante où f(x, y, z) (respectivement g(x, y, z)) désigne la valeur du niveau de gris de l'image f (respectivement g) au point de coordonnées (x, y, z):

$$R^{2} = \frac{\left(\sum f(x, y, z).g(x, y, z)\right)^{2}}{\left(\sum f^{2}(x, y, z)\right).\left(\sum g^{2}(x, y, z)\right)}.$$

 $^{^{4}}$ Il se peut que l'on trouve un déterminant de valeur -1, dans ce cas, l'automorphisme obtenu est la composée d'une réflexion et d'une rotation.

Cette mesure peut s'effectuer aussi bien dans le domaine de départ que dans le domaine de Fourier. Plus la corrélation entre deux images (ou deux régions seulement des images) est élevée et plus les images (ou les régions) seront similaires. Notons que nous n'avons pas une distance au sens strict du terme (*i.e.* une fonction définie positive telle que d(x, x) = 0 et $d(x, y) \leq d(x, z) + d(z, y)$) mais une fonction affine d'une fonction distance.

L'article [87] compare différentes mesures de similarité à partir des niveaux de gris de deux images médicales 3D (une première image p(x, y, z) de modalité TEP, et une deuxième m(x, y, z) obtenue par résonance magnétique).

Les différentes mesures comparées sont :

1. Le coefficient d'inter corrélation γ .

Il est normalisé en fonction du volume de recouvrement (partie a priori commune aux deux images, choisie pour la mesure de la qualité du recalage) des deux images 3D et de l'intensité moyenne des images sur le volume de recouvrement.

$$\gamma = \frac{\sum_{xyz} \left[p(x, y, z) - \bar{p} \right] \left[m(x, y, z) - \bar{m} \right]}{\sqrt{\sum_{xyz} \left[p(x, y, z) - \bar{p} \right]^2 \sum_{xyz} \left[m(x, y, z) - \bar{m} \right]^2}}$$

où \bar{p} et \bar{m} représente l'intensité moyenne des images TEP et IRM sur le volume de recouvrement. En terme de distribution 2D d'intensités de voxels, cette expression se ramène à :

$$\gamma = \frac{\sum_{m,p} (m - \bar{m})(p - \bar{p})h(m, p)}{\sqrt{\sum_{m,p} h(m, p)(p - \bar{p})^2} \sqrt{\sum_{m,p} h(m, p)(m - \bar{m})^2}}$$

où h(m, p) est le nombre de voxels ayant respectivement comme valeur m et p dans les deux modalités et se correspondant mutuellement à l'intérieur du volume de recouvrement. h(m, p) est donc un histogramme 2D. Lorsque les deux images sont recalées, γ est maximal. Notons que γ ne dépend pas du nombre de voxels de recouvrement mais nécessite une certaine corrélation des intensités des voxels d'une modalité à une autre, ce qui est en général le cas puisque les appareil imageurs ne font que rendre compte de réalités physiques.

2. Le minimum de variance d'intensité.

Cette mesure est proposée par Woods dans l'article [100]. Elle considère que pour une intensité donnée dans l'image IRM le nombre d'intensités correspondant point à point dans l'image TEP doit être minimal au recalage optimal. Si $\bar{p}(m)$ est la valeur moyenne des points de l'image TEP correspondant aux intensités m de l'image IRM, et $\sigma_p(m)$ est l'écart-type pour les même points, l'écart-type normalisé des valeurs TEP correspondant à une valeur m de l'IRM est :

$$\sigma'_n(m) = \sigma_p(m)/\bar{p}(m),$$

ce qui, en terme de distribution 2D d'intensité de voxels donne

$$\sigma_p'(m) = \frac{1}{\bar{p}(m)} \sqrt{\frac{\Sigma_p h(m, p)(p - \bar{p}(m))^2}{\Sigma_p h(m, p)}}$$

Cette valeur est minimale lorsque les images sont parfaitement corrélées. Une somme pondérée de cet écart-type normalisé sur toute les intensités IRM est alors une mesure de similarité sur le volume de recouvrement :

$$\sigma_p^{\prime\prime} = \Sigma_m \left[\sigma_p^{\prime}(m) \cdot \frac{\Sigma_p h(m, p)}{V} \right],$$

où $V = \sum_{m,p} h(m,p)$ est le nombre de voxels dans le volume de recouvrement. La pondération assure que la mesure de similarité est surtout influencée par la variance des valeurs de l'image TEP pour les valeurs les plus courantes de l'image IRM.

3. Moments de la distribution des valeurs sur l'histogramme 2D.

Cette approche est proposée par Hill (coauteur de l'article [87]). Il s'agit d'étudier, en calculant des moments d'ordre supérieur, la forme du pic principal de l'histogramme 2D des correspondances des voxels d'une image à une autre. Il remarque que lorsque l'on approche du recalage optimal, les pics de l'histogramme augmentent en hauteur et les régions à faibles valeurs se raréfient. Ceci a pour effet une augmentation de la distorsion de la distribution des valeurs de h(m,p), qui peut être mesurée par les moments d'ordre supérieur de cette distribution. Si u(h) est le nombre d'occurrences d'une valeur particulière de h pour toutes les valeurs de m et de p, alors le moment d'ordre n est défini comme :

$$M_n = \sum_h u(h).h^n,$$

ce qui peut être normalisé en divisant par le moment d'ordre zéro et en divisant h(m, p) par le volume V de recouvrement, ce qui à pour effet justement de rendre la mesure indépendante de ce volume de correspondance.

$$M_n = \frac{\Sigma_h u(h) \cdot \left(\frac{h}{V}\right)^n}{\Sigma_h u(h)}.$$

4. Entropie de l'histogramme 2D.

Enfin, la dernière mesure étudiée à l'article [87] est l'entropie de l'histogramme 2D qui mesure la stochasticité de la distribution (cette notion sera de nouveau abordée page 28 et plus amplement définie dans une note de bas de page). Cette mesure est définie de la manière suivante :

$$E = -\sum_{m,p} \frac{h(m,p)}{V} \log\left(\frac{h(m,p)}{V}\right).$$

Cette mesure normalisée sur le volume de recouvrement doit être minimisée pour obtenir le recalage optimal.

Ces mesures ont été utilisées dans le cadre de la recherche d'un recalage rigide (six paramètres) entre images cérébrales 3D. Les résultats obtenus sont discutés dans [87].

On retient que la corrélation γ est une mesure en général peu efficace, car très gourmande en temps de calcul, mais qui peut être utilisée dans une recherche multirésolution aux résolutions les plus basses.

La minimisation des moments d'ordre supérieur (dans l'article [87], c'est l'ordre 3 qui est utilisé) a donné de mauvais résultats couplée avec un algorithme d'optimisation de type descente de gradient (cf. section 1.5). Les résultats obtenus étaient en fait très liés au point de départ que l'on donnait à l'algorithme de recherche (initialisation), ce qui n'est exploitable que pour affiner une solution et non pour faire une recherche sans connaissance a priori.

La maximisation de l'entropie de l'histogramme 2D donna également des résultats peu fiables, en fait du même ordre que ceux obtenus avec la maximisation des moments d'ordre supérieur.

Par contre, la minimisation de la variance des correspondances entre intensités (niveaux de gris) a procuré les résultats les plus intéressants. Pour plus de renseignement se reporter à l'article [87], mais notons tout de même que cette mesure a permis également de faire, de manière assez robuste, des recalages d'images IRM et scanner partiellement tronquées.

Dans le même ordre d'idée, on peut citer l'article [69] de Christophoros NIKOU. Celui-ci présente une méthode de recalage sous-voxels d'images médicales multimodales en utilisant un M-estimateur⁵ robuste minimisant une variance inter-région d'intensité suite à un partitionnement des images. Cette approche, tout comme les cas 2 et 4 des corrélations citées précédemment (variance d'intensité, entropie de l'histogramme 2D), utilise l'hypothèse d'homogénéité interimage⁶, hypothèse au moins vraie dans le cadre du recalage monomodal et en général valable pour la plupart des recalages multimodalités.

La corrélation de caractéristiques géométriques

La plupart des recalages multimodaux simplifie en première approche le problème en transformant les images à niveaux de gris en un ensemble de caractéristiques, par exemple géométriques (points particuliers, courbes, plans, surfaces, volumes). Le recalage s'effectue alors sur des caractéristiques (p.ex. deux surfaces numériques), et le problème de la mesure de similarités entre des images de types différents se résume à la comparaison de caractéristiques communes.

⁵En analyse statistique, le but de l'estimation robuste est de trouver le vecteur de paramètres $\hat{\Theta}$ qui fait correspondre un modèle $M(X_i, \Theta)$ aux observations y_i , lorsque les données X_i sont bruitées. Un M-estimateur consiste à rechercher les paramètres d'un modèle qui minimisent une erreur définie comme la somme d'une fonction de résidus. D'une manière générale $\hat{\Theta} = \arg\min_{\Theta} \sum_i \rho(y_i - M(\Theta, X_i), \sigma)$, où ρ est appelé un Mestimateur puisque cette minimisation correspond à l'estimation du Maximum de vraisemblance, si ρ est interprété comme l'inverse de la vraisemblance en logarithme conditionnelle aux observations.

⁶Une zone homogène dans une image doit être également homogène dans l'autre image.

On présentera plus particulièrement dans ce paragraphe les travaux réalisés par l'équipe de recherche *Computer Vision Group, University Hospital Utrecht*⁷, menés par Petra A. VAN DEN ELSEN; la technique de recalage étudiée repose sur l'extraction de caractéristiques géométriques [59, 94].

L'intérêt des travaux de cette équipe a été de privilégier l'approche automatique, et ce sans avoir recours à aucune sorte d'information extrinsèque. Les travaux de thèse de J.B.Antoine MAINTZ ont porté sur le recalage multimodalité utilisant les caractéristiques géométriques des images⁸.

À partir des images niveaux de gris des différentes sources (dans le cas présent les modalités étudiées étaient le scanner et l'IRM), on procède d'abord à une extraction des caractéristiques communes aux deux modalités. Dans le cadre de l'étude du cerveau, on observe que le graphe d'une coupe scanner du crâne présente un ensemble de crêtes (ou ridges en anglais) correspondant à des canaux (troughs anglais); la figure 1.3 montre ces correspondances; leurs homologues IRM constituent des crêtes négatives. Étant donné que le cerveau (enfermé dans la boîte crânienne) est relativement rigide, la donnée spatiale des crêtes et des canaux suffit à décrire de façon suffisante le recalage à effectuer.



(a) Coupe scanner (TDM). L'os cortical correspond à des intensités élevées de niveaux de gris.



(b) Coupe IRM correspondante. L'os correspond à des intensités faibles.

Figure 1.3 — Paire de coupes TDM et IRM recalées; les intensités fortes sur l'image TDM (crêtes) correspondent à des intensités faible sur l'image IRM (canaux).

La première étape est donc l'extraction des crêtes et canaux, et la création de deux images binaires ou à niveaux de gris symbolisant les emplacements de ces caractéristiques géométriques. Le recalage sera ensuite piloté par la mesure de similarité entre ces deux nouvelles images.

Dans le cadre du recalage d'images 2D (tranches) d'un cerveau (cf. [94]), où les transformations envisageables sont seulement les transformations rigides (i.e. rotations et translations), la méthode utilisée est la suivante :

1. Lissage.

Les images scanner ou IRM du cerveau sont souvent très bruitées, aussi la détection de contours (crêtes et creux) ne peut être effectuée qu'après un lissage des images. Ce lissage est mis en œuvre à partir de la convolution d'un filtre Gaussien de taille (écart-type) σ . La largeur de cet opérateur est fonction de la résolution intrinsèque des images. Par exemple, pour une coupe scanner de 256×256 pixels, la taille du filtre a été choisie à $\sigma = 4$ pixels. Une expression mathématique du lissage peut être la suivante : $L(x,\sigma) = (L_0 * \mathcal{G})(x,\sigma)$, où L_0 est l'image d'intensité initiale, \mathcal{G} est le noyau Gaussien, x est le vecteur de coordonnées, σ est le facteur de lissage, et L est l'image lissée⁹ (voir figure 1.4).

2. Détection des crêtes et des creux.

Pour chaque point de l'image on définit un repère local (\vec{v}, \vec{w}) orienté dans le sens trigonométrique direct. \vec{w} est dirigé selon le gradient local et orienté selon la ligne de plus grande pente (vers le haut), et \vec{v} est choisi

⁷Ville des Pays-Bas, chef-lieu de la province du même nom.

⁸cf. l'URL http://gallery.cv.ruu.nl/Research/ComputerVision/Registration/Matching.

⁹Notons que le calcul d'une dérivée de L revient à remplacer la Gaussienne \mathcal{G} avec sa dérivée appropriée dans l'opération de convolution : $(L_{i_1...i_n})(x,\sigma) = (L_0 * \mathcal{G}_{i_1...i_n})(x,\sigma)$, où l'indice i_j désigne l'ordre de la dérivation par rapport aux variables spatiales $i \in \{x, y, z\}, n \in \mathbb{N}^+, j = 1...n$.



(a) Image d'intensité initiale : $L_0(x)$. (b) Image lissée : $L(x, \sigma)$.

Figure 1.4 — (a) Une coupe scanner (matrice de 256 × 256 pixels) représentant une surface d'intensité (les pixels ayant une grande intensité forment des montagnes). Les crêtes correspondant au crâne, à la peau et au cadre de maintien de la tête sont visiblement très bruitées. (b) Après avoir lissé l'image avec un filtre de taille approprié (ici, $\sigma = 4$ pixels), on découvre une crête bien définie. (Cette image est issue de *Brain Topography*, vol. 5, no. 2, pp 153–158 et *Proc. SPIE*, vol. 1808, pp 172–186).



Figure 1.5 — Croquis d'une image contenant une crête, représenté comme une surface d'intensité. La direction de \vec{v} est perpendiculaire au gradient \vec{w} et le couple (\vec{v}, \vec{w}) forme un repère orienté dans le sens trigonométrique direct. Le gradient est en général orienté vers la crête, sauf sur la crête elle même où il est orienté selon la direction de la crête. Le profil de l'intensité (L) selon la direction de \vec{v} pour les points situés dans un proche voisinage de la crête (ex : point P) est très convexe par rapport au profil obtenu pour les points éloignés de la crête (ex : point Q).

orthogonalement à \vec{w} . Lorsque l'on traverse un extremum, le gradient inverse sa direction. Au voisinage d'un point situé sur une crête, le profil de l'image selon la direction de \vec{v} (\perp à la crête ou au creux) est beaucoup plus convexe que si l'on se trouvait au voisinage d'un point quelconque (cf. figure 1.5). Cette remarque se traduit par une dérivée seconde dans la direction de \vec{v} beaucoup plus négative (resp. positive) au voisinage d'une crête (resp. d'un creux). Un opérateur appelé L_{vv} efficace pour la détection des creux et des crêtes utilisant les dérivées secondes a été introduit. Pour adopter une notation plus mathématique, $w_i = L_i$, et $v_i = \epsilon_{ij}L_j$ en notation tensorielle. Aussi L_{vv} représente la dérivée seconde dans la direction perpendiculaire au gradient local. La valeur de L_{vv} peut être calculée de la manière suivante :

$$L_{vv} = \frac{1}{\|\vec{v}\|^2} (\vec{v}.\vec{\nabla})^2 L$$

= $(L_y^2 L_{xx} - 2L_x L_y L_{xy} + L_x^2 L_{yy}) (L_x^2 + L_y^2)^{-1}$

La notion de ligne de crêtes à été généralisée par Thirion [91] au cas des isosurfaces au sein d'une image 3D, où le critère d'extrémalité fait intervenir les courbures principales locales $(\vec{k_1} \text{ et } \vec{k_2})$; le calcul de ces dernières, faisant appel aux dérivées 3^{es}, est particulièrement sensible au bruit et rend sont utilisation délicate. Remarquons également à ce stade que l'opérateur considéré (en 2D ou en 3D) étant local, il ne permet pas de résoudre les points ou les régions critiques, c'est-à-dire les points où $L_w = 0$ et, en 3D, $k_1 = k_2$ (également appelés ombilics).

La généralisation de l'opérateur L_{vv} à un espace à trois dimensions n'est pas triviale. Dans le cas 2D, la direction de \vec{v} est parfaitement déterminée, alors qu'en 3D, le sous-espace perpendiculaire au gradient est un plan, ce qui donne une infinité de directions possibles pour \vec{v} . La généralisation fait intervenir les surfaces « isophotes »¹⁰ ainsi que les lignes de flux (qui sont en tout point perpendiculaires aux surfaces isophotes). La figure 1.6 montre la disposition géométrique des lignes de flux et des isophotes. Le gradient local de l'image est en tout point tangent aux lignes de flux. La mesure de l'appartenance à un contour (L_{vv} dans le cas 2D) est en fait définie comme étant la dérivée seconde minimale de l'intensité, prise parmi toutes les directions perpendiculaires au gradient local.



 $Figure \ 1.6$ — Disposition géométrique des isophotes et des lignes de flux (flowline) vis-à-vis de l'arête (ridge).

D'autres opérateurs permettant l'extraction des crêtes et des creux dans les images sont discutés dans l'article [58]. On y trouve notamment :

- L_{vv}/L_w , souvent noté -k et appelé la courbure isophote. La normalisation de L_{vv} par la norme du gradient permet à l'opérateur de mieux réagir dans les régions de l'image où les variations d'intensité sont faibles.
- $L_{vv}L_w^{\alpha}$, qui est une généralisation directe des deux opérateurs d'extraction de crêtes cités précedemment (avec $\alpha = 0$ puis $\alpha = -1$). Les valeurs usuelles du coefficient α que l'on trouve également sont $\alpha = 2$ (mesure d'angularité ou cornerness), alpha = -1/2 et $\alpha = 1/2$.
- L_w, opérateur de détection d'arêtes (et non de crêtes), le plus efficace (pour ce qui concerne l'application au recalage d'images selon les résultats de l'article [58]), et également le plus rapide.
- L_{vw}/L_w qui représente la courbure le long des lignes de flux (flowline, cf. figure 1.6), et où une ligne de flux est définie comme l'intégrale curviligne du gradient.
- L_{ww} qui est le résultat de la densité isophote (L_{ww}/L_w) multipliée par L_w . Cette mesure est liée à la mesure du Laplacien et du L_{vv} par la relation : $L_{ww} = L_{ii} L_{vv}$. Cet opérateur n'ayant pas permi d'obtenir de bons résultats a été écarté par les auteurs.

La méthode de recalage utilisée ensuite pour mettre en correspondance deux volumes de caractéristiques (les images initiales traitées avec un ou plusieurs des opérateurs présentés précédemment) est basée sur l'inter-corrélation des niveaux de gris. En utilisant directement ces niveaux de gris, les auteurs évitent une segmentation des images de caractéristiques. Si nous notons L_1 et L_2 les volumes de caractéristiques des images de départ, la fonctionnelle c(t) à optimiser (maximiser dans ce cas) s'écrit pour une transformation t :

$$c(t) = \sum_{(x,y,z) \in L_1} L_1(x,y,z) L_2(t(x,y,z)).$$

Un deuxième aspect de leur recherche est l'approche multirésolution. Le but avoué de cette approche est d'accélérer la recherche de la solution optimale, puisque qu'une méthode de type exhaustif aurait un coût calculatoire prohibitif. À partir des images des caractéristiques géométriques, on crée de nouvelles images sous-échantillonnées. Les images sont donc plus petites et le calcul de corrélation entre ces images est grandement accéléré. Le processus de recalage s'opère alors comme suit :

- 1. Création d'une pyramide d'images sous-échantillonnées;
- 2. Partir de la résolution la plus faible et effectuer classiquement le recalage ;
- 3. Effectuer le recalage à l'ordre supérieur en se basant sur les résultats obtenus par l'étage inférieur;
- 4. Continuer jusqu'à la résolution originelle.

¹⁰En 2D les isophotes sont les lignes d'égale intensité, et en 3D ce sont les surfaces à intensité constante.

Parmi les diverses techniques de sous-échantillonnage existantes, les techniques retenues sont celles qui remplacent un ensemble de voxels de l'image de départ (jusqu'à 8 voxels maximum) par le minimum ou le maximum en valeur de ces voxels, ou encore la valeur médiane qui est plus robuste au bruit. C'est la solution qui permet de conserver au mieux l'information géométrique, contrairement aux méthodes consistant par exemple à prendre la valeur moyenne.

Enfin ces auteurs mentionnent également la possibilité d'adapter un tel algorithme multirésolution sur une machine parallèle.

En résumé il y a au moins deux avantages à utiliser une approche multirésolution (ou multiéchelle) :

- 1. Cela diminue considérablement le temps de traitement. Il est en effet possible de faire beaucoup plus d'itérations à une faible résolution (moins de détails à traiter) qu'à une résolution fine;
- 2. Pour l'anatomie humaine, les caractéristiques macroscopiques sont généralement plus stables que les détails microscopiques qui sont souvent très bruités.

La corrélation sur des images binarisées

Un cas particulier de l'extraction de caractéristiques géométriques est la transformation des images médicales en images binaires. Cette binarisation ne sera pas un simple seuillage de l'image mais plus une segmentation qui permettra d'en faire ressortir des éléments caractéristiques (surtout les contours, ou les lignes de crêtes [58]). Une image binaire sera donc composée d'un ensemble de points ayant chacun deux valeurs possibles. Une non nulle pour les points faisant partie des objets, et zéro partout ailleurs. Disposant des images binarisées, les calculs de distance sont alors simplifiés. La corrélation par exemple se traduit par une simple multiplication point à point suivi d'un comptage des points $actifs^{11}$. Considérons fet g comme étant deux images binaires représentant les contours d'un même objet calculés à partir d'une image à recaler et d'une image de référence. La qualité c(T) de la transformation T envisagée pourra alors être donnée par la formule suivante :

$$c(T) = \sum_{(x,y,z) \in \mathcal{D}_f} f(x,y,z).g(T(x,y,z))$$

1.4.2.2 Fonctionnelles à base de distances euclidiennes

Étant donné que l'extraction de caractéristiques géométriques (amers, courbes, surfaces) est la méthode la plus couramment utilisée, beaucoup de problèmes de recalage se résument à la minimisation d'une distance entre des objets géométriques. Pour ce faire, on introduit souvent la notion de distance euclidienne, ou d'approximation de distance euclidienne, entre les objets géométriques.

Distance euclidienne sur une série de points

Lorsque l'on travaille avec des informations extrinsèques comme les marqueurs cutanés, des cadres ou des masques, ceux-ci présentent des points particuliers en général facilement repérables sur les images. La mesure de distance revient alors à effectuer des mesures uniquement sur ces points. On peut ainsi définir une distance euclidienne entre la transformée de ces points et leur position réelle sur l'image. Par exemple, si nous disposons de N marqueurs cutanés de coordonnée X_i sur l'image de référence et Y_i sur l'image à recaler, et si T désigne la fonction de transformation de l'image à recaler vers l'image de référence alors le carré de la distance D peut s'exprimer de la manière suivante :

$$D^{2} = \sum_{i=1}^{N} (X_{i} - T(Y_{i}))^{t} \cdot (X_{i} - T(Y_{i})) avec X_{i} = (x_{i}, y_{i}, z_{i})^{t}$$

¹¹Un point actif est un point de valeur non nulle.

Cette méthode peut également s'appliquer dans le cas où l'utilisateur désigne interactivement une série de points sur une image et leur correspondant sur une autre image [72]. Le principal inconvénient est que la saisie manuelle de la position des points sur les images ajoute une imprécision supplémentaire, et rend le recalage moins performant.

Le recalage de caractéristiques géométriques

Un certain nombre d'approches a été proposé pour prendre en compte les informations de structure dans le processus de recalage, comme par exemple les courbes, les plans et les surfaces. Toutes ces méthodes utilisent en général les informations intrinsèques des images et génèrent des approximations globales. Le but est de construire un modèle physique des appareils imageurs et des objets visualisés et de chercher ensuite le meilleur ajustement des paramètres de ce modèle. Le recalage est effectué en comparant une image transformée et une image fixe. Pelizarri et d'autres chercheurs [72] ont développé une méthode populaire de recalage paramétrique 3D applicable aux images tomographiques du cerveau. Un modèle surfacique de la « tête » (*head*) extrait depuis une image est mis en correspondance avec un ensemble de points formant le « chapeau » (*hat*), extraits des contours d'une autre image, par une transformation rigide¹². Le « chapeau » est fixé sur la « tête » à l'aide d'une stratégie cherchant à minimiser la distance quadratique moyenne des points du chapeau à la surface de la tête. Il est à noter que l'évaluation de la fonction de performance (ou fonction de coût) peut être optimisée en utilisant une transformée distance sur les images des propriétés extraites.

Le principe de cette méthode du « head and hat » vient de la question : « Comment placer de manière optimale un chapeau d'une forme précise sur une tête ? ». Ce qui revient à trouver la position optimale du chapeau pour que sa distance moyenne à la tête soit minimale, ou autrement dit, pour qu'il ait une surface de contact avec la tête maximale. Bien sûr ce problème n'aurait pas de solution unique dans le cadre du recalage entre des formes pures et régulières (une calotte sphérique sur une sphère par exemple). Le problème de PELIZARRI est donc de positionner un modèle du chapeau qui est un ensemble de points distincts, sur un modèle de la tête qui est une surface. Dans le cadre des images médicales, le « chapeau » est un ensemble de points issus de la segmentation de la surface externe du crâne sur une image IRM 3D, et la « tête » est la surface obtenue par segmentation de la surface externe du crâne sur une image scanner 3D.

Le calcul de la distance entre le chapeau et la tête est effectué non pas en calculant pour chacun des points du chapeau la distance exacte à la surface représentant la tête mais en considérant la distance qui sépare ce point et le point de la tête qui est à l'intersection de la surface avec le segment de droite allant du centroïde de la tête au point du chapeau (cf. figure 1.7).



Figure 1.7 — Calcul de la distance de la «tête» au «chapeau».

Si on pose d_i la distance entre le point *i* du chapeau et le point d'intersection correspondant sur la surface, alors la fonctionnelle à minimiser s'écrit $D = \sum_{i=1}^{n} d_i$.

¹²ou éventuellement affine

L'exploration de l'espace de recherche, ou la minimisation de la fonctionnelle s'effectue grâce à la méthode de PowELL (voir section 1.5), qui ne nécessite pas la connaissance des dérivées de la fonction.

La robustesse de cette méthode a été validée par son application sur des images de fantômes possédant des points de référence fixes facilement repérables dans les deux modalités.

Un classique en matière de recalage : le *chamfer matching* ou recalage par chanfrein

Afin de présenter une des utilisations du recalage par la méthode du chanfrein¹³, nous faisons ici référence à l'article [102] et à leurs auteurs, J. YOU et W.P. ZHU, chercheurs à l'*University of South Australia*, ainsi qu'à E. PISSALOUX, professeur à l'Université de Paris XI et H.A. COHEN, chercheur à *La Trobe University, Australia*.

L'article en question présente un nouvel algorithme de recalage hiérarchique par chanfrein, fondé sur la détection de points « intéressants ». Cet algorithme est en fait un prolongement de la méthode traditionnelle de recalage par chanfrein en utilisant la notion de points « intéressants » à la place des *edge points* (ou points d'arêtes) dans le calcul de la transformée distance pour la mesure de la qualité du recalage. Une série d'images avec un nombre différent de points intéressants pour caractériser l'image originelle est créée de façon pyramidale grâce à une méthode de seuillage dynamique. Le recalage est effectué à partir de cette pyramide depuis le niveau brut (niveau le plus grossier dans la pyramide, celui où le seuillage est le plus fort) jusqu'au niveau fin (le niveau où il y a beaucoup plus de détails dans les images) en minimisant un critère de recalage en terme de distance entre les points dits intéressants. Le but de cette structure hiérarchique est de diminuer le coût de calcul. L'algorithme est simple à implémenter et assez robuste au bruit et autres perturbations. De plus une telle méthode de recalage hiérarchique peut facilement être mise en œuvre sur un réseau de machines avec PVM (*Parallel Virutal Machine*) simulant à un prix très raisonnable une machine parallèle et accélérant considérablement le traitement.

Le recalage par chanfrein a été proposé pour la première fois en 1977 par H.G. BARROW [6] pour trouver la meilleure association de points d'arêtes issus de deux images différentes. L'algorithme est fondé sur l'utilisation de transformée distance. Son utilisation repose sur la description des images au moyen de ces fameux points d'arêtes, ceux-ci permettant de quantifier l'erreur de recalage entre les deux images et ainsi de piloter l'évolution des équations paramétriques représentatives de la transformation géométrique. L'application d'une telle technique est limitée puisqu'elle nécessite une importante hypothèse de départ, à savoir une méthode de repérage des points d'arêtes. Néanmoins, cette méthode a montré sa capacité à traiter correctement des données imparfaites ou bruitées. Aussi, une extension du recalage par chanfrein fût proposée par G. BORGEFORS [11, 12]. Elle introduisit un nouveau critère de recalage (movenne quadratique pour la mesure de correspondance entre les objets) ainsi que la notion de pyramide de résolution. Chacun des objets à recaler doit être vaguement défini par un polygone (polyèdre dans le cas 3D). L'utilisation d'une transformée distance permet de calculer rapidement et à tout moment la qualité d'un recalage. Cette qualité est donnée en terme de distance entre deux polygones (ou polyèdres). Cette distance est fonction de la somme, pour chacun des points d'un polygone, des distances de son transformé au polygone de l'autre image.

Le recalage par chanfrein consiste donc à créer à partir d'une carte binaire de caractéristiques géométriques (typiquement des points, des droites et des surfaces) une carte distance,

¹³ Chamfer Matching est le terme anglais désignant le recalage par chanfrein. Le mot chanfrein provient du vieux français chant signifiant côté, et du verbe fraindre signifiant briser. Un chanfrein est par conséquent la surface obtenue en abattant une arête d'un objet anguleux.

i.e. une image dont la valeur de chacun des points représente la distance euclidienne à la plus proche caractéristique. Concrètement, sur cette image tout les points de valeur nulle sont les points représentatifs des caractéristiques géométriques à recaler (points, droites et surfaces dans le cas de la 3D). Il est souvent considéré comme trop coûteux de calculer la valeur exacte de la distance euclidienne c'est pourquoi BARROW propose un algorithme en donnant une valeur approchée en un temps de calcul dépendant de manière linéaire du nombre de points et non de manière exponentielle.

La première étape de l'algorithme consiste à créer une image dont les points caractéristiques ont pour valeur zéro et les autres une valeur infinie (ou non définie). Puis en deux passes (dans le cas des images 2D), une « avant » et une « arrière » on affecte à chaque point une valeur fonction de la valeur précédente du point courant (F[i, j]) et de quatre de ses voisins.

Pour la première passe (« avant »), l'algo-	et lors de la deuxième passe (« arrière »), on
rithme est le suivant :	a :
Pour $i = 2$ à nb_lignes	Pour $i = nb_{lignes-1}$ à 1
Pour $j = 2$ à nb_colonnes	$\mathbf{Pour} \ \mathbf{j} = \mathbf{nb} \ \mathbf{colonnes} \mathbf{-1} \ \mathbf{\hat{a}} \ 1$
$\mathrm{F}[\mathrm{i},\mathrm{j}]=\mathrm{min}(\mathrm{F}[\mathrm{i},\mathrm{j}],$	$\mathrm{F}[\mathrm{i},\mathrm{j}]=\mathrm{min}(\mathrm{F}[\mathrm{i},\mathrm{j}],$
$\mathrm{F}[ext{i-1,j}]+2,$	F[i+1,j] + 2,
F[i-1,j-1] +3,	${ m F[i+1,j+1]}+3,$
$\mathrm{F}[\mathrm{i},\mathrm{j} ext{-}1]+2,$	$\mathrm{F[i,j+1]}$ + 2,
F[i+1,j-1] + 3)	F[i-1,j+1] + 3)
Fin Pour j	Fin Pour j
Fin Pour i	Fin Pour i

Le chiffre 3 permet de donner une approximation entière de $2\sqrt{2}$. Cet algorithme porte le nom d'algorithme de chanfrein 2/3.

Il existe d'autres versions de l'algorithme de chanfrein, notamment une avec des coefficients de distance horizontale, verticale, et diagonale de 3/4. Pour une image carrée de $M \times M$ pixels, la plus grande erreur observable entre la distance de chanfrein 2/3 et la vraie distance euclidienne est d'environ $0.134 \times (M - 1)$ et de $0.081 \times (M - 1)$ entre le chanfrein 3/4 et la vraie distance euclidienne.

Plus proches voisins itératif (ICP)

Une méthode populaire également utilisée dans le recalage d'images est le calcul des plus proches voisins itératif (ou *Iterative Closest points*, ICP). Cet algorithme a été mis au point par BESL et MCKAY au début des années 90 [8]. Cette approche ne définit pas une fonctionnelle à proprement parler mais plutôt un algorithme complet de recalage. Il est néanmoins intéressant de constater qu'il existe des méthodes fondées sur le calcul de distances euclidiennes entre deux ensembles d'entités géométriques comme les points, les segments de droites, les courbes implicites et paramétriques, les ensembles de triangles, les surfaces implicites et paramétriques.

L'algorithme ICP est en général utilisé pour rechercher les six paramètres d'une transformation rigide 3D mettant au mieux en correspondance deux ensembles d'entités géométriques. Pour des raisons pratiques, l'ensemble de données à recaler sur le modèle sera un ensemble de points (échantillonnage ponctuel d'une entité géométrique).

Le calcul d'une distance est équivalent à rechercher le minimum d'une norme sur un ensemble d'espaces convexes et est particulièrement adapté à l'utilisation de la méthode de Newton. Cette méthode part d'une première approximation du minimum et oriente sa recherche en fonction de la valeur du gradient et de la dérivée seconde (le Hessien dans le cas 3D) de la fonctionnelle à minimiser.

Par exemple si $\vec{r}(\vec{u})$ est une entité géométrique paramétrique (dont \vec{u} est le paramètre), et si \vec{p} est un point, la fonction scalaire à minimiser pour trouver la distance du point à l'entité géométrique est :

$$f(\vec{u}) = \|\vec{r}(\vec{u}) - \vec{p}\|^2$$

En posant ensuite $\vec{\nabla} = [\partial, \partial \vec{u}]^t$ l'opérateur de gradient vectoriel (où ^t désigne la transposition vectorielle), f atteint son minimum lorsque $\vec{\nabla}f = 0$. Pour le cas où \vec{r} est une surface de paramètres (\vec{u}, \vec{v}) , le gradient 2D devient $\vec{\nabla}f = [f_u, f_v]^t$ et la matrice Hessienne 2D est définie par

$$\vec{\nabla}\vec{\nabla}^t(f) = \begin{bmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{bmatrix}.$$

L'algorithme de Newton s'applique de la manière suivante : soit \vec{u}_0 une première approximation du minimum de la fonction f, on calcule \vec{u}_n par la formule de récurrence suivante

$$\vec{u}_{k+1} = \vec{u}_k - [\vec{\nabla}\vec{\nabla}^t(f)(\vec{u}_k)]^{-1}\vec{\nabla}f(\vec{u}_k).$$

Il est a noter qu'il existe une version accélérée de l'algorithme ICP qui utilise une prédiction parabolique plutôt que linéaire lors des dernières itérations. On remarque, en effet, que la plupart des fonctionnelles à optimiser sont continues tout comme leurs dérivées première et seconde. Il en résulte un profil en forme de parabole autour du minimum cherché.

La pondération de caractéristiques géométriques : WGF

WGF qui signifie *Weighted Geometrical Features* est une technique introduite par MAURER et FITZPATRICK dans l'article [33]. L'article présente en effet une méthode de recalage d'images cérébrales d'une part en modalité scanner-X et d'autre part en IRM. Cet article compare en outre trois procédés de recalage à partir de structures segmentées. Les auteurs constatent que l'algorithme le plus significativement efficace est celui qui cherche à apparier un ensemble de points avec une surface plutôt qu'associer deux ensembles de points ou deux surfaces uniquement.

Le transformation T recherchée ici est une transformation rigide, c'est-à-dire composé d'une rotation et d'une translation.

La mise en œuvre est la suivante. Soit $\{P_i\}, \forall i \in [1, N_s]$, un ensemble de formes d'une image, à recaler sur une autre image comprenant l'ensemble de formes $\{X_i\}, \forall i \in [1, N_s]$. On commence par décomposer chaque forme P_i en un ensemble de points $P_i = \{p_{ij}\}, \forall j \in [1, N_{p_i}]$ (si $\{P_i\}$ n'était pas déjà un ensemble de points). Soit $\{w_{ij}\}$ un ensemble de points associés avec les points $\{p_{ij}\}$, permettant d'attribuer un coefficient de confiance à chaque point. Nous cherchons alors la transformation rigide T qui minimise la fonctionnelle suivante

$$l(T) = \sqrt{\sum_{i=1}^{N_s} \sum_{j=1}^{N_{p_i}} w_{ij} ||y_{ij} - T(p_{ij})||^2}$$

avec

 $y_{ij} = C_i[T(p_{ij}), X_i]$ le point de la forme X_i le plus proche possible du transformé de p_{ij} (donc C_i est l'opérateur « plus proche voisin »).

De la même manière que pour l'algorithme de recalage fondé sur l'ICP simple, la recherche du minimum se fait de manière itérative. L'article [33] donne une preuve de la convergence simple de cet algorithme.

La distance de Haussdorf

D'autres types de distances, comme la distance de *Haussdorf* ont été mises à l'épreuve et ont permis de concevoir des algorithmes encore plus robustes aux différentes perturbations que sont les imprécisions dans le positionnement des points d'arêtes ou tout simplement le bruit que peuvent subir les images. Au paragraphe 1.4.2.2 nous avons parlé de l'utilisation des cartes distances (qui est une approximation par échantillonnage du concept des surfaces de VORONOÏ). L'inconvénient de cette méthode est qu'elle mesure une distance entre une image et un espace (celui de la carte distance) au lieu de mesurer directement une distance entre deux images. La distance de Hausdorff permet justement d'éviter cela.

Le calcul de la distance de Hausdorff entre deux images s'effectue en considérant les images comme des ensembles de points binaires. La mesure sur des images à niveaux de gris doit par conséquent s'effectuer après une segmentation de celles-ci [43].

Étant donnés deux ensembles finis de points $A = \{a_1, \ldots, a_p\}$ et $B = \{b_1, \ldots, b_q\}$, la distance de Hausdorff est définie par :

$$H(A, B) = \max(h(A, B), h(B, A)) \text{ avec } h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||,$$
(1.1)

où $\|.\|$ désigne une norme¹⁴ sur les points de A et de B.

La fonction h(A, B) est appelée la distance de Hausdorff orientée de A vers B. En effet, h(A, B) classe chaque point de A en fonction de sa distance au plus proche point de B, et le point ayant alors le plus grand classement (le point de A ayant la plus grande distance) spécifie la valeur de la distance de Hausdorff orientée. Intuitivement, si h(A, B) = d, alors chaque point de A doit être éloigné au plus de d de chacun des points de B. Notons qu'en général h(A, B)et h(B, A) peuvent atteindre des valeurs très différentes (les distances orientées ne sont pas symétriques).

Le calcul de H(A, B) ne nécessite pas de déterminer une correspondance explicite point à point entre les deux ensembles (images). Au contraire, il arrive souvent qu'un certain nombre de points d'une image soient tous proches d'un même point de l'autre image. En terme de rapidité ou de coût de calcul, c'est un atout important par rapport aux méthodes utilisant des modèles et calculant point à point les correspondances possibles.

De plus amples détails le calcul de la distance de Hausdorff ainsi que quelques digressions quant à son utilisation pour le recalage en imagerie médicale sont présentés à la section 2.3.1

1.4.2.3 Une fonctionnelle utilisant la théorie de l'information

Il existe des méthodes de recalage multimodal n'ayant pas recours à l'extraction de caractéristiques géométriques ni au calcul d'une distance Euclidienne (ou d'une approximation). Ces méthodes utilisent directement les niveaux de gris des images à étudier, dans le cas où ces valeurs ont la caractéristique d'être corrélées (images d'un même patient en IRM et TDM par exemple). Nous présenterons plus particulièrement dans cette section les travaux de recherche qui ont été effectués par William M. WELLS¹⁵¹⁶, Paul VIOLA¹⁶¹⁷, Hideki ATSUMI¹⁸, Shin NAKAJIMA¹⁵ et Ron KIKINIS¹⁵. Leur technique de recalage est une maximisation stochastique de l'information mutuelle entre les différentes modalités (voir référence [97]).

L'idée de base est la suivante : considérons le problème du recalage de deux images issues de la même modalité (disons deux images par résonance magnétique, ou IRM) et de la même

 $^{^{14}}$ En général on utilise la norme L_2 ou norme euclidienne, cependant les résultats sont les mêmes avec n'importe qu'elle norme L_p .

¹⁵ Harvard Medical School and Brigham and Women's Hospital, Department of Radiology.

¹⁶ Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

¹⁷ The Salk Institute, Computational Neurobiology Laboratory.

¹⁸ Harvard Medical School and Brigham and Women's Hospital, Department of Neurosurgery.
personne. Quand ces deux images sont parfaitement alignées, les signaux doivent être identiques. Une mesure simple de la qualité d'un recalage hypothétique est la somme quadratique des différences voxel à voxel. Cette mesure peut s'expliquer d'un point de vue probabiliste. Si les images IRM sont entachées d'un bruit gaussien, de distribution uniforme et indépendante entre les deux images, alors la somme quadratique des différences voxel à voxel est directement proportionnelle à la probabilité que les deux images soient correctement recalées. Malheureusement, cette opération ne peut s'appliquer directement au cas de la multimodalité. Même lorsqu'elles sont parfaitement recalées, des images IRM et scanner d'un même individu sont assez différentes. C'est justement ce qui fait que la conjonction de ces images est intéressante d'un point de vue biologique.

Cependant les images IRM et scanner ne sont pas totalement indépendantes, aussi il est possible d'imaginer qu'avec beaucoup d'expérience on puisse construire une fonction F(.) qui soit capable de prédire (ou tout du moins d'approximer) la valeur d'un voxel dans une modalité connaissant son correspondant dans l'autre modalité. À partir de cette fonction F, il est alors possible d'évaluer un recalage comme dans le cas de la monomodalité en comparant F(IRM)et l'image scanner par exemple. Mais même si théoriquement il est possible de trouver une telle fonction F et de l'utiliser de cette manière, en pratique c'est un problème très difficile et sous-déterminé, et c'est pourquoi on essaye de quantifier une information mutuelle a posteriori entre les images, plutôt que d'essayer de trouver la relation F a priori.

Étant donné qu'une image IRM et une image scanner sont corrélées, il existe une information mutuelle entre ces deux modalités. La méthode proposée par l'équipe de Harvard est de trouver et d'utiliser une fonction F en se servant directement de l'information mutuelle qui existe entre les images. Une telle technique essaye donc de trouver le recalage en maximisant l'information qu'une image volumétrique donne à propos de l'autre image. La technique proposée ne requiert pas de connaissance *a priori* sur la relation entre les modalités, elle considère seulement qu'un volume donne le plus d'information sur l'autre volume lorsqu'ils sont recalés.

Mentionnons enfin que l'information mutuelle est définie en terme d'entropie, dont une approximation est calculée de manière stochastique. La méthode d'optimisation utilisée ensuite pour rechercher le minimum global de l'entropie, correspondant à un maximum global de l'information mutuelle est une descente de gradient stochastique (cf. section 1.5).

En voici les principes mathématiques :

Étant données une image de référence f et une image à recaler g, désignons par T la transformation recherchée permettant de faire correspondre chaque point de g avec son homologue sur f. La transformation optimale \hat{T} est obtenue en maximisant l'information mutuelle I entre f et $g \circ T$. $\hat{T} = \underset{T}{\operatorname{argmax}} I(f(x, y, z), g \circ T(x, y, z))$. L'information mutuelle

I de deux variables aléatoires est définie comme étant relative $\hat{\dot{a}}$ la somme des entropies individuelles de chacune des variables aléatoires, moins l'entropie conjointe du couple de variables :

$$I(f(x, y, z), g \circ T(x, y, z)) \equiv h(f(x, y, z)) + h(g \circ T(x, y, z)) - h(f(x, y, z), g \circ T(x, y, z))$$

h() est l'entropie d'une variable aléatoire et est définie par :

$$h(x) \equiv -\int p(x) \ln p(x) dx$$

et par l'entropie conjointe de deux variables aléatoires x et y qui s'exprime de la manière suivante :

$$h(x,y) \equiv -\int p(x,y)\ln p(x,y)dxdy$$

La densité de probabilité p(z) est approximée par la formule de PARZEN $p(z) \approx P^*(z) = \frac{1}{N_A} \sum_{z_j \in A} R(z - z_j)$ où A est un échantillon de N_A éléments de z et R est une fonction porte de largeur unité. En pratique on lui préfère une fonction de densité gausienne (voir expression de $G_{\psi}(z)$, où ψ est la matrice de covariance) ce qui simplifie l'étude, mais n'est pas obligatoire. Notons que n'importe quelle fonction différentiable aurait pu être utilisée.

$$G_{\psi}(z) \equiv (2\pi)^{\frac{-n}{2}} |\psi|^{\frac{-1}{2}} \exp(-\frac{1}{2}z^{t}\psi^{-1}z)$$

D'autre part, l'expression de l'entropie est difficilement utilisable sous sa forme intégrale, aussi l'approximation suivante est utilisée :

$$h(z) \approx -E_z[\ln P^*(z)] = -\int \ln P^*(z)dz$$
$$\approx -\frac{1}{N_B} \sum_{z_i \in B} \ln P^*(z_i)$$

où N_B est la taille d'un second échantillon B. La moyenne de l'échantillon converge vers la vraie espérance à un taux proportionnel à $1/\sqrt{N_B}$. La dernière étape consiste à établir une expression de $\frac{d}{dT}I(T)$. En assumant que la matrice de covariance ψ_w des fonctions de PARZEN multi-dimensionnelles est une matrice diagonale, $\psi_w = \text{Diag}(\psi_{ff}, \psi gg)$, où Diag désigne la construction d'une matrice diagonale à partir d'une liste d'éléments, on obtient :

$$\begin{split} \widehat{\frac{dI}{dT}} &= \frac{1}{N_B} \sum_{x_i \in B} \sum_{x_j \in A} (g_i - g_j)^{\mathrm{T}} \left[W_g(g_i, g_j) \psi_{gg}^{-1} - W_w(w_i, w_j) \right] \frac{d}{dT} (g_i - g_j) \\ &\text{avec} \\ W_g(g_i, g_j) &\equiv \frac{G_{\psi_v}(g_i - g_j)}{\sum_{x_k \in A} G_{\psi_v}(g_i - g_k)}, \quad W_w(w_i, w_j) \equiv \frac{G_{\psi_w}(w_i - w_j)}{\sum_{x_k \in A} G_{\psi_w}(w_i - w_k)} \\ &f_i \equiv f(x_i), \ g_i \equiv g(T(x_i)), \ \text{et} \ w_i \equiv [f_i, g_i]^{\mathrm{T}} \end{split}$$

Disposant maintenant d'une approximation de la variation de l'information mutuelle en fonction de ses paramètres, la recherche du maximum local de l'information mutuelle se fait en utilisant un analogue stochastique de la descente de gradient.

Répéter :			
	A	\leftarrow	$\{N_A \text{ échantillons de } x\}$
	B	\leftarrow	$\{N_B \text{ \' echantillons de } x\}$
	T	\leftarrow	$T + \lambda \frac{\widehat{dI}}{dT}$

Le paramètre λ est le taux d'apprentissage. La précédente procédure est répétée un nombre de fois fixé préalablement ou bien jusqu'à ce que la convergence soit détectée. Il faut cependant prendre certaines précautions dans l'utilisation d'une telle procédure. Si l'on voulait par exemple trouver la meilleure rotation en utilisant la notation matricielle T, il faut veiller à ce que $T + \lambda \frac{dI}{dT}$ reste une matrice de rotation.

L'équipe de recherche de Willian M. WELLS et Paul VIOLA n'est pas la seule à avoir utilisé la théorie de l'information pour effectuer des recalages. On peut noter par exemple dans le domaine médical les articles [69, 87] (voir paragraphe 1.4.2.1), les articles de COLLIGNON [20, 21, 22], l'article de BOKLYE et MEYER [52].

1.4.2.4 Une fonctionnelle à base de démons

Nous mentionnons ici l'approche originale des travaux effectués au sein de l'équipe de recherche de l'INRIA, et menés par Jean-Philippe THIRION et Nicolas AYACHE. Ces études ont porté sur « une méthode rapide de recalage non-rigide d'images médicales 3D ». Ce qui différencie les travaux de cette équipe est l'utilisation de « démons¹⁹ », qui déforment localement le modèle pour le faire tenir dans un moule, contrairement aux techniques plus traditionnelles basées sur des « attracteurs ».

Avant de développer l'originalité de la méthode (utilisation des démons), il faut préciser que les techniques classiques de recalage non-rigide font appel à un compromis entre l'action de forces intérieures (tenant compte de la rigidité du modèle) et celles des forces extérieures

¹⁹Dans un sens très similaire aux démons de MAXWELL.

(les attracteurs entre caractéristiques similaires). On voit bien que l'on est obligé de définir cette rigidité sans laquelle on serait capable de transformer n'importe quel objet en n'importe quel autre. On ne ferait alors plus du recalage mais du morphing! Cette notion de rigidité est pourtant largement subjective, bien qu'elle puisse être définie statistiquement à partir d'une base de données de recalages, mais nous entrons alors dans un cercle vicieux difficile à résoudre.

Parmi les recherches privilégiant l'aspect « attracteur » certaines invoquent la théorie de la mécanique des fluides pour modéliser l'élasticité et la déformation des volumes $3D^{20}$. L'inconvénient de cette méthode de recalage élastique est qu'il n'existe pas de raison a priori pour que la mécanique des fluides soit une bonne modélisation des différences inter-patients. Les similarités inter-patients proviennent de leur phylogénie²¹ commune ainsi que d'une ontogénie²² similaire. Ceci n'est pas mentionné ici à titre gratuit, nous voulions simplement attirer l'attention sur le fait que les modèles utilisés pour décrire les transformations 3D, qu'elles soient rigides ou non, ne sont souvent pas justifiés. La modélisation de la transformation par le biais de démons n'est en effet pas plus pertinente que celle considérant les organes étudiés comme des fluides visqueux. Cependant, l'utilisation d'une modélisation dont les paramètres sont bien maîtrisés et dont le comportement est proche de celui du phénomène observé simplifie considérablement le problème étudié, tout en conservant une bonne qualité aux solutions qui en dérivent.

La suite de ce paragraphe est donné uniquement à titre indicatif sur le fonctionnement de la méthode originale, et n'est pas fondamentalement nécessaire à la compréhension du reste de la thèse.

Le fonctionnement des démons

Afin de décrire la méthode des démons employée à l'INRIA, nous allons considérer le cas du recalage en 2D d'un objet aux contours bien délimités. Le modèle déformable est un contour dont différents échantillons (points) sont connus. Dans le cas des méthodes classiques utilisant les attracteurs (voir figure 1.8), ces points échantillonnés sont attirés par des points caractéristiques de la scène (ici ce sont les points du contour de l'objet). Les forces d'attraction peuvent être calculées simplement à partir d'une fonction « distance aux plus proches points de la scène », ou bien à partir de considérations plus évoluées comme la mesure de similarité de courbures.



Figure 1.8 — Modèle déformable et attracteurs.

Dans le cas des démons, il est nécessaire de connaître en plus du contour de l'objet son orientation locale (afin de pouvoir décider de quel côté se situe l'intérieur et l'extérieur de l'objet), et ceci à la fois pour le modèle et pour la scène. Cette fois-ci c'est le contour de l'objet (dans la scène) qui est échantillonné en un certain nombre de points associés à une orientation et à un démon. Cette dernière entité se rapporte aux démons de MAXWELL; elle a été introduite

²⁰Images projetées sur des grilles 3D entièrement déformables.

 $^{^{21}}$ La phylogénie est l'étude de la formation et de l'évolution des espèces animales et végétales en vue d'établir leur parenté.

 $^{^{22}{\}rm L'ontogénie}$ est l'étude du développement d'un organisme, de puis l'état embryonnaire jusqu'au stade adulte.

en physique pour modéliser l'idée paradoxale qu'une membrane semi-perméable pourrait violer le deuxième principe de la thermodynamique²³. Situés dans cette membrane, ces démons sont supposés trier localement deux différentes sortes de molécules provenant d'un mélange gazeux, produisant alors une diminution du désordre, et par conséquent une diminution d'entropie. La clé de ce paradoxe fut donnée par Léon BRILLOUIN, s'inspirant d'un travail antérieur de SZILLAR : il démontra que l'exercice de ses fonctions cognitives par le démon devait nécessairement consommer une certaine quantité d'énergie qui, dans le bilan de l'opération, compensait précisément la diminution d'entropie du système.

De même que les démons de MAXWELL trient les molécules de gaz, les démons utilisés pour recaler nos images agissent localement pour pousser le modèle déformable dans la direction normale au contour de la scène, et ce déplacement est orienté en fonction de la nature (intérieur ou extérieur) de l'estimée du point dans le modèle. Si le point est à l'intérieur du modèle, le point correspondant du modèle sera poussé vers l'intérieur, et vice versa (voir figure 1.9).



Démons poussant vers l'extérieur

Figure 1.9 — Modèle déformable et démons.

Intuitivement, cela revient à pousser le contenu du modèle à l'intérieur de l'objet, et le reste à l'extérieur. En d'autre termes, cela revient à trier les points intérieurs et extérieurs du modèle.

Une comparaison entre une méthode basée sur les attracteurs et une autre sur les démons est donnée à la figure 1.10. Le recalage est un recalage rigide entre deux disques. Pour le cas des attracteurs (méthode globale, ligne du haut) les forces ont pour origine le modèle déformable (le disque qui se déplace), et sont dirigées vers le point le plus près du disque fixe. Pour la méthode utilisant les démons (méthode locale, ligne du bas), les forces ont pour origine la frontière de la scène (le disque fixe), et sont orientées vers l'intérieur ou vers l'extérieur de ce disque selon que les points en correspondance sont à l'intérieur ou à l'extérieur du modèle. Remarquons que pour les deux méthodes l'intensité des forces décroît régulièrement. Notons également que la direction des forces est différente, ce qui montre que les deux méthodes sont bien différentes, et que cette différence n'est pas du à un changement du système de référence. Il faut voir également que la méthode utilisant les démons, mais en contrepartie les démons doivent estimer l'intérieur et l'extérieur de l'objet, chose inutile dans le cadre des attracteurs. La généralisation de cette méthode 2D au cas de la 3D est faite par l'utilisation de démons répartis sur la surface des objets et déformant le modèle dans une direction orthogonale à cette surface.

Modélisation mathématique des démons

Voyons maintenant dans le cadre de l'imagerie médicale comment ont été modélisés mathématiquement les démons. Pour les images médicales, les contours d'intensités constantes sont souvent très liés aux formes des objets. Ceci vient du fait que l'intensité représente la densité (C'est notamment le cas pour les images scanner et IRM). Désignons par

 $^{^{23}}$ Il est peut-être nécessaire de rappeler succinctement ce qu'est ce deuxième principe : Dans sa première forme, purement thermodynamique (énoncée par CLAUSIUS en 1850, comme généralisation du théorème de CARNOT) le deuxième principe prévoit que, dans une enceinte énergétiquement isolée, toutes les différences de température doivent tendre à s'annuler spontanément. Ou encore, et cela revient au même, le principe stipule qu'au sein d'une telle enceinte où la température serait uniforme, il est impossible qu'apparaissent des différences de potentiel thermique entre différentes régions du système. Or dans une enceinte à température uniforme, où ne subsiste plus aucune différence de potentiel, aucun phénomène (macroscopique) ne peut avoir lieu. Le système est inerte. C'est dans ce sens que l'on dit que le deuxième principe prévoit la dégradation inéluctable de l'énergie au sein d'un système isolé, tel que l'univers. L' « entropie » est la quantité thermodynamique qui mesure le niveau de dégradation de l'énergie d'un système. Selon le deuxième principe par conséquent, tout phénomène, quel qu'il soit, s'accompagne nécessairement d'un accroissement d'entropie au sein du système où il se déroule.



Figure 1.10 — Comparaisons sur 6 itérations de la convergence d'un algorithme fondé sur les attracteurs (méthode globale, ligne du haut) et d'un autre utilisant les démons (méthode locale).

f (resp. g) les fonctions 3D représentant l'image d'intensité de la scène (resp. du modèle). Nous associons un démon à chaque voxel P de f aux endroits où le gradient $|\vec{\nabla}f|$ n'est pas nul. Une surface de niveau dont l'équation implicite est f = f(P), et dont la normale orientée est le vecteur $\vec{\nabla}f(P)$, passe par ce voxel P. Le démon situé en P pousse l'image modèle dans la direction $\vec{\nabla}f(P)$ si f(P) < g(P) et dans la direction $-\vec{\nabla}f(P)$ si f(P) > g(P) (voir la figure 1.11). C'est ainsi qu'une grille 3D complète de démons agit pour déformer le modèle.



Figure 1.11 — Le modèle g est poussé (\vec{p}) par le démon dans la direction $\vec{\nabla}f(P)$ si f(P) < g(P) et dans la direction $-\vec{\nabla}f(P)$ si f(P) > g(P).

Pour être encore un peu plus précis sur l'expression mathématique des démons dans le cadre de l'imagerie médicale, prenons l'hypothèse de la conservation de l'intensité des points à déplacer (i.e. i(x(t), y(t), z(t)) = const). En dérivant cette équation nous obtenons :

$$\frac{\partial i}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial i}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial i}{\partial z}\frac{\partial z}{\partial t} = -\frac{\partial i}{\partial t}, \quad \text{ou alors} \quad \vec{\nabla}i.\vec{v} + \frac{\partial i}{\partial t} = 0.$$

En ce qui nous concerne, nous avons uniquement deux images f et g à comparer, et nous recherchons le déplacement \vec{v} qui amène g le plus proche possible de f, aussi nous considérons que f et g sont séparés par une unité de temps : $\frac{\partial i}{\partial t} = f - g$, et $\vec{v}(dx/dt, dy/dt, dz/dt)$ est la vitesse instantanée de g à f. En conclusion :

$$\vec{v}.\vec{\nabla}f = g - f.$$

Cependant cette équation ne suffit pas à calculer \vec{v} localement. Nous utilisons alors la méthode suivante (voir figure 1.12). Une approximation au premier ordre de la surface d'intensité passant par le point P en f est un hyperplan (P, f(P)) ayant pour normale $(-\vec{\nabla}f(P), 1)$. L'équation $\vec{v}.\vec{\nabla}f = g - f$ implique que l'extrémité du vecteur \vec{v} coïncide avec l'intersection de cet hyperplan et du plan horizontal passant par (P, g(P)). Sans aucune autre information, il est naturel de choisir pour \vec{v} le point le plus proche de P dont l'équation est :

$$\vec{v} = \frac{(g-f)\vec{\nabla}f}{\vec{\nabla}f^2}.$$

Intuitivement c'est la plus petite translation amenant le point g(P) sur l'hyperplan (P, f(P)) ayant pour normale $(-\nabla f, 1)$, ce qui fait que g est localement proche de f. Pour nos démons, nous pourrions prendre comme force poussante

p ce vecteur-vitesse \vec{v} , ce qui correspond bien au critère d'orientation que nous avions mentionné au début. Cependant cette équation devient instable lorsque la norme du gradient est faible. Dans un pareil cas, une faible variation de l'intensité pourrait rejeter l'extrémité de \vec{v} à l'infini, dans n'importe quelle direction, ce qui est loin de l'effet désiré, d'où l'idée de multiplier par un coefficient correctif $(\vec{\nabla}f^2 + k^2))$ où k est une constante dépendant du contraste de l'image. Afin d'éviter de rajouter un nouveau paramètre au problème on préfère choisir pour k la valeur g - f, ce qui produira le même effet de limitation sur l'intensité des forces.



Figure 1.12 — Vitesse instantanée de g à f.

En résumé la force poussante \vec{p} associée à un démon est donnée par la formule suivante :

$$\vec{p} = \frac{(g-f)\vec{\nabla}f}{\vec{\nabla}f^2 + (g-f)^2}$$

On pourrait trouver des expressions plus complexe pour les forces exercées par les démons, afin de leur donner un comportement encore plus stable face aux irrégularités que l'on peut rencontrer dans les images (par exemple lorsque certains détails dans une image n'ont pas d'homologue dans l'autre image). Néanmoins, il faut faire un compromis entre une expression compliquée et robuste, et une expression moins robuste mais beaucoup plus rapide. D'autres expressions possibles sont détaillées dans l'article [92].

L'implémentation des démons

À partir de deux images 3D, la scène f et le modèle g, voyons comment utiliser concrètement les démons. À chaque itération i du processus, g_i est l'image g déformée par la transformation courante T_i . On commence avec une transformation initiale $T_0 = Id$, et l'on cherche la transformation T_n qui rende g_n et f les plus similaire possible, n étant l'itération finale. Chaque itération i est décomposée comme suit :

- 1. Pour chaque démon P de f, calculer la poussée du démon en fonction de la forme locale de g_i en P, qui est g au point $P_i = T_i^{-1}(P)$ et de la forme locale de f en P.
- 2. Calculer une déformation élémentaire δ_T pour l'image entière, à partir du champ vectoriel de forces poussantes (ou δ_T⁻¹ à partir des forces inverses). C'est précisément l'étape déterminante dont dépend la stabilité de la méthode (en particulier vis-à-vis de la manière dont est traité la rigidité). Actuellement, δ_T est déterminé à partir d'un lissage gaussien (dont l'écart-type permet de paramètrer la rigidité de la déformation) du champ vectoriel de forces poussantes.
- 3. Appliquer la déformation élémentaire δ_T pour obtenir la nouvelle déformation globale : $T_{i+1} = \delta_T \circ T_i$ (ou $T_{i+1}^{-1} = T_i^{-1} \circ \delta_T^{-1}$).

Notons également que l'utilisation de démons se prête très bien à une implémentation multiéchelle puisque cela permet de rendre l'action locale des démons beaucoup moins locale.

Quelques restrictions dans l'utilisation des démons

Puisqu'aucune méthode n'est parfaite, et comme le font remarquer les auteurs de la méthode développée à l'INRIA, il y a essentiellement deux restrictions à l'utilisation des démons pour le recalage élastique. Il faut quand même noter que ces restrictions sont partagées par la majorité des techniques de recalage élastique dans le domaine de l'imagerie médicale.

- Tout d'abord, la méthode des démons est très sensible au positionnement initial des images. Des tests ont permis en effet de montrer que si les images étaient initialement très éloignées, le recalage était quasiment impossible à retrouver. Certaines méthodes comme celles fondées sur une diminution adaptative de l'espace de recherche (cf. la méthode développée à l'ENSTBR [81]) sont plus robuste face à un mauvais positionnement initial. Remarquons toutefois que la robustesse au positionnement initial se fait au détriment de la rapidité de recherche de la solution optimale. Si les démons ne sont pas adaptés au problème du mauvais positionnement initial, ils convergent en revanche plus rapidement que les autres algorithmes lorsqu'ils ne sont pas trop éloignés de la solution optimale. D'un autre côté, il faut voir que la recherche d'une position initiale pertinente présente également un coût significatif. Si l'on veut utiliser la méthode des démons, il faudra certainement y adjoindre en Le deuxième facteur limitatif est que la méthode des démons ne peut s'appliquer que pour des images dont les intensités sont très similaires, voire identiques. Ceci est un problème en général commun à tous les algorithmes de recalage inter-modalités. Malheureusement, dans le cas d'images réelles, les contrastes inter-tissus varient largement d'une modalité à une autre. On pourrait envisager, en se servant de connaissances a priori, de corriger les images et rendre ces intensités similaires, mais même si l'on connaissait exactement les courbes des réponses des appareils imageurs face à chaque variété de tissus, le problème resterait insoluble. En effet il n'y a que très rarement des correspondances bijectives entre deux modalités différentes. Cela se constate simplement en remarquant que certains tissus peuvent être facilement différenciés par le biais d'une modalité tout en étant complètement indifférentiables sur une autre.

1.5 Stratégie d'exploration de l'espace de recherche

Maintenant que nous connaissons différentes fonction d'appréciation du recalage et que nous avons défini les différents types de transformations que l'on peut rechercher, l'étape suivante consiste à trouver les paramètres correspondant à un extremum global de cette fonction. Lorsque l'espace de recherche n'est pas très grand (par exemple si la transformation ne dépend que d'un paramètre) on peut envisager une recherche exhaustive sur tout le domaine. Mais cela est rarement le cas dans un cadre général.

Il faut donc se munir d'algorithmes d'exploration des solutions sur l'espace de recherche. Bien que de tels algorithmes existent, ils ne peuvent en général pas assurer que l'on trouve *la solution optimale*, et ceci pour des raisons de rapidité de convergence. Il y a donc un compromis à faire entre recherche globale et rapidité de convergence locale.

1.5.1 Recherche locale

1.5.1.1 La méthode dite de descente de gradient (ou Hill climbing)

Descente standard

En supposant les paramètres de la transformation tous indépendants il est possible de faire une recherche sur chacun des paramètres les uns après les autres. Il suffit pour cela d'estimer l'influence de la variation d'un paramètre sur la fonction distance, et de modifier ce paramètre en conséquence. Par exemple, si l'augmentation d'un paramètre tend à faire diminuer la fonction distance, on va continuer à l'augmenter, et ce jusqu'à ce que la distance cesse de diminuer. L'inconvénient majeur d'une telle méthode, est que l'obtention d'une solution dépend fortement des valeurs initiales des paramètres. Par conséquent, il y a de fortes chances de tomber sur un minimum local (cf. figure²⁴ 1.13). La descente de gradient ne permet pas en pratique d'obtenir de bons résultats. Un autre inconvénient, est que cette méthode ne fonctionne pas dans le cas où les variables sont corrélées. Une valeur optimale d'un paramètre changerait en fonction de l'évolution des valeurs des autres paramètres. Néanmoins, cette méthode est parfois utile. Elle permet, lorsque les valeurs de tous les paramètres sont proches de l'optimal, de converger rapidement vers la solution.

En effet, le problème de cet algorithme de descente de gradient est qu'il s'arrête au premier minimum local rencontré. Cela signifie que toujours descendre en suivant les lignes de plus grande pente n'est pas forcément la meilleure heuristique pour trouver le minimum absolu. Deux adaptations de cette heuristique sont alors possible pour améliorer les performances de la descente de gradient.

²⁴La plupart des figures et schémas suivants sont extraits de l'excellent ouvrage [1].



Figure 1.13 — Hill Climbing : convergence vers un minimum local.

Descente avec inertie

La descente de gradient avec inertie consiste à pousser un peu plus loin la métaphore mécanique du processus de descente. En effet, imaginons un vélo dans une descente. Lorsque le vélo arrive en bas, il ne s'arrête pas subitement sous prétexte que la pente est inversée. Le phénomène qui lui permet de continuer encore sur sa lancée est appelé l'inertie. Cette inertie peut être dimensionnée à volonté. en pratique, on part de l'hypothèse qu'il y a de plus fortes chances de trouver des minima locaux au début de la descente (on dimensionne donc une forte inertie). On suppose de même qu'à la fin de cette descente, c'est-à-dire lorsque l'on est proche du minimum absolu, on a moins de chance de trouver un minimum local (on dimensionne alors une inertie faible). Il faut bien remarquer qu'une inertie forte permet de remonter des pentes, ce qui est utile pour sortir d'un minimum local, mais ce qui est nuisible lorsque l'on a trouvé le minimum absolu car dans ce cas là on préfère ne pas en ressortir.



Figure 1.14 — La descente avec inertie.

Descente stochastique

Cette extension de l'algorithme de descente de gradient est abordé au paragraphe 1.5.2.2.

1.5.1.2 Méthode de Powell

La méthode dite de Powell est un algorithme introduit en 1964 [73] par M.J.D. POWELL pour trouver le minimum d'une fonction de plusieurs variables sans avoir recours au calcul ou à l'estimation des dérivées de la fonction.

Cet algorithme est en fait une extension de l'algorithme dit de changement de variable tour après tour. Le changement est tel que si on l'applique pour trouver le minimum d'une quadrique dans un espace à n dimensions, il choisit des directions de recherche qui permettent de trouver l'optimum en un nombre d'itérations minimal²⁵.

L'algorithme est itératif dans le sens où la recherche du minimum se fait en partant d'un point p_0 et en cherchant le point final p_n en créant une trajectoire de points p_k , k allant de 0 à *n* de manière récurrente. À chaque itération k, on commence par effectuer une recherche locale le long de n directions indépendantes de l'espace $\xi_1, \xi_2, \ldots, \xi_n$ autour de p_k .

Là où la méthode diffère de la technique du « changement de variable tour à tour » est qu'après chaque itération on définit une nouvelle direction de recherche ξ et on propose de faire la recherche à l'étape suivante sur les directions $\xi_2, \xi_3, \ldots, \xi_n, \xi$. On choisit la direction ξ de telle sorte que pour la minimisation d'une quadrique, la recherche soit optimale, c'est-àdire qu'après k itérations, les k dernières directions de recherches soient toutes mutuellement $\operatorname{conjuguées}^{26}$.

Une itération de l'algorithme se présente ainsi : 1. Pour r = 1, 2, ..., n calculer λ_r qui minimise $f(p_{r-1} + \lambda_r \xi_r)$, et définir $p_r = p_{r-1} + \lambda_r \xi_r$.

- 2. Pour r = 1, 2, ..., n 1 remplacer ξ_r par ξ_{r+1} . 3. Remplacer ξ_n par $(p_n p_0)$. 4. Choisir λ tel que $f(p_n + \lambda \{p_n p_0\})$ soit minimum.
- 5. Finalement remplacer p_0 par $p_0 + \lambda \{p_n p_0\}$.

Les preuves de convergence absolue et un exemple d'application de l'algorithme sont donnés à l'article [73]. Notons enfin que cette méthode d'optimisation a été appliquée avec succès pour le recalage d'images médicales avec la technique du « head and hat » de PELIZARRI et al. [72], présentée au paragraphe 1.4.2.2.

1.5.2Recherche globale

1.5.2.1**Recherche exhaustive**

Les méthodes d'optimisation dites de recherche exhaustives sont très peu employées car souvent l'espace de recherche considéré est trop important. Néanmoins il faut savoir que cette méthode existe et peut être employée par exemple dans le cadre de l'association de points (ou *point-matching*) quand le nombre de points caractéristiques à faire correspondre est peu important. C'est notamment le cas pour une sélection interactive de points caractéristiques sur des images 2D ou 3D. Une fois connus un certain nombre de points intéressants sur une image de référence et sur une image à recaler, l'algorithme peut essayer de trouver les correspondances entre certains de ces points. La combinatoire de ce problème restant faible (si le nombre de points n'est pas trop élevé), une recherche exhaustive est envisageable. Toutefois, les approches utilisant une sélection interactive de points caractéristiques sur des images médicales imposent souvent à l'utilisateur (ou plutôt à l'expert) de préciser directement quelles sont les correspondances entre les points des deux images, ce qui évite ainsi toute étape d'optimisation, et réduit l'algorithme de recalage à un algorithme de calcul d'une transformation, globale ou locale, à partir d'une liste d'amers.

²⁵En fait on démontre que l'algorithme converge absolument vers le minimum global pour une quadrique. ²⁶C'est la condition pour avoir une convergence absolue.

Méthode des moments

À partir de l'image de départ, on se propose de décrire les formes qui y sont incluses. Pour cela, toute une théorie sur les moments a été élaborée. Notamment toute image 3D est parfaitement décrite soit par la donnée de la valeur de ses niveaux de gris point à point, et ceci par l'intermédiaire de sa fonction f(x, y, z), soit par l'ensemble de ses moments [80] $\{m_{pqr}\}$ avec :

$$m_{pqr} = \int_x \int_y \int_z x^p y^q z^r f(x, y, z) dx dy dz.$$

L'intérêt est que si nous connaissons les descripteurs de deux images ayant des similitudes, alors des relations existent entre ces descripteurs. Donc en comparant les descripteurs de deux images ayant a priori des similitudes, nous sommes théoriquement capables de retrouver les relations entre ces images (*i.e.* translations, rotations et facteurs d'échelle).

Il est à noter que cette méthode peut aussi bien s'appliquer à des images en niveaux de gris qu'à des images binaires. Cependant les résultats sont moins biaisés sur des images binaires que sur des images en niveaux de gris puisque la distributions des niveaux de gris n'est pas forcément la même pour les deux images (surtout pour des images de modalités différentes). Pour une image binaire, l'expression des descripteurs devient :

$$m_{pqr} = \sum_{x} \sum_{y} \sum_{z} x^{p} y^{q} z^{r} f(x, y, z) \text{ avec } f(x, y, z) = 0 \text{ ou } 1.$$

Recalage utilisant les informations de structure

Un certain nombre d'approches ont été proposées pour prendre en compte les informations de structure dans le processus de recalage, comme par exemple les courbes, les plans et les surfaces. Toutes ces méthodes utilisent en général les informations intrinsèques des images et génèrent des approximations globales. Le but est de construire un modèle physique des appareils imageurs et des objets visualisés et de chercher ensuite le meilleur ajustement des paramètres de ce modèle. Le recalage est effectué en comparant une image transformée et une image fixe. Pelizarri et d'autres chercheurs [72] ont développé une méthode populaire de recalage paramétrique 3D applicable aux images tomographiques du cerveau. Un modèle surfacique de la « tête » (head) extrait depuis une image est mis en correspondance avec un ensemble de points formant le « chapeau » (hat) extraits des contours d'une autre image par une transformation rigide²⁷. Le « chapeau » est fixé sur la « tête » à l'aide d'une stratégie cherchant à minimiser la distance quadratique moyenne des points du chapeau à la surface de la tête; cette stratégie que nous avons présenté au paragraphe 1.5.1.2 constituait une recherche locale, néanmoins, les mêmes informations de structure (« head and hat » peuvent être utilisées avec un algorithme déterministe global comme l'algorithme de recherche exhaustive, à condition que la taille de l'espace de recherche ne soit pas trop importante. Comme nous le verrons plus loin, il est également possible d'utiliser les informations de structures extraites des images avec un algorithme global mais cette fois stochastiques (pour les grands espaces de recherche) comme par exemple les algorithme génétiques. Enfin, Il est à noter que l'évaluation de la fonction de performance (ou fonction de coût) peut être optimisée en utilisant une transformée distance sur les images des propriétés extraites.

 $^{^{\}rm 27}{\rm ou}$ éventuellement affine

1.5.2.2 Algorithmes stochastiques

Descente stochastique

La descente stochastique a été proposée par WIDROW et HOFF dans les années 60. Au lieu de minimiser directement la distance en minimisant la fonction somme, on préfère minimiser indépendemment les termes de la somme (dépendant d'un paramètre à la fois), et ceci de manière itérative. On espère ainsi introduire un peu de hasard (et donc ne pas rester sur un minimum local) en passant d'un paramètre à un autre de manière indépendante. Il est démontré que cette démarche revient à suivre en moyenne les lignes de plus grande pente.



Figure 1.15 — La descente stochastique.

Notons que WELLS [97] (voir paragraphe 1.4.2.3) utilise un tel algorithme de descente stochastique pour effectuer la maximisation de l'information mutuelle entre deux images médicales à recaler.

Cela dit, si la fonction distance est vraiment très irrégulière, alors toutes les méthodes issues d'une descente de gradient échouent. Ce serait le cas pour une fonction distance du type de celle de la figure 1.16.



Figure 1.16 — Fonctionnelle très bruitée.

Algorithmes génétiques

Les algorithmes génétiques sont des outils d'optimisation fondés sur les mécanismes d'évolution des êtres vivants dans un milieu (théorie de l'évolution de Darwin). En raison de leurs bonnes capacité à résoudre des problèmes généraux sur un grand espace de recherche en un temps raisonnable, nous avons décidé d'utiliser ces algorithmes pour la recherche du recalage optimal. Le chapitre 3 étant dédié à la présentation et à l'étude des algorithmes génétiques, nous ne nous étendrons pas plus sur leur sujet dans cet état de l'art.

Notons quand même l'existence dans la littérature de travaux concernant la recherche de paramètres de recalage à l'aide d'algorithmes génétiques [16, 46, 48, 81, 86].

Recuit simulé

Le recuit-simulé (ou *simulated annealing* [53]) propose de faire une analogie avec le recuit, un principe de physique du solide à l'origine de son nom. L'explication en détail d'une telle théorie, fondée sur la thermodynamique statistique, dépasserait le cadre de ce rapport. Cependant nous pouvons montrer le synoptique général d'un tel algorithme (cf. figure 1.17).



Figure 1.17 — Principe du recuit-simulé, « filtre de hasard ».

Le fait que les algorithmes de recuit-simulé fonctionnent également comme un filtre du hasard les rendent assez proche conceptuellement des algorithmes génétiques. Ces algorithmes sont en effet assez rapides et robustes face au problème des minima locaux. La principale différence avec les algorithmes génétiques est que ces derniers ne considèrent pas un point isolé de l'espace de recherche migrant vers de meilleurs solutions, mais plutôt une population entière d'individus échangeant de l'information et s'adaptant à l'environnement au fur et à mesure. Cette notion de population est intéressante car elle est responsable d'un parallélisme implicite de l'algorithme génétique, ce qui le rend encore plus rapide.

Les algorithmes de recuit-simulé, comme les algorithmes génétiques restent encore un peu marginaux dans le domaine du recalage en imagerie médicale. On trouve quand même l'article de FITZPATRICK [32] (traitant d'ailleurs également d'algorithmes génétiques), et l'article, plus récent, de Nikou [69] sur l'application du recuit-simulé au recalage sous-voxel en imagerie médicale. Sa méthode utilise la définition de M-estimateurs robustes comme fonctionnelle, qui en fait revient à minimiser une variance inter-régions de niveaux de gris (voir la fin du paragraphe 1.4.2.1, page 15).

1.5.3 Algorithmes hybrides

Bien que nous ayons présenté des méthodes d'optimisation différentes, il est à noter que la plupart des traitements en imagerie ne se contente pas d'utiliser un seul algorithme d'optimisation. Beaucoup de méthodes consistent à diviser les problèmes en une série plus ou moins indépendante de sous problèmes dont chacun peut se résoudre avec des méthodes différentes.

Notons par exemple la possibilité de mélanger des méthodes de recherche globale comme le recuit simulé, les algorithmes génétiques, voire la méthode des moments avec des méthodes de recherche plus locale comme le *Hill Climbing*. Ceci présente l'avantage d'améliorer le temps de convergence de l'algorithme, tout en conservant une bonne qualité au résultat.

1.6 Validation des résultats

Une fois que l'on dispose d'une transformation 3D optimale vis-à-vis d'un certain critère de recalage, il nous reste encore à estimer la validité, ou la qualité extrinsèque, d'une telle transformation.

Cette section est consacrée à la présentation de quelques méthodes de validation des résultats d'un algorithme de recalage. Certaines méthodes permettent d'obtenir un résultat chiffré, d'autre sont basées sur une estimation visuelle du recalage.

Étant donné la grande diversité des algorithmes de recalage, il est souvent difficile de les comparer quantitativement. Certains sont rapides, d'autres extrêmement lents. Certains effectuent des recalages rigides d'autres des recalages élastiques. Les rares cas où l'on peut comparer à juste titre quantitativement deux algorithmes de recalage différents se limitent à l'usage d'une même fonction de coût. Mis à part ce cas précis, toutes les autres comparaisons sont souvent biaisées.

La robustesse des algorithmes est également difficile à comparer d'une méthode à une autre car elle est issue d'une prise de décision (résultat correct ou résultat incorrect) intrinsèque à chaque algorithme de recalage.

Un projet de validation de différentes méthodes (cf. chapitre 5) a été mis en place pour évaluer rétrospectivement la précision de méthodes de recalage (*i.e.* directement à partir des images brutes) en se référant à un recalage rigide considéré optimal et obtenu à l'aide de marqueurs préalablement fixés sur le patient avec un casque stéréotaxique (méthode prospective).

Enfin notons bien sûr que le meilleur moyen de comparer deux techniques de recalage est de faire estimer visuellement par un expert les résultats obtenus.

La suite de cette section sera divisée en deux parties : une première présentant quelques méthodes de validations visuelles, puis une deuxième partie traitant de la validation à l'aide de marqueurs.

1.6.1 Validation visuelle

1.6.1.1 Visualisation de coupes 2D

On affiche non pas le volume entier mais une série de vues en coupe 2D du volume.

la figure 1.18 représente un exemple de vue en coupe d'une image scanner de référence et d'une image IRM recalée sur la référence. Les niveaux de gris de ces images ont été normalisés pour utiliser la pleine échelle.



Figure 1.18 — Vue en coupe d'une image scanner et d'une image IRM.

Mélange des images. La valeur d'un pixel d'une coupe est dépendante de la valeur du pixel de même coordonnées sur les deux images à confronter. Soient A et B les deux images de même dimensions à mixer pour obtenir C. Pour chaque pixel (x, y) de C, C(x, y) = f(A(x, y), B(x, y)).

Différents modes d'affichages (ou fonction f(u, v)) sont envisageables :

1. Différence en valeur absolue

$$f_1(u,v) = |u-v|$$

Cette fonction est surtout utile pour les images de mêmes modalité ayant une intensité moyenne comparable. Ainsi un même niveau de gris dans les deux images représente la même réalité physique.

La figure 1.19 présente mode d'affichage avant et après recalage.



Figure 1.19 — $f_1(\text{Scanner}, \text{IRM})$ et $f_1(\text{Scanner}, \text{IRM recalée})$.

2. Différence centrée

$$f_2(u,v) = \frac{\operatorname{nv_gris_max} + u - v}{2}$$

 f_2 est une variante de f_1 permettant de savoir en plus, si le fond de l'image est de niveaux de gris constant, quel est précisément le volume qui excède l'autre.

De même, la figure 1.20 est une illustration de ce mode d'affichage avant et après recalage.



Figure 1.20 — $f_2(\text{Scanner}, \text{IRM})$ et $f_2(\text{Scanner}, \text{IRM recalée})$.

3. ET logique

$$f_3(u,v) = u \lor v$$

Ce mode, surtout utile pour comparer deux images binaires, permet d'afficher seulement l'intersection des images A et B.



Figure 1.21 — $f_3(\text{Scanner}, \text{IRM})$ et $f_3(\text{Scanner}, \text{IRM recalée})$.

4. OU logique

$$f_4(u,v) = u \wedge v$$

Ce mode est similaire au mode f_3 ; f_4 permet d'obtenir l'union logique des images A et B.

5. OU Exclusif

$$f_5(u,v) = u \oplus v$$

Il correspond à l'union de A et B à laquelle on enlève l'intersection de A et B. Le OU Exclusif permet de n'afficher que les zones où les deux images différent.



Figure 1.22 — f_4 (Scanner, IRM) et f_4 (Scanner, IRM recalée).



Figure 1.23 — f_5 (Scanner, IRM) et f_5 (Scanner, IRM recalée).

6. Affichage en damier

$$f_6(A(x,y), B(x,y)) = \begin{cases} A(x,y) & \text{si } \left[\frac{x}{\delta x}\right] + \left[\frac{y}{\delta y}\right] \text{ est pair,} \\ B(x,y) & \text{sinon.} \end{cases}$$

L'image résultante est similaire à un damier où les cases blanches contiendrait les pixels de l'image A et les cases noires les pixels de l'image B. La notation [.] signifie partie entière, et les quantités δx et δy mesure la taille des cases du damier.

Un tel procédé d'affichage permet d'une manière statique (sur une seule image de la série de coupes) d'estimer la continuité entre les structures de l'image A et celles de l'image B au niveau des frontières des cases.

Notons que si nous affichons rapidement des coupes en damier en inversant une image sur 2 le rôle des cases noires et le rôle des cases blanches, alors la persistance rétinienne permet dynamiquement de superposer le contenu de l'image A avec celui de l'image B.

La figure 1.24 illustre un exemple d'affichage en mode damier.

7. Surimpression de structures

$$f_7(A(x,y), B(x,y)) = \begin{cases} \operatorname{Vert}(A(x,y)) & \text{si } A(x,y) \in [a,b], \\ B(x,y) & \text{sinon.} \end{cases}$$

On détermine préalablement sur une image (A(x, y)) un intervalle de niveaux de gris correspondant à une structure particulière (segmentation par niveaux de gris), comme par exemple les niveaux de gris correspondant aux structures osseuses de l'image scanner.



Figure 1.24 — f_6 (Scanner, IRM) et f_6 (Scanner, IRM recalée).

On affiche ces niveaux de gris en niveaux de vert en surimpression avec les niveaux de gris de l'image B(x, y). Si les images sont parfaitement recalées, les structures des deux images seront corrélées.

La figure 1.25 montre un exemple d'affichage en surimpression d'une structure osseuse (pixels verts) extraite d'une image TDM, avec les niveaux de gris de l'image IRM correspondante, avant et après recalage.





Figure 1.25 — f_7 (Scanner, IRM) et f_7 (Scanner, IRM recalée).

1.6.1.2 Visualisation du volume entier

Une autre manière de valider visuellement le résultat d'un recalage est d'afficher, par une méthode de rendu multi-volumique direct [47], une série de projections 2D d'une combinaison des volumes 3D.

Cette méthode est spécialement étudiée en vue de la validation visuelle d'un recalage 3D. Elle repose sur une modélisation physique des milieux permettant de formuler des règles de composition des modalités en présence. Les informations structurelles peuvent ainsi être visuellement combinées par le jeu d'opérateurs logiques portant sur des surfaces floues (voir description plus détaillée en annexe C).

La figure 1.26 présente quatre rendus, deux avant recalage, deux après recalage, en mode union puis intersection des données, obtenus par rendu volumique direct [47]. Notons également que pour mieux visualiser l'ensemble du volume nous pouvons créer de petites séquences animées représentant le volume vu sous différents angles.



Figure 1.26 — Rendus effectués avec l'outil dvr.

1.6.2 Validation à l'aide de marqueurs

Une première façon d'obtenir une estimation chiffrée est d'utiliser deux ensembles de marqueurs anatomiques se correspondant mutuellement d'une modalité à une autre. Ces marqueurs ne doivent pas être utilisés par l'algorithme de recalage mais servent juste d'estimation, en terme de distance euclidienne entre deux ensembles de points, de la qualité du recalage.

Le calcul de la distance se fait comme au paragraphe 1.4.2.2.

Un exemple d'utilisation de cette méthode de quantification du résultat final est proposé par FITZPATRICK et MAURER dans l'article [33]. Par ailleurs, cette méthode est utilisée à des fins d'estimation rétrospective de la précision des méthodes de recalage dans le projet *RREP* (voir chapitre 5).

1.6.3 Test de robustesse statistique

Un point important dans la validation d'une méthode de recalage est non seulement de montrer qu'ils fonctionnent dans des cas idéaux mais aussi dans les cas plus difficile. De plus, pour les algorithmes stochastiques, il faut s'assurer que les résultats sont reproductibles (ceci permet de donner un degré de confiance dans l'algorithme). Le seul moyen de tester qu'un algorithme stochastique est robuste est d'effectuer plusieurs recalages et comparer les résultats.

1.7 Synthèse du chapitre

À l'issue de ce premier chapitre, nous avons abordé à titre introductif et de manière générale les divers aspects techniques d'un problème de recalage en imagerie médicale. Nous profitons de ce tour d'horizon pour situer notre démarche par rapport aux travaux que l'on trouve généralement dans la littérature. Le problème que nous nous étions fixés initialement était tout simplement d'effectuer un recalage élastique multimodalité relativement rapide entre des images volumiques TDM et IRM du crâne humain.

La résolution d'un tel problème s'est concrétisée par les réponses aux questions suivantes :

- Quel type de prétraitement ? Une segmentation des images pour l'extraction de caractéristiques communes afin de transformer le problème du recalage multimodal en un recalage monomodal.
- Quel type de fonctionnelle à optimiser? Une distance euclidienne stochastique entre deux surfaces numériques.
- Quelle type de transformation ? Une transformation globale élastique trilinéaire. Bien qu'il n'y ait pas de raison *a priori* que ce modèle corresponde à une réalité physique, sa généralité nous permet de penser qu'elle constitue une approximation suffisante du modèle réel de déformation. D'autre part, parmi toutes les modélisation élastiques, c'est une des plus simples à mettre en œuvre (contrairement aux modélisation fluides impliquant des méthodes d'éléments finis).
- Quel algorithme d'optimisation? Étant donné la taille de l'espace de recherche (nombre de paramètres), et les impératifs d'efficacité, nous nous sommes rapidement tournés vers les algorithmes génétiques qui, moyennant un couplage avec un algorithme d'optimisation locale in fine, s'avère être un outil très robuste et rapide.
- **Comment valider les résultats** Nous avons opté pour une validation visuelle, à la fois 2D (coupes à coupes) et 3D (rendu volumique) car nous pensons que c'est la seule mesure objective, et aussi parce que nous disposions justement d'un outil de rendu multivolumique direct permettant des modes d'affichages très discriminants (comme le mode intersection).

Maintenant que nous nous sommes positionnés dans le cadre très général des algorithmes de recalage, nous allons détailler les différents points importants de notre algorithme, à savoir : quelle modélisation nous avons utilisée (chapitre 2), comment fonctionne un algorithme génétique et quelle implémentation nous avons réalisée (chapitre 3), quelles sont les différentes étapes de l'algorithme qui nous ont permis de réaliser une méthode robuste et rapide (chapitre 4). Enfin nous verrons, au chapitre 5, l'application de notre algorithme de recalage sur une base de données partagée d'images médicales ayant servi à la validation et à la comparaison de diverses méthodes de recalage.

CHAPITRE 2 Recalage élastique 3D : Modélisation

Ce chapitre est consacré à la modélisation du problème de recalage que l'on se propose de résoudre. Nous allons en premier lieu détailler les modèles de transformations 3D génériques rigides et élastiques, avant de présenter plus précisément le choix du modèle élastique retenu pour le recalage d'images cérébrales (images servant d'application au chapitre 4). Les concepts mathématiques seront d'abord présentés puis des exemples de déformations 3D seront mis en œuvre.

Par ailleurs, nous modéliserons et définirons dans ce chapitre une distance géométrique (ou fonction de coût) caractérisant la qualité de la mise en correspondance de deux images 3D (vis-à-vis d'une certaine transformation 3D).

Enfin, ce chapitre, comme les suivants, comportera une synthèse des points essentiels abordés et à retenir.

2.1 Modèles de transformation géométrique 3D

Les images médicales que nous voulons étudier, et qui servent de support à cette thèse sont des images cérébrales acquises avec deux types d'imageurs 3D différents : scanner-X ou Tomo-DensitoMétrie (TDM) d'une part, et Imagerie par Résonance Magnétique (IRM) de l'autre (pour plus d'informations sur l'acquisition de ces images, voir l'annexe A). Les problèmes de recalage entre images de modalités différentes se situent à deux niveaux. Un niveau positionnel qui consiste en la mise en correspondance globale des objets observés par une transformation rigide (rotation et translation, voire facteur d'échelle), et un niveau d'affinement qui a pour but de corriger les non-linéarités (souvent locales) d'un imageur à un autre (transformation élastique qui est une extension de la transformation rigide).



Figure 2.1 — Principe général de compensation des erreurs d'acquisition par transformation rigide et élastique.

La figure 2.1 illustre, en 2D, le principe général de la transformation rigide et élastique en imagerie médicale. La transformation rigide (T^R) permet de recentrer et d'aligner les objets imagés les uns par rapport aux autres; la transformation élastique (T^E) , quant à elle, compense les non-linéarités induites d'une part par les appareils imageurs, et d'autre part par l'élasticité éventuelle des objets observés.

2.1.1 Transformation 3D rigide et globale

La transformation 3D globale la plus simple à envisager est la transformation rigide. Cette transformation est a priori appropriée au recalage d'images cérébrales (le crâne étant considéré comme un objet 3D rigide) monomodalité et monopatient¹.

Une transformation rigide 3D est une isométrie de \mathbb{R}^3 conservant l'orientation (transformation directe par opposition aux symétries qui sont des isométries indirectes). Une transformation rigide est donc tout simplement la composée d'une rotation et d'une translation. Dans la littérature du recalage, certains auteurs considèrent que les transformations composées d'une rotation, d'une translation et d'un ou plusieurs facteurs de mise à l'échelle (donc isotrope ou anisotrope) sont aussi des transformations rigides. Cependant si l'on s'appuie sur la définition exacte d'une transformation rigide ce n'est pas vrai.

Pour les transformations globales, on considère comme rigides, toutes les isométries du plan ou de l'espace qui conservent l'orientation. Pour les transformations locales, on considère comme rigides, toutes les compositions de transformations rigides au sens de la rigidité globale.

Si f est une transformation rigide, alors f est une fonction de \mathbb{R}^3 dans \mathbb{R}^3 , qui à tout point (x, y, z) d'une image associe le point transformé (x', y', z'), et est définie par

$$f(x, y, z)^{t} = (x', y', z')^{t} = \mathbf{T} + \mathbf{R}(x, y, z)^{t},$$
(2.1)

où \mathbf{T} est un vecteur de translation et \mathbf{R} est une matrice (3×3) de rotation, donc orthogonale et de déterminant unitaire (*i.e.* orthonormale) pour satisfaire la conservation de l'orientation (pas de facteur d'échelle non plus).

Une telle transformation présente l'avantage d'être globale. Non seulement, c'est une transformation continue certes à support discret puisque l'image d'une courbe est une courbe, mais en plus cette transformation conserve les angles et les formes (l'image d'un tétraèdre est un tétraèdre de même nature). Des exemples d'application de cette transformation seront présentés au paragraphe 2.2.

On peut aussi définir la matrice de rotation par le vecteur de quaternions $\vec{\mathbf{Q}} = (q_0 q_1 q_2 q_3)^t$ (4 paramètres mais seulement trois sont indépendants car on impose la relation $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$).

La relation entre $\vec{\mathbf{Q}}$ et \mathbf{R} est la suivante :

$$\mathbf{R} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix}.$$

Cette transformation rigide peut donc être utilisée uniquement pour recaler deux images ne présentant pas d'évolution de structures (croissance du crâne, ou modification de la position ou du volume d'une sous-structure) et dont les imageurs n'ont pas induit de distorsions. Néanmoins, on peut se servir d'une telle transformation comme *bonne approximation d'une transformation faiblement élastique*.

¹À condition qu'il n'y ait pas d'évolution sensible des structures, ce qui pourrait être le cas sur des images cérébrales de jeunes patients en cours de croissance.

2.1.2 Transformations 3D élastiques

Par défaut on considère comme élastiques toutes les transformations qui ne sont pas purement rigides. Cependant il est quand même juste de considérer les transformations rigides comme des cas particuliers de transformations élastiques.

Les transformations élastiques que nous considérons ici sont les transformations affines, les transformations polynomiales, les transformations basées sur une interpolation de type spline et les transformations fluides.

2.1.2.1 Transformations affines

Pour simplifier les notations, nous ne parlons ici que des transformations de \mathbb{R}^2 dans \mathbb{R}^2 .

Soient (u, v) les coordonnées d'un point dans une image de référence, on note alors (x, y) les coordonnées de son homologue dans l'autre image.

Une transformation affine 2D est définie par six coefficients a_{ij} . La relation entre les coordonnées dans une image et les coordonnées dans l'autre image est donnée par le système d'équations :

$$\begin{cases} x = a_{11}u + a_{21}v + a_{31} \\ y = a_{12}u + a_{22}v + a_{32} \end{cases} \iff (x, y, 1)^t = \underbrace{\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ 0 & 0 & 1 \end{bmatrix}}_{A} (u, v, 1)^t.$$

La première propriété des transformations affines est que celles-ci conservent l'alignement et le parallélisme ; autrement dit, deux droites parallèles le restent à l'issue de la transformation.

On peut décomposer la matrice de transformation affine A en la matrice $B = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$ et le vecteur $T = (a_{31}, a_{32})^t$ qui est un vecteur de translation.

- Si B = Id (matrice identité), la transformation se réduit évidemment à une translation. Dans le cas particulier où B est une matrice de rotation, nous retrouvons une transformation rigide.
- Si $B = \text{Diag}(S_u, S_v)$, la transformation est une mise à l'échelle (avec les facteurs S_u selon u et S_v selon v.
- Enfin le dernier cas particulier est celui de la matrice de distorsion ou $T = \vec{0}$ et $B = \begin{pmatrix} 1 & Hv \\ 0 & 1 \end{pmatrix}$ ou $B = \begin{pmatrix} 1 & 0 \\ Hu & 1 \end{pmatrix}$. Cette transformation est également connue sous le nom anglais de *shearing*.

Notons que la composition de deux ou plusieurs transformations affines est une transformation affine, et qu'inversement, toute transformation affine est la composée de transformations affines élémentaires.

Une transformation affine 2D possède six degrés de liberté; elle est donc parfaitement définie par la donnée de trois points dans une image et ses trois homologues dans l'autre image. Pour une transformation affine 3D, il y a douze degrés de liberté donc il faut connaître quatre points et ses transformés pour définir pleinement la transformation.

L'extension la plus directe des transformations affines sont les transformations inférant

une perspective (ou une projection). La matrice de transformation est la suivante :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} . \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

avec

$$\begin{cases} x = \frac{x'}{w'} = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \\ y = \frac{y'}{w'} = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \end{cases}$$

Pour les transformations projectives, l'alignement n'est conservé que pour les horizontales et les verticales, donc l'image d'une droite n'est une droite que si elle est verticale ou horizontale. Dans un cadre général, l'image d'une droite est une conique. Notons enfin que la transformation projective conserve tout de même les rapports de longueurs sur une courbe.

2.1.2.2 Transformations bilinéaires et trilinéaires

La transformation bilinéaire est une extension de l'interpolation bilinéaire qui consiste à estimer la valeur d'un point en fonction d'une combinaison linéaire de ses quatre plus proches voisins.

Étant donnés quatre points (u_0, v_0) , (u_1, v_1) , (u_2, v_2) et (u_3, v_3) ainsi que leur valeurs correspondantes x_0, x_1, x_2 , et x_3 , n'importe quelle valeur intermédiaire X(u, v) peut se calculer comme suit :

$$X(u, v) = a_0 + a_1 u + a_2 v + a_3 u v,$$

où les coefficients a_i sont obtenus en résolvant

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & u_0 & v_0 & u_0 v_0 \\ 1 & u_1 & v_1 & u_1 v_1 \\ 1 & u_2 & v_2 & u_2 v_2 \\ 1 & u_3 & v_3 & u_3 v_3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

L'avantage de la déformation bilinéaire est d'être une transformation séparable, dans le sens où elle est le produit de 2 déformations (dans le cas 2D) unidimensionnelles.

L'analogie 3D de la transformation bilinéaire est la transformation trilinéaire. Dans un cadre plus général, ces transformations sont des transformations polynomiales (voir paragraphe suivant), décrites par l'équation (2.2) où l'image d'un point p de \mathbb{R}^3 est un point p' de coordonnées :

$$p = (x, y, z)^{T} \quad p' = (x', y', z')^{T} = T(p)$$

$$\begin{cases}
x' = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{l} a_{i,j,k} x^{i} y^{j} z^{k} \\
y' = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{l} b_{i,j,k} {}^{i} y^{j} z^{k} \quad \text{avec } i+j+k \leqslant J. \\
z' = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{l} c_{i,j,k} x^{i} y^{j} z^{k}
\end{cases}$$
(2.2)

Nous restreignons cette transformation à une transformation polynomiale trilinéaire (*i.e.* J = 3, n = m = l = 1) en espérant modéliser de manière suffisante les faibles distorsions induites par les appareils imageurs. La dimension de l'espace des transformations trilinéaires (qui sera l'espace de recherche S de notre algorithme de recalage) est par conséquent $L = \dim S = 3 \times 2^3 = 24$.

Les 24 coefficients $a_{i,j,k}$, $b_{i,j,k}$, et $c_{i,j,k}$ sont les paramètres de la transformation. Notons que le vecteur $(a_{0,0,0}, b_{0,0,0}, c_{0,0,0})$ correspond à un vecteur de translation; les 9 coefficients $(a_{1,0,0}, a_{0,1,0}, a_{0,0,1})$ et équivalents pour b et c définissent une transformation affine, et dans le cas particulier d'une transformation rigide, ces 9 coefficients sont ceux d'une matrice de rotation et les 12 derniers coefficients sont tous nuls.

La transformation trilinéaire 3D est la transformation que nous avons retenu pour modéliser la géométrie du recalage des images cérébrales multimodales car elle peut être faiblement élastique, et aussi car elle est simple à mettre en œuvre. L'utilisation de cette transformation sous la forme de la relation (2.2) étant délicate nous proposons dans un paragraphe ultérieur (2.1.4) une approche utilisant la notion d'amers.

2.1.2.3 Transformations polynomiales

En conservant les mêmes notations que dans le paragraphe précédent, une transformation polynomiale 2D de degré N est définie par :

$$\begin{cases} x = \sum_{i=0}^{N} \sum_{j=0}^{N-i} a_{ij} \ u^{i} v^{j} \\ y = \sum_{i=0}^{N} \sum_{j=0}^{N-i} b_{ij} \ u^{i} v^{j} \end{cases}$$

Notons que pour N = 1 nous retrouvons le cas des transformations affines.

Ces transformations servent à modéliser les défauts élastiques globaux, et non locaux. En effet, surtout si l'ordre N choisi est faible, la transformation ne peut modéliser les changements seulement locaux (ce qui correspond à des hautes fréquences dans le spectre de l'image différence).

La composée de deux transformations polynomiales n'est en général pas une transformation polynomiale (ou du moins pas du même ordre). Pour une transformation polynomiale 2D d'ordre N, il faut au minimum K = (N+1)(N+2)/2 couples de points pour définir de manière unique les coefficients a_{ij} et b_{ij} .

2.1.2.4 Les plaques minces, splines généralisées

Théorie

L'idée de base de cette théorie sur l'interpolation des fonctions de deux variables est due à DUCHON [28] en 1976, et la formalisation a été mise en place par MEINGUET [64, 65] en 1979.

Une manière de considérer une transformation entre deux images 2D est de définir deux surfaces $2D^{1/2}$ (graphe d'une fonction) que nous appellerons X(u, v) et Y(u, v) dont la hauteur au point de coordonnées (u, v) nous donne les coordonnées transformées (x, y) d'un point de coordonnées (u, v).

Le problème réside dans la définition de la forme des deux surfaces. Si nous connaissons les correspondances de N points entre nos deux images, nous connaissons alors N points P_i de coordonnées (u_i, v_i) sur l'image de référence correspondant aux N points Q_i de coordonnées (x_i, y_i) sur l'image à recaler. Si nous imposons pour tout $i \in 1 \dots N$

$$\begin{cases} X(u_i, v_i) &= x_i, \\ Y(u_i, v_i) &= y_i \end{cases}$$

alors les surfaces X() et Y() sont des interpolants de \mathbb{R}^2 dans \mathbb{R}^2 . Par analogie avec la théorie des plaques minces, ces surfaces peuvent être restreintes en imposant la propriété de torsion minimale, ce qui revient à minimiser l'expression :

$$E(X) = \iint \left(\left(\frac{\partial^2 X}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 X}{\partial u \partial v} \right)^2 + \left(\frac{\partial^2 X}{\partial v^2} \right)^2 \right) du dv$$

On montre que X() et Y() sont également solutions de l'équation biharmonique $\Delta^2 X = 0$, dont une solution particulière est la fonction $U(r) = -r^2 \log r^2$ (où $r^2 = u^2 + v^2$). Par conséquent les surfaces d'interpolation correspondant à l'appariement des P_i aux Q_i s'expriment comme une combinaison linéaire des solutions élémentaires U:

$$\begin{cases} X(u,v) &= \sum_{i=0}^{N} x_i U\left(|(u,v) - P_i| \right), \\ Y(u,v) &= \sum_{i=0}^{N} y_i U\left(|(u,v) - P_i| \right) \end{cases}$$

Application au recalage en imagerie médicale

L'article de F.L. BOOKSTEIN [9] présente en 1989 une application de la théorie des plaques minces pour le recalage d'images binaires simples 2D, ainsi qu'une deuxième application pour la détection d'anomalies de développement de la mâchoire de l'homme. Un deuxième article du même auteur [10] écrit en 1991 utilise cette même méthode pour le recalage d'images médicales sur un atlas dans un but d'étude morphométrique.

Citons enfin l'article de K. ROHR [79], exposant également une application de la théorie des plaques minces au recalage d'images cérébrales.

2.1.2.5 Modélisation fluide

Finalement, notons également que l'on trouve dans la littérature des algorithmes de recalage considérant les images comme des fluides qui se déforment en s'écoulant [14]. D'autres approches encore, considèrent des champs locaux de déformations répondant à des critères d'équilibre de forces extérieures et de propriétés d'élasticité [71]. De tels modèles tentent d'approcher, sans toutefois y arriver complètement, les propriétés physiques de déformation des objets. L'inconvénient majeur est que le traitement sur des images 2D est déjà coûteux, et est sujet à l'instabilité numérique pour des méthodes telles que les éléments finis. L'extension au cas 3D accentue encore ces problèmes.

2.1.3 Choix du modèle de transformation élastique

Le but que nous nous étions fixé étant de réaliser un recalage élastique, il est bien entendu nécessaire de choisir un modèle de déformation approprié. Sur les images cérébrales que nous traitons, principalement deux types de déformations élastiques peuvent entrer en compte :

- Le premier type est la déformation inhérente à la non-linéarité des appareils imageurs et notamment de l'IRM. Cette élasticité est considérée comme globale et faible car en général elle est corrigée en grande partie par l'appareil imageur lui-même.
- La deuxième source de déformation est due aux variations des données elles-mêmes entre les différentes acquisitions. Ces changements sont d'une part de nature positionnelle (position et orientation du patient, mouvements articulaires faibles) et d'autre part de nature évolutive (évolution d'une tumeur par exemple).

Lorsque nous considérons un recalage élastique, nous voulons en premier lieu corriger les changements positionnels (position et orientation de l'objet visualisé) et, en même temps, corriger les faibles distorsions induites par les appareils imageurs. Notons bien que nous ne corrigeons pas, grâce au caractère non-local de la transformation de recalage, les changements évolutionnels comme l'apparition ou le développement de tumeurs².

Nous avons choisi comme modèle de transformation élastique la transformation trilinéaire présentant les avantages d'être d'une part globale et d'autre part simple (*i.e.* avec relativement peu de paramètres). Le fait que la transformation soit simple est un gage de performance lorsque l'on utilise des algorithmes génétiques car un grand nombre de transformations doit être testé tout au long du processus d'optimisation afin d'évaluer en permanence la qualité des solutions intermédiaires (cf. chapitre 3). Par ailleurs, une transformation globale permet de modéliser les effets macroscopiques de non-linéarité des appareils imageurs. Entendons bien que la transformation trilinéaire ne correspond pas à un modèle physique des déformations réellement inhérentes aux appareils imageurs, mais néanmoins, cette transformation reste assez générale pour corriger avec une bonne précision ces déformations, puisque nous l'avons utilisée avec succès en conjonction avec divers algorithmes d'optimisation portant sur les recalages monomodalité [46] et multimodalité [81, 82].

2.1.4 Description de la transformation par amers

Avec l'équation directe (2.2), il est n'est pas aisé de poser des frontières à l'espace de recherche. Néanmoins, la connaissance de 8 points de l'image F_1 correspondants à 8 points de l'image F_2 permet de donner une description indirecte du problème. Soit \mathbf{X}_1 (respectivement \mathbf{Y}_1 et \mathbf{Z}_1) le vecteur représentant les coordonnées en x (respectivement y et z) des 8 points

 $^{^{2}}$ Si nous corrigions les évolutions des tumeurs, nous ne pourrions plus les observer !

de $F_1(x_1(i), y_1(i), z_1(i)), i = 1, ..., 8$. De même, soit \mathbf{X}_2 (respectivement \mathbf{Y}_2 et \mathbf{Z}_2) le vecteur représentant les coordonnées en x (respectivement y et z) des 8 points de $F_2(x_2(i), y_2(i), z_2(i)),$ i = 1, ..., 8 correspondants aux 8 points de F_1 précédemment définis. Notons également \mathbf{M} la matrice 8×8 de transformation déduite de \mathbf{X}_1 , \mathbf{Y}_1 et \mathbf{Z}_1 et de la relation (2.2). La *i*-ième ligne de $\mathbf{M}(X_1, Y_1, Z_1)$ s'écrit :

$$[1, x_1(i), y_1(i), z_1(i), x_1(i)y_1(i), x_1(i)z_1(i), y_1(i)z_1(i), x_1(i)y_1(i)z_1(i)].$$

On obtient alors la relation reliant les amers de F_1 et de F_2 suivante :

$$\begin{cases} \mathbf{X}_{2} = \mathbf{M}(\mathbf{X}_{1}, \mathbf{Y}_{1}, \mathbf{Z}_{1}) \cdot \mathbf{A} \\ \mathbf{Y}_{2} = \mathbf{M}(\mathbf{X}_{1}, \mathbf{Y}_{1}, \mathbf{Z}_{1}) \cdot \mathbf{B} \\ \mathbf{Z}_{2} = \mathbf{M}(\mathbf{X}_{1}, \mathbf{Y}_{1}, \mathbf{Z}_{1}) \cdot \mathbf{C} \end{cases}$$
avec
$$\begin{cases} \mathbf{A} = [a_{0,0,0}, a_{1,0,0}, a_{0,1,0}, a_{0,0,1}, a_{1,1,0}, a_{1,0,1}, a_{0,1,1}, a_{1,1,1}]^{t} \\ \mathbf{B} = [b_{0,0,0}, b_{1,0,0}, b_{0,1,0}, b_{0,0,1}, b_{1,1,0}, b_{1,0,1}, b_{0,1,1}, b_{1,1,1}]^{t} \\ \mathbf{C} = [c_{0,0,0}, c_{1,0,0}, c_{0,1,0}, c_{0,0,1}, c_{1,1,0}, c_{1,0,1}, c_{0,1,1}, c_{1,1,1}]^{t} \end{cases}$$

$$(2.3)$$

Le calcul des coefficients polynômiaux \mathbf{A} , \mathbf{B} et \mathbf{C} est réalisé par une simple inversion du système linéaire précédent. Il est important de noter que pour une série de 8 couples de points (ou 8 correspondances entre F_1 et F_2), il existe une unique transformation trilinéaire T mettant en correspondance ces points. Par ailleurs, la transformation ne peut exister, ou plutôt le système linéaire n'est inversible, que si les 8 amers de F_1 sont distincts, et de même pour les amers de F_2 .

Avec cette nouvelle description de la transformation trilinéaire, deux types d'algorithmes sont envisageables :

- Un algorithme cherchant, à partir de 8 points fixes sur une image (par exemple les sommets du volume d'intérêt), 8 correspondances sur l'autre image dans 8 voisinages prédéfinis. L'espace de recherche est alors un hypercube de dimension 24 (cf. figure 2.2 montrant l'équivalent de la méthode en 2D).
- Une méthode qui consiste à faire du *point-matching*, c'est-à-dire identifier des points caractéristiques sur les deux images, et chercher quelles sont les associations possibles entre ces points. Une fois ces associations trouvées, il est facile d'interpoler la transformation sur la totalité du champ de l'image de référence en en déduisant les 24 coefficients de la transformation trilinéaire globale.

En anticipant sur la discussion du chapitre 4 page 81, notons que, même si 8 paires de points est le nombre exact nécessaire au calcul des 24 paramètres de manière déterministe, la prise en compte de paires supplémentaires demeure possible. Le système est alors résolu au sens des moindres carrés, ce qui donne l'avantage d'une meilleure insensibilisation au bruit de positionnement des points.

2.1.5 Application de la transformation par interpolation

Prenons le cas de deux images $F_1(p), p \in D_1 \subset \mathbb{R}^3$ et $F_2(p), p \in D_2 \subset \mathbb{R}^3$. Recaler F_2 sur F_1 revient a trouver une transformation géométrique T^* mettant en correspondance F_2^* avec F_1 . T^* est alors définie comme une application de D_1 vers \mathbb{R}^3 et F_2^* est déterminée pour $p \in D_1$ par :



Figure 2.2 — Illustration 2D du recalage par amers fixes et espace de recherche contraint (hypercube).

$$F_2^{\star}(p) = \begin{cases} F_2(T(p)) & \text{si } T(p) \in D_2, \\ 0 & \text{sinon.} \end{cases}$$
(2.4)

 T^{\star} maximise la similarité entre $F_1(p)$ et $F_2^{\star}(p)$. Étant donné que nous travaillons avec des signaux discrets (images numériques), et puisque la forme des transformations géométriques, qu'elles soient rigides ou élastiques, implique une transformation continue, la création de F_2^{\star} à partir de F_2 et de T implique une interpolation. En effet, il n'y a aucune raison particulière pour que $\forall p \in D_1, T(p)$ appartienne à la grille d'échantillonnage de F_2 . Même si le mot transformation au sens mathématique suppose que l'espace de départ est l'image à transformer, et l'espace d'arrivée est l'image transformée, on préfère utiliser une transformée inverse (la transformée T que nous avons introduite au début de ce paragraphe est orientée du domaine de référence D_1 vers le domaine de l'image à recaler D_2 , et pourtant est utilisée de manière à reconstruire le contenu de l'image F_2 sur le domaine D_1 ; si nous partions de chaque point de l'image à transformer, et en appliquant une fonction orientée naturellement (au sens mathématique) de D_2 vers D_1 , nous ne serions pas sûrs de remplir suffisamment l'image d'arrivée (il peut y avoir des trous de reconstruction). En prenant la transformée T orientée comme à l'équation (2.4) de D_1 vers D_2 , on parcourt l'image destination point à point et on regarde quelles sont les images par T du point courant. Puisque ces points sur l'image de F_2 sont calculés, ils peuvent tomber à cheval entre plusieurs voxels, et une interpolation linéaire est nécessaire pour éviter des phénomènes d'aliasing, mais au détriment d'un effet de lissage.

2.2 Application des modèles de transformation 3D

Maintenant que nous avons présenté les modèles de déformation rigide et élastique que nous utiliserons, nous nous proposons de visualiser leur effets sur des images 2D puis sur des images 3D.

2.2.1 Déformation 2D

La déformation 2D déduite de la transformation trilinéaire de l'équation (2.2) est résumée par la transformation bilinéaire suivante :

$$F_{2}^{\star}(p = (x, y)) = F_{2}(p' = T(p) = (x', y')) \quad \text{avec}$$

$$\begin{cases} x' = t_{x} + a_{x}x + a_{y}y + a_{xy}xy \\ y' = t_{y} + b_{x}x + b_{y}y + b_{xy}xy \end{cases}$$

Les points de coordonnées p de l'image transformée F_2^{\star} sont déduits de la valeur des points de coordonnées p' = T(p) sur l'image originale F_2 . La figure 2.3 montre un exemple d'application de cette transformée élastique 2D sur une image de manchot (l'image originale F_2 se situe en haut à gauche, 2.3(a)). Les différentes transformées F_2^{\star} sont calculées en fonction de la valeur des paramètres $(t_x, t_y, a_x, a_y, ..., b_{xy})$. Nous notons que lorsque a_{xy} et b_{xy} sont nuls, et lorsque la matrice $[[a_x, a_y][b_x, b_y]]$ est orthonormale et à déterminant unitaire, nous sommes en présence d'une transformation rigide. La transformation rigide est donc bien un cas particulier de la transformation élastique polynomiale envisagée.

Remarquons également que les transformations données en exemple à la figure 2.3 ont été calculées avec un système de coordonnées centré (ce qui n'est pas habituel en traitement d'image).

Les différentes transformations que nous observons dans cet exemple sont d'une part des transformations rigides (cas particulier de transformations élastiques) avec dans l'ordre l'identité (image 2.3(a)), la translation (image 2.3(b)), la composition d'une translation et d'une rotation (image 2.3(c)). Ensuite nous avons des transformation élastiques linéaires (ou perspectives) dans lesquelles l'image d'une droite reste une droite, mais où les distances et les angles ne sont pas conservés (images 2.3(d) et 2.3(e)). Enfin, l'image 2.3(f) montre un exemple de transformation élastique bilinéaire.

2.2.2 Déformation 3D

Nous allons maintenant appliquer différentes transformations polynomiales trilinéaires sur un objet 3D (une superquadrique en forme de savonnette³). Les rendus qui vont suivre sont des rendus volumiques directs basés sur une notion de lancer de rayons [47, 49] (Nous appellerons par la suite cette technique DVR (pour *Direct Volume Rendering*) ainsi que DMVR (pour *Direct Multi-Volume Rendering*) en ce qui concerne son extension au rendu simultané de plusieurs volumes. Le cadre théo-



rique de cet outil est plus amplement détaillé à l'annexe C. La figure ci-contre montre l'image initiale de la savonnette sur laquelle nous avons, de plus, superposé les directions des axes Ox, Oy et Oz du support. À partir d'une image volumique initiale (figure 2.4(a)) nous effectuons diverses déformations pour obtenir autant de nouveaux volumes. Ces volumes sont ensuite affichés en mode DVR. Il faut bien garder à l'esprit que ce ne sont pas les rendus 2D qui sont déformés, mais les volumes déformés en 3D qui sont projetés et affichés en 2D.

$$|\frac{x}{a}|^m + |\frac{y}{b}|^n + |\frac{z}{c}|^l = 1, \quad \text{avec} \quad m, n, l \geqslant 2.$$

Pour les superquadriques que nous montrons ici, m = n = l = 8.

³La forme réduite de l'équation implicite d'une superquadrique est du type







- (a) Image Initiale.
- (b) Transformation avec $t_x = t_y = 50$ pixels, les autres paramètres étant nuls.
- (c) Transformation en fixant $a_x = b_y = cos(50^\circ)$ et $a_y = -b_y = -sin(50^\circ)$.



(d) Transformation avec $a_x = 2 * \cos(50^\circ)$, et les autres paramètres conservés.

(e) Transformation avec $a_x = b_y = 2 * \cos(50^\circ)$, et les autres paramètres conservés.

(f) Transformation avec $a_{xy} = b_{xy} = -0.008$, et les autres paramètres conservés.

Figure 2.3 — Exemples de déformations 2D sur l'image du manchot.

Nous appliquons d'abord seulement des translations et des rotations par rapport à l'axe des x (figures 2.4(b), 2.4(c), 2.4(d)) puis des déformations trilinéaires non-rigides (figures 2.4(e), 2.4(f), 2.4(g), 2.4(h)).

En ce qui concerne les images 2.4(e), 2.4(f), 2.4(g), 2.4(h), la différence porte sur les coefficients de la partie élastique (les 12 coefficients de xy, xz, yz et xyz pour chaque coordonnée). La première image est calculée avec un seul coefficient non nul (en xy pour la coordonnée x, ou $a_{1,1,0}$ selon la notation de l'équation (2.2)), la deuxième image possède en plus un coefficient non nul en xz pour la coordonnée y ($b_{1,0,1} \neq 0$), la troisième présente un coefficient en yzpour la coordonnée y ($b_{0,1,1} \neq 0$). Enfin la dernière image possède également un coefficient non-nul en xyz pour la coordonnée z ($c_{1,1,1} \neq 0$).

La réalité physique de ces coefficients, considérés isolément, est difficilement interprétable. Aussi nous préférons utiliser la description par amers évoquée dans la section 2.1.4.



Figure 2.4 — Exemples de déformations 3D sur un volume synthétique.

2.3 Distance géométrique et fonction de coût

La recherche d'une transformation élastique, ou plus précisément la recherche des 24 coefficients de la transformation trilinéaire (ou des 6 coefficients de la transformation rigide) se fait par maximisation d'un critère d'optimalité. On ajuste les paramètres jusqu'à trouver la fonction qui met en correspondance de manière optimale les points d'une image avec leur homologue sur l'autre image. C'est ce critère que nous appellerons fonction d'appariement (équivalent français de « *fitness function* »), ou de manière analogue fonction distance. La distance d entre deux images à recaler est inversement proportionnelle à la qualité du recalage (ou *fitness f*). La transformation optimale T^* qui maximise f et en même temps minimise d, est définie de la manière suivante⁴ :

$$T^* = \arg\min_{T \in \mathcal{S}} d(T(F_1), F_2) = \arg\max_{T \in \mathcal{S}} f(T(F_1), F_2).$$

L'estimation de la distance d ou de la fonction d'appariement f n'est pas une chose triviale ; par ailleurs, nous avons la contrainte supplémentaire d'obtenir une fonction distance rapide à calculer (ou du moins à estimer) afin de ne pas porter atteinte à l'efficacité de la recherche

⁴On notera indifféremment par la suite la distance $d(T(F_1), F_2)$ ou la distance $d(F_1, F_2 \circ T)$, puisque cela revient au même de se demander où les points p de F_1 se transforment sur l'image F_2 , que de comparer les points de F_1 avec leurs homologues par T sur F_2 (notés $F_2 \circ T$). Dans le premier cas, le calcul de la distance est effectué sur le domaine de l'image F_2 , alors que dans le second cas, le domaine considéré est celui de F_1 .

par algorithme génétique (cf. chapitre 3). Les paragraphes suivants détailleront les différentes étapes qui nous ont amené à définir une fonction distance adaptée à nos besoins.

2.3.1 Fonction distance et information structurelle

On trouve dans la littérature de nombreux critères d'estimation d'une distance géométrique entre deux images médicales avec différents avantages et inconvénients. Les mesures fondées sur la corrélation des niveaux de gris ne sont pas utilisables dans notre contexte (algorithmes génétiques) car, bien que précises en général, ces mesures sont longues à calculer. La mesure la plus rapide que nous ayons trouvée est une mesure utilisant une description structurelle de l'information; autrement dit, une mesure sur des images segmentées. La principale caractéristique commune entre les images IRM et scanner du cerveau et du crâne humain (images que nous nous proposons de recaler) est la présence de l'interface air/peau. Un prétraitement visant à l'extraction de ces points caractéristiques est appliqué (cf. chapitre 4) et permet alors de résumer le recalage 3D à un problème général de recalage de surfaces numériques. Une conséquence importante de cette approche structurelle (surfacique) est que la fonction distance se résume au calcul d'une distance entre deux surfaces numériques.

Soient S_1 et S_2 , les deux surfaces numériques extraites des images à recaler F_1 et F_2 . Notons $d_E(p_1, p_2)$ la distance euclidienne (dans \mathbb{R}^3) entre deux points p_1 et p_2 . Ainsi nous définissons la pseudo-distance⁵ d_a entre S_1 et S_2 respectivement à la transformation T par

$$d_a(S_1, S_2|T) = \frac{1}{\operatorname{card} S_1} \sum_{p_1 \in S_1} \min_{p_2 \in S_2} (d_E(T(p_1), p_2)).$$
(2.5)

Notons que si nous avions utilisé, à la place de la somme sur $p_1 \in S_1$ un opérateur max sur $p_1 \in S_1$, nous aurions été en présence d'une distance de Hausdorff orientée (voir chapitre 1). La dissymétrie naturelle de la distance de Hausdorff qui se décompose suivant deux distances : la distance directe $h(T(S_1), S_2)$ de $T(S_1)$ vers S_2 , et la distance indirecte $h(S_2, T(S_1))$ (avec les notations de l'équation (1.1) de la page 24), peut être interprétée comme correspondant aux aspects « hypothèses » et « tests » d'une méthode. Les transformations T qui minimisent la distance directe sont un ensemble d'hypothèses, hypothèses qui sont ensuite testées en calculant la distance indirecte.

Une manière de calculer la distance de Hausdorff entre un objet A et un objet B fait appel à la notion de surface de Voronoï. Une surface de Voronoï est ce que nous obtenons lorsque nous effectuons le calcul d'une transformée distance d'une image binaire. En posant $d(x) = \min_{b \in B} ||x - b||$ et $d'(x) = \min_{a \in A} ||a - x||$, on a alors :

$$H(A,B) = \max\left(\max_{a \in A} d(a), \max_{b \in B} d'(b)\right).$$

Ceci implique que l'on peut obtenir H(A, B) en calculant d(a) et d'(b) pour tout $a \in A$ et tout $b \in B$. Il se trouve que le graphe $\{(x, d(x)) | x \in \mathbb{R}^3\}$ correspond à une extension 3D d'une surface de Voronoï de l'image B. Il suffit alors pour connaître la distance orientée h(A, B) de

⁵Nous l'appelons pseudo-distance car elle n'est pas définie comme une distance au sens mathématique strict (symétrique, définie, positive). En particulier, la pseudo-distance d_a que nous introduisons n'est pas symétrique; elle ne satisfait pas la relation d(a, b) = d(b, a). Nous avons seulement d(a, a) = 0 et $d(a, b) \ge 0$. Il manque la relation $d(a, b) = 0 \implies a = b$.

retrouver la valeur maximale du graphe parmi les endroits où coïncident les points de A. La distance inverse est quant à elle calculée de manière identique.

Le calcul de la distance de Hausdorff étant décomposé en deux sous-distances, la distance directe et la distance inverse, le nombre d'opérations à réaliser, et par conséquent le coût de calcul sera bien plus élevé.

Même si le calcul de la distance directe de l'image à recaler transformée vers la référence est relativement facile (on dispose déjà de la transformée distance de l'image de référence), l'autre distance est bien plus complexe à évaluer.

Par ailleurs, la distance de Hausdorff est assez sensible à la présence de points aberrants ; par conséquent, nous préférons utiliser une moyenne (avec l'opérateur somme dans la relation (2.5)) sur tous les éléments présents.

D'un autre côté, avec cette définition de distance (relation (2.5)), nous pouvons rencontrer des problèmes en présence d'occlusions partielles. Lorsque des points d'une image n'ont pas de contrepartie sur l'autre image, la distance est biaisée. Nous verrons à la section 2.3.3 comment s'affranchir d'un tel problème.

2.3.2 Utilisation d'une carte de distances

Nous avions vu au chapitre 1 que le calcul de distances euclidiennes pouvait être simplifié en ayant recours à la technique du chanfrein (voir paragraphe 1.4.2.2) établissant une carte précalculée de distances entre des formes géométriques. Étant donnée la forme de l'équation (2.5), il est évident que l'on peut accélérer le calcul de la distance en pré-calculant les termes $\min(d_E(T(p_1), p_2))$ sous forme d'une carte de distances. Cette carte, relative à la surface S_2 extraite du volume de données F_2 , sera en fait un volume noté S_2 défini sur le domaine D_2 de F_2 , et dont la valeur de chaque voxel est calculée par la formule suivante :

$$\forall p \in D_2, \ \mathcal{D}_2(p) = \min_{p_2 \in S_2} \left(d_E(p, p_2) \right).$$
 (2.6)

La figure 2.5 montre un exemple d'une coupe 2D d'une carte de distances (figure 2.5(b)) calculée à partir d'une surface numérique issue d'une segmentation 3D (figure 2.5(a)). Les niveaux de gris correspondent à la distance à la surface (tracée également en surimpression).

Utilisant une carte de distances \mathcal{D}_2 de la surface S_2 , les équations (2.5) et (2.6) aboutissent à la définition de la distance d_m qui est calculée comme suit :

$$d_m(S_1, S_2|T) = d_m(S_1, S_2|T, \mathcal{D}_2) = \frac{1}{\operatorname{card} S_1} \sum_{p_1 \in S_1} \mathcal{D}_2(T(p_1)).$$
(2.7)

2.3.3 Utilisation d'une distance stochastique

La fonction d_m utilisant une carte de distances pré-calculée permet d'améliorer grandement le temps de calcul, néanmoins, ce temps reste proportionnel à la taille de la surface S_1 (card S_1). Avec les images que nous utilisons, la surface S_1 contient typiquement entre 30,000 et 60,000 points, ce qui rend la fonction d'appariement encore difficilement utilisable pour un algorithme d'optimisation stochastique, comme les algorithmes génétiques, où le nombre



(a) Coupe $n^{\circ}9$ du volume IRM + segmentation.

(b) Coupe n°9 de la carte + segmentation.

Figure 2.5 — Exemple de carte de distances d'une surface 3D.

d'évaluations de la performance des solutions potentielles prévaut sur la qualité de cette évaluation.

Aussi nous accélérons une fois de plus le calcul de cette fonction distance en implémentant une méthode stochastique. Le calcul ne se fait plus désormais sur l'ensemble des points de S_1 mais sur un nombre constant n d'échantillons choisis aléatoirement⁶. Si l'on note S_1^n une réalisation de la sélection aléatoire de n éléments de S_1 , notre nouvelle fonction distance peut s'écrire :

$$d_s(S_1, S_2|T) = \frac{1}{n} \sum_{p_1 \in S_1^n} \mathcal{D}_2(T(p_1)).$$

Néanmoins, comme nous l'évoquions plus haut, une distance géométrique entre deux ensembles de points reste toujours inhérente à la qualité de la segmentation d'une part, et à la présence d'occlusions partielles d'autre part. Remarquons également que ce biais est d'autant plus important lorsque l'on utilise une distance stochastique avec peu de points (typiquement moins d'une centaine). Aussi, nous avons recours à un filtrage non-linéaire (pas au sens d'une convolution mais plutôt d'une réévaluation des niveaux de gris) de la fonction distance afin d'améliorer la robustesse de la distance (2.5). Désormais, ne seront pris en compte dans la distance que les points compris dans un faible rayon de distance⁷, ce qui permet d'éliminer facilement les points ne semblant pas avoir d'homologues (problème d'occlusions partielles et parties manquantes). Une distance de 0 donne une contribution maximale de 1, et une distance infinie donne une contribution nulle. De plus, nous écartons du calcul de la performance tout les points se transformant en dehors de la carte de distances. In fine, nous effectuons le calcul de la distance sur n^0 points (avec $n^0 < n$ le nombre de points se transformant dans le champ de la carte de distances que nous avons notée \mathcal{D}_2) et avec un facteur de lissage

⁶Le tirage est renouvelé à chaque itération de l'algorithme génétique.

⁷La notion de « faible rayon de distance » sera réalisée à l'aide d'un coefficient de lissage σ .
σ choisi à $\sigma = 4$ millimètres (à mettre en relation avec la taille des images, soit environ $250mm \times 250mm \times 150mm$). L'expression mathématique de cette fonction d'appariement est donnée par l'équation (2.8) tandis que la figure 2.7 montre l'effet de l'application d'un facteur de lissage (avec $\sigma = 4$) sur la carte de distances 3D. Si les n^0 points choisis aléatoirement dans S_1 se transforment tous en des points de S_2 alors la fonction d'appariement atteindra la valeur maximale 1.

$$f(S_1, S_2|T) = \frac{1}{n^0} \sum_{p_1 \in S_1^0} \exp{-\frac{(\mathcal{D}_2(T(p_1)))^2}{\sigma^2}}$$
(2.8)

La mesure de similarité stochastique (2.8) s'apparente alors à un opérateur robuste de reconnaissance de forme et, comme tel, suppose que les structures S_1 et S_2 possèdent en commun au moins 50 % de leurs sous-structures. En prenant $n_0 = \operatorname{card} S_1$, nous obtenons une version déterministe de cette mesure de similarité, que l'on notera f^* pour les utilisations ultérieures.

Étant donnée la forme de la courbe $\exp(-x^2/\sigma^2)$ sur \mathbb{R}^+ (figure 2.6), les points se transformant à grande distance de la surface n'incrémentent pas la fonction de performance, alors que les points situés à faible distance participent à cette fonction. Par ailleurs, on constate que l'influence des points situés à une distance de 2 à 6 millimètres (avec $\sigma = 4$) varie grandement ce qui fait que la fonction de performance introduite (équation (2.8)) est très discriminante pour les points se transformant dans un proche voisinage de la surface à recaler.



Figure 2.6 — Forme de la fonction de lissage exponentielle avec $\sigma = 4$.

Sur la figure 2.7, nous avons représenté à gauche (image 2.7(a)) une coupe du volume de la carte de distances. Les pixels de valeur nulle (noir absolu) sont les points issus de la segmentation de la tête. Chaque pixel de cette carte de distances a pour valeur (niveau de gris) la distance euclidienne qui sépare le point considéré au plus proche point de la surface segmentée. Plus le niveau de gris est élevé, et plus on est loin de la surface segmentée. À droite (figure 2.7(b)), nous avons représenté le résultat de l'application de la fonction exponentielle sur cette coupe. Cette transformation revient à établir une bijection entre les niveaux de gris initiaux de la carte de distances avec une nouvelle distribution de valeurs. Sur l'image résultat, les points voisins de la surface segmentée ont désormais un niveau de gris proche de 1 (ou plutôt de 255 si l'on considère des images à 256 niveaux de gris), alors que les points situés au delà de 2σ ont un niveau de gris proche de 0 (noir).



(a) Carte distance originale. (b) Carte distance transformée.

Figure 2.7 — Vue en coupe d'une carte de distances avant et après changement des niveaux de gris $\sigma = 4$.

2.4 Synthèse du chapitre

Ce chapitre vient de traiter des divers modèles sous-jacents à notre problème de recalage élastique 3D multimodalité, tant au niveau de la transformation géométrique 3D, qu'au niveau de l'évaluation de la qualité d'une transformation 3D pour répondre à un problème de recalage.

Nous avons présenté les modèles paramétriques de déformation que nous allons utiliser :

- D'une part, le modèle de transformation rigide est régi par l'équation (2.1), et possède
 6 degrés de liberté (ou paramètres);
- D'autre part, la déformation élastique 3D est modélisée par une transformation trilinéaire selon l'équation (2.2), puis de manière plus pratique par la relation (2.3) possédant 24 degrés de liberté.

Ces deux modèles nous permettent à partir d'un jeu de paramètres donné de déformer, rigidement ou élastiquement, une image médicale 3D.

En outre, ce chapitre a permis de définir une fonction distance, ou fonction d'appariement stochastique, qualifiant l'aptitude d'une transformation (un jeu de paramètres du modèle de déformation) à recaler deux images médicales données dont le contenu est préalablement décrit sous forme structurelle. Cette fonction d'appariement stochastique est définie de manière à être calculée rapidement à partir d'une carte distance 3D; son expression mathématique est établie par la formule (2.8).

CHAPITRE 3 Recalage élastique 3D : Optimisation génétique

Ce chapitre aborde la partie optimisation du recalage 3D élastique. Nous expliquerons pourquoi nous avons choisi les algorithmes génétiques, puis nous présenterons le fonctionnement général de ces algorithmes.

Ensuite nous verrons quelles instances ont été utilisées (algorithmes fonctionnant avec des paramètres réels, ou avec des index). Les algorithmes génétiques évolutionnistes testés sans succès seront également évoqués. Enfin, nous aborderons le problème de l'inefficacité des algorithmes stochastiques s'ils ne sont pas couplés avec des algorithmes déterministes (recherche globale / recherche locale).

3.1 Pourquoi un algorithme génétique

Une fois modélisé le problème du recalage en imagerie médicale, nous nous rendons compte qu'il se résume à un problème d'optimisation : Comment trouver les paramètres optimaux qui répondent au problème posé. Différentes manières d'aborder le problème débouchent en général sur différents espaces de recherche, mais la philosophie est toujours la même; elle consiste à minimiser ou à maximiser une fonction de coût ou d'appariement.

Comme nous l'avons vu au chapitre 1, il existe diverses méthodes d'optimisation (déterministes ou stochastiques), et bien qu'elles aient en théorie le même but, toutes n'ont pas la même efficacité pour atteindre ce but. D'une part, les divers algorithmes de recalage élastique en imagerie médicale sont en général lents, aussi, nous avons cherché à améliorer cet état de fait en axant notre priorité sur les algorithmes d'optimisation rapides. Un algorithme de recherche exhaustive est par conséquent à proscrire. D'autre part, l'espace de recherche considéré étant assez vaste et pouvant comporter une pléthore d'optima locaux, nous avons cherché du côté des algorithmes robustes. Seuls les algorithmes génétiques semblaient répondre simplement à ces critères. Ils sont en effet bien adaptés aux espaces de recherche vastes, ils sont en général rapides, et de plus peu sensibles à la qualité intrinsèque de la fonction de coût (du fait de la stochasticité). D'ailleurs, HOLLAND écrit dans son livre [42] Adaptation in Natural and Artificial Systems « Il a été prouvé théoriquement et expérimentalement que les algorithmes génétiques sont des procédures robustes d'exploration d'espaces complexes. ».

Finalement, n'oublions pas non plus de mentionner que ce qui nous a poussé à utiliser des algorithmes génétiques c'est aussi la curiosité et la fascination. Il est en effet assez étonnant de voir comment des algorithmes aussi simples de conception et se contentant de reproduire des phénomènes naturels peuvent être efficaces pour résoudre des problèmes assez complexes.

3.2 Origines et Principes généraux

Les algorithmes génétiques sont inspirés d'observations concernant les mécanismes de la sélection naturelle. Ils ont été conçus à l'origine pour résoudre des problèmes d'adaptation, ce qui a amené les chercheurs, et en particulier John Holland de l'université du Michigan, à tenter de reproduire certains phénomènes naturels. En effet, tout se passe comme si la Nature avait conçu des êtres vivants selon des caractéristiques leur conférant des aptitudes à prendre en compte des conditions très difficiles. Les premières réflexions niant l'existence d'une « conception originelle de systèmes naturels figés », déjà adaptés pour toujours à toutes les conditions extérieures, ont été introduites par DARWIN, vers 1860, avec sa théorie de l'évolution des espèces.

L'idée de base de cette théorie est que, sous l'influence des contraintes extérieures, les êtres vivants se sont graduellement auto-modifiés, et ce, au travers du processus de reproduction, pour élaborer des générations d'êtres de mieux en mieux adaptés à leur milieu naturel. Cette idée est porteuse d'une notion d'auto-organisation, permettant aux espèces d'évoluer spontanément du simple vers le complexe.

Une approche de modélisation plus poussée de cette évolution au travers du processus de reproduction a été formalisée par Jacques MONOD, dans « le hasard et la nécessité » [66], dans les années 1960. Cette approche purement biologique insiste sur la diversité originelle qui s'appuie sur la combinatoire d'un petit nombre de caractéristiques de base. Le filtrage du hasard opéré par les conditions extérieures laisse vivre ou fait mourir alors un être, avant ou après qu'il se soit reproduit.

Des chercheurs en informatique ont ensuite étudié la possibilité de donner à des machines l'aptitude d'évoluer spontanément, de manière à s'adapter automatiquement à de nouvelles conditions de la même manière que les être vivants dans leur milieu naturel.

Lors de la conception de ces algorithmes génétiques, les chercheurs remarquèrent que ceux-ci présentaient des aptitudes remarquables à l'optimisation. Leur comportement semblait même plus robuste que celui de certains algorithmes classiquement dédiés à cette discipline. Cette propriété a été formalisée, et depuis les algorithmes génétiques ont fait l'objet de nombreux travaux de recherche opérationnelle.

3.2.1 Fonctionnement

L'algorithme génétique étant vu comme un algorithme d'optimisation globale, s'inscrit dans une chaîne de traitement entre le problème d'optimisation et la solution de ce problème (voir figure 3.1). D'un point de vue optimisation, l'algorithme génétique se comporte comme une boîte noire prenant en entrée un domaine de recherche (l'espace de recherche sur lequel se situent les paramètres du problème), une fonction de valuation (fonction de coût ou de performance), et enfin une information concernant le type d'optimisation à effectuer (maximisation ou minimisation). L'algorithme génétique peut nécessiter le réglage de certains paramètres (voir paragraphe 3.2.2). En sortie, l'algorithme génétique fournit une liste de solutions potentielles valuées (en fonction de leur performance). Libre à l'utilisateur d'utiliser cette liste d'une façon ou d'une autre (par exemple utilisation du meilleur élément uniquement, ou couplage avec un algorithme d'optimisation déterministe).



Figure 3.1 — Utilisation d'un algorithme génétique dans un problème d'optimisation.

Il existe quatre points importants qui rendent les algorithmes génétiques différents des méthodes d'optimisation classiques :

- 1. Les AG opèrent une mise en correspondance entre un espace de recherche (les paramètres du problème dans le monde réel), et un espace artificiel de chaînes codées (espace des gènes). Leur efficacité vis-à-vis d'un problème spécifique dépend de la pertinence de cet encodage.
- 2. Les AG fonctionnent sur une population de points, et non pas un point isolé de l'espace de recherche. Cette propriété est à l'origine du parallélisme implicite des AG.
- 3. Les AG utilisent seulement les valeurs de la fonction de coût. Ils ne supposent pas l'existence, ni ne calculent, la dérivée de cette fonction ou une autre connaissance auxiliaire (toutes les connaissances a priori sont inclues dans l'encodage).

 Les AG utilisent des règles de transition probabilistes et non déterministes, appliquées au sein de l'espace codé¹.

Le fonctionnement d'un algorithme génétique est régit par quatre règles que sont d'une part, la reproduction, la mutation, la recombinaison, et d'autre part la propriété d'évolution itérative (répétition cyclique des opérateurs génétiques jusqu'à ce qu'une condition d'arrêt soit satisfaite).

3.2.1.1 La sélection et la reproduction

De même que la théorie de l'évolution considère une population et non pas un individu isolé, les algorithmes génétiques agissent sur des populations d'individus (N chromosomes), chacun étant codé par un génotype (information génétique), par analogie avec la biologie. Ce génotype est codé traditionnellement par une chaîne de bits, où chacun des l paramètres du problème (l gènes) sont codés sur q bits (q allèles), ce qui fait un total de $L = l \times q$ bits, comme le montre la figure 3.2.



Figure 3.2 — L'encodage naturel d'un chromosome : un vecteur binaire.

Le mécanisme de l'algorithme génétique va opérer des reproductions de ces individus, à cardinal N constant (voir figure 3.3), et donc des chaînes de bits. Cette reproduction est réalisée par un opérateur utilisant une fonction sélective (fonction de *fitness* ou de performance) qui régit le nombre de duplications de chaque individu, en fonction de son aptitude à satisfaire un critère donné (évalué par le biais de cette fonction). D'un point de vue déterministe, un chromosome i ayant une performance individuelle f_i , possède un nombre caractéristique D(i) défini par :

$$D(i) = \frac{f}{\overline{f}}$$
 avec $\overline{f} = \frac{1}{N} \sum_{i=1}^{N} f_i.$

Afin de réduire l'erreur stochastique induite par un échantillonnage direct, nous utilisons la méthode à deux passes introduite par Goldberg dans [34]. La première étape conserve N_0 géniteurs en dupliquant chaque individu de la population un nombre de fois égal à la partie entière de D(i). Ensuite, on utilise l'erreur de troncature $\epsilon_r(i)$ pour choisir aléatoirement (en fonction de la probabilité $p_r(i)$ définie par l'équation (3.1)) les $N - N_0$ individus complétant la population engendrée (la notation |D(i)| désigne la partie entière de D(i)).

¹Ceci est vrai pour les algorithmes génétiques que nous appellerons canoniques (mutation, recombinaison, reproduction, sélection), par opposition aux algorithmes génétiques évolutionnistes pouvant utiliser des opérateurs déterministes en plus des opérateurs de base.

$$p_r(i) = \frac{\epsilon_r(i)}{\sum_{i=1}^N \epsilon_r(i)} \quad \text{avec} \quad \epsilon_r(i) = D(i) - \lfloor D(i) \rfloor$$
(3.1)

D'autres contraintes comme la mise à l'échelle et la normalisation peuvent être utilisées pour transformer la performance brute (f) donnée par la fonction de *fitness* en une performance corrigée (f_c) avant le processus de sélection. Ces transformations peuvent aider à éviter que les meilleurs individus des premières générations ne dominent trop rapidement la population et impliquent alors une convergence prématurée de la population vers un extremum local. Toutes ces techniques de correction de performance sont expliquées plus en détails dans [34].



Figure 3.3 — La reproduction.

Maintenant, afin de pouvoir évoluer parmi l'ensemble des solutions recherchées, on ne va pas se contenter de seulement effectuer des copies conformes, mais plutôt d'introduire des phénomènes parasites de mutation et de recombinaison (ou *crossover*).

3.2.1.2 La mutation

À chaque itération génétique, un tirage aléatoire conditionnellement à une probabilité de mutation p_m (notons que même si p_m est une grandeur extrinsèque, on choisit souvent dans la littérature $p_m = 1/L$ où L désigne le nombre de bits du chromosome [35]) est effectué pour sélectionner un sous-ensemble de chromosomes devant subir une mutation génétique.

Le cas le plus simple de mutation consiste à générer une erreur sur un bit de la chaîne. Celle-ci est réalisée sur un bit tiré au hasard d'un individu tiré au hasard. On lui attribue une probabilité de mutation. Cette probabilité est utilisée par l'algorithme génétique pour générer les erreurs de recopie. On génère ainsi un nouvel individu qui n'existait pas auparavant et on explore aléatoirement l'espace de recherche.

Il est à noter qu'il existe d'autres types de mutations, proposant par exemple de modifier un paramètre dans un voisinage du paramètre actuel (mutation locale). La mutation uniforme ou gaussienne se propose de modifier successivement tous les paramètres d'un chromosome élu pour la mutation, conditionnellement à une probabilité, justement uniforme ou gaussienne.

3.2.1.3 La recombinaison entre individus

De même que la mutation est basée sur le hasard, la recombinaison effectue, à chaque itération génétique, un tirage aléatoire (conditionnellement à une probabilité de recombinaison



Figure 3.4 — La mutation.

 p_c , de valeur typique 80 %). Ce tirage sélectionne un sous-ensemble de paires de chromosomes devant subir une recombinaison génétique.

La recombinaison est, à la base, un phénomène biologique qui intervient lors de la méiose² (en prophase).

On peut voir cet opérateur de recombinaison (appelée en anglais *crossover*) comme une « erreur de lecture » due à la superposition partielle de certaines chaînes lors de la recopie : deux chaînes se croisent et la lecture considère qu'elles ne se croisent pas.

La recombinaison entre individus permet de fabriquer de nouveaux individus ayant des caractéristiques communes avec leurs parents. Cet « héritage » concerne le plus souvent des bonnes caractéristiques, puisque seuls les individus les mieux adaptés vivent suffisamment longtemps pour se reproduire, ce qui peut conduire à des individus encore mieux adaptés.



Figure 3.5 — La recombinaison (ou crossover).

La figure 3.5 montre l'exemple standard de la recombinaison monopoint. Ce type de recombinaison possède l'inconvénient de dépendre de l'ordre dans lequel les paramètres sont encodés. En effet, pour chaque chromosome subissant une recombinaison, le premier et le dernier paramètre seront toujours dissociés, alors que le premier et le deuxième par exemple ont très peu de chances de se séparer. Un premier pas vers la résolution de ce problème consiste à

- Métaphase : les chromosomes se fixent sur les fibres du fuseau chromatique et se disposent en plaque, la plaque équatoriale, à égale distance des deux pôles de la cellule.
- Anaphase : les chromosomes se scindent en deux lots qui se séparent vers les pôles (ascension polaire).
- Télophase : les chromosomes perdent leur individualité et forment à nouveau un réseau autour duquel se reforme une enveloppe nucléaire; les nucléoles se réorganisent. Le cytoplasme de la cellule mère se répartit en 2 masses égales autour des noyaux des 2 cellules filles qui se séparent : cytodiérèse.

 $^{^{2}}$ La méiose est le processus de division des cellules vivantes qui précède la formation des cellules reproductrices (gamètes). Elle est constituée de 4 phases :

⁻ Prophase : à partir du réseau de chromatine s'individualisent des filaments enchevêtrés, les chromosomes au nombre constant pour une même espèce. Les nucléoles (Corps sphérique très riche en A.R.N., situé à l'intérieur du noyau des cellules) se désorganisent, l'enveloppe nucléaire disparaît tandis que s'organise un fuseau fait de microtubules qui oriente le sens de la division. C'est à ce moment-là que les chromosomes sont regroupés par paires et qu'ils se scindent en deux chromatides.

introduire le concept de recombinaison cyclique (ou bi-points) comme le montre la figure 3.6. Au lieu de considérer le chromosome comme un segment, on considère une boucle où le premier et le dernier gène sont jointifs. La recombinaison consiste alors a diviser la boucle en deux segments (choix de deux points de *crossover*) puis à échanger l'un des segments entre les deux chromosomes.



Figure 3.6 — Crossover cyclique ou bi-points.

Enfin, de la même manière qu'il existe des opérateurs de mutation uniforme, on trouve des recombinaisons uniformes. Chacun des gènes de la paire de chromosomes sont successivement aléatoirement échangés en fonction d'une loi de probabilité (uniforme ou gaussienne par exemple).

3.2.1.4 Cycle génétique complet

Maintenant voyons, à l'aide de la figure 3.7, comment tout cela peut s'enchaîner. Une population initiale est élaborée, puis l'opérateur de sélection/reproduction filtre les individus de manière à ne conserver que ceux qui présentent de bonnes aptitudes pour répondre au problème posé; les chromosomes sont ensuite regroupés par paires et un certain nombre de ces paires (selon la probabilité p_c) subira une recombinaison. Une part (selon p_m) des chromosomes peuvent ensuite subire une mutation. Enfin cette population finale vient remplacer la population initiale, et le cycle peut recommencer; ce dernier se termine uniquement lorsqu'une condition d'arrêt est atteinte.

Pour résoudre un problème en utilisant un algorithme génétique, il s'agit de choisir une représentation des solutions sous la forme d'un codage binaire. Comme précisé plus haut, l'algorithme génétique ne manipule pas les solutions directement, mais les représentations (c'est-à-dire des codes binaires représentant les solutions d'un problème). Il est donc important que la représentation binaire des solutions respecte une notion de topologie.

De plus, il faut remarquer la ressemblance entre les algorithmes génétiques et celui du recuit-simulé. En effet, tout comme ce dernier, les algorithmes génétiques sont des filtres de hasard : le tirage aléatoire d'une modification trouve son homologue dans deux modifications aléatoires : la mutation et la recombinaison. La fonction de sélection probabiliste dépendant d'une température (algorithme du recuit-simulé) est généralisée à une fonction de sélection de sélectio



Figure 3.7 — Cycle complet définissant un algorithme génétique.

aux algorithmes génétiques comme par exemple la robustesse face aux optima locaux. Le principal point qui différencie ces deux types d'algorithmes est la manipulation intrinsèque d'une population, propre aux algorithmes génétiques, alors que le recuit-simulé manipule un seul individu. Cette manipulation d'une population est concrétisée par trois principes : la recombinaison entre individus, le partage de la fonction de coût ou d'appariement, et ce que l'on appelle le parallélisme implicite.

De plus amples digressions sur les origines et les particularités des algorithmes génétiques sont développées dans le livre *Genetic Algorithms* de GOLBERG [35]. De même, il existe une étude intéressante du comportement des algorithmes génétiques (appliqués à l'optimisation de fonctions) en fonction de l'initialisation et des paramètres des différents opérateurs [13]. Le but de la thèse n'étant bien sûr pas de faire un cours sur les algorithmes génétiques, nous nous contenterons de cette brève et générale présentation. Néanmoins, avant de détailler un exemple d'utilisation d'algorithmes génétiques sur un problème concret et simple (maximisation d'une fonction continue à deux dimensions, sur un domaine où elle ne possède qu'un maximum global, et divers maxima locaux), nous voulons insister sur le fait que les algorithmes génétiques ne sont pas seulement des outils expérimentaux, que leur robustesse a fait l'objet d'études approfondies (p.ex. FITZPATRICK [31], lequel a même poussé cette étude à l'application au recalage rigide en imagerie médicale [32]).

D'autres apparitions intéressantes d'algorithmes génétiques appliqués au recalage d'images (médicales ou non) que nous pouvons noter dans la littérature sont le recalage rigide multimodalité [86], le recalage d'objets numériques [16], le recalage élastique monomodalité d'os longs à partir de l'information en niveaux de gris [45, 46], et son extension directe en multimodalité [81]. Enfin notons aussi l'utilisation d'algorithmes génétiques pour la modélisation de surfaces articulaires [48, 50].

3.2.2 Les principaux paramètres des AG

Les paramètres les plus importants d'un algorithme génétique sont les suivants :

- Taille de la population. Le nombre N de chromosomes constituant une population génétique. Plus ce nombre est élevé, et plus le parallélisme implicite de l'algorithme est important. La valeur optimale de N dépend du problème considéré et surtout de la taille de l'espace de recherche. Une population trop petite vis-à-vis du problème à traiter aboutira nécessairement vers une convergence prématurée de l'algorithme.
- **Type d'encodage.** Celui-ci désigne la méthode utilisée pour représenter les valeurs des paramètres du problème au sein des chromosomes. Classiquement, on utilise une représentation sous forme de chaîne de bits, avec ou sans codage de gray, représentant des réels ou des entiers, selon un certain espace de recherche borné, etc. On trouve également des AG utilisant un encodage direct des paramètres du problème sous forme de nombres réels ou entiers (voir paragraphe 3.4.2).
- **Polarité de la recherche.** Indique si l'algorithme doit maximiser (fonction de performance) ou plutôt minimiser (fonction de coût) la fonction d'évaluation des chromosomes.
- Nombre de générations. Le nombre de générations est le nombre maximal de cycles (voir figure 3.7) que l'algorithme peut effectuer. Plus ce nombre est élevé, plus l'algorithme aura le temps de converger, voir d'affiner les solutions trouvées. Ce nombre est une borne maximale car l'algorithme génétique peut être interrompu prématurément si, par exemple, un critère de performance suffisante est atteint.
- Longueur du chromosome. La longueur L du chromosome correspond, dans le cas d'un encodage binaire, au nombre de bits (ou allèles) constituant le chromosome. Sinon elle correspond au nombre de paramètres directement encodés dans le chromosome (aussi appelés les gènes).
- **Type de recombinaison.** Ce paramètre permet de choisir quel genre de recombinaison sera effectuée. Cette recombinaison peut être soit monopoint, soit bi-points, soit uniforme.
- **Probabilité de recombinaison.** La probabilité de recombinaison p_c typiquement égale à 80 % désigne la probabilité qu'une paire de chromosomes de la population soit élue pour la recombinaison (ce qui correspond donc au pourcentage de la population atteint par la recombinaison à chaque itération du cycle génétique).
- **Type de mutation.** Ce paramètre permet de choisir quel genre de mutation sera effectuée, à savoir soit une mutation sur un bit, soit uniforme sur l'ensemble du chromosome, soit locale sur l'espace de recherche.
- **Probabilité de mutation.** La probabilité de mutation p_m désigne la possibilité qu'un chromosome de la population soit élu pour la mutation (ce qui correspond donc au pourcentage de la population atteint par la mutation génétique à chaque itération du cycle génétique).
- **Initialisation.** Ce paramètre désigne la méthode de formation de la population initiale de chromosomes. En général cette étape correspond juste à un tirage aléatoire de points sur l'espace de recherche, c'est-à-dire, dans le cas d'un encodage binaire, à un remplissage aléatoire des L bits des N chromosomes initiaux.

3.3 Exemple d'algorithme génétique sur un problème concret

Cette section a pour but de montrer un exemple d'application d'un algorithme génétique sur un problème d'optimisation simple. Nous nous proposons de trouver le maximum d'une fonction analytique à deux paramètres. Bien qu'un algorithme génétique soit un outil surdimensionné pour résoudre un problème aussi simple cela permettra d'expliquer comment ils sont utilisés.



(c) Fonction 2D à maximiser z = u(x) * v(y).

Figure 3.8 — Fonction test 2D à maximiser.

La fonction 2D (cf. figure 3.8) sur laquelle nous nous proposons de rechercher le maximum est la fonction analytique z = u * v avec $u = 1 - \cos(x)\cos(x/2)$, et $v = 1 + \cos(y)\cos(y/3)$. C'est donc une fonction à deux paramètres z = f(x, y). Nous limiterons l'espace de recherche à $[-\pi, 3\pi]$ sur x et $[-\pi, 2\pi]$ sur y de telle sorte qu'on l'étudie sur une seule période et qu'il n'existe qu'un seul maximum global en $(x, y) = (2\pi, 0)$ avec $f(2\pi, 0) = 4$.

Pour résoudre ce problème, nous avons utilisé un algorithme génétique simple dont les

paramètres sont résumés au tableau 3.1.

Paramètre	valeur
Taille de la population	50 chromosomes
Nombre de générations	150 cycles
Longueur d'un chromosome	L = 32 bits
Type de recombinaison	Monopoint
Probabilité de recombinaison	85~%
Type de mutation	1 seul bit
Probabilité de mutation	$p_m = 1/L = 1/32 = 3 \%$
Type d'encodage	Réels codés sur 16 bits en code de gray
Polarité de l'optimisation	Maximisation
Population initiale	Aléatoire

Tableau 3.1 — Paramètres de l'algorithme génétique permettant de maximiser $u \times v$.

Les deux figures 3.9(b) et 3.9(a) montrent l'évolution, au cours des générations génétiques, de la performance du meilleur élément, de la performance moyenne ainsi que des valeurs des paramètres du meilleur élément. La population initiale est constitué par tirage aléatoire de 50 points dans l'espace de recherche considéré. Nous constatons que la performance du meilleur élément augmente par saut successifs de générations en générations. Cette performance ne peut pas décroître car d'une part la fonction à maximiser est déterministe, et d'autre part la sélection est élitiste, et par conséquent conserve, quoi qu'il arrive, toujours le meilleur élément. La performance moyenne quant à elle augmente continuement, au début, et une fois que l'algorithme semble s'être rapproché fortement de l'optimum global, cette moyenne devient chaotique car la probabilité de trouver de nouveaux éléments avec une meilleure performance est faible, et par conséquent à chaque nouveau tirage d'individu, on crée de mauvais éléments.



(a) Performance moyenne et max sur la population.

(b) Valeur des paramètres du meilleur élément (x, y).

Figure 3.9 — Évolution de l'algorithme génétique pour la maximisation d'une fonction 2D.

Par ailleurs, notons que le brusque saut que l'on observe sur la performance du meilleur individu est parfaitement corrélé avec le saut de la valeur du paramètre en y vers la 30^{e} génération. Le meilleur élément de la population génétique était initialement *pris* au voisinage

de l'optimum local $x = 2\pi$, $y = \pi$, mais malgré cela, l'algorithme a réussi à le faire *sortir* de cet optimum local pour aller se caler sur l'optimum global.

Pour ce qui est du temps d'exécution de cet algorithme, nous ne pouvons rien conclure si ce n'est qu'avec un problème aussi simpliste, l'algorithme met quelques secondes pour le résoudre, alors qu'un algorithme déterministe disposant de l'équation de la fonction à maximiser serait quasiment instantané. Des évaluations de rapidité de l'algorithme plus à propos seront données dans la suite de ce manuscrit, lorsqu'on abordera l'application aux images médicales de l'algorithme de recalage que nous avons développé.

3.4 Les AG utilisés

Comme nous allons le voir au chapitre suivant, nous avons eu recours à deux algorithmes d'optimisation globale différents. Le premier, qui consiste à estimer un recalage rigide initial, opère directement sur l'espace des nombres réels (codés par la machine qui sert de support bien entendu). Un second algorithme génétique travaillera ensuite sur un espace de recherche indexé (discret).

3.4.1 Algorithmes génétiques et paramètres réels

Pour effectuer un recalage rigide d'images volumiques et en utilisant le modèle de déformation rigide décrit par l'équation 2.1 nous devons rechercher 6 paramètres réels. Ces 6 paramètres sont les 3 valeurs de la translation 3D, ainsi que les 3 angles de rotation 3D. L'espace de recherche qui pourrait être en théorie $\mathbb{R}^3 \times [-\pi, \pi]^3$ est restreint à des valeurs vraisemblables pour des images médicales. Les translations sont alors restreintes à une quarantaine de centimètres, et les angles de rotation sont respectivement compris dans $[\pi, \pi] \times [-\pi, \pi] \times [0, \pi]$. Nous aurions pu utiliser un algorithme génétique canonique (opérant sur des chaînes de bits représentant des nombres) tel que celui que nous avons utilisé pour résoudre le problème à deux dimensions (voir paragraphe 3.3). Néanmoins, pour éviter des problèmes d'échantillonnage de l'espace de recherche continu, nous avons directement encodé les chromosomes comme une succession de six nombres réels.

Par ailleurs, nous avons introduit une contrainte forte de répartition de la population génétique sur l'espace de recherche en utilisant le principe des carrés Latins [83]. Cette dernière contrainte assure que depuis la population initiale et après chaque itération de l'algorithme génétique un certain pourcentage de la population se trouve dans certains intervalles de l'espace de recherche. Autrement dit, on s'assure que la projection de la population sur chacun des axes de recherche reste supérieure à un minimum donné sur chaque intervalle de l'axe ; la répartition initiale étant uniforme. Pour ce faire on divise les six axes de l'espace de recherche en intervalles réguliers, et on s'assure qu'au moins un minimum de chromosomes se projettent dans chacun des intervalles. La figure 3.10 montre un exemple de répartition avec projection uniforme sur un espace 2D (donc pour un problème ayant seulement 2 paramètres). Nous voyons sur cette figure que chacune des colonnes et des lignes comportent un minimum de 2 éléments, ce qui veut dire que 20 (2 éléments pour 10 sous espaces) individus sont contraints sur la population de 21 éléments. En général la contrainte que nous imposons n'est pas aussi forte dans le sens où nous appliquons cette contrainte de répartition à, au plus 20 % de la population (le reste est positionné aléatoirement dans l'espace de recherche).

Le fait d'utiliser un encodage sous forme de nombres réels, et par ailleurs d'apporter une contrainte sur la répartition de la population dans l'espace de recherche implique la redéfinition



Figure 3.10 — Exemple de répartition uniforme sur le principe des carrés Latins en 2D.

des opérateurs génétiques.

3.4.1.1 Nouvel opérateur de mutation

Le principe de la mutation est d'explorer aléatoirement l'espace de recherche. Un chromosome subissant une mutation a de fortes chances de conduire à un point de l'espace de recherche qui n'avait jusqu'ici pas été étudié. La mutation est donc responsable : soit de l'affinement, soit de l'exploration³. Étant donnée la contrainte de répartition par carrés Latins, le soucis d'exploration globale n'a plus de raison d'être ; dès lors, le rôle de la mutation n'est plus que d'affiner les solutions sur leur voisinage. Avec une description sous forme de carrés Latins, et avec un encodage sur des nombres réels, la mutation peut simplement avoir pour effet de remplacer un ou plusieurs paramètres du chromosome par des points du même sous-espace de recherche (cf. figure 3.11 sur laquelle les cases correspondent aux sous-espace contraints, et ou les mutations déplaçant un chromosome d'une case à une autre sont interdites car non-locales et perturbant l'homogénéité de la répartition). On dit alors que la mutation est une mutation locale ; elle ne perturbe pas la répartition globale de la population sur l'espace de recherche tout en conservant son rôle d'affinement.

3.4.1.2 Nouvel opérateur de recombinaison

En ce qui concerne l'opérateur de recombinaison, son rôle est d'assurer un mélange et une transmission de parties des gènes des chromosomes. Ainsi, couplé avec un opérateur de sélection laissant se dupliquer les bons éléments et supprimant les mauvais, les gènes sont filtrés, et les *bons* sont transmis et recombinés. Cette propriété est intéressante car elle

 $^{{}^{3}}$ Exploration dans le sens où la mutation permet de donner une chance à tous les points de l'espace de recherche d'être atteints.



Figure 3.11 — Mutation locale dans un espace de recherche contraint.

permet à partir de deux chromosomes de moyenne performance de créer un chromosome ayant une performance supérieure (cumul des bonnes propriétés des parents). En terme de codage avec des nombres réels, cette propriété de l'opérateur de recombinaison peut être conservée en échangeant des paramètres entiers⁴. Notons de plus qu'une telle définition de la recombinaison ne modifie pas la répartition globale des chromosomes sur l'espace de recherche.

Parmi les différents types de recombinaison (monopoint, cyclique ou bi-points, uniforme), nous avons opté pour la version uniforme ayant l'avantage de ne pas dépendre de l'ordre d'encodage des paramètres du problème, et restant simple à utiliser.

3.4.1.3 Nouvel opérateur de sélection

L'opérateur de sélection dans le cadre d'un encodage par nombre réel n'a pas de raison d'être différent de l'opérateur de sélection habituel (avec des chaînes de bits). On se contente de calculer la performance d'un chromosome, quel que soit la manière dont les paramètres sont codés. Néanmoins, la sélection, par le biais de la duplication et de la regénération d'individus, modifie la distribution des individus dans l'espace de recherche. Aussi nous avons été obligés d'introduire une technique de contrôle dynamique de l'homogénéité de l'espace de recherche. Après chaque cycle de reproduction ou de sélection, on déplace certains individus des régions surreprésentées vers les régions ne possédant pas le minimum d'échantillons imposé.

3.4.2 Algorithmes génétiques et espace de recherche indexé

Nous avons vu au paragraphe 2.1.4 qu'une transformation élastique trilinéaire pouvait se définir à l'aide de 8 correspondances entre des points de S_1 et des points de S_2 . Aussi, le deuxième algorithme génétique que nous avons utilisé n'opère pas sur des paramètres réels

⁴Notons que ceci ne serait pas possible dans le cas d'un problème monodimensionnel, et fortement biaisé pour un problème à seulement deux ou trois paramètres.

mais sur un espace discret. Le deuxième type d'optimisation que nous avons à effectuer consiste en effet à rechercher parmi deux listes de points (la liste des sites de S_1 et celle des sites de S_2) 8 associations de points d'une liste à une autre. Le critère de performance calculé pour un choix donné de 8 couples sera exprimé au chapitre 4.

Pour résoudre le problème du choix de 8 associations, nous commençons par créer une liste des couples *possibles* (nous expliquerons au chapitre suivant ce que nous définissons comme *possible*), puis nous laissons le soin à l'algorithme génétique de trouver 8 index de couples distincts, représentant une solution potentielle de transformation élastique 3D globale, ayant une performance maximale. L'espace de recherche est donc l'espace des couples qui est à 8 dimensions et discret.

Les index sont encodés dans les chromosomes tout simplement comme des nombres entiers puisque l'espace de recherche n'est pas ordonné (deux index consécutifs ne correspondent pas forcément à deux couples de points proches). Autrement dit, puisque l'on a pas à exploiter un espace de recherche ordonné, il n'y a pas de raison d'utiliser un encodage par chaîne de bits. La figure 3.12 montre la constitution d'un chromosome comme un 8-uplet de couples.



Figure 3.12 — Encodage du chromosome sur l'espace de recherche indexé.

La seule contrainte que nous apportons à cet encodage est que les chromosomes, pour correspondre à une solution calculable, doivent toujours comporter 8 couples *différents*. Cette condition est nécessaire mais pas suffisante, car, comme nous l'avons dit, un chromosome fait correspondre 8 points d'une liste avec 8 points d'une autre liste, et que chacun des 8 points de chaque liste doivent être distincts. Un chromosome sera donc cohérent à la condition que les 8 couples utilisent 8 points distincts dans chacune des listes (8 points distincts pour les 8 amers de chaque surface). Aussi, un soin particulier quant au choix de la mutation et de la recombinaison a du être apporté.

La mutation revient, dans notre cas, à changer certains index du chromosome par d'autres index de l'espace des couples. Si les nouveaux index introduits rompent la condition de cohérence du chromosome, un nouveau tirage à lieu, et ce jusqu'à ce que le chromosome soit cohérent. La figure 3.13 illustre l'opérateur de mutation uniforme utilisé. Notons que la probabilité de mutation uniforme pour chaque index est tirée uniformément pour chaque chromosome élu pour la mutation. Ce tirage décide en fait du nombre d'index qui en moyenne sont changés dans le chromosome.

En ce qui concerne la recombinaison, le principe est presque le même. On utilise un opéra-



Figure 3.13 — Principe de la mutation uniforme sur un espace de recherche indexé.

teur de recombinaison uniforme, et à chaque échange probabiliste de gènes (rappelons qu'un gène correspond à un couple d'index), on effectue réellement l'échange si et seulement si le changement des couples n'engendre pas de dégénérescence. Ainsi, la figure 3.14 montre le procédé de recombinaison uniforme sur un espace indexé.



Figure 3.14 — Principe de la recombinaison uniforme sur un espace de recherche indexé.

3.5 Les limites des algorithmes génétiques

Ce qui fait qu'un algorithme génétique est efficace ou non vient certainement de l'encodage utilisé. Toute la difficulté consiste à trouver un encodage bien approprié au problème à traiter. David Goldberg [34] décrit divers problèmes courants ayant chacun des encodages spécifiques adaptés.

Mis à part le problème de l'encodage, de mauvais résultats peuvent être observés avec un algorithme génétique lorsque celui converge prématurément, ou trop vite, vers un optimum local. Plusieurs causes peuvent être à l'origine d'un tel phénomène. Tout d'abord, une très mauvaise initialisation de l'espace de recherche (par exemple une large désertification de l'espace englobant l'optimum global, et en même temps une grande densité de points autour d'un bon optimum local). Ensuite, un mauvais choix de paramètres (probabilités des opérateurs génétiques, taille de la population, etc.) peut également influer sur le résultat. Néanmoins, lorsque les paramètres sont fixés à des valeurs non extrêmes, la seule réelle influence du choix des paramètres que l'on observe est l'impact sur le temps global d'exécution, et pas sur la qualité du résultat obtenu, ce qui confirme la robustesse des AG.

Enfin, le dernier problème que l'on rencontre à l'usage d'un algorithme génétique est le choix du critère d'arrêt. Comment décider que l'algorithme a suffisamment convergé ? Doit on s'attacher à observer la performance moyenne de la population, la performance du meilleur individu, ou le nombre global d'itérations ? Ce problème n'a pas de réponse unique. En ce qui nous concerne, nous avons opté, de manière simpliste⁵, pour deux conditions d'arrêt :

- Le meilleur individu a atteint la performance théorique maximale⁶ (si cette dernière possède une borne maximale), où lorsque le meilleur individu a atteint une performance jugée suffisante par l'utilisateur (lorsque l'optimisation est utilisée pour faire de l'approximation rapide par exemple);
- Un certain nombre d'itérations globales ont été effectuées.

3.6 Algorithmes génétiques évolutionnistes

On trouve dans la littérature des utilisations d'algorithmes génétiques possédant des opérateurs évolués, éventuellement déterministes. Parmi tout ces opérateurs citons par exemple l'emploi de l'opérateur de recombinaison à trois parents selon l'algorithme du simplexe [101]. Nous avons essayé d'introduire de tels opérateurs dans notre algorithme de recalage rigide (fonctionnant sur les réels) sans y trouver un grand intérêt. Nous avions introduit un opérateur d'optimisation locale telle que la descente de gradient permettant de chercher dans un voisinage du chromosome considéré une direction de plus forte pente, et de déplacer le chromosome dans cette direction. Nous espérions ainsi améliorer la convergence de l'algorithme.

Néanmoins, avec cet opérateur de descente de gradient local, et pour différents paramètres (probabilité de mutation, de recombinaison, de descente de gradient, taille de la population génétique, nombre d'itérations) nous avons constaté que nous n'obtenions pas *in fine* de meilleurs résultats qu'avec l'algorithme génétique tel que celui décrit au paragraphe 3.4.1. Le nouvel algorithme nécessitait certes moins de cycles génétiques au total, mais avait un temps global d'exécution plus long (l'opérateur introduit étant plus complexe). De même, il semblait difficile d'introduire des opérateurs évolutionnistes pour la recherche de solutions sur un espace indexé, et c'est pour cela que notre deuxième algorithme génétique n'utilise pas non plus ce genre d'opérateurs.

La conclusion que nous tirons d'une telle expérience est que les algorithmes génétiques sont efficaces tant qu'ils restent simples. Il vaut mieux exécuter un grand nombre de fois des opérations simples, que peu de fois des opérations compliquées. Le fait d'utiliser des opérateurs compliqués revient quasiment à faire de l'optimisation avec un algorithme déterministe, ce qui est contraire à la philosophie des algorithmes génétiques.

3.7 Couplage avec un algorithme d'optimisation locale

Autant les algorithmes génétiques sont efficaces pour s'approcher rapidement de *la* solution globale d'un problème présentant une fonctionnelle de distribution chaotique, autant ils ne sont pas efficaces pour résoudre un problème où la fonction de coût est convexe. Or, pour tout les problèmes continus, il existe une échelle d'observation à laquelle le voisinage de l'optimum global (de même que pour les optima locaux) peut être considéré comme convexe. De cette

⁵Notons que nous nous limitons à des critères d'arrêt simplistes car l'espace des solutions obtenues par AG est ensuite affiné par un algorithme de recherche locale et déterministe (voir paragraphe 3.7).

⁶Ce qui est quasiment impossible à obtenir pour le recalage de données réelles.

petite réflexion, on en déduit que les algorithmes génétiques sont, par nature, efficaces pour dégrossir un problème chaotique continu, mais inefficaces pour affiner la solution finale. Aussi, nous considérons que tout les algorithmes génétiques doivent être couplés avec un algorithme de recherche locale déterministe ou non.

En poursuivant ce raisonnement, il est intéressant de constater que les algorithmes génétiques peuvent être utilisées comme **échantillonneurs stochastiques robustes** de l'espace de recherche. L'algorithme génétique fournit, en effet, une liste de points voisins d'optima locaux contenant, avec une forte probabilité, des voisins de l'optimum global du problème considéré. Cette liste peut être utilisée comme un ensemble de points de départ pour des algorithmes d'optimisation déterministes dont la principale faiblesse réside justement dans l'initialisation.

Le chapitre suivant, consacré à la description de la chaîne de traitement globale mise en œuvre pour le recalage d'images médicales, expliquera plus en détail la mise en cascade d'algorithmes génétiques et d'un algorithme d'affinement local.

3.8 Synthèse du chapitre

Ce chapitre vient de traiter du problème d'optimisation lié au recalage d'images médicales, ou plus précisément de la définition d'algorithmes responsable de la recherche des paramètres optimaux d'une transformation géométrique 3D.

Les algorithmes génétiques ont été choisis pour résoudre ce problème car ils sont bien adaptés aux larges espaces de recherche et parce qu'ils sont efficaces.

Une présentation générale du fonctionnement, ainsi qu'un exemple sur un problème concret simple et régulier ont été énoncés.

Par la suite, nous avons présenté deux instanciations d'algorithmes génétiques que nous avons utilisés dans le cadre, d'une part, du *recalage rigide* (recherche de 6 paramètres), et d'autre part, dans le cadre du *recalage élastique par appariement de points* (recherche de 8 couples de points).

Enfin, l'analyse des limites des AG nous a permis d'insister sur le fait que de tels algorithmes, pour être utilisés de manière optimale, devraient toujours être couplés avec un algorithme de recherche locale; autrement dit, on devrait toujours s'abstenir de ne prendre en compte *in fine* que le meilleur élément d'une population génétique comme *la* solution du problème à optimiser.

CHAPITRE 4 Approche hiérarchique robuste

Cette partie présente l'approche hiérarchique robuste mise en place pour effectuer le recalage élastique 3D à l'aide d'algorithmes génétiques.

La structure globale de l'algorithme de raffinement hiérarchique (de grossier à fin) ainsi que la philosophie de la méthode utilisée seront en premier lieu exprimées. Ensuite, nous détaillerons les trois étapes de la méthode, à savoir l'approximation globale (recalage rigide analytique), le recalage élastique structurel (appariement de points ou *point matching*) et enfin l'affinement local.

Pour finir, nous présenterons une analyse comparée de l'influence des paramètres des différentes étapes sur la robustesse de l'algorithme.

4.1 Structure hiérarchique

4.1.1 Pourquoi la structure hiérarchique

La mise en œuvre d'un algorithme de traitement d'image, et plus exactement le développement d'une méthode, s'effectue par un cheminement complexe. L'idée simple sur laquelle nous sommes partis initialement était seulement d'utiliser les bonnes propriétés de robustesse et de rapidité d'un algorithme génétique pour répondre à un problème qui de par nature était long à résoudre avec des méthodes classiques. Néanmoins, ces travaux [46] étaient d'une part restreints au recalage monomodalité, et d'autre part, la formulation du problème (recherche *directe* des 24 paramètres de la déformation élastique en explorant les déplacements possibles des 8 sommets du volume d'intérêt) n'exploitait pas du tout la topologie des images (description surfacique).

4.1.1.1 Passage de la monomodalité à la multimodalité

La fonction distance utilisée initialement utilisait directement les niveaux de gris des images. Le passage aux images multimodales pouvait se faire soit en utilisant des techniques comme la maximisation d'information mutuelle [22, 52, 97], ou par une transformation descriptive des images permettant de les comparer, quel que soit le type d'imageur qui les a créées. Si l'on s'en tient à utiliser des algorithmes génétiques, la première solution n'était pas envisageable car le calcul de l'information mutuelle est trop lourd vis-à-vis du grand nombre d'évaluations de performance exigé par les AG. Aussi, nous avons pensé extraire des informations communes entre les images, telle la surface air/peau. Le calcul de la fonction distance entre deux images (ou la fonction performance) se résumait alors au calcul de distance entre deux surfaces numériques de \mathbb{R}^3 , et pouvait être effectué rapidement selon la méthode expliquée au paragraphe 2.3.3 page 59.

4.1.1.2 Nouvelle formulation du problème

Puisque nous disposions alors d'images segmentées (cf. annexe B), la réflexion se porta sur l'exploitation d'une telle surface pour contraindre la recherche à quelque chose de pertinent. On n'utilise alors plus les sommets du volume d'intérêt, mais on cherche à effectuer un appariement entre 8 points d'une surface et 8 de l'autre. Cette idée est l'idée centrale actuellement utilisée, le reste de la méthode présentée ici ne constitue que des heuristiques d'accélération du traitement et d'amélioration de la robustesse. En effet, le nombre d'associations de 8 points d'une surface avec 8 points de l'autre surface (sachant que les surfaces des premières images que nous avons traitées font respectivement $n_a \approx 35000$ et $n_b \approx 65000$ points en IRM et en TDM) est d'environ :

$$C = \frac{(n_a.n_b)!}{p!(n_a.n_b - p)!} \approx 1,7.10^{70}, \text{ avec } p = 8.$$

Parmi ces C transformations, très peu correspondent à une transformation envisageable. En effet, la transformation que nous recherchons est faiblement élastique, ce qui revient à dire que chaque point ne peut se transformer qu'avec les points d'un certain voisinage (une trentaine environ). Le nombre N de transformations viables peut donc s'estimer par

$$N < \frac{(30.n)!}{p! (30.n-p)!} = \left(\frac{n!}{p!(n-p)!}\right) 30^p, \text{ avec } n = \max(n_a, n_b),$$

ce qui donne comme nombre maximal de transformations viables $N_{max} \approx 5, 18.10^{45}$. En terme de statistique, nous obtenons sur l'espace de recherche uniquement 2,91.10⁻²³ % de transformations viables. Évidemment aucun algorithme d'optimisation n'est capable de résoudre un tel problème.

4.1.1.3 Vers la réduction de l'espace de recherche

Plusieurs idées visant la réduction de l'espace de recherche ont été envisagées, comme par exemple la division de l'image 3D en sous-espaces (division en *octree*) où chaque point d'un sous-espace ne peut s'associer qu'avec des points d'un sous-espace homologue. Cette première idée est restée un échec car pas assez sélective, et posant des problèmes de définitions des sous-espaces duals. La deuxième idée exploitée était un calcul *a posteriori* d'élasticité de la transformation en considérant le déplacement induit par la partie élastique de la transformation (les coefficients en xy, xz, yz et xyz). Encore une fois, cette méthode n'a pas abouti car une décision *a posteriori* n'est pas un gage de performance, et de plus, cette méthode induisait la définition de seuils forfaitaires.

La première solution retenue a été celle de l'utilisation d'une classification en géométrie différentielle des points de la surface. Puisque l'on dispose de surfaces numériques, moyennant une résolution des images suffisantes, un calcul de courbure locale (min, max ou Gaussienne) peut être effectué en chaque point. Aussi, une classification des points en fonction par exemple du signe des courbures peut être entreprise. En posant l'hypothèse assez forte, par exemple, qu'un point d'un certain signe de courbure d'une image doit correspondre à un point de même signe de courbure sur l'autre image¹, l'espace de recherche se trouve considérablement réduit. La définition d'un nouvel espace de recherche (que l'on appellera par la suite *espace couple*) où chaque élément correspond à un couple de points que l'on envisage d'apparier, nous a amené à la mise en place d'une nouvelle contrainte topologique : un point ne peut désormais s'associer qu'avec des points géométriquement proches. Aussi, une estimation d'un recalage rigide initial a été nécessaire pour définir une estimée du transformé d'un point, et également son voisinage.

Enfin, la dernière amélioration apportée à notre algorithme se situe au niveau de l'exploitation de la convergence des algorithmes génétiques. Comme il est nécessaire de coupler un algorithme génétique avec un algorithme de recherche locale (pour optimiser la vitesse et la qualité de convergence), nous avons eût l'idée d'utiliser l'ensemble de l'information génétique (tous les éléments de la population finale) plutôt qu'un élément isolé (le meilleur individu). En regroupant ainsi l'information génétique, nous opérons rapidement une optimisation locale.

4.1.1.4 Structure hiérarchique

Pour résumer, l'algorithme que nous allons détailler est divisé en trois étapes :

- 1. Estimation génétique d'un recalage rigide et d'un coefficient de confiance (recalage rigide);
- 2. Estimation génétique des associations possibles entre 8 points d'une surface et 8 points de l'autre (échantillonnage stochastique robuste de l'espace de recherche);
- 3. Utilisation de l'information génétique pour en déduire les 24 coefficients de la transformation globale élastique (affinement local).

¹Hypothèse seulement réalisable avec des transformations polynomiales faiblement élastiques.

La structure hiérarchique de l'algorithme apparaît donc à trois niveaux. Tout d'abord elle découle de l'enchaînement d'étapes fournissant des solutions de plus en plus précises (affinement hiérarchique grossier vers fin). Ensuite elle apparaît dans l'utilisation à différents niveaux des structures de l'image (surfaces, puis couples de points, puis voxels). Enfin on trouve une structure hiérarchique à propos du nombre de degrés de liberté de chaque étape (6 pour le recalage rigide, 8 pour le recalage sur l'espace-couple, et enfin 24 pour l'affinement local).

4.1.2 Les différentes étapes



Figure 4.1 — Structure hiérarchique de l'algorithme mis en œuvre.

La figure 4.1 montre la structure générale de l'algorithme employé depuis l'acquisition des données, la segmentation des données par un algorithme de Ligne de Partage des Eaux (cf. annexe B), la création éventuelle de classes de courbure locale, et l'enchaînement des trois étapes évoquées précédemment jusqu'à l'obtention des 24 coefficients de la transformation élastique.

Nous allons détailler maintenant plus précisément le fonctionnement de chacune des 3 étapes de la structure hiérarchique. Tout au long des explications nous verrons sur un exemple

d'images médicales comment s'améliore, après chaque étape, la qualité du recalage obtenu. Les images qui nous serviront de support sont des images issues du laboratoire TIMC de Grenoble et ont été mises à disposition par Stéphane Lavallée.

Les images du TIMC ne sont pas les seules images qui nous ont servi de support. Nous avons également eût à notre disposition des images en provenance de l'université de Clermont-Ferrand. Ces images sont celles qui ont été utilisées par A. Colin [19], et nous ont été prêtées par J.-Y. Boire. Malheureusement, la nature de ces images (Scanner, et IRM séquence Flash), et notamment la composition de l'image IRM nous ont posé des problèmes de segmentation et ont rendu le recalage 3D très difficile, voire impossible (voir figure 4.2).



Figure 4.2 — Images Scanner et IRM séquence flash mises à disposition par l'université de Clermont-Ferrand. Les artéfacts de reconstruction présents sur les images IRM posent des problème de segmentation, et par voie de conséquence rendent le recalage difficile.

Les images du TIMC que nous avons utilisées sont un volume TDM et un volume IRM (cf. annexe A pour plus de précisions sur ces types d'images médicales) de la tête d'une même personne. Le scanner (TDM) a pour dimension $(256 \times 256 \times 91)$ voxels alors que l'IRM ne fait que $(256 \times 256 \times 68)$ voxels. Il est intéressant de noter que les voxels envisagés ne sont pas cubiques et possèdent une taille physique différente pour les deux volumes. L'extension physique de l'image TDM est de $232, 82 \times 232, 82 \times 165, 81 \text{ mm}^3$, ce qui fait une résolution effective de $0,91 \times 0,91 \times 1,82 \text{ mm}^3$, alors que celle du volume IRM est de 265, 45 \times 265, 45 \times $141,06 \text{ mm}^3$, d'où une résolution de $1,04 \times 1,04 \times 2,07 \text{ mm}^3$. Les algorithmes que nous avons développés sont donc capables de traiter des images avec de faibles anisotropies, ce qui nous évite d'effectuer une étape de rééchantillonnage visant à rendre les voxels utilisés cubiques. La figure 4.16 de la page 106 (les figures à propos des images médicales sont regroupées à la fin de ce chapitre pour faciliter la comparaison des résultats aux différentes étapes de l'algorithme) illustre les données utilisées avant tout recalage. Notons au passage que les artéfacts volumiques à l'arrière de l'image scanner font partie de l'image brute. Ces volumes auraient pu être retirés par un traitement spécial, mais nous les avons gardés de manière à tester la robustesse de l'algorithme. Évidemment comme il n'y a aucune raison pour que la 30^e coupe des 91 de l'image TDM corresponde avec la 30^e coupe des 61 de l'image IRM, les données n'y sont pas recalées. Notons qu'une fois les images recalées, l'image subissant la déformation (ici ce sera l'image IRM) est reconstruite dans la géométrie de l'image cible, et par conséquent la i^{e} coupe de la cible doit correspondre parfaitement à la i^{e} coupe de l'image recalée.

Les rendus de la figure 4.16 ont été réalisés à l'aide de l'outil de rendu direct multivolume (voir annexe C). Cet outil permet notamment d'afficher une union et une intersection floue des deux volumes en présence. Enfin, sur le bas de l'image, nous avons représenté deux vues en coupe (30^e coupe) du volume TDM et IRM, ainsi qu'une superposition (au centre) de la coupe IRM (en niveaux de gris), et de l'os cortical segmenté par seuillage de la coupe TDM (affiché en vert).

4.2 Étape 1 : Approximation globale (recalage rigide)

4.2.1 Description

Comme nous venons de l'expliquer précédemment, la première phase de l'algorithme correspond à la création d'un espace de recherche pertinent pour l'appariement de points. On se propose de rechercher directement, en utilisant un algorithme génétique fonctionnant sur les réels, 6 paramètres d'une transformation rigide (déplacement de \mathbb{R}^3) minimisant la distance stochastique entre nos deux surfaces segmentées. Nous faisons l'hypothèse que l'élasticité de la transformation que nous cherchons est faible puisque la structure osseuse de la boîte crânienne est rigide, et aussi parce que les distorsions induites par les appareils imageurs ne sont pas très violentes. Ainsi, nous pensons qu'un recalage rigide même approximatif, est une bonne estimation de la déformation élastique recherchée. Une fois donnée une telle estimation, et connaissant un coefficient de confiance (inversement proportionnel à la distance médiane entre les surfaces recalées rigidement), on peut créer une liste de couples de points susceptibles d'être appariés, et ainsi éliminer de l'espace de recherche les associations aberrantes de points.

La figure 4.3 illustre l'algorithme génétique mis en œuvre pour la recherche des 6 coefficients du recalage rigide.



Figure 4.3 — Principe de fonctionnement du recalage rigide par algorithme génétique.

4.2.2 Mise en œuvre

Avec les images de la figure 4.16, cette étape de recalage rigide s'applique en une vingtaine de secondes². Les paramètres que nous avons utilisés pour cette étape de recalage rigide, ainsi que les résultats en terme de performance sont établis par le tableau 4.1; les rendus obtenus avec un tel recalage rigide sont illustrés à la figure 4.17 reportée à la fin de ce chapitre (page 106) avec les rendus des autres recalages effectués.

Étape 1 : Recalage Rigide			
Paramètres	Nombre de chromosomes	500	
	Nombre d'itérations	100	
	Probabilité de mutation	1 %	
	Probabilité de recombinaison	80~%	
	Dynamique translation autorisée	$\pm 60 \text{ mm}$	
	Dynamique rotation	$\pm 60 \deg$	
	Nombre d'intervalles (carrés latins)	10	
	Nombre minimal de représentants par intervalle	5	
	Lissage distance (σ)	4	
	Points pris en compte pour la distance stoch. (n)	100	
Résultats	Temps d'exécution	$16s (\pm 1 s)$	
	Performance stochastique (%)	$74,909~(\pm 3,6)$	
	Translation rigide trouvée (mm)	[11, 30, 22, 49, -51, 86]	
	Angles de rotation trouvés (radians)	$\left[0, 135, 0, 093, -0, 072 ight]$	
	Référence de la figure de résultats	4.17	

Tableau 4.1 — Paramètres utilisés et résultats du recalage rigide.

Les résultats chiffrés présentés au tableau 4.1 sont liés à une observation moyenne sur une vingtaine de recalages rigides des mêmes images. Le temps affiché (16s) est par conséquent un temps moyen avec un écart-type d'environ 1 seconde. Le temps d'exécution varie d'un recalage à un autre car le nombre d'opérations nécessaires est aléatoire (du fait du fonctionnement même des algorithmes génétiques). De même, la performance stochastique finale observée varie d'un recalage à un autre. Une moyenne de 75 % ne signifie pas que 75 % des points de la surface à recaler se transforment parfaitement sur la surface de référence; ce chiffre est juste représentatif de la proximité moyenne de ces points transformés à la surface de référence. Bien évidemment, une performance de 100 %, qui est la performance théorique maximale, signifierait que l'on a trouvé une transformation géométrique faisant correspondre parfaitement chaque point de la surface segmentée à recaler avec la surface de référence. De la même manière, une performance de 0~% signifierait tout simplement que chacun de ces points se transforment à une distance relativement élevée de cette surface de référence. Il est bien entendu difficile d'apprécier la qualité du recalage seulement à partir de la donnée de cette performance stochastique. Cette performance est notamment sujette à l'existence de parties manquantes, et ne peut atteindre 100 % qu'uniquement lorsque les surfaces à mettre en correspondance possèdent rigoureusement la même extension géométrique, ce qui est rarement le cas. Pour cela, nous proposons une validation visuelle coupe à coupe et par DMVR (cf. annexe C) de ces volumes (voir figure 4.17, page 106).

 $^{^2 {\}rm Le}$ PC utilisé est un bi-Pentium Pro cadencé à 233 MHz, ayant 192 M
o de mémoire et tournant sous Linux 2.2.

4.3 Étape 2 : Recalage élastique (appariement de points)

4.3.1 Description

Les 24 paramètres de la transformation globale trilinéaire que nous recherchons sont parfaitement définis par la connaissance de 8 correspondances entre points des deux surfaces numériques à recaler. Aussi nous avons mis en œuvre un algorithme d'appariement de points (plus communément désigné avec le terme anglais *point-matching*) cherchant les correspondances possibles 8 à 8 entre les points des deux surfaces. L'optimisation étant effectuée sur la base d'un algorithme génétique fonctionnant sur un espace d'index, le résultat de l'algorithme est en fait une population génétique. Cette population génétique peut être interprétée comme une liste de 8-uplets intéressants et constituant un nouvel espace de recherche (pour un algorithme de recherche locale par exemple). Aussi, nous pouvons dire que ce deuxième algorithme génétique fonctionne comme un échantillonneur robuste de l'espace de recherche.

4.3.1.1 Critères de création d'un couple

La construction de la liste des couples susceptibles d'être appariés (\mathcal{L}) se fait à partir de la meilleure transformation rigide T^r trouvée par l'algorithme génétique de la première étape, ainsi que de la classification des points de S_1 et S_2 par géométrie différentielle.

En premier lieu, nous essayons de déterminer la distribution des distances entre les points de S_1 transformés par T^r et la surface S_2 . Cette distribution étant sujette à la présence de points sans homologues, nous déterminons en second lieu la valeur médiane de la distribution, que nous notons d_{md} .

On pose alors la règle suivante : un point p_1 de S_1 et un point p_2 de S_2 peuvent constituer un couple si et seulement si p_1 et p_2 appartiennent à la même classe de courbure et si de plus la distance euclidienne séparant le transformé de p_1 par T^r et p_2 est inférieure à 2 d_{md} .

Notons que l'on a choisi le double de la distance médiane comme étant un bon compromis entre la rapidité de l'algorithme (plus on est sélectif, plus on réduit l'espace de recherche et plus ce dernier est rapide), et la tolérance à l'élasticité (plus on autorise l'association de points éloignés vis-à-vis de la transformation rigide, et plus on favorise les transformations élastiques).

L'ensemble des couples possibles est alors décrit par la relation :

$$\mathcal{L} = \Big\{ (p_1, p_2) \in S_1 \times S_2, \mathcal{C}(p_1) = \mathcal{C}(p_2), \text{ et } d^{\star}(T^r(p_1), p_2) < 2d_{md} \Big\}.$$

La figure 4.4 montre un exemple 2D d'appariement de points selon une transformation rigide T^r , une distance moyenne d_{md} et une partition des courbures en deux classes (courbure 2D à signe positif ou négatif). Sur cet exemple, le point p_1 peut s'associer, en terme de distance, qu'avec les points c, d, e et f, mais la contrainte de courbure, rejette les deux derniers points. Finalement p_1 ne peut s'associer qu'avec c ou d, et donc seulement deux couples dans la liste \mathcal{L} feront référence à p_1 .

Notons au passage, que l'utilisation des propriétés de courbures locales de la surface en chaque point n'est pas obligatoire (on peut décider qu'il n'existe qu'une classe de points). Néanmoins, tant qu'il est possible d'utiliser de l'information a priori sur les images, et ainsi restreindre de manière intelligente l'espace de recherche, on s'y attache.



Figure 4.4 — Principe du choix des couples de points.

4.3.1.2 Détermination de la courbure locale

La détermination des propriétés de courbure locale de la surface est effectuée selon les considérations suivantes : étant donnée une surface V de \mathbb{R}^3 , en chaque point $p \in V$ on peut calculer le plan tangent T_pV composé de l'ensemble des tangentes à V en p. On prendra comme base orthogonale de T_pV une base $(\overrightarrow{r_u}, \overrightarrow{r_v})$. Soit \overrightarrow{n} le vecteur normal à V en p,

$$\overrightarrow{n} = \frac{\overrightarrow{r_u} \wedge \overrightarrow{r_v}}{||\overrightarrow{r_u} \wedge \overrightarrow{r_v}||}.$$

Un vecteur tangent t sur V permet alors de déterminer de façon unique un plan Π_t qui définit une courbe gauche unique sur V, $\Pi_t \cap V$, appelée section normale, et notée \mathcal{L}_t (cf. figure 4.5).



Figure 4.5 — Section normale.

On peut alors chercher la courbure relative de cette courbe au point p, ce qui définit une fonction $\overrightarrow{t} \mapsto k(\overrightarrow{t})$ sur les vecteurs de base de l'espace tangent T_pV . Sans rentrer dans les détails, $k(\overrightarrow{dr})$ ($\overrightarrow{dr} \in T_pV$) est aussi le rapport de la deuxième forme fondamentale sur la première forme fondamentale (pour plus de détails voir, par exemple, la référence [76]). Remarquons simplement qu'en chaque point p de V, la fonction $k(\overrightarrow{dr})$ ($\overrightarrow{dr} \in T_pV$) prend une valeur maximale et minimale. On note k_1 la valeur la plus grande en valeur absolue et k_2 la valeur la plus petite en valeur absolue. Leurs signes indiquent le sens de la courbure par rapport au repère lié à l'élément de surface, c'est-à-dire le plan tangent et la normale. GAUSS posa $K = k_1k_2$ et $H = (k_1 + k_2)/2$. La courbure totale K aussi communément appelée courbure de GAUSS est une valeur intrinsèque à la surface (elle ne dépend pas de la paramétrisation r = r(u, v) choisie) et est invariante par isométrie.

La somme des courbures est une valeur extrinsèque à la surface, elle en donne sa courbure moyenne « vue de l'extérieur ». Contrairement à K, H n'est pas invariant pour toutes les isométries mais seulement pour les translations et les rotations³.

4.3.1.3 Classification en fonction de la courbure

L'établissement de classes de courbures s'effectue en fonction des valeurs, ou seulement des signes des réels H et K. En général, nous n'avons utilisé que 3 classes de courbures. Ces trois classes sont définies seulement à partir de statistiques sur les courbures gaussiennes calculées. Posons μ_q la moyenne des courbures gaussiennes, et σ_q l'écart-type.

Deux premières classes ont été définies pour les points dont la courbure $K < \mu_g - \sigma_g/3$ et pour les points où $K > \mu_g + \sigma_g/3$. Une troisième classes est définie pour les points où $|K - \mu_g| < \sigma_g/40$. Les autres points sont éliminés⁴. Avec ces valeurs de seuils, arbitraires certes, nous restreignons considérablement l'espace de recherche, et par là même accélérons le temps de traitement. Notons que l'on aurait pu utiliser d'autres critères de classification en séparant explicitement par exemple les points à courbure gaussienne positive des points à courbure gaussienne négative, puis les points à courbure moyenne (H) positive des points à courbure moyenne négative⁵.

4.3.1.4 Algorithme de création de la liste

La création de la liste des couples intéressants \mathcal{L} aurait pu se faire, de manière simple, selon l'algorithme suivant :

 $^{^{3}}$ En d'autres termes, H est invariant par isométrie n'inversant pas l'orientation d'une base de vecteurs.

⁴Ce qui fait par conséquent un total de 4 classes de courbures en comptant la classe rejet. Néanmoins, seulement 3 classes de points seront utilisées par la suite.

⁵Ceci peut être réalisé en posant par exemple $\mu^- = \mathbb{E}(\{K_i, K_i < 0\})$ et $\sigma^- = \mathbb{E}(\{(K_i - \mu^-)^2, K_i < 0\})$ (et équivalents pour μ^+ et σ^+); deux classes sont extraites : celle satisfaisant $K < \mu^- - \lambda \sigma^-$; celle satisfaisant $K > \mu^+ + \lambda \sigma^+$; chacune de ces deux classes pouvant ensuite être dérivée en deux sous-classes en fonction du signe de H.

Création de la liste de couples \mathcal{L} $\mathcal{L} = \{\emptyset\}$ Pour chaque point P_i de S_1 Faire Pour chaque point Q_j de S_2 Faire Si $\mathcal{C}(P_i) = \mathcal{C}(Q_j)$ et $d^*(T^r(P_i), Q_j) < 2d_{md}$ Alors $\mathcal{L} \leftarrow \mathcal{L} \cup \{(P_i, Q_j)\}$ Fin Si Fin Pour Fin Pour

Cet algorithme implique le déroulement de deux boucles imbriquées et est, par conséquent, relativement lent lorsque le nombre de points des deux surfaces S_1 et S_2 est élevé. Un ordonnancement des points de chaque surface selon leur coordonnée z, et la détermination d'un tableau de z_{min} et de z_{max} d'index de S_2 , et ce pour chaque classe de courbures \mathcal{C} permet d'accélerer grandement la création de la liste de couples \mathcal{L} . En d'autres termes, à partir d'un point p_1 de S_1 , se transformant en p'_1 sur D_2 , on regarde l'altitude z'_1 de p'_1 ; deux tableaux ordonnés de correspondances entre les index des points de S_2 et leur altitude z nous permet de savoir entre quels index j_{min} et j_{max} de S_2 les points auront une altitude z comprise entre $z'_1 - 2d_{md}$ et $z'_1 + 2d_{md}$ (voir figure 4.6). Le tri des index étant très rapide, le principal gain de cette méthode consiste à ne parcourir pour chaque point de S_1 qu'une partie (en moyenne un huitième sur les images que nous avons utilisées) de S_2 . L'accélération de la méthode se situe par conséquent dans un rapport un pour huit.



 $Figure \ 4.6$ — Principe d'accélération du parcours des listes de points pour la création de l'ensemble des couples \mathcal{L} .

Notons enfin que l'algorithme de création des couples aurait encore pu être accéléré en profitant de l'architecture multiprocesseur de la machine sur laquelle est exécuté le recalage. Le parcours de la liste des points de S_1 peut se diviser en n (autant que de processeurs) sous-ensembles traités en parallèle, et ainsi créer n sous-listes de couples qu'il ne restera plus qu'à fusionner pour créer \mathcal{L} .

4.3.1.5 Utilisation de l'algorithme génétique

La figure 4.7 illustre l'algorithme génétique opérant l'échantillonnage stochastique de l'espace de recherche. Il est important, en effet, de comprendre que ce deuxième algorithme génétique n'a pas pour but de trouver une fois pour toutes les 8 meilleurs associations de points possibles et par conséquent les 24 paramètres d'une transformation élastique globale, mais plutôt d'explorer les possibilités d'associations et de fournir *in fine* une liste de 8-uplets susceptibles de donner de bons résultats. Aussi cette étape n'a de sens que si l'on considère qu'elle sera couplée avec un algorithme d'optimisation utilisant l'information génétique obtenue.



 $\label{eq:Figure 4.7} \emph{Figure 4.7} \emph{Principe de fonctionnement de l'échantillonneur stochastique de l'espace de recherche.}$

4.3.2 Mise en œuvre

L'établissement du recalage rigide (première étape) nous permet de déterminer l'intervalle de confiance pour la création des couples $(2d_{md})$. Dans le cas présent, nous obtenons $2d_{md} =$ 2.6mm. Avec les 35000 points de la surface IRM et les 63000 points de la surface TDM, et après environ deux minutes de recherche, plus de trois cent mille couples satisfont les critères d'association (si l'on n'utilise pas de critères de courbure, donc avec une seule classe C); néanmoins, si nous utilisons les trois classes de courbures définies précédemment, il ne reste plus que cinquante mille couples créés en une quarantaine de secondes. Une fois la liste \mathcal{L} de couples créée, on applique notre échantillonneur stochastique robuste (deuxième algorithme génétique). Les paramètres utilisés et les résultats obtenus lors de cette deuxième étape sont résumés par le tableau 4.2, et les rendus volumiques correspondant à la meilleure solution génétique de cette étape sont appréciables à la figure 4.18 page 107.

Étape 2 : Recalage élastique			
Paramètres	Nombre de chromosomes	500	
	Nombre d'itérations	100	
	Probabilité de mutation	1 %	
	Probabilité de recombinaison	$80 \ \%$	
	Points pris en compte pour la distance stoch. (n)	200	
Une seule classe de courbures			
Résultats	Création des couples	$2 \mathrm{mn} \ 44 \mathrm{s}$	
	Temps d'exécution	51s	
	Performance stochastique $(\%)$	$79,545~(\pm 1,191)$	
	Trois classes de courbures		
	Création des couples	$46\mathrm{s}$	
	Temps d'exécution	48s	
	Performance stochastique $(\%)$	$79,\!126~(\pm 2,526)$	
	Référence de la figure de résultats	4.18	

Tableau 4.2 — Paramètres utilisés et résultats du recalage élastique.

Les résultats résumés par le tableau 4.2 sont représentatifs d'une dizaine de recalages élastiques effectués sur la même paire d'images médicales. Là encore, les résultats en terme de performance stochastique moyenne sont difficilement interprétables directement. Un résultat intéressant à constater se situe au niveau des différences de temps de calcul selon que l'on utilise où non des critères de courbure locale. La prise en compte d'une classification selon la courbure gaussienne locale permet de réduire considérablement l'espace de recherche, ce qui permet, outre la création plus rapide de la liste \mathcal{L} des couples possibles comme évoqué plus haut, une convergence plus rapide de l'algorithme génétique (on divise environ par trois le temps de calcul global). La légère différence de performance stochastique (0.4 %) selon que l'on utilise ou non cette classification en géométrie différentielle ne s'observe pas sur les rendus volumiques (voir figure 4.18 page 107), aussi nous n'avons présenté que les rendus du recalage avec 3 classes de courbures. Enfin nous constatons sur ce tableau de résultats que la variabilité de la performance du recalage avec 3 classes de courbures est plus grande que lorsque l'on n'utilise pas d'information de courbure. Ce résultat indique vraisemblablement une convergence prématurée de l'algorithme génétique vers divers minima locaux. La taille de l'espace de recherche étant beaucoup plus petite, l'apparition d'une bonne solution est plus rapide et peut, parfois, devenir prépondérante. Modérons toutefois notre propos en observant

qu'une différence d'un ou deux pour-cent en terme de performance stochastique, ne correspond pas à une amélioration ou une détérioration très sensible de la qualité du recalage visuel.

4.4 Étape 3 : Affinement local (post-analyse)

4.4.1 Description

Utiliser directement le meilleur élément (*i.e.* chromosome) de la population génétique finale est un choix arbitraire qui n'est certainement pas optimal. En effet, en terme d'algorithmes génétiques il n'y a pas d'évolution continue et lisse des points de recherche vers les optima locaux. Aussi, rien ne garantit qu'un algorithme génétique, utilisant une fonction de performance stochastique, pris à un instant t ait comme meilleur élément le meilleur des éléments rencontrés aux instants antérieurs à t. Autrement dit, l'évolution de la performance des meilleurs éléments en fonction du temps (ou des générations génétiques), n'a pas de raison d'être monotone (ce qui serait le cas avec une sélection génétique de type élitiste et une fonction de performance déterministe). Aussi, il est absurde de penser que le meilleur chromosome de la dernière itération génétique donne une solution stable. Ce chromosome ne peut donner qu'une estimation, certes bonne, d'une solution fortement probable ; par conséquent, le meilleur moyen d'extraire au maximum l'information donnée par un algorithme génétique est d'utiliser le contenu de tous les chromosomes de la dernière itération (voir figure 4.8).



Solution male

Figure 4.8 — L'algorithme génétique vu comme un échantillonneur stochastique de l'espace de recherche, couplé à une procédure d'optimisation locale utilisant l'ensemble de l'information génétique.

La méthode que nous avons mise au point pour extraire au mieux l'ensemble de l'information génétique contenu dans la population doit être considérée par ailleurs comme une optimisation locale déterministe. Nous savons que huit est le nombre exact de couples de points nécessaires au calcul direct des 24 paramètres de la transformation trilinéaire globale que nous cherchons. Néanmoins, avec un nombre de couples plus élevé, il est toujours possible de calculer un jeu de 24 paramètres satisfaisant au mieux les conditions imposées par l'ensemble des $N_c > 8$ couples. La seule différence est qu'utiliser 8 couples revient à calculer une interpolation, alors qu'utiliser plus de couples revient à effectuer une approximation. Néanmoins, il apparaît clairement que l'utilisation d'une méthode d'approximation plutôt que d'interpolation permet de réduire le bruit positionnel inhérent à la qualité de la segmentation des surfaces, tout en permettant implicitement d'utiliser facilement l'ensemble de l'information génétique de la dernière itération. Cette assertion est en grande partie responsable de la robustesse de notre algorithme de recalage.

Avant de détailler plus amplement la manière dont est effectuée l'exploration de la population finale, nous allons expliciter la méthode utilisée pour résoudre l'approximation des 24 paramètres avec un nombre de couples $N_c > 8$.

4.4.1.1 Résolution de systèmes surdéterminés

Lorsque nous voulons calculer les 24 paramètres $a_{i,j,k}$, $b_{i,j,k}$, $c_{i,j,k}$ (i, j, k prenant les valeurs 0 ou 1) de la transformation trilinéaire élastique à partir de 8 couples, nous avons à résoudre 3 systèmes linéaires de 8 équations à 8 inconnues (voir chapitre 2, section 2.1.4). Lorsque nous disposons d'un nombre de couples $N_c > 8$, nous devons résoudre 3 systèmes (un pour chaque coefficient a, b, c) de N_c équations à 8 inconnues. Pour la résolution de tels systèmes surdéterminés au sens des moindres carrés, on utilise une méthode peu sensible au mauvais conditionnement du système (par exemple lorsque la configuration des points rend imprécise la connaissance de la solution suivant un des axes de l'espace de recherche) comme celle de la décomposition en valeurs singulières⁶. Prenons l'exemple de la résolution du système $X'_{[N_c \times 1]} = M_{[N_c \times 8]} \cdot A_{[8 \times 1]}$ où les inconnues sont les coefficients du vecteur A.

Si $N_c = 8$, le système n'est pas surdéterminé, mais possède une solution unique (dans le cas d'un déterminant non nul bien entendu), et nous pouvons le résoudre avec la méthode du pivot de GAUSS par exemple. Par contre dès que $N_c > 8$, cette méthode n'est plus applicable. Nous utilisons une décomposition de la matrice $M_{[N_c \times 8]}$ selon ses valeurs singulières.

Toute matrice A de taille $[M \times N]$ dont le nombre de lignes M est supérieur ou égal au nombre de colonnes N, peut être écrite comme le produit d'une matrice U de taille $[M \times N]$ et orthogonale selon les colonnes, d'une matrice diagonale W de taille $[N \times N]$ dont les éléments sont positifs ou nuls (les valeurs singulières), et la transposée d'une matrice V orthogonale et de taille $[N \times N]$.

$A = U.W.V^T$

Les matrices U et V sont orthogonales dans le sens où leurs vecteurs colonnes sont orthonormaux.

$$U^T.U = V^T.V = Id_{[N \times N]}$$

Par ailleurs, notons W^{-1} la matrice diag $(1/w_1, 1/w_2, \ldots, 1/w_n)$. Notons que si une valeur singulière est nulle nous l'écartons purement et simplement du calcul puisqu'elle est alors représentative d'une solution dégénérée qui ne nous intéresse pas. À ce moment là nous avons $W.W^{-1} = Id_{[N \times N]}$.

Remarquons au passage que puisque V est une matrice carrée, on a aussi $V.V^T = V^T.V \propto Id_{[N \times N]}$. Dans notre cas, nous avons toujours $M \ge N$.

 $^{^{6}}$ Les considérations mathématiques qui vont suivre sont extraites de [74], Chi-square fitting (p.659), SVD decomposition (p.59), General least square (pp.671–676)

Pour résoudre le système suivant $b_{[M \times 1]} = A_{[M \times N]} \cdot x_{[N \times 1]}$ nous procédons de la manière suivante :

En résumé, on a l'équivalence suivante :

$$b_{[M\times 1]} = A_{[M\times N]} \cdot x_{[N\times 1]} \iff x = V \cdot W^{-1} \cdot U^T \cdot b,$$

donc la solution aux moindres carrés est $x = (V.W^{-1}.U^T).b$, et l'erreur résultante est mesurée par ||A.x - b||.

4.4.1.2 Stratégie d'exploration de la population génétique finale

Une fois connue la méthode de résolution de systèmes surdéterminés et par conséquent une fois rendue possible la prise en compte d'un ensemble de $N_c > 8$ couples de points, il reste à adopter une stratégie d'exploration (sélection) des couples intéressants.

La population résultante après un cycle génétique est en général formée de plusieurs « amas » autour des différentes solutions locales (et de la solution globale) du problème (voir figure 4.8). Afin d'optimiser encore la recherche de la solution globale, il serait intéressant de ne prendre en compte pour le calcul de la solution optimale, que les chromosomes faisant partie de *la classe amassée autour de la solution globale*. Bien sûr cela n'est pas directement possible, puisque l'on ne connaît pas a priori la solution globale et donc on ne peut pas décider quel amas est représentatif de cette classe. C'est pour cela que nous proposons une méthode adaptative de recherche de la sous-population donnant la meilleure solution.

À l'origine, nous pensions utiliser un algorithme tel celui des *nuées dynamiques* pour effectuer l'extraction de la classe principale de chromosomes. Le problème est que cet algorithme impose de savoir a priori combien de classes sont présentes dans la population. Or ce n'est pas du tout le cas. Si l'algorithme génétique converge fortement, nous n'aurons qu'une seule classe dans la population finale, et s'il converge faiblement, nous en aurons autant qu'il y a de minima locaux. Cet algorithme ne pourra donc pas être adopté, et c'est pour cela que nous proposons une méthode itérative et adaptative de recherche de la sous-population représentant la classe principale⁷.

Le principe général de l'algorithme mis en œuvre, que nous appellerons de par sa position finale dans l'organigramme général de la procédure de recalage, algorithme de *post-analyse*, est décrit par la figure 4.9. À partir de la population finale de chromosomes, on commence par extraire des chromosomes la liste \mathcal{L}^* de couples qui y sont utilisés (il y a 8 couples par chromosome) en enlevant les doublons. Partant des 8 couples du meilleur chromosome,

⁷Nous avions examiné une autre méthode de recherche de sous-population représentant la classe principale fondée sur la *méthode des hiérarchies*, mais les résultats n'étaient pas stables. Il s'agissait de partir du meilleur élément de la population, et de créer la sous-population en lui adjoignant au fur et à mesure les éléments les plus proches (au sens de la distance euclidienne dans \mathbb{R}^{24}) si ceux-ci permettaient d'augmenter la performance générale. Malheureusement on se retrouvait finalement souvent avec une sous-population composée d'un seul voire de deux éléments seulement. Notons quand même que l'algorithme que l'on a finalement retenu est fondé sur une approche qui s'apparente à celles de la famille des statistiques robustes (SR) fonctionnant en tant que classifieur. La procédure peut en effet être appliquée sur les chromosomes rejetés après la première classification et trouver ainsi de suite les amas de chromosomes autour des minima locaux.
on regarde itérativement dans \mathcal{L}^* si l'adjonction de nouveaux couples permet d'accroître la performance générale, et ce jusqu'à ce que l'adjonction d'aucun autre couple ne puisse améliorer la solution actuelle. Étant donné que l'ordre dans lequel sont rangés les couples dans la liste \mathcal{L}^* peut influer sur le résultat, il est parfois possible d'améliorer le résultat global en essayant de supprimer quelques éléments de la liste des couples retenus, ce qui est effectué lors d'une deuxième passe. L'exécution de ces deux passes permet d'exploiter pleinement et rapidement (typiquement quelques secondes), l'ensemble de l'information génétique contenu dans la population finale. Par ailleurs, ces deux passes peuvent s'inscrire dans une boucle (répétition des deux passes d'ajout et de rejet), jusqu'à ce que la performance n'augmente plus significativement.



Figure 4.9 — Affinement de la solution finale (post-analyse).

L'exécution en synergie des trois étapes (recalage rigide, échantillonnage stochastique de l'espace de recherche du point-matching, et optimisation locale) constitue la clé de la robustesse et de l'efficacité de notre algorithme d'optimisation.

4.4.2 Mise en œuvre

L'application de l'optimisation locale sur les chromosomes de la population finale permet d'augmenter, à faible coût opératoire, la performance finale de l'algorithme. L'espace de recherche est constitué des 500 chromosomes de la population génétique finale. Sur ces 500 chromosomes (donc $500 \times 8 = 4000$ couples), on extrait les couples différents. Sur les 4000 couples, et avec nos images médicales considérés, environ 140 couples différents⁸ sont présents dans la dernière population. Le traitement de ces couples s'effectue en à peu près une seconde, et permet de dégager un groupe d'une quinzaine de couples dont la performance globale est 6 à 7 % meilleure (passage de 75 % à 80 % environ). Les figures 4.19 et 4.20 présentent les rendus finaux obtenus sur les images médicales du laboratoire TIMC. Par ailleurs le tableau 4.3 résume les différents résultats obtenus tout au long de l'algorithme de recalage. Les valeurs statistiques de ce tableau ont été établies à partir de 20 recalages des mêmes images avec les mêmes paramètres. D'autres statistiques (voir paragraphe 4.5) ont été menées pour étudier le comportement de l'algorithme en fonction de différentes valeurs des paramètres des algorithmes génétiques.

Opération	Temps CPU	Performance (%)	Figure
Segmentation	< 2 mn	-	4.16
Recalage Rig. (1)	16 s	$74,909~(\pm 3,6)$	4.17
Une	e seule classe de	courbures	
Création couples	2 mn 44 s	-	4.4
Élastique (meilleur)	$51~{ m s}$	$79,545~(\pm 1,191)$	-
Optim. locale	1 s	$80,810\ (\pm 0,872)$	4.19
Trait. Global	$3 \mathrm{~mn} ~ 36 \mathrm{~s}$	80,810	4.19
Av	m ec~3~classes~de	courbures	
Création couples	46 s	-	4.4
Élastique (meilleur)	48 s	$79,\!126~(\pm 2,526)$	4.18
Optim. locale	1 s	$81,117 (\pm 2,600)$	4.20
Trait. Global	$1 \mathrm{~mn} ~ 35 \mathrm{~s}$	81,117	4.20

Tableau 4.3 — Résultats généraux du recalage élastique sur les images du laboratoire TIMC.

Notons, par rapport aux résultats présentés dans ce tableau et par rapport aux rendus volumiques et coupe à coupe proposés à la fin de ce chapitre (pages 106 à 108), que nous observons bien une différence de qualité entre le recalage rigide et le recalage élastique (avec ou sans post-analyse), mais que nous observons à peine la différence (il faudrait sans doute l'avis d'un expert) entre les recalages avec ou sans post-analyse, et encore moins entre les recalages avec ou sans classification en géométrie différentielle. Néanmoins, il reste irréfutable que d'une part la classification en géométrie différentielle permet d'améliorer de manière notoire le temps de calcul nécessaire, et que d'autre part l'étape de post-analyse (étape 3) permet d'améliorer, à faible coût opératoire, les résultats en terme de performance stochastique.

4.5 Influence comparée des différentes étapes

Une analyse détaillée de l'enchaînement de ces trois étapes menant à une réflexion sur le choix des paramètres des divers algorithmes utilisés a été effectuée.

Comme nous venons de le voir, les trois étapes que nous envisageons sont deux algorithmes génétiques constituant un échantillonneur stochastique de l'espace de recherche et un

⁸Ce nombre de couples (140) comparé aux 4000 couples initiaux semble indiquer une convergence prononcée de l'algorithme génétique de l'étape 2. Le rôle de cet algorithme n'étant pas précisément de produire une solution fine, mais plutôt un ensemble complet de solutions potentielles, une convergence exagérée de ce dernier peut pénaliser l'optimisation réalisée par l'étape numéro 3. Il serait peut être plus judicieux de diminuer le nombre d'itérations de cet AG, ou alors d'augmenter la taille de la population génétique.

algorithme d'optimisation locale. Rappelons que pour un algorithme génétique, deux points intéressants sont à noter. Il faut premièrement éviter les convergences prématurées d'un AG en autorisant librement la diversité (nombre de chromosomes et probabilités des opérateurs d'exploration suffisamment élevés, ainsi qu'un opérateur de sélection pas trop pointu). Deuxièmement, il faut permettre la convergence de l'algorithme en définissant un opérateur de sélection pertinent, et en laissant l'algorithme explorer l'espace de recherche pendant un nombre d'itérations suffisant. Toute la question est maintenant d'évaluer les termes « suffisants » et « élevés » en gardant en tête que ce qui nous intéresse est également de minimiser un facteur temps. Pour résumer, nous cherchons à minimiser le rapport temps, performance en fonction des divers paramètres de nos deux algorithmes génétiques. Cette optimisation sera menée par une étude statistique (analyse comportementale) du recalage.

4.5.1 Protocole de l'analyse

Le principe d'enchaînement des trois étapes est décrit par la figure 4.10. Les paramètres d'entrée que nous pouvons faire varier et que nous allons analyser sont le nombre de générations des recalages rigides et élastiques, et le nombre de chromosomes utilisés en recalage rigide et en recalage élastique, ce qui fait un total de 4 paramètres.



Figure 4.10 — Mise en cascade des trois étapes du recalage.

Par ailleurs, les indicateurs que nous décidons d'observer lors du déroulement de l'algorithme de recalage, pour chaque configuration des paramètres en entrée, sont les performances maximales des recalages rigides et élastiques (en fait les performances du meilleur chromosome du recalage rigide puis du recalage élastique), la performance finale obtenue après post-analyse de la population génétique du recalage élastique, et enfin le temps global d'exécution.

Il est à noter que nous avons exclu de cette analyse l'étude du choix des paramètres probabilité de mutation et probabilité de croisement. Les valeurs que nous avons choisies sont des valeurs standards (à savoir pm = 0.01 et pc = 0.8) et, bien qu'il soit peut-être possible d'améliorer l'algorithme en réglant finement ces paramètres, voire en les faisant varier automatiquement au cours des itérations, il n'est pas sûr que le gain soit sensible. De plus, notons qu'ajouter des possibilités à l'algorithme comme la variation automatique des paramètres de mutation et croisement en fonction de la performance, n'est peut-être pas intéressant. En effet, il vaut mieux parfois avoir un algorithme simple effectuant beaucoup d'itérations rapides, que peu d'itérations lentes et complexes où l'on se rapproche de la philosophie des algorithmes déterministes⁹.

Avant de commencer l'étude proprement dite, nous avons essayé de déterminer la performance maximale que l'on pouvait obtenir aux diverses étapes. Ces performances ont été obtenues en faisant tendre le nombre d'itérations vers l'infini et en constatant que la performance, en moyenne, n'évoluait plus au delà d'un certain nombre d'itérations. Le tableau 4.4 résume les performances que l'on peut obtenir ainsi que le nombre d'itérations nécessaire en

⁹La phase déterministe reste néanmoins nécessaire pour affiner les recherches d'un algorithme stochastique rapide.

moyenne pour les obtenir¹⁰. Bien entendu, nous calculons ici une valeur moyenne puisque l'aspect stochastique des algorithmes génétiques implique que chaque exécution de l'algorithme est unique et que les seules choses que nous pouvons observer sont leur comportement moyen ainsi que leur stabilité (*i.e.* l'absence de mauvaise solution dans l'absolu).

Recalage	Rigide	Élastique	Post-analyse
Perf. max.	75	82	82
Nb. it. rigides	50	50	50
Nb. it. élastique	-	100	100

Tableau 4.4 — Performance maximale moyenne et nombre d'itérations nécessaires.

Nous remarquons à l'aide du tableau 4.4 que la performance maximale obtenue avec affinement est la même que la performance maximale que l'on obtient avec un algorithme génétique. Néanmoins, remarquons bien que le temps nécessaire à la convergence de l'algorithme génétique est beaucoup plus long que celui nécessaire à la post-analyse. Aussi, il faut bien voir qu'il est plus intéressant d'utiliser l'algorithme génétique du recalage élastique (deuxième étape) juste comme un échantillonneur de l'espace de recherche donnant des solutions suffisamment bonnes pour que la post-analyse puisse dériver une solution finale optimale. Toute la difficulté réside maintenant dans la définition du « suffisamment » pour la convergence de la deuxième étape. Nous remarquons, en outre, que dans le cas des données médicales du TIMC, une modélisation élastique de la transformation est nécessaire puisque la performance maximale que l'on arrive à obtenir avec un recalage rigide est inférieure à celle que l'on obtient avec une méthode élastique.

4.5.2 Mise en œuvre

Nous avons essayé d'étudier l'influence du nombre de générations génétiques sur la valeur des performances intermédiaires et finales. Nous avons pour cela choisi de fixer le nombre de chromosomes pour les deux algorithmes génétiques à des valeurs usuelles, à savoir 100 chromosomes pour le recalage rigide et 500 chromosomes pour le recalage élastique¹¹. Bien que pour être exhaustif dans l'analyse des paramètres de l'algorithme génétique, on aurait du également traiter du choix du nombre de chromosomes en relation avec le nombre d'itérations, nous avons préféré fixer le nombre de chromosomes à une valeur constante et seulement analyser l'influence du nombre d'itérations sur la convergence de l'algorithme. On a constaté qu'à partir d'un certain nombre (environ 50 en rigide et 300 en élastique), augmenter le nombre de chromosomes trop faible, l'algorithme n'est plus robuste (obtention de mauvaises solutions). Nous avons choisi 100 chromosomes en rigide, et 500 chromosomes en élastique pour prendre une marge par rapport à la limite observée (50, 300) et être sûrs de ne pas perturber l'analyse de l'influence du nombre d'itérations.

¹⁰Nous rappelons que les chiffres que nous montrons n'ont qu'une valeur relative car ils correspondent à un jeu de données bien particulier et ne reflète en aucun cas la généralité des entrées qu'est capable d'accepter l'algorithme. Cette étude nous permet juste de constater à quel titre le recalage élastique ou rigide est nécessaire ainsi que dans quelles proportions on doit s'attacher à les exécuter.

¹¹En tenant compte du fait que plus un problème est difficile (grand espace de recherche, fonction de coût variant rapidement au voisinage d'un paramètre), plus il est intéressant d'avoir un grand nombre de chromosomes et vice versa. Pour le recalage rigide, l'espace de recherche est continu et lisse, et possède seulement 6 paramètres; pour le recalage élastique, l'espace de recherche est non continu (discret) et possède 8 degrés de liberté.

Nous avons alors choisi de faire varier nos paramètres (nombre d'itérations) entre 1 et ItMax où ItMax représente le nombre d'itérations nécessaires en moyenne pour obtenir la performance maximale.

Les tests ont été menés sur un même jeux de données (toujours les mêmes images du TIMC), et pour chaque couple de valeurs (itérations en rigide, itérations en élastique) une centaine de recalages différents a été effectuée.

Pour mesurer l'impact du choix du nombre d'itérations sur le temps, nous avons tout d'abord vérifié que l'algorithme se comportait en O(n). Cette vérification s'est effectuée en même temps que la recherche des performances maximales. Nous constatons que le temps d'exécution est globalement linéaire à l'étape 1 et à l'étape 2 par rapport au nombre d'itérations. Les quelques fluctuations que nous observons viennent tout simplement du fait que les algorithmes génétiques sont des algorithmes stochastiques et que, par définition le nombre d'opérations effectuées par itération est sujet à quelques variations d'une exécution à une autre.

Avant de se lancer dans l'analyse des résultats regardons d'abord quels sont les points qui nous intéressent, ce que l'on essaye de faire ressortir. Les différentes questions que l'on se pose sont :

- À partir de quel état de convergence de la première étape la deuxième étape devient-elle efficace ?
- Sommes nous obligés d'attendre au maximum la convergence des algorithmes génétiques avant de passer aux étapes suivantes ?
- La post-analyse est-elle plus efficace sur une population génétique uniforme (qui a complètement convergé) ou encore aléatoire? (a priori la post-analyse est efficace si toutes les solutions ne sont pas identiques certes, mais suffisamment bonnes quand même !)
- Comment prendre en compte le temps global nécessaire aux divers recalages? Ce que l'on veut, c'est pour deux jeux de paramètres amenant à la même performance moyenne choisir celui qui permet d'aller le plus vite. On ne cherche pas à décider quand estce qu'une solution est jugée satisfaisante d'un point de vue recalage. Donc finalement on choisira le nombre d'itérations à partir duquel des itérations supplémentaires n'apportent pas d'amélioration, ou alors, le nombre d'itérations en deçà duquel on obtient une perte de qualité.

4.5.3 Analyse des résultats

4.5.3.1 Convergence du recalage rigide

Commençons tout simplement par constater l'évolution de la performance rigide maximale obtenue en moyenne (à la sortie de l'étape 1 décrite à la figure 4.10) en fonction du nombre d'itérations rigides.

Sur ce graphe, nous constatons que la performance¹² maximale moyenne du recalage rigide est corrélée au nombre d'itérations (résultat prévisible). Plus le nombre d'itérations est important, et plus la performance augmente en tendant assymptotiquement vers la valeur optimale du recalage rigide (donnée par le tableau 4.4).

Notons que nous n'avons pas représenté ici l'évolution de l'écart-type de notre statistique. En théorie, plus le nombre d'itérations est élevé, et plus l'écart-type tend vers l'erreur minimale induite par la fonction distance qui est stochastique.

¹²La performance maximale moyenne est la performance moyenne du meilleur chromosome de l'algorithme génétique sur la série de recalages effectués.



Figure 4.11 — Performance maximale du recalage rigide.

Un autre point intéressant à noter sur la figure 4.11 est le nombre d'itérations nécessaires pour atteindre un certain pourcentage de la valeur limite. Avec nos simulations (en regardant sur la courbe), on remarque qu'il faut en moyenne 18 itérations pour atteindre 90% de la valeur limite, et 25 itérations pour atteindre 95% de la valeur limite. Donc on voit bien qu'en sacrifiant 5% de performance sur le recalage rigide, on diminue de moitié le temps de calcul. Aussi il est important de se poser la question sur la nécessité de pousser ou non au maximum la recherche génétique.

4.5.3.2 Ajout du recalage élastique

Pour étudier le recalage élastique, nous nous contentons ici de montrer les différences de convergence selon que le nombre d'itérations élastiques est faible ou élevé (1 ou 100). La figure 4.12 montre l'évolution de la performance à l'issue de l'étape 2 en fonction :

- Du nombre d'itérations rigides (de 1 à 50);
- Du nombre d'itérations élastiques (1, 50 ou 100).



Figure 4.12 — Performance maximale moyenne avec recalage élastique.

La courbe du bas (—) correspond à la performance du recalage rigide (sans recalage élastique donc), la courbe suivante (—) montre la performance maximale moyenne obtenue en n'effectuant qu'une itération de recalage élastique. Puis la courbe (—) montre la performance maximale moyenne obtenue après 50 générations génétiques, et enfin la courbe du haut (—) dénote les performance obtenues après 100 générations génétiques de recalage élastique.

Ces courbes sont utiles pour analyser l'intérêt de l'algorithme génétique de recalage élastique. On remarque nettement l'apport, en terme de performance, d'une déformation élastique par rapport à une déformation rigide. Ce qui est moins clair est de savoir combien d'itérations élastiques sont nécessaires. Dans le cas présent, nous n'utilisons pas de post-analyse (historiquement c'est ce que l'on a fait en premier). Nous voyons clairement que, quelque soit le nombre d'itérations en recalage rigide, plus on fait d'itérations élastiques et meilleur est le résultat (il tend vers la valeur maximale telle que donnée dans le tableau 4.4).



Figure 4.13 — Performance du recalage élastique à nombre d'itérations rigides constant.

Par ailleurs, la figure 4.13 nous renseigne sur l'évolution de la performance du recalage élastique au cours des itérations élastiques à nombre d'itérations du recalage rigide constant (sont représentés les courbes pour 1 itération rigide (—), 10 itérations (—), 20 itérations (—), 30 itérations (—) et enfin 50 itérations rigides (courbe du haut)). On note grâce à ce graphique que l'apport du recalage élastique en terme de performance n'est pas très grand. L'intérêt de cette étape est donc seulement d'augmenter la robustesse de l'algorithme (bon échantillonnage de l'espace de recherche) plutôt que la recherche d'une solution optimale.

De plus, on remarque que si l'on s'en tient à n'utiliser que les étapes 1 et 2 de notre algorithme (pas de post-analyse), la performance limite du système est guidée par le nombre d'itérations rigides. Avec peu d'itérations rigides et un nombre infini d'itérations élastiques on n'arrive pas à atteindre la performance maximale¹³. On voit qu'il faut donc un minimum de 25 itérations rigides (dans le cas présent) pour que le système soit efficace. Il reste encore à voir si cette règle est encore valable lorsque l'on utilise la post-analyse.

Ce que cette courbe ne montre pas, c'est la robustesse de la convergence (*i.e.* valeurs de l'écart-type), et le temps nécessaire. Par ailleurs, nous verrons au paragraphe suivant que l'utilisation d'un algorithme de *post-analyse* permet d'améliorer nettement les performances du recalage élastique, tout en diminuant le nombre d'itérations nécessaires.

4.5.3.3 Utilisation de la post-analyse

Nous reprenons les mêmes expériences qu'au paragraphe précédant mais nous ajoutons une post-analyse sur la population finale (c'est le système complet). Nous observons alors sur la figure 4.14 les résultats obtenus après enchaînement des trois étapes pour différentes valeurs du nombre d'itérations rigides, et pour trois nombre d'itérations élastiques différents (1 itération pour la courbe (---), 50 itérations pour la courbe (---), et enfin 100 itérations pour

¹³Puisque l'approximation donnée par le recalage rigide conduisant la construction de l'espace des couples est mauvaise.

la courbe (--)). Nous avons de plus représenté en série de points les courbes de la performance du recalage élastique de la figure 4.12.



Figure 4.14 — Performance maximale moyenne avec recalage élastique et post-analyse.

La première constatation évidente est que pour toutes les configurations testées, la postanalyse apporte une amélioration de performance (utilité de la troisième étape après la deuxième étape). L'affinement du résultat est donc d'une part nécessaire, et d'autre part efficace.



Figure 4.15 — Performance avec post-analyse à nombre d'itérations rigides constant.

Par ailleurs nous constatons que quelque soit le nombre d'itérations de l'étape 2 (recalage élastique), après post-analyse les résultats sont tous à peu près identiques (voir également la figure 4.15 représentant la performance après post-analyse en fonction du nombre d'itérations élastiques à nombre d'itérations rigides constant. Les couleurs sont les mêmes que celles employées à la figure 4.13, à savoir (—) 1 itération rigide, (—) 10 itérations, (—) 20 itérations, (—) 30 itérations et enfin 50 itérations pour la dernière courbe). À la simple vue de cette figure, on pourrait avoir tendance à penser que l'étape 2 n'est pas nécessaire. Néanmoins, au moins une génération élastique est nécessaire pour créer les 8-uplets qui forment l'espace de recherche de l'algorithme de post-analyse. De plus ces 8-uplets, qui ne sont rien d'autre que des chromosomes, sont désormais évalués par une fonction de performance ce qui est nécessaire pour initialiser l'étape 3 (post-analyse). Enfin, même si ce n'est pas représenté par cette figure, nous notons que l'algorithme nécessite plusieurs itérations de l'étape 2 afin d'être robuste (ce qui doit se manifester par une diminution de l'écart-type de la distribution des performances). Donc cette deuxième étape est primordiale pour le bon fonctionnement de la chaîne globale de traitement.

4.5.4 Discussion et interprétation des Résultats

Les premiers résultats que nous pouvons déduire de cette analyse est que l'utilisation de trois étapes est justifiée. Il n'est pas efficace d'utiliser un algorithme génétique, donc de recherche global, sans le coupler *in fine* à un algorithme de recherche local. En outre, il n'est pas efficace de commencer directement par une recherche locale. Donc les deux algorithmes doivent fonctionner en synergie. Les deux premières étapes (génétiques) constituent un échantillonneur stochastique de l'espace de recherche dont les échantillons forment le nouvel espace de la recherche locale.

Quant à la valeur optimale des paramètres, nous voyons par exemple que si le nombre d'itérations rigides est compris entre 25 et 50, la performance *in fine* reste a peu près constante. Pour le nombre d'itérations élastiques, il semble inutile de pousser trop loin l'algorithme. Pour résumer, à la vue des statistiques menées et selon l'expérience, nous pensons que les paramètres optimaux, pour les images que nous avons considérées lors de l'étude, sont d'environ 25 itérations rigides, suivies d'une vingtaine d'itérations élastiques et enfin de l'algorithme de post-analyse.

4.6 Synthèse du chapitre

Ce chapitre vient de justifier et décrire l'approche hiérarchique que nous avons employée pour résoudre de manière robuste notre problème de recalage élastique 3D multimodalité.

La décomposition en 3 étapes visant à, d'une part réduire progressivement la taille de l'espace de recherche, et d'autre part à affiner la qualité du recalage obtenu est robuste dans le sens ou nous écartons au cours du traitement seulement les solutions ne pouvant mener à un bon résultat, et dans le sens ou nous utilisons l'algorithme génétique comme un échantillonneur stochastique d'un espace de recherche couplé avec une procédure d'optimisation locale. La robustesse de notre algorithme tient donc à deux facteurs : la robustesse intrinsèque des algorithmes génétiques à bas niveau, et la robustesse de la structure hiérarchique globale.

Même si, en théorie, il existe un jeu de paramètres unique pour chaque recalage d'images médicales, le nombre de degrés de libertés du modèle élastique implique qu'en pratique il existe beaucoup de jeux de paramètres, sensiblement différents en valeur, amenant visuellement au même résultat.

Finalement, ce chapitre nous a permis également de montrer une application directe des différentes étapes de l'algorithme de recalage sur des images médicales réelles, et ainsi de présenter des résultats et leurs interprétations, tant au niveau numérique qu'au niveau visuel.



Figure 4.16 — Données médicales avant recalage.



Figure 4.17 — Résultats du recalage rigide (étape 1).



Figure 4.18 — Résultats du recalage élastique (étape 2).



Figure 4.19 — Résultats finaux (une seule classe de courbures).



Figure 4.20 — Résultats finaux (trois classes de courbures).

CHAPITRE 5 Application et évaluation du recalage élastique 3D

Ce chapitre montre une application de notre méthode de recalage 3D élastique sur une base de données d'images IRM et scanner. Cette base de données a été établie au sein du projet RREP « *Retrospective Registration Evaluation Project* » (projet d'évaluation rétrospective de méthodes de recalage) projet numéro 1 R01 NS33926-01, au National Institutes of Health, Vanderbilt University, Nashville, TN., et dont le principal responsable est J.-M. Fitzpatrick [98].

Au cours de ce chapitre, nous allons en premier lieu présenter la base de données RREP que nous avons utilisée. Par la suite, nous définirons une stratégie d'analyse de cette base, afin d'évaluer la robustesse de notre algorithme de recalage.

Nous donnerons par conséquent les résultats, tant au niveau numérique que visuel, de divers recalages effectués à partir des images de cette base. À ce propos, nous discuterons des problèmes à l'égard de l'utilisation d'une telle base et de la confiance que l'on peut accorder aux résultats.

Enfin, nous aborderons la comparaison des résultats de la technique de recalage que nous avons présentée au chapitre précédent par rapport aux résultats des autres équipes de recherche utilisant cette base de données RREP.

5.1 La base de données RREP

5.1.1 Historique

Le but du projet RREP est de fournir une évaluation clinique de la précision des techniques rétrospectives de recalage d'images volumiques multimodales du cerveau humain. Le terme rétrospectif s'entend dans le sens ou des images (de diverses modalités) sont acquises à différents moments, et ultérieurement, plusieurs techniques de recalage peuvent être employées et comparées. Ce projet est articulé autour de trois modalités le PET (*Positon Emission Tomography*, utilisé en imagerie fonctionnelle), l'IRM et le TDM. Les recalages à effectuer sont le PET sur l'IRM pour la visualisation de l'activité neuronale sur un contexte anatomique (IRM), et le recalage IRM sur TDM pour la visualisation des tissus mous dans un contexte de structure rigide (os extrait des images TDM).

La qualification objective des différentes techniques de recalage par le biais du projet RREP devrait permettre d'établir un degré de confiance auquel peuvent s'attacher les neurologistes et les neurochirurgiens en matière de recalage rétrospectif d'images cérébrales.

5.1.2 Principe d'utilisation et de comparaison

La base de données est constituée d'images d'une dizaine de patients. Le principe original était uniquement de comparer des méthodes de recalage rigide, aussi la qualité et la précision du recalage sont déterminées par comparaison avec un recalage rigide prospectif. Ce recalage rigide est déterminé à l'aide d'un ensemble d'amers fixés sur le crâne avant l'acquisition des images avec un casque stéréotaxique. La localisation de ces points étant aisée sur les images TDM, PET et IRM, un algorithme déterministe peut facilement, et avec une bonne précision, calculer les paramètres de la transformation rigide correspondante. Les images mises dans la base de données ont été traitées de manière à supprimer l'information donnée par ces marqueurs¹.

Par ailleurs, pour déterminer l'impact des distorsions des images IRM sur la précision du recalage, chaque image IRM est disponible en deux versions : une version ayant subi une correction géométrique (calibration), et une version n'ayant subi aucune correction.

La comparaison avec les résultats des autres techniques (rétrospectives) peut donc se faire soit sur les valeurs des paramètres de la transformation rigide (translation 3D et angles de rotation), soit sur la position transformée d'un ensemble de points.

5.1.3 Les images de la base de données

Les différentes images dont nous disposons dans cette base, ainsi que leurs caractéristiques sont répertoriées dans le tableau 5.1. Notons que sur les 9 patients ayant servi à la construction de cette base de données, seuls les 7 premiers disposent d'une image scanner (TDM). Puisque l'algorithme que nous avons développé fonctionne sur la base d'un recalage entre surfaces numériques, et puisque la résolution des images PET est assez faible, la segmentation de ces dernières est un sujet délicat ; aussi nous restreignons l'application de notre algorithme au cas des recalages TDM-IRM. Par conséquent, seuls les 7 premiers patients seront utilisés.

¹Néanmoins, comme nous allons le constater bientôt, les empreintes de ces marqueurs restent fortement présentes dans la segmentation par LPE.

patient	modalité	$\operatorname{correction}$	nombre	taille	épaisseur	résolution
		géométrique	coupes	coupe	coupe	coupe (mm)
		$\operatorname{disponible}?$		(pixels)	(mm)	
	TDM	Non	28	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Oui	26	256×256	4	$1,25 \times 1,25$
1	IRM DP	Oui	26	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	26	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$
	TDM	Non	29	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Oui	26	256×256	4	$1,25 \times 1,25$
2	IRM DP	Oui	26	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	26	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$
	TDM	Non	34	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Oui	20	256×256	4	$1,25 \times 1,25$
3	IRM DP	Oui	20	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	20	256×256	4	$1,25 \times 1,25$
	TDM	Non	27	512×512	4	0,653595 imes 0,653595
	IRM T1	Oui	20	256×256	4	$1,25 \times 1,25$
4	IRM DP	Oui	20	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	20	256×256	4	$1,25 \times 1,25$
	TDM	Non	33	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Oui	26	256×256	4	$1,25 \times 1,25$
5	IRM DP	Oui	26	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	26	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$1,943042 \times 1,943042$
	TDM	Non	30	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Non	26	256×256	4	$1,25 \times 1,25$
6	IRM DP	Oui	26	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	26	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$
	TDM	Non	28	512×512	4	$0,653595 \times 0,653595$
	IRM T1	Oui	26	256×256	4	$1,25 \times 1,25$
7	IRM DP	Oui	26	256×256	4	$1,25 \times 1,25$
	IRM T2	Oui	26	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$
	IRM T1	Non	24	256×256	4	$1,25 \times 1,25$
8	IRM DP	Non	24	256×256	4	$1,25 \times 1,25$
	IRM T2	Non	24	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$
	IRM T1	Non	24	256×256	4	$1,25 \times 1,25$
9	IRM DP	Non	24	256×256	4	$1,25 \times 1,25$
	IRM T2	Non	24	256×256	4	$1,25 \times 1,25$
	PET	Oui	15	128×128	8	$2,590723 \times 2,590723$

Tableau 5.1 — Les images disponibles dans la base de données.

Des rendus volumiques ainsi que des vues en coupe des volumes médicaux du patient numéro 1 sont représentés à la figure 5.1. Nous constatons sur ces images que les IRM pondérées en T2 ne fournissent pas une image très contrastée de l'interface air/chair. Cette dernière est même absente sur certaines coupes. Par conséquent, ces images sont difficiles à segmenter et donc ne peuvent pas être directement prises en compte par notre algorithme de recalage. Nous ne considérons donc pas ici les recalages des images IRM T2, mais seulement les recalages entre images scanner TDM et IRM T1 ou DP.



Figure 5.1 — Quelques rendus volumiques et quelques coupes des volumes médicaux du patient numéro 1 dans la base de données RREP.

Par ailleurs, la figure 5.2 montre un panel de rendus volumiques des 7 patients en TDM et en IRM, ce qui permet d'apprécier les extensions physiques (volumes d'intérêt) des volumes considérés. Par la suite, bien que les recalages portent sur l'ensemble des données constituées par ces 7 patients, nous ne montrerons que des rendus volumiques relatifs aux données du patient numéro 1 et du patient numéro 5 (ceux-ci ayant l'avantage d'offrir un volume d'intérêt en IRM non restreint à la seule partie supérieure du crâne, et donc sont visuellement plus intéressants).



Figure 5.2 — Rendus volumiques directs des volumes TDM et IRM T1 des 7 premiers patients de la base de données RREP.

5.2 Méthodologie d'utilisation de la base

Le projet RREP visait à une évaluation des techniques rétrospectives de recalages rigides. Aussi, une première étape (cf. section 5.3.1) dans notre approche sera d'apprécier la qualité des divers recalages rigides que nous obtenons. Notre algorithme n'a pas été optimisé pour donner des recalages rigides idéaux, mais seulement des approximations fiables d'un recalage rigide global faisant correspondre aux mieux les points des images. De par sa position dans la chaîne de traitement (voir chapitre 4), le recalage rigide se contentait de donner rapidement des résultats approchés. Si nous considérons maintenant l'étape de recalage rigide comme la seule étape du traitement, nous pouvons laisser à l'algorithme génétique plus de champ de recherche en augmentant par exemple significativement la taille de la population génétique. Cette méthode, bien que sous-optimale (il faudrait en fait coupler cet algorithme génétique de recalage rigide avec un algorithme d'optimisation locale), permet de donner, à moindre effort, une estimation correcte du recalage rigide recherché.

Afin de valider plus précisément l'apport du recalage élastique, le deuxième point auquel nous nous attachons (cf. section 5.3.2) est la comparaison entre les résultats (chiffrés en terme de performance stochastique) du recalage rigide et du recalage élastique. Autrement dit, la deuxième étape est constituée par l'application de notre algorithme de recalage élastique sur les images de la base.

Enfin, nous discuterons les possibilités de comparaison des résultats des recalages entrepris, avec les résultats d'autres équipes de recherche.

5.3 Principaux résultats

Pour chaque patient, deux types de recalages sont effectués : un recalage rigide optimal, puis un recalage élastique. Les volumes que nous utilisons pour chaque patient sont les volumes TDM, IRM T1, IRM T1 corrigés, IRM DP et IRM DP corrigés, ce qui fait en tout 4 recalages rigides et 4 recalages élastiques pour chaque patient. Le domaine physique des images TDM étant plus étendu que celui des images IRM, nous avons choisi d'effectuer le recalage dans le sens IRM vers TDM². Notons que ce choix permet en outre d'obtenir *in fine* les deux images recalées sur le plus grand des deux domaines³. Notons enfin que le choix du sens de recalage que nous avons fait ne correspond pas au choix qui a été retenu pour le projet RREP, ce qui rajoute des contraintes, comme nous le verrons plus loin, lors de la comparaison avec les résultats des autres équipes de recherche.

5.3.1 Recherche du recalage rigide optimal

Pour effectuer un recalage rigide rapide des données, nous restreignons notre algorithme à seulement la première étape de la chaîne globale de traitement (cf. section 4.1 du chapitre précédent). Cette étape est constituée d'un algorithme génétique fonctionnant sur un espace de recherche borné inclus dans \mathbb{R}^6 . Ces 6 paramètres sont, d'une part, les 3 réels formant les angles de rotation (autour des axes du repère principal centré de l'image), et d'autre part les 3 réels relatifs à une translation 3D. Cet algorithme génétique (cf. section 3.4.1) recherche directement les 6 paramètres réels maximisant la fonction de performance stochastique définie par la relation (2.8) de la page 61.

La résolution des images de la base RREP étant du même ordre que celle des images utilisées au chapitre 4, nous pouvons facilement nous assurer qu'utiliser des paramètres de l'algorithme génétique (voir section 3.2.2) identiques donne de bons résultats.

Nous partons donc sur la base de 50 générations génétiques, de 100 marqueurs ou points

²Le recalage dans le sens IRM vers TDM est plus difficile pour l'algorithme de recalage rigide, car ce dernier calculant la transformation dans le sens Référence \rightarrow Image à recaler, se retrouve avec beaucoup de points n'ayant pas d'homologue (des points du TDM qui n'ont pas de correspondant sur l'image IRM), ce qui induit un biais dans la fonction distance.

 $^{^{3}}$ Notons également que dans le cadre d'un recalage élastique IRM/TDM, les images IRM étant plus sujettes aux non-linéarités d'acquisition, il est préférable d'aligner élastiquement l'IRM sur le scanner que le contraire.

utilisés pour le calcul de la performance stochastique, d'une probabilité de mutation $p_m = 0.01$, et de recombinaison $p_c = 0.8$. Par ailleurs, dans l'utilisation de cet algorithme au chapitre 4, nous avions utilisé une population génétique de 500 chromosomes, ce qui avait l'avantage de donner rapidement des résultats, mais pas forcément très précis. Ici nous augmentons cette population, pour le recalage rigide, à 1000 individus en espérant ainsi obtenir des résultats de recalage plus précis qu'une simple approximation.

Pour chaque patient de la base de données, 4 images IRM sont recalées vers l'image TDM choisie comme référence. Par ailleurs, pour chaque image IRM une dizaine de recalages⁴ sont effectués, et nous nous intéresserons ici, non pas à un recalage particulier, mais à une statistique sur l'ensemble des recalages.

Les résultats que nous représentons ici pour les recalages rigides seront de deux natures : d'une part nous présenterons une évaluation chiffrée des divers recalages avec notamment la moyenne et l'écart-type (sur la dizaine de recalages effectués) de la performance stochastique moyenne. De plus nous donnerons les valeurs moyennes⁵ et les écart-types sur les paramètres (translation en millimètres et rotation en degrés) trouvés. Le temps de calcul moyen observé sera également reporté sur les tableaux de résultats. En outre, nous présenterons quelques rendus coupe à coupe et volumiques (cf. annexe C) des recalages médians obtenus sur deux patients particuliers (patient numéro 1 et numéro 5) présentant l'avantage d'avoir une extension verticale plus importante que pour les autres patients, et par conséquent donnant une visualisation volumique plus facile à interpréter.

5.3.1.1 Résultats chiffrés

Le tableau 5.2 à la page suivante regroupe les résultats des recalages rigides par optimisation génétique des images médicales des 7 premiers patients de la base de données.

La première constatation que nous pouvons faire se situe au niveau des performances globales moyennes des recalages des divers patients. Nous observons que la performance obtenue varie grandement d'un patient à un autre (aux alentours de 93 % pour le patient numéro 1 et seulement 75 % pour le patient numéro 3). Cette différence de performance de recalage rigide peut s'expliquer par la nature de la fonction d'appariement (mesurant la qualité de la correspondance entre les images), et par le fait que les images ne présentent pas toutes les mêmes extensions verticales. Pour tous les patients, le nombre de coupes IRM disponibles est inférieur au nombre de coupes TDM. La nature de la fonction d'appariement impose que plus le nombre de coupe IRM est élevé et le nombre de coupes TDM est réduit⁶, alors plus la fonction d'appariemment peut atteindre des valeurs élevées. Si l'on classe les patients en fonction du rapport nombre de coupes IRM sur nombre de coupes TDM (par exemple pour le patient 1, 26/28 \approx 0.93), l'ordre est le suivant : 1,7,2,6,5,4,3. Si l'on regarde maintenant les résultats de performance stochastique moyenne après recalage rigide, on constate, comme on aurait pu le prédire, que nous obtenons le même classement.

Toujours à propos des performances stochastiques des recalage rigides, on constate que les volumes IRM DPR aboutissent à un meilleur recalage rigide que les volumes IRM DP.

⁴Nous effectuons en fait 11 recalages. Nous avons choisi un nombre impair de manière à pouvoir facilement extraire un recalage médian (le sixième par ordre de performance) afin de montrer une validation visuelle. Nous ne montrons ainsi pas le meilleur ni le moins bon des 11 recalages effectués, et nous ne calculons pas non plus un recalage hybride comme une moyenne des 11 recalages.

⁵La transformation calculée à partir des coefficients moyens ne doit pas refléter une transformation moyenne réalisable ; néanmoins, nous indiquons cette valeur pour renseigner sur la distributions des coefficients trouvés.

#	modalité	Performance	Rotation	Rotation écart-	Translation moyenne	Translation	Temps
	recalée	moyenne et	moyenne (de-	type (degrés)	(millimètres)	écart-type (mil-	CPU
		écart-type	grés)			limètres)	(sec.)
	IRM T1	91.92(1.00)	(0.59, 0.18, -0.85)	(0.40, 0.51, 0.57)	(1.00, -23.23, -5.65)	(0.16, 0.28, 0.51)	15.7
	IRM T1R	$94.31 \ (0.51)$	(0.46, 0.43, -0.34)	(0.52, 0.86, 0.39)	(0.78, -24.33, -5.49)	(0.24, 0.41, 0.51)	18.5
	IRM PD	89.54(1.30)	(0.72, 0.34, -1.03)	(0.58, 0.48, 0.59)	(1.22, -23.76, -4.23)	(0.31, 0.59, 0.77)	15.6
	IRM PDR	95.45(0.76)	(0.69, 0.37, -0.41)	(0.77, 0.60, 0.52)	(0.85, -24.27, -3.92)	(0.19, 0.42, 0.54)	15.5
	IRM T1	87.47 (2.69)	(5.01, -1.21, 0.11)	(4.15, 1.14, 0.63)	(1.12, -27.88, -21.48)	(0.63, 1.42, 1.94)	15.1
0	IRM T1R	86.39(2.57)	(0.14, -0.15, -0.27)	(5.34, 1.45, 0.74)	(1.14, -29.16, -21.88)	(0.30, 2.29, 2.52)	15.0
	IRM PD	87.27(1.99)	(5.52, -1.08, -0.60)	(4.53, 2.00, 0.58)	(0.88, -34.69, -19.09)	(0.78, 1.60, 2.06)	15.1
	IRM PDR	88.75(2.37)	(2.01, -0.34, -0.20)	(2.58, 2.03, 0.62)	(0.72, -33.39, -18.75)	(0.48, 0.97, 0.93)	15.2
	IRM T1	77.42 (3.20)	(2.69, -1.47, 1.65)	(2.95, 3.78, 2.37)	(-0.18, -29.31, -42.53)	(1.09, 1.13, 2.80)	16.0
က	IRM T1R	76.18(3.96)	(1.34, -0.99, 1.19)	(5.27, 2.82, 1.52)	(-0.16, -32.15, -42.39)	(0.86, 1.84, 2.81)	14.2
	IRM PD	73.90(5.11)	(-3.45, -0.29, -0.25)	(3.74, 4.25, 1.39)	(-0.80, -27.76, -38.93)	(0.79, 1.08, 1.67)	16.5
	IRM PDR	78.84 (4.15)	(-3.13, 0.67, -0.95)	(3.09, 4.17, 1.57)	(-0.75, -30.33, -36.78)	(0.65, 0.98, 1.74)	16.6
	IRM T1	82.69(3.01)	(-0.25, 0.06, 0.16)	(4.11, 2.59, 1.37)	(0.15, -28.49, -36.34)	(0.60, 2.12, 0.86)	14.0
4	IRM T1R	83.07 (2.12)	(1.11, 2.93, 0.09)	(4.34, 2.91, 0.93)	(0.79, -31.82, -36.64)	(0.56, 2.08, 0.81)	14.2
	IRM PD	83.78(3.32)	(0.71, 0.50, 0.65)	(4.91, 3.08, 1.51)	(0.29, -27.98, -35.63)	(0.67, 2.30, 0.85)	13.7
	IRM PDR	84.49(1.94)	(0.20, 1.08, 0.57)	(5.22, 4.80, 1.52)	(0.29, -30.28, -34.35)	(1.08, 1.52, 1.19)	13.9
	IRM T1	85.40(2.13)	(0.67, -0.29, 0.59)	(2.07, 1.81, 0.67)	(-5.24, -31.73, -22.23)	(0.43, 0.89, 0.60)	18.3
ы	IRM T1R	86.78(2.73)	(0.72, 1.13, 0.53)	(2.28, 1.84, 1.02)	(-4.97, -32.47, -20.69)	(0.56, 0.76, 0.81)	18.5
	IRM PD	83.44(0.00)	(1.26, -0.40, -1.32)	(0.00, 0.00, 0.00)	(-6.17, -31.31, -21.16)	(0.00, 0.00, 0.00)	18.0
	IRM PDR	85.58(2.38)	(4.81, -2.05, -0.23)	(3.45, 2.77, 0.67)	(-6.32, -33.98, -19.57)	(0.86, 0.82, 0.87)	18.7
	IRM T1	84.45(3.50)	(-1.24, -0.82, -0.72)	(2.22, 2.55, 0.90)	(-8.11, -29.60, -27.33)	(0.66, 0.90, 0.78)	17.6
9	IRM T1R						
	IRM PD	83.39(1.91)	(-1.17, -0.36, -1.85)	(3.55, 2.30, 1.26)	(-8.59, -30.29, -26.88)	(0.58, 1.03, 0.85)	17.3
	IRM PDR	88.16(2.58)	(1.18, 0.16, -0.95)	(1.35, 2.57, 0.86)	(-7.94, -32.38, -23.46)	(0.68, 0.62, 0.80)	17.5
	IRM T1	87.16(1.34)	(2.15, 0.45, -0.15)	(1.78, 0.80, 0.84)	(-5.57, -22.24, -7.77)	(0.38, 0.80, 0.54)	16.3
1-	IRM T1R	85.70(0.91)	(1.25, 0.99, 0.01)	(1.43, 1.38, 0.67)	(-5.34, -24.69, -8.09)	(0.35, 0.61, 0.58)	16.2
	IRM PD	89.00(2.14)	(0.41, 1.00, -0.03)	(1.19, 1.43, 0.78)	(-5.38, -21.89, -7.71)	(0.20, 0.40, 0.87)	16.0
	IRM PDR	91.36(1.30)	(0.39, 0.48, 0.06)	(1.02, 0.76, 0.72)	(-5.66, -24.08, -6.70)	(0.41, 0.41, 0.34)	16.0

Tableau 5.2 -	 Résultats des r 	ecalages rigides	optimaux	obtenus	avec optimisation	géné-
---------------	-------------------------------------	------------------	----------	---------	-------------------	-------

tique.

Cette constatation ne semble pas vérifiée entre les volumes IRM T1 et IRM T1R, ce qui tend à montrer que la correction géométrique apportée sur les images IRM est surtout appréciable sur les images de type DP que sur les images pondérées en T1.

Une autre constatation intéressante concerne les écart-types relevés à propos des performances stochastiques et des paramètres du recalage rigide. Pour les patients 2, 3, 4, 5, et 6 ces écart-types sont relativement importants, ce qui signifie qu'il existe une grande variabilité d'un recalage rigide à un autre. Cette grande variabilité est un indice en faveur de la nécessité d'un recalage élastique. Pour les patients 1 et 7, les paramètres obtenus varient peu, ce qui tend à penser que le recalage rigide optimal est très proche du recalage élastique optimal

Les temps de calcul (temps CPU) reportés sur le tableau 5.2, d'une valeur moyenne de 16 secondes, correspondent au temps CPU s'écoulant durant l'exécution de l'algorithme génétique (étape 1) opérant le recalage rigide. Ce temps ne prend donc pas en compte, ni les entrées-sorties (lecture sur le disque dur et chargement en mémoire), ni le temps nécessaire à la segmentation morphologique des images et à la création d'une carte distance (de l'ordre de quelques minutes par image 3D). Enfin, ce temps ne prend pas en compte non plus le temps nécessaire à la création proprement dite du volume recalé et son écriture sur le disque dur. Tout ce qu'il comprend c'est l'exécution de l'algorithme génétique et l'extraction des paramètres de la transformation rigide recherchée. La machine sur laquelle le programme a tourné est un bi-processeur Intel Pentium-Pro cadencé à 233 MHz, avec Linux-2.2 comme système d'exploitation. Le programme lui-même a été réalisé de manière à prendre en compte les deux processeurs de la machine. Le rendement du parallélisme n'est bien entendu pas de 100 %, et l'on a estimé que l'on aurait un temps de calcul seulement environ 50 % plus grand sur une machine identique en mode mono-processeur.

5.3.1.2 Validation visuelle du recalage rigide optimal

La figure 5.3 à la page suivante (respectivement 5.4 à la page suivante) représente une coupe (30^{e} coupe axiale) de la fusion des volumes TDM et IRM avant et après recalage rigide des images du patient 1 (respectivement 5). Le recalage rigide sélectionné, tant pour le patient 1, que pour le patient 5, parmi les 11 recalages effectués est le recalage médian (le 6^{e} par ordre de performance stochastique), donc ni le meilleur, ni le plus mauvais.

Sur ces coupes nous avons représenté d'une part en vert l'os cortical extrait de l'image TDM par seuillage des niveaux de gris (valeurs prises entre 60 et 140), et d'autre part, en gris, les pixels issus des images IRM.

Les 4 premières vues en coupe des images du patient 1 montrent l'adéquation des données initiales (avant tout recalage). Nous constatons clairement qu'il n'y a aucune correspondance entre la position de l'os cortical (affiché en vert) extrait des images TDM et les niveaux de gris de l'image IRM originale. Par contre, sur les 4 vues en coupe du bas, nous constatons que le recalage rigide permet une bonne première approximation de la mise en correspondance de ces images. Ces remarques restent valables pour les vues en coupes du patient numéro 5 et de son recalage rigide médian.

En observant de plus près ces coupes, et surtout celles du patient 5, on se rend compte qu'il subsiste de petites imperfections avec le recalage rigide justifiant l'utilisation d'un recalage élastique.

La figure 5.5 page 119 (respectivement 5.6 page 120) représente des rendus volumiques multivolume, en mode union puis intersection, des résultats du recalage rigide du patient 1 (respectivement du patient 5). Ces visualisations sont bien plus discriminantes d'un point de



Figure 5.3 — Vues en coupe axiale des recalages rigides médians des IRM du patient 1 sur l'image TDM.



Figure 5.4 — Vues en coupe axiale des recalages rigides médians des IRM du patient 5 sur l'image TDM.

vue validation du recalage que les vues coupe à coupe, car elles prennent en compte l'ensemble du volume. Par ailleurs, elle permettent de visualiser le recouvrement et l'intersection floue des surfaces externes des volumes recalés (donc de la seule information connue de l'algorithme de recalage).



Figure 5.5 — Rendus volumiques sous deux angles d'incidence et deux modes de visualisation (union et intersection) des recalages rigides médians des différentes IRM du patient 1 sur son image TDM.

À partir des rendus volumiques, on observe nettement que le recalage rigide obtenu avec



Figure 5.6 — Rendus volumiques sous deux angles d'incidence et deux modes de visualisation (union et intersection) des recalages rigides médians des différentes IRM du patient 5 sur son image TDM.

le patient numéro 1 est meilleur que celui obtenu avec le patient 5, ce qui correspond aux résultats numériques en terme de performance stochastique. D'autre part on observe que les résultats ne sont pas parfaits et que l'on peut espérer sans doute de meilleurs résultats avec un recalage élastique.

5.3.2 Recalage élastique par optimisation génétique

La deuxième phase de validation consiste à effectuer des recalages élastiques complets des volumes IRM vers le volume TDM de chaque patient. De même que pour les recalages rigides présentés au paragraphe 5.3.1, une dizaine de recalages pour chaque paire de volumes à mettre en correspondance a été effectuée. Ces divers recalages nous permettront d'établir des statistiques visant à évaluer la robustesse de l'algorithme.

Les recalages rigides effectués précédemment et constituant la première des trois étapes du recalage élastique global sont réutilisés ici. Aussi l'algorithme commence directement par la deuxième étape en ayant comme connaissance a priori les résultats du recalage rigide (on utilise le recalage médian parmi les 11 recalages rigides effectués). En agissant ainsi, nous espérons mieux apprécier l'apport du recalage élastique par rapport au recalage rigide, aussi bien au niveau des performances stochastiques moyennes obtenues qu'au niveau des validations visuelles.

Les paramètres que nous avons utilisés, et en particulier les paramètres de l'algorithme génétique de l'étape numéro 2 (cf. chapitre 4) sont similaires à ceux de l'algorithme génétique optimisant le recalage rigide. Nous avons par conséquent une population génétique de 1000 individus, un nombre maximal de 50 itérations génétiques, une probabilité de mutation $p_m =$ 0.01 et une probabilité de recombinaison $p_c = 0.8$. Le nombre de marqueurs utilisés pour le calcul de la distance stochastique est toutefois augmenté à 200 (au lieu de 100 pour le premier algorithme génétique) afin de diminuer l'erreur stochastique inhérente au choix des marqueurs.

De la même manière que pour la validation du recalage rigide, nous présenterons ici deux types de résultats : d'une part, un résumé des moyennes et des écart-types obtenus sur la performance stochastique des divers recalages, et d'autre part une série de coupes et de rendus volumiques correspondant aux recalages médians des volumes des patients numéro 1 puis numéro 5.

5.3.2.1 Résultats chiffrés

Le tableau 5.3 présente les performances stochastiques moyennes, ainsi que leur écart-type, obtenus pour les 4 recalages de chaque patient. De plus, ce tableau présente le temps moyen qui a été nécessaire pour effectuer ces recalages élastiques.

En ce qui concerne les performances, et si l'on compare par rapport aux résultats obtenus pour le recalage rigide (tableau 5.2), on note une systématique augmentation entre le recalage rigide et le recalage élastique (étape 3), ce qui justifie l'emploi de ce dernier type de recalage.

Nous avons volontairement écarté de ce tableau les valeurs moyennes et les écarts-types des 24 paramètres des recalages élastiques qui n'apporteraient pas beaucoup d'information et noieraient le lecteur dans les chiffres. Retenons simplement qu'on observe une relativement grande variabilité sur ces paramètres, ce qui semble vouloir dire que le modèle de déformation élastique employé, bien que simple, ne doit pas modéliser parfaitement la déformation réelle des données et que plusieurs jeux de paramètres (24 paramètres) peuvent mener à des résultats similaires visuellement, et similaires en terme de performance stochastique (on observe en effet que l'écart-type des performances stochastiques finales est en général faible et souvent inférieur à 2 %). Retenons également de cette observation que l'algorithme de recalage est robuste puisque les résultats obtenus, bien que possédant des paramètres légèrement différents (du fait de la stochasticité de l'algorithme), aboutissent sur des rendus visuellement satisfaisants.

#	Modalité	Perf. élas-	Perf. élas-	Temps
	recalée	tique avant	tique finale	CPU
		post-analyse	$(étape \ 3)$	sec.
		$(étape \ 2)$		
	IRM T1	$94,\!89\ (0,\!49)$	$96,\!69\ (0,\!30)$	$99,\!2$
1	IRM T1R	95,77(0,46)	97,08(0,24)	$96,\! 6$
	IRM PD	$91,\!95\ (1,\!20)$	$93,\!97(1,\!04)$	$92,\!1$
	IRM PDR	$95,\!81\ (0,\!42)$	$97,\!23\ (0,\!39)$	94,2
	IRM T1	88,33 $(1,14)$	$91,\!52\ (1,\!46)$	$75,\!6$
2	IRM T1R	$85,\!92\ (2,\!05)$	$89,\!81\ (1,\!69)$	$74,\!9$
	IRM PD	87,71(1,39)	$90,\!36\ (1,\!30)$	$73,\!9$
	IRM PDR	$88,\!13\ (0,\!78)$	$90,\!56\ (1,\!07)$	$65,\!8$
	IRM T1	$77,\!21\ (3,\!05)$	$81,\!84(3,\!60)$	$62,\!8$
3	IRM T1R	$77,\!21\ (1,\!58)$	$81,\!68\ (2,\!56)$	$63,\!8$
	IRM PD	74,92 (1,76)	$80,\!15\ (1,\!92)$	$63,\!9$
	IRM PDR	$74,\!26\ (2,\!93)$	$78,\!87\ (3,\!09)$	61,4
	IRM T1	$85,\!33\ (1,\!51)$	$89,\!05\ (1,\!55)$	$61,\!1$
4	IRM T1R	$84,\!88\ (1,\!51)$	$88,\!80\ (1,\!74)$	$63,\!1$
	IRM PD	$84,\!92(1,\!91)$	88,10(1,42)	61,6
	IRM PDR	$82,\!94\ (1,\!37)$	$87,\!42\ (1,\!66)$	$63,\!1$
	IRM T1	88,99(1,14)	$91,75\ (1,12)$	$85,\!5$
5	IRM T1R	$89,\!19\ (1,\!03)$	$91,\!84\ (1,\!12)$	$91,\!4$
	IRM PD	$88,\!48\ (0,\!96)$	$91,\!84\ (0,\!82)$	$92,\!6$
	IRM PDR	$88,\!01\ (1,\!60)$	$90,\!43\ (1,\!57)$	71,9
	IRM T1	$84,\!95\ (1,\!90)$	$88,\!36\ (1,\!99)$	66,2
6				
	IRM PD	$82,\!60\ (2,\!06)$	$86,\!56\ (2,\!19)$	$67,\!3$
	IRM PDR	85, 13(1, 33)	$88,\!\overline{09}\ (1,\!41)$	65,5
	IRM T1	$91,\!58(1,\!41)$	93,92(1,07)	87,1
7	IRM T1R	$88,\!68(1,\!10)$	$91,\!79\ (1,\!65)$	84,1
	IRM PD	$91,\!81\ (0,\!75)$	$94,\!00\ (0,\!69)$	89,7
	IRM PDR	$9\overline{3,}33\ (0,61)$	$9\overline{5},\!32\ (0,\!52)$	99,0

Tableau 5.3 — Résultats des recalages élastiques obtenus avec optimisation génétique.

En ce qui concerne les temps de calcul affichés, ils sont relatifs à la même machine que celle présentée pour le recalage rigide (bi-Pentium Pro sous Linux). Le temps affiché correspond au cumul du temps nécessaire à l'algorithme génétique de la deuxième étape et au temps nécessaire à la post-analyse (étape 3). Ce temps de calcul ne tient pas en compte l'étape de création de la liste des couples \mathcal{L} , qui nécessite, selon la précision du recalage rigide, entre 20 secondes et une minute.

Le temps nécessaire pour effectuer le recalage élastique varie donc d'un patient à un autre

car la taille des volumes et la taille de la liste \mathcal{L} changent. Néanmoins, nous tournons aux alentours de 80 secondes.

In fine, si nous voulons estimer le temps global moyen nécessaire au traitement complet d'un couple d'images de la base de données RREP nous aurions environ 3 minutes pour la segmentation et le recalage rigide, et une minute trente pour le recalage élastique, donc au total 4 minutes 30.



5.3.2.2 Validation visuelle du recalage élastique

Figure 5.7 — Vues en coupe axiale des recalages élastiques du patient 1.



Figure 5.8 — Vues en coupe axiale des recalages élastiques du patient 5.

Les figures 5.7 et 5.8 présentent les vues en coupes des résultats des recalages rigides (coupes du haut) puis des recalages élastiques (coupes du bas) pour les patients numéro 1 et numéro 5. Bien que les résultats chiffrés nous enseignent que les recalages élastiques aboutissent à de meilleurs résultats que les recalages rigides, il est difficile de bien voir la différence sur les vues coupe à coupe. C'est pourquoi nous proposons la validation sur les rendus volumiques en mode union et surtout intersection.

La figure 5.9 (respectivement 5.10 à la page ci-contre) représente des rendus volumiques multivolume, en mode union puis intersection, des résultats du recalage élastique du patient 1 (respectivement du patient 5).



Figure 5.9 — Rendus volumiques sous deux angles d'incidence et deux modes de visualisation (union et intersection) des recalages élastiques des différentes IRM du patient 1 sur son image TDM.



Figure 5.10 — Rendus volumiques sous deux angles d'incidence et deux modes de visualisation (union et intersection) des recalages élastiques des différentes IRM du patient 5 sur son image TDM.

Si nous comparons les rendus volumiques relatifs au recalage rigide et ceux du recalage élastique, nous observons surtout que les intersections des volumes présentent moins de « trous », ce qui signifie que les surfaces des volumes coïncident mieux après recalage élastique qu'après recalage rigide. Ceci peut également se voir sur les images en mode union du patient numéro 5 sur la modalité DPR. En effet, le rendu rigide de la figure 5.6 (page 120) montre un défaut d'alignement pour la modalité DPR au niveau du nez, et ce défaut n'est plus présent sur les rendus du recalage élastique.

Notons enfin que, même si les rendus relatifs aux autres patients ne sont pas présentés dans ce mémoire, les résultats du recalage élastique sont toujours visuellement meilleurs que ceux du recalage rigide.

5.3.3 Résumé des résultats

Les résultats que nous venons de présenter, tant en terme de performance stochastique, qu'en terme de qualité visuelle du recalage nous ont permis de montrer l'application de notre algorithme sur une base variée d'images médicales.

Par ailleurs, pour chaque paire d'images nous avons effectué plusieurs recalages, ce qui nous permet de valider la robustesse de l'algorithme stochastique.

Le point le plus fort et le plus remarquable de l'algorithme de recalage présenté est sa rapidité. En effet, les temps de calculs présentés (moins de 5 minutes pour un recalage élastique, et moins de deux minutes pour un recalage rigide) sont très satisfaisants si l'on compare aux temps moyens rencontrés dans la littérature (entre 15 minutes et 20 heures pour le recalage élastique, et entre 5 minutes et 12 heures pour le recalage rigide).

Enfin, pour être complet, il reste encore à confronter la précision de nos recalages par rapport aux recalages effectués par les autres équipes de recherche (ce qui était le but initial du projet RREP). Cette comparaison fait l'objet du paragraphe suivant.

5.4 Comparaison

Le projet RREP était initialement destiné à comparer les diverses techniques de recalage rigide. Comme notre algorithme effectue des recalages élastiques, et que le calcul d'une transformation rigide (étape 1) n'est qu'une approximation intermédiaire, il est difficile de confronter nos résultats avec ceux des autres équipes de recherche. Néanmoins, par curiosité, nous avons envoyé les résultats obtenus par notre étape d'estimation du recalage rigide aux gestionnaires de la base de données. Ces résultats sont transmis sous la forme des positions (en millimètres) de 8 couples de points (sommets du VOI) se correspondant sur les deux images recalées. À partir de ces 8 couples de points, la transformation rigide les faisant correspondre est calculée, puis appliquée aux amers de contrôle qui sont uniquement connus par les administrateurs de la base de données. Une erreur de recalage est alors calculée par comparaison entre les positions transformées de ces amers et les positions réelles optimales (là encore, seulement connues des administrateurs de la base de données).

Le tableau 5.4 montre les résultats obtenus pour 20 méthodes différentes (dont la notre). Ces résultats sont donnés en terme de distance moyenne, distance médiane et distance maximale sur les amers en millimètres.

Les différentes méthodes employées par les autres auteurs sont résumées à la page http://www.vuse.vanderbilt.edu/~jayw/results.html. Les résultats que nous obtenons révèlent des erreurs moyennes de l'ordre de 3 à 4 mm, ce qui d'une part semble beaucoup par rapport aux estimations que l'on avait calculées (environ 2 mm), mais d'autre part largement acceptables si l'on considère que l'algorithme de recalage n'est utilisé que pour donner une approximation rapide⁷ pour initialiser l'algorithme d'appariement de points (étape 2).

⁷L'approximation rigide s'effectue en une trentaine de secondes avec notre algorithme génétique, à comparer

TDM	1	vers IR	M DP	TDM	vers IRN	A DPR	TDM	vers IR	M T1	- MQT	vers IRI	M T1R
moy. med. max. mo	med. max. mo	max. mo	mo	y.	med.	max.	moy.	med.	max.	moy.	med.	max.
2,38 $1,92$ $6,93$ $2,5$	1,92 $6,93$ $2,5$	6,93 2,5	2,2	8	1,71	5,95	2,13	1,62	6,35	1,91	1,41	5,86
2,04 $2,09$ $3,83$ $0,8$	2,09 $3,83$ $0,8$	3,83 0,8	0,8	39	0,81	2,50	1,90	1,53	6,69	1,03	0,72	3,81
2,54 $2,01$ $6,55$ $1,0$	2,01 $6,55$ $1,0$	6,55 1,	1,	69	1,11	5,32	$2,\!12$	1,63	6,05	1,22	0,93	2,61
10,86 $3,12$ $49,60$ $9,$	3,12 49,60 9,	49,60 9,	<u>б</u>	66	3,06	45,86	10,46	3,39	51, 81	11,68	3,38	48,26
3,14 $2,37$ $10,45$ $1,$	2,37 10,45 1,	10,45 1.	-	78	1,66	3,69	2,68	1,37	10,97	1,08	1,00	2,12
2,00 $1,94$ $4,05$ 0	1,94 $4,05$ 0	4,05 0	0	89	0,73	2,36	1,36	1,17	2,78	0,87	0,71	2,35
1,86 $1,67$ $5,07$ $1,$	1,67 5,07 1.	5,07 1.	Ţ	47	$1,\!46$	2,72	2,73	2,51	7,05	2,43	2,38	5,78
2,16 $2,01$ $5,03$ 1	2,01 $5,03$ 1	5,03 1	1	,13	1,01	2,93	1,81	1,64	4,87	1,66	1,52	3,26
1,76 $1,71$ $3,56$ 1	1,71 $3,56$ 1	3,56 1	1	.08	0,97	2,66	1,22	1,10	2,99	1,15	1,03	2,81
5,41 $4,15$ $18,97$ 3	4,15 18,97 3.	18,97 3	က်	,78	2,97	10,15	5,68	5,05	12,85	5,05	4,94	14,33
10,41 $4,00$ $59,00$ $10,$	4,00 $59,00$ $10,$	59,00 10,	10,	22	4,04	62,66	10,08	4,32	61, 43	11,43	$5,\!42$	60,64
3,06 $2,60$ $5,80$ $3,$	2,60 $5,80$ $3,$	5,80 3,	က်	00	2,95	5,35	2,72	2,56	$6,\!43$	2,45	2,75	4,59
2,67 $2,31$ $6,18$ $2,$	2,31 $6,18$ $2,$	6,18 2,	Ъ,	01	1,86	5,07	1,93	1,50	4,36	1,75	$1,\!43$	4,54
6,89 $7,80$ $13,86$ $5,$	7,80 13,86 $5,$	13,86 5,	າບູ	93	4,61	11,57	4,58	3, 32	10,39	4,71	3,40	9,61
1,96 $1,93$ $4,30$ 2.	1,93 $4,30$ 2.	4,30 2.	C)	,16	2,04	4,65	2,79	2,74	7,27	2,34	2,23	5,95
7,06 $5,46$ $22,19$ $6,$	5,46 22,19 6,	22,19 6,	ô,	74	5,48	22,14	6,73	5,24	21,77	7,67	5,90	22,24
4,36 $3,88$ $15,25$ $4,$	3,88 15,25 4;	15,25 4;	4	27	4,18	9,97	3,39	2,75	12,48	5,60	4,52	20,34
2,01 $2,04$ $4,56$ $0,8$	2,04 $4,56$ $0,8$	4,56 0,8	0,8	35	0,83	1,71	1,69	$1,\!40$	4,67	1,04	0.95	4,02
1,94 $1,69$ $5,19$ $1,$	1,69 $5,19$ $1,$	5,19 1,	Ι,	15	1,07	4,62	1,72	1,56	5,98	1,07	0,84	4,15
2,15 $1,93$ $4,18$ $0,3$	1,93 $4,18$ 0,	4,18 0,9	0,0	66	1,01	1,64	1,61	1,53	4,17	1,05	0,89	3,11

Tableau 5.4 — Comparaison des résultats de recalage rigides pour diverses méthodes.

À propos des résultats, notons simplement quatre remarques :

- En premier lieu, les méthodes qui donnent les moins bons résultats sont celles qui utilisent seulement une description surfacique des scènes (celles qui n'utilisent pas l'information volumique). Cette observation ne permet bien entendu pas de conclure que les méthodes orientées « surface » sont moins efficaces, mais seulement que les images misent à disposition sur la base de données ne sont pas adaptées à de telles méthodes. Notons de plus que les images de la base de données ont été retouchées pour supprimer les informations liées aux marqueurs, et que ces retouches perturbent considérablement plus les données surfaciques que les données volumiques.
- La comparaison se fait par rapport à une transformation optimale calculée par un ensemble de marqueurs. Cette transformation optimale n'est par conséquent qu'optimale au degré de confiance sur la position des marqueurs près. De plus, la transformation calculée est une transformation rigide, alors que rien n'assure que les marqueurs se correspondent physiquement par une transformation rigide.
- La transformation rigide demandée pour la validation se faisait avec les images IRM comme références, alors que nous nous avons considéré le problème plus difficile du recalage vers l'image TDM. Cette différence est sans doute responsable en grande part de l'imprécision de recalage rigide que nous obtenons vis-à-vis des autres équipes de recherche⁸.
- Selon les conclusions d'une publication récente [98] du groupe responsable de la base de données RREP, les confiances chiffrées obtenues par la méthode des amers pour chaque recalage ne sont pas suffisamment précises pour tirer des conclusions cliniques. La seule comparaison acceptable est une comparaison visuelle de la mise en correspondance des données, et cette comparaison ne peut s'effectuer que par des experts.

5.5 Synthèse du chapitre

Ce cinquième et dernier chapitre vient de présenter l'application de notre méthode de recalage 3D élastique par optimisation génétique sur une base de données partagée d'images cérébrales (projet RREP).

Après avoir présenté le projet RREP, et les images de la base, nous avons défini la méthodologie d'utilisation de la base. Le choix des images à recaler, puis l'application de deux types de recalage (rigide puis élastique) nous a permis de justifier l'apport du recalage élastique par rapport au recalage rigide.

La robustesse de l'algorithme d'optimisation stochastique est également démontrée puisque les résultats d'une série de recalages d'une même paire d'images sont comparables et satisfaisants.

Une comparaison de la précision des recalages avec les résultats d'autres équipes, avec toutes les réserves que l'on a pu y apporter, a également fait l'objet de ce chapitre. Malheureusement, cette comparaison n'a pu s'effectuer que sur le recalage rigide, alors que le rôle premier de notre algorithme est plutôt le recalage élastique. Par ailleurs, les recalages que nous avons effectués pour la comparaison l'ont été faits dans le sens IRM vers scanner et non

avec les autre méthodes mettant entre 10 minutes et 12 heures.

⁸Nous n'avons pas eu le temps d'effectuer la validation des recalages dans le sens scanner \rightarrow IRM, mais nous avons déjà remarqué qu'en effectuant le recalage de l'IRM vers le scanner, la précision du recalage rigide peut être améliorée en supprimant du volume d'intérêt du scanner les parties n'ayant pas d'homologue sur le volume IRM (les coupes du bas). Ce résultat a été de plus confirmé sur l'utilisation d'un jeu de données d'entraînement de la base RREP.

scanner vers IRM comme les autres équipes de recherche, ce qui présente un cas plus difficile. La comparaison dans le sens scanner vers IRM est actuellement en cours et, faute de temps, n'a pas pu être insérée dans ce mémoire.

Conclusion

L'EXPLOSION de la quantité d'informations disponibles en imagerie médicale, liée au développement des techniques d'acquisition et notamment de la précision des imageurs impose la mise en œuvre d'outils d'analyse adaptés. Parmi ces outils, le recalage 3D élastique tient une place nouvelle importante. En effet, le recalage élastique 3D, souvent considéré comme très coûteux en temps de calcul ne connaît un essor que très récent, notamment pour deux raisons. La première est bien entendu dérivée de l'accroissement incessant de la puissance de calcul et de la mémoire des ordinateurs. La deuxième, mise à contribution dans cette thèse, est liée à l'utilisation de nouvelles méthodes d'optimisation comme l'optimisation stochastique génétique.

Après avoir rappelé, au cours du premier chapitre, les diverses problématiques du recalage en imagerie médicale, cette thèse propose une mise en œuvre originale d'algorithmes génétiques, et de traitement de l'information structurelle extraite des images permettant de répondre au problème du recalage.

Les objectifs initiaux de ce travail qui étaient d'établir une technique de recalage multimodal élastique 3D robuste et rapide à l'aide d'algorithmes génétiques sont atteints. En effet, le problème de la multimodalité a été contourné élégamment par l'utilisation d'une segmentation appropriée des images, rendant possible la comparaison d'informations spatiales de natures différentes. La chaîne globale de traitement est articulée selon une structure hiérarchique permettant d'explorer et de réduire, étape après étape, l'espace de recherche du problème. La description des surfaces segmentées en terme de géométrie différentielle est, lorsqu'elle est possible, un atout supplémentaire en faveur de la rapidité et de la précision de l'algorithme. Enfin, l'utilisation en série d'un algorithme génétique effectuant un échantillonnage stochastique robuste de l'espace de recherche, et d'un algorithme de recherche locale est apparue comme étant non seulement obligatoire, mais optimale d'un point de vue temps de calcul.

La validation d'un recalage, aussi bien rigide qu'élastique, est toujours un problème délicat. Même si en dernier recours on devrait toujours se reporter à l'avis d'un expert médical, une technique de validation élégante et pertinente a été utilisée pour cette thèse. En effet, la validation de la qualité d'un recalage au niveau de l'interface air-chair est illustrée dans ce travail par le biais de rendus d'intersections surfaciques floues (rendu direct multivolume). Ces rendus sont très discriminants quant à l'adéquation des données surfaciques, et ce sans avoir recours à la segmentation utilisée pour le recalage proprement dit.

Le dernier chapitre de ce mémoire a mis l'accent sur l'application de la méthode développée sur une base de données de validation rétrospective de techniques de recalage (RREP). Cette application a été utile à deux niveaux. Tout d'abord elle a permis de tester la robustesse de la méthode sur un vaste échantillon d'images IRM et scanner. Ensuite, elle a permis de comparer la précision de la méthode par rapport à celles des autres équipes de recherche. Malheureusement, cette comparaison n'a pu être effectuée que sur des recalages rigides (type de recalage qui n'était pas l'objectif premier de notre algorithme), et dans des conditions sous-optimales (nous avons effectué le recalage de l'IRM vers le scanner, ce qui pour notre algorithme est plus compliqué que dans le sens contraire). Même si une validation et une comparaison au niveau des recalages élastiques reste encore à établir, cette comparaison des résultats du recalage rigide confirme néanmoins le comportement globalement robuste de l'algorithme.

Différentes perspectives sont envisageables pour développer et améliorer la méthode que nous venons de présenter. La voie la plus intéressante à explorer est sans doute, à mon avis, l'intégration d'informations sur les structures internes du cerveau au lieu de seulement considérer l'interface externe air-chair. L'amélioration constante de la résolution des scanner et des IRM devrait permettre, dans un futur proche, d'extraire facilement des informations de structures internes pertinentes communes aux deux modalités. Par ailleurs, les efforts peuvent également se diriger vers une amélioration de la distance inter-surfaces, ou inter-images rapide. Enfin, une dernière piste à explorer se situe sans doute du côté de la mise en œuvre d'un algorithme d'appariement par optimisation génétique avec plus de 8 couples (8 étant le nombre minimal pour calculer la transformation élastique globale de notre modèle), et éventuellement des pondérations de confiance pour chaque couple de l'espace de recherche.

Dans un contexte où il est nécessaire de filtrer et d'interpréter rapidement un flux toujours grandissant d'informations médicales, ce travail de recherche est susceptible d'apporter une solution intéressante. De plus, l'accroissement continuel de la puissance des machines et de la précision des appareils imageurs ouvre la voie vers des perspectives nouvelles de traitement de l'information médicale. Puisse ce travail de recherche servir et être réutilisé pour améliorer l'analyse des données médicales.
A Caractérisation des images médicales

Cette annexe est consacrée à la présentation des deux types d'images médicales (deux modalités) que nous avons utilisés pour cette thèse. Le premier type est la tomographie X et le deuxième est l'imagerie par résonance magnétique.

A.1 Tomographie X

ANNEXE

La tomographie X, également appelée TomoDensitoMétrie (TDM), ou plus usuellement scanner (en anglais *CT-scan* pour *Computed Tomography Scanner*), est la modalité d'images médicales volumiques la plus couramment utilisée. Elle utilise le même principe que les radiographies à rayons X, à savoir la mesure de la densité d'absorption des rayons X dans les tissus traversés. L'inconvénient majeur des radiographies est que les images ne sont que des projections 2D de mesures effectuées sur un volume. La philosophie de la tomographie X est de multiplier les mesures sous différentes incidences d'un objet et d'en déduire, par résolution d'un problème inverse,



les densités d'absorption en chaque point d'une coupe transversale de l'objet étudié. Par ailleurs, l'accumulation de coupes successives permet de créer, informatiquement, des données volumiques.

Les scanners X sont utilisés pour obtenir de l'information sur la tête (le cerveau), les poumons, le foie, la vésicule biliaire, le pancréas, la colonne vertébrale, les reins, et bien d'autres organes. Cette information est utile pour mettre en évidence des traumatismes, des caillots sanguins, des tumeurs ou autres abcès.

A.1.1 Acquisition des images scanner

Un scanner est normalement constitué d'un tube émetteur de rayons X et d'une série de récepteurs comme montré sur le schéma A.1 (ce schéma est inspiré de la thèse d'Arnaud Colin [19], ainsi que les explications à propos de l'acquisition d'images scanner). Des photons sont émis par le tube en direction de l'objet à observer. Ces photons peuvent, en fonction du matériau rencontré, soit être absorbés, soit être déviés (phénomène de diffusion), soit traverser librement l'objet. Les capteurs situés face au cône de radiation relèvent la quantité de photons transmis sous l'incidence courante.



Figure A.1 — Représentation schématique du fonctionnement d'un scanner.

Si l'on note E_0 l'énergie émise par le tube (énergie du faisceau collimaté en éventail), et E_t l'énergie captée par les récepteurs (ou énergie transmise), E_t est fonction de E_0 et de



l'atténuation (correspondant aux phénomènes d'absorption et de diffusion) induite par les tissus traversés. Cette atténuation peut être caractérisée par μ (coefficient d'atténuation linéaire), et la relation suivante peut être exprimée :

$$E_t = E_0 \exp\left(-\int\limits_0^L \mu(x)dx
ight)$$

où L correspond à la distance entre le tube émetteur et les récepteurs. Lors d'un examen scanner en mode pas à pas, on fait tourner successivement le couple tube émetteur, récepteurs autour de l'objet (sur 360 °) pour obtenir différentes mesures (projections). Un algorithme de reconstruction est ensuite nécessaire pour recréer la cartographie des densités d'absorption pour la section transversale en cours. Une fois cela effectué,

on déplace, si besoin est, l'objet perpendiculairement à la coupe acquise pour effectuer des mesures sur une nouvelle coupe.

Il est à noter que sur la plupart des scanners actuels, les récepteurs sont répartis sur toute la périphérie du scanner pour éviter d'avoir à les mettre en rotation (seul le tube émetteur doit bouger). L'acquisition d'une coupe sur un scanner nécessite en général moins d'une seconde. Un tel dispositif permet également d'effectuer des acquisitions par balayage spiralé (le lit du patient étant en translation constante.

A.1.2 Caractéristiques

Nous présentons ici les principales caractéristiques, les avantages et inconvénients de la tomodensitométrie X. Les informations présentées dans ce paragraphe ont été prises sur la page http://www.med.univ-rennes1.fr/cerf/edicerf/.

A.1.2.1 Unités Hounsfield

L'image fournie par reconstruction est une représentation indirecte calculée de la réalité. C'est-à-dire qu'elle distribue les coefficients d'atténuation calculés de la structure sur une représentation de la section d'une tranche anatomique de 1 à 10 mm d'épaisseur.

À chaque pixel est attribué un coefficient d'atténuation du rayonnement X à 140 kV. À ce coefficient, on a donné le nom de Hounsfield (H) avec pour repères :

$$- eau = 0 H$$

– air = -1000 H $\,$

Tous les tissus ont des valeurs caractéristiques :

- caillot : 5 à 25
- matière : grise 45 à 60
- matière : blanche 30 à 55
- muscle : 55 à 60
- -os médulaire : 80 à 150
- os cortical : 300 à 3000
- etc.

A.1.2.2 Artéfacts scanographiques

Les valeurs Hounsfield calculées peuvent ne pas correspondre à la réalité.

* Hétérogénéité apparente d'une plage homogène

Une zone large et anatomiquement homogène comme le foie ou la rate a une représentation hétérogène (pixel à pixel ou sur une région d'intérêt). Ceci est la conséquence de variations aléatoires de reconstruction, et du bruit lié au signal ; ce bruit est d'autant plus important que l'on cherche toujours à réduire la dose d'exposition et le nombre de photons utilisés. Ces variations seront lissées au mieux par un filtre de densité.

* Volume partiel

Le volume correspondant à un Pixel sur une épaisseur de coupe scanographique (10 mm) peut être hétérogène, air sur une partie, tissu mou sur l'autre. La valeur Hounsfield calculée sera donc intermédiaire aux valeurs des deux parties. L'effet de volume partiel est plus marqué lorsque la coupe est épaisse et les milieux en contact très différents.

* Flou de contours

Le volume partiel dû à la limite oblique ou courbe entre deux structures (vaisseaux, travée osseuse) crée un flou que l'on peut mettre en évidence en changeant de fenêtre d'affichage. Ce flou explique les différences de mesure de longueur lorsque l'on change la fenêtre d'affichage et donc la nécessité d'une rigueur de procédure pour mesurer des valeurs faibles comme l'épaisseur d'interligne articulaire, ou le calibre de vaisseaux.

 $\boldsymbol{*}$ Hyper ou Hypo-densité de voisinage

Le foie sur une coupe thoracique présente au voisinage des côtes des variations de densité bien vues sur une fenêtre étroite :

- -en arrière d'une côte = hypo-densité,
- face à l'espace intercostal = hyper-densité.

Cela est dû à un effet excessif d'un filtre de convolution censé corriger l'effet d'une trop grande différence d'atténuation locale. N'oublions pas que l'image scanographique est une reconstruction et non pas une vue directe des coupes.

* Image en étoile

Un élément anatomique très dense comme une prothèse orthopédique ou une obturation dentaire présente dans la coupe, apparaît entourée d'opacité en étoile qui peuvent interdire toute analyse lorsque la cause est de grande taille ou proche de la région intéressante. Le coefficient μ d'atténuation linéaire prend une valeur énorme, laquelle est hors de portée du calcultateur effectuant la reconstruction.

A.1.2.3 Visualisation 3D

La succession de coupes (en deux dimensions X et Y) parallèles, jointives, chevauchantes ou éventuellement peu séparées fournit des informations dans un espace à trois dimensions (X, Y et Z). De cet espace pourront être tirées des reconstructions dans deux ou trois dimensions.

- Tout d'abord on peut effectuer des reconstructions obliques ou courbes. Dans l'ensemble de données, en X, Y, Z, sont sélectionnés un plan ou une courbe ; la courbe peut passer, par exemple, par les pédicules vertébraux ou une surface articulaire.
- Ensuite on peut envisager des reconstructions surfaciques. Des isosurfaces du volume 3D peuvent être extraites (en utilisant la technique du « marching cube » par exemple) puis définies en terme de surfaces numériques (ensembles de voxels connectés par des relations topologiques de voisinage). Selon une méthode 3D la surface correspondante est isolée, elle peut alors être manipulée et projetée selon divers angles.
- Enfin on peut effectuer des reconstructions volumiques (voir également à ce propos l'annexe C). L'ensemble des données est rapporté à un certain nombre de voxels formant un volume total qui peut être alors traité comme un tout et soumis à des sections planes, courbes, de gradients de densité, etc. Cette procédure est beaucoup plus complexe que les précédentes et permet d'associer sur une même vue, volume et section complexe.

Un cas particulier est celui de l'imagerie MIP ou par transparence : le volume 3D peut être projeté dans toutes directions mais soit en faisant la somme de tous les pixels traversés reproduisant une radio classique, soit en ne retenant que le pixel de plus haute densité (Maximum Intensity Projection).

A.2 Imagerie par Résonance Magnétique (IRM)

Cette section présente le principe de l'imagerie par résonance magnétique. Les informations présentes dans cette annexe sont en majeure partie extraites des thèses d'Arnaud Colin [19], Fabien Feschet [30], ainsi que de la page Internet http://www.med.univ-rennes1.fr/cerf/edicerf/.

La résonance magnétique nucléaire (ou en anglais Magnetic Resonance Imaging) est basée sur la particularité de certains noyaux d'atomes, comme l'atome d'hydrogène (proton ¹H), de se présenter comme de petits dipôles qui réagissent donc à la présence d'un champ magnétique externe. Après avoir polariser avec un aimant les atomes d'hydrogène (fortement présents dans le corps humain), on les excite avec un champ magnétique perpendiculaire oscillant. Il se produit alors un phénomène de résonance renseignant sur la nature des matériaux auxquels appartiennent les protons observés (absorption d'énergie par les protons). Ces protons deviennent excités, et dans un deuxième temps réémettent de l'énergie pour revenir à un état stable. La variation de l'orientation



magnétique des protons est directement responsable de l'émission d'énergie de résonance magnétique. Il existe deux grands types d'images IRM. Les images dites pondérées en T1 et celles pondérées en T2, correspondant à deux moments différents de la relaxation d'énergie.

Cette modalité permet donc de connaître la densité et la nature des protons contenus dans les tissus observés.

A.2.1 Acquisition des images IRM



Un atome est constitué d'un noyau, lui-même constitué de nucléons (neutrons et protons) et d'électrons qui gravitent autour de ce noyau. Le noyau d'hydrogène est constitué d'un seul nucléon : 1 proton. Il possède une masse m et une charge e^+ . Ce noyau possède une propriété de « spin », assimilable au fait qu'il peut tourner sur lui-même, ce qui lui confère d'une part un moment cinétique \vec{J} , qui dépend de sa masse, et d'autre part un moment magnétique $\vec{\mu}$ qui dépend de sa charge. Ces deux moments sont alignés. Le facteur multiplicatif qui les unit γ est appelé le rapport gyromagnétique de spin nucléaire $(\vec{\mu} = \gamma \vec{J})$. Chaque noyau atomique, ayant par définition un nombre de nucléons qui lui est propre, a un rapport gyromagnétique qui lui est propre lui aussi. Placé dans un champ

magnétique $\vec{B_0}$ extérieur, continu, ce moment magnétique nucléaire $\vec{\mu}$ s'oriente par rapport à $\vec{B_0}$ dans des directions privilégiées (quantifiées) et décrit autour de $\vec{B_0}$ un mouvement de précession avec une vitesse angulaire $\omega_0 = \gamma B_0$ (cf. figure A.2).



Figure A.2 — Mouvement de précession de $\vec{\mu}$ autour de $\vec{B_0}$ à la vitesse ω_0 .

Pour une population donnée de noyaux, la somme de toutes ces petites aimantations qui précessent à des vitesses angulaires identiques, est susceptible de donner une aimantation non nulle $\vec{M_0}$ dans la direction du champ magnétique $\vec{B_0}$. Le noyau d'hydrogène dans un champ B_0 de 1 tesla a une fréquence de précession f_0 qui est de 42,57 MHz.

L'application d'un courant sinusoïdal à la fréquence de résonance de l'hydrogène dans une bobine d'axe perpendiculaire à $\vec{B_0}$ crée un champ magnétique $\vec{B_1}$ alternatif, qui perturbe cette

aimantation résultante $\vec{M_0}$: tout se passe comme si $\vec{M_0}$ tournait autour de $\vec{B_1}$ dans un plan perpendiculaire à celui-ci. L'angle de basculement est proportionnel à la quantité d'énergie électrique transmise à la bobine d'excitation, c'est-à-dire à la durée et à l'amplitude de cette excitation. On peut parfaitement déterminer un angle de basculement de 90°, ou 180°, en adaptant ces deux paramètres. Si la fréquence d'excitation n'est pas proche ou égale à la fréquence de résonance, le système ne modifiera pas sa position d'équilibre.

À l'arrêt de l'excitation, en supposant par exemple que l'on ait fait un basculement de 90°, l'aimantation résultante n'est plus dans la direction du champ $\vec{B_0}$ mais perpendiculaire à celuici. On a effectivement créé une composante transversale qui n'existait pas auparavant. À l'arrêt de l'excitation, cette aimantation revient à sa position d'équilibre en décrivant un mouvement complexe à la vitesse ω_0 . On peut la décomposer en une composante longitudinale qui va croissant vers sa position d'équilibre $\vec{M_0}$ et une composante transversale qui va décroissant vers sa valeur d'équilibre, c'est-à-dire 0. C'est cette composante transversale de l'aimantation (Mxy) qui, en tournant à la vitesse ω_0 , devant la bobine qui a servi à l'excitation et qui travaille maintenant en réception, va induire dans cette bobine un signal sinusoïdal amorti à la fréquence ω_0 : c'est le signal RMN.

Une expérience de RMN se résume donc à des opérations assez simples (cf. figure A.3) : il suffit de placer l'échantillon dans l'entrefer d'un aimant qui produit le champ $\vec{B_0}$. La fréquence de résonance d'un élément donné est alors parfaitement connue. Si on s'intéresse à l'hydrogène, il suffira d'envoyer dans une bobine dont l'axe est perpendiculaire au champ $\vec{B_0}$ un courant sinusoïdal à la bonne fréquence. À l'arrêt de l'excitation, cette même bobine recueillera un signal sinusoïdal amorti correspondant au retour à l'équilibre de la composante transversale ainsi créée lors de l'excitation initiale. La Transformée de Fourier permet d'avoir le spectre ou la raie de résonance.

Différentes séquences d'acquisition existent selon les temps de relaxation T1 et T2, et selon le temps de répétition TR (ces temps seront définis ultérieurement). Il existe par exemple la séquence dite *spin écho*, la séquence d'écho de gradient, la séquence d'inversion, récupération.

A.2.2 Reconstruction de l'image (écho de gradient)

La reconstruction de l'image IRM peut s'effectuer de diverses façons en utilisant les propriétés de la résonance magnétique nucléaire. Parmi les diverses méthodes de reconstruction, nous parlons ici de la méthode la plus utilisée, à savoir la séquence d'écho de gradient. Etant donnée la relation $\omega_0 = \gamma B_0$, on constate que pour chaque valeur du champ B_0 il existe une fréquence de résonance ω_0 pour le signal d'excitation. Si l'on crée un champ radio-fréquence d'excitation local B_1 à la fréquence $\omega_0 = \gamma (B_0 + \Delta B_0)$ on peut choisir quelle coupe (coordonnée sur l'axe 0z) du volume à observer va contribuer au signal RMN (cf. figure A.4), à condition de faire varier B_0 linéairement et de manière statique suivant l'axe z. Ceci nous donne une information planaire. Pour coder l'information dans une deuxième direction, on applique un gradient de champ magnétique après l'excitation, pendant l'acquisition du courant induit par la relaxation¹. Ceci à pour effet de faire varier la vitesse de précession des moments magnétiques linéairement selon la direction du gradient (relation de Larmor). Après analyse du spectre RMN on en déduit une information ligne à ligne dans l'image 3D (ce codage correspond à un codage par la fréquence). Enfin, pour avoir l'information RMN en chaque voxel du volume étudié, il faut avoir recours à une troisième technique de codage : le codage par la phase. Si l'application d'un gradient du champ B_0 dans une première direction

¹D'où le nom d'écho de gradient.



Figure A.3 — Principe d'une mesure RMN : L'échantillon placé dans le champ magnétique Bo acquiert une aimantation M_0 dans l'axe du champ B_0 . Une excitation B_1 à la fréquence de résonance du noyau étudié modifie cet équilibre. À l'arrêt de l'excitation, le retour de l'aimantation à sa valeur d'équilibre induit dans cette même bobine un signal à la fréquence de résonance : c'est le signal RMN ou FID (Free Induction Decay). Sa transformée de Fourier est le spectre RMN.

a permis d'associer à chaque fréquence du spectre RMN une position sur l'axe 0x, l'application avant excitation et après excitation d'un gradient perpendiculaire au premier, ayant pour conséquence de déphaser le signal linéairement en fonction de la position sur l'axe 0y permet d'associer à chaque phase du signal une position sur l'axe 0y.



Figure A.4 — Sélection d'une coupe particulière du volume à observer par sélection d'une fréquence de résonance et par modification de l'aimantation globale.

Pour l'acquisition de chaque ligne $z = z_0, x = x_0$, le gradient s'applique pendant une durée appelée temps de répétition TR. Le temps séparant l'excitation de l'acquisition est appelé temps d'écho TE, et enfin l'angle de bascule de l'aimantation est noté α .

L'image informatique obtenue a un contenu fonction des deux temps de relaxation T1 et T2. L'importance relative de ces deux composantes dépend principalement des valeurs de TR, TE et α .

A.2.3 Caractéristiques

A.2.3.1 Interprétation des contrastes en images RMN

Pour un élément de volume, l'intensité du signal acquis lors de l'enregistrement de l'écho de spin dépend de multiples facteurs :

- * **De paramètres intrinsèques** liés à la nature même des tissus contenus dans cet élément de volume :
 - Les paramètres principaux ρ , T1, T2 ont déjà été définis.
 - D'autres tels que la diffusion, la présence d'agents de contraste ou la vitesse des fluides circulants, peuvent intervenir.
- * De paramètres extrinsèques liés à l'appareil de mesure lui-même et à la séquence d'acquisition :
 - L'intensité du champ principal B_0 qui définit pour partie l'intensité du signal RMN,
 - Les séquences utilisées et dans chacune d'elles les paramètres de séquence.
 - Il convient de garder présent à l'esprit deux notions générales :
 - Le contraste image représente les variations relatives d'intensité de signal en différents points de l'image. Pour deux points d'une même image acquise dans une séquence donnée, les paramètres extrinsèques sont les mêmes pour tous les points de l'image et le contraste image dépend donc essentiellement des paramètres intrinsèques.
 - L'intensité du signal en un point image correspond en fait à la moyenne des intensités des signaux dans les volumes échantillons et cette moyenne tient compte de l'épaisseur de la coupe et elle peut être responsable d'effets de volume partiel.

A.2.3.2 La durée d'une séquence

La durée TA d'une séquence IRM dépend du temps de répétition TR, de la taille de l'image acquise (typiquement N = 256) et du nombre d'accumulations (utilisé pour augmenter la résolution). Pour un temps de répétition TR = 0.5s, une matrice de taille 256 et un nombre d'accumulation de 2, le temps d'acquisition est TA = $0.5 \times 256 \times 2 = 256$ s soit 4 minutes et 16 secondes.

A.2.3.3 Les produits de contraste en IRM

Le principe de ces agents que l'on injecte par voie veineuse aux patients est clair : il s'agit de modifier localement les paramètres intrinsèques des tissus là où ces agents de contraste iront se fixer préférentiellement. Ces produits doivent bien sûr être non toxiques, efficaces en concentration faible et si possible posséder une certaine spécificité au moins compartimentale, si ce n'est tissulaire.

Il apparaît difficile *in situ* d'envisager des modificateurs de la quantité d'hydrogène et c'est donc les temps de relaxation des tissus qui seront visés.

Il est important de souligner que contrairement aux produits de contraste iodés utilisés en tomodensitométrie par exemple, et où le paramètre mesuré (l'absorption aux rayons X) est effectivement modifié par la présence du produit iodé, en imagerie par RMN ce n'est pas le produit de contraste lui-même qu'on observe, mais ses effets sur l'aimantation des noyaux d'hydrogène qui se trouvent dans son environnement. Ceci est important pour comprendre qu'il n'y a pas nécessairement proportionnalité entre les concentrations de ces agents de contraste et les modifications de signal observées en RMN.

On distingue essentiellement deux grandes classes de produits de contraste, l'une largement validée aujourd'hui avec des agents dits paramagnétiques qui sont des complexes du gadolinium, et d'autres en développement dits super-paramagnétiques, à base d'oxyde de fer.

Les premiers vont induire dans leur environnement des hypersignaux sur des images pondérées en T1, exactement comme le faisait la methémoglobine dans les hématomes. Cet hypersignal se verra là où le produit de contraste diffuse : dans les vaisseaux bien sûr, dans les tissus inflammatoires où il existe une hyperhémie, ainsi que dans les lésions cérébrales où il existe des ruptures de la barrière hémato-encéphalique.

Des produits super-paramagnétiques ayant une spécificité hépatique, captés par le système réticulo-endothélial, ont été développés. Ils ont pour effet d'éteindre le signal du foie sain et de permettre une bien meilleure visualisation des nodules tumoraux métastatiques au sein de celui-ci.

ANNEXE B Segmentation morphologique 3D

Cette annexe a pour but de présenter la méthode de segmentation morphologique développée par Jean-José Jacq [48]. Cette méthode s'appuie essentiellement sur la *segmentation morphologique par lignes de partage des eaux (LPE)* dans le cadre de la théorie de la morphologie mathématique [85, 96].

La segmentation morphologique constitue la première étape de notre algorithme de recalage. Le but de cette segmentation est d'obtenir, à partir des images volumiques brutes (IRM et scanner), d'une part deux surfaces numériques correspondant à l'interface air/chair visible sur les deux modalités, et d'autre part une carte distance de la surface numérique à recaler.

Ces deux informations (surface numérique et carte distance¹) sont obtenues grâce à un procédé de propagation de fronts dans un milieu tri-dimensionnel.

B.1 Présentation générale

Segmenter une image (à *n* dimensions et à niveaux de gris) revient souvent à rechercher les brusques variations des niveaux de gris, et à regrouper les zones à valeurs homogènes. La technique la plus élémentaire consiste donc soit à seuiller directement les images (segmentation sur les niveaux de gris), soit à rechercher les maxima locaux du gradient de l'image (recherche des brusques changements de niveaux de gris). Les méthodes usuelles effectuent la recherche des extrema locaux à l'aide d'opérateurs locaux. En réalité, il n'est pas toujours possible de décider de l'appartenance d'un point à un minimum ou un maximum régional sur la seule base de son voisinage. Une méthode robuste doit donc être globale. Une méthode globale de recherche des extrema locaux est le partitionnement en bassins versants. Elle consiste à étiqueter des bassins versants, en faisant monter régulièrement le niveau d'eau dans chacun des bassins en débutant à partir de l'élévation correspondant au minimum global. Lorsque deux bassins versants se rencontrent, on met en évidence un maximum local (cf. figure B.1).

La construction, non seulement efficace mais aussi précise², d'une ligne de partage des eaux doit faire appel à des processus d'advection linéaire et non-linéaire. En d'autres termes, un algorithme de construction de la LPE résultera d'une généralisation des algorithmes calculant les distances par propagation de fronts.

¹Le calcul de la carte distance peut également s'effectuer à l'aide de l'algorithme de Saïto [84].

²En vue de tenir compte, notamment, de l'existence de plateaux.



Figure B.1 — Principe de l'immersion pour la détection de maxima locaux.

B.2 Définitions de base - graphe morphologique

B.2.1 Image numérique

Une image numérique I à niveaux de gris à n dimensions est définie comme une application de $D_I \in \mathbb{Z}^n$ à valeurs dans \mathbb{Z} .

Le domaine de définition (D_I) résulte d'une grille d'échantillonnage régulière isotrope ou anisotrope. Pour un échantillonnage isotrope on parle de grille « carrée ». Les échantillons sont à valeurs dans un intervalle fini de \mathbb{Z} , correspondant souvent à un encodage portant sur 8 ou 16 bits, signés ou non.

À des fins de généralisation, une image numérique peut également être définie par son ensemble X des points p de la grille discrète d'échantillonnage. Plus précisément, on peut associer à chaque image numérique un graphe non orienté G = (X, A) où A désigne les arêtes définies implicitement par le type de relation de voisinage (connexité) de la grille.

B.2.2 Voisinage et connexité

En 2D comme en 3D, différents types de voisinages ou de connexités peuvent être envisagés. Sur une image 2D, un point possède 4 voisins dans les directions des axes de la grille, et 4 autres voisins dans les diagonales. La 4-connexité considère qu'un point ne possède comme voisins que les 4 premiers voisins (direction des axes), alors que la 8-connexité considère en plus les points diagonaux.

En 3D, on définit couramment 3 types de relation de voisinage : la 6-connexité (voisins en directions des 3 axes de l'image), la 18-connexité considère les 18 voisins des plans orthogonaux, et la 26-connexité qui considère tous les voisins immédiats.

La choix a priori de la connexité est important car il détermine la topologie des frontières entre les régions connexes. En 2D, des régions construites en 8-connexité ont des frontières 4-connexes et vice versa. De même en 3D, des régions construites en 6-connexité génèrent des frontières (surfaces numériques) 26-connexes. Pour la segmentation morphologique des images du crâne humain, nous avons choisi des régions 6-connexes afin que les frontières générées par la LPE soient 26-connexes.

Par la suite nous dirons que deux sommets p et q appartenant à un graphe non-orienté G = (X, A) sont voisins, si l'arête (p, q) appartient au sous-ensemble A de $\mathbb{Z}^n \times \mathbb{Z}^n$. De même, l'ensemble des sommets voisins de p au sens du graphe G est défini par

$$N_G(p) = \{q \in \mathbb{Z}^n, (p,q) \in A\}, \forall p \in \mathbb{Z}^n.$$

Enfin, l'ensemble des sommets voisins à un ensemble de sommets D est défini, au sens du graphe G, par

$$N_G(D) = \{ p \in \mathbb{Z}^n \setminus D, \exists q \in D, (p,q) \in A \}, \forall D \in \mathbb{Z}^n.$$

Étant donné un ensemble de sommets $U \subset \mathbb{Z}^n$ et un sommet $p \in U$, nous appelons composante connexe $C_p(U)$ de U l'union de tous les chemins d'origine p inclus dans U.

B.2.3 Distance et transformation géodésique

Une distance ou une transformation sont dites géodésiques lorsque leurs domaines d'application se limitent à un sous-ensemble Y du domaine d'étude X; de manière plus imagée, le sous-ensemble Y joue le rôle d'espace de travail, le reste de l'espace est assimilable à une zone interdite. De façon plus générale, dans le domaine de l'image, le but de la géodésie est la mesure des formes exactes d'un objet. Elle opère en prenant pour espace de travail la composante connexe associée à l'objet.

En présence d'une métrique permettant de valuer les éléments de A, la longueur du chemin peut être évaluée comme la somme des longueurs des arêtes le composant (voir paragraphe B.3.1). La distance géodésique $d_G(p,q)$, au sens du graphe G au sein duquel est défini un domaine d'étude Y, entre deux sommets p et q de G est donnée par la longueur minimale des chemins (un chemin C de longueur l(C) = m est un (m+1) upplet de points de G liés deux à deux par des arêtes de G).

$$\forall p, q \in G / \quad d_G(p, q) = \begin{cases} \inf\{l(C), C \text{ chemin inclu dans } Y \text{ joignant } p \text{ et } q \text{ dans } G \} \\ +\infty \text{ si } \{C, \text{ chemin inclu dans } Y \text{ joignant } p \text{ et } q \text{ dans } G \} = \emptyset \end{cases}$$

La distance géodésique est un outil de base lors de l'extraction de caractéristiques propres à chacune des composantes connexes de l'espace de travail.

B.2.4 Graphe morphologique

La dualité image numérique, graphe non-orienté peut s'étendre à la définition d'un graphe morphologique (donnée par Luc Vincent [96]) dont la structure (X, A) est constituée d'un ensemble de sommets valués X et d'un ensemble d'arêtes A.

La distance entre sommets du graphe est induite par le contenu de l'ensemble des arêtes A (on parle de distance au sens du graphe sous-jacent). Ces arêtes peuvent éventuellement être valuées selon une métrique appropriée.

B.3 Distances

B.3.1 Distance et métrique

Nous avons défini précédemment le concept de distance $d_G(p,q)$ au sein d'un graphe G = (X, A). Cette longueur obéit à la définition mathématique de distance dans un ensemble X, à savoir : elle est positive, définie et symétrique ($\forall (p,q) \in X \times X, p \neq q \Leftrightarrow d_G(p,q) > 0$, $d_G(p,q) = d_G(q,p)$ et $d_G(p,p) = 0$).

La mesure de la longueur d'un chemin requiert une valuation des arêtes (p,q) du graphe (métrique). Lorsque A peut être défini de manière à englober $X \times X$, on se trouve dans le cadre particulier de la métrique euclidienne, et dans tous les autres cas on doit se contenter de métriques non-euclidiennes (par exemple la métrique de Chanfrein [12]).

B.3.2 Transformation distance

La transformation distance porte sur un graphe morphologique binaire. Elle a pour but d'affecter a tout sommet appartenant au fond du graphe, la distance qui le sépare du plus proche sommet appartenant à une structure. Bien qu'il existe des techniques rapides d'obtention de transformées distances, le moyen le plus simple est d'effectuer une succession d'advections linéaires sur un front initial formé par les contours des composantes connexes (dilatations) et d'étiqueter les nouveaux éléments de manière incrémentielle.

B.4 Squelette par zone d'influence (SKIZ)

B.4.1 Définitions

Soit G = (X, A) un graphe morphologique binaire réunissant n composantes connexes $\{X_i\}$ $i \in [1, n]$, la zone d'influence $iz(X_i)$ de la i^e composante connexe X_i est l'ensemble des sommets de X plus proches de X_i que de toute autre composante connexe de X:

$$iz(X_i) = \{ p \in X, \forall j \neq i, d_G(p, X_i) < d_G(p, X_j) \}.$$
(B.1)

Par définition, le squelette par zone d'influence de G est l'ensemble constitué des frontières des différentes zones d'influence. Cette définition, non équivoque dans le cas continu, peut en première approche être reformulée dans le cas discret comme l'union des frontières de chaque zone d'influence. Comme les sommets équidistants d'au moins deux composantes connexes n'appartiennent à aucune zone d'influence (relation (B.1)), ces derniers doivent être explicitement intégrés à l'ensemble des sommets frontière de sorte que la définition discrète donnée en première approche devienne complète. Le squelette par zone d'influence du graphe morphologique binaire G = (X, A) est alors défini par :

$$\left(\bigcup_{i=1,n} (D^1(\mathrm{iz}(X_i)) \setminus \mathrm{iz}(X_i))\right) \cup \{p \in X, \exists j \neq i, d_G(p, X_i) = d_G(p, X_j) \neq \infty\}.$$
 (B.2)

B.4.2 Étiquetage des composantes connexes

La construction du squelette par zone d'influence requiert un étiquetage préalable de chacune des composantes connexes du graphe morphologique binaire; cet étiquetage peut s'effectuer par propagation d'un front à partir d'un élément d'une composante connexe. À chaque rencontre d'une nouvelle composante connexe, on incrémente un compteur d'événements et on numérote ainsi différemment la région balayée par le prochain front. En procédant de cette manière, le nombre de visites des sommets pour l'étiquetage est minimisé : le voisinage d'un sommet du fond n'est jamais examiné et le voisinage d'un sommet appartenant à une composante connexe n'est visité qu'une seule fois.

B.5 Ligne de partage des eaux

B.5.1 Définitions

La segmentation par ligne de partage des eaux (LPE) consiste à simuler une immersion dans un espace à *n*-dimensions (pour nous, n = 3), immersion avec élévation du niveau de l'eau à vitesse constante. On montre que cette immersion peut se décomposer, dans le cadre des images numériques à niveaux de gris quantifiés (graphe morphologique), en une série d'advections linéaires. L'advection linéaire correspond à la propagation d'un front à vitesse constante et caractérise l'aspect dynamique de la construction du squelette par zone d'influence, c'est-à-dire la brique élémentaire de construction de la LPE.

Le squelette par zone d'influence géodésique a déjà été formulé au paragraphe précédent (voir relations (B.1) (B.2)). Cependant, l'espace de travail étant redéfini ici au sein du même graphe morphologique, il convient de rendre son existence explicite au niveau des notations en ré-exprimant les relations (B.1) (B.2) sous la forme :

$$iz_Y(B_i) = \{ p \in Y, \forall j \neq i, d_Y(p, B_i) < d_Y(p, B_j) \},$$

SKIZ_Y(B) = Y \IZ_Y(B) avec IZ_Y(B) =
$$\bigcup_{i=1,n} iz_Y(B_i),$$

où apparaît un espace de travail (ou espace géodésique) $Y \subseteq X$ du graphe morphologique G = (X, A) et où les *n* composantes connexes B_i vérifient $B_i \subseteq Y \forall i$.

Une immersion progressive du relief, telle que proposée par Luc Vincent, résulte en une découverte ordonnée des minima régionaux. Le seuil de G au niveau h est noté $T_h(G) = \{p \in X, G(p) \leq h\}$ et $Min_h(G)$ désigne l'ensemble des sommets appartenant aux minima situés à l'altitude h. En supposant que les valuations du graphe morphologique numérique s'étendent sur la plage $[h_{min}, h_{max}]$, à l'issue de l'application de la relation récurrente,

$$\begin{cases} X_{h_{min}} = T_{h_{min}}(G) \\ X_{h+1} = Min_{h+1} \cup IZ_{T_{h+1}(G)}(X_h), \quad \forall h \in [h_{min}, h_{max} - 1], \end{cases}$$

 $X_{h_{max}}$ constitue la réunion de tous les bassins versants de G = (X, A). Dans cette relation de récurrence, $IZ_{T_{h+1}(G)}(X_h)$ représente les zones d'influence géodésiques, ou plus précisément le sous-ensemble des points de $T_{h+1}(G)$ tels qu'il soit plus proche des composantes connexes X_h que n'importe quelle autre composante connexe $X_i, i \in [h_{min}, h-1]$. Autrement dit,

$$IZ_{T_{h+1}(G)}(X_h) = \{ p \in T_{h+1}(G), \forall i \neq h, d_{T_{h+1}(G)}(p, X_h) < d_{T_{h+1}(G)}(p, X_i) \}.$$

La figure B.2 décrit une immersion 2D portant sur 4 niveaux et 5 minima régionaux.

Par définition, la LPE est constituée de X privé de $X_{h_{max}}$. Ainsi, une LPE peut être discontinue et doit être complétée par un post-traitement à l'issue duquel des portions de frontières peuvent avoir une épaisseur supérieure à 1. La localisation d'une région (ou bassin) n'est pas ambiguë, par contre l'emplacement des frontières est toujours ambigu (on travaille sur un support discret).



(a) Niveau h = 0 : découverte de 3 minima régionaux.



(c) Niveau h = 2: Un bassin étend sa région d'influence; puis découverte d'un nouveau minimum régional.



(b) Niveau h = 1: Deux bassins étendent leurs régions d'influence; puis découverte d'un nouveau minimum régional.



(d) Niveau h = 4: Extension de tous les bassins; la partition est complète.

Figure B.2 — Exemple de formation récurrente de bassins versants et de la LPE lors d'une immersion portant sur 4 niveaux.

B.5.2 Mise en œuvre sur des images médicales

B.5.2.1 Problème de sursegmentation

Le principe de la segmentation morphologique, ou plus précisément de la détection de contours, comme nous l'avons expliqué précédemment, consiste à appliquer un algorithme de LPE sur une image gradient. Néanmoins, l'application directe d'une LPE sur une image gradient, même filtrée, induit souvent une sursegmentation.

La figure B.3 montre l'exemple (vues en coupe d'un volume) d'une application prématurée d'une segmentation par LPE sur une image volumique TDM de la tête. Le masque morphologique utilisé, servant à créer un volume d'interêt, est obtenu à l'aide d'un seuillage sur l'image à niveaux de gris (donnant une image binaire), suivi d'un gradient morphologique (dilatée érodée).



Figure B.3 — L'application prématurée de la LPE induit une sursegmentation.

La stratégie mise en place pour éviter les problèmes de sursegmentation est constituée de deux étapes. La première est l'application d'une reconstruction géodésique (en utilisant le masque morphologique) du gradient 3D avant l'application de la LPE. Ceci permet d'imposer la localisation et le nombre de minima régionaux à l'aide de *marqueurs*. La deuxième étape consiste à regrouper les quelques bassins versants résiduels issus de la LPE pour recréer une surface numérique continue.

B.5.2.2 Reconstruction géodésique

Le choix des marqueurs et, de manière duale, des bassins versants à inonder, est effectué par sélection des régions offrant une valuation (par exemple le gradient) inférieure à un certain seuil h_0 (le seuil de création du masque morphologique). Notons G_0 le graphe morphologique associé à l'image brute. L'ensemble $M(G_0)$ des sommets formés par les marqueurs est identifié par

$$M(G_0) = \{ p \in X, G_0(p) \leq h_0 \}.$$

Une fois appliquée la reconstruction géodésique (cf. figure B.4 pour une application en 1D), ces marqueurs constituent alors les minimaux régionaux de l'image.

Notons G_m le graphe binaire où les sommets correspondant aux marqueurs sont valués à zéro et les autres à l'infini. Le graphe reconstruit sur lequel sera finalement appliqué la LPE est alors défini par $G_r(p) = E^{\infty}_{(G_0 \wedge G_m)}(G_m(p))$ (érosion géodésique de taille infinie de G_m conditionnellement à $(G_0 \wedge G_m)(p) = \inf\{G_0(p), G_m(p)\}$. Cette reconstruction revient à inonder les bassins versants dans les régions du masque, et par conséquent de supprimer les minima locaux responsables de la sursegmentation.



Figure B.4 — Reconstruction géodésique du gradient pour éliminer les minima locaux parasites.

B.5.2.3 Regroupement de bassins versants

Le regroupement des bassins repose sur un critère de volume : les bassins possédant un petit volume fusionnent avec le bassin vis-à-vis duquel, ils possèdent la frontière la plus perméable. En pratique, le regroupement des bassins repose sur le résultat de la LPE calculée précédemment. Une dilatation morphologique de petite taille (typiquement 2) de cette LPE est utilisée comme masque de reconstruction géodésique du gradient préalablement reconstruit (G_r) . Nous obtenons alors une nouvelle reconstruction du gradient sur laquelle de nouveaux bassins versants potentiels ont été éliminés (ceux dont le volume n'est pas totalement comblé par la dilatation). Une LPE appliquée à ce nouveau gradient ne portera plus trace de ces petits bassins.



La figure B.5 explique l'algorithme global de segmentation utilisé.

Figure B.5 — Synopsis de la procédure de segmentation 3D des images TDM et IRM.

B.5.3 Autres exemples

Une autre application de la segmentation que nous avons utilisée est la détection et la modélisation automatique de surfaces articulaires [48, 50]. La figure B.6 montre un exemple de rendu volumique de l'épaule gauche avant et après segmentation LPE.



(a) Données brutes avant segmentation.



- (b) Données segmentées.
- $Figure \ B.6$ Segmentation LPE de l'épaule gauche.

ANNEXE C Visualisation : Rendu Direct Multi-Volume

Le rendu volumique constitue une branche jeune et très active de la visualisation scientifique, cette dernière s'inscrivant dans le vaste domaine de l'infographie. Historiquement, elle est née du besoin de procéder à l'analyse visuelle de données scientifiques multidimensionnelles, portant sur des phénomènes ou des structures naturelles, issues de nouvelles générations d'appareillage de mesure. D'ailleurs, l'imagerie médicale, et notamment la tomographie par rayon X (cf. annexe A), a joué un rôle prépondérant dans le développement du rendu volumique. Avec le développement en terme de précision, et la diversité des capteurs sur le marché, le volume important des données, la visualisation devient un point essentiel du traitement. Ce traitement, contrairement aux autres techniques d'infographie, a pour but d'extraire de la masse de données l'information pertinente afin d'aider à l'analyse et à la compréhension des images. La notion de réalisme est alors dépourvue de sens et doit être écartée au profit de la clarté du rendu et de la fidélité aux données.

Cette annexe a pour but de présenter une des catégories de rendus volumiques, le rendu volumique direct (DVR, signifiant *Direct Volume Rendering*), et plus particulièrement l'outil de rendu direct multivolume que nous avons utilisé pour la validation visuelle des recalages au cours de cette thèse. Cet outil à été développé à l'ENSTBR et a fait l'objet d'articles en revues internationales [47, 49].

Le rendu volumique direct opère directement sur les données brutes (voir paragraphe C.1), typiquement un champ de données scalaires, sans passer par une représentation géométrique intermédiaire (représentation surfacique par exemple). La projection est élaborée au travers d'un modèle physique (voir paragraphe C.2) de transfert des radiations. Le rendu volumique direct vise principalement les données texturées et/ou se prêtant difficilement à une description explicite sous forme de surface; certains auteurs le caractérise en parlant d'absence de segmentation explicite.

C.1 Caractérisation des données

Lorsque l'on parle de rendu volumique, il est important de commencer par caractériser les sources de données. Les données, scalaires ou vectorielles, définies dans un espace à trois dimensions (ou plus), proviennent essentiellement de simulations et de mesures issues du domaine des sciences naturelles, de la physique chimie, et de la médecine ou de la biologie.

L'imagerie biomédicale constitue la principale source de données volumiques. En tant que telle, elle représente la discipline d'où sont issues les contributions majeures en matière de technique de rendu volumique. Les principaux imageurs biomédicaux, tout au moins ceux concernés par cette thèse ont déjà été décrits à l'annexe A. Ajoutons néanmoins que d'une manière générale, les données disponibles sont décrites à l'aide d'une grille d'échantillonnage définie dans un espace à N dimensions; à chacun des nœuds est associé un échantillon scalaire ou vectoriel. Si, sous un angle de rendu volumique, la grille cartésienne constitue le cas idéal, les grilles réellement rencontrées sont malheureusement empreintes de la géométrie du dispositif de mesure, du modèle de description ou de l'inhomogénéité de la densité de l'information dans l'espace.

Pour les images médicales, le maillage est, en général, de type régulier, c'est-à-dire que les échantillons sont alignés suivant les axes d'un repère orthogonal et possèdent un espacement régulier propre à chacun des axes.

Un dernier point à souligner concerne le fait que les données médicales numériques disposées sous forme de grilles régulières ne correspondent pas forcément à l'agencement spatial des données brutes. Celles-ci peuvent répondre à un maillage physique plus complexe propre au dispositif de mesure, ou, dans le domaine de la tomographie, résulter de calculs portant sur des signaux 1D indépendants. Comme, en règle général, les données brutes demeurent inaccessibles à l'utilisateur, il est hasardeux de faire des hypothèses fortes sur la nature de l'échantillonnage régulier 2D ou 3D disponible. De plus, en imagerie médicale 3D, la troisième dimension est souvent rapportée par des artifices (TDM et IRM : coupes parallèles d'épaisseur fixe, à intervalles réguliers, obtenues soit par déplacement du support (TDM), soit par modification du champ d'aimantation (IRM)) introduisant une anisotropie de la résolution. En conséquence, aucune hypothèse solide ne peut être faite sur le degré réel de continuité de la fonction sous-jacente aux données et, par exemple, la validité d'une interpolation entre échantillons voisins n'est pas toujours garantie. Le rééchantillonnage propre à l'algorithme de rendu peut alors engendrer des artéfacts supplémentaires.

C.2 Fondements physiques et théoriques du DVR

En infographie, le cœur de tout algorithme de rendu volumique est le processus de projection : celui-ci réalise une intégration complexe (suivant des modèles non linéaires) visant la projection d'un jeu de données multidimensionnelles sur un plan 2D; d'un point de vue pratique, l'opération revient à définir la couleur des pixels, disposés suivant un maillage régulier, formant le plan de projection (*i.e.* l'image numérique). La formation des pixels peut se faire de deux manières différentes : dans un ordre défini par l'objet ou dans un ordre défini par l'image.

Dans le cas d'un ordre défini par l'objet, les éléments constitutifs sont projetés un à un sur le plan formé par l'image et, suivant l'étendue de la projection, modifient simultanément et non uniformément un ou plusieurs pixels.

Dans le cas d'un ordre défini par l'image, l'espace décrivant les données est rééchantillonné en des sites situés sur des rayons émis du point d'observation et émergeant des pixels du plan de projection ($Raycasting^1$). En première approximation, un seul rayon est émis par pixel; l'étude des échantillons situés sur la trajectoire du rayon permet de déterminer isolément l'énergie lumineuse parvenant au pixel.

Pour contourner les défaillances potentielles de toute décision binaire attachée à la

¹On utilise le terme de raycasting plutôt que de raytracing (en usage dans le domaine de la synthèse d'images), car ici seuls sont considérés les rayons primaires (pas de réflexions); ces derniers permettent d'échantillonner le champ de données 3D.

construction de modèles explicites de surface et, plus généralement, celles de la segmentation (ou de la classification) binaire, des approches basées sur une modélisation statistique ont été proposées : la description des structures demeure implicite et, grâce à la classification floue sous-jacente, la cohérence intrinsèque des données émerge lors du processus de projection au fil de la sommation des contributions élémentaires d'un grand nombre de sources pseudo-ponctuelles. Les règles de sommation reposent sur une modélisation inspirée de la théorie physique du transfert radiatif. Ce type d'approche particulaire a donné naissance aux algorithmes de rendu volumique direct (ainsi qu'à des modèles plus réalistes dans le domaine de la synthèse d'images). Dans un rendu de type DVR, l'observateur injecte la connaissance a priori (nombre et nature des classes en présence) dans le modèle attaché au comportement des particules à l'aide d'un paramétrage de propriétés physiques fictives attachées à chacun des matériaux.

C.2.1 Transport de particules : cas continu

La problématique attachée à la physique du transport de particules consiste à caractériser leurs distributions idéalisées en prenant en compte leur déplacements et leurs interactions avec le milieu; de tels problèmes se rencontrent dans l'étude de la propagation de la chaleur ou du rayonnement (par transport de photons) ou en physique nucléaire (déplacement des neutrons). Les équations gouvernant ces processus dérivent de l'équation de Boltzmann, ellemême issue de la théorie cinétique des gaz. Dans sa forme simplifiée, l'équation de Boltzmann linéaire peut s'écrire :

$$\Phi(P) = S(P) + \int_{\Omega} K(P' \to P) \Phi(P'),$$

où P représente un position; $\Phi(P)$ est la densité de radiation régnant en P; elle résulte d'une part de la source locale S et d'autre part d'une fraction des radiations issues de toutes les autres $P' \neq P$ décrites par l'angle solide Ω parvenant, au travers du milieu intermédiaire, en P; la fonction K est la fonction de diffusion. Dans les cas concrets, la résolution analytique de cette équation est très ardue, et seule des solutions approchées sont en général obtenues.

Afin de mieux cerner son adaptation au rendu volumique direct, ce paragraphe examine dans un premier temps l'application du modèle physique sous-jacent à la description du transfert du rayonnement lumineux au sein d'un nuage de densité inhomogène constitué de particules virtuelles sphériques lumineuses.

Les billes lumineuses sont supposées être de caractéristiques identiques et d'albédo négligeable (pas de réflexion). En négligeant ainsi les réflexions, nous limitons les interactions possibles à l'absorption (donc pas de diffusion non plus), et le problème du transport peut se reformuler de manière mono-dimensionnelle. La propagation du rayon lumineux ne dépend localement que de la densité $\rho(x, y, z)$ des billes et de leur géométrie ; afin de pouvoir traiter la densité comme un fonction continue de l'espace, la taille des billes est supposée suffisamment petite par rapport au volume élémentaire considéré.

D'après la définition de la densité $\rho(x, y, z)$, le nombre N_{Ω} de particules présentes dans une portion d'espace Ω s'écrit :

$$N_{\Omega} = \int_{\Omega} \rho(x, y, z) dx dy dz$$

À un rayon $[t_1 \rightarrow t_2]$ est attaché une portion d'espace cylindrique de section constante σ assimilée à la section efficace d'interaction des particules (voir figure C.1. Le nombre de

particules N_t associées à la portion d'espace $v_t = (t - t_1)\sigma$ représente la densité optique (ou, de manière équivalente, profondeur optique) du chemin correspondant et s'exprime sous la forme

$$N_t = \sigma \int_{t_1}^{t_2} \rho(t') dt'$$

en tant qu'intégrale des coefficients d'extinction $\sigma \rho(t) dt$.

On considère le rayonnement incident $dI_0(t)$ émis par les particules appartenant au sousvolume d'abscisse t et d'épaisseur dt. La fraction dI(t) transmise jusqu'au point d'observation est proportionnelle à la probabilité de ne rencontrer aucune particule sur la trajectoire $[t_1 \rightarrow t]$. En supposant la densité suffisamment faible, cette probabilité de zéro collision peut être modélisée à partir d'une distribution de Poisson

$$P(0; v_t) = e^{-N_t}$$

et représente la transmittance $\theta(t, t_1)$ du volume v_t associé à la portion de trajectoire $[t_1 \to t]$. Par définition, l'opacité correspondante du même intervalle est :

$$\alpha(t_1, t) = 1 - \theta(t_1, t) = 1 - \exp\left(-\sigma \int_{t_1}^t \rho(t') dt'\right).$$
 (C.1)

La brillance au point d'observation est la somme des contributions atténuées de toutes les sections :

$$B = \int_{t_1}^{t_2} P(0; v_t) dI_0(t) = \int_{t_1}^{t_2} \exp\left(-\sigma \int_{t_1}^t \rho(t') dt'\right) dI_0(t).$$
(C.2)

Etant donné que l'on ne considère que les interactions d'absorption (pas de réflexion), l'intensité lumineuse générée par un volume élémentaire s'exprime simplement comme le produit du nombre d'émetteurs inclus dans ce volume élémentaire par l'intensité κ d'un émetteur; l'expression précédente peut s'écrire sous la forme

$$B = \int_{t_1}^{t_2} \exp\left(-\sigma \int_{t_1}^t \rho(t') dt'\right) \sigma \rho(t) dt,$$

en établissant une relation de proportionnalité $dI_0(t) = \kappa \sigma \rho(t) dt \propto \sigma \rho(t) dt$ avec la densité locale.

Dans une modélisation moins simplificatrice (modèle de Kajiya et Blinn), $dI_0(t)$ dépend lui-même du rayonnement incident en t; sa prise en compte nécessite l'introduction d'un terme de phase précisant la fraction de la lumière incidente réémise dans la direction de Rpar l'élément de volume situé en t. La lumière provenant de l'unique incidence L_t est réfléchie par un élément de surface associé à chacune des billes (cf. figure C.1). Le modèle ne considère qu'un seul niveau de réflexion (les rayons secondaires sont seulement sujets à l'absorption); l'intensité de la source lumineuse est transmise intégralement suivant L_t (pas d'occlusion ni de dispersion) et le rayonnement ambiant interagit seulement par absorption totale lors de la rencontre d'une bille. Ce modèle constitue le fondement des algorithmes de rendu volumique direct. Son expression générale s'obtient en introduisant le terme de phase $\phi(t, R, L_t)$ dans la relation (C.2); d'où la relation

$$B = \int_{t_1}^{t_2} \exp\left(-\sigma \int_{t_1}^t \rho(t') dt'\right) \sigma\rho(t)\phi(t, R, L_t) dt$$
(C.3)



Figure C.1 — Modèle de Kajiya et Blinn.

consécutive à la nouvelle forme différentielle $dI_0(t) = \phi(t, R, L_t)\sigma\rho(t)dt$. L'expression analytique du terme de phase $\phi(t, R, L_t)$ est liée au choix du *modèle physique d'illumination* (en général on utilise le modèle standard de Phong); le terme de phase représente la quantité de lumière émise par une bille.

C.2.2 Passage au cas discret

La discrétisation du modèle de Kajiya et Blinn comporte deux aspects indépendants : la discrétisation du plan de projection et la discrétisation de l'interaction entre les éléments de volume. La discrétisation du plan de projection consiste principalement en un suréchantillonnage adaptatif du plan de projection. Autant de rayons sont tirés aux alentours d'un pixel du plan de projection que nécessaire pour obtenir une image sans artéfacts (notamment pour éliminer l'*aliasing*, ou crénelage). Plus délicat est le problème de la discrétisation le long du rayon (sujet de ce paragraphe), *i.e.* la discrétisation de l'équation du transport.

Suivant le modèle de Kajiya et Blinn, la discrétisation du processus physique le long du rayon peut être décrite de manière idéalisée en trois étapes :

- Reconstruction de la fonction suivant le rayon R par rééchantillonnage des données brutes;
- Transformation de la fonction suivant les grandeurs physiques locales (densité, paramètres du modèle d'illumination et émission lumineuse résultante) associées à la modélisation physique;
- Application de l'équation du transport.

Le rééchantillonnage le long du rayon R se fait en considérant le rayon comme une suite de *n* cellules de taille constante. La dimension des cellules est déterminée à partir de la résolution des données. Le chemin $[t_1 \rightarrow t_2]$ ainsi reconstitué suivant le rayon R est formé d'une suite de *n* volumes élémentaires d'épaisseur δt_i indexés par *i* variant de 1 à *n* pour *t* variant de t_1 à t_2 . Le pas d'échantillonnage δt_i choisi est lié à la résolution de l'information initiale ; dans la pratique, on considère un pas égal (ou inférieur) à la moitié de l'extension d'un voxel suivant la direction R.

Comme la contribution lumineuse d'un voxel virtuel n'est plus localisée sur une section infiniment mince, la discrétisation de l'équation du transport ne peut porter directement sur la relation (C.3); par contre, cette relation permet le calcul de la contribution lumineuse propre $I_0(\delta t_i)$ au voxel virtuel. Ce calcul donne, en supposant la densité et les paramètres du modèle d'illumination constants sur l'intervalle δt_i :

$$I_0(\delta t_i) = \int_0^{\delta t_i} \exp(-\sigma \int_0^t \rho(t')dt')\sigma\rho(t_i)\phi(t_i, R, L_{t_i})dt$$
$$= (1 - \exp(-\sigma\rho(t_i)\delta t_i))\phi(t_i, R, L_{t_i})$$
$$= \alpha(\delta t_i)\phi(t_i, R, L_{t_i})$$

Pour des pas consécutifs de longueur δt_i , la formulation discrète de la relation (C.2) est :

$$B = \sum_{i=1}^{n} \exp\left(-\sigma \sum_{j=1}^{i-1} \rho(t_i) \delta t_j\right) \alpha(\delta t_i) \phi(t_i, R, L_{t_i})$$

$$= \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \exp(-\sigma \rho(t_j) \delta t_j)\right) \alpha(\delta t_i) \phi(t_i, R, L_{t_i}).$$
 (C.4)

En introduisant l'identité (C.1) et après allégement de la notation, la relation (C.4) exprime la luminosité résultante du rayon R sous la forme fondamentale

$$B = \sum_{i=1}^{n} I_0(i) \left(\prod_{j=1}^{i-1} 1 - \alpha(j) \right),$$
(C.5)

avec $\alpha(i) = 1 - exp(-\sigma\rho(t_i)\delta t_i)$ et $I_0(i) = \alpha(i)\phi(t_i, R, L_{t_i})$. Cette dernière relation peut être reformulée de manière récursive de deux manières, selon que l'on considère une progression du rayon de type arrière vers avant (Back To Front ou BTF) ou de type avant vers arrière (Front To Back ou FTB). L'approche avant vers arrière est plus intéressante car dès que l'opacité cumulée le long du rayon est voisine de l'unité, on peut économiser des calculs en ignorant tout les voxels situés plus loin sur le rayon (qui ne seront de toutes façons pas visibles).

La formulation récurrente BTF peut s'écrire :

$$I(i) = I_0(i) + (1 - \alpha(i))I(i - 1)$$
 avec $i = 1, n$ pour $t = t_2, t_1$.

De même, la formulation récurrente FTB peut s'écrire :

$$I(i) = I(i-1) + (1 - \alpha_c(i-1))I_0(i) \quad \text{avec } i = 1, n+1 \text{ pour } t = t_1, t_2 + 1,$$

et $\alpha_c(i) = \alpha_c(i-1) + \alpha(i)(1 - \alpha_c(i-1))$ et $\alpha_c(0) = 0$.

Comme dans toutes les formules, σ est un paramètre constant – les particules diffèrent seulement par les propriétés attachées à leur micro-réflecteur – son existence demeurera implicite dans la suite de cette annexe. Afin d'éviter l'introduction d'une nouvelle variable, le produit $\sigma \rho$ sera noté ρ et sera identifiable à une *profondeur* (ou *densité*) *optique linéique*.

C.2.3 Composition locale

La composition locale vise le mapping des paramètres du modèle du nuage de particules luminescentes en chacun des échantillons du volume. Sous cette dénomination sont regroupés deux prétraitements locaux et consécutifs. Le premier positionne les paramètres indépendants du point de vue ; le second procède, sur la base de ces paramètres, à l'évaluation des quantités locales fonctions du point d'observation. Avant de pouvoir appliquer la relation de transport discrète (C.5), la profondeur optique linéique ρ ainsi que la luminosité ϕ d'une particule doivent être préalablement connues en chacun des échantillons du rayon. Notons S le vecteur des paramètres du modèle d'illumination². La dépendance de la luminosité particulaire peut être rendue explicite au travers de la notation $\phi(S, N, L, R)$, où R est le rayon d'observation et où N et L désignent respectivement la normale évaluée au centre de l'échantillon et la direction unitaire de la source lumineuse telle que vue depuis le centre de l'échantillon.

Pour un échantillon homogène, les quantités pouvant être déterminées indépendamment de l'angle de vue sont :

- la densité physique ρ_0 ,
- la normale N du champ de densité ρ_0 ,
- la profondeur optique liné
ique $\rho,$
- les paramètres S du modèle d'illumination.

L'extension δt que présente l'échantillon le long du rayon courant est, dans le cas général, fonction du point d'observation et/ou du pas d'échantillonnage propre au rayon courant. Les quantités dérivées, fonctions du point d'observation sont :

- l'opacité $\alpha = 1 \exp(-\rho \delta t)$ de l'échantillon suivant R,
- la contribution lumineuse $\phi(S, N, L, R)$ d'une particule de l'échantillon suivant R (que l'on peut abréger sous la forme $\phi(S)$ voire ϕ),
- la contribution lumineuse $I_0 = \alpha \phi(S, N, L, R)$ de l'échantillon suivant R.

Les paramètres du modèle se résument donc à $\{\rho_0, \rho, S\}$, dont la donnée constitue une signature de la substance locale. L'objectif premier de la phase de composition locale est d'établir, point à point, une correspondance entre les données brutes (f(x, y, z)) et la signature $\{\rho_0(x, y, z), \rho(x, y, z), S(x, y, z)\}$.

En se replaçant dans le cadre du modèle physique, on peut observer que les quantités ρ et ϕ sont par elles-mêmes des quantités moyennes et peuvent être vues comme résultant d'une combinaison linéaire de celles de m substances pures, de signatures $\{\rho_{0,i}, \rho_i, S_i, i = 1, ..., m\}$, potentiellement présentes dans l'élément de volume associé à un échantillon. Les caractéristiques d'un tel mélange peuvent être déterminées par l'intermédiaire d'une classification floue; la composition d'un volume élémentaire quelconque est alors définie par l'éventail des fonctions d'appartenance μ

$$\{\mu_i, i = 1, \dots, m\}, \quad \text{avec} \sum_{i=1}^m \mu_i = 1,$$

associé à ce même volume élémentaire. La fonction d'appartenance μ peut être obtenue par classification automatique ou être rapportée de façon extérieure à l'algorithme de rendu. Dans le domaine de l'imagerie biomédicale, sous réserve que l'échantillonnage soit effectué avec une résolution suffisante, il existe de manière générale au plus deux substances au sein d'un même voxel; pour répondre à cette spécificité, le classifieur doit alors incorporer des contraintes précisant les substances potentiellement adjacentes. La figure C.2 donne un exemple typique de classification floue en quatre substances d'une fonction scalaire représentant des données tomographiques X. En notant par V les valeurs que peut prendre une fonction scalaire f(x, y, z)et par $P_i(V)$ la probabilité que la substance i prenne la valeur V, la probabilité qu'un voxel

 $^{^{2}}$ Typiquement ces paramètres sont les triplets RGB des composantes ambiante, diffuse, et spéculaire des micro-réflecteurs des particules représentant le milieu.

quelconque possède cette valeur est

$$P(V) = \sum_{i=1}^{m} \mu_i P_i(V).$$

En supposant connue a priori la distribution $P_i(V)$ de chacune des substances, la fonction d'appartenance d'une valeur V peut être obtenue par estimation Bayesienne sous la forme

$$\mu_i(V) = \frac{P_i(V)}{\sum_{j=1}^m P_j(V)}.$$

De manière approchée, les fonctions d'appartenance peuvent être décrites sous la forme de fonctions linéaires par morceaux (profil trapézoïdal); pour se donner une classification a priori, les seuls paramètres requis sont, dans ces conditions, les largeur et position des segments support et noyau définissant la distribution trapézoïdale de chaque substance (cf. figure C.2(c)).



(c) Classification floue.

Figure C.2 — Classification floue sur 4 substances (coupes scanner X).

Sur la base de la classification floue μ_i , une estimée de la densité physique moyenne ρ_0 peut être obtenue en chacun des voxels par pondération de la densité physique propre $\rho_{0,i}$ des substances en présence

$$\rho_0 = \sum_{i=1}^m \mu_i \rho_{0,i}.$$

La normale N, ainsi que son module normalisé G, nécessaire à la mise en œuvre du modèle de réflexion, sont obtenus par application d'un opérateur gradient à la fonction densité :

$$ec{N} = rac{ec{
abla}
ho_0}{ec{
abla}
ho_0ec} \qquad G = rac{ec{
abla}
ho_0ec}{ec{
abla}
ho_0ec|_{
m max}}$$

Suivant des règles opératoires analogues à celles de la densité physique, la profondeur optique linéique ρ résulte d'une somme pondérée de celles (ρ_i) des substances en présence :

$$\rho = \sum_{i=1}^{m} \mu_i \rho_i.$$

Le problème central de la visualisation est de traduire la présence d'interfaces (existence d'un gradient non-négligeable au sein d'un champ de densité physique) entre milieux homogènes. Afin d'effectuer des rendus surfaciques flous, et dans le cadre de cette thèse, afin de vérifier visuellement la bonne corrélation entre deux « surfaces », la profondeur optique linéique moyenne du milieu associé à l'échantillon peut être modulée par l'intensité du gradient local G suivant la relation :

$$\rho = G \sum i = 1^m \mu_i \rho_i.$$

Toutes les particules présentes au sein d'un même échantillon constituent un milieu de propriété homogène; dans le cas de données biomédicales, on suppose que ce mélange se réduit localement à celui de deux substances prépondérantes. Plutôt que de former un vecteur paramètre S du modèle de réflexion en tant que vecteur moyen de celui des substances pures en présence, il est préférable, pour faciliter l'interprétation du rendu, de choisir le vecteur paramètre d'une de ces substances. La solution retenue pour ce choix, est de prendre le vecteur paramètre de la substance apportant la contribution majoritaire à l'opacité de l'échantillon, soit :

$$S = S_j$$
 avec j tel que $\mu_i \rho_i \leqslant \mu_j \rho_j, \forall i \in [1, m]$

C.3 Exemples de rendus

Le rendu volumique étant utilisé comme outil de validation pour le recalage dans cette thèse, on trouve de nombreux exemples de rendus volumiques au cours de ce document (voir par exemple, la figure 4.16 de la page 106).

Notons par ailleurs, qu'en modifiant les paramètres définissant les matériaux présents, on peut effectuer des rendus avec un effet de transparence. La figure C.3 représente un volume Scanner de la main avec 3 substances dont la chair semi-transparente.

C.4 Extension au rendu multi-volumique direct

Le passage au traitement de volumes multiples se fait par extension directe du traitement mono-volumique en considérant les différents matériaux en présence dans chacun des volumes et en définissant des règles de mélange entre les différents volumes. L'ensemble des paramètres du problème est donc désormais le suivant :



Figure C.3 — Visualisation avec effet de transparence d'un volume scanner de la main.

- M, nombre de modalités (ou nombre de volumes);
- m(j), nombre de matériaux dans la modalité j;
 - $\rho_{0,i}^{j}$, densité physique du matériau *i* de la modalité *j*;
 - ρ_i^j , densité optique linéique du matériau *i* de la modalité *j*;
 - μ_i^j , fonction d'appartenance associée au matériau *i* de la modalité *j*;
 - S_i^j , vecteur des paramètres décrivant les propriétés des micro-réflecteurs associés au matériau i de la modalité j.

À l'encontre du cas mono-volumique, le rendu multi-volumique ne peut procéder à l'intégration du rayonnement au fil de la progression du rayon. L'ensemble de la procédure liée au traitement d'un rayon [47, 49] doit être scindé en trois étapes successives :

- collecte séparée des signaux bruts 1D associés à chacune des intersections du rayon avec un volume; chacun des signaux est échantillonné avec un pas spécifique découlant de la résolution du volume considéré; les règles de composition locale sont ensuite appliquées au sein de chacun des échantillons;
- 2. rééchantillonnage synchrone de l'ensemble des signaux, puis application des règles de composition inter-modalité en chacun des échantillons;

3. application de l'équation du transport au signal physique 1D résultant de la composition locale.

La première étape constitue donc une traversée passive de l'ensemble de l'espace associé au repère du monde. Il s'agit d'un rééchantillonnage séparé de chacune des intersections du rayon courant avec chacun des volumes. Il en résulte une suite de P ($\leq M$) signaux 1D alignés et pouvant partiellement se chevaucher. À l'issue de l'application des règles de composition locale, les P segments représentent une suite d'échantillons vectoriels $\langle \rho, \phi \rangle$. L'ensemble de ces échantillons est totalement défini après évaluation de l'ensemble

$$\{l(p), \delta t^p, t^p_0, t^p_1, \{\phi^p(S(k)), \rho^p(k), k = 1, \dots, l(p)\}, p = 1, \dots, P\}$$

avec :

P, nombre de segments (< nombre de volumes);

l(p), nombre d'échantillons le long du segment p;

 δt^p , extension d'un échantillon le long du segment p;

 t_0^p, t_1^p , abscisse des deux extrémités du segment p suivant le rayon d'observation ;

 $\phi^p(S(k))$, intensité d'un émetteur de l'échantillon k du segment p;

 $\rho^p(k)$, densité optique linéique effective de l'échantillon k du segment p.

L'étape suivante vise la reconstruction d'un signal 1D unique, représentatif de la composition des P segments, auquel pourra être appliqué l'équation de transport. De manière imagée, ce signal résultant peut être vu comme la concaténation de caissons de taille identique contenant chacun un gaz luminescent homogène de densité et de spectre différents. L'origine, l'extrémité et l'épaisseur des échantillons du signal 1D équivalent sont respectivement déterminés par :

$$t_0 = \min_{p=1,\dots,P} t_0^p; \ t_1 = \max_{p=1,\dots,P} t_1^p; \ \delta t = \min_{p=1,\dots,P} \delta t^p.$$

Une fois chacun des rayons rééchantillonnés (la profondeur optique totale de chaque segment demeurant inchangée), il reste à établir des règles de mélange opérant point à point sur les échantillons homologues des segments. Il s'agit d'une opération dont la définition est principalement gouvernée par l'objectif visuel à atteindre. Il est toutefois possible d'établir les bases d'une *algèbre de recomposition* à partir de quelques opérations élémentaires, portant simultanément sur la profondeur optique linéique et la luminescence d'une particule du mélange, telles que la minimisation, la maximisation, et la moyenne :

$$S(k)) = \phi^{p_1}(S(k)) \text{ avec } p_1 = \arg(\max_{p=1,\dots,P} \rho^p(k))$$
(C.6)

maximum des contributions :
$$\begin{cases} \rho(k) &= \max_{p=1,\dots,P} \rho^p(k) \\ \phi(S(k)) &= \phi^{p_1}(S(k)) \text{ avec } p_1 = \arg(\max_{p=1,\dots,P} \rho^p(k)) \end{cases} (C.6)$$

minimum des contributions :
$$\begin{cases} \rho(k) &= \min_{p=1,\dots,P} \rho^p(k) \\ \phi(S(k)) &= \phi^{p_2}(S(k)) \text{ avec } p_2 = \arg(\min_{p=1,\dots,P} \rho^p(k)) \end{cases} (C.7)$$

moyenne des contributions :
$$\begin{cases} \rho(k) &= \frac{1}{p} \sum_{p=1}^{p} \rho^p(k) \\ \phi(S(k)) &= \frac{\sum_{p=1}^{P} \rho^p(k) \phi^p(S(k))}{\sum_{p=1}^{P} \rho^p(k)} \end{cases} . (C.8)$$

(C.7)

En tant que cas particuliers, les règles non linéaires (C.6) et (C.7) engendrent des effets similaires aux opérations logiques sur des ensembles flous (union et intersection).

Un des intérêts majeurs du rendu multi-volumique direct réside, en imagerie biomédicale, dans la faculté de procéder à la validation visuelle du recalage entre différentes modalités d'une même scène tout en considérant une segmentation (floue, et donc plus représentative de la complexité intrinsèque des données brutes) découplée de celle, explicite, sur laquelle s'est appuyé l'algorithme de recalage. On peut ainsi contrôler le degré d'ajustement d'un modèle en procédant à une visualisation multi-volumique directe des données et de la voxellisation floue du modèle [48]. De même, il est possible d'obtenir une validation visuelle très fine du recalage de deux volumes à l'aide d'une règle de mélange (C.7) simulant une intersection floue des structures recalées [82].

Là encore, la thèse présente de nombreux exemples d'utilisation des rendusmultivolumiques pour vérifier la qualité des divers recalages. Nous montrons à la figure C.4 une autre utilisation de l'outil de rendu multivolumique, à savoir la validation de la modélisation de la surface articulaire. Deux objets sont alors représentés : d'une part les données brutes issues d'un scanner, et d'autre part un volume artificiel modélisant l'articulation [48, 50].



 $Figure \ C.4$ — Rendu multivolumique en mode union de données scanner et d'un modèle théorique de l'articulation de l'épaule.

Liste des figures

1.1	Synoptique général d'un algorithme de recalage	7
1.2	Exemples de transformations 2D	9
1.3	Paire de coupes TDM et IRM recalées	16
1.4	Application du lissage sur une coupe scanner.	17
1.5	Détection d'une crête.	17
1.6	Disposition géométrique des isophotes et des lignes de flux (<i>flowline</i>) vis-à-vis de l'arête (<i>ridge</i>).	18
1.7	Calcul de la distance de la « tête » au « chapeau »	20
1.8	Modèle déformable et attracteurs	27
1.9	Modèle déformable et démons	28
1.10	Comparaison de la convergence avec et sans démons.	29
1.11	Fonctionnement des démons.	29
1.12	Vitesse instantanée de g à f	30
1.13	Hill Climbing : convergence vers un minimum local	32
1.14	La descente avec inertie.	32
1.15	La descente stochastique.	35
1.16	Fonctionnelle très bruitée.	35
1.17	Principe du recuit-simulé, « filtre de hasard »	36
1.18	Vue en coupe d'une image scanner et d'une image IRM	38
1.19	$f_1(\text{Scanner}, \text{IRM})$ et $f_1(\text{Scanner}, \text{IRM recalée})$.	38
1.20	$f_2(\text{Scanner}, \text{IRM})$ et $f_2(\text{Scanner}, \text{IRM recalée})$	39
1.21	$f_3(\text{Scanner}, \text{IRM})$ et $f_3(\text{Scanner}, \text{IRM recalée})$	39
1.22	$f_4(\text{Scanner}, \text{IRM})$ et $f_4(\text{Scanner}, \text{IRM recalée})$	40
1.23	$f_5(\text{Scanner}, \text{IRM})$ et $f_5(\text{Scanner}, \text{IRM recalée})$	40
1.24	$f_6(\text{Scanner}, \text{IRM})$ et $f_6(\text{Scanner}, \text{IRM recalée})$.	41
1.25	$f_7(\text{Scanner}, \text{IRM})$ et $f_7(\text{Scanner}, \text{IRM recalée})$.	41
1.26	Rendus effectués avec l'outil dvr.	42

2.1	Principe général de compensation des erreurs d'acquisition par transformation rigide et élastique.	46
2.2	2 Illustration 2D du recalage par amers fixes et espace de recherche contraint (hypercube)	54
2.3	B Exemples de déformations 2D sur l'image du manchot	56
2.4	Exemples de déformations 3D sur un volume synthétique	57
2.5	5 Exemple de carte de distances d'une surface 3D	60
2.6	Forme de la fonction de lissage exponentielle avec $\sigma = 4$	61
2.7	Vue en coupe d'une carte de distances avant et après changement des niveaux de gris $\sigma = 4$	62
3.1	Utilisation d'un algorithme génétique dans un problème d'optimisation.	65
3.2	2 L'encodage naturel d'un chromosome : un vecteur binaire	66
3.3	B La reproduction.	67
3.4	La mutation.	68
3.5	6 La recombinaison (ou <i>crossover</i>).	68
3.6	6 Crossover cyclique ou bi-points.	69
3.7	7 Cycle complet définissant un algorithme génétique.	70
3.8	Fonction test 2D à maximiser.	72
3.9	Évolution de l'algorithme génétique pour la maximisation d'une fonction 2D.	73
3.1	0 Exemple de répartition uniforme sur le principe des carrés Latins en 2D	75
3.1	11 Mutation locale dans un espace de recherche contraint	76
3.1	2 Encodage du chromosome sur l'espace de recherche indexé	77
3.1	3 Principe de la mutation uniforme sur un espace de recherche indexé	78
3.1	4 Principe de la recombinaison uniforme sur un espace de recherche indexé	78
4.1	Structure hiérarchique de l'algorithme mis en œuvre	84
4.2	2 Images scanner et IRM séquence flash mises à disposition par l'université de	
	Clermont-Ferrand.	85
4.3	B Principe de fonctionnement du recalage rigide par algorithme génétique	86
4.4	Principe du choix des couples de points.	89
4.5	Section normale.	89
4.6	Frincipe d'accélération du parcours des listes de points pour la création de l'ensemble des couples \mathcal{L}	91
4.7	Principe de fonctionnement de l'échantillonneur stochastique de l'espace de	0.0
	recherche.	92
4.8	L'algorithme genetique vu comme un échantillonneur stochastique.	94
4.9	Amnement de la solution finale (post-analyse).	97
4.1	U Mise en cascade des trois étapes du récalage.	99
4.1	II Performance maximale du recalage rigide	102

4.12	Performance maximale moyenne avec recalage élastique	102
4.13	Performance du recalage élastique à nombre d'itérations rigides constant	103
4.14	$Performance\ maximale\ moyenne\ avec\ recalage\ \'elastique\ et\ post-analyse.\ .\ .$	104
4.15	Performance avec post-analyse à nombre d'itérations rigides constant. \ldots	104
4.16	Données médicales avant recalage.	106
4.17	Résultats du recalage rigide (étape 1).	106
4.18	Résultats du recalage élastique (étape 2).	107
4.19	Résultats finaux (une seule classe de courbures).	107
4.20	Résultats finaux (trois classes de courbures)	108
5.1	Quelques rendus volumiques et quelques coupes des volumes médicaux du pa- tient numéro 1 dans la base de données RREP.	112
5.2	Rendus volumiques directs des volumes TDM et IRM T1 des 7 premiers patients de la base de données RREP.	113
5.3	Vues en coupe axiale des recalages rigides médians du patient 1	118
5.4	Vues en coupe axiale des recalages rigides médians du patient 5	118
5.5	Rendus volumiques des recalages rigides médians du patient 1	119
5.6	Rendus volumiques des recalages rigides médians du patient 5	120
5.7	Vues en coupe axiale des recalages élastiques du patient 1	123
5.8	Vues en coupe axiale des recalages élastiques du patient 5	123
5.9	Rendus volumiques des recalages élastiques du patient 1	124
5.10	Rendus volumiques des recalages élastiques du patient 5	125
A.1	Représentation schématique du fonctionnement d'un scanner.	134
A.2	Mouvement de précession de $ec{\mu}$ autour de $ec{B_0}$ à la vitesse ω_0	137
A.3	Principe d'une mesure RMN	139
A.4	Sélection d'une coupe dans un volume IRM	139
B.1	Principe de l'immersion pour la détection de maxima locaux.	144
B.2	Exemple de formation de bassins versants et de la LPE	148
B.3	L'application prématurée de la LPE induit une sursegmentation.	149
B.4	Reconstruction géodésique du gradient pour éliminer les minima locaux	150
B.5	Synopsis de la procédure de segmentation 3D des images TDM et IRM	151
B.6	Segmentation LPE de l'épaule gauche.	152
C.1	Modèle de Kajiya et Blinn.	157
C.2	Classification floue sur 4 substances (coupes scanner X)	160
C.3	Visualisation avec effet de transparence d'un volume scanner de la main	162
C.4	Rendu multivolumique en mode union de données scanner et d'un modèle théo- rique de l'articulation de l'épaule	164
Liste des tableaux

3.1	Paramètres de l'algorithme génétique permettant de maximiser $u \times v$	73
4.1	Paramètres utilisés et résultats du recalage rigide	87
4.2	Paramètres utilisés et résultats du recalage élastique	93
4.3	Résultats généraux du recalage élastique sur les images du laboratoire TIMC.	98
4.4	Performance maximale moyenne et nombre d'itérations nécessaires	100
51	Los imagos disponibles dans la base de données	111
J.1	Les mages disponibles dans la base de données.	111
5.2	Résultats des recalages rigides optimaux obtenus avec optimisation génétique.	116
5.3	Résultats des recalages élastiques obtenus avec optimisation génétique	122
5.4	Comparaison des résultats de recalage rigides pour diverses méthodes	127

Bibliographie

- J.-L. AMAT et G. YAHIAUOI, Techniques avancées pour le traitement de l'information, Cépaduès Éditions, avril 1995.
- [2] F. ARMAN, Model-based object recognition in dense-range images a review, ACM Computing Surveys, in press, vol. 25, no. 1, pp. 5–43, March 1993.
- [3] K. ARUN et T. HUANG, Least-squares fitting of two 3-D point sets, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, (pp. 698–700), PAMI, September 1987.
- [4] R. BAJCSY et S. KOVACIC, Multiresolution elastic matching, Computer Vision, Graphics, and Image Processing, vol. 46, pp. 1–21, 1989.
- [5] G. BAREQUET et M. SHARIR, Partial surface and volume matching in 3 dimensions, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, PAMI, September 1997.
- [6] H. G. BARROW et AL, Parametric correspondence and chamfer matching: Two new techniques for image matching, dans Proceedings of the 5th International Joint Conference on Artificial Intelligence (edité par R. REDDY), (pp. 659–663), MA, Cambridge, August 1977.
- [7] P. J. BESL, *The free-form surface matching problem*, Machine Vision for Three-Dimensional Scenes, (pp. 25–71), 1990. Edited by Herbert FREEMAN.
- [8] P. J. BESL et N. MCKAY, A method of registration of 3-D shapes, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 14, (pp. 239–256), PAMI, 1992.
- [9] F. L. BOOKSTEIN, Principal warps : thin plate splines and the decomposition of deformations, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, PAMI, June 1989.
- [10] F. L. BOOKSTEIN, Thin plate spline and the atlas problem for biomedical images, Information Processing in Medical Imaging, (pp. 326–342), 1991.
- [11] G. BORGEFORS, An improved version of the chamfer matching algorithm, dans IA-PR'84, 7th International Conference on Pattern Recognition, numéro 2, 1984.
- [12] G. BORGEFORS, Hierarchical chamfer matching : A parametric edge matching algorithm, vol. 10, (pp. 849-865), PAMI, novembre 1988.
- [13] M. BRAMELETTE, Initialization, mutation and selection methods in genetic algorithms for function optimization, dans Fourth International Conference on Genetic Algorithms, San Diego, CA, 1991.
- [14] M. BRO NIELSEN et C. GRAMKAW, Fast fluid registration of medical images, dans VBC'96, (pp. 267–276), 1996.

- [15] L. BROWN, A survey of image registration techniques, ACM Computing Surveys, in press, vol. 24, no. 4, pp. 325–376, 1992.
- [16] BRUNSTRÖM et STODDART, Genetic algorithms for free-form surface matching, dans ICPR'96, March 1996.
- [17] R. CHELLAPPA et C. LIN, Classification of partial 2d shapes using finite fourier descriptors, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, (pp. 686-690), PAMI, September 1987.
- [18] P. CHIRON, Mise en correspondance et critères de similitude en imagerie médicale, Thèse de doctorat, Université de Nantes, 1989.
- [19] A. COLIN, Étude de méthodes de recalage et de fusion d'images 3D du cerveau. Application au suivi d'une pathologie, Thèse de doctorat, Université de Clermont Ferrand, 1997.
- [20] A. COLLIGNON, 3-D multimodality medical image registration using feature space clustering, dans Proc of CVRMed'95, vol. 905, (pp. 195–204), 1995.
- [21] A. COLLIGNON, Automated multimodality image registration using information theory, dans Information Processing in Medical Imaging, (pp. 263–275), Brest, 1995.
- [22] A. COLLIGNON, Multimodality image registration by maximisation of mutual information, IEEE Transaction on Medical Imaging, vol. 16, no. 2, March 1997.
- [23] COLLINS et EVANS, Multi-resolution image registration and brain structure segmentation, dans IEEE EMBS France – Rennes, 1992.
- [24] G. COX et G. DE JAGER, A survey of point pattern matching techniques and a new approach to point pattern recognition, dans Proc of the 1992 South Africa Symposium on Communication and Signal Processing, (pp. 243–248), 1992.
- [25] E. DE CASTRO et C. MORANDI, Registration of translated and rotated images using finite fourier transforms, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, (pp. 700-703), PAMI, September 1987.
- [26] A. DHAWAN, Iterative principal axes registration method for analysis of MR-PET brain images, dans IEEE Trans on Biomedical Engineering, vol. 42, November 1995.
- [27] A. DHAWAN et ARATA, 3-D anatomical model-based segmentation of MR brain images through principal axes registration, dans IEEE Trans on Biomedical Engineering, vol. 42, November 1995.
- [28] J. DUCHON, Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, RAIRO Analyse Numérique, vol. 10, pp. 5–12, 1976.
- [29] S. DURBEC, C. ROUX et R. GUILLEMAUD, Mise en correponsdance d'images multimodales, par maximisation d'information mutuelle, et par recalage de phase, Rapport technique, CEA/Grenoble LETI/DSYS/SCSI, ENST de Bretagne, Septembre 1997.
- [30] F. FESCHET, Techniques d'imagerie pour l'aide au positionnement en dosimétrie conformationnelle, Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 1999.
- [31] J. FITZPATRICK et J. GREFENSTETTE, Genetic algorithms in noisy environments, Machine Learning, vol. 3, no. 2-3, pp. 101–120, 1988.
- [32] J. FITZPATRICK et V. MANDAVA, Adaptive search space scaling in digital image registration, IEEE Transaction on Medical Imaging, vol. 8, no. 3, june 1989.
- [33] J. FITZPATRICK et MAURER, Registration of 3-D images using weighted geometrical features, dans IEEE Trans. on Medical Imaging, vol. 15, 1996.

- [34] D. GOLDBERG, Genetic Algorithms in search, Optimization, and Machine Learning, Addisson-Wesley, 1989. Reading, MA.
- [35] D. GOLDBERG, Algorithmes Génétiques. Exploration, optimisation et apprentissage automatique, Addisson-Wesley, June 1994. Traduit de l'anglais par VincentCORRUBLE.
- [36] M. GOOSSENS, F. MITTELBACH et A. SAMARIN, The LATEX companion, Addison-Wesley, mars 1997.
- [37] A. GOSHTASBY, Image registration by local approximation methods, Image Vision Computer, vol. 6, no. 4, pp. 255–261, November 1988.
- [38] A. GUÉRIN, Fusion de données morphologiques IRM-MM 3-D basée sur des informations de courbure, Rapport technique, ENSTBr, 1995.
- [39] C. GUERRA, A survey of parallel algorithms for structural pattern matching, dans IA-PR'94, 12th International Conference on Pattern Recognition, vol. 3, (pp. 275–278), October 1994.
- [40] A. HILGER, The physics of medical imaging, S. WEBB, 1988.
- [41] D. L. HILL et al., Registration of MR and CT images for skull base surgery using pointlike anatomical features, The British Journal of Radiology, vol. 64, pp. 1030–1035, Nov. 1991.
- [42] J. HOLLAND, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.
- [43] D. P. HUTTENLOCHER et W. J. RUCKLIDGE, A multi-resolution technique for comparing images using the hausdorff distance, Rapport technique 1321, Department of Computer Science, Cornell University, Ithaca, NY 14853, 1992. Version électronique sur le ftp.cs.cornell.edu.
- [44] J. ILLINGWORTH et J. KITTLER, The adaptive hough transform, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 9, (pp. 690-698), PAMI, September 1987.
- [45] J.-J. JACQ et C. ROUX, Automatic registration of 3-D images using a simple genetic algorithm with a stochastic performance function, dans Proceeding of EMBS 93 – invitted paper, San diego, (pp. 90–91), October 1993.
- [46] J.-J. JACQ et C. ROUX, Registration of 3-D images by genetic optimization, Pattern Recognition Letters, vol. 16, no. 8, pp. 823-841, August 1995.
- [47] J.-J. JACQ et C. ROUX, A direct multi-volume rendering method. Application to visual assessment of 3-D image registration algorithms, dans Visualization in Biomedical Computing, Lecture notes in computer science - 1131, (pp. 53-62), Springer-Verlag, Sept. 1996.
- [48] J.-J. JACQ et C. ROUX, Automatic detection of articular surfaces in 3-D image through minimal subset random sampling, dans Computer Vision, Virtual Reality and Robotics in Medicine. CVRMed'97 (edité par N. AYACHE), vol. 1205 de Lecture notes in computer science, (pp. 73-82), Springer-Verlag, Grenoble, France, mars 1997.
- [49] J.-J. JACQ et C. ROUX, A direct multi-volume rendering method aiming at comparisons of 3-D images and models, IEEE Trans. on Information Technology in Biomedecine, vol. 1, no. 1, pp. 30-43, Mar. 1997.
- [50] J.-J. JACQ, C. ROUX, E. STINDEL et C. LEFEVRE, Surface recognition in 3-D medical images using genetic matching. application to the extraction of articular surfaces in 3-D ct data sets., International Journal of Imaging System and Technology - Special Issue on 3-D imaging (Ed. Gabor T. Herman). à paraître.

- [51] J. G. JIMÉNEZ et A. BATURONE, Estimación de la posición de un robot móvil, Informatica y automatica, vol. 29, no. 4, pp. 3–18, Dec. 1996.
- [52] B. KIM et al., Mutual information for automated multimodal image warping, dans VBC'96, (pp. 349-354).
- [53] S. KIRKPATRICK, Optimization by simulated annealing, Science, vol. 220, pp. 671–680, May 1983.
- [54] S. LAVALLÉE, Computer integrated surgery and therapy : State of the art, dans Three Dimensionnal Biomedical Imaging (edité par C. ROUX et J. COATRIEUX), IOS Press, Nov. 1996.
- [55] S. LAVALLEE, R. SZELISKI et L. BRUNIE, Anatomy-based registration of 3-D medical images, range images, X-ray projections, 3-D models using Octree-Splines, dans Computer Integrated Surgery (edité par R. TAYLOR, S. LAVALLEE, G. BURDEA et R. MOSGES), (pp. 115–143), MIT Press, Cambridge, MA, Nov. 1996.
- [56] T. LIAO et Z. ZHANG, A review of similarity measures for fuzzy systems, dans IEEE Int. Conference on Fuzzy Systems. Fuzz-IEEE'96, vol. 2, (pp. 930-335), June 1996.
- [57] G. Q. MAGUIRE et al., Graphics applied to medical image registration, IEEE Computer Graphics and Appllications magazinze, vol. 11, no. 2, pp. 20–28, 1991.
- [58] J. MAINTZ, P. VAN DEN ELSEN et M. VIERGEVER, Comparison of feature-based matching of CT and MR brain images, dans Computer Vision, Virtual Reality and Robotics in Medicine. First International Conference, CVRMed'95 (edité par N. AYACHE), vol. 905 de Lecture notes in computer science, (pp. 219–228), Springer-Verlag, Nice, France, april 1995.
- [59] J. MAINTZ, P. A. VAN DEN ELSEN et M. VIERGEVER, Evaluation of ridge seeking operators for multimodality medical image matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 4, pp. 353–365, april 1996.
- [60] J. MAINTZ et M. VIERGEVER, A survey of medical image registration, Medical Image Analysis, vol. 2, no. 1, pp. 1–36, 1998.
- [61] G. MALANDAIN, Registration of 3-D medical images using a mechanical based method, dans IEEE EMBS France - Rennes, 1992.
- [62] G. MALANDAIN et J.-M. ROCCHISANI, Matching of 3D medical images with a potential based method, Rapport technique 1890, INRIA Sophia Antipolis, Mars 1993.
- [63] D. MANFAAT et al., A review of pattern matching approaches, The knowledge engineering review, vol. 11, no. 2, pp. 161–189, June 1996. Cambridge University Press.
- [64] J. MEINGUET, Multivariate interpolation at arbitrary points made simple, Journal of Applied Mathematics and Physics (ZAMP), vol. 30, pp. 292–304, 1979.
- [65] J. MEINGUET, Surface spline interpolation : Basic theory and computational aspects, Approximation Theory and Spline Functions, (pp. 127–142), 1984.
- [66] J. MONOD, Le hasard et la nécessité, Points, vol. 43, 1960.
- [67] P. NEMLER et P. A. VAN DEN ELSEN, A quantitative comparison of residual error for three different multimodality registration techniques, dans Information Processing in Medical Imaging, (pp. 251–262), Brest, 1995.
- [68] C. NIKOU, Contribution au recalage d'images médicales multimodales : approches par fonctions de similarité robustes et modèles déformables sous contraintes statistiques, Thèse de doctorat, Université Lousi-Pasteur-Strasbourg I, mai 1999.

- [69] C. NIKOU et al., Recalage sous-voxel d'images médicales multimodales par une approche robuste, dans GRETSI'97, vol. 2, Grenoble, Septembre 1997.
- [70] M. OGHABIAN et K. TODD-POKROPEK, Registration of brain images by a multiresolution sequential method, Information Processing in Medical Imaging, (pp. 326-342), 1991.
- [71] W. PECKAR, C. SCHNÖRR, K. ROHR, S. STIEHL et U. SPETZGER, Linear and incremental estimation of elastic deformation in medical registration using prescribed displacements, Machine GRAPHICS & VISION, vol. 7, no. 4, pp. 807–829, 1998.
- [72] C. PELIZZARI et al., Accurate three-dimensional registration of CT, PET and/or MR images of the brain, J Comput Assist Tomogr, vol. 13, no. 1, pp. 20–26, 1989.
- [73] M. POWELL, An efficient method for finding the minimum of a function of several variables without calculating derivatives, The computer journal, vol. 4, pp. 155–162, 1964.
- [74] W. H. PRESS, B. P. FLANNERY, S. A. TEULOSKY et W. T. VETTERLING, *Numerical recipes in C*, Cambridge University Press, Cambridge, 1988.
- [75] A. RANGARAJAN et al., A robust point matching algorithm for autoradiograph alignment, dans VBC'96, (pp. 277-286), 1996.
- [76] F. REINHARDT et H. SOEDER, Atlas des Mathématiques, chapitre Geométrie différentielle, (pp. 411–415), dans Livre de Poche [77], 1997.
- [77] F. REINHARDT et H. SOEDER, Atlas des Mathématiques, Livre de Poche, 1997.
- [78] A. ROCHE, G. MALANDAIN, X. PENNEC et N. AYACHE, Multimodal image registration by maximization of the correlation ratio, Rapport technique 3378, INRIA Sophia Antipolis, Août 1998.
- [79] K. ROHR et al., Point-based elastic registration of medical image data using approximating thin-plate splines, dans VBC'96, (pp. 297–306).
- [80] A. ROSTAMPOUR, Algorithms for image registration, moving object detection and identification, Thèse de doctorat, Purdere University, 1983.
- [81] J.-M. ROUET, J.-J. JACQ et C. ROUX, Recalage 3-D élastique de surfaces numériques par optimisation génétique, dans GRETSI'97, vol. 2, Grenoble, Septembre 1997.
- [82] J.-M. ROUET, J.-J. JACQ et C. ROUX, 3D elastic multimodality image registration through a genetic algorithm, dans 20th Annual International Conference of the IEEE EMBS, vol. 20, (pp. 663–666), Hong-Kong, Octobre 1998.
- [83] H. RYSER, Mathématiques Combinatoires, Monographies DUNOD, 1969.
- [84] T. SAÏTO et J. TORIWAKI, New algorithms for Euclidean distance transformation of a n-dimensional digitized picture with applications, Pattern Recognition, vol. 27, no. 7, pp. 66–88, July 1994.
- [85] M. SCHMITT et J. MATTIOLI, Morphologie mathématique, Masson, Paris, 1994.
- [86] L. H. STAIB et X. LEI, Intermodality 3-D medical imaging registration with global search, dans IEEE Worksop on Biomedical Analysis, (pp. 225–234), 1994.
- [87] C. STUDHOLME, D. HILL et D. HAWKES, Multiresolution voxel similarity measures for MR-PET registration, dans Information Processing in Medical Imaging, (pp. 287–298), Brest, 1995.
- [88] V. STURN et al., Stereotactic computer tomography with a modified riechert-mundinger device as the basis for integrated stereotactic neuroradiological investigations, Acta Neurochir, vol. 68, pp. 11–17, 1983.

- [89] R. SZELISKI et S. LAVALLEE, Matching 3-D anatomical surfaces with non rigid deformations using octree-splines, IEEE Workshop on Biomedical Image Analysis, (pp. 144–153), 1994.
- [90] Y. TANG, New algorithm for fixed and elastic geometric transformation models, 1994.
- [91] J.-P. THIRION, Automatic registration of 3-D cat-scan images using surface curvature, 1992.
- [92] J.-P. THIRION, Fast non-rigid matching of 3-D medical images, Rapport technique 2547, INRIA Sophia Antipolis, may 1995.
- [93] P. VAN DEN ELSEN, Medical image matching a review with classification, IEEE EMBS Magazine, (pp. 26–39), March 1993.
- [94] P. VAN DEN ELSEN, J. MAINTZ, E.-J. POL et M. VIERGEVER, Automatic registration of CT and MR brain images using correlation of geometrical features, IEEE Transaction on Medical Imaging, vol. 14, no. 2, pp. 384–396, june 1995.
- [95] P. VAN DEN ELSEN et al., Accurate matching of electromagnetic dipole data with CT and MR images, Brain Topogr, vol. 3, no. 4, pp. 425–432, 1991.
- [96] L. VINCENT et P. SOILLE, Watersheds in digital spaces : an efficient algorithm based on immersion simulations, dans IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 13, (pp. 583-598), June 1991.
- [97] W. M. WELLS, Multi-modal volume registration by maximization of mutual information, http://splweb.bwh.harvard.edu :8000/pages/papers/ swells/mia-html/mia.html, March 1996.
- [98] J. WEST, J. FITZPATRICK, M. WANG, B. DAWANT, C. R. MAURER, R. M. KESSLER et R. J. MACIUNAS, Retrospective intermodality registration techniques : surface-based versus volume-based, dans Computer Vision, Virtual Reality and Robotics in Medicine. CVRMed'97 (edité par N. AYACHE), vol. 1205 de Lecture notes in computer science, (pp. 151-160), Springer-Verlag, Grenoble, France, mars 1997.
- [99] G. WOLBERG, Digital Image Warping, IEEE Computer Society Press, Los Alamos CA, 1990.
- [100] R. P. WOODS, J. MAZZIOTA et S. CHERRY, MRI-PET registration with automated algorithm, Journal of Computer Assisted Tomography, vol. 17, no. 4, pp. 536–546, Jul. 1993.
- [101] J. YEN, J. LIAO, B. LEE et D. RANDOLPH, A hybrid approach to modeling metabolic systems using genetic algorithm and simplex method, SMC'95, 1995.
- [102] J. YOU, E. PISSALOUX, W. ZHU et COHEN. H.A., Efficient image matching : A hierarchical chamfer matching scheme via distributed system, Real Time Imaging, vol. 1, pp. 245-259, 1995.