THÈSE

préparée à

L'INRIA Sophia Antipolis

et présentée à

L'UNIVERSITÉ de NICE - SOPHIA ANTIPOLIS

pour obtenir le grade de

DOCTEUR

Spécialité

Sciences de l'ingénieur

par

Alexis ROCHE

Sujet de la thèse :

RECALAGE D'IMAGES MÉDICALES PAR INFÉRENCE STATISTIQUE

Soutenue le 2 février 2001 devant le jury composé de :

MM.	**Jacques**	**DARCOURT**	Président
	Nicholas	**AYACHE**	Directeur
	Michael	**BRADY**	Rapporteurs
	Dirk	**VANDERMEULEN**	
	Grégoire	**MALANDAIN**	Examinateurs
	Max	**VIERGEVER**	

Remerciements

Aux membres éminents de mon jury :

- Nicholas Ayache, mon directeur de thèse, qui m'a orienté vers un sujet de recherche à la fois riche et stimulant. En m'accueillant au sein de l'équipe Epidaure de l'INRIA Sophia Antipolis, il m'a permis de travailler dans un environnement scientifique de haut niveau et dans des conditions matérielles uniques au monde.
- Mike Brady, dont j'ai toujours grand plaisir à partager la bonne humeur et la compréhension exhaustive du domaine de la vision par ordinateur. Ses encouragements réguliers m'ont été précieux.
- Dirk Vandermeulen, dont les travaux pionniers sur le recalage par information mutuelle ont été l'une des principales sources d'inspiration de cette thèse. Sa présence dans le jury est une marque de reconnaissance dont je suis flatté.
- Max Viergever, qui m'a fait lui aussi l'honneur d'évaluer mon travail à travers sa compétence mondialement reconnue en vision par ordinateur.
- Jacques Darcourt, qui a manifesté pour cette thèse l'intérêt d'un médecin et me laisse ainsi espérer qu'elle puisse être utile.
- Grégoire Malandain, qui a co-encadré cette thèse et a été un soutien permanent au cours de ces trois années, tant sur le plan scientifique que sur le plan humain.

Un grand merci à Éric Bardinet, qui a consacré beaucoup de son temps à relire ce mémoire et dont les remarques ont permis d'en améliorer significativement la qualité.

Alexandre Guimond et Xavier Pennec ont pris une part active à cette recherche. Qu'ils soient assurés de ma reconnaissance et du plaisir que j'ai eu à travailler avec eux.

A l'issue de ces trois années, ma principale découverte aura peut-être été qu'aucun travail de recherche n'est possible sans un échange constant avec d'autres chercheurs. Je remercie tous les membres actuels et passés de l'équipe Epidaure pour le temps qu'ils m'ont consacré, avec une pensée particulière pour Sylvain Prima et Sébastien Ourselin. Je n'oublie pas les discussions fructueuses que j'ai eues avec Octave Migneco, Jean-François Mangin, David Sarrut. J'ai trouvé en Ève Coste-Manière et en Jérôme Declerck deux modèles de chercheurs dont je veillerai à m'inspirer.

Mon séjour sur la côte d'Azur ne s'est pas mal passé, merci. Valérie est une amie attentive. Je regretterai les virées de fin de semaine avec Stab, Nico, Coralie, Natacha, Marc, Max, Karine et les autres. Merci Bénédicte et Christian. Merci à l'Italie et ses couleurs.

Et puis les amis de longue date : ma sœur Aurélie, mon frère Ivan, maman, papa, la famille Roche au sens large, la famille Schblah, Urtul et ses prophètes, Bob & co. Florence. Longue vie à Louise, Sarah, Justin, Ninon, Margot, Elsa et William.

À Gwendoline,

Table des matières

Table des figures

Chapitre 1

Introduction

1.1 Le rôle du recalage dans l'analyse des images médicales

« For NASA, space is still a high priority »
(George W. Bush, Jr., 5 septembre 1993)

Après l'invention de la photographie, le XX^{e} siècle a vu l'émergence de nouvelles techniques d'imagerie qu'il n'est pas excessif de qualifier de révolutionnaires. En effet, plutôt que de réduire la représentation d'un objet à une surface plane, celles-ci l'appréhendent dans sa géométrie naturelle, c'est-à-dire en trois dimensions. Ces techniques comprennent le scanner, l'imagerie ultrasonore ou échographie, l'imagerie nucléaire ou scintigraphie et l'imagerie par résonance magnétique (IRM). Leur champ d'observation privilégié est l'être vivant et leur domaine d'application principal est la médecine.

La grande nouveauté qu'apportent ces systèmes optiques modernes n'est d'ailleurs pas tant de produire des images tridimensionnelles que de permettre de voir à travers le corps de façon quasiment non-invasive. Ils sont en cela une source d'information inestimable pour le diagnostic et la pratique thérapeutique. Aujourd'hui, des dizaines de milliers d'images médicales sont acquises quotidiennement dans le monde.

Cependant, une fois acquises, les images doivent encore être examinées et interprétées par le praticien. Bien souvent, celui-ci réalise l'examen à partir d'une succession de coupes bidimensionnelles imprimées sur un film, d'où une difficulté certaine à percevoir les objets dans leur globalité, d'autant plus lorsque plusieurs images doivent être étudiées *simultanément*. L'inconfort de l'examen ne facilite pas l'interpétation. Le développement parallèle des sciences informatiques a fait naître l'idée que des outils logiciels pourraient aider le praticien dans ce travail. Ainsi s'est constituée, depuis la fin des années 80, une nouvelle branche de la vision artificielle dont l'objectif général est d'automatiser l'accès aux informations pertinentes fournies par les images médicales (Ayache, 1998; Duncan et Ayache, 2000).

Parmi les besoins les plus courants du praticien se trouvent une variété de tâches de *comparaison* : comparer des images d'un même patient acquises à des instants différents ou selon des modalités différentes, comparer des images de patients différents, ou encore comparer une image avec un atlas anatomique ou fonctionnel. Ces tâches de comparaison relèvent toutes d'une même problématique : le recalage. La présente thèse décrit un certain nombre de contributions méthodologiques au recalage, dans le contexte particulier de l'analyse des images médicales.

1.1.1 Le recalage, un problème à géométrie variable

Le recalage est un vieux problème de vision artificielle encore incomplètement résolu. Il consiste essentiellement à établir une relation géométrique entre les objets représentés par deux images. De manière quelque peu réductrice, le terme *recalage* est synomyme d'expressions telles que *mise en correspondance* ou *alignement*. En donner une définition précise relève cependant de l'utopie. Il n'existe pas un problème unique de recalage mais plutôt une variété de tâches visuelles que le bon sens nous incite à regrouper sous la même rubrique.

Un exemple classique de recalage est celui de la compensation d'un mouvement apparent entre deux images représentant un même objet rigide. En effet, lors d'acquisitions espacées dans le temps, l'objet occupe rarement une position fixe par rapport au système d'imagerie et, ainsi, ses représentations successives seront elles-mêmes localisées différemment dans les images. Dans ce contexte, le recalage s'identifie parfaitement à la mise en correspondance : il s'agit de calculer la transformation géométrique qui permet d'apparier les pixels ou voxels représentant des points physiquement homologues.

Dans le cas d'images tridimensionnelles, cette transformation est idéalement rigide dans la mesure où elle caractérise un simple déplacement dans l'espace. Sa description mathématique fait alors intervenir trois paramètres de translation et trois paramètres de rotation. Remarquons que paradoxalement, la situation se complique dans le cas d'images bidimensionnelles acquises au moyen d'une caméra ou d'un appareil photographique ; on parle alors de recalage stéréoscopique. La transformation reliant le monde réel et le plan d'image n'est pas rigide mais de type projectif. C'est un phénomène dont la manifestation visuelle bien connue est l'effet de perspective. De ce fait, la relation géométrique entre les images n'est pas, a priori, une application : un pixel d'une image peut n'avoir aucun correspondant dans l'autre image et réciproquement.

Jusqu'ici, le recalage apparaît comme un problème objectif et, pourrait-on dire, froidement mathématique. C'est loin d'être toujours vrai. D'abord, parce que rares sont les objets du monde réel qui sont parfaitement rigides – a fortiori s'agissant d'organes vivants comme le cerveau, le cœur, le foie... De tel objets peuvent se déformer de manière complexe dans le laps de temps qui sépare deux acquisitions (celui-ci pouvant varier de quelques secondes à plusieurs mois).

Dès lors, le problème du recalage peut avoir différents sens. Si l'on s'intéresse uniquement à la composante rigide du mouvement, celle qui est due à la différence de positionnement de l'objet par rapport au système d'imagerie, on cherche comme précédemment une transformation rigide. Cependant, certaines applications nécessitent de caractériser les déformations propres de l'objet et, dans ce cas, la transformation reliant les images devra être recherchée dans un espace beaucoup plus grand que celui des transformations rigides. Ces deux problèmes distincts peuvent être considérés l'un et l'autre comme du domaine du recalage, mais seul le dernier relève à proprement parler de la mise en correspondance.

Ajoutons à cela que les procédés d'imagerie ne sont pas parfaits : ils peuvent eux-mêmes créer des distorsions géométriques. Ce sont cette fois des déformations virtuelles dans la mesure où elles ne caractérisent pas une évolution physique de l'objet. Le simple fait qu'une image numérique soit échantillonnée à une certaine résolution spatiale implique une incertitude sur la localisation de l'objet représenté ; c'est en soi une source de distorsion géométrique. D'autres types de distorsions, parfois beaucoup plus sévères, sont dues à l'inexactitude des modèles physiques qui sous-tendent les procédés d'imagerie (hypothèse de célérité constante en imagerie ultrasonore, hypothèse de linéarité du champ magnétique en IRM, etc.).

En théorie, il faudrait calculer la transformation réalisant la mise en correspondance parfaite des images pour en isoler ultérieurement les différentes causes physiques (déplacement rigide, déformations propres, déformations virtuelles). Une telle analyse conduit cependant à la résolution d'un problème inverse dont la solution n'est généralement pas déterminée de manière unique. En pratique, un problème de recalage est souvent traité directement, c'est-à-dire en appariant «au mieux» un certain nombre de structures d'intérêt que l'on appelle *primitives*, et moyennant une certaine classe de transformations spatiales adaptées, dans la mesure du possible, à la nature du mouvement que l'on cherche à estimer. Ceci avec toute la subjectivité que l'expression «au mieux» laisse entendre.

Cette subjectivité apparaît de façon encore plus manifeste lorsqu'il s'agit de recaler des images représentant des objets non plus identiques mais (présumés) comparables. Dans un contexte médical, on parle alors de recalage inter-sujets ou inter-patients. L'exemple très actuel d'un tel problème, ayant de nombreuses implications dans le domaine des neurosciences, est le recalage inter-sujets d'images du cerveau. Du fait de l'extrême variabilité du cerveau d'un individu à l'autre, et plus particulièrement de l'anatomie corticale, il s'agit d'un problème dont la signification physique est sujette à différentes interprétations, pour ne pas dire controverses. On pourra se reporter à (Mangin, 1995; Toga, 1999; Lester, 1999; Hellier, 2000) pour différents points de vue sur le sujet.

Ainsi, certaines approches se fondent sur la détection de points anatomiques stables, comme par exemple ceux du quadrillage proportionnel de (Talairach et Tournoux, 1993). D'autres méthodes utilisent une extraction préalable des sillons du cortex dans le but de privilégier la mise en correspondance des surfaces corticales. Enfin, certaines méthodes assimilables au «morphing» n'utilisent aucune primitive géométrique et cherchent à maximiser une mesure de corrélation entre les intensités des images. La comparaison et la validation de ces différentes techniques suppose de définir des critères objectifs d'évaluation. Ici encore, ces critères pourront être différents selon les domaines d'applications visés : neurochirurgie, sciences cognitives, radiologie...

1.1.2 Quelques exemples illustratifs (dont certains illustrés)

Les différents problèmes de recalage rencontrés en analyse d'images médicales peuvent être classifiés schématiquement selon les deux catégories suivantes :

- Le type de transformation géométrique recherchée : *rigide/non-rigide*,

- Les modalités d'acquisition des images : *mono/multi-modal*.

Il s'agit d'une classification d'origine méthodologique : elle tient moins compte des caractéristiques médicales du problème posé que des niveaux de difficulté que sa résolution implique du point de vue de l'ingénieur. On pourrait notamment faire intervenir une troisième catégorie, déterminée par la conformité des objets représentés, et il conviendrait alors de distinguer le recalage intra-sujet du recalage inter-sujets.

Les problèmes de recalage rigide monomodal sont typiquement des pré-traitements nécessaires à l'étude de séries temporelles d'images. On peut citer l'exemple d'images IRM d'un patient atteint de sclérose en plaques ; dans ce cas, les acquisitions sont échelonnées sur plusieurs mois. Le recalage rigide permet de repositionner les images dans un même référentiel spatial, ce qui facilite grandement l'étude de l'évolution des lésions. Un autre exemple est celui de la correction de mouvement dans les séquences d'IRM fonctionnelles et, dans ce cas, la fréquence d'acquisition est de l'ordre de la seconde. Il s'agit cette fois de détecter des activations neuronales mais le problème du recalage se pose de la même façon.

Si l'on cherche à caractériser le mouvement d'un organe, le recalage rigide doit être complété par un recalage non-rigide. Un bon exemple est l'étude de la dynamique du cœur à partir d'images nucléaires ou par IRM marquée (Declerck, 1997). Le recalage non-rigide peut également servir à estimer un mouvement *virtuel*, entre deux sujets différents comme mentionné plus haut, ou entre deux images d'un même sujet. Pour reprendre l'exemple de la sclérose en plaques, l'approche proposée par (Rey et al., 1999) consiste à effectuer un recalage non-rigide entre deux images successives pour détecter les zones où les lésions évoluent.

Le recalage multimodal est utile lorsque l'on souhaite tirer profit des informations complémentaires apportées par différentes modalités d'acquisition. Un exemple caractéristique est celui du recalage rigide scanner/IRM pour la radiothérapie des tumeurs cérébrales. Il s'agit d'optimiser la position, l'orientation et la puissance des faisceaux de manière à irradier suffisamment le volume tumoral en détruisant un minimum de tissus sains. L'imagerie scanner convient naturellement au calcul des doses car elle mesure localement le coefficient d'absorption aux rayons X. Mais contrairement à l'IRM, elle ne permet pas une localisation précise des différentes structures anatomiques de l'encéphale. Après recalage des deux modalités, le médecin dispose des informations nécessaires à une planification précise de la radiothérapie (voir la figure 1.1).

Enfin, le recalage multimodal non-rigide peut être vu comme le «cas général» des problèmes de recalage. C'est un problème difficile qui est souvent traité indirectement, en combinant un recalage rigide multimodal et un recalage non-rigide monomodal. Dans l'exemple précédent, on peut envisager de recaler l'image scanner avec une image IRM de référence pour laquelle on dispose d'un étiquetage anatomique (réalisé à la main par un expert). La transformation géométrique ainsi calculée permet de «propager» les étiquettes anatomiques dans le repère du scanner et d'y faire apparaître certaines structures d'intérêt pratiquement invisibles par scannographie. Le résultat illustré par la figure 1.2 a été obtenu par G. Malandain et P.-Y. Bondiau en composant un recalage rigide scanner/IRM et un recalage non-rigide entre l'IRM d'un patient et l'IRM d'un sujet de référence.

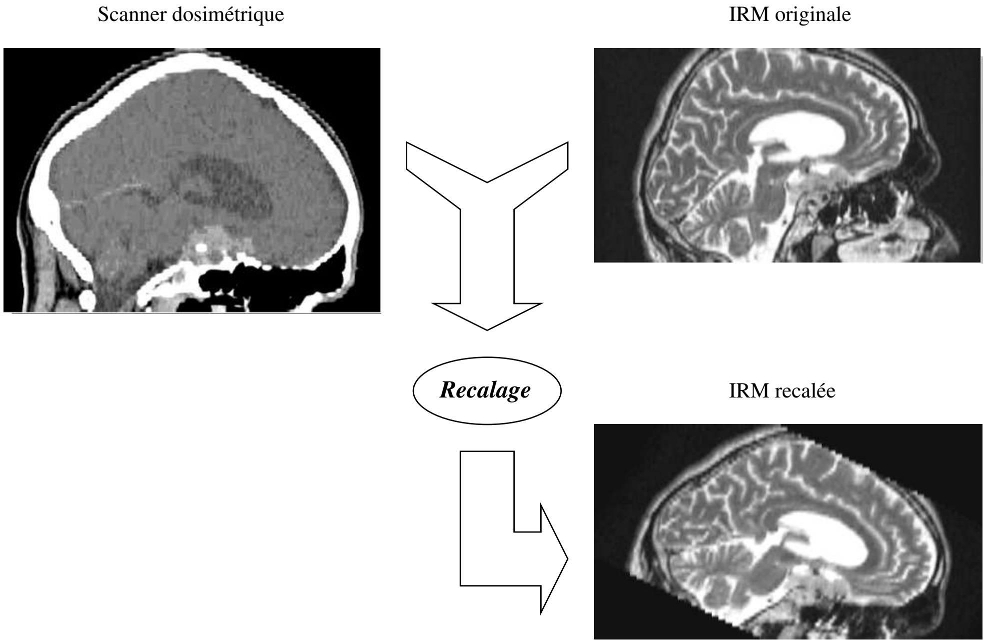


FIG. 1.1 – *Recalage rigide d'une IRM (acquise en protocole T2) et d'un volume scanner dosimétrique. Images grâcieusement fournies par Pierre-Yves Bondiau, Centre Antoine Lacassagne (Nice).*

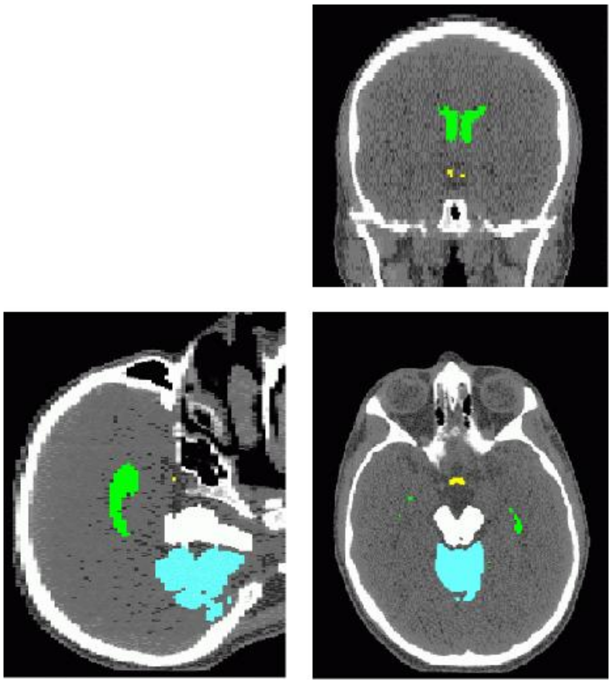


FIG. 1.2 – *Fusion d'un atlas étiqueté et d'un volume scanner. Images grâcieusement fournies par Pierre-Yves Bondiau, Centre Antoine Lacassagne (Nice). Les structures incrustées en couleur sont les ventricules (en vert), le cervelet (en bleu), le tronc cérébral (en blanc) et le chiasma optique (en jaune).*

1.1.3 Le recalage ultrason/IRM

Le problème du recalage d'images ultrasonores et IRM du cerveau a été le point de départ de nos travaux et il sera le fil rouge de ce mémoire. Ce problème apparaît dans le contexte de la neurochirurgie assistée par ordinateur, où l'on souhaite tirer partie d'images médicales pour fournir au chirurgien des outils de visualisation temps-réel. L'opération est typiquement planifiée sur une image IRM pré-opératoire. Lors de l'intervention, on acquiert des images ultrasonores qu'il faut recaler avec l'image pré-opératoire pour permettre au chirurgien de visualiser la trajectoire des instruments et contrôler ainsi son geste thérapeutique.

Les principales causes du mouvement apparent entre l'image ultrasonore et l'image IRM sont les suivantes :

- La position et l'orientation de la sonde ultrasonore dans le reprère de l'image pré-opératoire sont inconnues a priori. Ceci implique déjà l'estimation d'un déplacement rigide.
- Les distorsions géométriques inhérentes à l'imagerie ultrasonore provoquent une déformation apparente du cerveau.
- Pendant la crâniotomie, et en particulier après ouverture de la dure-mère, le cerveau peut subir un déplacement de l'ordre du centimètre et se déformer de manière conséquente. Ce phénomène (*brain shift*) est dû essentiellement aux variations de pression du liquide céphalo-rachidien dans lequel baigne le cerveau.

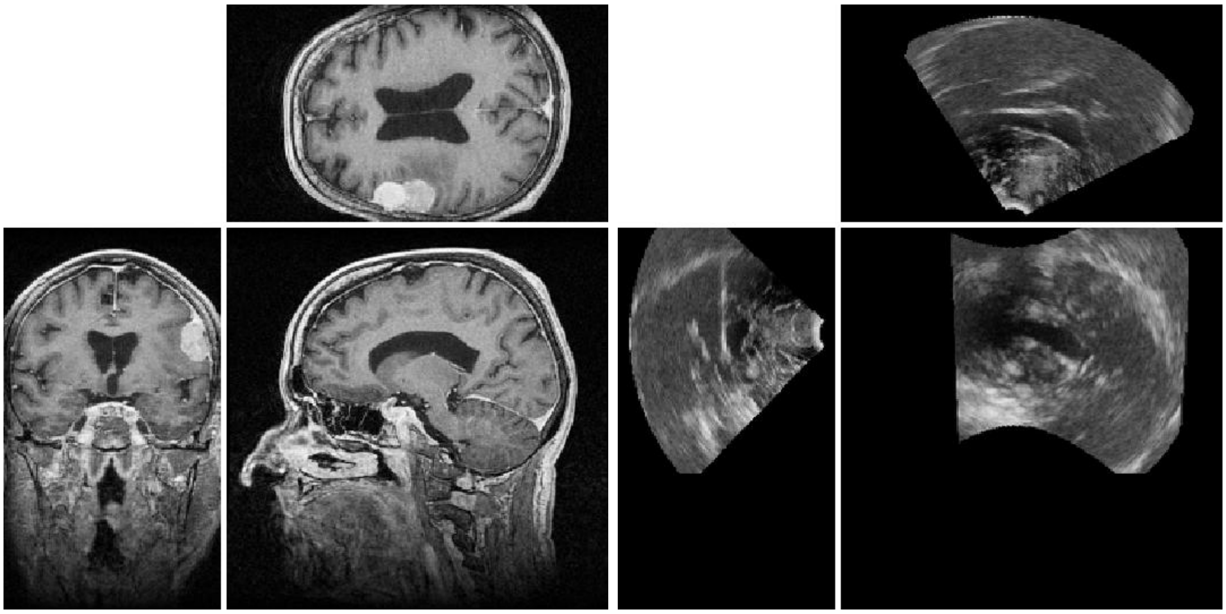
Nous sommes donc en présence d'un problème de recalage multimodal et, en principe, non-rigide. Sa difficulté particulière tient au faible rapport contraste/bruit des images ultrasonores. De nombreux chirurgiens considèrent que l'interprétation de ce type d'images nécessite un long apprentissage et on peut comprendre pourquoi en observant la figure 1.3. Nous gageons que même un système visuel aussi perfectionné que l'esprit de notre lecteur peine à reconnaître les structures antomiques représentées sur l'image ultrasonore.

Le recalage automatique ultrason/IRM fait encore partie des quelques problèmes de recalage d'images médicales dont on peut considérer qu'ils n'ont pas de solution satisfaisante à l'heure actuelle. Nous espérons y avoir apporté quelques éléments de réponse à travers ces travaux de recherche.

1.2 Approches algorithmiques du recalage

Il est absolument impossible de recenser toutes les méthodes de recalage d'images qui ont été proposées dans la littérature. Les états de l'art écrits par (Brown, 1992; Zhang, 1993), couvrant plusieurs domaines de la vision artificielle, ainsi que ceux de (van den Elsen et al., 1993; Lavallée, 1995; Maintz et Viergever, 1998), spécifiques au traitement d'images médicales, constituent d'excellentes synthèses techniques. Il existe deux grandes familles de méthodes de recalage : les méthodes géométriques[1] et les méthodes iconiques[2].

1. En anglais, l'expression *feature-based* est plus usitée.
2. Cette traduction de *intensity-based* est semble-t-il due à (Zhang, 1993).



(a) Image IRM pré-opératoire (b) Image ultrasonore per-opératoire

FIG. 1.3 – *Trois vues orthogonales correspondantes d'une image IRM et d'une image ultrasonore. Les images sont-elles recalées ? La réponse est en fait : non (voir la figure 6.4 du chapitre 6).*

1.2.1 Méthodes géométriques

Les méthodes géométriques sont peut-être les plus naturelles car elles procèdent de façon analogue à l'esprit humain conscient. Dans un premier temps, il s'agit d'identifier dans les images des caractéristiques géométriques communes ; caractéristiques que l'on appelle des primitives et qui peuvent être des points, des lignes, des surfaces, des volumes, des repères orientés..., voire des descriptions structurées mettant en jeu des relations entre éléments. Ayant ainsi réduit le signal visuel à une information géométriquement pertinente, on cherche la transformation spatiale qui apparie «au mieux» les différentes primitives.

Les méthodes géométriques reposent donc sur deux étapes bien distinctes, qui sont d'ailleurs généralement menées de façon complètement indépendante. La première étape, dite de segmentation, est souvent la plus problématique dès lors que l'on souhaite la réaliser automatiquement. Elle requiert d'extraire des primitives qui soient significatives, ce qui peut s'avérer difficile si les images sont fortement bruitées (par exemple, des images ultrasonores). Mais ce n'est pas tout car il faut encore que ces primitives correspondent aux *mêmes* réalités physiques dans les deux images ; cette deuxième contrainte est particulièrement délicate dans un problème de recalage multimodal. Ainsi, la mise en œuvre des méthodes géométriques nécessite une forte connaissance *a priori* sur la nature des objets imagés, ce qui rend leur champ d'application très spécifique.

Notons qu'il n'est pas interdit de demander à l'utilisateur (le médecin) de réaliser lui-même l'étape de segmentation. Bien que nous ayons tendance à penser que les outils de vision artificielle devraient toujours être complètement automatiques, nous n'en reconnaissons pas moins que l'esprit humain est sans aucun doute l'algorithme de recalage le plus fiable au monde – peut-être faut-il s'en réjouir. Dans un même ordre d'idée, on peut prévoir d'inclure dans les images des marqueurs stéréotaxiques fixés rigidement au patient et dont la détection sera plus aisée. Il s'agit alors, pour reprendre la terminologie de (van den Elsen et al., 1993), de primitives extrinsèques par opposition à intrinsèques.

Nous pouvons donner quelques exemples de primitives détectées automatiquement dans un but de recalage : des points extrêmaux (Thirion, 1996), des lignes de crête (Subsol et al., 1998), des surfaces (Pelizzari et al., 1989), des volumes (Malandain et al., 1994) ou, à un niveau sémantique supérieur, des graphes relationnels distribués (Mangin, 1995; Amit et Kong, 1996; Fernández-Vidal et al., 2000). On trouvera bien d'autres références dans les états de l'art précédemment cités.

Après la segmentation des primitives, la deuxième étape est celle du recalage proprement dit, c'est-à-dire le calcul d'une transformation spatiale «optimale». Il convient de distinguer le cas où les correspondances entre primitives sont connues à l'avance (primitives labélisées) du cas contraire (primitives non-labélisées), très fréquent lorsque la segmentation est automatique. Les méthodes adaptées au premier cas sont typiquement des méthodes d'ajustement aux moindres carrés ou au sens d'une métrique non-euclidienne. Dans le second cas, les méthodes les plus populaires sont, pêle-mêle, les algorithmes de type prédiction-vérication, la minimisation d'une carte de distances (Borgefors, 1988), l'algorithme ICP[3] (Besl et McKay, 1992; Zhang, 1994) ou les algorithmes basés sur des correspondances floues (Rangarajan et al., 1997; Rangarajan et al., 1999).

1.2.2 Méthodes iconiques

Si les méthodes géométriques peuvent être considérées comme des approches haut-niveau, les méthodes iconiques sont des approches bas-niveau. Contrairement aux premières, celles-ci ne nécessitent aucune segmentation préalable des images. Elles consistent essentiellement à optimiser un critère de ressemblance ou *mesure de similarité* fondé uniquement sur des comparaisons locales d'intensités. Dans ce cas, les primitives guidant le recalage sont les données brutes, c'est-à-dire des vecteurs 4D contenant la position et l'intensité des voxels. Pour faire un parallèle avec la vision naturelle, nous pouvons dire que les primitives iconiques sont de l'ordre de la *sensation* alors que les primitives géométriques sont de l'ordre de la *perception*, voire de la *cognition* dans le cas de primitives structurées comme les graphes relationnels (voir la figure 1.4).

On reconnaît une méthode iconique à deux propriétés essentielles. D'une part, le choix des primitives est complètement arbitraire : tous les voxels sont *a priori* des candidats valables. D'autre part, les primitives ne sont pas des entités géométriques ; elles appartiennent à un espace figuratif, différent du monde réel. Ainsi, le critère utilisé pour comparer ces primitives est une mesure de similarité reflétant *indirectement* une distance géométrique.

3. *Iterative Closest Point.*

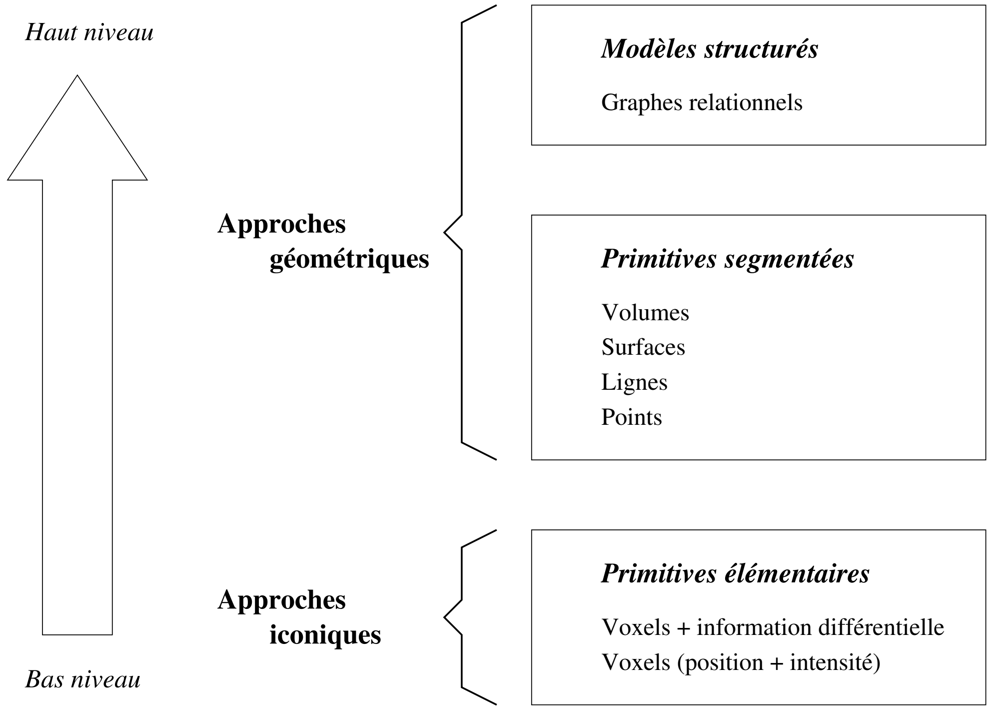


FIG. 1.4 – *Classification générale des méthodes de recalage d'images selon le type de primitives qu'elles utilisent.*

Remarquons qu'il est possible d'enrichir l'information d'intensité par des informations différentielles d'ordre supérieur (par exemple, le gradient). De telles approches permettent de «simuler» un recalage géométrique fondé sur des appariements de contours, mais elles restent des méthodes purement iconiques. En revanche, rien n'empêche de choisir des primitives de types différents dans les images à recaler, notamment des primitives iconiques dans l'une et géométriques dans l'autre (Montagnat, 1999). Les méthodes de ce type peuvent être qualifiées d'hybrides.

1.2.3 Lesquelles choisir ?

Nous sommes sans doute encore très loin d'avoir exploré complètement les potentialités de chacune de ces deux grandes familles de méthodes. Aussi serait-il hasardeux de porter des jugements de valeur sur leurs mérites comparés. Les nombreuses études de validation permettent de dégager certaines tendances, mais il ne faut pas oublier que lorsque l'on parle des méthodes géométriques ou iconiques d'un point de vue expérimental, il s'agit en fait des méthodes géométriques ou iconiques développées *jusqu'à présent*.

Il y a une dizaine d'années, les méthodes iconiques étaient quasiment tombées en désuétude, éclipsées notamment par la rapidité et la robustesse des techniques de recalage de surfaces du type *chamfer-matching* (Borgefors, 1988). Elles ont bénéficié depuis d'un très net regain d'intérêt grâce à la puissance croissante des ordinateurs et au succès des approches fondées sur l'information mutuelle (Collignon et al., 1995a; Viola, 1995) dans de nombreux problèmes de recalage multimodal. Vers la fin de l'année 1997, date à laquelle ont commencé nos travaux, le projet «Vanderbilt» (West et al., 1997) venait de publier un article comparant une douzaine de méthodes pour le recalage scanner/IRM et TEP[4]/IRM. Les résultats donnaient un avantage certain aux méthodes iconiques.

C'est dans ce contexte que nous avons abordé le problème du recalage d'images ultrasonores et IRM à travers une approche iconique. Chemin faisant, nous avons acquis la conviction qu'aucune mesure de similarité existante n'était adaptée à ce problème. Cette thèse décrit la démarche que nous avons suivie pour en construire une.

1.3 Organisation du mémoire et contributions

La suite de ce mémoire est divisée en sept chapitres organisés de la manière suivante :

- Le chapitre 2 est consacré à un état de l'art des méthodes iconiques, l'accent étant mis sur celles qui ont été proposées en analyse d'images médicales. Plutôt que de dresser un inventaire exhaustif, nous avons cherché à faire ressortir les idées majeures, quitte à simplifier parfois la présentation de certains travaux. Ce chapitre ne contient pas de contribution originale, en dehors d'un effort de synthèse et d'un éclairage peut-être nouveau sur le lien qui unit le recalage iconique non-rigide et la théorie des splines multivariées (paragraphe 2.4.6).
- En constatant le caractère sous-contraint des mesures de similarité du type de l'information mutuelle, nous justifions au chapitre 3 une formulation du recalage d'images dont dérive une nouvelle classe de mesures liée à la notion de *rapport de corrélation*. Nous montrons en quoi notre formulation est effectivement nouvelle, notamment par sa capacité à gérer les effets de recouvrement partiel entre les images. Ce chapitre développe les aspects théoriques abordés dans (Roche et al., 1998b; Roche et al., 2000d), deux articles présentés aux éditions 1998 et 2000 de la conférence internationale MICCAI[5].
- Le chapitre 4 propose une évaluation objective de différentes mesures de similarité à partir de la base de données «Vanderbilt». Il répond à un double objectif : valider les mesures fondés sur le rapport de corrélation dans le cadre du recalage multimodal, et comparer quantitativement leurs performances avec les résultats publiés dans (West et al., 1997). Une des conclusions de ce chapitre est que, dans le contexte d'une optimisation locale, certaines versions du rapport de corrélation donnent lieu à une estimation plus robuste et au moins aussi précise que l'information mutuelle.
- Le chapitre 5 est peut-être le plus important de ce mémoire. Notre objectif est de déterminer à quelles hypothèses correspondent les mesures de similarité usuelles, afin de mieux comprendre leurs conditions d'utilisation et finalement d'aider au choix d'une mesure appropriée à un problème donné. Dans la continuité de (McGillem et Svedlow, 1976; Mort et Srinath, 1988; Costa et al., 1993; Viola, 1995; Leventon et Grimson, 1998; Bansal et al., 1998), nous définissons le recalage comme un problème d'estimation par maximum de vraisemblance associé à un modèle

4. Tomographie par émission de positron
5. *Medical Image Computing and Computer-Assisted Intervention.*

générique d'acquisition d'images. Nous calculons alors les mesures correspondant à différentes instanciations du modèle et retrouvons, en particulier, l'information mutuelle et le rapport de corrélation. De façon plus générale, cette approche permet de générer automatiquement une mesure de similarité en fonction d'*a priori* concernant la nature des images. Une version préliminaire de ce travail a été présentée à MICCAI'99 (Roche et al., 1999) et publiée dans un numéro spécial du journal IJIST[6] consacré à l'imagerie 3D (Roche et al., 2000b).

- Le chapitre 6 aborde spécifiquement le problème du recalage *rigide* d'images ultrasonores et IRM du cerveau. La mesure de similarité que nous proposons est un rapport de corrélation bivarié utilisant l'information d'intensité et de gradient de l'image IRM. Elle découle d'un modèle simpliste de dépendance entre les images et des résultats théoriques obtenus au chapitre 5. Nous présentons une validation préliminaire de la méthode qui fait appel à des techniques d'estimation de l'incertitude développées par Xavier Pennec (paragraphes 6.4 et 6.5). Ce chapitre est une version augmentée de l'article présenté à MICCAI'00 (Roche et al., 2000d).
- Le chapitre 7 est une étude de faisabilité du recalage multimodal non-rigide. Notre méthode reprend la philosophie du rapport de corrélation en y incorporant quelques adaptations pratiques dont on montrera qu'elles sont nécessaires dans le contexte du recalage non-rigide. Nous présentons quelques résultats de recalage inter-sujets d'images du cerveau combinant des modalités différentes (atlas labélisé, trois protocoles d'acquisition IRM, scanner). Cette méthode nécessite encore quelques développements pour aboutir au recalage non-rigide ultrason/IRM. Il nous est agréable de souligner que ce travail est le fruit d'une collaboration avec Alexandre Guimond dont on trouvera une présentation légèrement différente dans sa thèse (Guimond, 1999), dans les actes de la conférence ECCV'00[7] (Roche et al., 2000a) et dans un numéro à paraître du journal TMI[8] (Guimond et al., 2001). Par rapport à ces publications, nous avons reformulé la partie consacrée au calcul des déplacements, en précisant sa relation à une minimisation d'énergie et en incluant une comparaison avec les modèles fluides de (Christensen et al., 1996).
- Enfin, le chapitre 8 fait le bilan des travaux présentés et esquisse quelques directions de recherche future.

6. *International Journal of Imaging Systems and Technology.*
7. *European Conference on Computer Vision.*
8. *IEEE Transactions on Medical Imaging.*

Chapitre 2

Recalage par comparaison d'intensités : état de l'art

2.1 Introduction

Contrairement aux méthodes géométriques, les méthodes iconiques de recalage ne font pas appel à une segmentation préalable des images. Leur principe commun est de rechercher la transformation spatiale qui optimise un certain critère de ressemblance, ou mesure de similarité, dont le calcul ne fait intervenir que des comparaisons locales d'intensités.

2.1.1 Premiers pas

Un exemple très simple permet d'introduire ces approches. Supposons que l'on souhaite recaler deux images identiques à une transformation spatiale près. On comprend que si on superpose les images de manière à compenser cette transformation, leurs intensités sont identiques en tout point. Par contre, en appliquant une transformation qui ne correspond pas au véritable décalage, on risque d'apparier des points d'intensités différentes. Ainsi, si l'on est capable de mesurer le degré de recouvrement entre points de même intensité, on obtient indirectement une mesure de l'écart entre la transformation considérée et la transformation véritable, quand bien même cette dernière est inconnue ! Or une telle mesure est très facile à calculer ; on peut prendre, par exemple, la somme des différences d'intensité au carré. Ce critère atteint son minimum absolu (zéro) lorsque les images sont recalées.

Bien entendu, la situation où les images sont parfaitement identiques à une transformation près n'existe pas en pratique. Mais, tant qu'elles restent relativement semblables (en particulier, si elles sont acquises selon la même modalité), le principe très simple que nous venons de décrire s'avère redoutablement efficace. Il semble que les premières méthodes exploitant le concept de mesure de similarité entre intensités soient apparues au début des années 70. Certains auteurs ont cherché alors à relier les mesures de similarité à des modèles stochastiques d'acquisition. Ainsi, il a été démontré que le critère de somme des différences au carré que nous avons mentionné est théoriquement justifié si l'on suppose que les images à recaler sont identiques à une transformation spatiale près et à un bruit blanc gaussien additif près (McGillem et Svedlow, 1976).

De nombreuses mesures analogues ont été proposées, telles que le coefficient de corrélation (Svedlow et al., 1978), la somme des différences d'intensités en valeurs absolues (Barnea et Silverman, 1972), ou encore le nombre de changements de signe dans l'image de différence (Venot et al., 1984). On se reportera à l'état de l'art écrit par (Brown, 1992) pour une liste plus détaillée ainsi que pour une discussion des avantages comparés de ces différentes mesures.

Développées à l'origine dans un contexte de recalage global (rigide ou affine), les méthodes iconiques ont par la suite été étendues au cas du recalage non-rigide, où elles sont à la base d'une variété d'algorithmes de flux optique (Horn et Schunck, 1981). On se reportera à (Barron et al., 1994; Lester, 1999) pour deux états de l'art extrêmement détaillés, le premier d'ordre général et le second spécifique aux images médicales.

Le passage à des espaces de recherche plus grands que celui des transformations affines pose en fait une difficulté conceptuelle supplémentaire. Dès lors que l'on n'impose plus de contraintes explicites à la transformation recherchée, l'optimisation d'une mesure de similarité peut conduire à une solution qui n'a pas les bonnes propriétés de continuité et d'inversibilité attendues. Pour remédier à cette difficulté, l'approche la plus courante est de modifier le problème d'optimisation en ajoutant à la mesure de similarité un terme d'énergie permettant de contrôler la régularité de la transformation.

2.1.2 L'histoire contemporaine

Jusqu'à la fin des années 80, la recherche s'était concentrée sur des problèmes de recalage entre images de même modalité. C'est en partie l'émergence du traitement d'images médicales qui a fait apparaître la problématique du recalage multimodal[1]. Or, les mesures de similarité connues reposaient sur une hypothèse de conservation d'intensité, ou au mieux de dépendance linéaire entre les intensités, qui s'accomodaient mal de la relation complexe manifestée par des images issues de modalités différentes. Les mesures classiques étant prises en défaut, il fallait en trouver de nouvelles.

Les travaux de (Woods et al., 1993) ont permis une avancée significative dans cette direction. Pour s'affranchir de l'hypothèse de dépendance linéaire entre les intensités, Woods et al. proposent de partitionner une des images en ses ensembles d'iso-intensités. L'idée est de calculer, pour chacun de ces ensembles, un écart-type (normalisé) traduisant la dispersion en intensités des voxels correspondants dans la seconde image. La mesure de similarité à minimiser est alors obtenue en faisant une somme pondérée de ces écart-types. A l'origine, la méthode était appliquée au recalage rigide TEP/IRM d'images du cerveau et son bon fonctionnement nécessitait une étape de pré-traitement de l'IRM visant à masquer les structures extra-cérébrales. Certains auteurs l'ont améliorée par la suite au prix de modifications sensibles apportées à la mesure de similarité originale (Ardekani et al., 1995; Alpert et al., 1996; Nikou et al., 1998).

1. Le problème de la fusion d'images issues de sources différentes existait déjà en imagerie satellitaire.

Dans la foulée des travaux de Woods et al., un concept très utile allait être introduit par (Hill et al., 1994) : la distribution conjointe d'intensités, ou histogramme conjoint [2], que nous définirons formellement au paragraphe 2.3.1. Hill et al. observent que lorsque la transformation relative entre deux images s'écarte de la position idéale de recalage, leur histogramme conjoint tend à se «disperser». Ainsi, une mesure de dispersion de l'histogramme conjoint est une mesure de similarité. Dans l'article auquel nous faisons référence, c'est en l'occurrence le moment d'espace d'ordre 3 qui est utilisé, un choix que les auteurs présentent eux-mêmes comme arbitraire.

Plusieurs groupes ont alors envisagé d'autres choix et ont montré que des critères hérités de la théorie de l'information permettaient de mesurer plus efficacement la dispersion de l'histogramme conjoint. Les propositions les plus notables sont l'entropie conjointe (Collignon et al., 1995b; Studholme et al., 1995), l'information mutuelle (Collignon et al., 1995a; Viola, 1995) et le coefficient de corrélation entropique ou information mutuelle normalisée (Studholme et al., 1998). Ces trois mesures diffèrent entre elles essentiellement par leurs propriétés de robustesse au recouvrement partiel, une notion sur laquelle nous aurons l'occasion de revenir en détail.

En 1997, on disposait donc déjà d'une panoplie étoffée de mesures de similarités potentiellement adaptées au recalage multimodal rigide. Néanmoins l'information mutuelle s'était imposée dans l'esprit de beaucoup comme la mesure standard, sinon universelle. Une étude de validation du recalage rigide TEP/IRM et scanner/IRM, impliquant des méthodes développées par différents groupes, venait d'être publiée (West et al., 1997) et établissait expérimentalement l'efficacité du recalage par information mutuelle.

D'un point de vue conceptuel, plusieurs auteurs (Viola, 1995; Wells et al., 1996b; Maes et al., 1997) ont avancé des arguments similaires pour expliquer la supériorité de l'information mutuelle sur les mesures du type de celle de Woods et al. On peut les résumer de la manière suivante : ces dernières mesures reposent en fait sur une hypothèse implicite de dépendance *fonctionnelle* entre les intensités des images, hypothèse certes moins restrictive que celle d'une dépendance linéaire, mais encore trop grossière, selon eux, pour rendre compte de la réalité. L'information mutuelle, dont on peut considérer qu'elle repose sur une hypothèse d'ordre plus général, serait donc mieux adaptée à la nature des images médicales.

Néanmoins, certains travaux contemporains de ceux sur l'information mutuelle utilisaient explicitement l'hypothèse d'une dépendance fonctionnelle entre les intensités des images. L'idée était d'estimer simultanément la transformation de recalage et une fonction de correction d'intensité appliquée à l'une des images pour qu'elle «ressemble» à l'autre. Cette idée, présente dans (Barber, 1992; Feldmar et al., 1997) mais restreinte à l'estimation d'une fonction de correction linéaire, a été exploitée par (Friston et al., 1995) et Viola lui-même (Viola, 1995) dans le cas d'une fonction non-linéaire, voire spatialement instationnaire en ce qui concerne la première référence.

C'est une formulation similaire qui nous a conduit à proposer le rapport de corrélation comme mesure de similarité (Roche et al., 1998b). Il s'agit d'un critère de liaison connu en statistique comme une généralisation naturelle du *coefficient* de corrélation déjà évoqué. Le chapitre 3 est consacré à

2. Dans cet article, les auteurs utilisaient le terme *feature space* (espace des caractéristiques), qui n'a pas perduré.

sa description et à sa comparaison avec d'autres mesures.

2.1.3 Et maintenant ?

Nous pouvons considérer aujourd'hui que la faisabilité du recalage multimodal rigide est établie pour différentes combinaisons de modalités d'images médicales. Deux tendances actuelles semblent se dessiner. Plusieurs équipes ont commencé à appliquer les mesures de similarité précédemment évoquées au recalage multimodal non-rigide (Maintz et al., 1998; Gaens et al., 1998; Lau et al., 1999; Guimond et al., 2001; Hellier et Barillot, 2000). Les développements concernant le recalage rigide continuent cependant avec un effort croissant pour concevoir des mesures de similarité spécifiques aux applications visées. La problématique qui en découle est la prise en compte de modèles d'acquisition d'images (Costa et al., 1993; Viola, 1995; Leventon et Grimson, 1998; Bansal et al., 1998; Roche et al., 2000b).

2.1.4 Plan du chapitre

Les algorithmes de recalage iconique se distinguent entre eux par trois types de choix de mise en œuvre :

1. la mesure de similarité,
2. l'espace dans lequel est recherchée la transformation,
3. la méthode d'optimisation.

Ces trois aspects seront abordés dans cet ordre au cours des paragraphes 2.3, 2.4 et 2.5. Nous commençons par introduire quelques notations et éléments de vocabulaire que nous utiliserons fréquemment au cours de ce mémoire.

2.2 Notations et terminologie

Puisque un problème de recalage suppose la donnée de deux images, nous utiliserons les lettres I et J pour les désigner. Les images seront considérées comme des applications d'une partie finie de $\mathbb{R}^2$ ou $\mathbb{R}^3$ (grille) à valeurs dans $\mathbb{R}$,

$$I : \Omega_I \subset \mathbb{R}^d \to \mathbb{R}, \qquad J : \Omega_J \subset \mathbb{R}^{d'} \to \mathbb{R}, \qquad (d,d') \in \{2,3\}^2$$

Par souci de clarté, nous supposerons par défaut que les images sont tridimensionnelles, étant entendu que cette hypothèse n'entraîne pas nécessairement une perte de généralité.

La grille Ω_I (resp. Ω_J) est l'ensemble des vecteurs dont les coordonnées représentent les positions des voxels de I (resp. J) dans un certain repère. Par abus de langage, nous parlerons des voxels de Ω_I ou Ω_J. Le choix des repères est bien entendu arbitraire. Pour une image échantillonnée sur une grille cubique, il est pratique de choisir les axes naturels de la grille comme axes du repère, l'origine étant souvent prise par convention au coin supérieur gauche de la grille. Lorsque cela sera important, nous préciserons si les repères considérés sont *réels*, les coordonnées d'un point tenant compte de la résolution spatiale de l'image, ou si ils sont *virtuels*, les voxels étant alors repérés par des indices entiers. Pour une grille cubique, la matrice de passage d'un repère à l'autre est une simple matrice diagonale.

La lettre $\mathbf{x}$ sera réservée aux éléments de Ω_I et la lettre $\mathbf{y}$ aux éléments de Ω_J. De plus, afin d'alléger au maximum les notations mathématiques, nous serons fréquemment amenés à indicer ces éléments. Les lettres k et ℓ seront respectivement réservées aux indices correspondant à Ω_I et Ω_J. Les intensités des voxels seront alors abrégées de la façon suivante :

$$\begin{cases} i_k \equiv I(\mathbf{x}_k), & k = 1, \dots, n \\ j_\ell \equiv J(\mathbf{y}_\ell), & \ell = 1, \dots, m \end{cases}$$

où, par définition, $n \equiv \mathrm{Card}\, \Omega_I$ et $m \equiv \mathrm{Card}\, \Omega_J$.

Le problème du recalage consiste à relier spatialement les deux grilles d'images. De manière peut-être restrictive, nous le voyons comme la recherche d'une transformation spatiale $T : \mathbb{R}^3 \to \mathbb{R}^3$ à appliquer à l'une des grilles pour la mettre en correspondance avec l'autre. Ce point de vue impose une répartition des rôles : l'image dont la grille subit la transformation sera appelée *source* et l'autre, dont la grille sert par conséquent de référence, sera appelée *cible*. La convention qui sera adoptée tout au long du mémoire sera de prendre I comme image source et J comme image cible.

Cette terminologie nous semble moins équivoque que la terminologie plus classique dans laquelle une des images est vue comme la «référence». En effet, puisque la transformation T est appliquée à Ω_I, nous serions tentés de considérer J comme image de référence. Mais le point de vue peut être renversé car, contrairement à l'intuition, la transformation T ainsi définie est celle qui sert habituellement à rééchantillonner J dans la géométrie de I. Nous avons opté pour ce changement de terminologie après avoir constaté qu'une majorité d'auteurs appelaient image de référence ce que nous appelons ici image source.

2.3 Mesure de similarité

Il existe un nombre considérable de mesures de similarité dans la littérature du recalage d'images. Notre objectif n'est pas de les recenser, mais de présenter les grandes familles d'une manière aussi synthétique que possible. Or, il s'avère que la plupart des mesures existantes peuvent être définies à partir d'un histogramme conjoint (Hill et al., 1994; Bro-Nielsen, 1997; Sarrut, 2000). C'est par conséquent le formalisme que nous avons choisi d'adopter au cours de ce chapitre.

Nous signalons que plusieurs mesures peuvent être appréhendées indépendamment de la notion d'histogramme conjoint. Celle-ci implique une certaine interprétation probabiliste, mais d'autres interprétations sont possibles et parfois plus naturelles. De plus, en termes de mise en œuvre, le calcul effectif d'un histogramme conjoint peut accroître inutilement la complexité calculatoire d'une mesure. C'est le cas, en particulier, pour toutes les mesures basées sur des différences d'images. Il faut donc considérer ce formalisme comme unificateur d'un point de vue conceptuel, mais pas nécessairement pertinent d'un point de vue algorithmique.

2.3.1 Le concept d'histogramme conjoint

Soient deux images I et J et soit une transformation de l'espace $T : \mathbb{R}^3 \to \mathbb{R}^3$. Nous supposons pour le moment que T fait parfaitement coïncider les voxels de I avec des voxels de J. En termes mathématiques, la restriction de T à Ω_I est supposée être une application de Ω_I dans Ω_J. L'histogramme conjoint est alors défini en fonction de T comme le tableau de contingence des couples d'intensités associés aux voxels appariés :

$$p_{i,j}(T) = \frac{1}{n} \operatorname{Card} \left\{ k \in [1, \ldots, n],\, i_k = i,\, j_k^{\downarrow} = j \right\}$$

où nous posons $j_k^{\downarrow} \equiv J(T(\mathbf{x}_k))$. L'utilisation d'une flèche orientée vers le bas pour souligner la dépendance en T constitue une notation abusive, mais nous pensons qu'elle allège la typographie mathématique de l'exposé.

L'histogramme conjoint associe à chaque couple d'intensités possible une valeur comprise entre 0 et 1 qui représente une proportion de voxels. On peut l'interpréter comme une densité de probabilité discrète, ce qui a le sens suivant : $p_{i,j}(T)$ est la probabilité qu'un voxel tiré aléatoirement dans Ω_I ait l'intensité i dans l'image I et que son correspondant par T ait l'intensité j dans l'image J.

La définition précédente n'a de sens statistique que si le nombre de valeurs d'intensités prises par les images est faible devant le nombre de voxels utilisés pour calculer l'histogramme conjoint. En pratique, cette condition est généralement réalisée pour des images encodées sur 1 octet, les intensités prenant alors un maximum de 256 valeurs ; typiquement $(i,j) \in \{0,1,\ldots,255\}^2$. Mais dès que l'on considère des images encodées sur 2 voire 4 octets, il est nécessaire de regrouper les intensités de valeurs proches. Ceci nous amène à une définition plus générale de l'histogramme conjoint. Pour deux réels $\alpha > 0$ et $\beta > 0$ donnés, considérons la fonction indicatrice du pavé $[-\frac{\alpha}{2}, \frac{\alpha}{2}[\times[-\frac{\beta}{2}, \frac{\beta}{2}[$:

$$K_{\alpha,\beta}(i,j) = \begin{cases} 1 & \text{si} \quad -\frac{\alpha}{2} \le i < \frac{\alpha}{2}, \quad -\frac{\beta}{2} \le j < \frac{\beta}{2} \\ 0 & \text{sinon} \end{cases}$$

La fonction $K_{\alpha,\beta}$ engendre une partition discrète de $\mathbb{R}^2$ en fenêtres d'aires égales et centrées sur les valeurs $(i,j) = (p\alpha, q\beta)$, pour (p,q) décrivant $\mathbb{Z}^2$. Nous pouvons définir l'histogramme conjoint sous forme d'une somme :

$$p_{i,j}(T) = \frac{1}{n} \sum_{k=1}^{n} K_{\alpha,\beta} \left(i - i_k, j - j_k^{\downarrow} \right), \qquad \forall (i,j) \in \left\{ (\alpha p, \beta q)\,,\, (p,q) \in \mathbb{Z}^2 \right\} \tag{2.1}$$

Dans la suite de l'exposé, les indices (p,q) seront assimilés aux couples d'intensités (i,j) qu'ils représentent. La restriction de $p_{i,j}(T)$ à ces valeurs possède clairement les propriétés d'une densité de probabilité discrète et généralise donc naturellement la précédente définition.

Les paramètres α et β contrôlent la taille des fenêtres, et on peut se poser la question de leurs valeurs optimales. Une solution courante est de prendre $\alpha = \beta = 1$ pour deux images encodées sur 1 octet, ce qui revient à ne pas regrouper des intensités différentes. Dans le cas général, on se ramène à une répartition des intensités en 256^2 classes par des choix appropriés de α et β. En se fondant

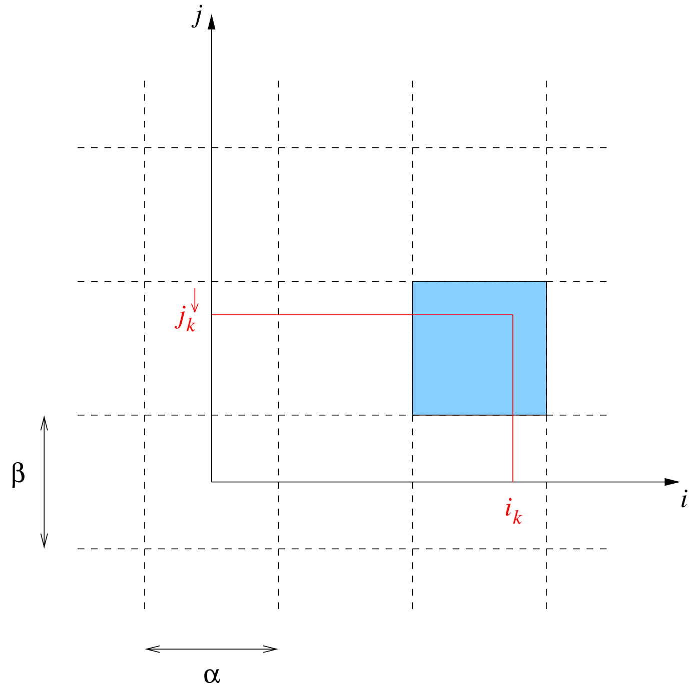


FIG. 2.1 – *Construction d'un histogramme conjoint. Le couple d'intensités* $(i_k,j_k^\downarrow)$ *augmente d'une unité l'effectif de la fenêtre coloriée en bleu.*

sur un argument statistique, (Jenkinson et Smith, 2000) introduisent une dépendance de ces valeurs en fonction du nombre de voxels utilisés.

Insistons sur le fait que la définition de l'histogramme conjoint considérée ici est discrète, au sens où les couples (i,j) varient dans un espace discret. Une définition continue est possible via une très légère modification de l'équation (2.1) : il suffit de normaliser la fonction $K_{\alpha,\beta}$ pour que son intégrale prise sur $\mathbb{R}^2$ soit unitaire. C'est le principe même du fenêtrage de Parzen (Duda et Hart, 1973). Il est également possible de choisir un autre noyau que la fonction indicatrice, par exemple un noyau gaussien. Cette approche permet de régulariser la dépendance en T de l'histogramme conjoint, mais au prix d'un accroissement important de la complexité calculatoire. Elle été explorée dans le contexte discret (Thévenaz et Unser, 1998) et continu (Viola, 1995).

Jusqu'à présent, nous avons travaillé sous l'hypothèse de correspondances spatiales parfaites entre les voxels, ce qui restreint notre définition de l'histogramme conjoint à une classe de transformations peu intéressante car n'ayant pas beaucoup de sens physique. La figure 2.2 illustre les difficultés qui apparaissent lorsque l'on considère des transformations spatiales plus générales. Le correspondant d'un voxel $\mathbf{x}_k$ peut ne pas coïncider exactement avec un nœud de la grille d'arrivée, de sorte que son intensité $j_k^\downarrow$ est indéfinie. Deux problèmes peuvent alors se poser :

1. un problème d'interpolation pour les points $T(\mathbf{x}_k)$ tombant à l'intérieur du domaine délimité par la grille Ω_J,
2. un problème d'extrapolation pour les points $T(\mathbf{x}_k)$ tombant à l'extérieur.

Bien qu'étroitement liés, ces deux problèmes sont généralement traités de manières radicalement différentes. Nous les abordons dans cet ordre au cours des deux prochains paragraphes, ce qui aboutira à généraliser la notion d'histogramme conjoint à toutes les transformations spatiales.

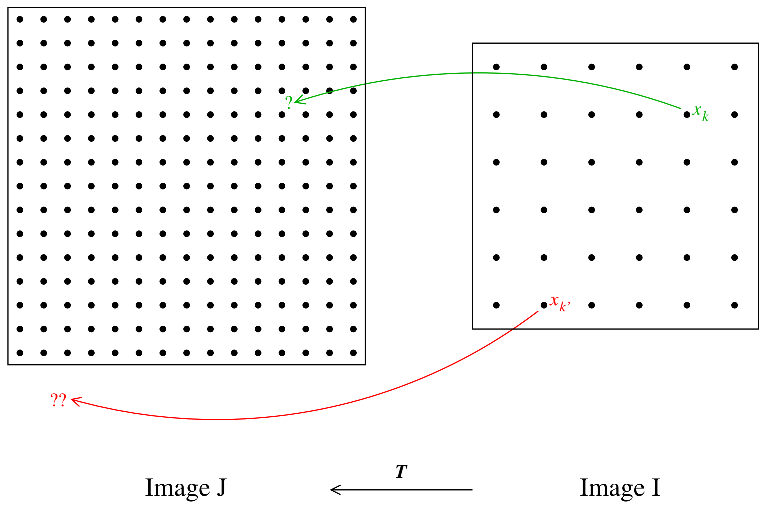


FIG. 2.2 – *Application d'une transformation spatiale à une image, et illustration des effets de bord.*

2.3.2 Interpolation

Le problème qu'on se pose ici est le suivant: comment estimer la contribution à l'histogramme conjoint d'un voxel $\mathbf{x}_k$ dont le correspondant $T(\mathbf{x}_k)$ est situé à l'intérieur du domaine délimité par la grille Ω_J ?

2.3.2.1 Interpolation de l'intensité

L'approche la plus répandue consiste à interpoler l'intensité $j_k^\downarrow$ du correspondant et de lui substituer cette valeur interpolée dans l'équation (2.1). En d'autres termes, il s'agit de représenter J par une fonction définie sur l'espace entier, $\tilde{J} : \mathbb{R}^3 \to \mathbb{R}$, et qui coïncide avec le signal échantillonné aux différents nœuds de la grille Ω_J. Dans le cadre général de l'interpolation linéaire, le signal reconstruit est obtenu par convolution du signal discret (vu comme une somme d'impulsions de Dirac) avec un noyau convenablement choisi.

Un cas particulier intéressant d'un point de vue calculatoire est celui de l'interpolation spline (Unser, 1999; Unser et al., 1993a; Unser et al., 1993b). L'interpolation spline d'ordre d est définie par sa décomposition sur la base des fonctions B-splines de degré d :

$$\tilde{J}(\mathbf{y}) = \sum_{\ell=-\infty}^{+\infty} c_\ell \, \beta_3^d(\mathbf{y} - \mathbf{y}_\ell), \qquad \forall \mathbf{y} \in \mathbb{R}^3 \tag{2.2}$$

où les vecteurs $\mathbf{y}_\ell$ représentent des coordonnées virtuelles (voir le paragraphe 2.2). La somme infinie qui apparaît dans l'équation (2.2) provient du fait que, pour garantir l'unicité de la spline interpolante, le signal doit être défini sur une grille d'extension infinie. Une manière judicieuse de réaliser cette condition est de dupliquer indéfiniment le signal original en lui appliquant des symétries successives. Cette opération, bien évidemment artificielle, ne joue que sur l'interpolation des points situés à proximité des bords de la grille.

Dans notre cas, le signal étant une image 3D, il importe de réaliser que les fonctions B-splines dont nous parlons sont des produits tensoriels de fonctions B-splines unidimensionnelles ; si (y^1,y^2,y^3) désignent les composantes d'un vecteur $\mathbf{y}$:

$$\beta_3^d(\mathbf{y}) = \beta_1^d(y^1)\beta_1^d(y^2)\beta_1^d(y^3).$$

Notons que cette définition n'est pas la seule généralisation possible des fonctions B-splines aux dimensions supérieures à 1. La «bonne» généralisation, celle qui permet de conserver les propriétés d'optimalité établies en dimension 1, trouve son expression dans la théorie de la régularisation et des splines multivariées (Duchon, 1977; Girosi et al., 1995). Néanmoins, les produits tensoriels de fonctions B-splines présentent un avantage calculatoire considérable lié au fait que leur support est compact.

Les coefficients c_ℓ de l'équation (2.2) sont choisis de manière à statisfaire aux contraintes d'égalité entre $\tilde{J}$ et J aux nœuds de la grille. Dans le cas de grilles régulières, ils peuvent être calculés par filtrage récursif (Unser et al., 1993b), ce qui peut être fait une fois pour toutes avant la procédure de recalage. Remarquons cependant que pour les splines de degré 0 et de degré 1, ce calcul n'a pas lieu d'être puisque les coefficients sont alors égaux aux valeurs d'intensité, $c_\ell = j_\ell$.

Une fois les coefficients c_ℓ connus, les intensités interpolées sont données par :

$$j_k^\downarrow \approx \sum_\ell c_\ell \, \beta_3^d\big(T(\mathbf{x}_k) - \mathbf{y}_\ell\big),$$

où un nombre restreint de nœuds voisins de $T(\mathbf{x}_k)$ interviennent dans la somme, étant donnée la taille finie du support des fonctions B-splines. Le nombre de voisins, et donc la complexité calculatoire, augmentent en fonction du degré de la spline selon $(d+1)^3$. Les cas $d=0$ et $d=1$ correspondent respectivement à l'interpolation par le plus proche voisin et à l'interpolation trilinéaire. Un résultat satisfaisant pour l'esprit est que, lorsque d tend vers l'infini, l'interpolation spline tend vers l'interpolation par sinus cardinal (sinc) préconisée par la théorie de l'échantillonnage de Shannon.

Une autre manière de réaliser approximativement une interpolation sinc est d'utiliser des noyaux de convolution obtenus par multiplication de la fonction sinc avec une fonction à support compact (Meijering et al., 1999). Le signal reconstruit est alors obtenu par convolution directe du signal discret :

$$\tilde{J}(\mathbf{y}) = \sum_{\ell=-\infty}^{+\infty} j_\ell \, h(\mathbf{y} - \mathbf{y}_\ell), \qquad h(\mathbf{y}) = \mathrm{sinc}(\mathbf{y})\, w(\mathbf{y})$$

Il s'agit vraiment de l'approche duale à l'interpolation spline. En effet, on peut voir les fonctions B-splines comme des approximations de la fonction sinc dans le domaine de Fourier (Unser, 1999) ; les fonctions sinc fenêtrées en sont des approximations dans le domaine spatial.

En pratique, par souci de réduire les temps de calcul, la plupart des auteurs se limitent à l'interpolation trilinéaire ($d=1$), pour laquelle la fonction B-spline élémentaire s'écrit $\beta_3^1(\mathbf{y}) =$

$(1-|y^1|)(1-|y^2|)(1-|y^3|)$ si $y \in]-1,1[^3$ et vaut zéro sinon. Dans ce cas, l'intensité interpolée peut être vue comme une moyenne pondérée des 8 intensités voisines :

$$j_k^{\downarrow} \approx \sum_{\ell} w_{k\ell}^{\downarrow} j_\ell \tag{2.3}$$

les poids $w_{k\ell}^{\downarrow} \equiv \beta_3^1 \left(T(\mathbf{x}_k) - \mathbf{y}_\ell\right)$ ayant une somme sur ℓ égale à 1, pour tout k.

Le prix à payer est une forte irrégularité de la mesure de similarité vis-à-vis des paramètres de transformation, ce qui se traduit typiquement par l'existence d'extrema locaux parasites (Pluim et al., 2000b; Maes et al., 1997). D'autres auteurs ont opté pour des méthodes d'interpolation plus fines, par splines cubiques (Thévenaz et Unser, 1998) ou par sinc fenêtré (Hill et al., 1994; Friston et al., 1995). Nous pensons néanmoins que les méthodes suivantes offrent un meilleur compromis entre robustesse et temps de calcul.

2.3.2.2 Interpolation intrinsèque

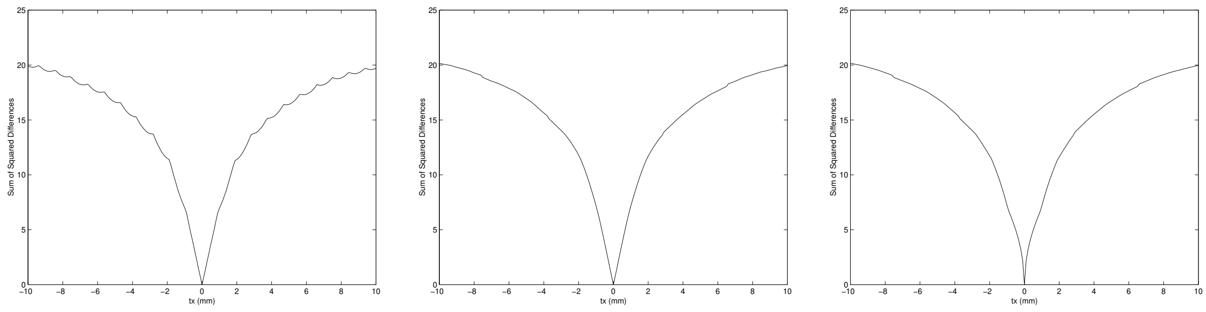
Interpoler l'intensité de l'image cible est une façon quelque peu détournée d'estimer l'histogramme conjoint. En partant de ce constat, (Collignon et al., 1995a; Maes et al., 1997) ont proposé une modification astucieuse de l'interpolation trilinéaire. Plutôt que de faire contribuer le point $T(\mathbf{x}_k)$ à un seul couple d'intensités, l'idée est de répartir son influence sur les intensités de ses voisins, de façon proportionnelle aux poids intervenant dans l'interpolation trilinéaire. Mathématiquement parlant, l'incrément $K_{\alpha,\beta}\big(i - i_k, j - \sum_\ell w_{k\ell}^{\downarrow} j_\ell\big)$ est remplacé par $\sum_\ell w_{k\ell}^{\downarrow} K_{\alpha,\beta}(i - i_k, j - j_\ell)$, d'où l'approximation suivante :

$$p_{i,j}(T) = \frac{1}{n} \sum_{k=1}^{n} \sum_{\ell=1}^{N} w_{k\ell}^{\downarrow} K_{\alpha,\beta}\left(i - i_k, j - j_\ell\right) \tag{2.4}$$

Cette technique est connue sous le nom d'interpolation par volume partiel (VP). On vérifie aisément que l'histogramme conjoint ainsi défini conserve les propriétés d'une densité de probabilité discrète. Ceci est dû au fait que les poids $w_{k\ell}^{\downarrow}$ sont positifs ou nuls et de somme unitaire.

La figure 2.3 illustre la robustesse de l'interpolation VP par rapport aux méthodes d'interpolation de l'intensité. Le problème, trivial, est de recaler une image avec elle-même. Nous représentons le profil de la somme des différences d'intensité au carré (SDC, voir paragraphe 2.3.5) en fonction d'un paramètre de translation variant autour de la transformation identité. On voit nettement que l'interpolation trilinéaire introduit plusieurs minima locaux. Ceux-ci disparaissent avec l'interpolation par splines cubiques mais le temps nécessaire au calcul de la mesure est quadruplé. L'interpolation VP permet d'obtenir une allure suffisamment régulière avec une complexité calculatoire équivalente à celle de l'interpolation trilinéaire.

D'autres stratégies d'interpolation peuvent être imaginées à partir de cette idée. Ainsi, (Sarrut, 2000; Sarrut et Feschet, 1999) proposent deux alternatives : l'interpolation par maximum d'entropie pour laquelle les poids sont choisis constants (1/8), et l'interpolation par différence d'intensité partielle pour laquelle les poids sont modulés en fonction de l'intensité des voisins. Ils rapportent en



(a) interpolation trilinéaire

(b) interpolation spline cubique

(c) interpolation par volume partiel

FIG. 2.3 – *Recalage d'une IRM cérébrale avec elle-même. Allure de la SDC en fonction de la translation horizontale, pour trois méthodes d'interpolation différentes. Les temps de calcul nécessaires pour (a) et (c) sont sensiblement les mêmes et ils sont environ 4 fois inférieurs à celui de (b).*

particulier que la première de ces méthodes s'avère remarquablement robuste compte tenu de son faible coût calculatoire.

L'équation (2.4) met en évidence le fait que l'interpolation VP n'est autre qu'une interpolation trilinéaire. Simplement, au lieu d'interpoler l'image J en $T(\mathbf{x}_k)$, c'est l'image fonction de J, $K_{\alpha,\beta}\,(i - i_k, j - J)$, que l'on interpole. Cette remarque suggère d'étendre l'interpolation VP aux interpolations splines d'ordres supérieurs ou par fonctions sinc fenêtrées. De telles stratégies de calcul mériteraient sans doute d'être explorées.

2.3.3 Le problème du recouvrement partiel

On est très souvent amené à considérer des transformations spatiales pour lesquelles les grilles d'images ne se recouvrent que partiellement. Plus précisément, ce que nous appelons recouvrement partiel correspond à la situation où la grille source transformée $T(\Omega_I)$ n'est pas intégralement contenue dans l'enveloppe convexe de la grille de l'image cible Ω_J.

De manière très générale, l'interpolation associée à un point $T(\mathbf{x}_k)$ fait intervenir ses plus proches voisins dans Ω_J, l'étendue du voisinage à considérer dépendant de la méthode choisie. Ainsi, pour peu que $T(\mathbf{x}_k)$ soit situé à l'extérieur de la grille (ou éventuellement à l'intérieur mais à proximité du bord), l'information de voisinage devient insuffisante pour mettre en œuvre l'interpolation. Un voxel $\mathbf{x}_k$ pour lequel c'est le cas sera appelé «point isolé». Par exemple, dans le contexte d'une interpolation VP ou trilinéaire, les points isolés sont ceux dont les correspondants ne possèdent pas 8 voisins de grille. Pour de tels points, notre problème d'interpolation devient un problème d'extrapolation. Il nous appartient alors de décider si nous cherchons à le résoudre ou si nous préférons le contourner.

Il semble qu'il existe un concensus général à ce sujet. Dans le contexte du recalage rigide ou affine, c'est de très loin la deuxième approche qui est la plus prisée. Elle consiste tout simplement à ne pas prendre en compte les points isolés dans le calcul de l'histogramme conjoint. L'ensemble $\Omega_I^{\downarrow} \subset \Omega_I$ des voxels appariés dépend alors de T :

$$p_{i,j}(T) = \frac{1}{n^{\downarrow}} \sum_{k \in \Omega_I^{\downarrow}} K_{\alpha,\beta}\left(i - i_k, j - j_k^{\downarrow}\right)$$

où $n^{\downarrow} \equiv \mathrm{Card}\Omega_I^{\downarrow}$. Bien entendu, cette formule doit être modifiée en fonction de la méthode d'interpolation choisie[3].

Pour le recalage déformable, en revanche, l'approche la plus répandue est de rééchantillonner intégralement l'image J dans la géométrie de I, ce qui revient à interpoler et extrapoler J. L'extrapolation consiste généralement à attribuer une intensité arbitraire aux voxels manquants dans Ω_J. Ces valeurs peuvent d'ailleurs également servir de conditions aux limites pour la fonction d'interpolation (voir le paragraphe 2.3.2). L'avantage de cette approche est de rendre indépendant de T l'ensemble des voxels contribuant à l'histogramme conjoint. Son inconvénient majeur est d'inventer des structures qui ont peu de chances de correspondre à la réalité et qui, de ce fait, risquent de fausser complètement la mesure de similarité.

Les raisons pour lesquelles le recouvrement partiel est traité de façon différente selon l'espace de recherche semblent d'ordre purement pratique. Le recalage affine implique en général la recherche de «grands» déplacements (typiquement, plusieurs dizaines de voxels). La zone de recouvrement entre les images est donc susceptible de varier fortement au cours du processus de recalage, d'où le caractère particulièrement dangereux du biais introduit par l'extrapolation.

Dans le contexte du recalage déformable, ce sont des déplacements de faibles amplitudes qui sont recherchés tout simplement parce qu'en règle générale un recalage affine a été effectué au préalable. Ainsi, même si la mesure de similarité est biaisée, on peut espérer que le biais varie peu au cours du recalage et qu'ainsi il affecte modérément l'estimation de la transformation. De plus, étant donné le caractère local de la transformation, les effets de l'extrapolation se traduiront de façon sporadique par des déplacements erronés aux bords de l'image, c'est-à-dire loin des régions d'intérêt.

2.3.4 Accélération du calcul

La complexité algorithmique du calcul d'un histogramme conjoint est approximativement proportionnelle au nombre de voxels n de l'image source. Pour accélérer le calcul, on peut se limiter à un sous-ensemble de la grille Ω_I.

Une manière courante de réaliser cette opération consiste à sous-échantillonner régulièrement Ω_I. On ne considère alors que les voxels dont les coordonnées virtuelles selon chaque axe sont des multiples de facteurs entiers (f_x, f_y, f_z). De cette manière, le temps de calcul sera approximativement divisé par $f_x.f_y.f_z$. Dans le cadre d'un algorithme stochastique d'optimisation de l'information mutuelle, (Viola, 1995; Viola et Wells, 1997) utilisent un ensemble de voxels tirés aléatoirement et

3. Dans le cas de l'interpolation VP, les voxels transformés à proximité du bord de l'image cible peuvent être partiellement pris en compte (Maes, 1998).

remis à jour à chaque itération. Une autre stratégie possible consiste à ne garder que les voxels proches de contours (Thirion, 1998).

2.3.5 Classification des mesures de similarité

La notion d'histogramme conjoint est maintenant définie ou, plus exactement, plusieurs définitions possibles de l'histogramme conjoint ont été données. Nous en venons à l'étude proprement dite des mesures de similarité.

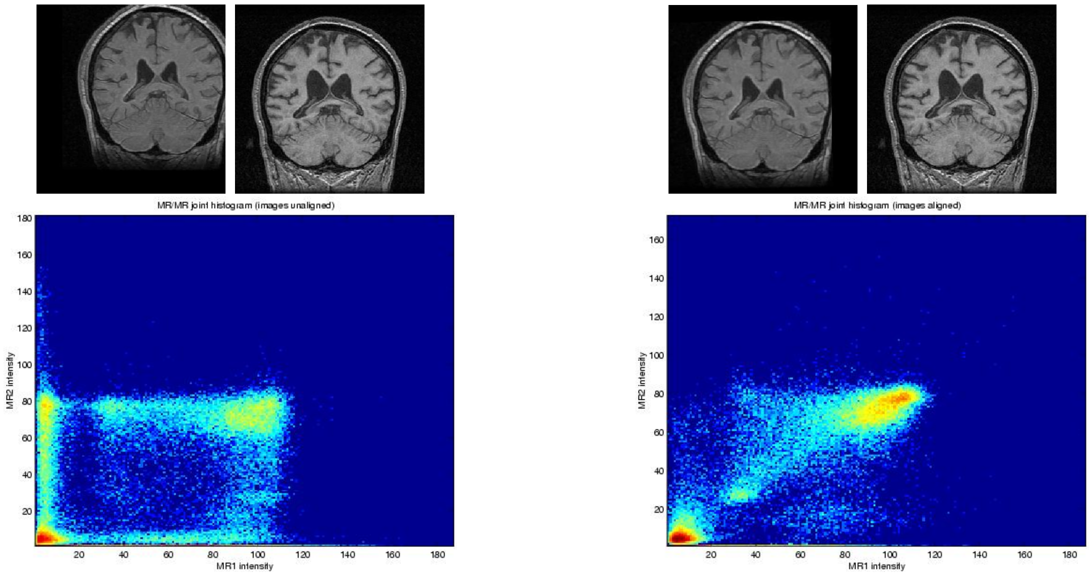
Dans le contexte de ce chapitre, une mesure de similarité est une fonction à valeurs réelles dont l'argument est l'histogramme conjoint, lui-même fonction de la transformation spatiale. Le fondement commun aux nombreuses mesures proposées dans la littérature est l'idée que les intensités de deux images manifestent une cohérence d'autant plus forte que les images sont bien alignées. Le rôle de la mesure de similarité est précisément de donner une signification quantitative à cette notion de cohérence.

Les figures 2.4 et 2.5 justifient ce principe à travers deux exemples concrets de recalage rigide intra-sujet. La figure 2.4 correspond au recalage monomodal de deux images IRM : nous représentons leur histogramme conjoint calculé pour deux transformations différentes, l'une traduisant un décalage visible (a), l'autre un recalage quasiment parfait (b). Dans la première situation, les couples d'intensité sont répartis dans le plan de façon relativement désordonnée. La cohérence est manifestement plus forte dans la deuxième situation, où on a la très nette impression que les intensités sont regroupées le long d'une droite.

La figure 2.5 représente de façon analogue le cas du recalage multimodal d'une image scanner et d'une image IRM. L'évolution de l'histogramme conjoint entre la situation de décalage (a) et la situation de recalage (b) donne encore l'impression visuelle d'un gain de cohérence. Mais, cette fois-ci, la relation entre les intensités n'est clairement plus de type affine lorsque les images sont recalées. La majorité des couples d'intensités est regroupée autour d'une courbe d'allure parabolique. On observe un amas secondaire, nettement plus dispersé, autour de l'intensité 50 dans l'IRM et 150 dans le scanner. Il n'est pas évident de caractériser une telle relation.

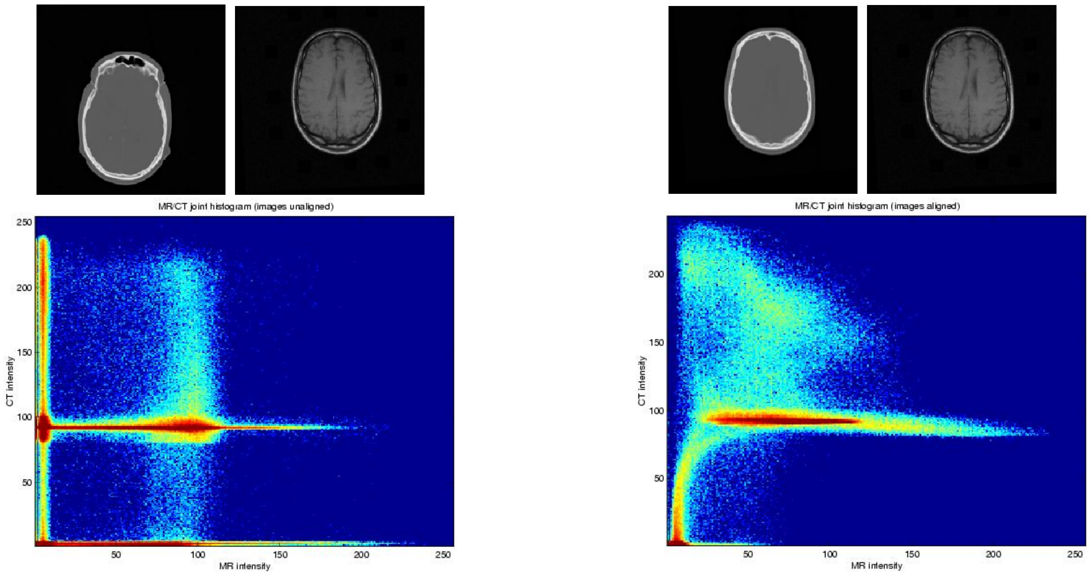
Le choix d'une mesure adaptée à un problème de recalage particulier n'est pas toujours clair, et ce pour au moins deux raisons. D'une part, établir la relation théorique qui existe entre les intensités des images nécessiterait de faire appel à des modèles physiques d'acquisition. Or, de tels modèles sont rarement disponibles et, quand c'est le cas, ils sont extrêmement complexes et difficiles à synthétiser. D'autre part, la plupart des mesures de similarité se fondent sur des hypothèses de dépendance qui elles-mêmes ne sont pas entièrement explicites. La classification suivante s'efforce de regrouper les mesures selon le type d'hypothèse auquel elles correspondent. Elle doit être considérée comme indicative. Pour la clarté de l'exposé, les formules mathématiques sont regroupées dans le tableau 2.1.

Conservation de l'intensité. Les mesures de cette classe considèrent comme idéale la situation où les images sont identiques. Elles sont le plus souvent calculées à partir d'une image de différence.



(a) images non-recalées (b) images recalées

FIG. 2.4 – *Histogramme conjoint de deux IRM cérébrales (coupes coronales représentées) calculé pour deux transformations différentes. Le patient souffre d'une production anormalement élevée de liquide céphalo-rachidien, ce qui explique la variation de volume des ventricules entre les deux acquisitions.*



(a) images non-recalées (b) images recalées

FIG. 2.5 – *Histogramme conjoint d'un couple scanner/IRM (coupes axiales représentées) calculé pour deux transformations différentes.*

Les plus usitées (Brown, 1992) sont la somme des différences au carré (SDC) et la somme des différences en valeurs absolues (SDA). Plus récemment, (Buzug et Weese, 1998) ont proposé différents critères d'entropie calculées à partir de l'histogramme simple de l'image de différence. Toutes ces mesures peuvent être calculées à partir de l'histogramme conjoint même si, répétons-le, ce n'est pas la manière la plus efficace de procéder. En ce qui concerne la SDC et la SDA, les formules que nous donnons correspondent à des versions normalisées (division par le nombre de voxels) : il s'agit d'une précaution saine lorsque l'on décide de rejeter les points isolés (voir le paragraphe 2.3.3). Signalons enfin quelques mesures pour lesquelles la connaissance de l'histogramme conjoint est insuffisante : le nombre de changements de signe dans l'image de différence (Venot et al., 1984) ou la mesure contextuelle[4] également proposée par (Buzug et Weese, 1998).

Dépendance linéaire ou affine. L'étape suivante est de supposer que les intensités sont reliées par une fonction affine, c'est-à-dire : $I \approx \alpha\, J + \beta$. Le cas où le paramètre β est supposé nul correspond à une hypothèse de dépendance linéaire. Ce type d'hypothèse permet de modéliser des contrastes différents entre les images. Les mesures qui lui sont adaptées sont pour la plupart des variantes du coefficient de corrélation (Brown, 1992). Nous pouvons également signaler une variante de la SDC due à (Alpert et al., 1996) ainsi que la variance du rapport des intensités (Woods et al., 1992), à ne pas confondre avec ce que nous appelons ici le critère de Woods. D'après notre expérience, le coefficient de corrélation est souvent un excellent choix pour un recalage monomodal.

Dépendance fonctionnelle. Pour des images de modalités différentes, en revanche, l'hypothèse de dépendance affine est souvent grossièrement fausse. On peut envisager de s'y ramener par un certain nombre de pré-traitements (van den Elsen et al., 1994) ou alors supposer l'existence d'une relation fonctionnelle plus générale (non-linéaire, non-monotone) entre les intensités. Les mesures adaptées à cette situation sont les différentes variantes du critère de Woods (Woods et al., 1993; Alpert et al., 1996; Nikou et al., 1998), et le rapport de corrélation (Roche et al., 1998b). Nous aurons l'occasion de discuter au chapitre 3 les différences pratiques et théoriques entre ces différentes mesures. Par nature, les mesures de cette classe sont asymétriques au sens où elles ne font pas jouer le même rôle aux lignes et aux colonnes de l'histogramme conjoint. En effet, il peut exister une relation fonctionnelle du type $I \approx f(J)$ sans que la relation inverse soit fonctionnelle : f peut ne pas être monotone. Il importe donc de choisir à l'avance dans quel sens la mesure est calculée.

Dépendance statistique. L'hypothèse d'une dépendance fonctionnelle, même très générale, reste éminemment discutable. Un exemple typique est celui du recalage d'images cérébrales scanner et IRM. En IRM T1, le liquide céphalo-rachidien et l'os donnent sensiblement les mêmes réponses alors qu'ils sont nettement différenciés dans le scanner ; c'est d'ailleurs ce qui explique la présence d'un amas minoritaire, que nous avons signalé plus haut, dans l'histogramme conjoint de la figure 2.5. A l'inverse, les tissus mous sont généralement différenciés en IRM T1 mais confondus en scanner. En

4. Cette mesure est appelée *pattern intensity*.

toute rigueur, il n'est donc pas possible de relier fonctionnellement les intensités d'une IRM T1 et d'un scanner. Un argument similaire a été donné par (Wells et al., 1996b).

C'est ici que la notion d'histogramme conjoint prend toute sa dimension. Puisque l'histogramme conjoint peut être interprété comme la densité de probabilité d'un couple de variables aléatoires, il devient naturel d'utiliser une mesure du degré de dépendance statistique entre ces variables comme mesure de similarité entre les images. La théorie de l'information (Blahut, 1987; Cover et Thomas, 1991) fournit une multitude de critères de liaison statistique, parmi lesquels l'entropie conjointe (Collignon et al., 1995b; Studholme et al., 1995), le coefficient de corrélation entropique (Studholme et al., 1998) mais surtout l'information mutuelle (Collignon et al., 1995a; Viola, 1995) sont devenus des mesures de similarité standard pour le recalage d'images médicales. La notion de dépendance statistique étant plus générale que la notion de dépendance fonctionnelle, les mesures de cette classe sont celles qui reposent sur les hypothèses les plus faibles.

Moyennant une généralisation directe de l'histogramme conjoint, les mesures précédentes peuvent être calculées entre des images dérivées des images originales, y compris des images vectorielles. Pour le recalage monomodal, le coefficient de corrélation des images de gradient est un choix courant (Collins et Evans, 1997; Rohling, 1998; Penney et al., 1998) qui offre souvent une robustesse supérieure aux mesures basées sur les intensités brutes. Dans le contexte du recalage multimodal, des mesures fondées sur les opérateurs différentiels ont été proposées par (van den Elsen et al., 1996). Plus récemment, (Pluim et al., 2000a) ont mis en évidence la robustesse d'une mesure combinant l'information mutuelle des intensités avec le coefficient de corrélation des directions de gradients. Nous avons utilisé une extension comparable du rapport de corrélation pour le recalage d'images ultrasonores et IRM (Roche et al., 2000d) ; voir le chapitre 6.

2.4 Espace de recherche

Nous décrivons ici les principales classes de transformations considérées dans les méthodes iconiques de recalage d'images médicales. Bien entendu, ces classes sont sensiblement les mêmes que pour d'autres types d'approches. Néanmoins, la présentation tient compte du contexte particulier des approches iconiques, plus particulièrement en ce qui concerne le paragraphe 2.4.6 consacré aux transformations locales non-paramétriques.

Contrairement au paragraphe 2.3.2, nous travaillons ici en coordonnées réelles (voir le paragraphe 2.2).

2.4.1 Transformations rigides

Les transformations rigides (rotation et translation) forment en quelque sorte le plus petit espace de recherche possible pour un algorithme de recalage. L'hypothèse de rigidité convient au cas où l'on cherche à compenser la différence de positionnement d'un patient par rapport aux capteurs sans

Nature de la dépendance	**Mesure**	**Formule**
Conservation de l'intensité	SDC*	$\sum_{i,j} p_{i,j}(i-j)^2$
	SDA*	$\sum_{i,j} p_{i,j}\lvert i-j\rvert$
Linéaire	Inter-corrélation normalisée	$\sum_{i,j} p_{i,j}\frac{ij}{e_I e_J}$
	Critère d'Alpert*†	$\sum_{i,j} p_{i,j}\left(\frac{\bar{J}}{\bar{I}}i-j\right)^2$
Affine	Coefficient de corrélation	$\sum_{i,j} p_{i,j}\frac{(i-\bar{I})(j-\bar{J})}{\sigma_I\sigma_J}$
Fonctionnelle	Critère de Woods*†	$\sum_{j} p_j\frac{\sigma_j}{\mu_j}$
	Rapport de corrélation†	$1-\frac{1}{\sigma_I^2}\sum_j p_j\sigma_j^2$
Statistique	Entropie conjointe*	$-\sum_{i,j} p_{i,j}\log p_{i,j}$
	Information mutuelle	$\sum_{i,j} p_{i,j}\log\frac{p_{i,j}}{p_i p_j}$

Notations

$$p_i=\sum_j p_{i,j} \qquad p_j=\sum_i p_{i,j} \qquad p_{i|j}=\frac{p_{i,j}}{p_j}\ (=0 \text{ si } p_j=0)$$

$$\bar{I}=\sum_{i,j} p_{i,j}\,i=\sum_i p_i\,i \qquad e_I^2=\sum_{i,j} p_{i,j}\,i^2 \qquad \sigma_I^2=e_I^2-\bar{I}^2$$

$$\bar{J}=\sum_{i,j} p_{i,j}\,j=\sum_i p_i\,j \qquad e_J^2=\sum_{i,j} p_{i,j}\,j^2 \qquad \sigma_J^2=e_J^2-\bar{J}^2$$

$$\mu_j=\sum_i p_{i|j}\,i \qquad \sigma_j^2=\sum_i p_{i|j}\,i^2-\mu_j^2$$

TAB. 2.1 – *Quelques mesures de similarité utiles. Les mesures marquées d'une astérisque (*) sont à minimiser, tandis que les autres sont à maximiser. Les mesures asymétriques sont marquées d'une dague (†).*

tenir compte d'éventuelles déformations des tissus imagés ou de distorsions géométriques créées par les procédés d'imagerie.

Les transformations rigides peuvent être définies par leur forme matricielle :

$$T(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t},$$

$\mathbf{t}$ étant un vecteur colonne 3×1 représentant la translation, et $\mathbf{R}$ étant une matrice orthogonale directe 3×3, c'est-à-dire vérifiant les conditions $\mathbf{R}\mathbf{R}^t = \mathbf{I}_3$ et $\det \mathbf{R} = 1$.

L'espace des transformations rigides n'est pas un espace vectoriel ; en revanche c'est une variété différentielle de dimension 6 (3 en 2D), ce qui signifie que toute transformation rigide peut être décrite par 6 paramètres libres. Si les paramètres de la translation sont évidents, il existe plusieurs représentations possibles de la rotation (Ayache, 1989; Pennec, 1996) : le vecteur rotation, les angles d'Euler ou le quaternion unitaire. Les deux premières consistent en 3 paramètres libres et la dernière en 4 paramètres liés par une contrainte de normalité.

2.4.2 Similitudes

Parfois appelées homothéties, les similitudes sont des transformations à 7 degrés de liberté (4 en 2D) généralisant les transformations rigides par l'adjonction d'un facteur d'échelle global :

$$T(\mathbf{x}) = s\mathbf{R}\mathbf{x} + \mathbf{t}, \qquad s > 0$$

Notons qu'à l'instar des premières, elles forment un groupe de Lie. Bien que ne conservant pas les distances en général, elles conservent les angles et les rapports de distances.

2.4.3 Transformations affines

De façon plus générale, les transformations affines sont définies par :

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t}$$

où, cette fois, $\mathbf{A}$ est une matrice 3×3 quelconque. Les transformations affines conservent le parallélisme. L'espace qu'elles forment a une structure d'algèbre et c'est en particulier un espace vectoriel de dimension 12 (6 en 2D).

La matrice $\mathbf{A}$ étant quelconque, ses coefficients peuvent servir de paramètres libres. Toutefois, on obtient une représentation géométriquement plus signifiante en exploitant la décomposition en valeurs singulières :

$$\mathbf{A} = \mathbf{R}\boldsymbol{\Delta}\mathbf{Q}$$

où $\mathbf{Q}$ et $\mathbf{R}$ sont des matrices orthogonales directes (matrices de rotation) et $\boldsymbol{\Delta}$ une matrice diagonale dont les éléments sont positifs ou nuls. Toute matrice carrée peut être écrite sous cette forme. Si $\mathbf{Q}$ est la matrice identité, la transformation est une affinité, c'est-à-dire une compression ou dilatation indépendante selon chaque axe suivie d'une transformation rigide. Ce type de transformations est utile lorsque la résolution spatiale de l'image source est méconnue. Plus généralement, les variations de $\mathbf{Q}$ par rapport à l'identité autorisent des effets de cisaillement de la grille source.

2.4.4 Transformations projectives

Les transformations projectives conviennent typiquement au cas où des images 3D sont recalées avec des images 2D acquises au moyen d'une caméra (par exemple, des images vidéo ou des images radiologiques). Elles prennent en compte les effets de perspective liés à la formation de l'image 2D. Il existe autant de familles de transformations projectives que de modèles de caméra, le plus utilisé étant le modèle à sténopé. Dans ce cas, les droites restent des droites mais le parallélisme n'est pas conservé. L'étude de ces transformations étant relativement complexe et sortant du cadre de notre recherche, nous renvoyons le lecteur à l'ouvrage de (Faugeras, 1993).

2.4.5 Transformations locales paramétriques

Les classes de transformation que nous avons évoquées jusqu'ici sont souvent qualifiées de globales dans la mesure où les variations de volume qu'elles autorisent sont soit nulles (cas des transformations rigides), soit constantes dans l'espace. De ce fait, elles ne permettent pas de modéliser convenablement les déformations complexes d'un ensemble de tissus ou les distorsions géométriques inhérentes à certaines modalités d'acquisition.

On conçoit intuitivement que des transformations d'ordre plus général pourront être obtenues en perturbant localement la composante globale du mouvement. Les coordonnées de la déformation résiduelle peuvent être modélisées comme des combinaisons linéaires de fonctions élémentaires :

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t} + \sum_{p=1}^{P} \mathbf{c}_p G(\mathbf{x} - \mathbf{x}_p)$$

Les paramètres $\mathbf{c}_p$ sont des vecteurs 3×1 à déterminer (un jeu de paramètres pour chaque coordonnée) et les positions $\mathbf{x}_p$ désignent des points de contrôle qui peuvent être fixés ou non. La fonction G étant naturellement choisie de manière à ce que les fonctions $G(\mathbf{x} - \mathbf{x}_p)$ forment une famille libre, il est clair qu'une transformation de ce type possède $12 + 3P$ degrés de libertés. Des choix classiques pour G sont le produit tensoriel de fonctions B-splines ou des fonctions à base radiale telles que la fonction «plaque mince» [5].

Le choix des points de contrôle est délicat dans la mesure où il présuppose la connaissance des régions dans lesquelles les déformations résiduelles seront prédominantes. Dans le contexte des approches iconiques, (Rueckert et al., 1999; Kybic et al., 1999; Hayton et al., 1999) utilisent des points de contrôle régulièrement espacés pour évaluer des déformations B-splines. Dans le cas de déformations «plaque mince», (Meyer et al., 1997; Meyer et al., 1998) considèrent les points de contrôle comme des paramètres additionnels de la mesure de similarité (en l'occurrence, l'information mutuelle) ; ils les repositionnent itérativement après les avoir initialisés à la main.

5. *Thin-plate* en anglais : $G(\mathbf{x}) = \|\mathbf{x}\|^2 \log \|\mathbf{x}\|$ en dimension 2 et $G(\mathbf{x}) = \|\mathbf{x}\|$ en dimension 3.

2.4.6 Transformations locales non-paramétriques

Pourquoi ne pas considérer des transformations quelconques ? Un raisonnement simple nous montre que le problème d'optimisation de la mesure de similarité est alors mal posé. En effet, celle-ci ne dépend de T que via un nombre fini de points $T(\mathbf{x}_k)$ correspondant aux nœuds de la grille source Ω_I. Or, deux transformations peuvent être différentes tout en coïncidant sur la grille, et on voit que la seule mesure de similarité ne suffira pas à les départager.

Il s'agit donc de régulariser notre problème d'optimisation. Cela peut être fait par analogie avec la théorie de l'approximation de Tikhonov, c'est-à-dire en ajoutant à la mesure de similarité (dont on supposera ici qu'elle est à minimiser) un terme d'énergie indépendant des images dont le rôle est de pénaliser les transformations à caractère «oscillatoire». Le problème de minimisation devient :

$$\hat{T} = \arg\min_T \left[S(T) + \lambda U(T) \right] \tag{2.5}$$

où $S(T)$ est la mesure de similarité et λ une constante strictement positive qui contrôle l'influence relative des deux termes d'énergie.

L'énergie intrinsèque $U(T)$ est souvent appelée stabilisateur. Pour des fonctions multivariées à valeurs réelles, les stabilisateurs quadratiques ont été intensivement étudiés (Wahba, 1990; Girosi et al., 1995). Ils peuvent être directement généralisés au cas de transformations spatiales (fonctions à valeurs dans $\mathbb{R}^3$) par simple sommation de stabilisateurs indépendants attachés à chaque application coordonnée (Terzopoulos, 1986). On parle alors de stabilisateurs séparables. L'idée étant de caractériser l'énergie «haute fréquence» d'un signal, les stabilisateurs sont aisément définis dans le domaine de Fourier [6] :

$$U(T) = \frac{1}{2} \sum_{i=1}^{3} \int_{\mathbb{R}^3} \frac{|\hat{T}^i(\mathbf{s})|^2}{\hat{G}(\mathbf{s})}\, d\mathbf{s}$$

où $\hat{T}^i(\mathbf{s})$ désigne la transformée de Fourier de la i-ème application coordonnée de T et $\hat{G}(\mathbf{s})$ est une fonction positive qui tend vers zéro lorsque $\|\mathbf{s}\|$ tend vers l'infini (G est un filtre passe-bas). On choisit généralement $\hat{G}$ symétrique, de sorte que sa transformée de Fourier inverse, $G(\mathbf{x})$, est également réelle et symétrique.

L'intérêt des stabilisateurs quadratiques tient en grande partie à la linéarité de l'opérateur $dU : T \mapsto \partial U / \partial T$. On montre en outre que G est une fonction de Green associée à cet opérateur, ce qui signifie que pour toute transformation f, on a :

$$\frac{\partial U}{\partial T} = f \quad \Rightarrow \quad T(\mathbf{x}) = T_0(\mathbf{x}) + G \otimes f(\mathbf{x}) \tag{2.6}$$

où $G \otimes f$ désigne la convolution par G de chaque application coordonnée de f. L'opérateur dU n'étant pas nécessairement inversible, la solution est définie à une transformation T_0 près pour laquelle $dU(T_0) = 0$. A titre illustratif, le tableau 2.2 donne quelques exemples de stabilisateurs avec l'expression de leur opérateur différentiel et d'une fonction de Green associée.

6. Il est bien sûr possible, et même courant, de les définir dans le domaine spatial.

L'équation (2.6) va nous permettre d'exprimer la solution du problème (2.5) sous forme paramétrique. De manière très générale, puisque la mesure de similarité ne dépend de T que par l'intermédiaire d'un nombre fini de correspondances $T(\mathbf{x}_k)$, nous pouvons écrire sa dérivée sous forme d'une somme finie. En posant $T_k \equiv T(\mathbf{x}_k)$, nous avons :

$$\frac{\partial S}{\partial T} = \sum_k \frac{\partial S}{\partial T_k} \frac{\partial T_k}{\partial T} = \sum_k \frac{\partial S}{\partial T_k} \delta(\mathbf{x} - \mathbf{x}_k)$$

Dans cette expression, chaque terme $\partial S / \partial T_k$ est un vecteur 3×1 qui représente le gradient de la mesure de similarité par rapport au déplacement individuel de $\mathbf{x}_k$. En exploitant la condition nécessaire à l'existence du minimum :

$$\frac{\partial S}{\partial T} + \lambda \frac{\partial U}{\partial T} = 0,$$

et en utilisant la propriété (2.6), nous voyons que la solution du problème (2.5) peut être recherchée, à une transformation T_0 près, comme une combinaison linéaire des fonctions G centrées en chacun des points de grille $\mathbf{x}_k$:

$$T(\mathbf{x}) = T_0(\mathbf{x}) + \sum_k G(\mathbf{x} - \mathbf{x}_k)\, \mathbf{c}_k \qquad \text{avec} \quad \mathbf{c}_k \equiv -\frac{1}{\lambda} \frac{\partial S}{\partial T_k}(\hat{T}) \tag{2.7}$$

Par ce résultat, nous montrons que l'effet de la régularisation est de nous ramener à une forme paramétrique analogue à celle qui a été donnée au paragraphe 2.4.5. On observe deux différences : d'une part, tous les voxels de la grille Ω_I sont ici des points de contrôle ; d'autre part, la transformation T_0 n'est pas nécessairement affine.

Pour déterminer T_0 de façon unique, il est nécessaire d'imposer certaines conditions aux limites comme celles que propose par exemple (Christensen, 1994). On notera cependant qu'il existe des conditions aux limites naturelles qui sont implicitement réalisées lorsque l'on se place dans un espace de Sobolev adapté au choix du stabilisateur (Wahba, 1990). Il s'agit en substance de l'espace des transformations pour lesquelles l'énergie $U(T)$ est bornée. Sur cet espace, les fonctions du noyau de dU sont, pour la plupart des stabilisateurs usuels, des polynômes voire de simples transformations affines.

La formule (2.7) est une équation implicite qui suggère une méthode de résolution itérative semblable à celle employée par (Thirion, 1998; Bro-Nielsen et Gramkow, 1996). D'autres techniques de résolution ont néanmoins été envisagées. Dans (Christensen, 1994; Friston et al., 1995), la solution est recherchée par sa décomposition sur une base tronquée de vecteurs propres de l'opérateur dU. Il s'agit d'une forme paramétrique différente de celle fournie par l'équation (2.7) car les fonctions de Green centrées $G(\mathbf{x} - \mathbf{x}_k)$ ne sont pas, sauf cas très particulier, des vecteurs propres de dU. Il est également possible de minimiser explicitement le critère par une descente de gradient (Bajcsy et Kovačič, 1989; Cachier et al., 1999; Pennec et al., 1999) ou d'intégrer l'équation d'Euler-Lagrange associée par différences finies (Christensen, 1994).

La décomposition précédente reste valable dans le cas de stabilisateurs quadratiques non-séparables. Un exemple devenu classique pour les images médicales est celui de l'élasticité linéaire

Stabilisateur	Forme différentielle dU	Fonction de Green $G(\mathbf{x})$
Plaque mince	Δ^2	$\|\mathbf{x}\|$
Volume	$-\Delta^3$	$\|\mathbf{x}\|^3$
Gaussien	$\sum_{m=0}^{\infty} \frac{(-1)^m}{m!a^m} \Delta^m$	$e^{-a\|\mathbf{x}\|^2}$
Élastique	$-\mu\Delta - (\lambda+\mu)\nabla(\nabla.)$	$\|\mathbf{x}\|\mathrm{I}_3 - \alpha \frac{\mathbf{x}\mathbf{x}^t}{\|\mathbf{x}\|}$
Spline tensoriel	$(-1)^m \partial_x^{2m} \partial_y^{2m} \partial_z^{2m}$	$\beta_3^{2m-1}(\mathbf{x})$

TAB. 2.2 – *Quelques stabilisateurs classiques en 3D et les fonctions de Green associées. A noter que la fonction de Green correspondant à l'élasticité linéaire est reprise de (Davis et al., 1997) et que c'est une approximation : G est solution de $dU(G) = 1/\|\mathbf{x}\|$ et non de $dU(G) = \delta(\mathbf{x})$. La constante α est reliée aux constantes λ et μ par $\alpha = (\lambda+\mu)/(3\lambda+7\mu)$.*

(Broit, 1981; Bajcsy et Kovačič, 1989; Gee et al., 1993; Christensen, 1994). Dans ce cas, la fonction de Green est une matrice 3×3 dont (Davis et al., 1997) et (Bro-Nielsen et Gramkow, 1996) ont donné des formules approchées.

Enfin, certains auteurs ont critiqué la tendance des stabilisateurs quadratiques à produire des champs de déplacements trop lisses. La préservation de discontinuités est une problématique classique en suivi de mouvement (Barron et al., 1994). Dans le contexte du recalage inter-sujets, (Hellier, 2000) a mis en œuvre des stabilisateurs non-quadratiques. Une approche différente mais motivée par des arguments similaires est celle des déformations fluides de (Christensen, 1994; Christensen et al., 1996) : la formulation du problème est alors purement dynamique au sens où elle ne peut être décrite comme la minimisation d'une énergie. Il serait intéressant d'étudier dans quelle mesure la forme paramétrique générale obtenue dans le cas de stabilisateurs quadratiques peut être étendue à de telles formulations.

2.5 Méthode d'optimisation

Après la mesure de similarité et l'espace de recherche, le troisième aspect fondamental d'un algorithme de recalage est le choix de la méthode d'optimisation. Pour les méthodes iconiques, ce choix est rendu critique par le fait qu'aucune mesure de similarité connue n'est convexe en tant que fonction des paramètres de transformation, et ce quel que soit l'espace de recherche considéré. Il est d'ailleurs permis de douter sérieusement de la possibilité même d'une mesure convexe qui soit fondée uniquement sur des comparaisons d'intensités. Cette lacune est sans doute la principale contrepartie des approches iconiques par rapport aux approches géométriques.

En optimisation, la convexité d'un critère est une propriété souhaitable. D'une part, elle assure l'existence d'un minimum unique (nous nous ramenons au cas d'un critère à minimiser). D'autre part, elle implique que si deux arguments T et Q sont tels que $S(T) < S(Q)$, alors T est plus proche de l'argument minimum que Q au sens d'une certaine distance. Ainsi, pour estimer numériquement la solution, il suffit de construire une suite d'approximations T_n pour laquelle $S(T_n)$ est une suite décroissante.

Cela n'est malheureusement pas vrai de nos mesures de similarité. En règle générale, il existe plusieurs minima locaux voire, dans certains cas pathologiques et heureusement rarissimes, plusieurs minima globaux. Il est raisonnable de supposer que la mesure est convexe au voisinage de chaque minimum local ; on parle alors du bassin d'attraction d'un minimum local. La capacité d'une méthode itérative à converger vers le «bon» argument minimum est soumise à la condition que l'estimée initiale appartienne au bassin d'attraction correspondant. La taille du bassin d'attraction dépend quant à elle de la mesure de similarité, de différents choix de mise en œuvre comme l'interpolation ou le sous-échantillonnage, mais aussi des images elles-mêmes.

Le projet «Vanderbilt» a permis notamment d'évaluer la précision de plusieurs méthodes d'optimisation de l'information mutuelle pour le recalage rigide multimodal d'images cérébrales. Voir (West et al., 1997) et la page web : `http://www.vuse.vanderbilt.edu/~jayw/results.html`. Plus récemment, (Jenkinson et Smith, 1999; Jenkinson et Smith, 2000) ont comparé la robustesse de différentes méthodes dans le contexte du recalage affine inter-sujets.

2.5.1 Méthodes standard sans gradient

Parmi les méthodes génériques d'optimisation (Press et al., 1992), certaines ne nécessitent pas de calculer le gradient du critère à minimiser ou à maximiser. C'est en particulier le cas de la méthode de Nelder-Mead, dite du simplexe, et de la méthode de Powell, qui ont toutes deux été intensivement utilisées pour le recalage rigide/affine (Maintz et Viergever, 1998). Ces deux méthodes présentent l'avantage d'être relativement rapides. En revanche, elles ne garantissent que la convergence vers un minimum local et sont de fait sensibles à l'initialisation des paramètres de transformation. Dans certains cas, il est possible d'atténuer ce manque de robustesse par une initialisation judicieuse. On peut adopter, par exemple, la transformation qui correspond à aligner les centres et/ou axes d'inertie des images (Alpert et al., 1990; Collins et al., 1994).

Dans une optique d'indépendance vis-à-vis des conditions initiales, certains auteurs (Hill et al., 1994; Kruggel et Bartenstein, 1995) ont proposé d'utiliser des algorithmes génétiques (Holland, 1975). Sous certaines conditions, on peut montrer que les algorithmes génétiques garantissent l'obtention du minimum global. Ces conditions impliquent malheureusement des temps de calcul considérables. Enfin, (Studholme et al., 1996) ont envisagé une approche quasi-exhaustive fondée sur une discrétisation adaptative de l'espace de recherche. Dans (Jenkinson et Smith, 2000), une approche similaire est couplée avec la méthode de Powell.

En règle générale, les méthodes sans gradient fournissent des temps de calcul acceptables tant que l'espace de recherche est de «faible» dimension. Elles sont réservées en pratique à la recherche de transformations globales voire de transformations locales impliquant un petit nombre de points de contrôle (Meyer et al., 1997).

2.5.2 Descentes de gradient

Le gradient d'une fonction donne localement la direction dans laquelle celle-ci décroît le plus. C'est grossièrement sur ce principe que se fondent les méthodes de descente de gradient, dont on trouvera une description détaillée dans (Lemaréchal, 1989; Press et al., 1992). Leur avantage est qu'elles permettent de construire rapidement une suite d'approximations de la solution pour laquelle on est certain (ou presque) que la valeur du critère diminue à chaque itération. Il existe plusieurs familles de descentes de gradient :

- les méthodes du premier ordre : on se déplace toujours dans la direction du gradient (cette famille inclut la méthode de plus profonde descente),
- les méthodes du deuxième ordre ou de quasi-Newton : la direction de descente est donnée par le produit du hessien inverse (éventuellement approché) et du gradient,
- les méthodes de type Levenberg-Marquardt, qui peuvent être vues comme des hybridations de descentes du premier et du deuxième ordre,
- les méthodes de gradient conjugué : la direction de descente combine le gradient et la direction précédente.

L'emploi de méthodes de descente suppose de savoir dériver la mesure de similarité par rapport aux paramètres de transformation. La dérivabilité des mesures fondées sur un histogramme conjoint dépend, par tautologie, de la dérivabilité de l'histogramme conjoint et donc, a priori : de la forme du noyau $K_{\alpha,\beta}$ (paragraphe 2.3.1), de la méthode d'interpolation (paragraphe 2.3.2) et de la méthode de traitement du recouvrement partiel (paragraphe 2.3.3). Selon les choix particuliers concernant ces trois aspects, les formules de dérivation sont différentes (Maes, 1998).

Différents types de descente de gradient ont été utilisés pour le recalage rigide/affine (Maes, 1998; Maes et al., 1999; Woods et al., 1998a; Woods et al., 1998b; Thévenaz et Unser, 1998). Dans le cas du recalage non-rigide, il est plus courant d'utiliser des techniques d'intégration d'équations aux dérivées partielles pour résoudre l'équation d'Euler-Lagrange associée au problème de minimisation (voir le paragraphe 2.4.6). C'est également ce qui est fait dans (Ashburner et Friston, 1997) pour un recalage rigide. Dans certains cas, ces approches reviennent en fait à des descentes de gradient particulières.

Par nature, les descentes de gradient sont adaptées à la recherche d'un minimum local. Afin d'en améliorer la robustesse, certains auteurs utilisent des méthodes de gradient stochastique (Viola et Wells, 1997; Nikou et al., 1998) dont on notera qu'elles ne sont pas nécessairement plus coûteuses en temps de calcul.

2.5.3 Méthodes spécifiques

Les méthodes d'optimisation que nous avons évoquées jusqu'ici sont standard au sens où elles peuvent être appliquées à des critères de natures très différentes. Si on peut apprécier le fait qu'elles soient génériques, on peut également se demander quel bénéfice il y aurait à exploiter les propriétés mathématiques spécifiques aux mesures de similarité que l'on veut optimiser.

Les méthodes suivantes exploitent le caractère «séparable» de certaines mesures de similarité, c'est-à-dire la propriété d'être définies par une somme voxel par voxel d'énergies individuelles.

2.5.3.1 Plus proche voisin itéré

Un exemple d'une telle approche nous est donné par (Feldmar, 1994; Feldmar et al., 1997). Nous nous permettons d'en dénaturer quelque peu la présentation pour la replacer dans le contexte de ce chapitre. Le problème particulier est de minimiser la mesure SDC (voir le paragraphe 2.3.5) sur un certain espace de transformations $\mathcal{T}$. Comme nous l'avons déjà signalé, cette mesure peut être définie indépendamment de la notion d'histogramme conjoint. En admettant que l'image cible J soit définie en tout point de l'espace, ce qui revient à faire une interpolation et une extrapolation (voir paragraphes 2.3.2 et 2.3.3), la SDC s'écrit [7] :

$$S(T) = \sum_{k=1}^{n} \left(i_k - j_k^{\downarrow} \right)^2$$

où on rappelle que, par définition, $j_k^{\downarrow} \equiv J(T(\mathbf{x}_k))$.

Supposons (naïvement) que T soit recherchée commme une transformation quelconque de Ω_I dans $\mathbb{R}^3$. On remarque alors que les différents termes de la somme peuvent être minimisés indépendamment. Il suffit d'apparier chaque voxel $\mathbf{x}_k$ avec n'importe quel point de l'image J dont l'intensité est la plus proche de i_k parmi toutes les valeurs prises par J (s'il en existe, tous les points d'intensité égale à i_k conviennent). Bien entendu, le problème est alors mal posé mais nous pouvons dire qu'il est facile d'en trouver une solution.

Le fait que $\mathcal{T}$ soit un ensemble restreint de transformations spatiales induit des couplages entre les termes $(i_k - j_k^{\downarrow})^2$. On peut néanmoins briser ces couplages en remarquant que notre problème équivaut à chercher une transformation quelconque $Q : \Omega_I \to \mathbb{R}^3$ minimisant $S(Q)$ sous la contrainte qu'il existe $T \in \mathcal{T}$ coïncidant avec Q sur Ω_I. La contrainte peut s'écrire :

$$\min_{T \in \mathcal{T}} \left[\sum_k \| T(\mathbf{x}_k) - Q(\mathbf{x}_k) \|^2 \right] = 0$$

Pour s'épargner l'exercice délicat d'une optimisation sous contrainte, on modifie légèrement le problème en insérant la contrainte dans le critère à minimiser via un multiplicateur de La-

7. Pour être tout à fait cohérents avec la présentation que nous avons adoptée, il faudrait préciser que cette formule correspond à une définition continue de l'histogramme conjoint, le noyau $K_{\alpha,\beta}$ étant choisi comme une impulsion de Dirac : $K_{\alpha,\beta} = \delta(i)\delta(j)$. Cela revient à faire tendre α et β vers zéro dans la définition originale.

grange $\kappa > 0$:

$$S(T,Q) = \sum_k \left\{ [i_k - J(Q(\mathbf{x}_k))]^2 + \kappa \, \|T(\mathbf{x}_k) - Q(\mathbf{x}_k)\|^2 \right\}$$

Dans cette formulation, la transformation Q joue le rôle d'une variable auxiliaire au sens de (Cohen, 1996), dont l'effet est de rendre le critère quadratique en T. Notons que lorsque l'on fait tendre κ vers l'infini, on retrouve le problème initial. En interprétant le terme de la somme comme une distance euclidienne dans $\mathbb{R}^4$, Feldmar et al. reconnaissent une formulation classique pour l'appariement de points non-labélisés, à laquelle correspond naturellement un schéma de résolution connu sous le nom de «plus proche voisin itéré» [8] (Besl et McKay, 1992; Zhang, 1994).

Dans le travail de Feldmar et al., l'algorithme ICP en 4D est appliqué aussi bien au recalage rigide qu'au recalage non-rigide (transformations B-splines) d'images par résonance magnétique. L'algorithme ICP consiste à minimiser alternativement $S(T,Q)$ en fonction de Q et de T : la première étape donne lieu à une recherche quasi-exhaustive relativement aisée, et la seconde revient à un problème d'estimation aux moindres carrés pour lequel on a généralement une solution analytique ; voir (Eggert et al., 1997) pour le cas des transformations rigides et (Declerck, 1997) pour le cas de transformations B-splines.

Bien que nous ne l'ayons pas fait, il serait intéressant de comparer l'ICP avec d'autres méthodes de minimisation de la SDC, notamment en termes de sensibilité aux minima locaux.

2.5.3.2 Appariement de régions

Les méthodes d'appariement de régions [9] sont généralement des processus itératifs en deux étapes qui ne sont pas sans rappeler l'algorithme ICP précédemment évoqué. Elles ont été développées à l'origine pour la compression d'images vidéo (Jain, 1981).

Ici encore, l'idée est d'introduire une transformation auxiliaire Q variant dans un espace sur lequel la mesure est «facile» à minimiser. En l'occurrence, on considère un espace $\mathcal{Q}$ de transformations constantes par morceaux (éventuellement, rigides ou affines par morceaux). Cela revient à réaliser une partition de l'image source en blocs, de mêmes tailles ou non, pour chacun desquels un vecteur translation doit être estimé. La taille des blocs étant amenée à diminuer progressivement en cours de recalage de manière à raffiner l'espace $\mathcal{Q}$, on parle aussi d'approches multi-grilles.

A un niveau de grille donné, les deux étapes sont schématiquement les suivantes :

1. *Appariement.* Minimiser la mesure de similarité en fonction de $Q \in \mathcal{Q}$. Dans le cas d'une mesure de similarité «séparable» du type de la SDC, le déplacement d'un certain bloc n'affecte la mesure globale que par l'intermédiaire des voxels de ce bloc. Par conséquent, pour mener à bien cette étape de minimisation, il suffit d'examiner les blocs à tour de rôle et une seule fois (l'ordre de traitement n'ayant aucune incidence sur le résultat). Ceci conduit le plus souvent à une série de recherches quasi-exhaustives car, en pratique, on se limite à des translations discrètes et bornées dont l'amplitude maximale est reliée au niveau de grille.

8. Encore mieux connu sous le nom *Iterative Closest Point* (ICP).
9. *Block-Matching* ou *Region-Matching*.

2. *Régularisation.* Calculer la transformation $T \in \mathcal{T}$ la plus proche de Q au sens d'une certaine métrique. Typiquement,

$$\hat{T} = \arg \min_{T \in \mathcal{T}} \left[\sum_p \| T(\mathbf{x}_p) - Q(\mathbf{x}_p) \|^2 \right]$$

où les positions $\mathbf{x}_p$ correspondent aux centres des différents blocs de l'image source. Dans le cas où $\mathcal{T}$ est un espace de transformations non-affines, ce critère peut éventuellement faire intervenir un potentiel stabilisateur (voir le paragraphe 2.4.6), et alors l'étape 2 revient à un filtrage du champ de déplacements calculé à l'étape 1.

Ce processus alternant entre la minimisation de deux énergies concurrentes fait penser aux algorithmes «minimax» de la théorie des jeux. Il faut remarquer que la convergence n'est pas assurée en toute circonstance (considérer par exemple le cas d'une image à recaler rigidement avec sa version symétrique par rapport à un plan). Par ailleurs, rien ne prouve qu'en cas de convergence, la transformation à laquelle on aboutit réalise effectivement un minimum de la mesure de similarité sur l'espace $\mathcal{T}$. Il est donc difficile de savoir en quoi les méthodes d'appariement de régions sont des méthodes d'optimisation proprement dites.

Notons que certaines approches multigrilles remédient à cette difficulté théorique en réduisant la recherche des appariements à une seule et même étape de minimisation (Mémin et Pérez, 1998; Hellier, 2000). Dans ce cas, une recherche quasi-exhaustive n'est pas envisageable car les blocs interagissent du fait de la contrainte explicite de régularisation.

La précédente minimisation en deux étapes peut être généralisée à des mesures de similarité n'ayant pas la propriété de «séparabilité». La première étape est alors plus difficile car l'optimisation de la translation associée à un bloc donné ne fait pas intervenir que les voxels de ce bloc. En général, il faut prendre en compte tous les voxels de l'image source à chaque nouvelle évaluation de la mesure, ce qui peut s'avérer très lourd en temps de calcul. De fait, il est courant de se limiter à un calcul local de la mesure de similarité. La première étape ne correspond plus alors à la résolution d'un problème d'optimisation global.

Dans le domaine de l'analyse des images médicales, les techniques d'appariement de régions ont été appliquées le plus souvent à des problèmes de recalage non-rigide, monomodal (Collins et Evans, 1997; Hellier et al., 1999; Lau et al., 1999) ou multimodal (Maintz et al., 1998; Gaens et al., 1998; Hellier et Barillot, 2000). D'autres travaux tendent à prouver qu'elles améliorent la robustesse des méthodes d'optimisation conventionnelles dans différents cas de recalage rigide (Ourselin et al., 2000a; Ourselin et al., 2000b; Prima et al., 2000).

2.5.4 Schéma multi-résolution

Bien souvent, les méthodes d'optimisation que nous avons signalées sont imbriquées dans un schéma multi-résolution. Le principe d'un tel schéma repose sur une transformée pyramidale de chacune des deux images (Moreau, 1995). On recale successivement les versions d'images correspondant à des niveaux cohérents de résolution et d'échelle, en allant des niveaux les plus grossiers vers les plus fins.

Il est généralement admis que les approches multi-résolution offrent le double avantage d'accélérer la convergence et de réduire la sensibilité aux minima locaux. En faisant une analogie avec les méthodes de continuation (Blake et Zisserman, 1987), on peut voir le niveau de résolution comme un paramètre de contrôle dont l'augmentation entraîne une perte de convexité graduelle du paysage d'énergie. Cette affirmation repose sur la conjecture que les mesures de similarité sont «d'autant plus convexes» que la résolution des images est faible. Il nous semble que les résultats théoriques qui permettraient de l'accréditer sont encore insuffisants à l'heure actuelle.

2.6 Conclusion

Après ce tour d'horizon des méthodes iconiques de recalage, nous espérons avoir convaincu le lecteur de la diversité des choix de mise en œuvre qui s'offre au concepteur d'un programme. Ces choix portent sur :

- la mesure de similarité, qui doit être adaptée à la relation physique qui existe entre les intensités des images à recaler,
- différents aspects liés au calcul de la mesure et pouvant influer sur la précision et la robustesse de l'estimation : l'interpolation, le traitement du recouvrement partiel, la forme du noyau intervenant dans le calcul de l'histogramme conjoint et la méthode de sous-échantillonnage de l'image source,
- l'espace de recherche, qui doit être adapté à la relation géométrique qui existe entre les images,
- et enfin, la méthode d'optimisation, dont la capacité à éviter les extrema locaux conditionne de façon déterminante la robustesse de l'estimation.

Classer ces choix par ordre d'importance relève d'un exercice périlleux et forcément subjectif. Tout dépend en fait de l'application visée, du niveau de précision attendu et du degré d'interaction humaine que l'on autorise. Selon les contraintes exprimées par le «cahier des charges», certains aspects pourront revêtir plus d'importance que d'autres.

Dans la suite de ce mémoire, nous nous intéressons essentiellement à l'étude des mesures de similarité. Les raisons qui nous ont poussés à cette orientation sont liées à l'une des principales applications que nous souhaitions réaliser, à savoir le recalage d'images ultrasonores et IRM. Il nous est apparu expérimentalement que la faisabilité de cette application reposait de manière déterminante sur le choix d'une mesure de similarité spécifique.

Chapitre 3

Une classe de mesures de similarité fondée sur le rapport de corrélation

3.1 Limitations de l'information mutuelle

Le chapitre précédent nous a permis de mettre en exergue plusieurs mesures de similarité proposées dans le contexte du recalage d'images médicales. Parmi les mesures adaptées au cas du recalage multimodal, l'information mutuelle (Collignon et al., 1995a; Viola, 1995) s'est imposée sans conteste comme la mesure standard au cours des dernières années.

Rappelons ici son expression : étant données deux images I (source) et J (cible), leur information mutuelle est définie en fonction de leur histogramme conjoint, lui-même dépendant de la transformation spatiale appliquée à I (voir chapitre 2, paragraphe 2.3) :

$$\mathcal{I}_{I,J}(T) = \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{p_i p_j} \tag{3.1}$$

Parce qu'elle repose sur des hypothèses faibles concernant la relation entre les intensités des images, l'information mutuelle a toutes les apparences d'une mesure de similarité universelle. Elle ne suppose ni une relation affine comme le coefficient de corrélation, ni même une relation fonctionnelle d'ordre général comme le critère de Woods ou le rapport de corrélation (voir le chapitre 2, paragraphe 2.3.5). Il s'agit, à l'origine, d'un concept de la théorie de l'information (Blahut, 1987; Cover et Thomas, 1991) dont la vocation est de mesurer l'écart d'un couple de variables aléatoires par rapport à la situation d'indépendance. Son nom dit bien ce qu'elle est : une mesure générale d'information.

Cette «universalité» s'accompagne néanmoins de contreparties. Sur un plan conceptuel, on constate que l'information mutuelle traite les intensités comme des variables purement qualitatives, c'est-à-dire sans faire intervenir une relation d'ordre dans l'espace des intensités. Dans les images médicales, hormi le cas des images labélisées, un tissu n'est jamais représenté par une intensité unique mais plutôt par un certain intervalle de valeurs. Ceci est dû au bruit d'acquisition et à ce qu'on appelle parfois le «bruit biologique» : la grandeur physique mesurée par le procédé d'imagerie manifeste une certaine variabilité dans un tissu donné.

Ainsi, deux voxels d'une même image appartiennent de façon d'autant plus vraisemblable au même tissu que leurs intensités sont proches. La proximité entre les intensités fournit donc une information spatiale dont on aurait envie que la mesure de similarité tienne compte. Un moyen d'imposer cette contrainte, que nous avons signalé au chapitre 2, paragraphe 2.3.1, est de faire appel à des techniques de fenêtrage de Parzen lors de la construction de l'histogramme conjoint. Cette approche équivaut à une convolution, discrète ou continue, de l'histogramme conjoint par un filtre passe-bas ; elle implique un accroissement important du temps de calcul. D'un point de vue plus philosophique, elle revient à estimer l'histogramme conjoint que l'on observerait en bruitant les images, ce qui suggère également sa tendance à faire «disparaître» de l'information.

3.1.1 Expériences synthétiques

Les deux expériences synthétiques à suivre donnent des exemples de recalages aberrants qui peuvent être obtenus par maximisation de l'information mutuelle.

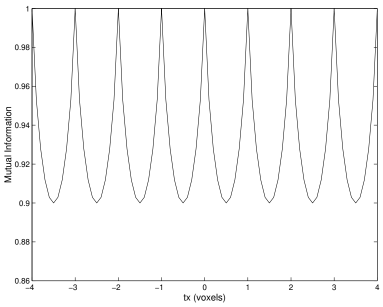
Expérience de la barre grise Nous avons créé deux images surfaciques représentant une barre grise imaginaire, voir la figure 3.1 : (a) une image binaire de taille 28×20 pixels et (b) une image en dégradé de taille 20×20 pixels dont chaque colonne possède une intensité différente. L'origine des repères étant par convention prise au centre des images, la transformation identité correspond au cas où les barres grises sont parfaitement superposées.

(a) image cible (b) image source

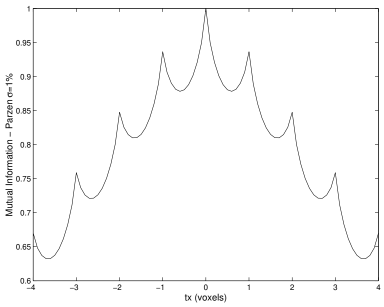
FIG. 3.1 – *Deux images syntétiques représentant (pourquoi pas ?) une barre grise.*

Sur la figure 3.2 sont représentés les profils de différents critères en fonction de la translation horizontale appliquée à l'image source. Les cadres (a), (b) et (c) correspondent à l'information mutuelle pour différentes stratégies de calcul de l'histogramme conjoint, respectivement : la méthode standard, qui correspond à une partition de l'espace des intensités en 256^2 classes, la méthode de Parzen (discrète) avec un noyau gaussien isotrope d'écart-type $\sigma = 2.5$ (1% des niveaux de gris), et la même méthode de Parzen avec un écart-type $\sigma = 12.8$ (5% des niveaux de gris). Le cadre (d) correspond au rapport de corrélation, pour lequel l'histogramme conjoint est calculé par la méthode standard. Dans tous les cas, la méthode d'interpolation utilisée est la version 2D de l'interpolation par volume partiel (VP) définie au chapitre 2, paragraphe 2.3.2.

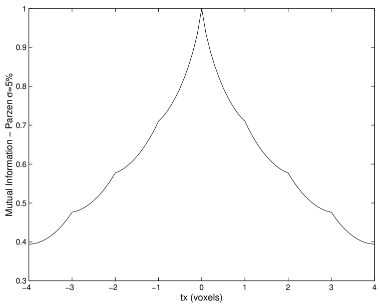
En ce qui concerne le cadre (a), nous observons que l'information mutuelle admet un maximum absolu pour chaque translation entière (à condition que l'image source soit intégralement contenue dans le domaine délimité par l'image cible). Pour les translations non-entières, on observe des valeurs plus faibles du fait de l'interpolation. Avec le fenêtrage de Parzen, l'information mutuelle ne conserve qu'un maximum absolu correspondant à la translation nulle. Les maxima correspondant aux autres translations entières deviennent seulement locaux avec $\sigma = 2.5$ et disparaissent pratiquement avec $\sigma = 12.8$. Le profil du rapport de corrélation est néanmoins le plus régulier.



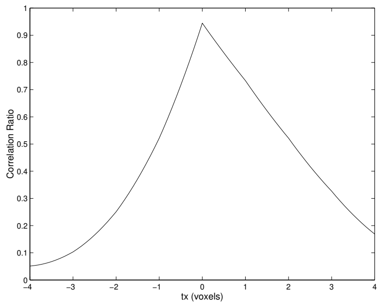
(a) Information mutuelle standard



(b) Information mutuelle avec fenêtrage de Parzen ($\sigma = 2.5$)



(c) Information mutuelle avec fenêtrage de Parzen ($\sigma = 12.8$)

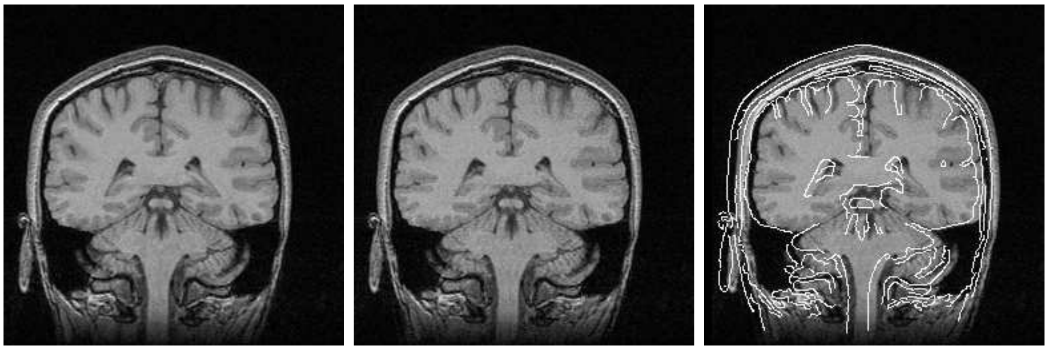


(d) Rapport de corrélation

FIG. 3.2 – *Expérience de la barre grise. Profil de différents critères en fonction de la translation horizontale.*

Expérience de la distorsion invisible Pour cette deuxième expérience, nous avons extrait une coupe coronale d'une IRM cérébrale encodée sur 1 octet. Nous avons ensuite falsifié la coupe de la manière suivante : l'intensité d'un pixel de coordonnées (x,y) est remplacée par celle du pixel de coordonnées $(x-6,y)$ si la différence absolue d'intensité est inférieure à 18.

L'image originale et l'image falsifiée sont représentées sur la figure 3.3 respectivement dans les cadres (a) et (b). Leurs intensités diffèrent pour 62% des pixels mais on constate qu'il est visuellement très difficile de les différencier. En les recalant de façon rigide, on s'attend intuitivement à obtenir la transformation identité. C'est le cas avec le rapport de corrélation mais pas avec l'information mutuelle, dont le maximum correspond à une translation horizontale pure de 6 pixels.



(a) image cible (b) image source (c) résultat du recalage par information mutuelle

FIG. 3.3 – *Expérience de la distorsion invisible. Le recalage par information mutuelle correspond à un décalage horizontal de* 6 *pixels.*

La figure 3.4 représente le profil de ces deux mesures en fonction de la translation horizontale appliquée à l'image falsifiée. Comme précédemment, nous avons utilisé la méthode d'interpolation VP et les histogrammes conjoints sont calculés par regroupement en 256^2 classes. On note que l'information mutuelle possède deux maxima locaux, dont un non global pour la translation nulle (transformation identité).

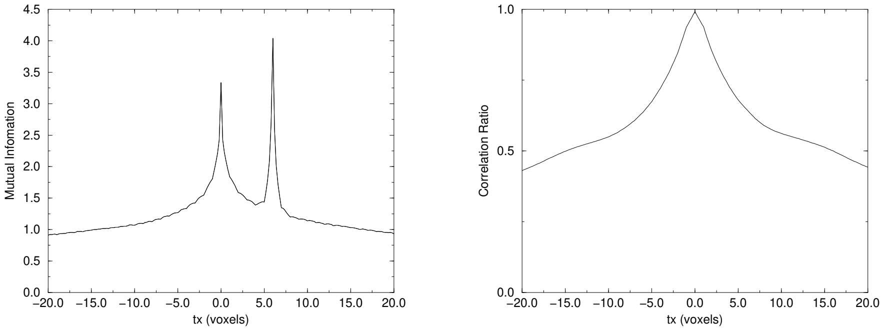


FIG. 3.4 – *Expérience de la distorsion invisible. Profil de l'information mutuelle (à gauche) et du rapport de corrélation (à droite) en fonction de la translation horizontale.*

Nous pouvons expliquer ce résultat en analysant les histogrammes conjoints de la figure 3.5. Celui de gauche correspond à la translation nulle. Comme on s'y attend, les couples d'intensités sont regroupés autour de la première bissectrice du plan, l'écart vertical étant toujours inférieur à 18. Sur l'histogramme de droite, qui correspond à la translation horizontale de +6 pixels, les couples ayant des intensités différentes sont moins nombreux mais ils occupent une région beaucoup plus grande du plan. L'information mutuelle «préfère» cette deuxième situation car elle n'utilise aucune notion de proximité entre les intensités.

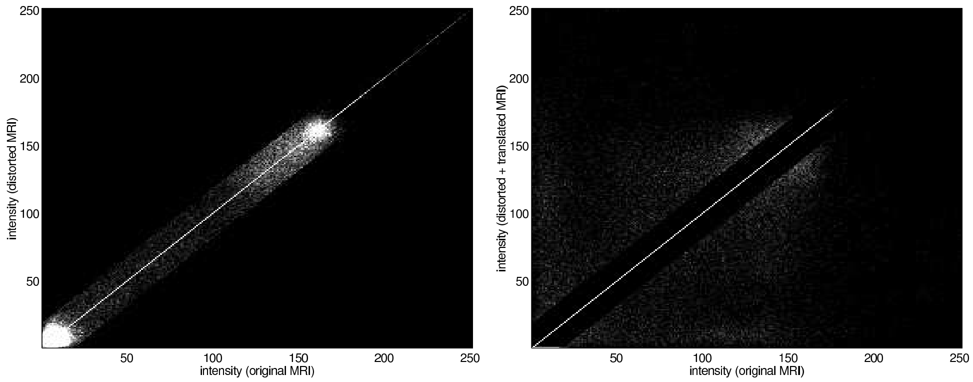


FIG. 3.5 – *Expérience de la distorsion invisible. Histogrammes conjoints correspondant respectivement à la transformation identité et à la translation horizontale pure de +6 pixels.*

3.1.2 Mode d'emploi du chapitre

Ces deux expériences, bien que synthétiques, suggèrent le caractère sous-contraint de l'information mutuelle. Plusieurs auteurs ont observé expérimentalement sa tendance à exhiber de nombreux maxima locaux ainsi que sa sensibilité au sous-échantillonnage et à l'interpolation (Maes et al., 1997; Pluim et al., 1999; Sarrut, 2000). Ces faiblesses contribuent à rendre difficile son optimisation pratique.

Deux réponses complémentaires peuvent être apportées à ce problème. D'une part, la capacité d'un algorithme à trouver le maximum global d'un critère dépend de la procédure d'optimisation qu'il utilise. Il est donc légitime de penser que les performances pratiques de l'information mutuelle seront d'autant meilleures que la procédure d'optimisation est robuste. L'approche que nous avons retenue est néanmoins différente. Plutôt que d'incriminer la méthode d'optimisation, nous avons remis en cause le problème d'optimisation lui-même, c'est-à-dire le choix de la mesure de similarité.

D'après les observations qui ont été faites plus haut, nous allons nous fixer comme objectif de construire une mesure de similarité qui respecte deux exigences apparemment contradictoires : premièrement, être adaptée au recalage multimodal ; deuxièmement, traiter les niveaux de gris comme des variables non pas qualitatives, mais quantitatives. Nous observons que cette deuxième contrainte est vérifiée par certaines mesures de similarité adaptées au recalage monomodal, et notamment par celles d'entre elles qui sont des distances au sens strict. Formellement, de telles mesures sont définies par la donnée d'une norme sur l'espace des intensités (notion que nous préciserons plus loin) :

$$S(T) = \|I - J \circ T\|$$

Par exemple, la somme des différences au carré (SDC) correspond à la norme euclidienne ; la somme des différences absolues (SDA) à la norme L_1 encore appelée distance de Manhattan, etc. (voir le chapitre 2). Si les mesures de ce type ne sont pas applicables au recalage multimodal, c'est parce qu'elles sont calculées à partir de la différence des images, ce qui revient à supposer les intensités comparables. Qu'à cela ne tienne : pour rendre les intensités comparables, il suffit de corriger l'une des images de manière à ce qu'elle «ressemble» à l'autre. Ainsi, le fondement théorique de notre approche est de remplacer l'opérateur de différence par un opérateur du type :

$$(I,J) \mapsto (I - f \circ J \circ T),$$

où $f : \mathbb{R} \to \mathbb{R}$ est une fonction de correction d'intensité, non nécessairement linéaire ou monotone, qu'il s'agira d'estimer. Ce type de transformation d'intensité apparaît, en somme, comme le moyen le plus simple de se ramener au cas du recalage monomodal.

Cette idée est formalisée au paragraphe 3.2. Pour définir rigoureusement notre critère de recalage, il apparaîtra nécessaire de tenir compte de deux aspects pratiques : l'interpolation et le recouvrement partiel. Outre un problème de définition, nous montrerons que le recouvrement partiel pose un problème d'ambiguïté du recalage : plusieurs transformations spatiales peuvent minimiser globalement le critère. Nous utilisons une approche axiomatique pour le normaliser, afin de rejeter a priori celles de ces transformations qui sont manifestement aberrantes. Nous aboutissons finalement à une mesure de similarité générique, robuste au recouvrement partiel, et dont les instanciations particulières dépendent du choix de la norme et de la forme paramétrique de la fonction de correction.

Le paragraphe 3.3 étudie le cas particulier où la norme est euclidienne et la fonction de correction est recherchée constante par morceaux. La mesure qui en découle est le rapport de corrélation «standard», que l'on peut interpréter de manière probabiliste dans le formalisme des histogrammes conjoints. Il s'ensuit une série de comparaisons avec quelques mesures de similarité usuelles : le coefficient de corrélation, le critère de Woods et l'information mutuelle.

Le paragraphe 3.4 généralise l'étude au cas de fonctions paramétriques quelconques et de «normes» robustes (M-estimateurs d'échelle). Enfin, nous abordons au paragraphe 3.5 les aspects liés à la mise en œuvre, en introduisant notamment deux schémas d'optimisation.

3.2 Formulation

Étant données deux images I et J, nous cherchons à construire une mesure de similarité fondée sur une distance entre les intensités de l'image I et de l'image J, cette dernière étant corrigée au moyen d'une fonction $\hat{f} : \mathbb{R} \to \mathbb{R}$. En supposant provisoirement que $\hat{f}$ est connue, nous pouvons définir notre mesure comme :

$$S(T) = \sigma\left(i_1 - \hat{f}(j_1^{\downarrow}), i_2 - \hat{f}(j_2^{\downarrow}), \ldots, i_n - \hat{f}(j_n^{\downarrow})\right),$$

où $\sigma : \mathbb{R}^n \to \mathbb{R}^+$ est une certaine norme de $\mathbb{R}^n$. Nous utilisons toujours les notations simplifiées qui ont été introduites au chapitre 2, paragraphe 2.2. En particulier, $i_k \equiv I(\mathbf{x}_k)$ et $j_k^{\downarrow} \equiv J(T(\mathbf{x}_k))$. Pour fixer les idées, nous supposerons dans ce paragraphe que σ est la norme euclidienne, de sorte que notre critère s'écrit :

$$S(T) = \sum_{k=1}^{n} \left[i_k - \hat{f}(j_k^{\downarrow}) \right]^2 \tag{3.2}$$

La généralisation à d'autres normes est reportée au paragraphe 3.4.2. En fait, nous considèrerons la notion plus générale d'*estimateur d'échelle*, à laquelle nous nous réfèrerons aussi par le terme *métrique*, à comprendre dans un sens plus large que sa définition mathématique.

En première approche, le recalage des images I et J peut être formulé comme la minimisation du critère $S(T)$ défini par l'équation (3.2). On voit que cette façon de résoudre un problème de recalage multimodal ne constitue pas une grande révolution théorique. Les choses vont cependant se compliquer un peu pour la simple raison qu'en pratique, la fonction $\hat{f}$ est le plus souvent inconnue. On peut envisager de la déterminer *a priori* à partir d'abaques fournis par les constructeurs de systèmes d'imagerie, indiquant les valeurs de réponses correspondant aux différents types de tissus anatomiques. Cependant, on dispose rarement de tels abaques. En leur absence, il faut estimer $\hat{f}$ en utilisant la seule information disponible, à savoir les images elles-mêmes.

On peut alors considérer le critère défini à l'équation (3.2) comme une fonction de T et de f, et faire la remarque suivante. Notons $\hat{T}$ la transformation de recalage inconnue. Si, comme nous l'avons implicitement supposé, il existe véritablement une relation fonctionnelle entre les intensités, il est clair que le critère $S(T,f)$ calculé en $T = \hat{T}$ et $f = \hat{f}$ vaut zéro à un ε près lié au bruit d'acquisition et au bruit biologique inhérents aux images. Pour des valeurs différentes de T et/ou de f, on s'attend à ce que $S(T,f)$ augmente, soit à cause des erreurs de recalage, soit à cause des erreurs de correction d'intensité. Il est donc naturel d'estimer $\hat{f}$ conjointement avec $\hat{T}$ en résolvant le problème suivant :

$$\min_{T,f} S(T,f), \qquad \text{avec} \quad S(T,f) \equiv \left\{ \sum_k \left[i_k - f(j_k^{\downarrow}) \right]^2 \right\} \tag{3.3}$$

Par la suite, nous noterons $\mathcal{T}$ et $\mathcal{F}$ les espaces de recherche respectifs de T et de f.

En pratique, le critère $S(T,f)$ ne peut être calculé exactement à cause de la nature finie de l'image cible. Les deux problèmes qui se posent sont les mêmes que ceux que nous avons rencontrés au chapitre 2, paragraphe 2.3, pour définir l'histogramme conjoint de deux images. D'une part, les points transformés $T(\mathbf{x}_k)$ ne coïncident généralement pas avec des nœuds de la grille Ω_J, d'où la nécessité d'une interpolation. D'autre part, certains points $T(\mathbf{x}_k)$ peuvent de surcroix être isolés, au sens où leur information de voisinage est insuffisante pour réaliser l'interpolation (voir chapitre 2, paragraphe 2.3.3).

3.2.1 Interpolation

Si un point $T(\mathbf{x}_k)$ possède suffisamment de voisins de grille, il est possible d'interpoler son intensité $j_k^{\downarrow}$. Une méthode classique et efficace en temps de calcul est celle de l'interpolation spline d'ordre 1 (interpolation trilinéaire en 3D, bilinéaire en 2D). Le nombre de voisins de grille nécessaires est alors 8 en 3D et 4 en 2D.

Une stratégie alternative consiste à interpoler, non pas l'intensité $j_k^\downarrow$, mais la contribution incrémentale du point, $[i_k - f(j_k^\downarrow)]^2$. Nous parlons alors d'interpolation intrinsèque. La différence entre ces deux approches est illustrée par la figure 3.6. Nous verrons un peu plus loin que la deuxième méthode revient en fait à l'interpolation par volume partiel (VP) proposée par (Collignon et al., 1995a; Maes et al., 1997) dans le contexte du calcul d'histogrammes conjoints.

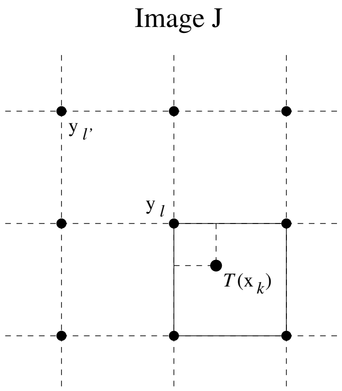
Nous aboutissons à une première adaptation du critère de recalage selon la méthode d'interpolation choisie :

- Pour une interpolation de l'intensité,

$$S(T,f) = \sum_k \Big[i_k - f\big(\sum_\ell w_{k\ell}^\downarrow j_\ell\big)\Big]^2 \tag{3.4}$$

- Pour une interpolation intrinsèque,

$$S(T,f) = \sum_k \sum_\ell w_{k\ell}^\downarrow \Big[i_k - f(j_\ell)\Big]^2 \tag{3.5}$$



Poids :

$w_{k\ell}^\downarrow = \beta(T(\mathbf{x}_k) - \mathbf{y}_\ell)$, avec :

$$\begin{cases} \beta(\mathbf{z}) = (1 - |z^1|)(1 - |z^2|), & \text{si } |z^1| \leq 1, |z^2| \leq 1 \\ \beta(\mathbf{z}) = 0, & \text{sinon} \end{cases}$$

Interpolation de l'intensité :

$$\Big[i_k - f(j_k^\downarrow)\Big]^2 \approx \Big[i_k - f\big(\sum_\ell w_{k\ell}^\downarrow j_\ell\big)\Big]^2$$

Interpolation intrinsèque :

$$\Big[i_k - f(j_k^\downarrow)\Big]^2 \approx \sum_\ell w_{k\ell}^\downarrow \Big[i_k - f(j_\ell)\Big]^2$$

FIG. 3.6 – *Illustration de l'interpolation spline d'ordre* 1 *dans le cas 2D (interpolation bilinéaire), et deux stratégies correspondantes d'interpolation du critère de recalage.*

3.2.2 Normalisation

Si nous décidons de ne pas tenir compte des points isolés dans le calcul du critère de recalage, une difficulté apparaît : $S(T,f)$ peut prendre des valeurs arbitrairement faibles pour des transformations spatiales aberrantes, et ce quelle que soit la méthode d'interpolation choisie. Imaginons le cas extrême d'une transformation pour laquelle seule une région parfaitement uniforme de I est appariée à l'intérieur du domaine délimité par J. On voit qu'il existe alors une fonction f pour laquelle $S(T,f) = 0$. Autrement dit, on obtient une correspondance d'intensité parfaite dans la zone de recouvrement entre les images, et le couple (T,f) réalise un minimum absolu du critère !

En pratique, on aboutit facilement à ce genre de configuration si l'algorithme de recalage part d'une position relativement éloignée de la position idéale. Il peut alors basculer dans un bassin d'attraction plus proche qui correspond, typiquement, à apparier le fond d'une image avec différentes strucures de l'autre image. Une simple translation peut faire l'affaire. La problématique que nous sommes en train d'esquisser est celle de la robustesse d'une mesure de similarité vis-à-vis du recouvrement partiel. On notera que c'est précisément pour garantir cette propriété que (Viola, 1995; Collignon et al., 1995a; Studholme et al., 1998) ont motivé l'emploi de l'information mutuelle au détriment d'autres critères entropiques.

Dans notre cas particulier, la difficulté provient du fait qu'imposer aux régions d'être liées par une relation de type $I \approx f(J)$ ne suffit pas à garantir une véritable dépendance fonctionnelle. Il existe des cas de dépendance «dégénérée» : ceux pour lesquels f est une fonction constante. Ces cas se produisent notamment lorsque I est uniforme dans la zone de recouvrement. Les transformations qui mènent à cette situation doivent être considérées comme aberrantes et il apparaît souhaitable de les rejeter *a priori*.

Pour une transformation T donnée, nous désignons par $I^{\downarrow}$ la sous-image de I qui est appariée à l'intérieur du domaine de J. Nous voulons tester si la dépendance fonctionnelle entre $I^{\downarrow}$ et J est dégénérée, autrement dit si la fonction réalisant le minimum de $S(T,f)$ à T fixé est constante ou non. Pour le faire, nous pouvons comparer la valeur minimale du critère à celle qu'on obtiendrait en cherchant f comme une fonction constante. Nous définissons :

$$\begin{cases} S(T) & \equiv & \min_f S(T,f) \\ S_0(T) & \equiv & \min_{f=\text{cte}} S(T,f) \end{cases}$$

A condition que l'ensemble des fonctions constantes soit inclus dans l'espace de recherche de f, ce que nous supposerons, l'inégalité suivante est toujours vraie :

$$0 \leq S(T) \leq S_0(T)$$

Le cas d'égalité se produit précisément lorsque la dépendance est dégénérée, c'est-à-dire lorsque la meilleure approximation de $I^{\downarrow}$ par J est une fonction constante. Nous allons maintenant normaliser le critère en cherchant une mesure g fonction de $S(T)$ et de $S_0(T)$ qui satisfait aux axiomes suivants :

($A1$) La fonction $g(S,S_0)$ prend ses valeurs entre 0 et 1

($A2$) A S_0 fixé, $g(S,S_0)$ est une fonction croissante de S

($A3$) Si $S = 0$ et $S_0 \neq 0$ alors $g(S,S_0) = 0$

($A4$) Si $S = S_0$, alors $g(S,S_0) = 1$

Pour respecter l'axiome ($A2$), nous pouvons chercher g comme une fonction affine de S, c'est-à-dire $g(S,S_0) = a(S_0)S + b(S_0)$. On voit alors que :

$$\begin{array}{lll} (A3) & \Rightarrow & b(S_0) = 0 \quad \text{si } S_0 \neq 0 \\ (A4) & \Rightarrow & a(S_0) = \dfrac{1}{S_0} \quad \text{si } S_0 \neq 0 \end{array}$$

Par conséquent, $g(S,S_0) = S/S_0$ si $S_0 \neq 0$. Si $S_0 = 0$, le fait que l'on ait toujours $0 \leq S \leq S_0$ implique $S = 0$ et l'axiome ($A4$) impose donc $g(S,0) = 1$. L'axiome ($A1$) est alors automatiquement vérifié, là encore parce que $0 \leq S \leq S_0$.

Notre mesure normalisée est donc :

$$\tilde{S}(T) = \frac{S(T)}{S_0(T)} = \frac{\min_f S(T,f)}{\min_{f=\mathrm{cte}} S(T,f)} \tag{3.6}$$

avec $\tilde{S}(T) = 1$ si $S_0(T) = 0$. Si, pour une transformation donnée, il existe une correspondance d'intensité parfaite entre les images, alors $\tilde{S}(T)$ vaut 0 à condition que le dénominateur soit non nul. Les transformations pour lesquelles la dépendance fonctionnelle est dégénérée donnent une valeur de 1 et correspondent donc à des *maxima* absolus.

Remarquons qu'en normalisant la mesure, nous avons aussi un peu altéré notre formulation initiale. La mesure non-normalisée $S(T,f)$ était vue comme une fonction de (T,f) à minimiser sur $\mathcal{T} \times \mathcal{F}$. La mesure normalisée ne dépend, elle, que de T : la variable f est cachée du fait que nous lui imposons de minimiser $S(T,f)$ à T fixé. Cette distinction n'est qu'apparente et nous y reviendrons au paragraphe 3.5.

3.2.3 Formulations analogues

Certains auteurs ont déjà exploité une estimation aux moindres carrés semblable à l'équation (3.3). Dans un contexte de recalage monomodal, (Barber, 1992; Feldmar et al., 1997) considèrent le cas où la fonction d'intensité est recherchée dans un espace de fonctions polynômiales ; ils se restreignent en pratique aux fonctions affines. De leur côté, (Friston et al., 1995) envisagent une forme paramétrique plus générale incluant des fonctions instationnaires, c'est-à-dire des fonctions $f(i,\mathbf{x})$ qui dépendent à la fois de l'intensité et de la position. La méthode est ensuite appliquée au recalage rigide TEP/IRM avec une forme paramétrique très spécifique. Enfin, (Viola, 1995) a étudié une formulation analogue sans doute avant ses travaux sur l'information mutuelle. Pour estimer la fonction d'intensité, il adopte une approche non-paramétrique avec régularisation.

Sans passer explicitement par cette formulation, d'autres auteurs ont proposé des mesures heuristiques qui s'avèrent étroitement liées à une estimation aux moindres carrés (Woods et al., 1993; Ardekani et al., 1995; Alpert et al., 1996; Nikou et al., 1999). En ce qui concerne la mesure proposée par (Ardekani et al., 1995), c'en est même un cas particulier. Nous reviendrons plus précisemment sur ces mesures au paragraphe 3.3.4.

La principale différence qui existe entre ces approches et la nôtre réside dans la prise en compte des effets de recouvrement partiel entre les images. C'est pour prévenir les effets d'occultation liés au recouvrement partiel que nous avons normalisé la mesure quadratique (paragraphe 3.2.2). Le problème est également traité, mais de façon différente, par (Feldmar et al., 1997). Leur approche repose sur une adaptation robuste de l'algorithme ICP 4D utilisé pour la minimisation [1]. Viola se borne pour sa part à constater le manque de robustesse de sa mesure quadratique vis-à-vis du recouvrement partiel. Il est d'ailleurs intéressant de noter que c'est l'une des raisons qu'il donne pour lui préférer l'information mutuelle.

1. Cet algorithme a été succintement décrit au chapitre 2, paragraphe 2.5.3.1.

3.3 Cas particulier : métrique quadratique, fonctions constantes par morceaux

Pour l'instant, nous n'avons pas spécifié l'espace de recherche $\mathcal{F}$ de la fonction d'intensité. Nous considérons dans ce paragraphe le cas de fonctions constantes par morceaux. Ce choix offre un double avantage, l'un pratique, l'autre théorique. D'une part, il mène à un calcul rapide de la mesure $\tilde{S}(T)$. D'autre part, il permet de l'appréhender à travers la notion d'histogramme conjoint. Elle apparaît comme un critère de liaison bien connu des statisticiens : le rapport de corrélation.

3.3.1 Calcul explicite

Quelle que soit la méthode d'interpolation choisie, trouver la fonction f qui minimise $S(T,f)$ à T fixé n'est autre qu'un problème de régression aux moindres carrés. On peut s'en convaincre en considérant les équations (3.4) et (3.5). On note que dans le cas de l'interpolation VP, les résidus au carré sont pondérés.

Dans le cas où f est recherchée comme une combinaison linéaire de fonctions élémentaires ($\mathcal{F}$ est un espace vectoriel), il est bien connu que le problème de minimisation revient à la résolution d'un système d'équations linéaires. En général, il faut calculer et inverser une matrice symétrique. Une simplification importante apparaît lorsque f est recherchée comme une fonction constante par morceaux. En introduisant une partition de $\mathbb{R}$ en intervalles de taille constante β, une telle fonction peut être définie de la manière suivante :

$$\forall q \in \mathbb{Z}, \qquad f(j) = \mu_q, \qquad \forall j \in \mathbb{R}, \quad q - \frac{1}{2} \leq \frac{j}{\beta} < q + \frac{1}{2}$$

Si f admet cette expression, le critère $S(T,f)$ peut être réécrit sous une forme simplifiée. Prenons le cas d'une interpolation trilinéaire de l'intensité (3.4). Pour une transformation T donnée, nous définissons les ensembles :

$$\Omega_I^{\downarrow q} = \left\{ \mathbf{x}_k \in \Omega_I, \quad q - \frac{1}{2} \leq \frac{j_k^{\downarrow}}{\beta} < q + \frac{1}{2} \right\}$$

où $j_k^{\downarrow}$ est l'intensité interpolée d'un point $T(\mathbf{x}_k)$. Ces ensembles forment une partition de l'ensemble $\Omega_I^{\downarrow}$ des voxels $\mathbf{x}_k$ dont les correspondants sont non-isolés (dans le cas où on extrapole l'image J, $\Omega_I^{\downarrow} = \Omega_I$ et est indépendant de T). On a :

$$\begin{aligned} S(T,f) &= \sum_{k \in \Omega_I^{\downarrow}} \left[i_k - f(j_k^{\downarrow}) \right]^2 \\ &= \sum_{q \in \mathbb{Z}} \sum_{k \in \Omega_I^{\downarrow q}} (i_k - \mu_q)^2 \end{aligned}$$

On voit alors que la minimisation du critère en fonction des paramètres μ_q conduit à plusieurs minimisations indépendantes :

$$\forall q \in \mathbb{Z}, \quad \Omega_I^{\downarrow q} \neq \emptyset \;\Rightarrow\; \hat{\mu}_q = \arg\min_{\mu} \sum_{k \in \Omega_I^{\downarrow q}} (i_k - \mu_q)^2$$

dont les solutions sont, en posant $n_q^\downarrow \equiv \mathrm{Card}\Omega_I^{\downarrow q}$:

$$\hat{\mu}_q = \frac{1}{n_q^\downarrow} \sum_{k \in \Omega_I^{\downarrow q}} i_k \tag{3.7}$$

Le fait que certains ensembles $\Omega_I^{\downarrow q}$ puissent être vides implique la non-unicité de la solution. Néanmoins, toutes les solutions donnent la même valeur minimale du critère :

$$\min_f S(T,f) = \sum_q \sum_{k \in \Omega_I^{\downarrow q}} (i_k - \hat{\mu}_q)^2 \tag{3.8}$$

Nous savons donc calculer le numérateur de la mesure $\tilde{S}(T)$ définie à l'équation (3.6). Le calcul du dénominateur s'avère très similaire. On vérifie sans peine que la «fonction» constante qui minimise $S(T,f)$ est la moyenne des intensités de I calculée sur l'ensemble $\Omega_I^\downarrow$. La valeur correspondante du critère est le produit de leur variance par $n^\downarrow = \mathrm{Card}\Omega_I^\downarrow$:

$$\min_{f=\mathrm{cte}} S(T,f) = n^\downarrow \mathrm{Var}(I^\downarrow) \tag{3.9}$$

Les formules (3.7), (3.8) et (3.9) sont vraies pour une interpolation trilinéaire et, plus généralement, pour toute méthode d'interpolation de l'intensité. Dans le cas de l'interpolation VP, qui correspond à la définition de $S(T,f)$ donnée par l'équation (3.5), seule (3.9) reste vraie. On vérifie que (3.7) et (3.8) sont adaptées de la façon suivante :

$$\begin{cases} \hat{\mu}_q = \displaystyle\sum_{k \in \Omega_I^\downarrow} \sum_{\ell \in \Omega_J^q} w_{k\ell}^\downarrow i_k \Big/ \sum_{k \in \Omega_I^\downarrow} \sum_{\ell \in \Omega_J^q} w_{k\ell}^\downarrow \\[2ex] \displaystyle\min_f S(T,f) = \sum_q \sum_{k \in \Omega_I^\downarrow} \sum_{\ell \in \Omega_J^q} w_{k\ell}^\downarrow (i_k - \hat{\mu}_q)^2 \end{cases}$$

où les ensembles Ω_J^q sont définis par $\Omega_J^q = \left\{ \mathbf{y}_\ell \in \Omega_J, \quad q - \frac{1}{2} \leq \frac{j_\ell}{\beta} < q + \frac{1}{2} \right\}$.

Pour la clarté de l'exposé, nous avons supposé ici que les intervalles étaient de taille constante. Il est clair que cette contrainte n'est pas fondamentale et que les formules précédentes restent valables dans le cas d'intervalles de taille variable, à condition qu'ils forment une partition de $\mathbb{R}$ (et notamment qu'ils ne se recoupent pas).

3.3.2 Interprétation probabiliste : le rapport de corrélation

Nous allons voir maintenant que notre mesure de similarité s'interprète comme un rapport de corrélation (Saporta, 1990). Supposons pour fixer les idées que les images sont encodées sur un octet, de sorte qu'elles prennent des valeurs entières comprises entre 0 et 255, et choisissons $\beta = 1$. Nous nous plaçons d'abord dans le cadre d'une interpolation de l'intensité. Le fait que f soit recherchée comme une fonction constante par morceaux de taille $\beta = 1$ et que les images prennent des valeurs entières nous permet d'écrire :

$$\begin{aligned} S(T,f) &= \sum_k \left[i_k - f(j_k^{\downarrow})\right]^2 \\ &= \sum_k \sum_{j\in\mathbb{Z}} K(j - j_k^{\downarrow})\,[i_k - f(j)]^2 \\ &= \sum_k \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} K(i - i_k)K(j - j_k^{\downarrow})\,[i - f(j)]^2 \end{aligned}$$

où K est la fonction indicatrice de l'intervalle $[-\frac{1}{2},\frac{1}{2}[$. Nous reconnaissons ici l'histogramme conjoint des images dans sa définition correspondant à une interpolation de l'intensité (voir chapitre 2, paragraphe 2.3) :

$$p_{i,j}(T) = \frac{1}{n^{\downarrow}} \sum_{k\in\Omega_I^{\downarrow}} K(i - i_k)K(j - j_k^{\downarrow})$$

D'où la relation :

$$S(T,f) = n^{\downarrow} \sum_{i,j} p_{i,j}\,[i - f(j)]^2 \tag{3.10}$$

On vérifie aisément que cette relation reste valable dans le cas d'une interpolation intrinsèque du critère $S(T,f)$, voir l'équation (3.5). La définition correspondante de l'histogramme conjoint est celle d'une interpolation VP :

$$p_{i,j}(T) = \frac{1}{n^{\downarrow}} \sum_{k\in\Omega_I^{\downarrow}} \sum_{\ell} w_{k\ell}^{\downarrow} K(i - i_k)K(j - j_\ell)$$

Nous nous rappelons maintenant que l'histogramme conjoint peut être interprété comme la densité de probabilité d'un couple de variables aléatoires. Notons $\tilde{I}$ et $\tilde{J}$ ces variables[2], en sous-entendant le fait qu'elles dépendent de T. Le critère $S(T,f)$ s'interprète à son tour, au terme $n^{\downarrow}$ près, comme une erreur quadratique moyenne (EQM) :

$$S(T,f) = n^{\downarrow}\, E\left[\tilde{I} - f(\tilde{J})\right]^2$$

où E désigne l'opérateur d'espérance mathématique. Les résultats suivants, issus de la théorie du conditionnement (Saporta, 1990; Papoulis, 1991), sont détaillés dans l'annexe A :

- La variable $f(\tilde{J})$ qui minimise l'EQM (bien entendu, à T fixé) est l'espérance conditionnelle de $\tilde{I}$ sachant $\tilde{J}$. On la note généralement $E(\tilde{I}|\tilde{J})$. La valeur correspondante de l'EQM est appelée variance conditionnelle et notée $\mathrm{Var}(\tilde{I}|\tilde{J})$.
- La constante qui minimise l'EQM est l'espérance de $\tilde{J}$. La valeur correspondante de l'EQM est la variance de $\tilde{I}$.

2. En effet, il ne serait pas tout à fait correct d'assimiler les variables aléatoires aux images elles-mêmes. On trouvera une définition plus rigoureuse de $\tilde{I}$ et $\tilde{J}$ dans (Roche et al., 1998a).

- Le rapport entre ces deux EQM minimales définit le rapport de corrélation :

$$\eta^2_{\tilde{I}|\tilde{J}} = 1 - \frac{\mathrm{Var}(\tilde{I}|\tilde{J})}{\mathrm{Var}(\tilde{I})} \tag{3.11}$$

Par conséquent, la mesure $\tilde{S}(T)$ définie par l'équation (3.6) est égale à $1 - \eta^2_{\tilde{I}|\tilde{J}}$. Minimiser $\tilde{S}(T)$ en fonction de T est donc équivalent à maximiser le rapport de corrélation. Dans le contexte où le rapport de corrélation est une mesure de similarité, il sera noté $\eta^2_{I|J}(T)$.

Le rapport de corrélation est un concept probabiliste qui se justifie dans la mesure où, par l'intermédiaire d'un histogramme conjoint, nous pouvons interpréter les images comme des variables aléatoires. A vrai dire, nous aurions pu épargner au lecteur cette incursion dans le monde des probabilités. Nous ne l'avons pas fait pour deux raisons : d'une part, justifier l'appellation «rapport de corrélation» ; d'autre part, établir un rapprochement entre notre formulation et la notion de critère de liaison probabiliste. Ce point de vue va nous être utile pour comparer le rapport de corrélation avec d'autres mesures d'origine probabiliste : le coefficient de corrélation, le critère de Woods et l'information mutuelle.

L'équivalence entre la maximisation du rapport de corrélation et le problème formulé par l'équation (3.6) impose une certaine façon de construire l'histogramme conjoint. Plus généralement, on peut calculer un rapport de corrélation dès qu'on a défini un histogramme conjoint en utilisant, par exemple, les formules suivantes :

$$1 - \eta^2_{I|J}(T) = \frac{1}{\sigma^2} \sum_j p_j \sigma_j^2 \tag{3.12}$$

avec :

$$\left\{ \begin{array}{ll} \sigma^2 = \sum_i p_i\, i^2 \; - \mu^2 & \mu = \sum_i p_i\, i \\ \sigma_j^2 = \sum_i p_{i|j}\, i^2 \; - \mu_j^2 & \mu_j = \sum_i p_{i|j}\, i \\ p_i = \sum_j p_{i,j} & p_{i|j} = \dfrac{p_{i,j}}{p_j} \end{array} \right.$$

On retiendra néanmoins qu'il est possible de s'affranchir du calcul explicite d'un histogramme conjoint, comme le montre le paragraphe 3.3.1. Les deux procédures de calcul ne sont pas équivalentes en termes de complexité calculatoire.

3.3.3 Lien avec le coefficient de corrélation

En tant que critère de liaison probabiliste, le rapport de corrélation est connu pour être la généralisation naturelle du coefficient de corrélation, mesure de similarité classique dans le contexte du recalage monomodal (Brown, 1992). La relation qui existe entre ces deux mesures est étudiée en détails dans l'annexe A.

Le coefficient de corrélation peut lui aussi être défini à partir de l'histogramme conjoint des images :

$$\rho^2_{I,J}(T) = \frac{\mathrm{Cov}(\tilde{I},\tilde{J})}{\sqrt{\mathrm{Var}(\tilde{I})\mathrm{Var}(\tilde{J})}}$$

Il existe une définition équivalente, peu intéressante d'un point de vue calculatoire, mais instructive d'un point de vue théorique (voir annexe A) :

$$1 - \rho^2_{I,J}(T) = \frac{\min_{a,b} E\left[\tilde{I} - a\tilde{J} - b\right]^2}{\mathrm{Var}(\tilde{I})}$$

Autrement dit, le coefficient de corrélation est la mesure que nous aurions obtenue en recherchant la fonction d'intensité f comme une fonction *affine*. Or, nous l'avons supposée constante par morceaux, ce qui, dans le contexte discret d'un histogramme conjoint, revient à la supposer quelconque. On conçoit ainsi que le coefficient de corrélation est une mesure de dépendance affine[3] alors que le rapport de corrélation est une mesure de dépendance fonctionnelle d'ordre général. Ce parallèle suggère l'existence d'un continuum de mesures que l'on peut obtenir en jouant sur l'espace de recherche $\mathcal{F}$.

3.3.4 Lien avec le critère de Woods et ses variantes

Le critère de (Woods et al., 1993) est sans doute la première mesure de similarité qui ait été proposée pour le recalage multimodal d'images médicales. En reprenant les notations introduites au paragraphe 3.3.2, il est défini par :

$$W(T) = \sum_j p_j \frac{\sigma_j}{\mu_j} \tag{3.13}$$

Cette expression est à comparer avec la caractérisation (3.12) du rapport de corrélation. On voit que les deux équations expriment fondamentalement la même idée. La première différence réside dans le fait que le rapport de corrélation fait intervenir une somme de variances (σ_j^2) alors que le critère de Woods somme des écarts-types normalisés (σ_j/μ_j). La division par la moyenne μ_j est justifiée par Woods et al. comme un moyen de garantir l'invariance du critère par transformation linéaire des intensités de l'image I. Le rapport de corrélation fait apparaître une normalisation globale par σ^2 qui conduit à la même propriété d'invariance (et même celle d'invariance affine), mais dont la raison d'être véritable est de garantir la robustesse au recouvrement partiel des images (voir paragraphe 3.2.2).

Récemment, (Jenkinson et Smith, 2000) ont introduit la notion de comportement asymptotique d'une mesure de similarité, notion destinée à caractériser la robustesse au recouvrement partiel. Ils montrent en particulier que le rapport de corrélation possède un comportement asymptotique idéal et que ce n'est pas le cas du critère de Woods. La normalisation par μ_j n'aurait donc pas une grande utilité, d'autant moins qu'elle a tendance à produire des instabilités numériques.

3. On parle souvent de dépendance linéaire, mais c'est un abus de langage.

Du reste, parmi les auteurs qui ont adapté le critère de Woods, certains ont supprimé la normalisation. Outre cet aspect, (Ardekani et al., 1995) remplacent les écart-types par des variances dans la formule (3.13). Ainsi, leur mesure est équivalente au numérateur du rapport de corrélation (3.12) et elle correspond donc à un cas particulier de la formulation générale aux moindres carrés dont nous sommes partis : celui des fonctions constantes par morceaux.

Une autre adaptation du critère de Woods a été proposée par (Nikou et al., 1999). Là encore, la division par μ_j est supprimée, et les écart-types σ_j sont estimés de façon robuste au moyen de M-estimateurs. Nous utilisons une idée similaire pour généraliser le rapport de corrélation au paragraphe 3.4.2.

3.3.5 Lien avec l'information mutuelle

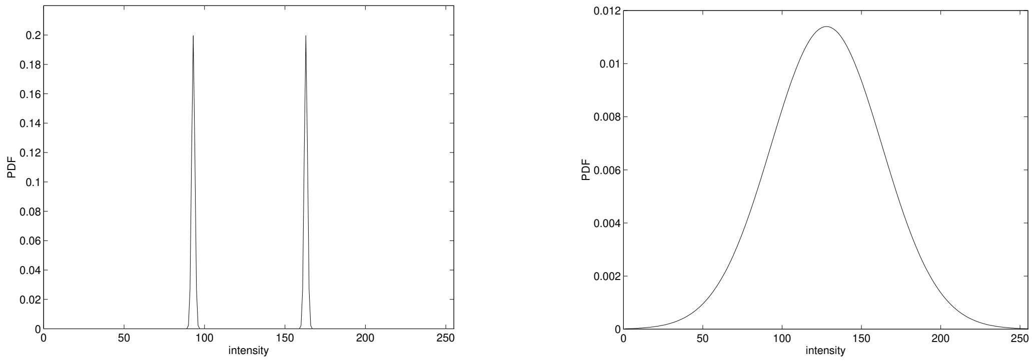
Il existe une analogie formelle remarquable entre le rapport de corrélation et l'information mutuelle. Nous avons vu que le rapport de corrélation était étroitement lié à la notion de variance. La variance est une mesure d'incertitude bien connue, mais il en existe d'autres : par exemple, l'entropie (Blahut, 1987; Cover et Thomas, 1991). Les expressions suivantes montrent que l'entropie joue pour l'information mutuelle un rôle très similaire à celui que joue la variance pour le rapport de corrélation :

$$\left\{ \begin{array}{lcl} 1 - \eta^2(\tilde{I}|\tilde{J}) & = & \dfrac{\mathrm{Var}(\tilde{I}|\tilde{J})}{\mathrm{Var}(\tilde{I})} \\ -\mathcal{I}(\tilde{I},\tilde{J}) & = & H(\tilde{I}|\tilde{J}) - H(\tilde{I}) \end{array} \right.$$

Aux changements de variable près, le rapport de corrélation se présente comme le rapport entre une variance conditionnelle et une variance marginale alors que l'information mutuelle est la différence entre une entropie conditionnelle et une entropie marginale. Ces deux mesures s'interprètent de manière analogue. D'une part, elles encouragent les transformations pour lesquelles l'image source est «bien expliquée» par l'image cible. D'autre part, elles font intervenir une mesure de dispersion marginale dont l'effet est de pénaliser les transformations pour lesquelles l'image source devient uniforme dans la zone de recouvrement.

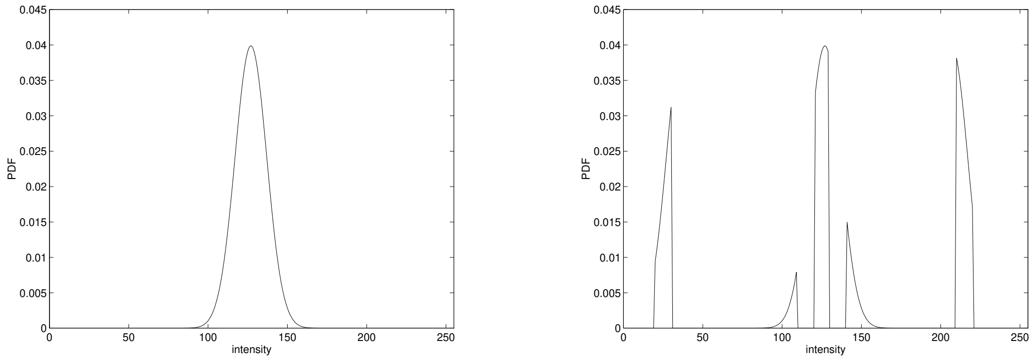
Il serait cependant hâtif de conclure à l'équivalence du rapport de corrélation et de l'information mutuelle. La raison en est que la variance et l'entropie ne sont pas des mesures d'incertitude équivalentes, comme cela est illustré par les figures 3.7 et 3.8. La variance caractérise les fluctuations d'une variable aléatoire autour d'une tendance moyenne. Elle est basée sur la distance quadratique dans $\mathbb{R}$. L'entropie n'utilise quant à elle aucune notion de distance ; c'est une mesure du caractère prédictible d'une variable aléatoire.

En dehors de ce distinguo, on peut également se demander pourquoi le rapport de corrélation repose sur une division alors que l'information mutuelle fait intervenir une différence. Définie sous cette forme, l'information mutuelle a la propriété satisfaisante pour l'esprit d'être une fonction symétrique de $\tilde{I}$ et $\tilde{J}$, ce qui n'est pas le cas du rapport de corrélation. Certains auteurs ont néanmoins envisagé des variantes de l'information mutuelle. Ainsi, (Sarrut et Miguet, 1999; Sarrut, 2000) remarquent que la mesure correspondant au rapport des deux entropies est un coefficient asymétrique



(a) Faible entropie (b) Forte entropie

FIG. 3.7 – *Ces deux densités de probabilité ont la même variance mais des entropies différentes.*



(a) Faible variance (b) Forte variance

FIG. 3.8 – *Ces deux densités de probabilité ont la même entropie mais des variances différentes.*

connu sous le nom de u de Theil :

$$u(\tilde{I},\tilde{J}) = 1 - \frac{H(\tilde{I}|\tilde{J})}{H(\tilde{I})}$$

Ils montrent que cette mesure, ainsi que le rapport de corrélation, sont des cas particuliers du concept plus général de mesure RPE (réduction proportionnelle de l'erreur de prédiction). Ils utilisent également la version symétrique du coefficient de Theil, le coefficient de corrélation entropique :

$$k(\tilde{I},\tilde{J}) = 1 - \frac{H(\tilde{I}|\tilde{J}) + H(\tilde{J}|\tilde{I})}{H(\tilde{I}) + H(\tilde{J})} = \frac{2\mathcal{I}(\tilde{I},\tilde{J})}{H(\tilde{I}) + H(\tilde{J})}$$

Leurs expériences de recalage tendent à prouver que ces trois versions de l'information mutuelle sont sensiblement équivalentes en pratique. L'utilisation du coefficient k avait été précédemment motivée par (Studholme et al., 1998) sur la base d'observations théoriques et empiriques indiquant une robustesse supérieure à l'information mutuelle standard vis-à-vis du recouvrement partiel des images[4].

3.4 Rapport de corrélation généralisé

Nous abordons ici deux manières possibles de construire une variante du rapport de corrélation :

1. En considérant un espace de recherche $\mathcal{F}$ autre que celui des fonctions constantes par morceaux.
2. En utilisant une métrique de comparaison d'intensité autre que la norme quadratique.

Ces deux généralisations sont motivées et développées dans cet ordre aux paragraphes 3.4.1 et 3.4.2.

3.4.1 Espace de fonctions

Plus généralement que des fonctions constantes par morceaux, nous pouvons considérer des fonctions de la forme :

$$f(j) = \sum_{p=1}^{P} c_p \, \phi_p(j)$$

où les fonctions élémentaires ϕ_p sont supposées former une famille libre. Le cas des fonctions constantes par morceaux correspond à choisir les ϕ_p comme des fonctions en créneau.

Considérons le problème général de minimiser $S(T,f)$ sur un tel espace de fonctions. Nous ne traitons ici que le cas d'une interpolation trilinéaire. Le cas de l'interpolation VP est similaire et il est détaillé dans l'annexe B.1. En dérivant l'équation (3.4) par rapport aux coefficients c_p, nous

4. Dans cet article, le coefficient k est appelé *information mutuelle normalisée*.

obtenons :

$$\begin{aligned}
\frac{1}{2}\frac{\partial S}{\partial c_p} &= \sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \phi_p(j_k^\downarrow) \left[i_k - \sum_{q=1}^{P} c_q\, \phi_q(j_k^\downarrow) \right] \\
&= \sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \phi_p(j_k^\downarrow)\, i_k - \sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \sum_{q=1}^{P} c_q\, \phi_p(j_k^\downarrow)\phi_q(j_k^\downarrow) \\
&= \sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \phi_p(j_k^\downarrow)\, i_k - \sum_{p=1}^{P} c_q \left[\sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \phi_p(j_k^\downarrow)\phi_q(j_k^\downarrow) \right]
\end{aligned}$$

Les coefficients sont donc solutions du système linéaire $\mathbf{Ac} = \mathbf{b}$, avec :

$$\mathbf{c} = \begin{pmatrix} c_1 & c_2 & \dots & c_P \end{pmatrix}^t$$

$$\mathbf{A} = \begin{pmatrix}
\sum_k \phi_1(j_k^\downarrow)^2 & \sum_k \phi_1(j_k^\downarrow)\phi_2(j_k^\downarrow) & \dots & \sum_k \phi_1(j_k^\downarrow)\phi_P(j_k^\downarrow) \\
\sum_k \phi_2(j_k^\downarrow)\phi_1(j_k^\downarrow) & \sum_k \phi_2(j_k^\downarrow)^2 & \dots & \sum_k \phi_2(j_k^\downarrow)\phi_P(j_k^\downarrow) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_k \phi_1(j_k^\downarrow)\phi_P(j_k^\downarrow) & \dots & \dots & \sum_k \phi_P(j_k^\downarrow)^2
\end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} \sum_k \phi_1(j_k^\downarrow)\, i_k & \sum_k \phi_2(j_k^\downarrow)\, i_k & \dots & \sum_k \phi_P(j_k^\downarrow)\, i_k \end{pmatrix}^t$$

Il existe plusieurs méthodes pour inverser la matrice $\mathbf{A}$. Pour des raisons de stabilité numérique, celle que nous utilisons est la méthode de décomposition en valeurs singulières (DVS). La décomposition est immédiate dans le cas des fonctions constantes par morceaux car $\mathbf{A}$ est alors diagonale. On conçoit que pour des fonctions ϕ_p qui ne sont pas à support compact, le calcul de $\mathbf{A}$ (plus que son inversion elle-même) peut s'avérer très coûteux. Il est alors souhaitable de réduire au maximum le nombre d'appels à la procédure de régression linéaire. Dans ce but, nous serons amenés à modifier notre schéma d'optimisation (voir le paragraphe 3.5).

Quel est le choix le plus judicieux pour l'espace de fonctions ? Le fait de considérer des fonctions constantes par morceaux revient à corriger indépendamment les différentes classes d'intensité de l'image «explicative». Cette approche nous paraît particulièrement indiquée lorsque les classes correspondent rigoureusement à des structures anatomiques distinctes. C'est typiquement le cas d'un atlas labélisé. Dans les images réelles, le regroupement des voxels en fonction de leur seule intensité ne suffit généralement pas à produire une segmentation anatomique idéale.

L'exemple emblématique d'une telle méthode de segmentation est l'algorithme des k-moyennes. Il peut éventuellement servir à optimiser le choix des intervalles d'intensité (Ardekani et al., 1995), mais il s'avère peu robuste aux différents artefacts que l'on rencontre inéluctablement dans les images :

bruit de mesure, bruit biologique, hétérogénéités diverses et variées (biais du champ magnétique en IRM, scintillement et effets d'ombrage en imagerie ultrasonore, etc.).

Des pré-traitements plus robustes peuvent être envisagés : lissage, de préférence anisotrope (Perona et Malik, 1990; Gerig et al., 1992; Krissian et al., 1996) et/ou segmentation avec correction de biais (Wells et al., 1996a; Guillemaud et Brady, 1997; Rajapakse et al., 1998; Van Leemput et al., 1999; Kapur, 1999; Zhang et al., 2000; Prima, 2000). Il importe de noter, d'une part, que ces techniques ne fonctionnent pas pour tous les types d'images. D'autre part, elles peuvent nécessiter des temps de calcul importants et on aimerait alors pouvoir s'en dispenser.

En l'absence d'une segmentation réaliste, il nous semble prudent de choisir des «petits» intervalles, ce qui revient à sur-segmenter l'image explicative. L'avantage de cette approche est de diminuer le risque de regrouper des voxels appartenant en réalité à des tissus différents. En pratique, nous utilisons souvent une partition uniforme des intensités en 256 intervalles et il apparaît que ce choix est efficace pour de nombreux problèmes réels de recalage. Sa contrepartie est un relatif manque de régularité de la mesure vis-à-vis des paramètres de transformation spatiale.

Nous avons constaté que ce problème se manifestait de façon critique pour le recalage d'images ultrasonores et IRM. Dans ce cas, nous choisissons toujours l'IRM comme image explicative. A cause du niveau de bruit généralement élevé dans l'image ultrasonore, l'estimation des moyennes correspondant à chaque intervalle de l'IRM peut s'avérer statistiquement inefficace. Cet effet est clairement lié au fait que la partition de l'IRM est trop fine. On conçoit alors la nécessité d'imposer une contrainte de continuité à la fonction d'intensité. Une façon simple de le faire est de la supposer polynômiale :

$$f(j) = \sum_{p=1}^{P} c_p \, j^{p-1}$$

Les polynômes ne sont certainement pas des fonctions physiquement très réalistes. Le principal avantage que nous leur trouvons est de pouvoir prendre des formes relativement complexes tout en reposant sur un petit nombre de degrés de libertés. Il y aurait peut-être un bénéfice important à étudier d'autres classes de fonctions, et nous pensons particulièrement à des fonctions B-splines non-uniformément réparties (Foley et Nielson, 1989).

3.4.2 Métrique robuste

Le critère $S(T,f)$ est une mesure quadratique de l'écart d'intensité entre l'image I et l'image J corrigée. Dans le cas d'une interpolation de l'intensité :

$$S(T,f) = \sum_k \left[i_k - f(j_k^{\downarrow}) \right]^2$$

Il est bien connu que la mesure quadratique est principalement influencée par les «grands» résidus. Étant donné un couple (T,f), il suffit qu'un voxel $\mathbf{x}_k$ soit responsable d'un résidu 10 fois plus élevé que tous les autres pour que sa contribution relative soit 100 fois supérieure. C'est une bonne nouvelle si tous les grands résidus sont dûs à des vrais négatifs, c'est-à-dire à des erreurs de recalage ou de correction d'intensité. En pratique, on rencontre malheureusement des faux négatifs, à savoir des points pour lesquels l'écart d'intensité est grand alors que les images sont parfaitement recalées et que la correction d'intensité est par ailleurs idéale.

Ces faux négatifs peuvent avoir plusieurs sources : occultations, distorsions géométriques, effets de volume partiel et, plus généralement, le fait que l'hypothèse de dépendance fonctionnelle est une approximation. Ils risquent de fausser la mesure de similarité et, par là-même, de biaiser le recalage. Il paraît souhaitable de limiter leur influence.

Idéalement, nous voudrions être capables de séparer les faux négatifs des vrais négatifs, mais il n'y a aucun moyen de le faire sans connaissance a priori : en effet, rien ne permet d'affirmer que les faux négatifs se caractérisent par des écarts d'intensité différents de *tous* les vrais négatifs. En revanche, il est raisonnable de supposer que les vrais négatifs correspondent à des écarts d'intensité plus variés, y compris des écarts modérés. Ainsi, en réduisant l'influence des écarts les plus forts, nous pouvons espérer ignorer les faux négatifs tout en conservant suffisamment de vrais négatifs pour que la mesure soit sensible aux erreurs de recalage, ce qui est évidemment indispensable.

Cette problématique rejoint celle de l'estimation robuste (Rousseeuw et Leroy, 1987; Meer et al., 2000). On peut interpréter le critère $S(T,f)$ comme une moyenne quadratique empirique (modulo une division par $n^{\downarrow}$) ; autrement dit, dans la terminologie des statisticiens, un estimateur d'échelle connu pour avoir un point de rupture à 0% : une seule donnée peut modifier arbitrairement sa valeur. Il devient alors naturel de le remplacer par un estimateur d'échelle robuste. Une idée similaire a été exploitée par (Nikou et al., 1998; Nikou et al., 1999; Nikou, 1999) dans le but d'améliorer le critère de Woods.

3.4.2.1 S-estimateurs

Étant donné un échantillon $\{r_k\}$, on peut définir une métrique robuste $\sigma(r_1,r_2,\ldots,r_n)$ par une équation implicite du type :

$$\frac{1}{n}\sum_k \rho\left(\frac{r_k}{\sigma}\right) = K \tag{3.14}$$

où ρ est la fonction de coût associée à un M-estimateur et K est une constante de normalisation. Une telle métrique est connue sous le nom de M-estimateur d'échelle (Rousseeuw et Leroy, 1987). La constante K est généralement choisie de manière à ce que σ soit un estimateur sans biais de l'écart-type lorsque les données suivent une loi normale centrée, ce qui se traduit par la contrainte :

$$K = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \rho(x) e^{-\frac{x^2}{2}} dx$$

En général, la métrique σ n'admet pas d'expression analytique (voir le tableau 3.1). On remarque néanmoins les cas particuliers suivants :

- Si $\rho(x) = x^2$ et $K = 1$, alors σ est la moyenne quadratique :

$$\sigma(r_1,r_2,\ldots,r_n) = \sqrt{\frac{1}{n}\sum_k r_k^2}$$

- Si $\rho(x) = |x|$, alors $K = (2\pi)^{-\frac{1}{2}}$ et σ est proportionnel à la moyenne des valeurs absolues :

$$\sigma(r_1,r_2,\ldots,r_n) = \frac{\sqrt{\pi}}{2n} \sum_k |r_k| \tag{3.15}$$

- Si $\rho(x) = 1 + \mathrm{sgn}(|x| - c)$ avec $c \approx 0.684$, alors $K = 1$ et σ est proportionnel à la médiane des valeurs absolues :

$$\sigma(r_1,r_2,\ldots,r_n) = \frac{1}{c}\,\mathrm{med}\,\{|r_k|\} = 1.4628\,\mathrm{med}\,\{|r_k|\} \tag{3.16}$$

Dans le contexte d'un problème d'estimation, les données r_k dépendent d'un ensemble de paramètres θ et ils sont appelés résidus. La valeur de θ réalisant le minimum de σ est appelée S-estimateur. Dans notre cas particulier, $\theta = (T,f)$ et, si on choisit une interpolation de l'intensité, $r_k = i_k - f(j_k^{\downarrow})$. L'interpolation VP est traitée en annexe B.2.

Le choix de la fonction ρ conditionne les propriétés de robustesse et d'efficacité du S-estimateur, deux notions pratiquement antinomiques. La robustesse d'un estimateur caractérise sa capacité à ne pas être influencé par les données aberrantes. L'efficacité a davantage à voir avec la notion de précision : elle traduit l'aptitude de l'estimateur à prendre en compte un maxium de données pertinentes. Par exemple, le S-estimateur correspondant à l'équation (3.16) est très robuste mais peu efficace (Rousseeuw et Leroy, 1987). A l'inverse, le S-estimateur défini par l'équation (3.15), dit estimateur L_1, ne possède pas des propriétés optimales de robustesse mais il est beaucoup plus efficace[5].

Il y a deux raisons pour lesquelles les S-estimateurs ne sont pas directement applicables à notre problème de recalage :

1. La métrique $\sigma(T,f)$ étant définie par une équation implicite, son calcul est généralement très lourd en pratique car il fait intervenir une suite d'approximations numériques. Cette remarque ne concerne pas les cas particuliers précédemment évoqués, notamment les métriques quadratique et L_1.
2. Le fait que $\sigma(T,f)$ soit robuste aux résidus aberrants ne garantit en rien qu'il soit robuste au recouvrement partiel des images. Le critère doit donc être normalisé de la même manière que précédemment (voir le paragraphe 3.2.2).

3.4.2.2 S-estimateur instantané

Dans le cas général, (Rousseeuw et Leroy, 1987) proposent de résoudre numériquement l'équation (3.14) en construisant la suite d'approximations :

$$\sigma_{i+1} = \sigma_i \sqrt{\frac{1}{Kn} \sum_k \rho\left(\frac{r_k}{\sigma_i}\right)}$$

5. D'ailleurs, la définition des S-estimateurs donnée par (Rousseeuw et Leroy, 1987) est plus restrictive que celle que nous considérons ici : la fonction ρ doit vérifier certaines conditions particulières, ceci afin de garantir une robustesse optimale (point de rupture à 50%). En ce sens, les estimateurs quadratique et L_1 ne sont pas de véritables S-estimateurs.

Métrique	$\rho(x)$	$\sigma(r_1,\dots,r_n)$
L_2	x^2	$\frac{1}{n}\sum_k r_k^2$
L_1	$\|x\|$	$\frac{\sqrt{2\pi}}{n}\sum_k \|r_k\|$
MDA	$1+\mathrm{sgn}(\|x\|-c)$	$\frac{1}{c}\mathrm{med}\|r_k\|$
Huber	$\min\left(x^2, 2c\|x\|-c^2\right)$	implicite
Cauchy	$c^2\log\left[1+(x/c)^2\right]$	implicite
Geman-McClure	$\frac{x^2}{1+(x/c)^2}$	implicite
Tukey	$\frac{c^2}{3}\min\left(1-\left[1-(x/c)^2\right]^3, 1\right)$	implicite

TAB. 3.1 – *Quelques fonctions ρ usuelles et les M-estimateurs d'échelle qui leur sont associés.*

On remarque que les fonctions ρ de la forme $\rho(x) = |x|^p$ permettent de converger en une seule itération quelle que soit la valeur initiale de la suite. Cependant, plusieurs itérations sont généralement nécessaires, ce qui peut se traduire par une complexité calculatoire prohibitive. Pour s'affranchir de cette contrainte, (Rousseeuw et Leroy, 1987) se limitent en toute circonstance à une seule itération. L'estimateur qui en résulte est appelé S-estimateur instantané[6]. Dans notre contexte, un S-estimateur instantané correspond à la fonction de coût suivante :

$$\sigma_1^2(T,f) = \frac{\sigma_0^2}{Kn^{\downarrow}} \sum_k \rho\left(\frac{i_k - f(j_k^{\downarrow})}{\sigma_0}\right) \tag{3.17}$$

où σ_0 est la valeur initiale de la suite d'approximations. Lorsque σ_0 est considérée comme une constante, le S-estimateur instantané est, par définition, un M-estimateur. A T fixé, la minimisation de $\sigma_1^2(T,f)$ en fonction de f peut alors être résolue par une procédure de moindres carrés itérés à condition que l'espace de recherche $\mathcal{F}$ soit un espace vectoriel (voir annexe B.2).

Ce qui distingue fondamentalement un S-estimateur instantané d'un M-estimateur est le fait que σ_0 dépend en principe des paramètres à estimer. Idéalement, σ_0 devrait être aussi proche que possible de la limite de la suite $\sigma(T,f)$ et devrait donc dépendre de (T,f). Un choix d'initalisation classique est la médiane des déviations absolues (MDA), dont nous avons d'ailleurs mentionné plus haut qu'elle correspond à un S-estimateur particulier :

$$\sigma_0(T,f) = 1.4826 \,\mathrm{med}\Big\{ \big| i_k - f(j_k^{\downarrow}) \big| \Big\} \tag{3.18}$$

Cette approche a deux inconvénients. D'une part, il faut recalculer σ_0 pour chaque nouveau couple (T,f) . D'autre part, la minimisation de $\sigma_1^2(T,f)$ selon f ne consiste plus en une simple procédure de moindres carrés itérés. C'est pourquoi en pratique, nous considérons σ_0 comme «périodiquement» constante : sa valeur est fixée pendant un certain nombre d'itérations de l'algorithme de minimisation, après quoi elle est réinitialisée au moyen de l'équation (3.18). Nous reviendrons plus en détails sur cet aspect au paragraphe 3.5.

On voit que cette stratégie de calcul revient à moduler progressivement la distance de coupure (paramètre c) de la fonction ρ. Il s'avère que c'est une astuce pratique fréquemment employée lors de la mise en œuvre d'un M-estimateur (Nikou et al., 1999). En fait, la similitude entre les S-estimateurs et les M-estimateurs est encore plus profonde : (Rousseeuw et Leroy, 1987) montrent que tout S-estimateur est un M-estimateur pour lequel la distance de coupure de la fonction ρ admet une certaine valeur, déterminée par les données et donc inconnue *a priori*. De ce point de vue, les S-estimateurs sont des «M-estimateurs intelligents» capables de régler eux-mêmes leur distance de coupure.

6. En anglais, *one-step S-estimator*.

3.4.2.3 Normalisation

Nous définissons la mesure normalisée correspondant à une métrique robuste de la même manière que dans le cas d'une métrique quadratique :

$$\tilde{S}(T) = \frac{\min\limits_{f} \sigma_1^2(T,f)}{\min\limits_{f=\text{cte}} \sigma_1^2(T,f)} \tag{3.19}$$

Il est clair que cette mesure vérifie les axiomes de normalisation $(A1)$, $(A2)$, $(A3)$ et $(A4)$ énoncés au paragraphe 3.2.2. Par la suite, nous appelerons *rapport de corrélation généralisé* toute mesure de la forme $1 - \tilde{S}(T)$, où $\tilde{S}(T)$ est du type (3.19). Un rapport de corrélation généralisé est défini par l'espace de recherche de la fonction d'intensité et par la métrique robuste qui lui sont associé. On vérifie aisément qu'il prend toujours ses valeurs entre 0 et 1.

La plupart des fonctions ρ usuelles (Huber, Geman-McClure, Cauchy, Tukey, ...) sont quadratiques au voisinage de zéro. En faisant tendre leur distance de coupure vers l'infini, on retrouve le rapport de corrélation «quadratique» comme un cas limite du rapport de corrélation généralisé.

3.4.2.4 Cas particulier : le rapport de corrélation L_1

Un exemple de fonction ρ pour laquelle cette propriété n'est pas vraie est $\rho(x) = |x|$ (métrique L_1). Ce cas partage néanmoins avec la métrique quadratique deux propriétés particulièrement intéressantes d'un point de vue calculatoire. D'une part, le paramètre σ_0 ne joue aucun rôle et il n'est donc pas nécessaire de l'estimer. D'autre part, lorsque f est recherchée comme une fonction constante par morceaux, la minimisation de $\sigma_1(T,f)$ en fonction de f admet une solution explicite.

L'expression analytique de cette fonction a été établie au paragraphe 3.3.1 dans le cas quadratique. Pour la transposer au cas de la métrique L_1, nous avons besoin du résultat suivant :

Théorème 3.1 *Soit un échantillon, c'est-à-dire un ensemble de réels $\{x_1,x_2,\dots,x_n\}$. Le réel $\hat{\mu}$ qui minimise $\sum_k |x_k - \mu|$ est la médiane de l'échantillon.*

Preuve. Voir par exemple (Papoulis, 1991), page 178. □

Le raisonnement que nous avons mené dans le cas quadratique est directement généralisable. En reprenant les notations du paragraphe 3.3.1, nous voyons que la recherche des valeurs prises par f sur les différents intervalles $[q - 1/2, q + 1/2[$ conduit, ici encore, à une série de minimisations indépendantes :

$$\forall q \in \mathbb{Z}, \quad \Omega_I^{\downarrow q} \neq \emptyset \;\Rightarrow\; \hat{\mu}_q = \arg\min_{\mu} \sum_{k \in \Omega_I^{\downarrow q}} |i_k - \mu_q|$$

Cette expression correspond au cas d'une interpolation de l'intensité. Par application du théorème 3.1, les valeurs recherchées sont :

$$\hat{\mu}_q = \text{med}\left\{ i_k,\, k \in \Omega_I^{\downarrow q} \right\} \tag{3.20}$$

De la même manière, la fonction constante qui minimise $\sigma_1(T,f)$ est la médiane des intensités de I calculée sur l'ensemble complet des voxels appariés, $\Omega_I^\downarrow$. Dans le cas d'une interpolation VP, ce dernier résultat reste vrai tandis que l'adaptation de l'équation (3.20) fait intervenir la notion de médiane pondérée (Brownrigg, 1984). Pour chaque ensemble Ω_J^q défini par

$$\Omega_J^q = \left\{ \mathbf{y}_\ell \in \Omega_J, \quad q - \frac{1}{2} \leq \frac{j_\ell}{\beta} < q + \frac{1}{2} \right\},$$

il faut calculer la fonction de répartition empirique F_q des intensités i_k correspondant aux voxels $\mathbf{x}_k$ appariés avec Ω_J^q. Mais, comme ici les voxels sont appariés de manière «floue», ils contribuent à la fonction de répartition proportionellement à leur poids $w_{k\ell}^\downarrow$. Les valeurs optimales $\hat{\mu}_q$ sont alors les médianes des différentes fonctions de répartition, c'est-à-dire les valeurs $\hat{\mu}_q$ vérifiant $F_q(\hat{\mu}_q) = 0.5$.

3.5 Schéma d'optimisation

3.5.1 Optimisation hiérarchisée et optimisation alternée

Revenons à la définition générale donnée par l'équation (3.19). Maximiser un rapport de corrélation généralisé équivaut à résoudre le problème suivant :

$$\min_T \left(\min_f \tilde{S}(T,f) \right) \qquad \text{avec} \quad \tilde{S}(T,f) \equiv \frac{\sigma_1^2(T,f)}{\min\limits_{f=\text{cte}} \sigma_1^2(T,f)}$$

Cette formulation correspond à ce que nous appelons une minimisation hiérarchisée : la recherche est menée dans le sous-espace de $\mathcal{T} \times \mathcal{F}$ défini par les couples (T,f) pour lesquels f réalise le minimum de $\tilde{S}(T,f)$ à T fixé. Le théorème suivant montre que la minimisation hiérarchisée est simplement une stratégie particulière de minimisation sur l'espace sans restriction.

Théorème 3.2 *Soient A et B deux ensembles et soit une fonction $h : A \times B \to \mathbb{R}$. Si h admet un minimum sur $A \times B$ et si, pour tout $x \in A$, la fonction $y \mapsto h(x,y)$ admet un minimum sur B, alors :*

$$\min_x \left(\min_y h(x,y) \right) = \min_{x,y} h(x,y)$$

Preuve. La relation suivante est toujours vraie :

$$\inf_x \left(\inf_y h(x,y) \right) = \inf_{x,y} h(x,y)$$

Par hypothèse, la fonction $H(x) = \min_y h(x,y)$ est définie et à valeurs dans $\mathbb{R}$. De même, il existe un couple (x_0,y_0) tel que $h(x_0,y_0) = \inf\limits_{x,y} h(x,y)$. On en déduit $h(x_0,y_0) = \inf\limits_x H(x)$. Or on a nécéssairement $\min\limits_y h(x_0,y) = h(x_0,y_0)$ et donc $H(x_0) = h(x_0,y_0)$. Par conséquent, $h(x_0,y_0)$ est un minimum de la fonction $x \mapsto H(x)$ atteint en x_0. □

En termes de mise en œuvre, la minimisation hiérarchisée impose de recalculer f à chaque fois qu'une nouvelle transformation T est considérée, donc de résoudre autant de problèmes de régression linéaire. Dans le cas du rapport de corrélation «standard», f est recherchée comme une fonction constante par morceaux et la métrique de comparaison est la norme quadratique. Le coût de de la régression linéaire est alors négligeable et la minimisation hiérarchisée peut être envisagée. Ce n'est pas le cas en général. Pour d'autres classes de fonctions f, par exemple des fonctions polynômiales (voir paragraphe 3.4.1), le système d'équations associée à la régression linéaire fait intervenir une matrice pleine dont le calcul est nettement plus fastidieux. Lorsque l'on fait appel à une métrique robuste, la résolution de tels systèmes linéaires doit de surcroix être itérée (typiquement, 5-6 fois).

On optera alors de préférence pour une minimisation alternée de $\tilde{S}(T,f)$, c'est-à-dire une procédure dont le schéma est le suivant :

1. A T fixé, minimiser $\tilde{S}(T,f)$ en fonction de f. Il s'agit d'un problème de régression linéaire dont la résolution nécessite «au pire» une procédure de moindres carrés itérés.
2. A f fixé, minimiser $\tilde{S}(T,f)$ en fonction de T. Le critère revient alors à une mesure de différence d'images normalisée. Comme il n'existe pas de solution analytique, la recherche doit être menée par un algorithme numérique du type de ceux qui ont été présentés au chapitre 2, paragraphe 2.5.
3. Retourner à l'étape 1 si T ou f ont évolué de manière significative.

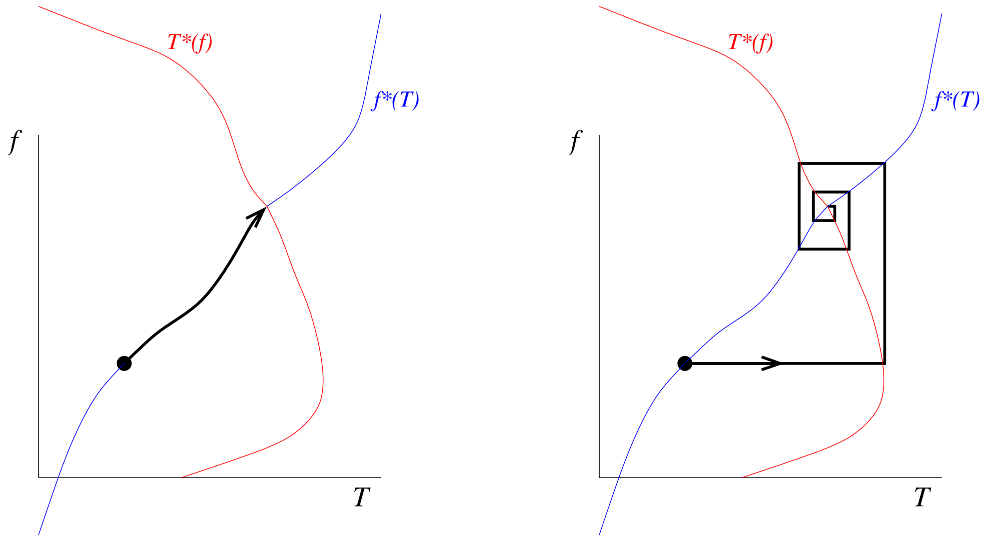
Le principe de la minimisation alternée est bien connu. On le retrouve notamment dans l'algorithme de Powell et dans l'algorithme ICP, tous deux évoqués au chapitre 2. La différence avec la minimisation hiérarchisée est illustrée par la figure 3.9. Les deux stratégies ne conduisent pas nécessairement à la même solution si, comme c'est fréquemment le cas, un algorithme d'optimisation locale est utilisé pour estimer T. La figure 3.9 représente la situation (optimiste) où $\tilde{S}(T,f)$ admet au plus deux minima locaux à f fixé. Selon les valeurs initiales de T et f, la minimisation hiérarchisée et la minimisation alternée peuvent donner des résultats différents.

3.5.2 Mise en œuvre

Nous avons implémenté deux algorithmes de recalage rigide/affine, l'un correspondant à une optimisation hiérarchisée et l'autre à une optimisation alternée. Le premier algorithme est similaire à celui de (Maes et al., 1997). Il utilise la méthode de Powell (Press et al., 1992) comme procédure d'optimisation. Plusieurs alternatives peuvent être envisagées (voir le chapitre 2.5). Nos premières expériences utilisant des descentes de gradient se sont révélés décevantes, vraisemblablement à cause d'un mauvais conditionnement des paramètres de transformation. Cependant, les résultats rapportés notamment par (Maes et al., 1999) pour l'information mutuelle laissent à penser que les méthodes de gradient ne sont pas nécessairement moins performantes que la méthode de Powell.

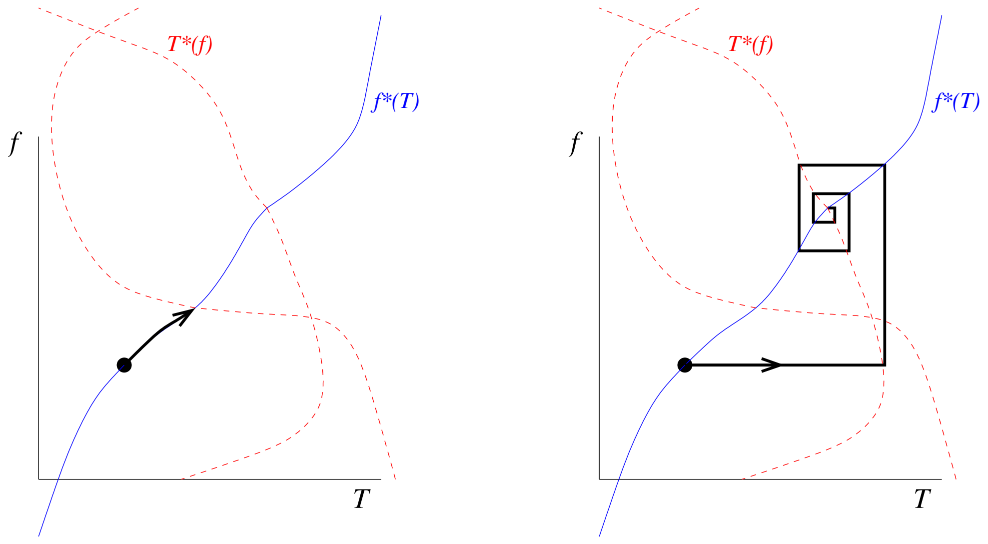
L'algorithme correspondant à une optimisation alternée est schématisé par la la figure 3.11. Un algorithme similaire sera présenté au chapitre 7 pour le recalage non-rigide. Ici aussi, nous utilisons la méthode de Powell pour optimiser les paramètres de transformation spatiale.

Chacune des deux stratégies d'optimisation a ses inconvénients. Celui de l'optimisation hiérarchisée est d'effectuer un grand nombre d'appels à la procédure d'estimation de f (régression linéaire). L'optimisation alternée permet de limiter considérablement le nombre de ces appels mais, en contrepartie, elle nécessite plusieurs appels à la méthode de Powell. En pratique, l'optimisation



(a) minimisation hiérarchisée (b) minimisation alternée

FIG. 3.9 – *Illustration des deux stratégies de minimisation de $\tilde{S}(T,f)$. La courbe bleue représente les positions des minima à T fixé. Inversement, la courbe rouge correpspond aux minima à f fixé.*



(a) minimisation hiérarchisée (b) minimisation alternée

FIG. 3.10 – *En présence de plusieurs minima locaux, les deux méthodes ne convergent pas nécessairement vers la même solution. Ici le minimum global peut correspondre à l'une ou l'autre des solutions.*

hiérarchisée est plus intéressante d'un point de vue calculatoire dès que les conditions suivantes sont remplies :

1. L'espace de recherche $\mathcal{F}$ est celui des fonctions constantes par morceaux.
2. La métrique de comparaison d'intensités est la métrique L_2 ou L_1, c'est-à-dire : $\rho(x) = |x|^p$ avec $p \in \{1,2\}$.

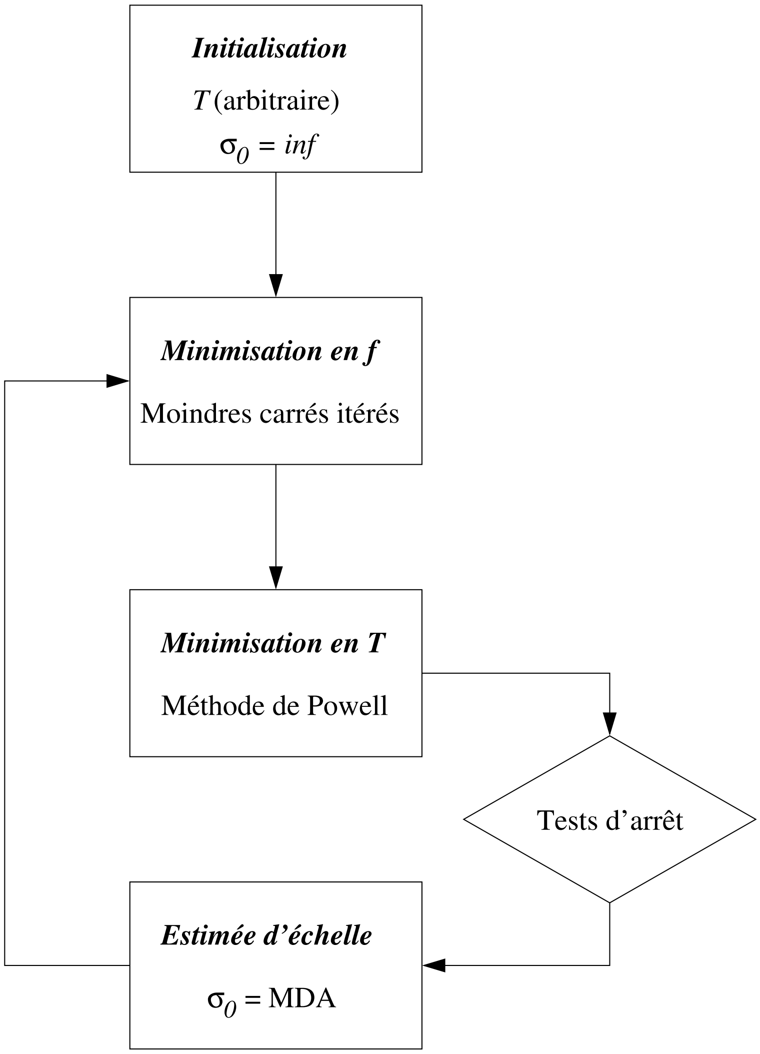


FIG. 3.11 – *Organigramme de l'algorithme de recalage rigide/affine correspondant à une optimisation alternée du rapport de corrélation généralisé.*

Comme nous l'avons déjà signalé, il existe alors une solution explicite au problème d'estimation de f et le paramètre d'échelle σ_0 est superflu. En dehors de ces cas particuliers, l'algorithme d'optimisation alternée est de loin le plus rapide.

En outre, il autorise une stratégie d'estimation adaptative du paramètre σ_0 (voir le paragraphe 3.4.2.2). Nous fixons sa valeur initiale à $\sigma_0 = \infty$, ce qui signifie que la première passe de l'algorithme revient à utiliser une métrique quadratique. Dans notre implémentation actuelle,

σ_0 est réinitialisée à chaque itération globale de la minimisation alternée, c'est-à-dire après une minimisation selon T suivie d'une minimisation selon f. En règle générale, σ_0 diminue au cours de la procédure et, ainsi, la métrique est d'autant moins tolérante vis-à-vis des grands écarts d'intensité. On peut légitimement penser que le fait d'autoriser σ_0 à varier hypothèque la convergence de l'algorithme. Nous ne voyons en tous cas pas d'argument pour la démontrer. Nous n'avons toutefois jamais rencontré de cas de divergence ou de comportement oscillatoire.

Une autre adaptation pratique concerne le dénominateur de l'équation (3.19). Ce terme provient de la normalisation et il représente une estimée d'échelle globale de l'image I (restreinte aux voxels appariés à l'intérieur du domaine cible, d'où la dépendance en T). Dans le cas des fonctions $\rho(x) = x^2$ et $\rho(x) = |x|$, il correspond respectivement à la variance classique $\mathrm{Var}(I^{\downarrow})$ et à la moyenne des écarts absolus par rapport à la médiane. En général, il n'existe pas de telle forme analytique et le calcul doit être mené par une procédure de moindres carrés itérés analogue à l'estimation de f. Mais comme cette procédure doit être réappliquée à chaque fois que T est mise à jour, il en découle un accroissement important du temps de calcul, injustifié dans la mesure où le terme de normalisation est présumé à variation lente. C'est pourquoi, nous le remplaçons par la variance quadratique $\mathrm{Var}(I^{\downarrow})$ dont le calcul est beaucoup plus rapide (excepté le cas de la métrique L_1).

3.6 Conclusion

Nous avons formulé le recalage de deux images comme un problème d'estimation aux moindres carrés et, plus généralement, au sens d'une certaine métrique de comparaison. Dans cette formulation, l'une des images est simultanément déplacée et corrigée en intensité pour «ressembler» autant que possible à l'autre. Par ailleurs, le critère d'estimation est normalisé de manière à tenir compte des effets de recouvrement partiel.

Lorsque la métrique est la norme euclidienne et que la fonction de correction est recherchée constante par morceaux, deux choix avantageux en termes d'efficacité calculatoire, le critère de recalage s'identifie au concept probabiliste de rapport de corrélation. Le rapport de corrélation est un critère asymétrique de liaison fonctionnelle connu pour généraliser le coefficient de corrélation, critère symétrique de liaison affine par ailleurs classiquement utilisé comme mesure de similarité pour le recalage monomodal (Brown, 1992). En jouant sur la métrique de comparaison et sur l'espace de recherche de la fonction de correction, nous avons défini une classe de mesures que, par analogie, nous regroupons sous la rubrique *rapport de corrélation généralisé*.

Ce travail s'inscrit dans la lignée de plusieurs auteurs qui ont formulé le recalage d'images comme un problème analogue d'estimation aux moindres carrés (Barber, 1992; Feldmar, 1994; Viola, 1995; Friston et al., 1995) ou qui ont proposé des mesures de similarité étroitement liées au rapport de corrélation (Woods et al., 1993; Ardekani et al., 1995; Alpert et al., 1996; Nikou et al., 1999). L'une des principales spécificités de notre approche tient à la prise en compte des effets de recouvrement partiel dans la construction de la mesure de similarité. Pour faire un parallèle historique, le rapport de corrélation est un peu aux critères mentionnés ci-dessus ce que l'information mutuelle est à l'entropie conjointe (Collignon et al., 1995b; Studholme et al., 1995).

Nous avons généralisé le rapport de corrélation à des métriques robustes en adaptant l'approche des S-estimateurs introduite par (Rousseeuw et Leroy, 1987) dans le contexte de la régression linéaire. A notre connaissance, il s'agit d'une contribution originale. Cependant, l'idée d'utiliser une métrique robuste n'est pas nouvelle en soi et la mise en œuvre pratique des S-estimateurs équivaut sensiblement à celle des M-estimateurs, déjà utilisés par (Nikou, 1999) pour généraliser le critère de Woods.

Enfin, nous avons introduit deux schémas d'optimisation associés au rapport de corrélation généralisé, l'un et l'autre pouvant être avantageux en temps de calcul selon les choix effectués pour la métrique de comparaison et l'espace de recherche de la fonction de correction. Le schéma alterné consiste à mettre à jour alternativement les paramètres de la transformation spatiale et les paramètres de la fonction de correction. Le schéma hiérarchisé impose à la fonction de correction de minimiser continuellement le critère d'estimation, ce qui conduit à optimiser une mesure de similarité au sens strict, c'est-à-dire un critère dépendant uniquement des paramètres de transformations spatiales. Remarquons que la notion de rapport de corrélation (généralisé) apparaît dans le contexte d'un schéma hiérarchisé. Dans la suite de ce mémoire, nous nous autorisons néanmoins à parler de l'optimisation alternée du rapport de corrélation.

Un rapport de corrélation est potentiellement adapté au recalage multimodal dans la mesure où il se fonde sur une hypothèse de dépendance fonctionnelle non nécessairement linéaire ou monotone entre les intensités des images. Bien qu'une telle hypothèse soit parfois simpliste, elle peut être assouplie par l'utilisation d'une métrique robuste. Contrairement à l'information mutuelle, qui repose sur des hypothèses de dépendance plus générales, le rapport de corrélation considère les intensités comme des variables quantitatives.

La question est maintenant de savoir si cette classe de mesures est d'une utilité réelle dans le cadre du recalage multimodal d'images médicales. Le chapitre suivant y répond par l'affirmative.

Chapitre 4

Comparaison expérimentale de mesures

4.1 La base d'images «Vanderbilt»

En 1995, l'université Vanderbilt (Nashville, États-Unis) a créé une base de données d'images cérébrales destinée à évaluer la précision de différentes techniques de recalage multimodal. Cette base a déjà permis l'étude comparative d'une vingtaine de techniques développées par différents groupes à travers le monde (West et al., 1997). La base originale contient un ensemble de modalités acquises pour dix patients différents [1] :

- Des images scanner (pour 8 patients sur 10) : elles consistent en une trentaine de coupes de dimension 512^2. La taille de voxel est $0.65 \times 0.65 \times 4$ mm^3.
- Des images TEP (pour 8 patients sur 10) : une quinzaine de coupes de dimension 128^2 ($2.59 \times 2.59 \times 8$ mm^3).
- Des images IRM correspondant aux protocoles d'acquisition T1, T2 et DP (densité de proton). Les images contiennent 20 à 26 coupes de dimension 256^2, avec une taille de voxel de $1.25 \times 1.25 \times 4$ mm^3.

Avant chaque série d'acquistions, des marqueurs ont été implantés sur les patients et un cadre stéréotaxique leur a été fixé. Les marqueurs ont servi à estimer les transformations rigides reliant les différentes modalités ; ce sont les transformations standard qui tiennent lieu de «vérité terrain». Aucune indication n'est fournie quant à leur précision mais on peut espérer que celle-ci soit nettement inférieure à la résolution spatiale des images.

Par la suite, le cadre et les marqueurs ont été effacés des images au moyen d'un traitement semi-automatique. On trouvera plus de détails sur les différentes étapes de la construction de la base en consultant l'article de référence (West et al., 1997) ainsi que d'autres documents disponibles sur le site web du projet : `http://www.vuse.vanderbilt.edu/~jayw/results.html`.

1. Nous incluons le jeu de données d'essai qui n'était pas pris en compte dans les expériences relatées par (West et al., 1997). Par ailleurs, de nouvelles images ont été rajoutées depuis mais nous ne les avons pas utilisées dans nos expériences.

4.2 Description des expériences

Nous nous sommes intéressés au recalage des images scanner et TEP avec chacun des 3 protocoles d'acquisition d'IRM (T1, T2, DP), ce qui nous amène donc à étudier 6 combinaisons de modalités différentes. Pour chacune d'entre elles, nous disposons de 8 couples d'images étant donné que deux patients n'ont pas subi de scanner et deux autres n'ont pas subi de TEP.

Notre objectif dans cette étude est, d'une part, de quantifier la précision de différentes mesures de similarité fondées sur le rapport de corrélation et, d'autre part, de la comparer avec celle de mesures de similarité classiques. Comme mesures de comparaison, nous avons choisi l'information mutuelle et le critère de Woods, qui nous semblent être les plus représentatives de l'existant.

4.2.1 Mesures de similarité

4.2.1.1 Information mutuelle

Nous avons implémenté l'information mutuelle de la façon préconisée par (Collignon et al., 1995a; Maes et al., 1997), c'est-à-dire en calculant l'histogramme conjoint par simple regroupement des couples d'intensité en 256^2 classes. Ainsi que nous l'avons vu au chapitre 2, paragraphe 2.3.1, cette méthode de construction de l'histogramme conjoint (*binning*) est un cas particulier de la méthode des fenêtres de Parzen discrète pour lequel le noyau est une fonction en créneau. Ici, la taille de fenêtre (unitaire dans les deux directions) est ajustée de manière à respecter la dynamique des images, ce qui correspond en quelque sorte à une règle de prudence. Nous n'avons pas testé d'autres valeurs dans ces expériences. L'utilisation de fenêtres plus grandes permet généralement de réduire la tendance de l'information mutuelle à produire des maxima locaux. Toutefois, le fait d'augmenter la taille de fenêtre n'est pas nécessairement bénéfique en termes de *précision* du recalage (transformation réalisant le maximum global).

Par ailleurs, comme l'a montré l'expérience de la «barre grise» présentée au chapitre 3, paragraphe 3.1, le choix d'un noyau continu (par exemple, gaussien) agit également dans le sens d'une diminution des maxima locaux. Il en résulte cependant une complexité calculatoire considérablement accrue et, pour une procédure d'optimisation donnée, ce type d'implémentation de l'information mutuelle n'est plus comparable en temps de calcul avec les mesures fondées sur le rapport de corrélation [2].

4.2.1.2 Critère de Woods

De manière analogue à l'information mutuelle, le critère de Woods est calculé via un partionnement de l'image cible en 256 classes d'intensité, tout comme dans sa version originale (Woods et al., 1993). On sait que le critère de Woods est une mesure assymétrique fondée sur l'hypothèse

2. Remarquons que puisque le rapport de corrélation (généralisé) peut être appréhendé dans le formalisme d'un histogramme conjoint, on pourrait également l'implémenter en utilisant des fenêtres de Parzen. Mais la philosophie de la méthode consiste justement à s'épargner un tel calcul (voir le chapitre 3).

d'une dépendance fonctionnnelle de l'image à partitionner *vers* l'autre image (voir le chapitre 3, paragraphe 3.3.4).

Dans cette étude, toutes les expériences de recalage impliquent, soit des couples scanner/IRM, soit des couples TEP/IRM. C'est à chaque fois l'image IRM que nous avons choisie comme l'image *explicative*, c'est-à-dire l'image à partitionner. Ce choix est naturellement dicté par l'observation empirique que les images IRM, quel que soit leur protocole d'acquisition, sont celles qui présentent le plus grand nombre de structures anatomiques différenciées.

4.2.1.3 Rapports de corrélation généralisés

Nous avons testé 6 versions différentes du rapport de corrélation généralisé. Elles sont résumées par le tableau 4.1. A l'instar du critère de Woods, ces mesures caractérisent une dépendance fonctionnelle et font jouer des rôles différents aux images à recaler. Pour les mêmes raisons que celles données plus haut, c'est systématiquement à l'image IRM que nous avons attribué le rôle d'image explicative.

Tel que nous l'avons défini au chapitre 3, un rapport de corrélation généralisé est caractérisé par la donnée d'une métrique de comparaison d'intensités et d'un espace de recherche de la fonction d'intensité. Sa mise en œuvre fait intervenir un troisième choix : le schéma d'optimisation (alterné ou hiérarchisé).

La métrique de comparaison d'intensités (voir chapitre 3, paragraphe 3.4.2). En toute généralité, c'est un M-estimateur d'échelle. Nous avons considéré ici trois cas particuliers : la métrique quadratique (RC0, RC1, RC2), la métrique L_1 (RC3) et la métrique fondée sur la fonction ρ de Geman-McClure (RC4, RC5) :

$$\rho(x) = \frac{x^2}{1 + \frac{x^2}{c^2}}$$

Pour cette dernière, la distance de coupure a été fixée à $c = 3.648$ dans toutes les expériences, ce qui correspond à une efficacité gaussienne de 95% (la constante de normalisation vaut alors $K = 0.832$). Nous comparons ainsi trois métriques ayant des propriétés de robustesse théoriquement différentes : une métrique non-robuste (quadratique), une métrique «faiblement» robuste (L_1) et une métrique «moyennement» robuste (Geman-McClure). Ces propriétés de robustesse sont liées au comportement asymptotique de la fonction ρ (Rousseeuw et Leroy, 1987).

L'espace de recherche de la fonction d'intensité (voir chapitre 3, paragraphe 3.4.1). Nous avons envisagé deux cas : celui des fonctions constantes par morceaux (RC0, RC1, RC3, RC4) et celui des fonctions polynômiales (RC2, RC5). Pour les premières, les intensités de l'image explicative (dans tous les cas, l'IRM) sont partitionnées en 256 intervalles uniformes. Dans la mesure où les images dont nous disposons sont encodées sur 1 octet, cela revient à corriger indépendamment les ensembles d'iso-intensité de l'IRM. Pour les secondes, nous avons fixé un degré maximal de 5 dans toutes les expériences.

Sigle	Métrique	Fonction d'intensité	Optimisation
RC0	quadratique	constante par morceaux	hiérarchisée
RC1	quadratique	constante par morceaux	alternée
RC2	quadratique	polynômiale	alternée
RC3	L_1	constante par morceaux	hiérarchisée
RC4	Geman-McClure	constante par morceaux	alternée
RC5	Geman-McClure	polynômiale	alternée

TAB. 4.1 – *Les différentes version du rapport de corrélation utilisées dans les expériences de recalage de la base « Vanderbilt».*

Notre motivation est d'étudier le compromis entre la souplesse du modèle fonctionnel et l'efficacité statistique de l'estimation. La première approche n'impose en fait aucune contrainte mais elle nécessite l'estimation d'un nombre de paramètres relativement grand (256). La seconde laisse moins de latitude mais elle repose sur un petit nombre de paramètres (6).

Le schéma d'optimisation (voir chapitre 3, paragraphe 3.5). Ici, nous avons deux possibilités : l'optimisation hiérarchisée (RC0, RC3) et l'optimisation alternée (RC1, RC2, RC4, RC5). A l'exception de RC1, nous avons choisi le schéma d'optimisation le plus rapide compte tenu des choix effectués pour la métrique et pour la fonction d'intensité. La comparaison de RC1 avec RC0 permet d'évaluer directement l'influence du schéma d'optimisation sur la précision du résultat.

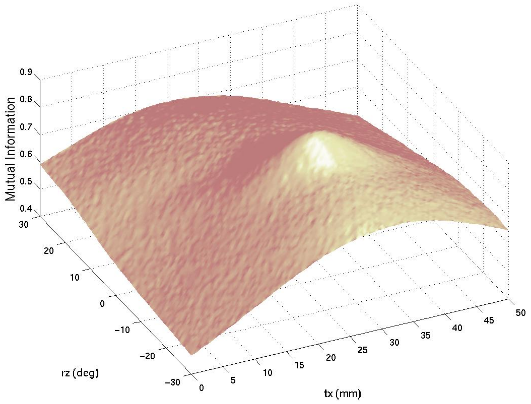
4.2.2 Implémentation commune

Le calcul de ces différentes mesures nécessite le choix d'une méthode d'interpolation (voir le chapitre 2, paragraphe 2.3.2). Nous avons opté pour l'interpolation par volume partiel (VP) dont nous avons observé qu'elle fournit, en règle générale, des paysages d'énergies plus réguliers que la méthode classique d'interpolation trilinéaire.

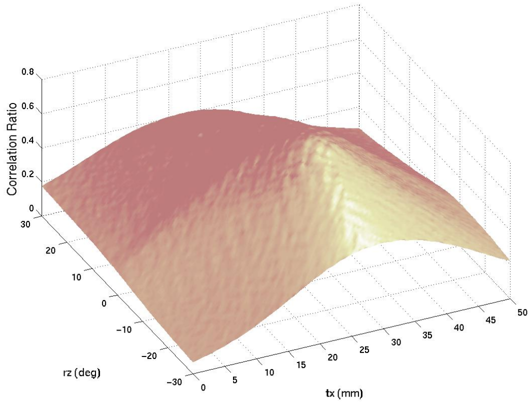
La figure 4.1 illustre l'effet de l'interpolation dans le cas d'un couple TEP/T1. Pour l'une et l'autre des deux méthodes d'interpolation évoquées, nous représentons les profils de l'information mutuelle et du rapport de corrélation RC0 obtenus en faisant varier un paramètre de translation et un paramètre de rotation par rapport à la transformation standard. On constate que les profils qui correspondent à l'interpolation trilinéaire sont plus «bruités» que ceux qui correspondent à l'interpolation VP, cet effet apparaissant d'ailleurs un peu plus prononcé dans le cas de l'information mutuelle.

Il s'agit d'une tendance générale que nous avons maintes fois observée pour d'autres couples de modalités et d'autres mesures de similarité. L'interpolation VP semble donc avoir une tendance moindre à «créer» des maxima locaux. Cette conjecture corrobore les observations de (Maes et al., 1997) dans le cas particulier de l'information mutuelle.

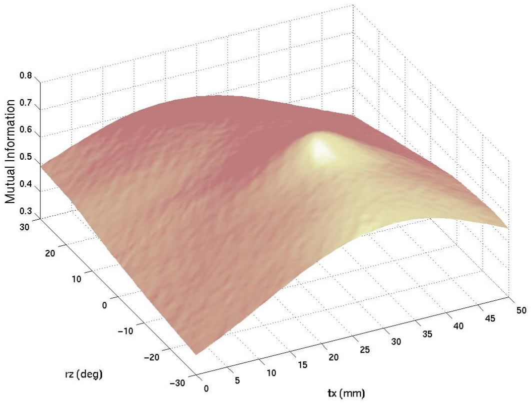
Un autre aspect essentiel de la mise en œuvre est la méthode numérique employée pour optimiser



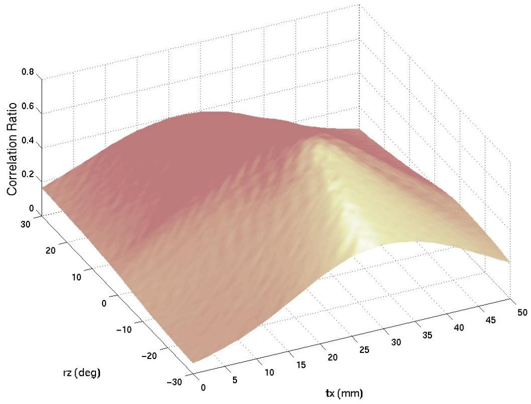
(a) Information mutuelle - interpolation trilinéaire



(b) Rapport de corrélation RC0 - interpolation trilinéaire



(c) Information mutuelle - interpolation VP



(d) Rapport de corrélation RC0 - interpolation VP

FIG. 4.1 – *Recalage rigide d'une image TEP avec une IRM-T1: profil de l'information mutuelle et du rapport de corrélation RC0 en fonction de la translation en x et de la composante en z du vecteur rotation, avec interpolation trilinéaire (première ligne) et interpolation par volume partiel (deuxième ligne).*

les mesures de similarité. Nous avons utilisé la méthode de Powell (Press et al., 1992). Notons que dans le cas d'une optimisation alternée (RC1, RC2, RC4 et RC5), plusieurs appels sont nécessaires jusqu'à la convergence (voir le chapitre 3, paragraphe 3.5). En dehors d'expériences préliminaires avec des descentes de gradient, nous n'avons pas testé d'autres méthodes d'optimisation.

La méthode de Powell est par nature une méthode *locale* : elle fournit un optimum local non nécessairement global. Le résultat de l'optimisation dépend donc de l'initialisation des paramètres de transformation. Or notre but est d'évaluer la précision de la transformation correspondant à l'optimum *global*. Autant que faire se peut, nous avons donc cherché à supprimer l'influence de l'initialisation. Ceci nous a conduit à définir des protocoles expérimentaux spécifiques aux combinaisons de modalités scanner/IRM d'une part, et TEP/IRM d'autre part.

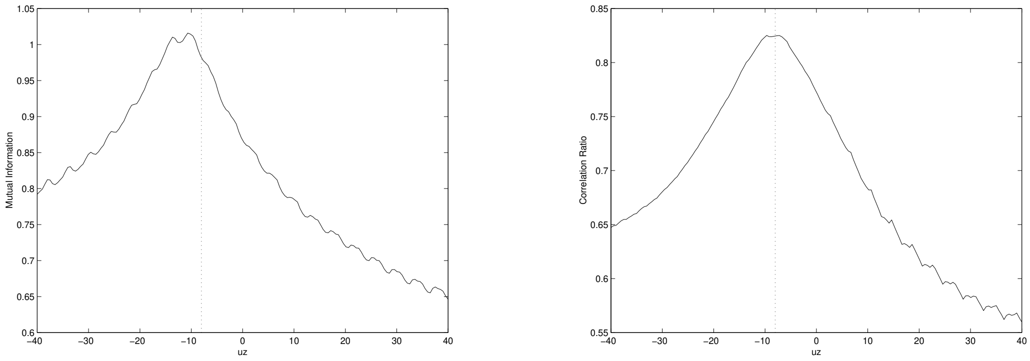
4.2.3 Recalages scanner/IRM

Une approche simple permettant d'améliorer la robustesse de l'optimisation est de l'intégrer dans un schéma multi-résolution (voir chapitre 2, paragraphe 2.5.4). Ici, nous avons mis en œuvre une pyramide gaussienne à 3 niveaux. Aux deux premiers niveaux, l'image scanner et l'image IRM sont convoluées par des filtres gaussiens d'écarts-type successifs 4 mm et 2 mm. Le dernier niveau considère les images originales (sans filtrage).

Protocole A : initialisation automatique. Dans une première série d'expériences, la transformation initiale de l'algorithme multi-résolution a été choisie comme la translation pure alignant les centres des images. Nous avons ainsi recalé les différents couples scanner/IRM avec chaque mesure de similarité. Les résultats que nous avons obtenus se sont révélés médiocres pour certaines mesures et, dans le cas de l'information mutuelle, incohérents avec ceux annoncés par plusieurs groupes dans (West et al., 1997).

Une analyse plus fine a fait apparaître que les principales sources d'erreurs concernaient de façon quasiment systématique l'estimation de la translation selon l'axe longitudinal (orthogonal aux coupes). En règle générale, les 5 autres paramètres de transformation étaient correctement estimés. Ce constat est vraisemblablement lié à la forte épaisseur de coupes des images (4 mm pour le scanner comme pour l'IRM) qui, à cause des effets de volume partiel, se traduit par un faible contraste dans la direction transverse. Cela pourrait expliquer le fait que certaines mesures de similarité manquent de sensibilité vis-à-vis de la translation longitudinale.

Pour illustrer ce propos, nous représentons sur la figure 4.2 les profils de l'information mutuelle et du rapport de corrélation RC0 obtenus en faisant varier la translation longitudinale par rapport à la position standard fournie par l'université Vanderbilt. Cette expérience correspond à un couple scanner/T1. Dans ce cas, la translation standard vaut -8.02 mm. On observe que les deux critères ont une allure heurtée et plus particulièrement l'information mutuelle. Les oscillations s'amplifient au fur et à mesure que le décalage augmente. De toute évidence, ceci est dû à l'effet cumulé du manque de contraste et du rétrécissement de la zone de recouvrement, ce dernier agissant de manière à rendre les mesures de moins en moins significatives.



(a) Information mutuelle

(b) Rapport de corrélation RC0

FIG. 4.2 – *Recalage entre une image scanner et une image T1. Profil de l'information mutuelle et du rapport de corrélation RC0 en fonction de la translation longitudinale. La ligne pointillée indique la position de la translation standard. La translation* $u_z = 0$ *est le paramètre initial de l'algorithme.*

Protocole B : initialisation manuelle. Nous avons mené une deuxième série d'expériences identique à la précédente excepté le fait que la translation longitudinale était cette fois initialisée à la main. Pour le faire, nous avons repéré visuellement les coupes homologues dans chaque couple d'images scanner/IRM ; ce travail quelque peu fastidieux garantit que les résultats des recalages ne soient pas biaisés par la connaissance des transformations standard.

4.2.4 Recalages TEP/IRM

La résolution spatiale des coupes TEP est moindre que celle des coupes scanner (2.59 mm contre 0.65 mm). C'est pourquoi nous nous sommes limités à un schéma multi-résolution à deux niveaux. Les images TEP et IRM sont convoluées par un filtre gaussien d'écart-type $\sigma = 4$ mm au premier niveau et laissées intactes au second niveau.

Les erreurs de recalage que nous avons obtenues avec une initialisation automatique (alignement des centres d'images) sont, dans la grande majorité des cas, inférieures à l'épaisseur de coupe des images TEP (8 mm). Cette observation suggère que les différentes mesures de similarité ont été optimisées globalement dans la plupart des cas. La raison en est probablement fortuite : les déplacements inter-coupes à trouver sont toujours relativement faibles, de l'ordre de une à deux coupes. Contrairement à ce que nous avons fait pour les couples scanner/IRM, nous avons donc jugé inutile de reprendre les expériences en initialisant la translation longitudinale.

4.3 Résultats

Deux exemples de recalage sont représentés sur les figures 4.3 et 4.4. Ils correspondent respectivement à un couple scanner/DP et à un couple TEP/T1. Ils ont été obtenus respectivement avec le rapport de corrélation RC4 et avec le rapport de corrélation RC3. Bien entendu, nous ne les avons pas choisis complètement au hasard. Ils correspondent à des erreurs sub-voxéliques par rapport aux transformations standard fournies par l'université Vanderbilt. En cela, ils sont représentatifs des meilleurs résultats de recalage que nous avons obtenus.

4.3.1 Calcul des erreurs de recalage

Dans (West et al., 1997), les erreurs des différents algorithmes de recalage sont calculées de la manière suivante. Un point de contrôle $\mathbf{c}$ est déterminé dans l'image cible (dans tous les cas, l'IRM). Il correspond au barycentre d'un certain nombre de voxels d'intérêt sélectionnés par un expert. Soit $\mathbf{c}'$ le point de l'image source apparié avec $\mathbf{c}$ par la transformation standard. En notant T_s la transformation standard et $\mathbf{R}_s$ la matrice de rotation qui lui est associée, on a :

$$T_s(\mathbf{x}) = \mathbf{R}_s(\mathbf{x} - \mathbf{c}') + \mathbf{c}$$

Un algorithme de recalage fournit une transformation T que l'on espère aussi proche que possible de T_s. L'erreur de contrôle[3] définie par (West et al., 1997) correspond à l'erreur commise sur la position de $\mathbf{c}$:

$$\varepsilon_c = \|T(\mathbf{c}') - \mathbf{c}\|$$

Nous avons utilisé la même approche à une nuance près : comme nous ne connaissions pas les voxels d'intérêt utilisés par West et al., nous avons choisi $\mathbf{c}$ comme un point «central» du cerveau, en l'occurrence le point le plus dorsal du corps calleux (voir la figure 4.5). Ainsi, les erreurs de contrôle que nous calculons ne sont pas exactement équivalentes à celles qui sont données dans (West et al., 1997) mais elles devraient leur être comparables. Les chiffres que nous obtenons avec l'information mutuelle nous confortent dans cette hypothèse : ils sont effectivement du même ordre que ceux que différents groupe ont annoncé dans le même article (voir les paragraphes suivants).

Si la rotation est mal estimée, l'erreur de contrôle n'est représentative de l'erreur de recalage qu'au voisinage de $\mathbf{c}$ et c'est pourquoi ce point est choisi au centre du cerveau. Pour caractériser pleinement l'écart entre la transformation trouvée et la transformation standard, il faut prendre en compte l'information complémentaire fournie par l'erreur de rotation. Nous la calculons ici comme l'angle de la rotation résiduelle $\mathbf{R}_s^t\mathbf{R}$, où $\mathbf{R}$ est la matrice de rotation associée à T :

$$\varepsilon_r = \|\delta r\|,$$

δr étant le vecteur rotation de $\mathbf{R}_s^t\mathbf{R}$. Cette mesure d'erreur semble pertinente car elle correspond à une véritable distance sur l'ensemble des rotations (Pennec, 1996). Sous l'hypothèse où ε_r est négligeable au deuxième ordre, un petit calcul suivi d'un développement limité montre que l'erreur de recalage en un point $\mathbf{x}$ est majorée par une fonction croissante de la distance $\|\mathbf{x} - \mathbf{c}'\|$:

$$\|T(\mathbf{x}) - T_s(\mathbf{x})\| \lesssim \varepsilon_c \left(1 + \frac{\varepsilon_r}{\varepsilon_c}\|\mathbf{x} - \mathbf{c}'\|\right),$$

3. *Target Registration Error.*

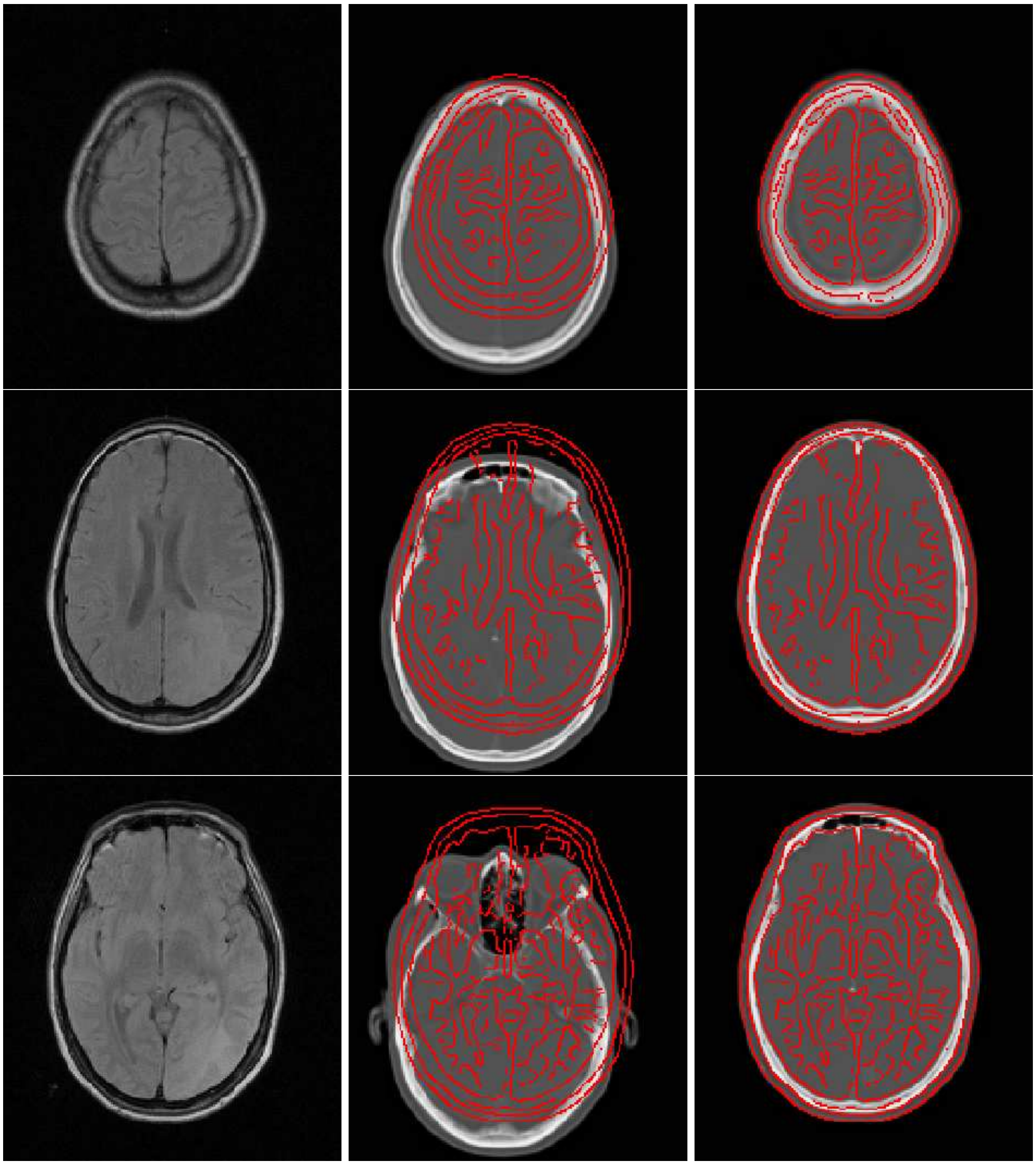


FIG. 4.3 – *Exemple de recalage rigide scanner/DP obtenu avec le rapport de corrélation RC4. Trois plans de coupe différents sont représentés. Pour chacun, on voit de gauche à droite : la coupe IRM, la coupe correspondante dans le volume scanner non-recalé et la coupe correspondante dans le volume scanner recalé. Nous superposons aux coupes scanner les contours extraits de l'IRM (par un détecteur de Canny-Deriche) pour mieux visualiser la qualité du recalage.*

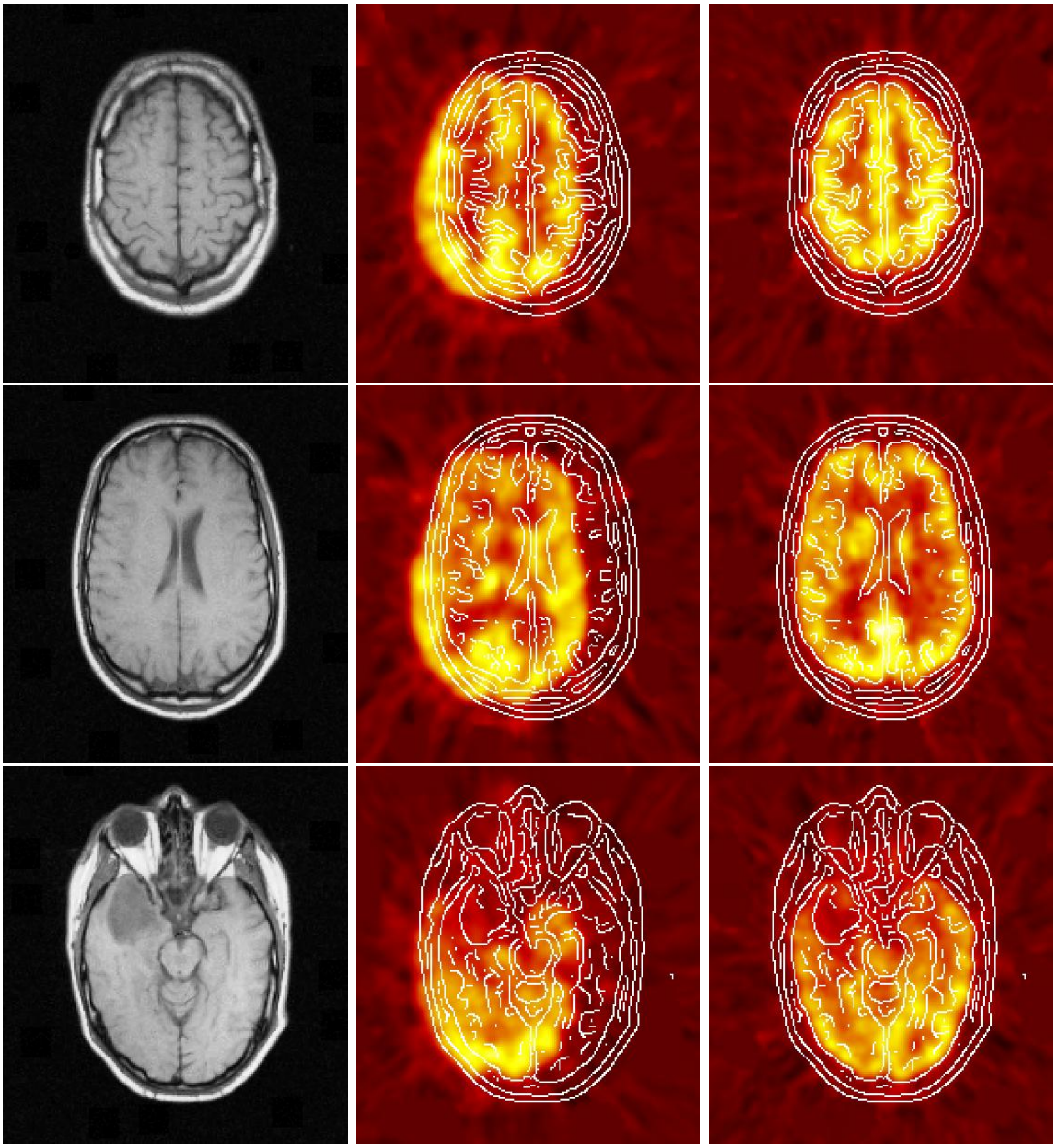


Fig. 4.4 – *Exemple de recalage rigide TEP/T1 obtenu avec le rapport de corrélation RC3. Trois plans de coupe différents sont représentés. Pour chacun, on voit de gauche à droite : la coupe IRM, la coupe correspondante dans le volume TEP non-recalé et la coupe correspondante dans le volume TEP recalé. Nous superposons aux coupes TEP les contours extraits de l'IRM (par un détecteur de Canny-Deriche) pour mieux visualiser la qualité du recalage.*

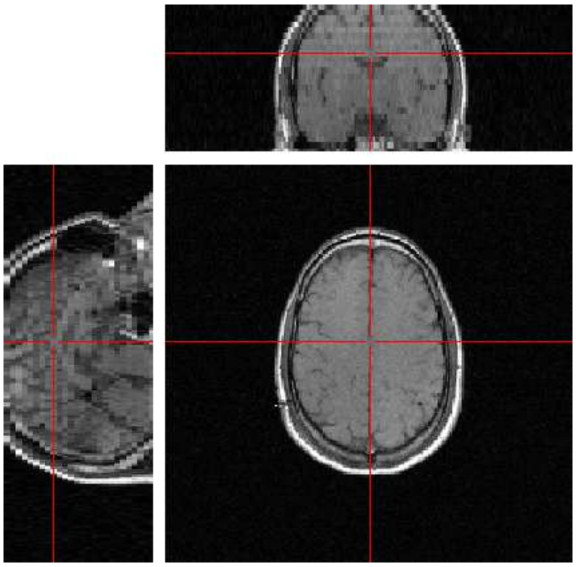


FIG. 4.5 – *Localisation dans une IRM-T1 du point de contrôle utilisé pour le calcul des erreurs de recalage. C'est approximativement le point le plus dorsal du corps calleux, une commissure inter-hémisphérique que l'on discerne très bien sur la vue sagittale. Merci à Jérôme Yelnik (U289 INSERM) pour ses éclaircissements.*

où ε_r est exprimée en radians. Ainsi, pour constater une erreur millimétrique deux fois supérieure à l'erreur de contrôle, il faut choisir un point $\mathbf{x}$ dont la distance à $\mathbf{c}'$ est au moins égale au rapport $\varepsilon_c/\varepsilon_r$ si ε_r est exprimée en radians, et donc environ 60 fois ce rapport si ε_r est exprimée en degrés.

4.3.2 Présentation des résultats

Les résultats des différentes séries d'expérience sont résumés par les tableaux 4.2, 4.3 et 4.4 correspondant respectivement aux recalages scanner/IRM (protocole A), scanner/IRM (protocole B) et TEP/IRM. Nous reportons, pour chaque couple de modalités et pour chaque mesure de similarité, la médiane et la moyenne quadratique[4] des erreurs de contrôle ε_c d'une part, et des erreurs de rotation ε_r d'autre part. Dans tous les cas, le calcul de ces valeurs fait intervenir 8 patients, d'où la nécessité de les comparer avec prudence.

Dans toutes les expériences de recalage scanner/IRM, le critère de Woods a donné des résultats grossièrement faux (erreurs de l'ordre de plusieurs centimètres). C'est pourquoi il est omis dans les tableaux 4.2 et 4.3. Ces échecs répétés s'expliquent par son caractère numériquement instable. Comme nous l'avons vu au 3, paragraphe 3.3.4, le critère de Woods fait intervenir une série de divisions par les moyennes d'intensité de l'image source (ici, le scanner) calculées sur les différents ensembles de voxels induits par la partition de l'image cible (ici, l'IRM). Les images scanner étant

4. *Root Mean Square* en anglais. La moyenne quadratique est plus adéquate que la moyenne arithmétique si on suppose, comme c'est naturel, que les erreurs au carré suivent une loi du χ^2. En tout état de cause, il n'y a pas de différence qualitative à considérer l'une plutôt que l'autre.

très peu bruitées, leur fond a une intensité proche de zéro. Or, pour une transformation spatiale proche du recalage idéal, le fond de l'image scanner est apparié avec le fond de l'image IRM. Il se produit alors une division par une valeur très faible qui pénalise lourdement la transformation considérée. Ce problème ne se pose pas dans le cas des images TEP car leur fond est nettement plus bruité.

Les erreurs de contrôle obtenues pour les recalages scanner/T1, scanner/T2 et scanner/DP sont représentées graphiquement sur les figures 4.6 à 4.8. Elles correspondent au protocole B. Le critère de Woods est également absent de ces figures. Afin d'alléger la présentation, nous ne fournissons pas de figures équivalentes pour les erreurs correspondant au protocole A, celles-ci étant moins révélatrices du niveau de précision des différentes mesures de similarité. Les figures 4.9 à 4.11 représentent les erreurs de contrôle obtenues pour les recalages TEP/T1, TEP/T2 et TEP/DP.

Ces erreurs doivent être comparées à la résolution spatiale des images, sachant que le paramètre critique est l'épaisseur des coupes. Dans le cas des couples scanner/IRM, les deux modalités ont la même épaisseur de coupes : 4 mm. Les erreurs inférieures à 4 mm indiquent donc une précision sub-voxélique au voisinage du point de contrôle $\mathbf{c}$. En ce qui concerne les couples TEP/IRM, la précision des recalages est intrinsèquement limitée par l'épaisseur de coupes des images TEP, qui est de 8 mm.

De manière générale, les valeurs médianes sont plus représentatives de la précision d'une mesure que les moyennes quadratiques. Il suffit en effet que l'algorithme ait nettement échoué dans un cas pour que la moyenne quadratique soit anormalement élevée. Un tel échec est généralement dû à la convergence vers un minimum local non global. La comparaison de la moyenne quadratique et de la médiane donne une indication sur la robustesse d'une mesure vis-à-vis de la méthode d'optimisation. Il est clair qu'en utilisant une méthode plus robuste que celle de Powell, nous n'aurions probablement pas constaté d'échecs flagrants. L'utilisation de la méthode de Powell est cependant pertinente d'un point de vue clinique dans la mesure où elle est relativement rapide.

4.3.3 Commentaires

4.3.3.1 Recalages scanner/IRM

La comparaison des tableaux 4.2 et 4.3 montre que les erreurs correspondant au protocole B sont nettement plus faibles que celles du protocole A. L'initialisation manuelle a donc permis une amélioration notable des performances d'ensemble. Toutefois, les gains en précision sont inégaux selon les mesures de similarité : ils sont particulièrement nets pour l'information mutuelle, moins sensibles pour RC0, RC1, RC2 et RC3, et quasiment inexistants pour RC4 et RC5. Ces résultats suggèrent que l'utilisation d'une métrique robuste permet de réduire l'influence de l'initialisation.

Concernant les couples scanner/T1 et scanner/DP, les figures 4.6 et 4.8 montrent clairement que toutes les mesures donnent des résultats comparables (erreurs de contrôle médianes de l'ordre de 2 mm) exception faite de la mesure RC2, légèrement moins précise que les autres. Cette dernière est fondée sur la recherche d'une fonction d'intensité polynômiale ; c'est le seul aspect qui la distingue de

RC1, qui repose sur une correction d'intensité constante par morceaux. Le choix d'un degré maximal de 5 s'avère donc légèrement insuffisant. Néanmoins, la comparaison de RC4 et RC5 n'étant pas de façon évidente à l'avantage de RC4, il semblerait que l'utilisation d'une métrique robuste ait tendance à combler le manque de souplesse du modèle polynômial.

Les recalages impliquant les acquisitions T2 résèrvent une petite surprise par rapport aux deux autres combinaisons scanner/IRM. La figure 4.7, en particulier, révèle que les rapports de corrélation fondés sur une métrique quadratique (RC0, RC1, RC2) sont nettement moins efficaces que les autres mesures. Dans ce cas, l'utilisation d'une métrique robuste paraît indispensable. Il convient donc de remettre en cause l'hypothèse d'une dépendance fonctionnelle entre les intensités d'un scanner et d'une IRM-T2. Le choix d'une métrique robuste permet de juguler cette difficulté en imposant qu'il y ait une correspondance fonctionnelle, non pas pour la totalité, mais pour la majorité des voxels appariés.

Sur l'ensemble des expériences de recalage scanner/IRM, les mesures qui se détachent sont l'information mutuelle et les rapports de corrélation robustes (RC3, RC4, RC5) avec peut-être un léger avantage à l'information mutuelle en termes de précision. En pratique, nous avons cependant tendance à préférer RC4 pour sa meilleure stabilité vis-à-vis de l'initialisation des paramètres de transformation. Nous remarquons également qu'il n'y a pas de différence flagrante entre RC0 et RC1, ce qui tend à prouver que les schémas d'optimisation hiérarchisé et alterné sont sensiblement équivalents en pratique.

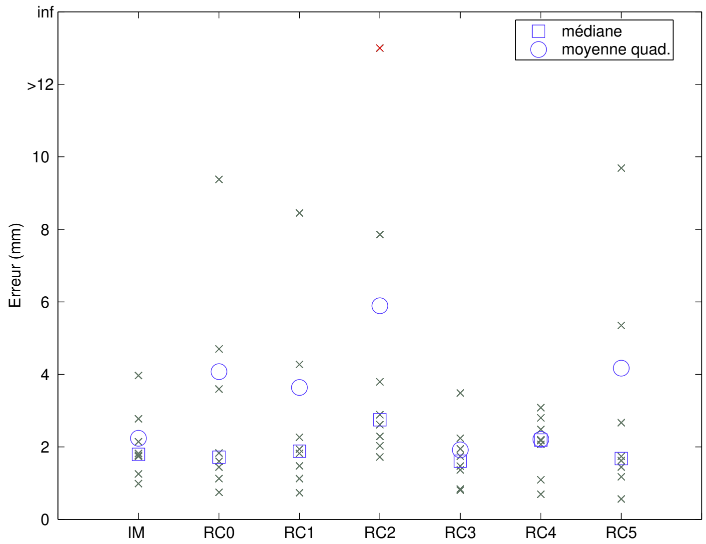


FIG. 4.6 – *Erreurs de contrôle pour les recalages scanner/T1 (protocole B).*

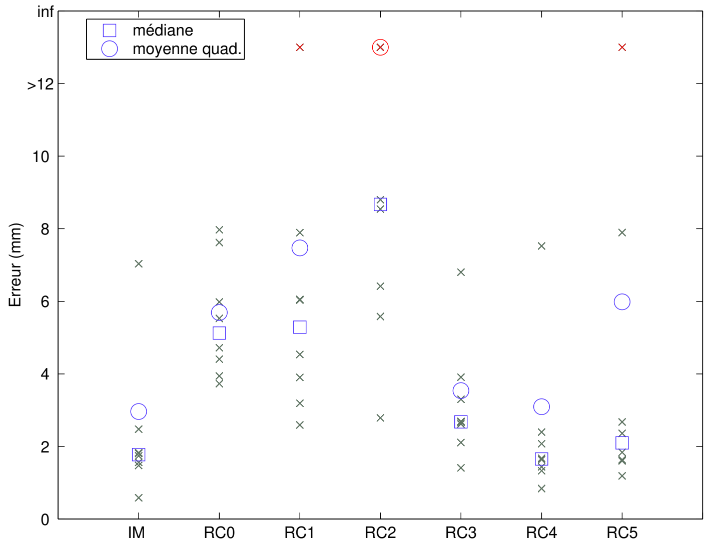


FIG. 4.7 – *Erreurs de contrôle pour les recalages scanner/T2 (protocole B).*

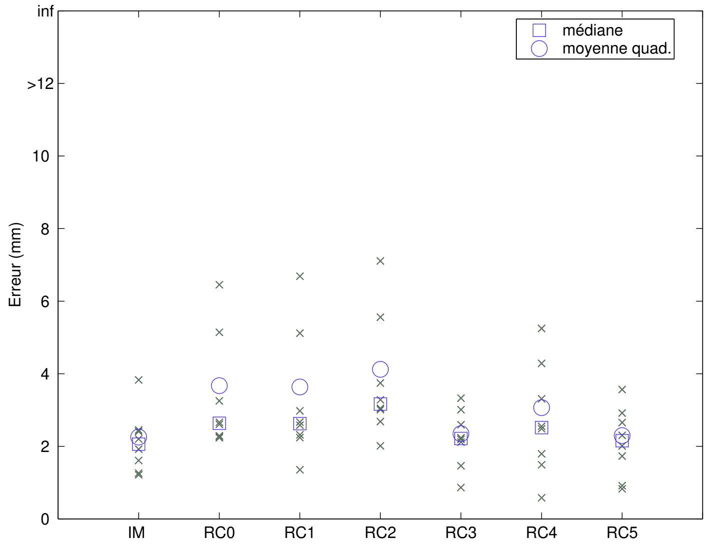


FIG. 4.8 – *Erreurs de contrôle pour les recalages scanner/DP (protocole B).*

4.3.3.2 Recalages TEP/IRM

L'analyse des recalages TEP/IRM fait apparaître une hiérarchie différente : voir le tableau 4.4 et les figures 4.9 à 4.11. Outre son manque de stabilité déjà constaté, l'information mutuelle ne fait partie des mesures les plus précises que pour les couples TEP/DP. De la même manière, les rapports de corrélation RC4 et RC5, pourtant efficaces dans les expériences de recalage scanner/IRM, sont cette fois très décevants. Cela est d'autant plus remarquable que RC3 s'affirme sans doute comme la meilleure mesure sur l'ensemble des couples TEP/IRM. Dans le cas des images TEP, une métrique robuste serait donc pertinente à condition de ne pas être «trop robuste».

L'explication de ce résultat est assez subtile. Comme nous l'avons souligné au paragraphe 3.4.2 du chapitre 3, la vocation d'un rapport de corrélation robuste est d'être à la fois sensible aux vrais négatifs (couples d'intensité traduisant une erreur de recalage) et insensible aux faux négatifs (couples d'intensités en désaccord avec l'hypothèse de dépendance fonctionnelle, du fait d'occultations, d'effets de volume partiel, ...). Le fait qu'une modalité comme la TEP ait un faible rapport contraste/bruit rend difficile la discrimination entre vrais et faux négatifs car les différences locales d'intensité dues aux uns et aux autres auront tendance à être du même ordre de grandeur. On conçoit alors qu'une métrique trop robuste risque de limiter systématiquement l'influence des vrais négatifs.

Les rapports de corrélation fondés sur la métrique quadratique (RC0, RC1, RC2) donnent des résultats corrects dans l'ensemble mais moins bons que RC3. Ici encore, RC2 est un peu moins précis que RC1, ce qui suggère le caractère limitatif d'un degré maximal de 5 pour les fonctions d'intensité polynômiales. La comparaison de RC0 et RC1 confirme qu'il n'y a pas d'influence manifeste du schéma d'optimisation.

Enfin, le critère de Woods s'avère particulièrement inconstant au cours des expériences et, de ce fait, assez imprécis en moyenne. Ce constat explique a posteriori pourquoi (Woods et al., 1993) ont cherché à améliorer leur technique par une étape préliminaire de segmentation.

4.4 Conclusion

Nous avons testé plusieurs mesures de similarité dans le cadre du recalage multimodal rigide d'images provenant de la base de données «Vanderbilt» : l'information mutuelle telle qu'elle est décrite dans (Maes et al., 1997), le critère de (Woods et al., 1993) et 6 variantes du rapport de corrélation généralisé (voir le chapitre 3). L'optimisation a été réalisée au moyen d'une implémentation commune utilisant l'interpolation par volume partiel et la méthode de Powell imbriquée dans un schéma multi-résolution. La comparaison des transformations calculées par notre algorithme avec les transformations standard fournies par l'université Vanderbilt nous a permis d'évaluer les erreurs commises par chaque mesure de similarité.

Conformément aux conclusions de (West et al., 1997), nos résultats confirment que l'information mutuelle est généralement précise pour les recalages scanner/IRM et TEP/IRM. Nous observons cependant qu'elle a échoué grossièrement dans un certain nombre de cas. Cela s'explique par le

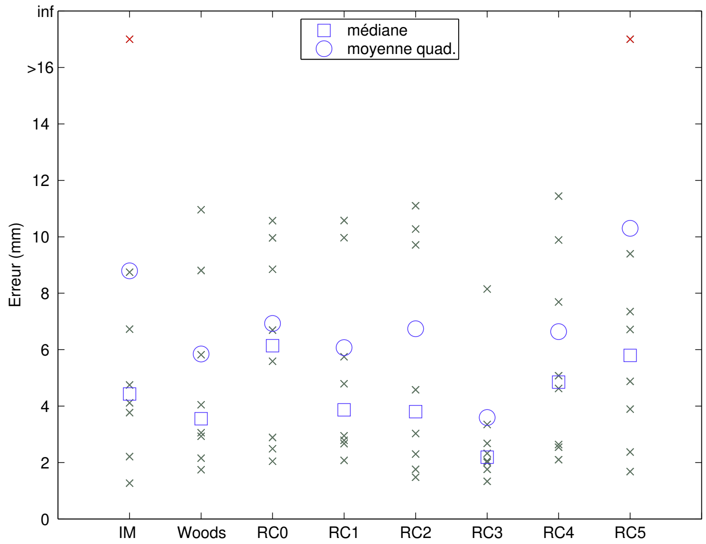


FIG. 4.9 – *Erreurs de contrôle pour les recalages TEP/T1.*

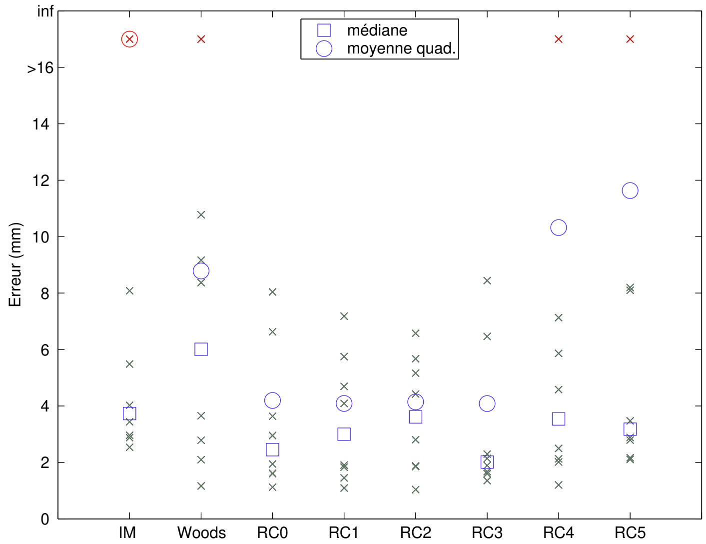


FIG. 4.10 – *Erreurs de contrôle pour les recalages TEP/T2.*

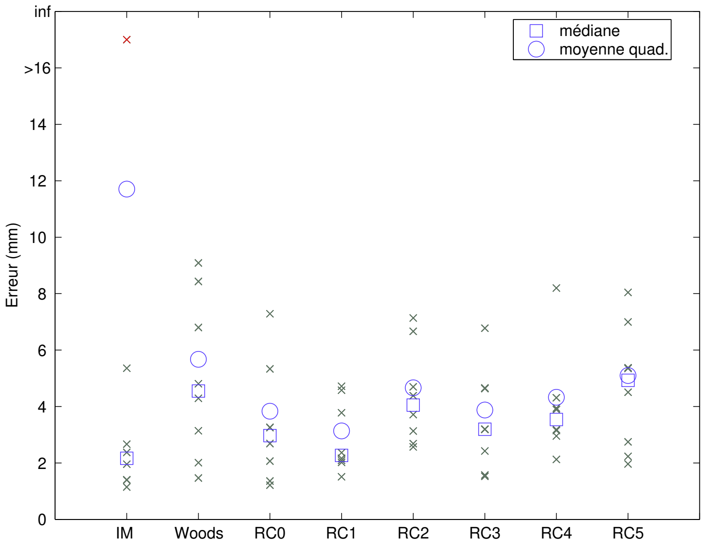


FIG. 4.11 – *Erreurs de contrôle pour les recalages TEP/DP.*

manque de robustesse de notre technique d'optimisation mais, plus profondément, par le fait que l'information mutuelle a tendance à présenter de nombreux maxima locaux, ce qui rend son optimisation globale particulièrement difficile. Le critère de Woods fournit quant à lui des erreurs très variables pour les recalages TEP/IRM et il se montre totalement inefficace pour les recalages scanner/IRM.

Concernant les variantes du rapport de corrélation, nous obtenons des niveaux de performances différents entre eux et différents selon les combinaisons de modalités, que ce soit en termes de précision moyenne ou de variabilité. Il semble que la caractéristique la plus déterminante à cet égard soit la métrique de comparaison d'intensité. Pour les recalages scanner/IRM, nous avons trouvé que les rapports de corrélation fondés sur la métrique de Geman-McClure étaient les plus précis alors que pour les recalages TEP/IRM, nos expériences suggèrent l'utilisation d'une métrique «moins robuste», en l'occurrence la norme L_1. L'espace de recherche de la fonction de correction semble lui aussi avoir un impact non négligeable sur les performances du recalage. De ce point de vue, il s'avère préférable de ne pas imposer trop de contraintes. Enfin, le schéma d'optimisation ne se révèle pas d'une importance décisive.

A condition de bien choisir la métrique de comparaison d'intensité, le rapport de corrélation généralisé peut donner des résultats au moins aussi précis que l'information mutuelle et incontestablement plus reproductibles, même si ce deuxième aspect reste problématique pour toutes les mesures que nous avons testées. Malgré la nature simplificatrice de l'hypothèse de dépendance fonctionnelle, le rapport de corrélation se révèle être un concept pertinent pour le recalage multimodal d'images médicales.

Couples de modalités	Mesure de similarité	Erreur de contrôle (mm)		Erreur de rotation (deg)	
		médiane	moy. quad.	médiane	moy. quad.
Scanner/T1	IM	5.24	19.72	0.79	25.11
	RC0	2.70	9.36	0.69	1.02
	RC1	2.86	4.78	0.74	0.87
	RC2	3.41	6.26	1.50	1.72
	RC3	4.39	9.71	0.92	1.08
	RC4	2.04	1.99	1.07	1.77
	RC5	1.75	2.62	1.03	1.95
Scanner/T2	IM	6.96	11.68	1.16	16.00
	RC0	10.17	12.47	1.07	3.53
	RC1	6.08	26.06	0.84	8.84
	RC2	13.97	22.85	2.41	14.96
	RC3	5.97	10.29	1.29	2.03
	RC4	1.58	2.93	1.59	1.78
	RC5	2.49	7.13	4.33	10.20
Scanner/DP	IM	6.98	10.85	0.99	1.30
	RC0	4.39	10.04	1.31	1.34
	RC1	4.38	7.89	1.11	1.16
	RC2	5.12	17.35	1.14	16.67
	RC3	2.80	9.98	1.30	1.46
	RC4	2.56	3.07	0.98	1.09
	RC5	2.15	2.29	1.00	1.55

TAB. 4.2 – *Erreurs de recalage Scanner/IRM pour le protocole A (initialisation automatique des paramètres de transformation par alignement des centres d'images).*

Couples de modalités	Mesure de similarité	Erreur de contrôle (mm)		Erreur de rotation (deg)	
		médiane	moy. quad.	médiane	moy. quad.
Scanner/T1	IM	1.79	2.24	0.85	2.27
	RC0	1.71	4.07	1.05	3.06
	RC1	1.88	3.64	0.97	2.74
	RC2	2.75	5.89	1.60	2.46
	RC3	1.60	1.92	0.92	1.46
	RC4	2.19	2.21	0.84	1.31
	RC5	1.68	4.17	0.93	1.84
Scanner/T2	IM	1.77	2.96	1.12	1.75
	RC0	5.12	5.69	2.83	6.26
	RC1	5.29	7.47	2.62	4.38
	RC2	8.67	18.44	5.68	15.07
	RC3	2.68	3.54	1.19	2.14
	RC4	1.66	3.10	1.67	1.87
	RC5	2.10	5.98	4.35	9.14
Scanner/DP	IM	2.06	2.25	1.19	1.42
	RC0	2.63	3.67	1.67	3.04
	RC1	2.63	3.64	1.24	2.03
	RC2	3.17	4.12	2.23	2.62
	RC3	2.21	2.35	1.14	1.21
	RC4	2.51	3.07	0.98	1.09
	RC5	2.15	2.29	1.00	1.55

TAB. 4.3 – *Erreurs de recalage Scanner/IRM pour le protocole B (les paramètres de transformation initiaux correspondent à aligner les coordonnées x,y des centres d'images et à estimer visuellement la translation longitudinale).*

Couples de modalités	Mesure de similarité	Erreur de contrôle (mm)		Erreur de rotation (deg)	
		médiane	moy. quad.	médiane	moy. quad.
TEP/T1	IM	4.43	8.79	3.09	21.66
	Woods	3.55	5.85	1.91	2.42
	RC0	6.14	6.93	2.08	2.80
	RC1	3.87	6.07	2.00	2.99
	RC2	3.80	6.74	3.49	5.33
	RC3	2.19	3.59	2.47	3.82
	RC4	4.85	6.64	2.88	5.03
	RC5	5.80	10.30	5.50	14.05
TEP/T2	IM	3.73	26.88	3.58	3.55
	Woods	6.01	8.79	1.85	3.14
	RC0	2.45	4.19	1.76	2.64
	RC1	3.00	4.09	2.43	3.09
	RC2	3.61	4.15	2.01	3.05
	RC3	2.02	4.09	1.24	1.93
	RC4	3.54	10.32	4.37	14.91
	RC5	3.18	11.63	5.54	15.88
TEP/DP	IM	2.16	11.70	2.49	21.00
	Woods	4.55	5.67	1.97	2.72
	RC0	2.97	3.83	1.63	2.21
	RC1	2.27	3.13	1.86	2.54
	RC2	4.05	4.66	2.47	4.14
	RC3	3.20	3.88	1.25	1.79
	RC4	3.54	4.33	3.51	5.97
	RC5	4.93	5.10	5.32	5.91

TAB. 4.4 – *Erreurs de recalage TEP/IRM.*

Chapitre 5

Le recalage d'images vu comme un problème d'inférence statistique

5.1 Introduction

Les mesures de similarité étudiées au cours des chapitres précédents nous ont permis de recaler de nombreux couples d'images médicales issues de diverses modalités d'acquisition. Il s'agit d'un constat pratique tout à fait agréable, qui illustre la puissance des approches iconiques mais qui ne saurait en aucun cas l'expliquer.

Il nous arrive aussi de constater l'échec d'une mesure de similarité à recaler un couple d'images particulier alors qu'elle fonctionne habituellement bien avec des images du même type. Le chapitre 4 a fourni un certain nombre d'exemples de ce genre. Ils ne sont pas l'apanage d'une mesure particulière, bien que certaines y soient plus disposées que d'autres. Bien souvent, ceci est dû à la convergence de l'algorithme d'optimisation vers un extremum local mais non-global. Il est alors possible d'obtenir un recalage satisfaisant en initialisant différemment les paramètres de la transformation spatiale. Reste à savoir pourquoi le premier essai n'a pas marché : est-ce dû à des paramètres d'acquisition atypiques, à l'anatomie particulière du sujet, ...?

Nous savons bien que les mesures de similarité fondées sur des comparaisons d'intensité sont fortement non-convexes. Aussi n'est-il pas surprenant que l'optimisation échoue de temps en temps. On peut imputer la responsabilité d'un échec au manque de robustesse de l'algorithme d'optimisation. Mais les défauts de convexité de la mesure peuvent aussi être considérés comme révélateurs d'une certaine inadéquation. Nos expériences de recalage d'images ultrasonores et IRM nous ont confortés dans cette idée. Sans anticiper sur le chapitre 6, nous avons constaté qu'avec un même algorithme d'optimisation, les mesures de similarité conventionnelles manifestent des taux d'échecs sensiblement plus élevés que la mesure que nous avons spécifiquement conçue pour ce problème.

5.1.1 De l'intérêt de la modélisation

La question qui se dissimule derrière ces remarques est celle de la reproductibilité de nos méthodes de recalage. Comment pouvons-nous garantir le bon comportement d'une mesure de similarité vis-à-vis de la versatilité des procédés d'acquisition et vis-à-vis de la variabilité anatomique des sujets ? Deux réponses complémentaires peuvent être données : l'une est pratique, l'autre théorique. La réponse pratique est bien sûr la validation. Seule l'utilisation d'une grande base de donnée d'images permet réellement d'évaluer les performances d'une méthode de recalage (précision, taux de succès, etc.). Toutefois, la validation ne permet pas en tant que telle d'améliorer une méthode dont on constate expérimentalement les insuffisances, et elle reste en cela une réponse incomplète.

Il nous semble donc indispensable de tenir compte de la réponse théorique qu'est la modélisation. Le recalage d'images médicales est fondamentalement un problème physique : nous disposons d'images dont les intensités mesurent des grandeurs physiques émanant de la même réalité anatomique (ou de réalités anatomiques supposées comparables). Si nous étions capables de modéliser les processus d'acquisition d'une part, et l'anatomie commune d'autre part, nous pourrions en déduire la relation théorique qui existe entre les intensités des images. De plus, nous pourrions prédire la façon dont cette relation est modifiée sous l'effet de variations anatomiques ou de changements dans les conditions d'acquisition.

Une telle approche peut sembler naïve au vu de la complexité des phénomènes mis en jeu dans la formation d'images médicales. Pourtant plusieurs groupes se sont intéressés à la modélisation de processus d'imagerie. Par exemple, un simulateur d'images par résonance magnétique fondé sur les équations de Bloch a été développé au Montreal Neurological Institute (Collins et al., 1998). Il génère des images cérébrales d'un réalisme remarquable à partir d'un atlas labelisé.

La modélisation anatomique fait également l'objet de recherches actives. Pour rester dans le contexte du cerveau, nous pouvons citer les travaux concernant la construction d'atlas probabilistes (Toga, 1999) ou les représentations structurées de la topographie corticale (Mangin et al., 1995; Régis et al., 1995). Nous pensons que les développements effectués dans le domaine de la modélisation auront, à terme, un impact considérable sur les techniques de traitement d'images et notamment sur les techniques de recalage.

Le grand paradoxe de nos mesures de similarité est qu'elles ne reposent sur aucune base physique. En apparence seulement car, en réalité, mesurer la similarité de deux images revient à mesurer l'adéquation de leur distribution conjointe d'intensités avec certaines hypothèses de dépendance plus ou moins explicites. Nous savons par exemple que le rapport de corrélation est adapté au cas d'une liaison fonctionnelle entre les intensités, alors que l'information mutuelle rend compte d'une forme de dépendance plus générale que l'on peut qualifier de statistique. Il faut cependant reconnaître que ces hypothèses de dépendance sont assez vagues et qu'il est difficile d'évaluer leur pertinence physique.

Par ailleurs, le fait qu'une mesure de similarité s'avère efficace pour recaler un couple d'images autorise tout au plus à conclure que les hypothèses sur lesquelles elle repose sont acceptables dans le cas particulier de ces images. Mais rien ne prouve que si les paramètres d'acquisition avaient été différents, si le patient ou les organes imagés avaient été différents, la relation entre les intensités aurait été de même nature. Autrement dit, même si l'on est capable d'expliciter le modèle physique auquel une mesure de similarité se réfère, celui-ci peut ne pas être suffisamment riche pour prédire les fluctuations auxquelles on risque d'être confronté en pratique. Plutôt que de se voir imposer un modèle par le choix d'une mesure de similarité, nous préfèrerions choisir la mesure de similarité en fonction d'un modèle.

Ce chapitre répond par conséquent à un double objectif :

1. relier les mesures de similarité usuelles à des hypothèses de dépendance clairement établies ;
2. définir une méthodologie permettant de construire une mesure de similarité en fonction d'hypothèses de dépendance.

Nous avertissons le lecteur que notre étude n'a pas vocation à décrire une technique de recalage fondée sur la physique des images médicales. Le problème qui nous intéresse est d'affiner le lien entre les modèles, aussi simplistes soient-ils, et les mesures de similarité. Ce n'est qu'un premier pas vers la prise en compte de modèles réalistes.

5.1.2 Plan du chapitre

Au paragraphe 5.2, nous commençons par nous donner un modèle générique de la formation de deux images représentant la même réalité anatomique. Outre le fait qu'il est rudimentaire, ce modèle a la caractéristique d'être probabiliste. Cela ne signifie pas que les images résultent fondamentalement de phénomènes aléatoires, ce qui est d'ailleurs faux. La modélisation probabiliste est simplement un outil mathématique commode pour décrire des phénomènes, même complètement déterministes, dont on ne maîtrise pas toute la complexité.

Le modèle fait intervenir des paramètres décrivant la transformation géométrique relative entre les images mais aussi des paramètres caractérisant la relation entre leurs intensités ; ces derniers sont appelés paramètres auxiliaires. Le problème du recalage apparaît naturellement comme un problème d'inférence statistique puisqu'il s'agit d'estimer l'ensemble des paramètres inconnus (paramètres géométriques et paramètres auxiliaires) à partir de l'observation des images.

Nous décrivons au paragraphe 5.3 le principe d'une estimation par maximum de vraisemblance, non sans évoquer d'autres stratégies possibles. Selon les paramètres auxiliaires que l'on considère comme connus ou non, on a affaire à différents problèmes d'optimisation. Il s'avère commode de traiter l'optimisation en fonction des seuls paramètres auxiliaires, les paramètres géométriques étant fixés, dans la mesure où ce problème réduit bénéficie de méthodes de résolution spécifiques, voire de solutions analytiques selon les cas.

Le paragraphe 5.4 s'attache à détailler un certain nombre de ces cas. Le rapport de corrélation et l'information mutuelle au sens de (Viola, 1995) apparaissent alors comme des instanciations particulières du critère de vraisemblance qui correspondent à des cas limites du modèle général. Nous montrons également que dans les cas intermédiaires, l'estimation des paramètres auxiliaires peut être réalisée au moyen d'un algorithme EM.

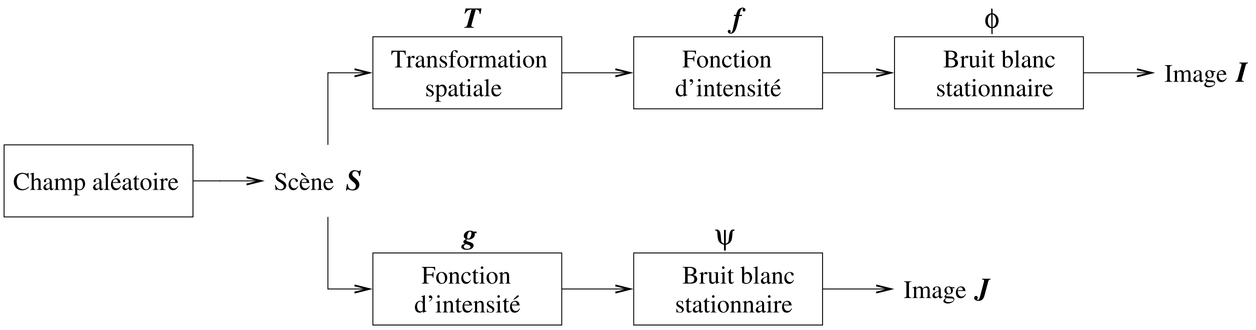


FIG. 5.1 – *Représentation schématique du modèle de dépendance inter-images.*

Enfin, parce que notre modèle repose sur une formulation spatialement discrète, la mise en œuvre pratique du recalage par maximum de vraisemblance fait intervenir des techniques ad-hoc d'interpolation. Nous discutons brièvement cet aspect au paragraphe 5.5.

5.2 Modèle probabiliste

Une image médicale est un ensemble de mesures locales d'une certaine grandeur physique. C'est la nature du procédé d'acquisition qui détermine les valeurs d'intensité prises par les différents types de tissus. En outre, le problème du recalage de deux images n'a de sens que si celles-ci sont reliées par une réalité anatomique commune. Ainsi, modéliser la relation qui existe entre les intensités des images doit faire intervenir, d'une part, la nature des procédés d'acquisition et, d'autre part, une description de l'anatomie sous-jacente.

Le modèle général que nous considérons dans ce chapitre est résumé par la figure 5.1. Les deux images à recaler $I : \Omega_I \rightarrow \mathbb{R}$ et $J : \Omega_J \rightarrow \mathbb{R}$ sont vues comme les réalisations de deux processus aléatoires dépendant d'une image labélisée S que nous appelons la *scène*. Cette dernière est censée représenter de façon parfaite l'anatomie commune aux deux images en attribuant à chaque voxel une valeur arbitraire mais caractéristique du tissu auquel il appartient. Puisque S est a priori inconnue, nous la considérons également comme la réalisation d'un processus aléatoire.

5.2.1 Modèles d'acquisition

Nous nous inspirons de modèles classiques en restauration d'images. Dans ces approches, le passage d'une scène parfaite à une image réelle fait intervenir une composante déterministe et une composante stochastique. La composante déterministe est une fonction d'intensité qui, à chaque label de S, associe une valeur réelle. Ces valeurs peuvent être interprétées comme les intensités caractéristiques des différents tissus anatomiques. La composante stochastique est un bruit additif destiné à prendre en compte différents artefacts communs : bruit de mesure, bruit de quantification, bruit biologique, etc. Ces modèles d'acquisition peuvent s'écrire formellement :

$$\left\{\begin{array}{lcl} I(\mathbf{x}_k) & = & f\Big(S\big(T(\mathbf{x}_k)\big)\Big) + \varepsilon_k \\ J(\mathbf{y}_\ell) & = & g\Big(S(\mathbf{y}_\ell)\Big) + \varepsilon'_\ell \end{array}\right. \tag{5.1}$$

où f et g sont respectivement les fonctions d'intensité correspondant à I et J, et T est la transformation spatiale qui relie les repères de Ω_I et Ω_J. Par convention, nous supposons que l'image J est alignée avec la scène, et il n'y a alors pas lieu de faire intervenir une transformation spatiale dans le modèle d'acquisition de J. Cette convention est justifiée par le fait que nous nous intéressons au déplacement relatif entre I et J. Ainsi, et comme au cours des chapitres précédents, notre formulation fait jouer des rôles différents aux images : I est l'image *source* tandis que J est l'image *cible*.

Concernant les processus aléatoires ε et ε', nous allons faire les hypothèses suivantes :

1. ε et ε' sont des bruits blancs. Autrement dit, leurs réalisations ponctuelles sont mutuellement indépendantes. La probabilité d'une réalisation est alors égale au produit des probabilités des réalisations ponctuelles :
$$\left\{\begin{array}{lcl} P(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n) & = & \prod_k P(\varepsilon_k) \\ P(\varepsilon'_1,\varepsilon'_2,\ldots,\varepsilon'_m) & = & \prod_\ell P(\varepsilon'_\ell) \end{array}\right.$$
2. ε et ε' sont stationnaires, ce qui signifie que la densité de probabilité d'une réalisation ponctuelle est indépendante de l'espace. Nous noterons ϕ et ψ les fonctions de densité de probabilité associées respectivement à ε et ε'. Nous nous réservons le droit de spécifier ultérieurement leur forme paramétrique (gaussienne, laplacienne, uniforme, ...). Il paraît cependant raisonnable de se limiter à des distributions unimodales.
3. ε et ε' sont indépendants de S et indépendants entre eux, d'où la relation :
$$P(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n,\varepsilon'_1,\varepsilon'_2,\ldots,\varepsilon'_m|S) = \prod_k \phi(\varepsilon_k) \prod_\ell \psi(\varepsilon'_\ell) \tag{5.2}$$

Sous ces hypothèses, la probabilité conditionnelle du couple d'images, connaissant la scène, s'écrit sous la forme d'un simple produit de probabilités ponctuelles. En définissant comme d'habitude $i_k \equiv I(\mathbf{x}_k)$, $j_\ell = J(\mathbf{y}_\ell)$ et $s_k^\downarrow \equiv S(T(\mathbf{x}_k))$, nous remarquons que les images sont conditionnellement indépendantes (ce qui ne veut pas dire qu'elles sont indépendantes) :

$$P(I,J|S) = \underbrace{\prod_k \phi\left(i_k - f(s_k^\downarrow)\right)}_{P(I|S)} \underbrace{\prod_\ell \psi\left(j_\ell - g(s_\ell)\right)}_{P(J|S)} \tag{5.3}$$

Bien entendu, la loi conditionnelle $P(I,J|S)$ dépend des paramètres T, f, g, ϕ et ψ, de sorte que nous pourrions aussi bien la noter $P(I,J|S,T,\ldots,\psi)$. Cette dernière notation n'est rigoureuse que si $T,\ldots,\psi$ sont considérés comme des variables aléatoires et non comme des paramètres. Il s'agit d'une question de point de vue et la distinction n'est pas vraiment fondamentale.

Il importe davantage de réaliser les limitations des modèles d'acquisition considérés ici. L'hypothèse 1 selon laquelle ε et ε' sont des bruits blancs est simpliste dans le cas de certaines modalités d'images. On sait par exemple que le scintillement[1] typique des images ultrasonores est spatialement corrélé. L'hypothèse de stationnarité 2 est tout aussi discutable au vu des différents types de biais spatiaux rencontrés dans les images médicales : biais du champ magnétique en IRM, effets d'ombrage en imagerie ultrasonore, etc. Enfin, l'hypothèse 3 ignore le fait que certains artefacts sont corrélés à l'anatomie. C'est typiquement le cas des effets de volume partiel.

Notre modèle général apparaît donc comme très simple, pour ne pas dire simpliste. Nous verrons cependant que la plupart des mesures de similarité utilisées pour le recalage d'images reposent implicitement sur des hypothèses de ce type.

5.2.2 Modèle anatomique

Lorsque la scène S est inconnue, ce qui est généralement le cas, l'équation (5.3) n'exprime pas une relation directe entre les images I et J. Pour obtenir une telle relation, il faut intégrer $P(I,J|S)$ sur l'ensemble des réalisations possibles de S :

$$P(I,J) = \int P(I,J|S)\, P(S)\, dS, \tag{5.4}$$

ce qui suppose bien évidemment d'avoir défini $P(S)$, la loi de probabilité a priori de la scène.

A ce stade, deux remarques s'imposent. D'une part, le rôle de S est de partionner l'espace en différentes classes anatomiques. Les valeurs des labels n'ont aucune importance en tant que telles et leur choix est arbitraire. Nous adopterons la convention d'assimiler les labels aux indices de classes, de sorte que S prend ses valeurs dans $\{1,2,\ldots,K\}$, K étant le nombre de classes.

L'autre remarque concerne l'ensemble de définition de S. Une difficulté survient du fait de l'existence d'une transformation spatiale T reliant S et l'image source I. Pour qu'une expression telle que l'équation (5.1) ait toujours un sens, il faut que pour toute transformation T et pour tout voxel $\mathbf{x}_k \in \Omega_I$, le point transformé $T(\mathbf{x}_k)$ appartienne à l'ensemble de définition de S. La seule manière de remplir cette condition est de définir S sur l'espace entier $\mathbb{R}^3$. Cette approche est sans doute envisageable mais elle nous est apparue techniquement trop complexe. Elle impose de manipuler des champs aléatoires à variables continues et à valeurs discrètes, étranges objets mathématiques qui nous semblent avoir été peu étudiés.

C'est pourquoi nous avons opté pour une description spatialement discrète. Puisque la scène est supposée alignée avec l'image cible J, il serait naturel de choisir son ensemble de définition comme étant la grille Ω_J. Cependant, rien n'interdit à la transformation spatiale de diriger certains voxels $\mathbf{x}_k$ en dehors du domaine délimité par Ω_J. Nous voudrions donc pouvoir considérer des positions extérieures à la grille. Dans ce but, nous définissons S comme une fonction de la grille $\tilde{\Omega}_J$ obtenue par prolongement infini de Ω_J :

$$S : \tilde{\Omega}_J \subset \mathbb{R}^3 \to \{1,2,\ldots,K\}$$

1. Speckle.

La contre-partie de l'approche discrète est qu'elle requiert une hypothèse concernant la transformation spatiale : T doit faire parfaitement coïncider les voxels de Ω_I avec des voxels de $\tilde{\Omega}_J$, ce qui implique bien évidemment une restriction de l'espace de recherche. La grille $\tilde{\Omega}_J$ étant le prolongement de Ω_J, ceci revient à discrétiser les déplacements à la résolution spatiale de l'image cible. Nous envisageons au paragraphe 5.5 des techniques d'interpolation permettant de s'affranchir de cette limitation.

La scène étant définie sur un espace discret, il est légitime de lui attribuer la loi de probabilité d'un champ de Gibbs-Markov (Li, 1995):

$$P(S) = \frac{1}{Z} e^{-U(S)}$$

La constante de normalisation Z, parfois appelée fonction de partition, est définie de manière à ce que la somme de $P(S)$ soit unitaire sur l'ensemble des configurations possibles de S. Cette contrainte se traduit par la relation :

$$Z = \int_{\tilde{\Omega}_J} e^{-U(S)} \, dS$$

L'énergie du champ $U(S)$ fait intervenir un système de voisinage de la grille $\tilde{\Omega}_J$. Dans ce contexte, les voxels sont encore appelés des sites. Pour tout entier $q > 0$, on appelle clique d'ordre q un q-uplet de sites voisins deux à deux. Leur ensemble est noté $\mathcal{C}_q$ et on a notamment $\mathcal{C}_1 = \tilde{\Omega}_J$. Lorsque le champ est supposé stationnaire (homogène), son énergie se décompose en une somme d'énergies attachées aux différentes cliques et ne dépendant que des positions *relatives* des sites :

$$U(S) = \sum_{\ell \in \mathcal{C}_1} V_1(s_\ell) + \sum_{(\ell,\ell') \in \mathcal{C}_2} V_2(s_\ell, s_{\ell'}) + \sum_{(\ell,\ell',\ell'') \in \mathcal{C}_3} V_3(s_\ell, s_{\ell'}, s_{\ell''}) + \dots$$

Si on se limite aux cliques d'ordre 1, alors $P(S)$ s'exprime comme un produit (infini) de probabilités ponctuelles et on peut choisir $Z = 1$:

$$P(S) = \prod_\ell P(s_\ell) \qquad \text{avec} \quad P(s_\ell) = e^{-V_1(s_\ell)}$$

Autrement dit, les classes anatomiques attribuées aux différents voxels sont considérées comme mutuellement indépendantes. La prise en compte de cliques d'ordre supérieur permet d'introduire des couplages entre les voxels voisins et, par ce biais, de discriminer les configurations en fonction du degré de continuité spatiale qu'elles manifestent.

5.2.3 Relation directe entre les images

Pour obtenir l'expression de la loi conjointe des images $P(I,J)$, nous substituons les expressions de $P(I,J|S)$ et de $P(S)$ dans l'équation (5.4). Le fait que S admette un ensemble dénombrable de configurations justifie de faire disparaître l'intégrale au profit d'une somme :

$$P(I,J) = \sum_S \left[e^{-U(S)} \prod_{\mathbf{x}_k \in \Omega_I} \phi\left(i_k - f(s_k^\downarrow)\right) \prod_{\mathbf{y}_\ell \in \Omega_J} \psi\left(j_\ell - g(s_\ell)\right) \right] \tag{5.5}$$

Il semble impossible de simplifier une telle équation à moins de supposer que la scène soit spatialement non-corrélée, c'est-à-dire que seules les cliques d'ordre 1 interviennent dans l'expression de $U(S)$. Il sera commode d'admettre en outre que T est injective : les correspondants de deux voxels distincts sont eux-mêmes distincts. Nous introduisons l'ensemble $\mathcal{A}$ des voxels de $\tilde{\Omega}_J$ qui admettent un antécédent par T, c'est-à-dire :

$$\mathcal{A} = \left\{ \mathbf{y}_\ell \in \tilde{\Omega}_J,\, \exists \mathbf{x}_k \in \Omega_I,\, T(\mathbf{x}_k) = \mathbf{y}_\ell \right\}$$

L'intersection de $\mathcal{A}$ et de Ω_J représente la zone de recouvrement entre les images. Sous réserve d'existence, nous désignerons par la notation $k:\ell$ l'indice de l'antécédent du voxel $\mathbf{y}_\ell$ dans la grille Ω_I. Cet antécédent est forcément unique du fait de l'injectivité présumée de T.

Sous ces hypothèses, on peut regrouper les différents termes de l'équation (5.5) en fonction du voxel $\mathbf{y}_\ell \in \tilde{\Omega}_J$ auquel ils correspondent. Ceci permet d'écrire le terme général de la somme comme un produit de fonctions indépendantes, et donc d'inverser les signes produit et somme :

$$P(I,J) = \sum_S \left[\prod_{\mathbf{y}_\ell \in \tilde{\Omega}_J} \alpha_\ell(s_\ell) \right] = \prod_{\mathbf{y}_\ell \in \tilde{\Omega}_J} \left[\sum_{s_\ell = 1}^{K} \alpha_\ell(s_\ell) \right], \tag{5.6}$$

avec, pour tout $\mathbf{y}_\ell \in \tilde{\Omega}_J$:

$$\alpha_\ell(s_\ell) \equiv e^{-V_1(s_\ell)} \times \begin{cases} 1 & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\ \psi\left(j_\ell - g(s_\ell)\right) & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\ \phi\left(i_{k:\ell} - f(s_\ell)\right) & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \\ \phi\left(i_{k:\ell} - f(s_\ell)\right)\, \psi\left(j_\ell - g(s_\ell)\right) & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \end{cases}$$

Les sommes intervenant dans l'équation (5.6) dépendent donc de l'information d'intensité disponible au point $\mathbf{y}_\ell$: soit il n'y en a pas, soit une seule des images en apporte, soit les deux images en apportent (voir la figure 5.2). La quantité $\alpha_\ell(s_\ell)$ s'interprète comme la probabilité a posteriori d'observer simultanément au point $\mathbf{y}_\ell$, la classe s_ℓ et les intensités présentes. Le fait de sommer $\alpha_\ell(s_\ell)$ sur l'ensemble des classes possibles traduit le fait que s_ℓ est inconnue.

Afin d'alléger les notations, posons : $g_p \equiv g(p)$, $f_p \equiv f(p)$ et π_p, la probabilité a priori que le voxel $\mathbf{y}_\ell$ appartienne à la classe p (cette probabilité est indépendante de la position du fait de l'hypothèse de stationnarité, voir le paragraphe 5.2.2) :

$$\pi_p = P(s_\ell = p) = e^{-V_1(p)}$$

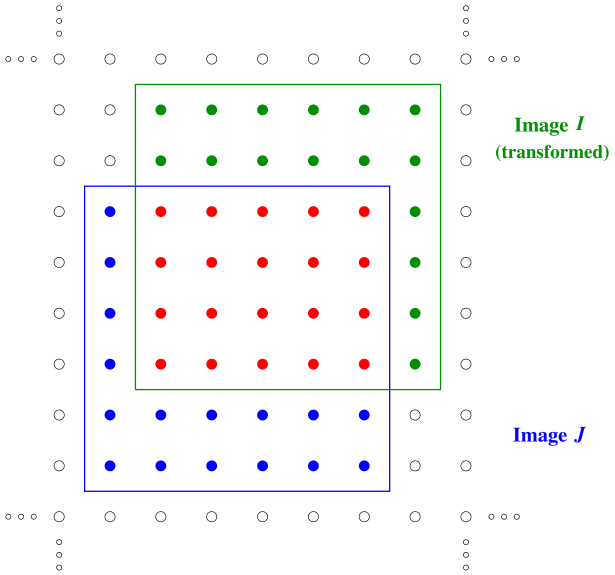


FIG. 5.2 – *Effet de recouvrement partiel entre les images. Différents sous-ensembles de la grille infinie $\tilde{\Omega}_J$ sont à considérer : $\Omega_J - \mathcal{A}$ (en bleu), $\mathcal{A} \cap \Omega_J$ (en rouge) et $\mathcal{A} - \Omega_J$ (en vert).*

Avec ces notations, on obtient selon les cas :

$$\sum_{s_\ell=1}^{K} \alpha_\ell(s_\ell) = \begin{cases} 1 & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\ \sum_{p=1}^{K} \pi_p \psi\left(j_\ell - g_p\right) & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\ \sum_{p=1}^{K} \pi_p \phi\left(i_{k:\ell} - f_p\right) & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \\ \sum_{p=1}^{K} \pi_p \phi\left(i_{k:\ell} - f_p\right) \psi\left(j_\ell - g_p\right) & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \end{cases}$$

En reportant ces quatre expressions possibles dans l'équation (5.6), nous observons que les points sans intensité ($\mathbf{y}_\ell \notin \Omega_J$ et $\mathbf{y}_\ell \notin \mathcal{A}$) n'ont aucune influence sur $P(I,J)$ car ils contribuent à hauteur d'une multiplication par 1. La loi conjointe $P(I,J)$ se présente finalement sous la forme d'un produit fini :

$$P(I,J) = \prod_{\Omega_J - \mathcal{A}} P(j_\ell) \times \prod_{\mathcal{A} \cap \Omega_J} P(i_{k:\ell}, j_\ell) \times \prod_{\mathcal{A} - \Omega_J} P(i_{k:\ell}), \tag{5.7}$$

où $P(j_\ell)$, $P(i_{k:\ell})$ et $P(i_{k:\ell}, j_\ell)$ sont d'authentiques lois de probabilité. Leurs formes paramétriques respectives sont données ci-dessous :

$$\left\{ \begin{array}{lcl} P(i) & = & \displaystyle\sum_{p=1}^{K} \pi_p \phi\,(i - f_p) \\ P(j) & = & \displaystyle\sum_{p=1}^{K} \pi_p \psi\,(j - g_p) \\ P(i,j) & = & \displaystyle\sum_{p=1}^{K} \pi_p \phi\,(i - f_p)\,\psi\,(j - g_p)\,, \end{array} \right. \tag{5.8}$$

et on vérifie sans peine les relations :

$$P(i) = \int_{\mathbb{R}} P(i,j)\,dj, \qquad P(j) = \int_{\mathbb{R}} P(i,j)\,di$$

Par la suite, il nous sera plus utile d'exprimer $P(I,J)$ sous une forme qui montre explicitement qu'il s'agit d'une fonction de T. A cet effet, nous introduisons les ensembles suivants :

- l'ensemble $\Omega_I^{\downarrow}$ des voxels de Ω_I ayant un correspondant dans Ω_J (correspondant à $\mathcal{A} \cap \Omega_J$)
- l'ensemble $\bar{\Omega}_I^{\downarrow}$ des voxels de Ω_I dont les correspondants sont situés en dehors de Ω_J ; c'est bien entendu le complémentaire de $\Omega_I^{\downarrow}$ dans Ω_I (correspondant à $\mathcal{A} \setminus \Omega_J$),
- l'ensemble $\bar{\Omega}_J^{\downarrow}$ des voxels de Ω_J n'ayant pas d'antécédent dans Ω_I (correspondant à $\Omega_J \setminus \mathcal{A}$).

L'utilisation de flèches orientées vers le bas est destinée à marquer la dépendance en T. Nous pouvons réécrire l'équation (5.7) à l'aide de ces ensembles, ce qui revient en somme à un changement de repère :

$$P(I,J) = \prod_{\mathbf{y}_\ell \in \bar{\Omega}_J^{\downarrow}} P(j_\ell) \;\times\; \prod_{\mathbf{x}_k \in \Omega_I^{\downarrow}} P(i_k, j_k^{\downarrow}) \;\times\; \prod_{\mathbf{x}_k \in \bar{\Omega}_I^{\downarrow}} P(i_k) \tag{5.9}$$

Cette expression possède une forme symétrique qui n'a rien de surprenant dans la mesure où notre formulation ne fait pas jouer de rôle préférentiel à l'une des images, en dehors du fait que la transformation T opère de I vers J. Ce défaut apparent de symétrie est compensé par l'hypothèse d'injectivité de T. Il importe également de remarquer que tous les voxels de I et de J contribuent à la loi de probabilité conjointe des images. Ceux qui n'ont pas de correspondant sont pris en compte par l'intermédiaire de leur seule probabilité marginale, c'est-à-dire $P(i_k)$ ou $P(j_\ell)$.

Un autre corollaire remarquable de l'équation (5.9) est que les images sont statistiquement indépendantes si leur zone de recouvrement est vide. Autrement dit, si deux voxels $\mathbf{x}_k$ et $\mathbf{y}_\ell$ ne se correspondent pas géométriquement, leurs intensités sont indépendantes. Ce résultat découle en fait de l'hypothèse de non-corrélation spatiale de la scène et du bruit.

Nous voyons là apparaître un premier lien avec les approches présentées dans les chapitres précédents. Finalement, la distribution ponctuelle $P(i,j)$ n'est rien d'autre qu'un modèle paramétrique de l'histogramme conjoint des images. En outre, l'équation (5.9) montre que la relation entre les images dépend uniquement de la connaissance de $P(i,j)$, résultat obtenu sous l'hypothèse que la scène S est spatialement non-corrélée. Cette hypothèse apparaît donc sous-jacente aux approches iconiques usuelles, dans lesquelles la relation entre les intensités est caractérisée par l'histogramme conjoint.

Lorsque l'énergie $U(S)$ fait intervenir des cliques d'ordre supérieur à 1, la forme factorisée (5.9) n'est plus valable. Des termes de couplages apparaissent entre des voxels qui ne sont pas appariés. Autrement dit, la relation entre les images n'est plus décrite par une simple comparaison d'intensités point à point. Cette dernière remarque doit être considérée comme prospective. Dans les paragraphes qui suivent, il devrait apparaître clairement que la modélisation de la scène par un champ de Markov non-trivial est difficilement abordable dans l'état actuel de la puissance calculatoire d'un ordinateur conventionnel. C'est pourquoi nous nous focalisons désormais sur le cas où la scène est supposée spatialement non-corrélée, tout en gardant à l'esprit la possible généralisation de cette hypothèse.

En exploitant la relation $P(i_k,j_k^\downarrow) = P(i_k|j_k^\downarrow)P(j_k^\downarrow)$, nous obtenons une expression équivalente de $P(I,J)$ dont l'intérêt est de faire apparaître la loi conditionnelle $P(I|J)$:

$$P(I,J) = \underbrace{\prod_{\mathbf{y}_\ell \in \Omega_J} P(j_\ell)}_{P(J)} \times \underbrace{\prod_{\mathbf{x}_k \in \Omega_I^\downarrow} P(i_k|j_k^\downarrow) \times \prod_{\mathbf{x}_k \in \bar{\Omega}_I^\downarrow} P(i_k)}_{P(I|J)} \tag{5.10}$$

Remarquons que l'ensemble $\bar{\Omega}_J^\downarrow$ n'intervient pas dans cette dernière expression et que $P(I,J)$ ne dépend de T que par l'intermédiaire de la probabilité conditionnelle $P(I|J)$. Il s'avère que $P(I|J)$ admet une forme factorisée similaire à $P(I|S)$; voir l'équation (5.3). Ce serait exactement la même forme si la grille transformée $T(\Omega_I)$ était intégralement contenue dans l'enveloppe convexe de Ω_J. D'une certaine manière, l'image J peut donc être considérée comme une scène anatomique au même titre que S. Il faut cependant réaliser que les expressions des densités conditionnelles $P(i_k|s_k^\downarrow)$ et $P(i_k|j_k^\downarrow)$ sont différentes. A la moyenne près, les premières s'identifient à la distribution du bruit :

$$P(i_k|s_k^\downarrow) = \phi\left(i_k - f(s_k^\downarrow)\right)$$

Il est raisonnable de choisir ϕ comme une distribution unimodale et alors, les distributions conditionnelles $P(i|s)$ sont elles-mêmes unimodales (pour tout s). En revanche, les distributions $P(i_k|j_k^\downarrow)$ sont des distributions de mélange comme le montre l'équation (5.8). Elles comportent a priori autant de modes qu'il existe de classes anatomiques :

$$P(i_k|j_k^\downarrow) = \sum_{p=1}^{K} \pi'_p(j_k^\downarrow)\, \phi\,(i_k - f_p) \qquad \text{avec} \quad \pi'_p(j_k^\downarrow) \equiv \frac{\pi_p \psi\left(j_k^\downarrow - g_p\right)}{\sum_{q=1}^{K} \pi_q \psi\left(j_k^\downarrow - g_q\right)}$$

5.3 Principe d'identification des paramètres

Étant données deux images, le problème qui se pose maintenant est d'estimer les paramètres du modèle de dépendance, à savoir :

1. Les paramètres d'acquisition. Il s'agit des fonctions d'intensité f et g donc des valeurs $(f_1,f_2,\ldots,f_K)$ et $(g_1,g_2,\ldots,g_K)$, mais aussi des paramètres des distributions ϕ et ψ liées

au bruit : typiquement, ce sont des paramètres d'échelle σ_ϕ et σ_ψ (qui s'interprètent comme des écart-types si ϕ et ψ sont gaussiennes).

2. Les probabilités a priori d'appartenance aux classes anatomiques : $\pi_1,\pi_2,\ldots,\pi_K$. Lorsque on suppose que la scène est spatialement non-corrélée, ces probabilités suffisent à caractériser la probabilité globale $P(S)$.
3. Le nombre de classes K. Ce paramètre joue un rôle particulier dans la mesure où il détermine la dimension de l'espace de recherche.
4. Et, bien entendu, les paramètres qui nous intéressent le plus dans le contexte d'un problème de recalage : ceux de la transformation spatiale T reliant les grilles Ω_I et Ω_J.

Afin d'alléger les notations, nous regroupons tous les paramètres autres que géométriques dans un même vecteur θ :

$$\theta = (K,f_1,\ldots,f_K,\sigma_\phi,g_1,\ldots,g_K,\sigma_\psi,\pi_1,\ldots,\pi_K)\,,$$

notation qui n'est d'ailleurs pas très rigoureuse si le nombre de classes K est un paramètre libre.

D'un point de vue subjectif, les paramètres θ peuvent être considérés comme auxiliaires étant entendu que c'est l'estimation de T qui nous intéresse avant tout. En réalité, ce sont des paramètres «comme les autres» et il n'y a pas de différence conceptuelle à faire entre θ et T en dehors de la nature des phénomènes qu'ils caractérisent.

En pratique, il est possible voire souhaitable de fixer certains paramètres auxiliaires dans la mesure où on est capable de les estimer sur la base d'observations courantes ou au moyen de méthodes indépendantes. Par souci de clarté, nous supposerons dans ce paragraphe que tous les paramètres sont libres. Nous envisagerons au paragraphe 5.4 différents cas particuliers pour lesquels certains paramètres sont fixés.

5.3.1 Estimateur du maximum de vraisemblance

Comme le montrent les équations (5.8) et (5.9), la loi de probabilité conjointe $P(I,J)$ dépend de (T,θ). Ici, nous cherchons à estimer les paramètres ; dans ce contexte, nous utiliserons la notation usuelle mais abusive : $P(I,J|T,\theta)$.

Notre problème se situe typiquement dans le cadre de l'inférence statistique : il s'agit de déterminer les valeurs des paramètres (T,θ) pour lesquelles le modèle est en adéquation maximale avec les données empiriques, à savoir les images I et J. Pour ce faire, il faut définir un estimateur de (T,θ). L'estimateur le plus courant dans la littérature statistique est celui du maximum de vraisemblance (MV). Il est fondé sur le principe intuitif que les paramètres les plus plausibles sont ceux pour lesquels la probabilité des données est maximale :

$$(\hat{T},\hat{\theta}) = \arg\max_{T,\theta}\; P(I,J|T,\theta)$$

Le principe du recalage d'images par maximum de vraisemblance est déjà une idée ancienne. Les premiers travaux s'attachaient à des problèmes de recalage monomodal pour lesquels un modèle de dépendance inter-images relativement simple suffisait (McGillem et Svedlow, 1976; Mort et Srinath, 1988). De ce fait, il n'y avait pas de paramètre auxiliaire du type de θ à estimer. Plus récemment, certains auteurs ont appliqué le principe du maximum de vraisemblance à des problèmes de recalage multimodal (Costa et al., 1993; Viola, 1995; Leventon et Grimson, 1998; Bansal et al., 1998). Les modèles de dépendance qu'ils proposent sont similaires au nôtre et leur mise en œuvre fait donc intervenir de manière plus ou moins cachée l'estimation de paramètres auxiliaires jouant un rôle analogue à θ.

Curieusement, en dehors de (Leventon et Grimson, 1998), aucun de ces auteurs ne réalise cette estimation auxiliaire par maximum de vraisemblance. Notons cependant que dans ce travail, les paramètres auxiliaires sont appris une fois pour toute à partir d'un premier couple d'images déjà recalées. Dans l'algorithme de recalage proprement dit, ils sont considérés comme connus et donc fixés. Pour leur part, (Costa et al., 1993) réalisent l'estimation auxiliaire par une méthode ad-hoc, bien que naturelle compte tenu du modèle qu'ils utilisent. L'approche retenue par (Viola, 1995) est non-paramétrique puisque elle consiste à estimer les distributions ponctuelles $P(i_k, j_k^{\downarrow})$ par fenêtrage de Parzen. Enfin, (Bansal et al., 1998) combinent l'approche de Viola avec une méthode originale fondée sur le principe du maximum d'entropie.

La particularité de l'approche que nous considérons ici est de traiter l'estimation des paramètres auxiliaires exactement de la même manière que celle des paramètres géométriques puisque T et θ maximisent un seul et même critère.

5.3.2 Autres estimateurs

« Le vrai peut quelquefois n'être pas vraisemblable »
(Boileau, L'Art Poétique)

Nous ouvrons ici une petite parenthèse pour signaler que l'estimateur du maximum de vraisemblance n'est qu'un estimateur parmi d'autres possibles. C'est sans doute le plus naturel, mais ce n'est pas nécessairement le plus simple à mettre en œuvre. Dans certains cas pratiques, des estimateurs ad-hoc peuvent s'avérer plus robustes; le chapitre 7 en apportera l'illustration.

On peut aussi envisager des estimateurs bayésiens tels que le maximum de vraisemblance a posteriori (MAP) ou l'espérance conditionnelle. Ceux-ci offrent la possibilité de contraindre l'espace de recherche de certains paramètres[2]. Dans le cas du recalage non-rigide, il s'avère bien souvent indispensable de pénaliser les transformations spatiales à caractère «oscillatoire» (voir le chapitre 2, paragraphe 2.4.6). Les approches les plus courantes consistent à ajouter un potentiel de régularisation à la mesure de similarité, ce qui revient à définir un estimateur MAP. Dans le même contexte, (Christensen, 1994) a étudié l'espérance conditionnelle.

Enfin, une autre série d'alternatives apparaît lorsque la scène est considérée comme une variable à estimer et non comme une variable cachée comme c'est le cas dans notre formulation. On peut

2. Plus exactement, ils considèrent les paramètres comme des variables aléatoires auxquelles on peut donc attribuer des lois de probabilité.

imaginer par exemple une estimation par maximum de vraisemblance a priori :

$$\max_{S,T,\theta} P(I,J|S,T,\theta)$$

ou, sans doute beaucoup mieux, par maximum a posteriori :

$$\max_{S,T,\theta} P(I,J,S|T,\theta) = \max_{S,T,\theta} P(I,J|S,T,\theta)P(S|\theta)$$

On notera que cette dernière stratégie permet d'incorporer une forme complexe pour $P(S)$ plus facilement que lorsque S est considérée comme une variable cachée (par exemple, une distribution de Gibbs-Markov). Ces approches se proposent ni plus ni moins de résoudre simultanément le problème du recalage et de la segmentation. En réalité, c'est également le cas lorsque S est considérée comme une variable cachée. Nous aurons l'occasion de revenir sur ce thème au paragraphe 5.4.1.

5.3.3 Log-vraisemblance

Pour des raisons purement calculatoires qui apparaîtront plus loin, nous allons manipuler la log-vraisemblance négative plutôt que la fonction de vraisemblance elle-même :

$$L(T,\theta) = -\log P(I,J|T,\theta)$$

Le changement de signe est arbitraire : il permet simplement de se ramener à la notion familière d'énergie puisque $L(T,\theta)$ est à minimiser. L'intérêt de passer au logarithme est de convertir le produit qui apparaît dans l'équation (5.9) en une somme :

$$L(T,\theta) = -\sum_{\mathbf{x}_k \in \Omega_I^{\downarrow}} \log P(i_k,j_k^{\downarrow}|\theta) \; - \sum_{\mathbf{x}_k \in \bar{\Omega}_I^{\downarrow}} \log P(i_k|\theta) \; - \sum_{\mathbf{y}_\ell \in \bar{\Omega}_J^{\downarrow}} \log P(j_\ell|\theta), \tag{5.11}$$

les formes paramétriques de $P(i_k,j_k^{\downarrow}|\theta)$, $P(i_k|\theta)$ et $P(j_\ell|\theta)$ étant fournies par l'équation (5.8). D'après l'équation (5.10), la log-vraisemblance négative s'écrit aussi :

$$\begin{aligned} L(T,\theta) &= -\sum_{\mathbf{x}_k \in \Omega_I^{\downarrow}} \log P(i_k|j_k^{\downarrow},\theta) \; - \sum_{\mathbf{y}_\ell \in \Omega_J} \log P(j_\ell|\theta) \; - \sum_{\mathbf{x}_k \in \bar{\Omega}_I^{\downarrow}} \log P(i_k|\theta) \\ &= -\sum_{\mathbf{x}_k \in \Omega_I^{\downarrow}} \log \frac{P(i_k|j_k^{\downarrow},\theta)}{P(i_k|\theta)} \; - \sum_{\mathbf{y}_\ell \in \Omega_J} \log P(j_\ell|\theta) \; - \sum_{\mathbf{x}_k \in \Omega_I} \log P(i_k|\theta) \end{aligned} \tag{5.12}$$

Dans l'équation (5.12), on remarque que $L(T,\theta)$ ne dépend de T que via le terme de gauche, c'est-à-dire le terme impliquant le rapport de probabilités $P(i_k|j_k^{\downarrow},\theta)/P(i_k|\theta)$. En revanche, les trois termes dépendent de θ et si on comprend bien que l'estimation de T dépend a priori de celle de θ, il doit être clair que chaque terme influence le recalage.

5.3.4 Information mutuelle empirique

Le critère de log-vraisemblance que nous obtenons est étroitement lié à la notion d'information mutuelle empirique popularisée par (Viola, 1995). Chacun des trois termes de l'équation (5.12) peut être interprété comme une entropie empirique à un facteur multiplicatif près. De gauche à droite, il s'agit respectivement de l'entropie relative des images (information mutuelle) calculée dans la zone de recouvrement et des entropies individuelles de I et J. En définissant $n^\downarrow \equiv \mathrm{Card}\Omega_I^\downarrow$, $n \equiv \mathrm{Card}\Omega_I$ et $m \equiv \mathrm{Card}\Omega_J$, nous pouvons préciser l'analogie :

$$L(T,\theta) = \underbrace{-\sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \log \frac{P(i_k|j_k^\downarrow,\theta)}{P(i_k|\theta)}}_{-n^\downarrow \mathcal{I}_{I,J}(T,\theta)} \underbrace{-\sum_{\mathbf{x}_k \in \Omega_I} \log P(i_k|\theta)}_{nh_I(\theta)} \underbrace{-\sum_{\mathbf{y}_\ell \in \Omega_J} \log P(j_\ell|\theta)}_{mh_J(\theta)} \tag{5.13}$$

Si les paramètres auxiliaires θ étaient connus, de sorte que nous n'aurions pas à les estimer, alors maximiser la vraisemblance en fonction de T (à θ fixé) reviendrait à maximiser $n^\downarrow$ fois l'information mutuelle, $n^\downarrow$ caractérisant la taille de la zone de recouvrement des images. Mis à part le fait qu'il apparaît ici une pondération par $n^\downarrow$ qui n'était pas prise en compte par Viola, notre formulation s'approche de la sienne.

La différence fondamentale réside dans le fait que les paramètres auxiliaires sont a priori inconnus. Puisque nous cherchons à les estimer par maximum de vraisemblance, nous optimisons un critère qui dépend de (T,θ). La formulation de Viola ne fait pas explicitement référence à des paramètres auxiliaires et, de ce fait, il optimise un critère qui dépend seulement de T. L'astuce de son implémentation consiste à utiliser la méthode des fenêtres de Parzen pour mettre à jour, en fonction de T, la distribution conjointe $P(i,j)$ nécessaire au calcul de l'information mutuelle. On conçoit que cette stratégie joue un rôle analogue à l'estimation des paramètres auxiliaires.

Nous montrerons au paragraphe 5.4.3 qu'il s'agit plus que d'une analogie : lorsque le nombre de classes K est considéré comme un paramètre libre, la distribution de Parzen est approximativement celle qui maximise la vraisemblance à T fixé. La méthode de Viola apparaîtra alors, à quelques approximations près, comme un cas particulier d'estimation par maximum de vraisemblance.

Le parallèle avec l'information mutuelle n'est pas tout à fait le fruit du hasard. Viola justifie en partie le recalage par information mutuelle au moyen d'une formulation par maximum de vraisemblance relativement proche de la nôtre, mais à deux nuances près : d'une part, elle ne fait pas appel à un modèle paramétrique de dépendance inter-images ; d'autre part, elle ne tient pas compte des effets de recouvrement partiel. A cause de ce deuxième aspect, Viola aboutit à un critère de recalage qui s'interprète comme une entropie conditionnelle et non comme une information mutuelle. C'est ensuite sur la base d'arguments heuristiques qu'il justifie l'emploi de l'information mutuelle au détriment de l'entropie conditionnelle.

La prise en compte du recouvrement partiel nous a permis d'établir un lien plus profond, traduit par l'équation (5.13), entre la log-vraisemblance et l'information mutuelle. Cependant, il ne s'agit pas d'une équivalence stricte, sans compter le fait que la notion d'information mutuelle est sujette à des acceptions tout aussi variées qu'équivoques.

5.3.5 Schéma d'optimisation

La minimisation de $L(T,\theta)$ en fonction de (T,θ) n'admet pas de solution explicite et nous serions tentés de la résoudre par n'importe quelle méthode numérique. Ce serait manquer cruellement d'optimisme, pour deux raisons. D'abord parce qu'il existe une méthode remarquablement adaptée à la minimisation de $L(T,\theta)$ en fonction des seuls paramètres θ (à T fixé) : c'est l'algorithme espérance/maximisation (EM) dû à (Dempster et al., 1977). Ensuite, parce que la minimisation à T fixée peut même, sous certaines conditions, être résolue analytiquement. Ces affirmations seront étayées au paragraphe 5.4.

Nous voyons ici ressurgir une problématique rencontrée au chapitre 3. Nous avions à minimiser un critère $\tilde{S}(T,f)$ dépendant d'une transformation spatiale T et d'une fonction d'intensité f, tout en sachant que la minimisation en f pouvait être résolue par une simple procédure de moindres carrés, éventuellement itérée (voir le paragraphe 3.5). Ceci nous avait conduit à proposer deux stratégies de minimisation, l'une fondée sur un schéma *hiérarchisé*, l'autre sur un schéma *alterné*. Rappelons que dans le schéma hiérarchisé, on forçait f à minimiser constamment le critère. Cette stratégie nous ramenait à la minimisation d'un critère dépendant uniquement de T :

$$\tilde{S}(T) = \min_f \tilde{S}(T,f)$$

Le schéma alterné consistait, sans surprise, à alterner la minimisation en f à T fixé avec la minimisation en T à f fixé. Dans les deux cas, l'optimisation en T était menée par une méthode purement numérique, en l'occurrence l'algorithme de Powell.

On voit que θ joue vis-à-vis de $L(T,\theta)$ un rôle similaire à celui que joue f vis-à-vis de $\tilde{S}(T,f)$. Nous pouvons donc envisager les deux mêmes schémas de résolution. Pour l'un comme pour l'autre, nous devons être capables de calculer la valeur de θ qui minimise $L(T,\theta)$ à T fixé. C'est le problème que nous abordons maintenant.

5.4 Estimation des paramètres auxiliaires

Notre but est de minimiser $L(T,\theta)$ en fonction de θ, la transformation T étant fixée. Nous noterons $\mathcal{L}(T)$ la log-vraisemblance négative minimale à T fixé ; dans le cadre d'un schéma hiérarchisé (voir le paragraphe 5.3.5), c'est le critère à minimiser en fonction de T :

$$\mathcal{L}(T) \equiv \min_\theta L(T,\theta),$$

ou, de manière équivalente :

$$\mathcal{L}(T) \equiv L(T,\hat{\theta}(T)), \qquad \text{avec} \quad \hat{\theta}(T) \equiv \arg\min_\theta \, L(T,\theta)$$

Cette dernière expression souligne le fait que les paramètres auxiliaires optimaux à T fixé sont fonction de T. C'est une notion qu'il faudra garder à l'esprit tout au long de ce paragraphe.

Nous allons considérer trois cas successifs.

1. Le nombre de classes K est fixé, tous les autres paramètres sont inconnus.

2. Idem 1, mais avec des hypothèses supplémentaires concernant l'image J : la fonction g est supposée injective et le niveau de bruit σ_ψ négligeable devant le contraste.
3. Les échelles σ_ϕ et σ_ψ sont fixées, tous les autres paramètres sont inconnus (notamment K).

Le paragraphe 5.4.1 montre que l'algorithme EM est une méthode de résolution adaptée au premier cas. Les cas 2 et 3, traités aux paragraphes 5.4.2 et 5.4.3, bénéficient d'une solution analytique approchée. Nous montrons en outre qu'ils sont étroitement liés à des mesures de similarité connues : respectivement, le rapport de corrélation et l'information mutuelle au sens de (Viola, 1995).

5.4.1 Algorithme EM

L'estimation de θ est typiquement ce qu'on appelle un problème à «données incomplètes». Ici, les données manquantes sont les labels anatomiques de la grille Ω_J, c'est-à-dire la scène S. Si S était connue, il serait possible d'estimer θ en maximisant la vraisemblance des données complètes :

$$\max_\theta P(I,J,S|T,\theta)$$

Lorsque les distributions ϕ et ψ sont convenablement choisies, ce problème admet une solution analytique. On aimerait donc se ramener au cas de données complètes ; c'est exactement le principe de l'algorithme EM (Dempster et al., 1977).

L'algorithme EM est une méthode générique pour résoudre des problèmes d'estimation par maximum de vraisemblance. Sa caractéristique principale est de combiner l'estimation des paramètres avec l'estimation de la loi de probabilité a posteriori des données manquantes. Il en découle deux étapes alternatives à itérer jusqu'à la convergence : l'étape E (espérance) et l'étape M (maximisation). Nous renvoyons le lecteur aux excellentes présentations de (Couvreur, 1996; Neal et Hinton, 1998) pour une justification théorique de cette terminologie.

Sous des hypothèses très générales, il est établi que l'algorithme EM converge vers un maximum local de la fonction de vraisemblance (Neal et Hinton, 1998). Toutefois, la convergence vers le maximum global n'est pas garantie et, de ce fait, l'initialisation des paramètres est un aspect crucial de la mise en œuvre. Nous ne connaissons pas de méthode automatique assurant une bonne initialisation. En pratique, les paramètres qui se révèlent critiques de ce point de vue sont les moyennes de classes f_p et g_p ; nous les initialisons au moyen d'une observation préalable des images. Nous choisissons par ailleurs les probabilités a priori d'appartenance aux classes de manière uniforme ($\forall p$, $\pi_p = 1/K$) et nous réglons arbitrairement les paramètres d'échelle σ_ϕ et σ_ψ à 10% de la dynamique des images.

5.4.1.1 Calcul itératif

Dans notre cas particulier, les étapes E et M sont instanciées comme suit.

Etape E. On dispose d'une estimée courante $\hat\theta$ que l'on considère comme fixée. Celle-ci nous permet d'estimer la loi de probabilité a posteriori des données manquantes :

$$\tilde P(S) = P(S|I,J,T,\hat\theta)$$

D'après la formule de Bayes, $P(S|I,J)$ est proportionnelle au produit des lois $P(I,J|S)$ et $P(S)$ en tant que fonction de S. Puisque chacune de ces deux lois admet une forme factorisée, c'est également le cas de $P(S|I,J)$:

$$P(S|I,J,T,\hat{\theta}) = \prod_{\mathbf{y}_\ell \in \tilde{\Omega}_J} P(s_\ell|I,J,\hat{\theta})$$

Bien entendu, ceci n'est vrai que sous l'hypothèse où la scène est spatialement non-corrélée. Dans le cas où la scène suit une véritable loi de Gibbs-Markov, on montre que sa loi conditionnelle $P(S|I,J)$ est également une loi de Gibbs-Markov dont l'expression analytique n'est cependant pas d'une grande utilité pratique. Dans son article d'introduction à l'algorithme ICM[3], (Besag, 1986) justifie une approximation de la loi conditionnelle par une forme factorisée du type précédent. Cette approximation porte le nom de pseudo-vraisemblance dans le contexte de la restauration d'images par maximum a posteriori (MAP).

Bien que cette remarque ouvre un certain nombre de perspectives intéressantes, nous nous en tiendrons ici à l'hypothèse d'indépendance spatiale. Dans ce cas, la probabilité a posteriori que le voxel $\mathbf{y}_\ell$ appartienne à la classe de label p :

$$\tilde{\pi}_{p\ell} \equiv P(s_\ell = p|I,J,\hat{\theta}),$$

ne dépend que de l'information d'intensité disponible au point $\mathbf{y}_\ell$. Comme au paragraphe 5.2.3, nous devons considérer l'ensemble $\mathcal{A}$ des voxels $\mathbf{y}_\ell \in \tilde{\Omega}_J$ ayant un antécédent dans Ω_I par T :

$$\tilde{\pi}_{p\ell} = \begin{cases} \hat{\pi}_p & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\[2ex] \dfrac{\hat{\pi}_p\, \psi(j_\ell - \hat{g}_p)}{\sum_q \hat{\pi}_q\, \psi(j_\ell - \hat{g}_q)} & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \notin \mathcal{A} \\ \dfrac{\hat{\pi}_p\, \phi(i_{k:\ell} - \hat{f}_p)}{\sum_q \hat{\pi}_q\, \phi(i_{k:\ell} - \hat{f}_q)} & \text{si} \quad \mathbf{y}_\ell \notin \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \\ \dfrac{\hat{\pi}_p\, \phi(i_{k:\ell} - \hat{f}_p)\, \psi(j_\ell - \hat{g}_p)}{\sum_q \hat{\pi}_q\, \phi(i_{k:\ell} - \hat{f}_q)\, \psi(j_\ell - \hat{g}_q)} & \text{si} \quad \mathbf{y}_\ell \in \Omega_J \quad \text{et} \quad \mathbf{y}_\ell \in \mathcal{A} \end{cases}$$

De cette manière, les probabilités d'appartenance aux classes combinent au mieux l'information apportée par les deux images. En l'absence d'information provenant de l'image I, seule l'image J intervient, et vice-versa. Les voxels qui ne présentent pas d'information d'intensité ($\mathbf{y}_\ell \notin \Omega_J$ et $\mathbf{y}_\ell \notin \mathcal{A}$) se voient attribuer une probabilité a posteriori égale à la probabilité a priori. Nous verrons cependant qu'ils n'interviennent pas dans l'étape M.

3. *Iterated Conditional Mode.*

Etape M. Les probabilités a posteriori $\tilde{\pi}_{p\ell}$ sont maintenant fixées, et ce sont les paramètres θ que l'on estime. L'idée est de maximiser l'espérance sur $\tilde{P}$ de la log-vraisemblance des données complètes (log-vraisemblance moyenne) :

$$E_{\tilde{P}}\left[\log P(I,J,S|T,\theta)\right]$$

La solution de ce problème dépend de la forme paramétrique des distributions ϕ et ψ. Un choix simple et partiellement justifié par le théorème de la limite centrale (Saporta, 1990; Papoulis, 1991) est de les supposer gaussiennes :

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma_\phi}\, e^{-\frac{x^2}{2\sigma_\phi^2}} \qquad \psi(x) = \frac{1}{\sqrt{2\pi}\sigma_\psi}\, e^{-\frac{x^2}{2\sigma_\psi^2}}$$

Dans ce cas, les formules d'estimation sont les suivantes (on les obtient par exemple en annulant le gradient de la log-vraisemblance moyenne) :

$$\left\{\begin{array}{lcl} \hat{f}_p & = & \sum_{\Omega_I} \tilde{\pi}^{\downarrow}_{pk}\, i_k \Big/ \sum_{\Omega_I} \tilde{\pi}^{\downarrow}_{pk} \\ \hat{g}_p & = & \sum_{\Omega_J} \tilde{\pi}_{p\ell}\, j_\ell \Big/ \sum_{\Omega_J} \tilde{\pi}_{p\ell} \\ \hat{\sigma}_\phi^2 & = & \frac{1}{n} \sum_p \sum_{\Omega_I} \tilde{\pi}^{\downarrow}_{pk} \left(i_k - \hat{f}_p\right)^2 \\ \hat{\sigma}_\psi^2 & = & \frac{1}{m} \sum_p \sum_{\Omega_J} \tilde{\pi}_{p\ell} \left(j_\ell - \hat{g}_p\right)^2 \\ \hat{\pi}_p & = & \frac{1}{m'} \sum_{\Omega_J \cup \mathcal{A}} \tilde{\pi}_{p\ell} \end{array}\right. \tag{5.14}$$

où $m' \equiv \mathrm{Card}(\Omega_J \cup \mathcal{A})$ et la notation $\tilde{\pi}^{\downarrow}_{pk}$ désigne la probabilité a posteriori que le correspondant $T(\mathbf{x}_k)$ du voxel $\mathbf{x}_k$ appartienne à la classe de label p : $\tilde{\pi}^{\downarrow}_{pk} = \tilde{\pi}_{p\ell}$ si $\mathbf{x}_k$ est l'antécédent de $\mathbf{y}_\ell \in \tilde{\Omega}_J$.

On obtient des formules d'estimation du même type dans le cas plus général où ϕ et ψ sont des lois exponentielles (Titterington et al., 1985). Notons également que rien n'empêche de supposer que ϕ et ψ ont des formes paramétriques différentes.

5.4.1.2 Collaboration recalage-segmentation

Au terme du processus complet de recalage (estimation de T et de θ), les cartes de probabilité a posteriori peuvent servir à réaliser une segmentation bi-canaux des images. Il suffit d'attribuer à chaque voxel $\mathbf{y}_\ell$ le label qui réalise le maximum de $\tilde{\pi}_{p\ell}$; c'est le principe d'une classification bayésienne. Les figures 5.3 et 5.4 illustrent ce propos dans le cadre du recalage rigide multimodal d'images du cerveau : elles correspondent respectivement à un recalage IRM/scanner (avec $K = 5$) et à un recalage d'images IRM issues des protocoles T2 et FLAIR (avec $K = 4$).

Dans les deux cas, nous avons mis en œuvre un schéma alterné (voir le paragraphe 5.3.5) de façon à limiter le nombre d'appels à l'algorithme EM. Les cartes de probabilité finales permettent de segmenter grossièrement le fond, la matière blanche, la matière grise et le liquide céphalo-rachidien. Dans le cas de la figure 5.3, nous parvenons également à isoler la boîte crânienne, très distincte dans le scanner. On remarque que les structures minoritaires sont mal étiquetées ; en particulier, celles qui forment le scalp. Même en élevant le nombre de classes, il est impossible de les segmenter correctement. Cela est dû au fait que leurs plages d'intensités ne sont pas suffisamment spécifiques pour que l'on puisse les différencier des structures majoritaires. Pour y parvenir, il faudrait sans doute imposer des contraintes spatiales.

L'algorithme EM fait aujourd'hui partie des méthodes standard pour la segmentation statique d'images IRM (Wells et al., 1996a; Guillemaud et Brady, 1997; Rajapakse et al., 1998; Van Leemput et al., 1999; Kapur, 1999; Zhang et al., 2000; Prima, 2000). Les modèles utilisés dans ces travaux sont plus spécifiques et plus complexes que le modèle général que nous avons considéré dans ce chapitre. Ils font généralement intervenir un biais spatial destiné à prendre en compte les hétérogénéités du champ magnétique. Dans (Kapur, 1999; Zhang et al., 2000), la scène est modélisée comme un «vrai» champ de Markov, c'est-à-dire un champ spatialement corrélé. Soulignons qu'il n'y a pas d'obstacle théorique à envisager de tels modèles dans le contexte du recalage par maximum de vraisemblance. Contrairement à un préjugé courant, l'algorithme EM ne repose nullement sur l'hypothèse d'indépendance statistique des données.

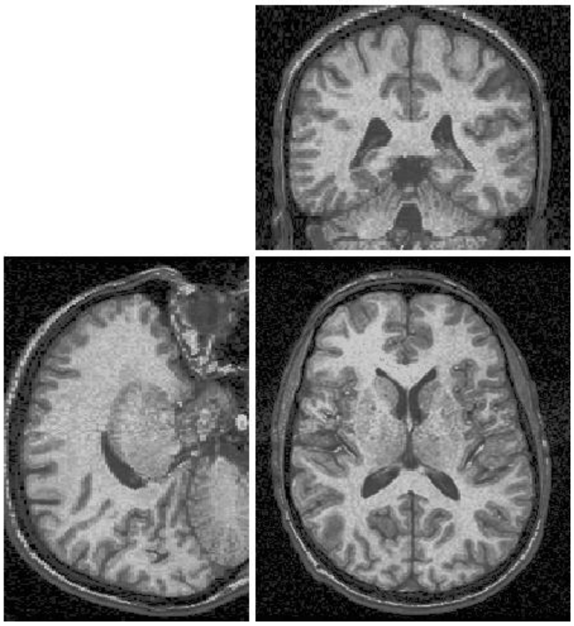
Dans notre implémentation, nous nous sommes cependant limités à un modèle sans corrélation spatiale. Même dans ce cadre, l'algorithme EM implique des temps de calcul considérablement plus longs que les méthodes de recalage fondées sur des mesures de similarité classiques (information mutuelle, rapport de corrélation, etc.). Nous espérons améliorer prochainement cet aspect en utilisant des variantes rapides de l'algorithme EM (Celeux et Govaert, 1991; Neal et Hinton, 1998). Du fait de cette lourdeur algorithmique, nos expériences de validation sont restées à un stade préliminaire. Une étude plus poussée est nécessaire pour évaluer la précision et la robustesse de la méthode.

L'utilisation de l'algorithme EM correspond à la situation où le modèle est contraint par un nombre limité de classes anatomiques. Nous allons maintenant envisager deux autres cas particuliers pour lesquels les contraintes sont respectivement plus fortes (paragraphe 5.4.2) et moins fortes (paragraphe 5.4.3). Leur intérêt est double. Il est algorithmique, d'une part, car nous verrons que la recherche des paramètres θ les plus vraisemblables bénéficie alors d'approximations analytiques raisonnables, ce qui rend l'algorithme EM superflu. D'autre part, ces cas font apparaître un lien théorique avec deux mesures de similarité que nous venons de mentionner : respectivement, le rapport de corrélation et l'information mutuelle.

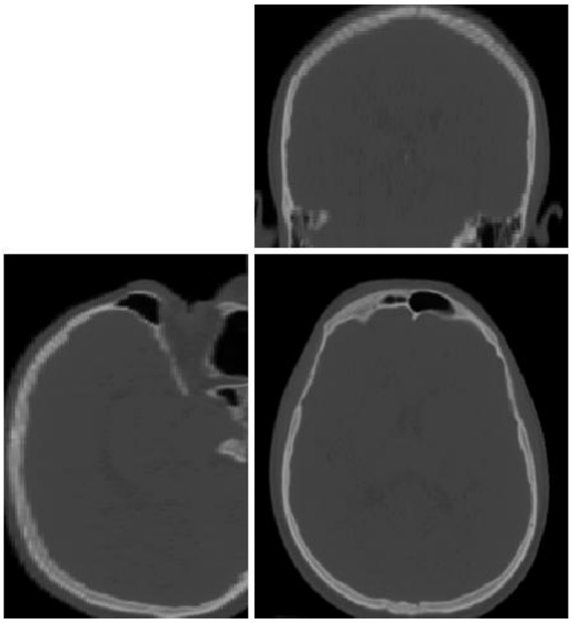
5.4.2 Hypothèse de redondance

Comme au paragraphe 5.4.1, nous supposons ici que tous les paramètres autres que le nombre de classes sont inconnus, mais en ajoutant deux hypothèses simplificatrices :

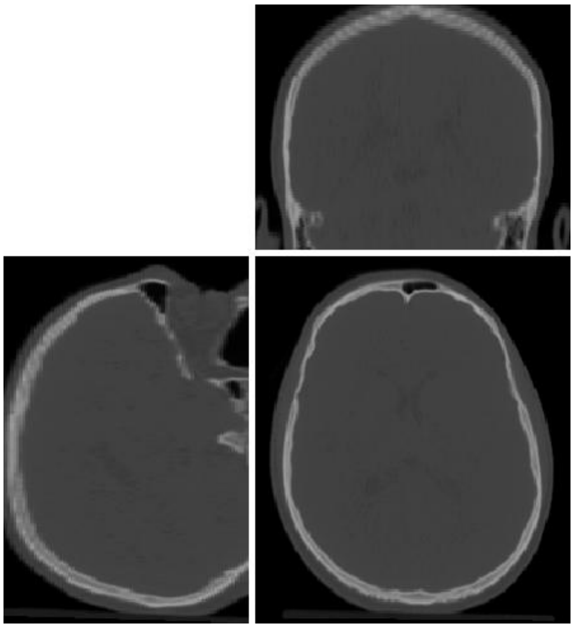
1. Pour l'une des deux images I ou J, la fonction d'intensité $f(S)$ ou $g(S)$ est injective. Sans



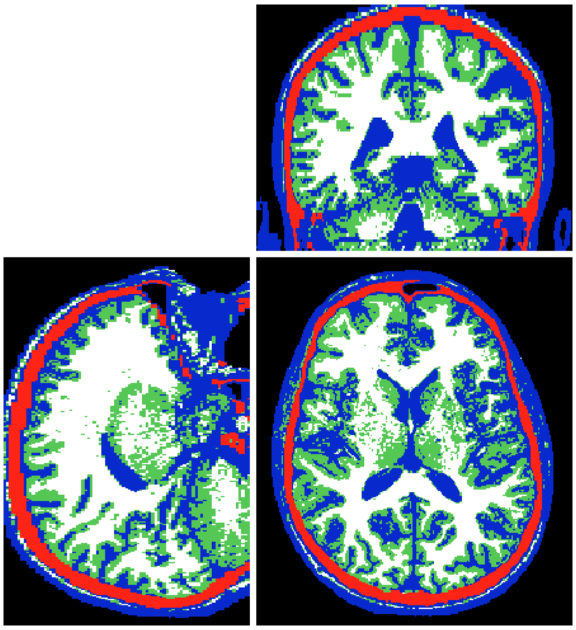
(a) Volume IRM-T1



(b) Volume scanner

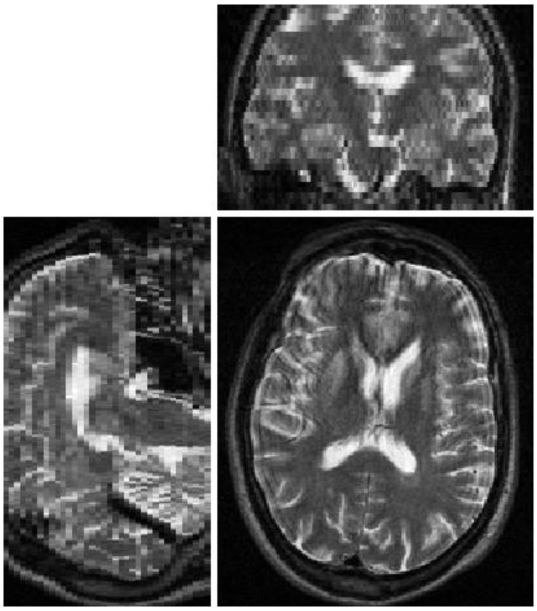


(c) Volume scanner recalé

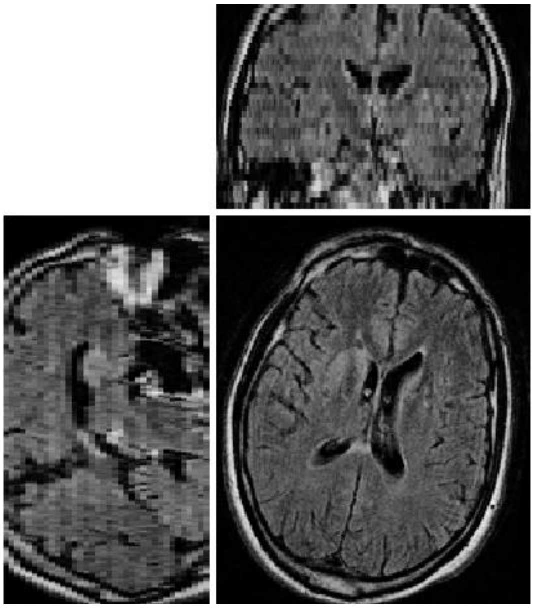


(d) Segmentation en 5 classes

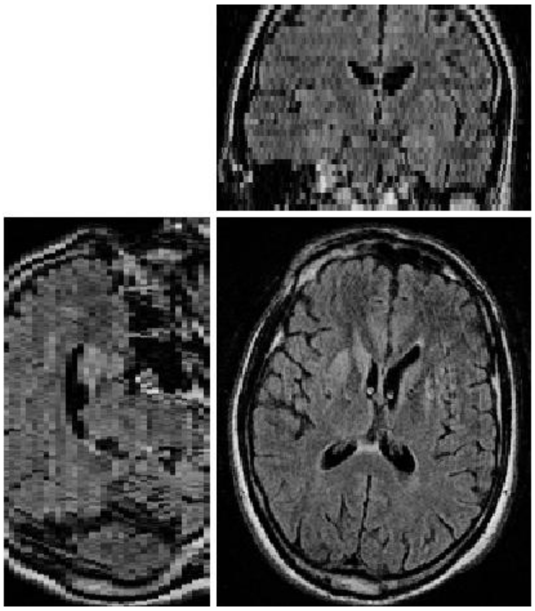
FIG. 5.3 – *Recalage rigide IRM/scanner par maximum de vraisemblance (données fournies par le Centre Antoine Lacassagne, Nice). Les paramètres auxiliaires θ sont estimés par un algorithme EM avec $K = 5$. La fusion des deux modalités permet de segmenter la matière blanche, la matière grise, le liquide céphalo-rachidien (structures bien distinctes dans l'IRM) et le crâne (très distinct dans le scanner, mais dont les intensités en IRM-T1 sont masquées par celles du liquide céphalo-rachidien).*



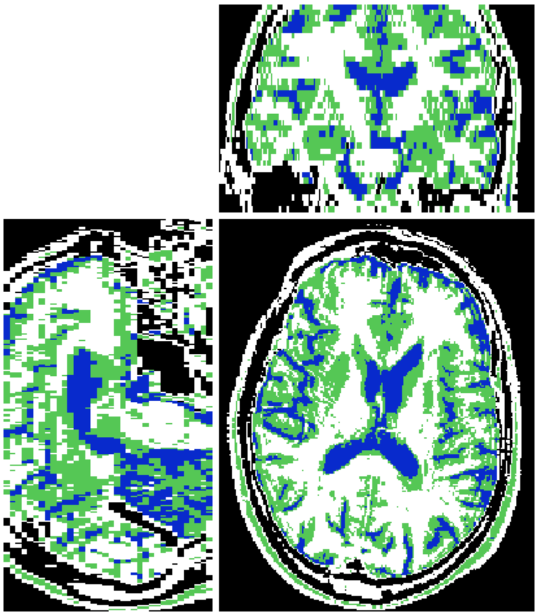
(a) Volume IRM-T2



(b) Volume IRM-FLAIR



(c) Volume FLAIR recalé



(d) Segmentation en 4 classes

Fig. 5.4 – *Recalage rigide T2/FLAIR par maximum de vraisemblance (données collectées par la* CJD Surveillance Unit*, Edimbourg, Royaume-Uni, dans le cadre du projet européen BMH4-98-6048 QAMRIC). Les paramètres auxiliaires θ sont estimés par un algorithme EM avec $K = 4$. Le patient est atteint de la forme sporadique de la maladie de Creutzfeldt-Jakob, ce qui se traduit notamment par l'existence d'hyper-signaux en T2 et en FLAIR dans le putamen et le noyau caudé. La fusion et la segmentation des deux modalités pourrait aider à la détection des hyper-signaux et, par là-même, au diagnostic précoce de la maladie.*

perte de généralité, nous supposerons qu'il s'agit de l'image J. Ainsi, deux labels anatomiques distincts p et p' ont des intensités caractéristiques distinctes : $g_p \neq g_{p'}$.

2. Dans cette même image, le niveau de bruit est négligeable devant le contraste :

$$\sigma_\psi \ll \min_{p \neq q} |g_p - g_q|$$

La conjonction des hypothèses 1 et 2 implique que le contenu informationnel de l'image I est redondant par rapport à celui de l'image J : deux structures anatomiques discernables dans I le sont nécessairement dans J. Ceci revient à supposer que la scène anatomique est connue à un bruit de faible niveau près. Par souci de clarté, nous allons explicitement restreindre l'étude au cas où les distributions ϕ et ψ sont gaussiennes. Il est alors possible de donner une approximation analytique des paramètres $\hat{\theta}(T)$ réalisant le minimum de $L(T,\theta)$ à T fixé.

5.4.2.1 Approximation de $\hat{\theta}(T)$

D'après l'équation (5.10), la log-vraisemblance négative peut être écrite comme la somme d'un terme conditionnel et d'un terme marginal lié uniquement à l'image J et indépendant de T :

$$L(T,\theta) = \underbrace{-\sum_{\Omega_I^\downarrow} \log P(i_k|j_k^\downarrow) - \sum_{\bar{\Omega}_I^\downarrow} \log P(i_k|\theta)}_{\log P(I|J,T,\theta)} \underbrace{- \sum_{\Omega_J} \log P(j_\ell|\theta)}_{\log P(J|\theta)} \tag{5.15}$$

Concernant le terme conditionnel, nous allons supposer que la somme sur $\bar{\Omega}_I^\downarrow$ est négligeable devant la somme sur $\Omega_I^\downarrow$. Cette approximation est justifiée si la majorité des voxels de Ω_I sont appariés. En gardant cette condition à l'esprit, nous pouvons écrire :

$$-\log P(I|J,T,\theta) \approx -\sum_{\Omega_I^\downarrow} \log P(i_k|j_k^\downarrow)$$

Les hypothèses 1 et 2 vont nous permettre de simplifier la forme paramétrique de la densité de probabilité conditionnelle $P(i|j)$. D'après l'équation (5.8), celle-ci est en principe une distribution de mélange dont les proportions dépendent de j :

$$P(i|j) = \sum_p \pi'_p(j)\phi(i - f_p)$$

avec :

$$\pi'_p(j) = \frac{\pi_p \psi\,(j - g_p)}{\sum_q \pi_q \psi\,(j - g_q)} = \frac{1}{1 + \sum_{q \neq p} \frac{\pi_q \psi\,(j - g_q)}{\pi_p \psi\,(j - g_p)}} \tag{5.16}$$

Pour deux labels distincts $p \neq q$, nous avons $g_p \neq g_q$ d'après l'hypothèse d'injectivité (hypothèse 1). Puisque ψ est gaussienne, un calcul direct donne :

$$\frac{\psi(j - g_q)}{\psi(j - g_p)} = \exp\left[\frac{1}{2\varepsilon^2} \underbrace{\left(\frac{g_p + g_q - 2j}{g_p - g_q}\right)}_{\alpha(j)}\right], \qquad \text{avec} \quad \varepsilon = \frac{\sigma_\psi}{g_p - g_q}$$

La quantité ε peut être considérée comme un infiniment petit du premier ordre en vertu de l'hypothèse 2, ce qui justifie un passage à la limite dans l'expression précédente. Selon le signe de la fonction $\alpha(j)$, on obtient :

$$\lim_{\varepsilon \to 0} \frac{\psi(j-g_q)}{\psi(j-g_p)} = \begin{cases} 0 & \text{si} \quad \alpha(j) < 0 \quad \Leftrightarrow \quad |j-g_p| < |j-g_q| \\ +\infty & \text{si} \quad \alpha(j) > 0 \quad \Leftrightarrow \quad |j-g_p| > |j-g_q| \end{cases}$$

En reportant ces limites dans l'équation (5.16), on voit que la valeur du poids $\pi'_p(j)$ dépend du fait que g_p soit ou non la valeur la plus proche de j dans l'ensemble $\{g_1,\ldots,g_K\}$. Ceci justifie d'introduire la partition de $\mathbb{R}$ formée par les intervalles ouverts suivants :

$$\mathcal{J}_p = \left\{ j \in \mathbb{R}, \quad \forall q \neq p, \quad |j-g_p| < |j-g_q| \right\}$$

Ces intervalles sont a priori de taille variable et il en existe autant que l'on impose de classes, c'est-à-dire : K. En ignorant les cas pathologiques où une intensité j est à la frontière de deux intervalles $\mathcal{J}_p$, on a :

$$\pi'_p(j) \approx \begin{cases} 1 & \text{si} \quad j \in \mathcal{J}_p \\ 0 & \text{si} \quad j \notin \mathcal{J}_p \end{cases}$$

Pour toute valeur d'intensité $j \in \mathbb{R}$, il existe donc un unique label p pour lequel $\pi'_p(j) \neq 0$ (sauf sur un ensemble de mesure nulle correspondant aux frontières des intervalles $\mathcal{J}_p$). Nous pouvons donc définir la fonction constante par morceaux $F : \mathbb{R} \to \mathbb{R}$ qui à j associe la valeur f_p correspondant au label p vérifiant $\pi'_p(j) \neq 0$; voir la figure 5.5.

Il découle de cette définition la relation : $P(i|j) = \phi(i - F(j))$, d'où l'approximation suivante de la log-vraisemblance conditionnelle :

$$-\log P(I|J,T,\theta) \approx -\sum_{\Omega_I^{\downarrow}} \log \phi\left(i_k - F(j_k^{\downarrow})\right)$$

Moyennant un simple reparamétrage, les inconnues du modèle sont F, g, π, σ_ϕ et σ_ψ (nous raisonnons à T fixé). Nous observons alors que la minimisation de $L(T,\theta)$ en fonction de θ revient à deux minimisations indépendantes puisque :

- Le terme conditionnel dépend uniquement de F et σ_ϕ.
- Le terme marginal dépend uniquement de g, π et σ_ψ :

$$-\log P(J|\theta) = -\sum_{\Omega_J} \log P(j_\ell|\theta) \qquad \text{avec} \quad P(j) = \sum_p \pi_p \psi(j - g_p)$$

Ce dernier terme peut être minimisé par un algorithme EM analogue à celui du paragraphe 5.4.1 si ce n'est que les données sont cette fois unidimensionnelles. Par ailleurs, $\log P(J|\theta)$ étant indépendant de T, ce sera également le cas des estimées par maximum de vraisemblance $\hat{g}$, $\hat{\pi}$ et $\hat{\sigma}_\psi$. En conséquence, l'algorithme EM peut n'être appelé qu'une seule fois, en préambule de la procédure de recalage proprement dite. Il convient par ailleurs de vérifier *a posteriori* que les estimées $\hat{g}$ et $\hat{\pi}$ vérifient bien les hypothèses 1 et 2 énoncées plus haut.

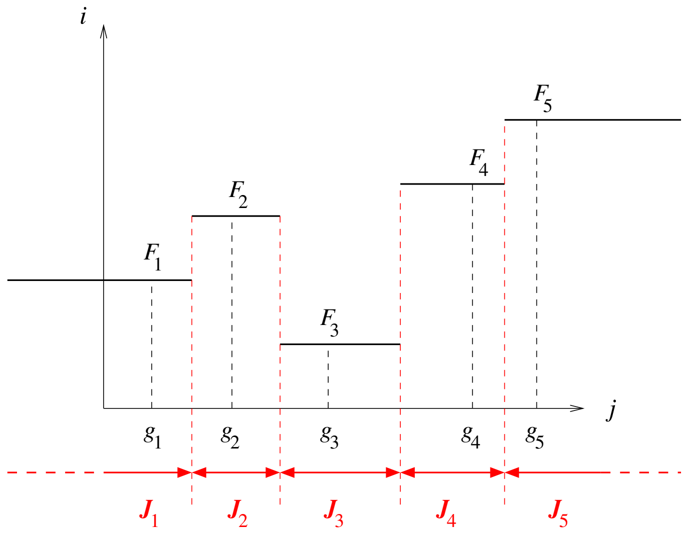


FIG. 5.5 – *Représentation graphique de la fonction F reliant les intensités des images I et J.*

La minimisation du terme conditionnel dépend en revanche de la transformation considérée, mais nous allons voir qu'elle admet une solution explicite. Puisque ϕ est supposée gaussienne, nous voyons apparaître un critère de moindres carrés qui nous est familier depuis le chapitre 3 :

$$-\log P(I|J,T,\theta) \approx n^{\downarrow} \log\left(\sqrt{2\pi}\,\sigma_\phi\right) + \frac{1}{2\sigma_\phi^2}\sum_{\Omega_I^{\downarrow}}\left(i_k - F(j_k^{\downarrow})\right)^2$$

Maximiser la vraisemblance conditionnelle (approchée) par rapport à F équivaut à résoudre un problème de régression aux moindres carrés du même type que celui que nous avons traité au chapitre 3, paragraphe 3.3.1. Ici, F est recherchée comme une fonction constante sur les intervalles $\mathcal{J}_p$ déterminés au préalable (par l'algorithme EM). Les paramètres naturels de la fonction F sont les valeurs F_p qu'elle prend sur les différents intervalles $\mathcal{J}_p$. Puisque nous voulons également estimer le paramètre d'échelle σ_ϕ, il nous faut dériver $\log P(I|J,T,\theta)$ par rapport à F_p et σ_ϕ :

$$\left\{\begin{array}{lclcl}
\dfrac{\partial L}{\partial F_p} & = & -\dfrac{1}{\sigma_\phi^2}\displaystyle\sum_{\Omega_I^j}(i_k - F_p) & \Longrightarrow & \hat{F}_p = \dfrac{1}{n_p^{\downarrow}}\displaystyle\sum_{\Omega_I^{\downarrow p}} i_k \\[2ex]
\dfrac{\partial L}{\partial \sigma_\phi} & = & \dfrac{n^{\downarrow}}{\sigma_\phi} - \dfrac{1}{\sigma_\phi^3}\displaystyle\sum_{\Omega_I^{\downarrow}}\left(i_k - F(j_k^{\downarrow})\right)^2 & \Longrightarrow & \hat{\sigma}_\phi^2 = \dfrac{1}{n^{\downarrow}}\displaystyle\sum_{\Omega_I^{\downarrow}}\left(i_k - \hat{F}(j_k^{\downarrow})\right)^2
\end{array}\right.$$

avec $\Omega_I^{\downarrow p} \equiv \left\{\mathbf{x}_k \in \Omega_I^{\downarrow},\, j_k^{\downarrow} \in \mathcal{J}_p\right\}$ et $n_p^{\downarrow} \equiv \mathrm{Card}\,\Omega_I^{\downarrow p}$.

Si nous avons pu nous ramener à un problème de régression aux moindres carrés, c'est parce qu'en première approximation, la distribution d'intensité de l'image I est conditionnellement gaussienne. Cela est dû non seulement à la forme gaussienne de ϕ mais également à l'hypothèse de faible bruit dans l'image J. La forme paramétrique de ψ, que nous avons également supposée gaussienne, ne joue en fait qu'un rôle mineur. Elle nous a permis de justifier, via un passage à la limite, l'existence d'une fonction constante par morceaux reliant les intensités de I et J. Il est clair, cependant, que d'autres types de loi autorisent le même passage à la limite.

Par contre, la forme paramétrique de ϕ influence directement la métrique du problème de régression, l'hypothèse gaussienne permettant d'obtenir un critère quadratique. On peut vérifier, par exemple, que si ϕ est une loi de Laplace :

$$\phi(x) = \frac{1}{2\sigma_\phi}\, e^{-\frac{|x|}{\sigma_\phi}},$$

la métrique associée est la norme L_1. Ce cas a été étudié au chapitre 3, paragraphe 3.4.2.4.

5.4.2.2 Lien avec le rapport de corrélation

Ayant estimé les paramètres auxiliaires $\hat{\theta}(T)$ au prix, rappelons-le, de quelques approximations, nous les substituons dans l'expression de la log-vraisemblance négative (5.15):

$$\mathcal{L}(T) \approx n^{\downarrow} \log\left(\sqrt{2\pi e}\,\hat{\sigma}_\phi\right) - \sum_{\bar{\Omega}_I^{\downarrow}} \log\left[\sum_p \hat{\pi}_p \phi(i_k - \hat{F}_p)\right] \underbrace{- \sum_{\Omega_J} \log\left[\sum_p \hat{\pi}_p \psi(j_\ell - \hat{g}_p)\right]}_{\text{cte}}$$

Nous pouvons faire quelques remarques concernant chacun des trois termes de cette expression. De gauche à droite :

1. Le premier terme est l'approximation précédente de la log-vraisemblance conditionnelle calculée avec les paramètres auxiliaires optimaux : il dépend de T car $n^{\downarrow}$ et $\hat{\sigma}_\phi$ en dépendent. Dans le contexte où nous étudions la dépendance en T, nous utilisons la notation fléchée $\hat{\sigma}_\phi^{\downarrow}$ plutôt que $\hat{\sigma}_\phi$.
2. Le second terme correspond à la partie de la log-vraisemblance conditionnelle que nous avions négligée précédemment. Il dépend également de T car $\bar{\Omega}_I^{\downarrow}$, $\hat{F}$ et $\hat{\sigma}_\phi$ en dépendent. On peut l'interpréter comme $(n - n^{\downarrow})$ fois l'entropie empirique $\bar{h}_I^{\downarrow}$ de l'image I, calculée sur l'ensemble $\bar{\Omega}_I^{\downarrow}$ des voxels non-appariés :

$$-\sum_{\bar{\Omega}_I^{\downarrow}} \log\left[\sum_p \hat{\pi}_p \phi(i_k - \hat{F}_p)\right] = (n - n^{\downarrow})\bar{h}_I^{\downarrow}$$

   Il serait cohérent de le négliger à nouveau dans le contexte de la minimisation par rapport à T, mais nous allons voir qu'il joue le rôle d'une énergie de rappel dont il serait dangereux de se passer.
3. Enfin, le troisième terme est la log-vraisemblance minimale de la seule image J et, comme nous l'avons signalé plus haut, il est indépendant de T.

En tenant compte de ces remarques, nous aboutissons à une écriture plus concise :

$$\mathcal{L}(T) \approx n^{\downarrow} \log\left(\sqrt{2\pi e}\,\hat{\sigma}_\phi^{\downarrow}\right) + (n - n^{\downarrow})\bar{h}_I^{\downarrow} + \text{cte}$$

Cette formule donne lieu à une interprétation satisfaisante : la log-vraisemblance négative minimale à T fixé est une fonction décroissante de $\hat{\sigma}_\phi^{\downarrow}$: l'erreur quadratique moyenne d'ajustement de

I par J. Cependant, minimimiser $\mathcal{L}(T)$ n'est pas strictement équivalent à minimiser $\hat{\sigma}_\phi^\downarrow$ car $\mathcal{L}(T)$ subit l'influence d'autres termes liés au recouvrement partiel entre les images. L'entropie $\bar{h}_I^\downarrow$ impose que la partie non-appariée de I ait un faible contenu informationnel. On voit qu'elle agit de manière à éviter que les images se déconnectent. Par ailleurs, le nombre $n^\downarrow$ de voxels appariés intervient explicitement dans le critère en assurant une pondération relative des deux énergies : plus la zone de recouvrement est resserrée, plus le terme de rappel est influent, et vice-versa.

Il y a là plus qu'une analogie avec le rapport de corrélation. Lors du chapitre 3, le rapport de corrélation est apparu dans le contexte général d'un problème de régression linéaire aux moindres carrés. C'est par définition, à une fonction décroissante près, le rapport entre l'erreur quadratique moyenne d'ajustement et la variance de l'image I calculée sur l'ensemble des voxels appariés $\Omega_I^\downarrow$; en appliquant une transformation logarithmique pour faciliter la comparaison avec $\mathcal{L}(T)$, nous avons :

$$\log\left(\sqrt{1-\eta^2_{I|J}(T)}\right) = \log\left(\hat{\sigma}_\phi^\downarrow\right) - \log\left(\sigma_I^\downarrow\right)$$

De cette manière, on voit que le rapport de corrélation et la log-vraisemblance sont deux versions normalisées de l'erreur quadratique moyenne $\hat{\sigma}_\phi^\downarrow$. Si l'on ne tient pas compte des effets de recouvrement partiel, maximiser le rapport de corrélation revient à minimiser $\hat{\sigma}_\phi^\downarrow$. Lorsque la zone de recouvrement est susceptible de varier en fonction de T, la variance simple $\sigma_I^\downarrow$ joue le rôle d'une énergie de rappel analogue à l'entropie $\bar{h}_I^\downarrow$. Cette variance est cependant calculée sur l'ensemble $\Omega_I^\downarrow$ des voxels *appariés* et non sur son complémentaire : elle est à rendre aussi grande que possible, contrairement à $\bar{h}_I^\downarrow$. Notons également que la taille relative des ensembles $\Omega_I^\downarrow$ et $\bar{\Omega}_I^\downarrow$ n'intervient pas explicitement dans le rapport de corrélation.

Ainsi, le rapport de corrélation apparaît étroitement lié au critère de vraisemblance dans le cas particulier des hypothèses de redondance énoncées au début du paragraphe 5.4.2. Toutefois, la normalisation découlant de la formulation par maximum de vraisemblance n'est pas la même que celle que nous avions obtenue de manière axiomatique au chapitre 3, paragraphe 3.2.2.

5.4.3 Hypothèse de complexité anatomique

Dans les deux paragraphes précédents, nous avons considéré le nombre de classes anatomiques K comme une constante connue. Nous supposons ici que K est inconnu. En revanche, les paramètres d'échelle σ_ϕ et σ_ψ seront fixés pour des raisons qui apparaîtront plus loin.

5.4.3.1 Approximation de $\hat{\theta}(T)$

Revenons un instant sur l'équation (5.11) définissant la fonction de vraisemblance $L(T,\theta)$. Minimiser $L(T,\theta)$ selon θ revient à chercher la densité de probabilité conjointe $P(i,j)$ qui explique au mieux la répartition des intensités que l'on observe pour une transformation T donnée. Les distributions $P(i,j)$ prédites par le modèle sont de la forme :

$$P(i,j) = \sum_{p=1}^{K} \pi_p \phi\left(i - f_p\right) \psi\left(j - g_p\right)$$

En termes algébriques, ce sont des combinaisons linéaires de fonctions du type $\phi(i-x)\psi(j-y)$. Lorsque K est un paramètre libre, notre problème revient à minimiser :

$$L(T,P) = -\sum_{\mathbf{x}_k \in \Omega_I^{\downarrow}} \log P(i_k,j_k^{\downarrow}) \; - \sum_{\mathbf{x}_k \in \bar{\Omega}_I^{\downarrow}} \log P(i_k) \; - \sum_{\mathbf{y}_\ell \in \bar{\Omega}_J^{\downarrow}} \log P(j_\ell), \tag{5.17}$$

sur l'espace vectoriel engendré par les fonctions $\phi(i-x)\psi(j-y)$:

$$\mathcal{E} = \mathrm{Vect}\left\{\phi(i-x)\psi(j-y),\, (x,y) \in \mathbb{R}^2\right\}$$

et sous la contrainte que $P(i,j)$ soit une fonction de densité. Nous verrons qu'il suffit d'imposer :

$$\int_{\mathbb{R}^2} P(i,j)\, di\, dj = 1 \tag{5.18}$$

Commençons par nous placer dans l'espace $L^2(\mathbb{R}^2)$ dont $\mathcal{E}$ est un sous-espace si ϕ et ψ sont à énergie finie, ce que nous supposerons. Dans $L^2(\mathbb{R}^2)$, la log-vraisemblance négative $L(T,P)$ est différentiable au sens des distributions :

$$\left\{\begin{array}{lcl} \dfrac{\partial}{\partial P} \log P(i_k,j_k^{\downarrow}) & = & \dfrac{\delta(i-i_k)\delta(j-j_k^{\downarrow})}{P(i_k,j_k^{\downarrow})} \\ \dfrac{\partial}{\partial P} \log P(i_k) & = & \dfrac{\delta(i-i_k)}{P(i_k)} \\ \dfrac{\partial}{\partial P} \log P(j_\ell) & = & \dfrac{\delta(j-j_\ell)}{P(j_\ell)} \end{array}\right.$$

$$\Longrightarrow \quad \frac{\partial L}{\partial P} = -\sum_{\Omega_I^{\downarrow}} \frac{\delta(i-i_k)\delta(j-j_k^{\downarrow})}{P(i_k,j_k^{\downarrow})} - \sum_{\bar{\Omega}_I^{\downarrow}} \frac{\delta(i-i_k)}{P(i_k)} - \sum_{\bar{\Omega}_J^{\downarrow}} \frac{\delta(j-j_\ell)}{P(j_\ell)} \tag{5.19}$$

On remarque par ailleurs que le gradient de la contrainte (5.18) est constant et unitaire. Il existe donc un multiplicateur de Lagrange $\lambda > 0$ tel que la relation

$$\frac{\partial L}{\partial P} + \lambda = 0$$

soit vraie au minimum. Mais nous parlons ici du minimum sur $L^2(\mathbb{R}^2)$. Pour obtenir le minimum sur $\mathcal{E}$, la condition d'existence est que $(\partial L/\partial P + \lambda)$ soit orthogonal à $\mathcal{E}$ au sens du produit scalaire défini sur $L^2(\mathbb{R}^2)$, ce qui se traduit par :

$$\forall (x,y) \in \mathbb{R}^2, \quad \iint \phi(i-x)\psi(j-y)\Big(\frac{\partial L}{\partial P} + \lambda\Big) di\, dj = 0$$

Cette équation fait intervenir une convolution bidimensionnelle par le noyau $\phi(-i)\psi(-j)$. Par souci de clarté, nous supposerons que ϕ et ψ sont symétriques, c'est-à-dire $\phi(-i) = \phi(i)$ et $\psi(-j) = \psi(j)$. Puisque ϕ et ψ sont des fonctions de densité, la convolution de la constante λ par $\phi(i)\psi(j)$ est identiquement égale à λ. On aboutit ainsi à la condition du premier ordre :

$$\lambda = \sum_{\Omega_I^{\downarrow}} \frac{\phi(i-i_k)\psi(j-j_k^{\downarrow})}{\hat{P}(i_k,j_k^{\downarrow})} + \sum_{\bar{\Omega}_I^{\downarrow}} \frac{\phi(i-i_k)}{\hat{P}(i_k)} + \sum_{\bar{\Omega}_J^{\downarrow}} \frac{\psi(j-j_\ell)}{\hat{P}(j_\ell)} \tag{5.20}$$

L'équation (5.20) définit la distribution optimale $\hat{P}(i,j)$ de manière implicite. Il semble malheureusement que sa résolution analytique soit impossible. Pour en évaluer la solution, nous allons faire les approximations suivantes :

1. Les sommes sur $\bar{\Omega}_I^{\downarrow}$ et $\bar{\Omega}_J^{\downarrow}$ sont négligeables vis-à-vis de la somme sur $\Omega_I^{\downarrow}$. Cette hypothèse est justifiée si le nombre d'appariements est grand devant le nombre de voxels sans correspondant.
2. La résolution serait plus aisée si on pouvait faire apparaître $\hat{P}(i,j)$ dans l'équation (5.20), ce qui motive l'approximation :

$$\frac{\phi(i-i_k)\psi(j-j_k^{\downarrow})}{\hat{P}(i_k,j_k^{\downarrow})} \approx \frac{\phi(i-i_k)\psi(j-j_k^{\downarrow})}{\hat{P}(i,j)}$$

   Cette approximation paraît raisonnable dans la mesure où ϕ et ψ sont des fonctions en cloche à décroissance rapide.

Moyennant ces deux approximations, nous obtenons une simplification de l'équation (5.20) :

$$\lambda \approx \sum_{\Omega_I^{\downarrow}} \frac{\phi(i-i_k)\psi(j-j_k^{\downarrow})}{\hat{P}(i,j)} \quad \Longrightarrow \quad \lambda \hat{P}(i,j) \approx \sum_{\Omega_I^{\downarrow}} \phi(i-i_k)\psi(j-j_k^{\downarrow}),$$

ce qui implique clairement $\lambda = \mathrm{Card}\Omega_I^{\downarrow} \equiv n^{\downarrow}$ et donc :

$$\hat{P}(i,j) = \frac{1}{n^{\downarrow}} \sum_{\Omega_I^{\downarrow}} \phi(i-i_k)\psi(j-j_k^{\downarrow}) \tag{5.21}$$

5.4.3.2 Lien avec les fenêtres de Parzen

En établissant l'équation (5.21), nous venons de retrouver une technique d'estimation bien connue : la méthode des fenêtres de Parzen (Duda et Hart, 1973), déjà proposée par (Viola, 1995) pour calculer l'information mutuelle. Il est commun d'opposer la méthode de Parzen à la méthode du maximum de vraisemblance, en considérant que la première est non-paramétrique et que la seconde est paramétrique. Le lien qui unit ces deux méthodes est sans doute beaucoup plus fort qu'on n'a tendance à le penser. L'estimateur de Parzen apparaît ici comme une approximation naturelle de l'estimateur du maximum de vraisemblance dans le contexte où la distribution à estimer comporte un nombre indéterminé de modes.

Nous ne connaissons pas de résultat établissant la qualité de l'estimateur de Parzen en tant qu'approximation du maximum de vraisemblance. Cette piste est peut-être à explorer, d'autant plus qu'elle dépasse de très loin le cadre du recalage d'images. Par contre, il existe des conditions générales sous lesquelles l'estimateur de Parzen converge en moyenne quadratique vers la *véritable* distribution des données (Duda et Hart, 1973). Ces conditions imposent notamment de choisir les paramètres d'échelle en fonction du nombre de données, ce qui peut servir ici à fixer σ_ϕ et σ_ψ.

On peut se demander pourquoi il n'est pas possible d'estimer ces paramètres par maximum de vraisemblance. C'est bien ce que nous avons fait au paragraphe 5.4.1, lorsque le nombre de classes était fixé. La réponse est que si K est un paramètre libre, les estimées par maximum de vraisemblance de σ_ϕ et σ_ψ sont nulles. On peut s'en convaincre en revenant à l'expression de la log-vraisemblance négative (5.17). Si on s'arrange pour que $P(i,j)$ prenne une valeur infinie à chaque fois que (i,j) coïncide avec un couple d'intensités disponible, alors la vraisemblance de P sera infinie ! Il existe de nombreuses façons de réaliser cet objectif ; prenons, par exemple, la distribution suivante :

$$\hat{Q}(i,j) = \frac{1}{nm} \sum_{k,\ell} \delta(i - i_k)\delta(j - j_\ell),$$

dont on note au passage qu'elle est indépendante de T. Le problème est que $\hat{Q}$ fait partie des distributions prédites par le modèle dès lors que σ_ϕ et σ_ψ ne sont pas fixés. Le choix d'une telle distribution revient à conclure qu'il n'y pas de bruit dans les images ($\sigma_\phi = \sigma_\psi = 0$) et donc que les variations d'intensités sont dues exclusivement à des changements de tissus anatomiques. Ceci est tout à fait contraire au bon sens. Il est pourtant difficile de trouver une distribution plus vraisemblable que $\hat{Q}$...

La contradiction provient du fait que si aucune contrainte n'est imposée à l'espace de recherche de la distribution, alors le nombre de paramètres du modèle devient plus grand que le nombre de données disponibles. Le problème d'estimation qui en découle est donc mal posé et il est indispensable de le régulariser. Le fait de fixer les paramètres d'échelle est une façon de le faire.

5.4.3.3 Lien avec l'information mutuelle au sens de Viola

Cette fois, le lien avec l'information mutuelle de (Viola, 1995) est plus profond qu'il ne l'était au paragraphe 5.3.4. En tenant compte de l'analogie exprimée par l'équation (5.13), la log-vraisemblance optimale à T fixé s'interprète comme une combinaison linéaire d'entropies empiriques calculées à partir de la distribution de Parzen $\hat{P}$:

$$\mathcal{L}(T) = -n^{\downarrow}\mathcal{I}^{\downarrow}_{I,J} + nh^{\downarrow}_I + mh^{\downarrow}_J, \tag{5.22}$$

où nous signalons par une flèche orientée vers le bas les termes qui dépendent de T.

Minimiser $\mathcal{L}(T)$ en fonction de T (maximiser la vraisemblance) *n'est pas* équivalent à maximiser l'information mutuelle car les quantités $n^{\downarrow}$, $h^{\downarrow}_I$ et $h^{\downarrow}_J$ dépendent a priori de T. Pour qu'elles n'en dépendent pas, il faudrait que T soit toujours une bijection parfaite entre les grilles Ω_I et Ω_J, ce qui ne se produira jamais en pratique. Nous avons néanmoins de bonnes raisons de penser que ces trois termes «parasites» sont à variation lente par rapport à T. Il y aurait alors une quasi-équivalence entre la vraisemblance et l'information mutuelle au sens de Viola.

Un cas particulier intéressant est celui où les fonctions ϕ et ψ sont choisies comme des fonctions en créneau, c'est-à-dire des lois uniformes. L'équation (5.21) montre alors que $\hat{P}$ est une version continue de l'histogramme conjoint des images tel qu'il est défini classiquement (voir le chapitre 2, paragraphe 2.3.1). La notion d'histogramme conjoint apparaît ainsi sous un jour nouveau, en tant que résultat d'un processus d'estimation. Cette remarque n'est pas neutre car elle implique un changement radical de point de vue. Dans les chapitres précédents, l'interprétation probabiliste de l'histogramme conjoint était parfaitement artificielle ; son seul intérêt était de justifier l'utilisation de coefficients de liaison issus de la littérature probabiliste comme mesures de similarité entre images.

Ici, l'histogramme conjoint intervient en tant qu'estimée de la loi de probabilité conjointe des intensités. Parler d'une telle loi a un sens dans la mesure où nous avons modélisé les images comme des réalisations de processus aléatoires stationnaires. Dans ce cadre, il existe effectivement une relation probabiliste entre les intensités, le caractère probabiliste traduisant la variabilité des procédés d'acquisition et la variabilité anatomique des sujets.

Dans les implémentations de l'information mutuelle telle que celle de (Maes et al., 1997), l'histogramme conjoint est discrétisé et, de ce fait, interprété comme une loi de probabilité discrète. Les différentes entropies $H_{I,J}$,$H_{I|J}$, H_I, H_J, ... qui en sont déduites ne correspondent pas exactement à la notion d'entropie empirique qui apparaît dans l'expression de la log-vraisemblance $\mathcal{L}(T)$. Ce sont de «véritables» entropies au sens où leur calcul fait intervenir uniquement la loi de probabilité (discrète) et non pas directement les données :

$$H_{I,J} = -\sum_{i,j} p_{i,j} \log p_{i,j}, \qquad H_{I|J} = -\sum_{i,j} p_{i,j} \log p_{i|j}, \qquad \text{etc.}$$

En définitive, l'entropie d'une loi discrétisée et l'entropie empirique sont deux approximations différentes d'une même quantité, à savoir l'entropie de la loi continue[4]. Ainsi, bien que l'information mutuelle de l'histogramme conjoint ne s'identifie pas exactement à l'information mutuelle empirique, elle lui est analogue et, par là-même, elle est étroitement liée au critère de vraisemblance.

5.5 Interpolation

Jusqu'à présent, nous avons admis que la transformation T est recherchée parmi des applications de la grille source dans le prolongement infini de la grille cible. Ceci implique que le correspondant $T(\mathbf{x}_k)$ d'un voxel $\mathbf{x}_k$ coïncide parfaitement avec un voxel $\mathbf{y}_\ell$. En pratique, cette limitation est rarement acceptable, non seulement pour des questions de précision mais également parce que la transformation à estimer est continue par nature.

Idéalement, nous aimerions prendre en compte des transformations spatiales quelconques tout en restant dans le cadre de l'estimation par maximum de vraisemblance. Or, la restriction est imposée par la nature même du modèle que nous avons adopté (voir le paragraphe 5.2). Elle découle du fait que le support de la scène est spécifié comme discret. Ainsi, la fonction de vraisemblance n'est tout simplement pas définie si T ne vérifie pas les conditions énoncées plus haut. Sauf à redéfinir le modèle, il n'est pas possible de lever l'hypothèse sur T autrement que de manière ad-hoc.

Une problématique analogue est déjà apparue à plusieurs reprises lors des chapitres 2 et 3. On peut la résumer à une seule question : comment évaluer la contribution d'un point $\mathbf{x}_k$ à la log-vraisemblance (5.11) sachant que l'intensité $j_k^\downarrow$ de son correspondant est inconnue ? Pour ce problème, nous disposons des techniques qui ont été décrites au chapitre 2, paragraphe 2.3.2.

4. En théorie de l'information, l'entropie d'une variable continue est souvent appelée *entropie différentielle*, l'emploi sans qualificatif du mot *entropie* étant réservé aux variables discrètes.

Rappelons qu'il existe deux types d'approches. La première consiste à estimer indirectement le critère de recalage en passant par une interpolation de l'intensité. Dans le cas d'une interpolation trilinéaire, l'intensité estimée d'un point $T(\mathbf{x}_k)$ est la moyenne pondérée de ses huit voisins immédiats dans la grille Ω_J :

$$j_k^{\downarrow} \approx \sum_{\ell} w_{k\ell}^{\downarrow} j_\ell$$

On peut alors calculer la log-vraisemblance en substituant les intensités interpolées $j_k^{\downarrow}$ dans l'équation (5.11).

La seconde approche considère le problème de façon globale. Puisque c'est fondamentalement le critère et non l'image J que l'on veut interpoler, autant le faire directement. La méthode directe qui correspond à l'interpolation trilinéaire est l'interpolation par volume partiel (VP) :

$$\log P(i_k,j_k^{\downarrow}|\theta) \approx \sum_{\ell} w_{k\ell}^{\downarrow} \log P(i_k,j_\ell|\theta)$$

La première approche n'implique aucune modification structurelle du critère de vraisemblance. En particulier, la minimisation de $L(T,\theta)$ à T fixé donne lieu aux même formules de calcul que celles obtenues au paragraphe 5.4. Ce n'est pas le cas avec les méthodes d'interpolation directe, dont l'effet est de pondérer la log-vraisemblance d'une manière telle qu'à un voxel $\mathbf{x}_k$ correspondent en fait plusieurs couples (i_k,j_ℓ), chacun se voyant attribuer un certain poids. Les formules d'estimation à T fixé peuvent être adaptées de manière à ce que la maximisation de la vraisemblance soit cohérente avec la méthode d'interpolation. Ceci donne lieu à une généralisation relativement immédiate des calculs du paragraphe 5.4.

Nous avons cependant constaté que si l'interpolation influence considérablement l'optimisation vis-à-vis des paramètres de transformation spatiale, elle ne semble pas avoir un effet déterminant sur l'estimation des paramètres auxiliaires. En pratique, il est généralement suffisant d'utiliser une interpolation par le plus proche voisin pour la phase d'optimisation par rapport à θ, ce qui permet en outre d'accéler les calculs.

5.6 Conclusion

Dans ce chapitre, nous avons formulé le recalage d'images comme un problème d'inférence statistique. Notre approche se fonde en premier lieu sur un modèle probabiliste de dépendance entre les images et en second lieu sur un estimateur destiné à identifier les paramètres du modèle, à savoir : les paramètres de la transformation spatiale mais également les paramètres «auxiliaires» caractérisant la relation entre les intensités. Les principales limitations du modèle générique que nous avons considéré sont les hypothèses d'indépendance spatiale et de stationnarité concernant les processus aléatoires qui interviennent dans la formation des images. L'estimateur auquel nous avons restreint notre étude est celui du maximum de vraisemblance.

Ce travail s'inscrit dans la continuité de publications antérieures traitant du recalage par maximum de vraisemblance (McGillem et Svedlow, 1976; Mort et Srinath, 1988; Costa et al., 1993; Viola, 1995; Leventon et Grimson, 1998; Bansal et al., 1998). Il nous semble que l'originalité de notre formulation tient essentiellement à deux aspects. D'une part, nous prenons en compte les effets de troncature liés à la nature finie des images. De ce fait, le critère de vraisemblance est naturellement robuste au recouvrement partiel et il n'est pas nécessaire de recourir à une normalisation heuristique pour garantir cette propriété. D'autre part, les paramètres auxiliaires sont estimés par maximum de vraisemblance, au même titre que les paramètres de transformation. Tous les paramètres maximisent donc un seul et même critère, ce qui confère à la méthode de bonnes propriétés de convergence.

En jouant sur les degrés de liberté du modèle, nous pouvons décliner l'estimation par maximum de vraisemblance en différentes stratégies de recalage. Le rapport de corrélation et l'information mutuelle apparaissent comme deux approximations raisonnables de la vraisemblance dans deux cas particuliers, le premier correspondant à des contraintes plutôt fortes (hypothèse de redondance), le second à des contraintes plutôt faibles (hypothèse de complexité anatomique). Ceci nous permet de mieux comprendre les *a priori* sous-jacents à ces deux mesures de similarité. De manière plus générale, le formalisme du maximum de vraisemblance fournit une méthode systématique pour concevoir des mesures de similarité en fonction d'hypothèses explicites de dépendance, notamment grâce à l'algorithme EM.

Les perspectives théoriques qu'ouvrent cette étude semblent très larges. Nous sommes convaincus que la prise en compte de corrélations spatiales pourrait avoir une influence bénéfique sur les performances de la méthode, plus particulièrement en termes de robustesse. Bien que nous ayons esquissé la possibilité de modéliser la scène anatomique comme un champ de Markov, nous n'avons pas implémenté cette stratégie et c'est peut-être une piste à explorer.

Un autre écueil théorique auquel nous nous sommes heurtés est la description spatialement continue de la scène anatomique. Nous avons contourné le problème en adoptant une formulation discrète qui nous oblige à employer des techniques ad-hoc d'interpolation dont nous maîtrisons mal l'impact sur le résultat de l'estimation. Le choix d'une «bonne» description continue est en soi un problème difficile. Nous pressentons qu'il y aurait un parti à tirer de segmentations plus fines telles que celles offertes par les modèles déformables.

Chapitre 6

Recalage rigide d'images ultrasonores et IRM

6.1 Introduction

Over the last years, the development of real-time 3D ultrasound (US) imaging has revealed a number of potential applications in image-guided surgery. The major advantages of 3D US over existing intra-operative imaging techniques are its comparatively low cost and simplicity of use. However, the automatic processing of US images has not gained the same degree of development as other medical imaging modalities, probably due to the low signal-to-noise ratio of US images, among other reasons.

The registration of US with pre-operative Magnetic Resonance (MR) images will allow the surgeon to accurately localize the course of instruments in the operative field, resulting in minimally invasive procedures. At present, few papers have been published on this particular registration problem (Maintz et Viergever, 1998). Most of the approaches that have been proposed are based on stereotactic systems. For instance, in (Pagoulatos et al., 1999) registration is achieved by tracking the US probe with a magnetic position sensor. Existing image-based methods match homologous features extracted from both the US and MR data. Features are user-identified in (Erbe et al., 1996), while semi-automatically extracted in (Hata et al., 1994). More recently, (Ionescu et al., 1999) registered US with Computed Tomography (CT) data after automatically extracting contours from the US using watershed segmentation.

The present registration technique expands on the generalized correlation ratio (GCR) method which has been described in chapter 3. It is an intensity-based approach as it does not rely on explicit feature extraction. Our early experiments revealed that none of the usual similarity measures (correlation ratio, mutual information, ...) was robust enough to enable automated registration of MR and US images. Part of these have been reported in a previous article (Roche et al., 2000b).

In this work, we have improved the GCR method based on the observation that ultrasound imagery is essentially concerned with the interfaces between anatomical structures. The idea is then to correlate the US intensity with both the intensity and the gradient of the MR image. This leads to an almost straightforward, multivariate extension of the GCR.

We detail in section 6.2 the theoretical background of the multivariate GCR. Section 6.3 describes the phantom and clinical data that were used in our experiments, while sections 6.4 and 6.5 propose original evaluations of the method accuracy and robustness, respectively.

6.2 Method

6.2.1 A rough US/MR dependence model

Ultrasound images are commonly said to be "gradient images" as they enhance the interfaces between anatomical structures. The physical reason is that the amplitudes of the US echos are proportional to the *difference* of acoustical impedance caused by successive tissue layers. Ideally, the US signal should be high at the interfaces, and zero within homogeneous tissues. This would be the case if perfectly specular reflection occurred. In practice, images obtained from conventional ultrasound imaging systems have interference patterns called speckle. Speckle is mainly responsible for the low contrast/noise ratio of ultrasound images. It is traditionally modelled as a multiplicative, Rayleigh distributed, white noise (Mulet-Parada, 2000; Rohling, 1998). We should notice however that speckle is in fact a purely deterministic phenomenon.

From these remarks, the radio-frequency ultrasound signal may be expressed as follows:

$$I_{rf}(\mathbf{x}) = |\nabla Z.\mathbf{u}|\,(\mathbf{x})\, s(\mathbf{x}),$$

where $\mathbf{x}$ is a given location in space, $Z(\mathbf{x})$ is the acoustical impedance, $\mathbf{u}(\mathbf{x})$ is the scan direction, and $s(\mathbf{x})$ is the outcome of speckle noise at $\mathbf{x}$.

Clinical imaging systems reduce the dynamic range of the radio-frequency signal using a machine-dependent process called logarithmic compression. As a consequence, the images we are provided with do not verifiy the previous relation. A rough model of the log-compression transfer function may be given as follows :

$$I(\mathbf{x}) = D \log \big[\, |\nabla Z.\mathbf{u}|\,(\mathbf{x}) \big] + G + \varepsilon(\mathbf{x}), \tag{6.1}$$

where D is the dynamic range of input, G is the linear gain, and $\varepsilon(\mathbf{x})$ is the log-compressed speckle noise. As an effect of compression, the statistics of speckle noise completely change. Deriving the actual distribution of ε is a difficult problem that has been tackled by (Dutt et Grenleaf, 1996). We will assume here that ε is normally distributed, which is consistent with a former model proposed by (Loupas et al., 1989).

Now, the question is how to relate the acoustical impedence to the signal produced by a magnetic resonance sensor. This might be a challenging problem. Based on the observation that most of the anatomical structures have specific intensity ranges in an MR image, we will make the completely heuristic assumption that there is a functional dependence between $Z(\mathbf{x})$ and the MR signal, $M(\mathbf{x})$. Ignoring the effect of noise in the MR, we may write :

$$Z(\mathbf{x}) = \varphi\big[M(\mathbf{x})\big] \quad \Longrightarrow \quad \nabla Z(\mathbf{x}) = \varphi'\big[M(\mathbf{x})\big]\, \nabla M(\mathbf{x}) \tag{6.2}$$

Of course, the function φ is unknown and it should depend on the particular protocol used for the MR acquisition (usually T1 in our experiments). From equations (6.1) and (6.2), it follows that a US image and an MR image representing the same anatomy are related by an equation of the form:

$$I(\mathbf{x}) = f_1\big[M(\mathbf{x})\big] + f_2\big[\,|\nabla M.\mathbf{u}|\,(\mathbf{x})\big] + G + \varepsilon(\mathbf{x}),$$

with $f_1 = D\log\circ\varphi'$, and $f_2 = D\log$. In practice, the "black box" parameters D and G are likely to be themselves dependent (at least) on $|\nabla Z.\mathbf{u}|$. Then it is safer to consider the more general model:

$$I(\mathbf{x}) = f\big[M(\mathbf{x}),\, |\nabla M.\mathbf{u}|\,(\mathbf{x})\big] + \varepsilon(\mathbf{x}), \tag{6.3}$$

where f is simply an unknown two-to-one mapping.

6.2.2 Bivariate correlation ratio

If equation (6.3) holds, then it certainly does up to an unknown spatial transformation, which is precisely what we are looking for. Introducing the double-valued image, $J = \big(M, |\nabla M.\mathbf{u}|\big)$, and assuming that J is the target image, we end up with a dependence model of the form :

$$I(\mathbf{x}) = f\,\left[J\big(T(\mathbf{x})\big)\right] + \varepsilon(\mathbf{x}),$$

where ε is assumed to be a stationary Gaussian additive white noise.

This is exactly the model we employed in chapter 5, section 5.4.2, to derive the correlation ratio as the similarity measure to be maximized for registration. More precisely, we have shown that maximizing the correlation ratio corresponds (approximately) to a maximum likelihood estimation of T. This argument is straightforwardly generalized to a multivariate template image and the formal definition of the bivariate correlation ratio is the same as in the monovariate case:

$$\eta^2_{I|J}(T) = 1 - \min_f \tilde{S}(T,f),$$

where the energy $\tilde{S}$ is a normalized least squares criterion. Using partial volume (PV) interpolation, $\tilde{S}$ is expressed as:

$$\tilde{S}(T,f) = \frac{1}{n^{\downarrow}\mathrm{Var}(I^{\downarrow})} \sum_{k,\ell} w^{\downarrow}_{k\ell}\,[i_k - f(j_\ell)]^2\,, \tag{6.4}$$

where the terms depending on T are marked with an arrow: $w^{\downarrow}_{k\ell}$ are the interpolation weights, $n^{\downarrow}$ is the number of points $\mathbf{x}_k$ such that $T(\mathbf{x}_k)$ has 8 neighbours in the template grid, and $\mathrm{Var}(I^{\downarrow})$ is the variance of I computed over these points.

If the template image J was the raw MR image, then the correlation ratio method would try to predict the intensities of the US by remapping those of the MR. Hence, uniform regions of the MR would remain uniform in the remapped image. One sees that this procedure is not able to account for interface-dependent variations which are observed in the US. Clearly, using the additional explanatory variable $|\nabla M.\mathbf{u}|$ enables a better prediction.

In practice, we will restrict our search to a polynomial fitting function f. Let $j_{1\ell}$ and $j_{2\ell}$ denote the intensity components of J at the voxel position $\mathbf{y}_\ell$. We are searching for a mapping of the form:

$$f(j_\ell) = \sum_{p+q \leq d} \theta_{pq} \, j_{1\ell}^p \, j_{2\ell}^q, \tag{6.5}$$

where d is the specified polynomial degree. The number of free parameters describing f then reduces to $(d+1)(d+2)/2$. In all the experiments presented below, the degree was set to $d = 3$, implying that 10 coefficients were estimated. It is shown in appendix B that minimizing equation (6.4) with respect to the polynomial coefficients brings us to a weighted least square linear regression problem. As is standard, this is solved by the method of singular value decomposition (SVD).

The main difficulty in estimating f is that $|\nabla M.\mathbf{u}|$ is unknown before registration since it actually depends on the position and orientation of the US probe in the MR coordinate system. A possible solution to overcome this problem would be to iteratively estimate $\mathbf{u}$ in terms of the current estimate of the spatial transformation T. For the time being, we have opted for a different stategy which simply consists of replacing the projected MR gradient with its modulus, $\|\nabla M\|$.

At first glance, using $\|\nabla M\|$ instead of $|\nabla M.\mathbf{u}|$ does not appear to be acceptable from the physics of US imaging. Still, we believe that there are good reasons not to take into account information from the gradient orientation, at least as a first-order approximation. Through diffraction of the ultrasound beam on interfaces, the received echo is actually less dependent on the direction propagation than would be the case with perfectly specular reflection. This, added to log-compression, tends to equalize the response values corresponding to different tissue orientations.

6.2.3 Robust intensity distance

Our method is based on the assumption that the intensities of the US may be well predicted in the least squares sense from the information available in the MR. This is related to our modelling of speckle as a stationary Gaussian white noise. Due to several ultrasound artefacts, we do not expect this assumption to be perfectly true. Shadowing, duplication or interference artefacts may cause large variations of the US intensity from its predicted value, and this even when the images are perfectly registered. Such bad intensity matches are false negative.

The sensitivity of the registration criterion to false negative may be reduced by replacing the squared intensity distance in equation (6.4) with a robust scale estimate. This is one of the generalizations of the correlation ratio that we have presented in chapter 3, section 3.4.2. The idea is to build a robust intensity distance from a one-step S-estimator (Rousseeuw et Leroy, 1987), yielding a bivariate generalized correlation ratio:

$$\tilde{S}^2(T,f) = \frac{\sigma_0^2}{K n^{\downarrow} \mathrm{Var}(I^{\downarrow})} \sum_{k,\ell} w_{k\ell}^{\downarrow} \, \rho\left(\frac{i_k - f(j_\ell)}{\sigma_0}\right), \tag{6.6}$$

where ρ is the objective function corresponding to a given M-estimator, K is a normalization constant to ensure consistency with the normal distribution, and σ_0 is some initial guess of the scale.

As a function of f, $\tilde{S}$ may be minimized by a simple iterative weighted least squares procedure as shown in appendix B. This generally requires no more than 5-6 iterations.

In our implementation, we have opted for the Geman-McClure ρ-function,

$$\rho(x) = \frac{x^2}{1 + \frac{x^2}{c^2}},$$

for its computational efficiency and good robustness properties, to which we always set a cut-off distance $c = 3.648$ corresponding to 95% Gaussian efficiency. The normalization constant is then $K = 0.832$. See chapter 3, sections 3.4.2 and 3.5, for implementation details.

6.3 Data

The experiments related here were performed within the framework of the EC-funded ROBOSCOPE project HC 4018[1]. The datasets were provided by ISM, Salzburg, Austria, for the US datasets, and the Max Planck Institute for Psychiatry, AG-NMR, Munich, Germany, for the MR datasets. We are especially grateful to our colleagues Dorothee Auer (Max Planck Institute), Michael Rudolph (ISM), and Pr. Ludwig Auer (ISM), for their dedication in this task.

The goal of the ROBOSCOPE project is to assist neuro-surgical operations using real-time 3D ultrasound images and a robotic manipulator arm. The operation is planned on a pre-operative MRI and 3D US images are acquired during surgery to track in real time the deformation of anatomical structures. In this context, the rigid registration of the pre-operative MR with the first US image (dura mater still closed) is a fundamental task to relate the position of the surgical instruments with the actual anatomical structure. This task being determinant for the global accuracy of the system, different datasets were acquired to simulate the final image quality and to perform accuracy evaluations.

It should be emphasized that all the US images provided in this project were stored in Cartesian format, which means that the actual (log-compressed) ultrasound signal is resampled on a regular cubic lattice. As a consequence, the images undergo sever interpolation artifacts (blurring) in areas which are distant from the probe. In the following, we will refer to US images as cubic images, but one has to keep in mind that this is somewhat artificial. Notably, the voxel size in Cartesian US images should not be confused with the real spatial resolution, which is in fact spatially dependent.

6.3.1 Phantom dataset

Within ROBOSCOPE, ISM developed an MR and US compatible phantom made of two balloons, one ellipsoid and one ellipsoid with a "nose" (complex ellipsoid), that can be inflated with known volumes in order to simulate deformations. Each acquisition consists in one 3D MR image and one 3D US image.

1. http://www.ibmt.fhg.de/Roboscope/home.htm

Both balloons were initially filled with 40 ml of fluid (acquisition 0). During the first series of 5 acquisitions, the ellipsoid was filled in steps of 10 ml, while the complex ellipsoid was kept constant. During the second series, the ellipsoid was deflated and the complex ellipsoid filled by steps of 10 ml:

Acquisition number	0	1	2	3	4	5	6	7	8	9	10
Ellipsoid volume (ml)	40	50	60	70	80	90	80	70	50	60	40
Complex ellipsoid volume (ml)	40	40	40	40	40	40	50	60	70	80	90

All MR images do have $256 \times 256 \times 124$ voxels of size $0.90 \times 0.90 \times 1\,\text{mm}^3$ except images 0 and 10 where the voxel size is $0.94 \times 0.94 \times 1\,\text{mm}^3$. All (Cartesian) US images have $184 \times 184 \times 184$ voxels of size $0.41\,\text{mm}^3$.

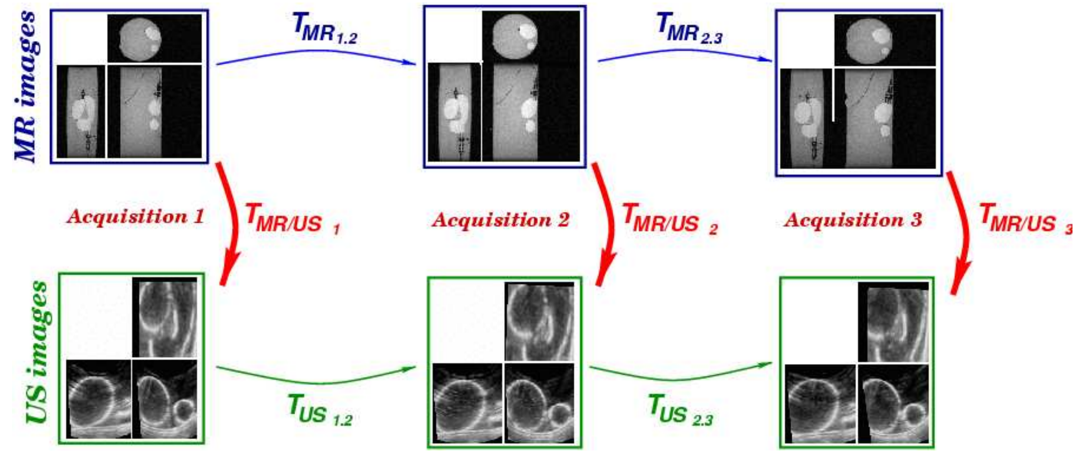


FIG. 6.1 – *Original MR and US images (before registration) of the Phantom and the rigid transformations that relate them.*

This dataset would be perfect for the validation of MR and US registration if all MR (resp. all US) images were in exactly the same coordinate system. Since the US probe cannot enter the MR machine, this is impossible: the phantom has to be moved and the US probe removed between MR acquisitions. Thus we have to register the MR (resp. the US) images together. Unfortunately, there were no rigid markers inserted in the phantom and the US/US rigid registrations are much less accurate than the MR/MR rigid registrations (see section 6.4).

6.3.2 Baby dataset

This clinical dataset was acquired to simulate the degradation of the US images quality with respect to the number of converters used in the probe. Here, we have one MR T1 image of a baby's head and 5 transfontanel US images with different percentages of converters used (40%, 60%, 70%, 80%, 90%, 100%).

The MR image has $256 \times 256 \times 124$ voxels of size $0.89 \times 0.89 \times 0.9\,\text{mm}^3$. The Cartesian US images have $184 \times 184 \times 184$ voxels of size $0.29\,\text{mm}^3$. Note the US images were truncated from the original acquisition, as can be observed in figure 6.3.

6.3.3 Patient dataset

This dataset is a real surgical case: two sagittal MR T1 images with and without a contrast agent were acquired before surgery. After craniotomy (dura mater still closed), three 3D US images were acquired to precisely locate the tumor to resect. The MR and US image sizes are summarised below:

	Image dimension			voxel size		
MR1	256	256	124	0.90	0.90	1.10
MR2	256	256	124	0.94	0.94	1.10
US1	236	136	164	0.61	0.61	0.61
US2	236	136	164	0.63	0.63	0.63
US3	190	136	194	0.95	0.95	0.95
US4	180	136	188	0.95	0.95	0.95

6.4 Accuracy study

We computed all the MR/US registrations using the previously described algorithm. The location of the US probe being linked to the pathology and its orientation being arbitrary (the rotation may be superior to 90 degrees), it was necessary to provide a rough initial estimate of the transformation. Here, this was done using an interactive interface that allows to draw lines in the images and match them. This procedure was carried out by a non-expert, generally taking less than 2 minutes. However this user interaction could be alleviated using a calibration system such as the one described in (Pagoulatos et al., 1999).

In all the experiments, the gradient norm of the MR image was computed by convolution with a Gaussian kernel (Deriche, 1992) with $\sigma = 1$ voxel. The maximization of the bivariate GCR was undertaken using the Powell method embedded in an alternate scheme (see chapter 3, section 3.5). After initialization, we observed that the algorithm found residual displacements in the range of 10 mm and 10 degrees.

6.4.1 Principle of the accuracy evaluation

To estimate the accuracy of the algorithm, one should ideally compare the result of a registration with a gold-standard. Up to our knowledge, there is no such gold-standard for MR/US registration. To get around this problem, our main idea is to use several MR and/or US images to compute registration loops and assess residual errors on test points. What we call a registration loop is a succession of composition of transformations leading to a global transformation that is theoretically the identity. A typical loop is a sequence of the form $US_i \rightarrow MR_i \rightarrow MR_j \rightarrow US_j \rightarrow US_i$ in the case of the Phantom data below (see figure 6.1).

If we were given perfectly registered images within each modality, this loop would only be disturbed from the identity by errors on the two MR/US registrations. Since variances are additive, the variance of the observed error should roughly be: $\sigma^2_{loop} = 2\sigma^2_{MR/US}$.

Unfortunately, we are not provided with a ground truth transformation within each modality: we need to estimate it. Therefore we need to take into account errors occurring at each stage of the loop. Since we are combining one MR/MR, one US/US and two MR/US registrations, the variance of the loop error will be roughly: $\sigma^2_{loop} \simeq 2\sigma^2_{MR/US} + \sigma^2_{MR/MR} + \sigma^2_{US/US}$. We thus define the *expected* accuracy as:

$$\sigma_{MR/US} \simeq \sqrt{\frac{1}{2}\left(\sigma^2_{loop} - \sigma^2_{MR/MR} - \sigma^2_{US/US}\right)}.$$

What we really measure in practice is σ_{loop}, from which we may infer the expected MR/US accuracy provided that $\sigma^2_{MR/MR}$ and $\sigma^2_{US/US}$ are faithfully estimated (see section 6.4.2). In any event, the worst-case or *conservative* MR/US accuracy is given by $\sigma_{MR/US} \simeq \sigma_{loop}/\sqrt{2}$. In order to minimize the influence of intra-modality registrations errors in this figure, we need to provide very accurate MR/MR and US/US registrations. For that purpose, we designed the following algorithm.

6.4.2 Multiple intra-modality registration

To relate n images together, we need to estimate $n-1$ rigid transformations $\bar{T}_{i,i+1}$. The principle of multiple registration is that an optimal accuracy will be obtained when registering all image pairs, and thus obtaining $n(n-1)$ transformations $\hat{T}_{i,j}$. From these, we may accept the transformations $\bar{T}_{i,i+1}$ that best explain our measurements in the least-square sense, i.e. that minimize the following criterion:

$$C(\bar{T}_{1,2}, \ldots \bar{T}_{n-1,n}) = \sum_{i,j} \mathrm{dist}^2(\bar{T}_{i,j},\hat{T}_{i,j}) \quad \text{where} \quad \bar{T}_{i,j} = \begin{cases} \bar{T}_{j-1,j} \circ \ldots \circ \bar{T}_{i,i+1} & \text{if } j > i \\ \bar{T}^{(-1)}_{j,i} & \text{if } j < i \end{cases}$$

To define the 'dist' function, we used a robust variant of the left invariant distance on rigid transformations (Pennec et al., 1998). Let χ^2 be a certain threshold and let σ_r and σ_t be typical scales corresponding to uncertainties, respectively on the rotation angle and on the translation magnitude. If (r,t) are the rotation vector and the translation of transformation T, the distance between two transformations is:

$$\mathrm{dist}^2(T_1,T_2) = \mathrm{dist}^2(T_2^{(-1)} \circ T_1,\mathrm{Id}) = \min\left(\|r_2^{(-1)} \circ r_1\|^2/\sigma_r^2 + \|t_1 - t_2\|^2/\sigma_t^2 \ , \ \chi^2\right)$$

The standard deviations σ_r and σ_t are manually adjusted to correspond roughly to the residual rotation and translation error after minimization. To obtain the minimum, we used a Newton gradient descent similar to the one described in (Pennec et al., 1998). Implementation details will be found in (Roche et al., 2000c). We end up with estimates of the rigid transformations $\bar{T}_{i,i+1}$ as well as estimates of their variance, from which we estimate the previously introduced $\sigma^2_{MR/MR}$ and $\sigma^2_{US/US}$.

6.4.3 Phantom dataset

In this experiment, we used 8 acquisitions with different balloons volumes, each acquisition consisting of one 3D MR and one 3D US image. However, we cannot directly compare the MR/US registrations as the phantom is moved between the acquisitions. Thus, the first step is to rigidly register all the MR images together and similarly the US images.

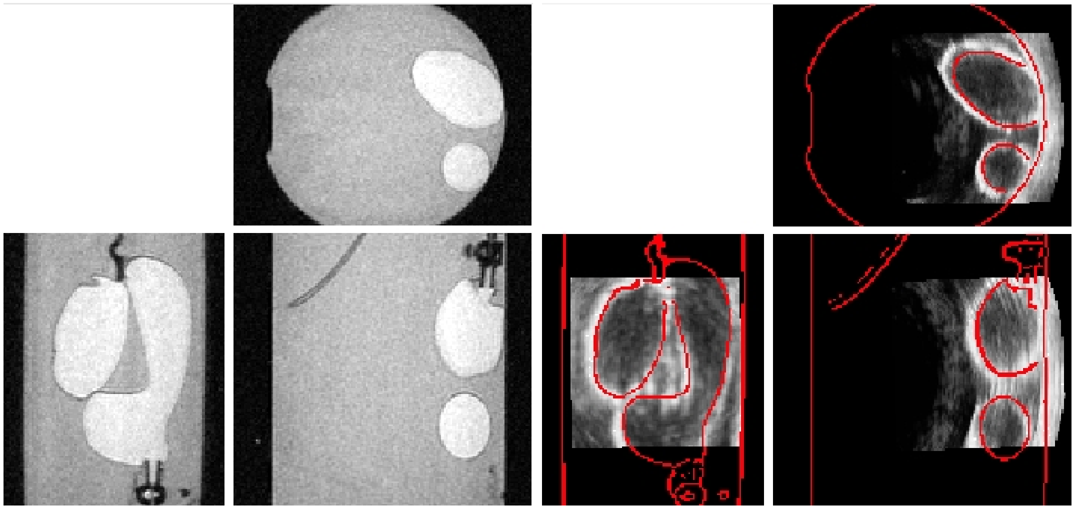


FIG. 6.2 – *Example registration of the MR and US images of the phantom. Left: original MR image. Right: registered US image with the MR contours superimposed.*

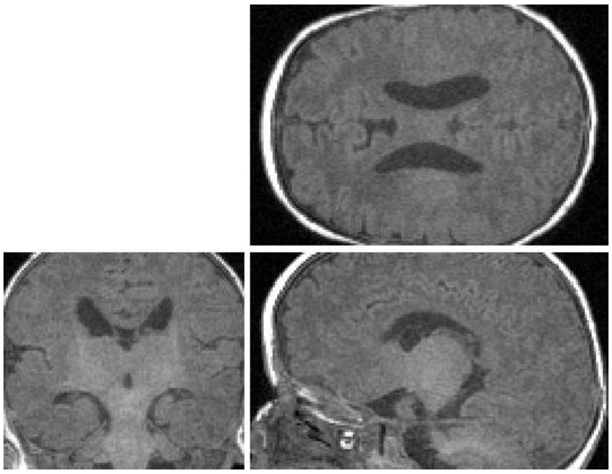
The main problem for the multiple intra-modality registration of the phantom images is that the acquisitions are intrinsically non-rigid except the outer part of the container and in a certain extent one of the two valves. Thus, intra-modal registrations of MR images were carried out using a feature-based registration algorithm known to handle a large amount of outliers (Pennec et al., 2000). As it is very difficult to extract meaningful features from the US images, we used a robust block matching technique based on the maximization of local correlation (Ourselin et al., 2000a).

The loops we used for the accuracy estimation are the $n(n-1)$ following ones: $US_i \rightarrow MR_i \rightarrow MR_j \rightarrow US_j \rightarrow US_i$. Of course, only $n-1$ loops are independent but since the ideal value is known (the identity) there is no need to correct the estimation for the number of free parameters.

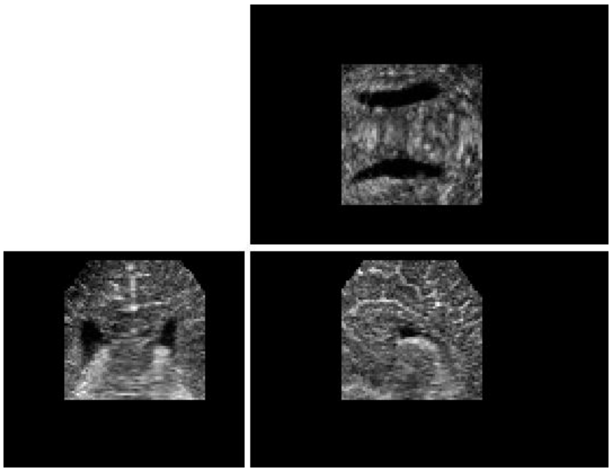
6.4.4 Baby dataset

This dataset was acquired to simulate the degradation of the US images quality with respect to the number of converters used in the probe. Here, we have one MR T1 image of a baby's head and 5 transfontanel US images with different percentages of converters used.

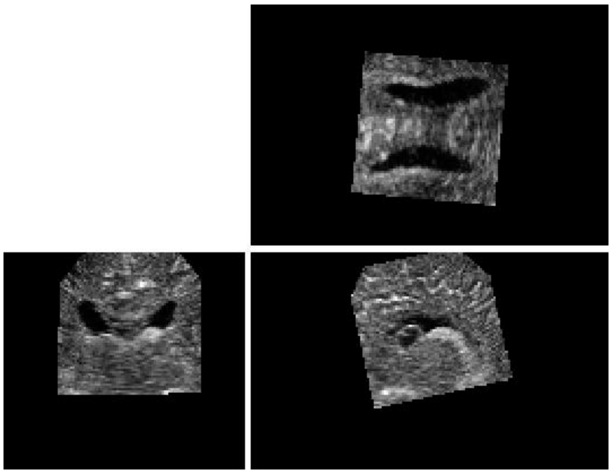
In this case, we have no or very few deformations inbetween acquisitions. Therefore, we can rigidly register all the US images onto our single MR and test the 30 following loops $US_i \rightarrow MR \rightarrow US_j \rightarrow US_i$ (only 5 of them being independent). For that, we still need to register the US images together. Our results suggest that the algorithm is much more efficient and accurate than for the phantom dataset. This is probably due to the fact that the rigidity assumption is better verified here.



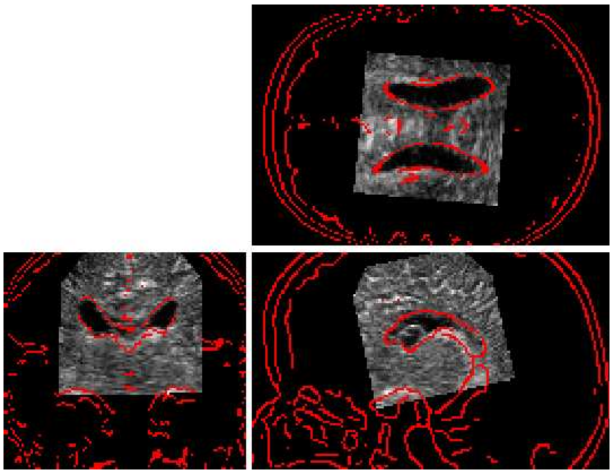
(a) MR volume (zoom)



(b) US volume (manually initialized)



(c) Registered US volume



(d) Registered US volume with MR contours overlaid

FIG. 6.3 – *Example registration of the baby US/MR images. Note that the US volume is truncated with respect to the original acquisition.*

Dataset	Registration	σ_{rot} (deg)	σ_{trans} (mm)	σ_{test} (mm)
Phantom dataset	Multiple MR/MR	0.06	0.10	0.13
MR: $0.9 \times 0.9 \times 1$ mm^3	Multiple US/US	0.60	0.40	0.71
US: 0.41^3 mm^3	Conservative MR/US	1.15	1.01	1.46
	Expected MR/US	1.06	0.97	1.37
Baby dataset	Multiple US/US	0.10	0.06	0.12
MR: 0.9^3 mm^3	Conservative MR/US	1.21	0.36	0.90
US: 0.3^3 mm^3	Expected MR/US	1.21	0.36	0.89
Patient dataset	MR/MR	0.02	0.02	0.03
MR: $0.9 \times 0.9 \times 1.1$ mm^3	Conservative MR/US	1.57	0.58	1.65
US: 0.63^3 and 0.95^3 mm^3	Expected MR/US	1.57	0.58	1.65

TAB. 6.1 – *Accuracy results. These figures have to be compared with the voxel sizes of the MR images.*

6.4.5 Patient dataset

This dataset is an actual surgical case: two MR T1 images with and without a contrast agent were acquired before surgery. After craniotomy (dura mater still closed), a set of 3D US images was acquired to precisely locate the tumor to resect. In this experiment, we only use the three US images that are large enough to contain the ventricles. We could not register the others using our technique.

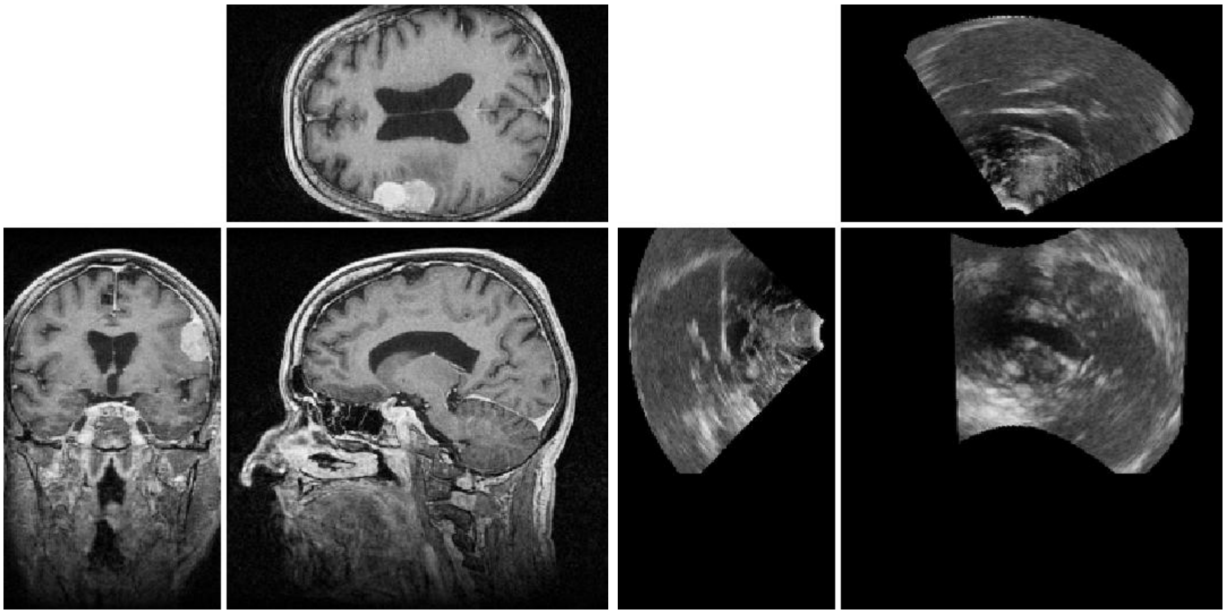
Here again, the two MR images were registered using the feature-based method (Pennec et al., 2000) with a very high accuracy (probably overestimated as we only have two images). We tested the loops $US_i \rightarrow MR_0 \rightarrow MR_1 \rightarrow US_i$.

6.4.6 Overall discussion

We put in table 6.1 the standard deviations of the residual rotation, of the residual translation and of the displacement of the test points for the different registration involved. Since we took the origin of the images at the center, the σ_{trans} value corresponds to the mean error at the center of the image while σ_{test} corresponds to the maximum registration error within the US image (the test points are taken at the corners of the Cartesian US image).

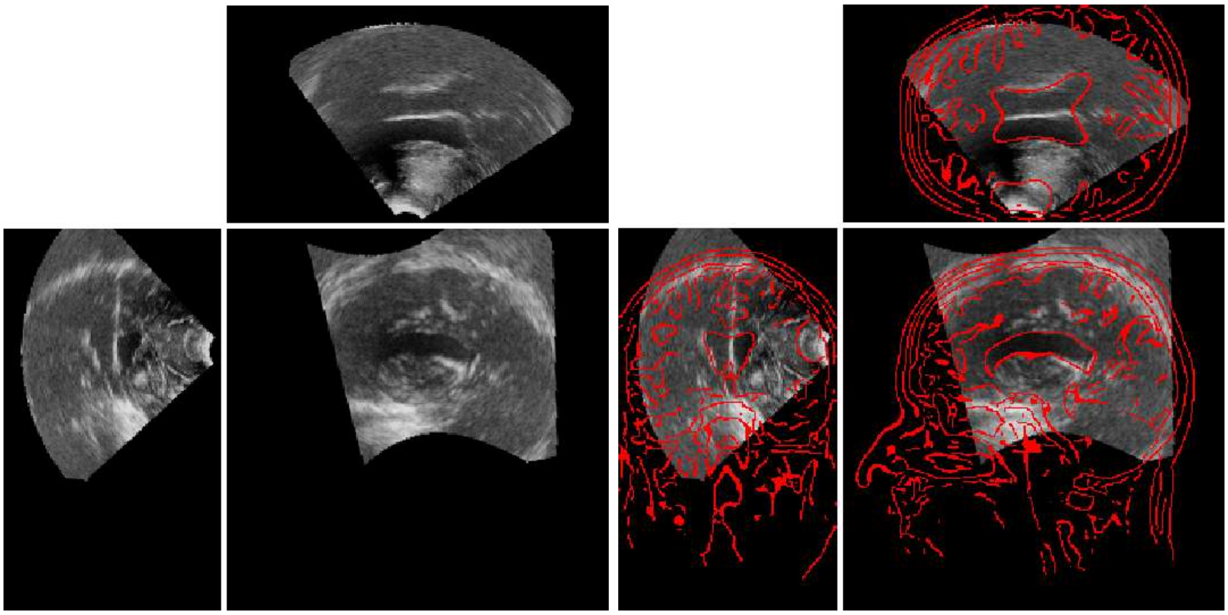
The results on the Phantom show that the MR/US registration accuracy is of the order of the MR resolution. One could probably expect a better conservative accuracy by acquiring larger US images including some rigid landmarks for multiple US/US registration. One finds the same type of results for the other datasets: slightly under the MR voxel size for the baby dataset and a bit larger for the patient dataset.

However, when we look more carefully at the patient results, we find that the loop involving the smallest US image (real size $150 \times 85 \times 100$ mm, voxel size 0.63^3 mm^3) is responsible for a



(a) Pre-operative MR

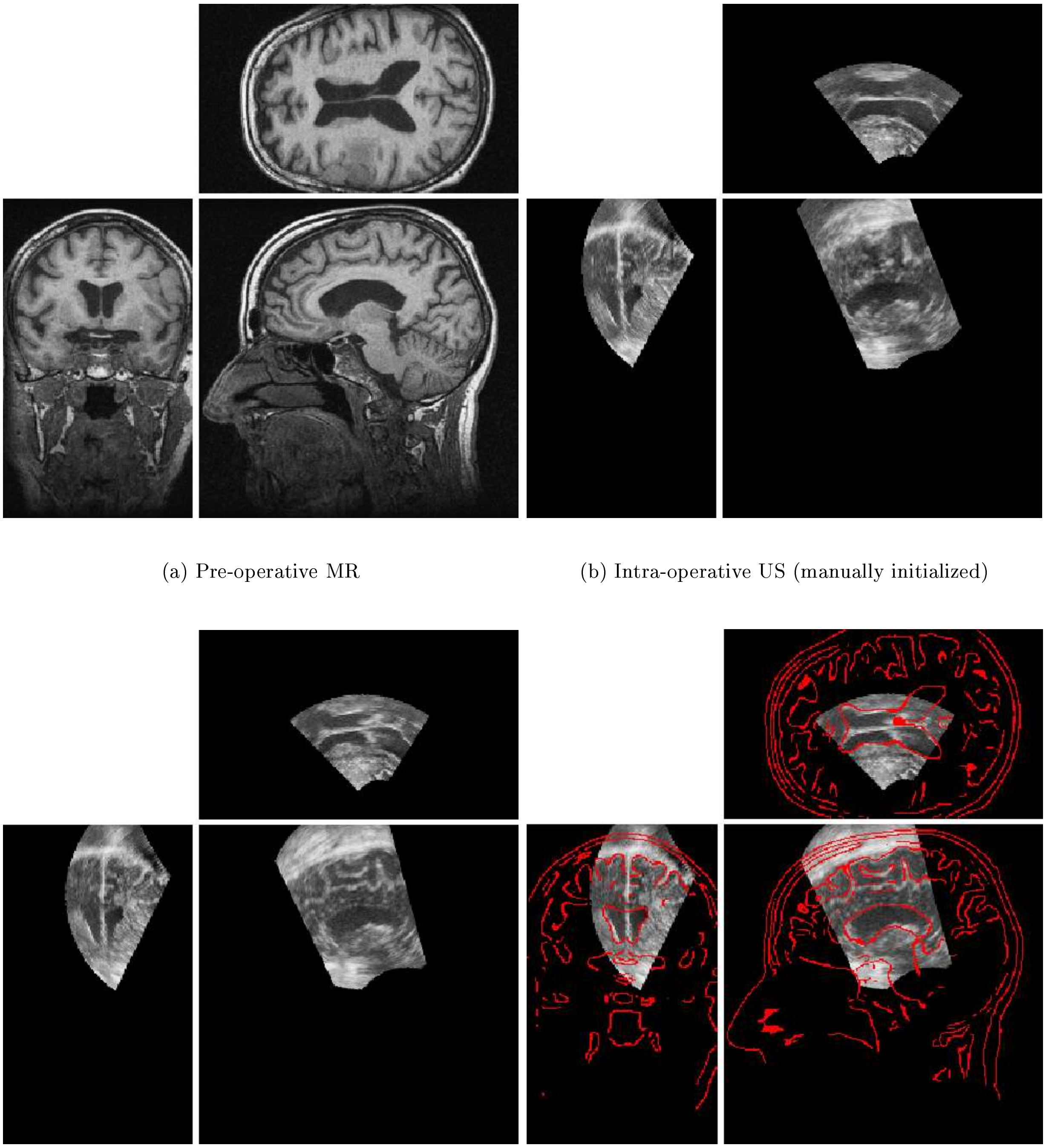
(b) Intra-operative US (manually initialized)



(c) Registered US volume

(d) Registered US volume with MR contours overlaid

FIG. 6.4 – *Example registration of the patient US/MR images. Notice that this MR image was acquired with a contrast agent to better delineate the tumor.*



(a) Pre-operative MR

(b) Intra-operative US (manually initialized)

(c) Registered US volume

(d) Registered US volume with MR contours overlaid

FIG. 6.5 – *Another example registration of the patient US/MR images. This MR is acquired without contrast agent. The US corresponds to the smallest acquisition cone among the dataset. This is the image we have found to be responsible for the largest registration error.*

corner error of 2.61 mm ($\sigma_{\text{trans}} = 0.84$ mm) while the loops involving the two larger US images (real size $170 \times 130 \times 180$, voxel size 0.95^3 mm^3) do have a much smaller corner error of about 0.84 mm ($\sigma_{\text{trans}} = 0.39$ mm). We suspect that a non-rigidity in the smallest US could account for the registration difference between the two MR images. Another explanation could be a misestimation of the sound speed for this small US acquisition leading to a false voxel size and once again the violation of the rigidity assumption.

6.5 Robustness study

6.5.1 Principle

The goal of this section is to study the consistency of our algorithm with respect to varying the initial transformation parameters. For that purpose, we chose one representative US and MR images in each dataset and registered them manually using an interactive matching tool. This gives us a rough "ground truth" transformation which is independent of any automated registration algorithm. Notice, although we performed the manual registration as carefully as possible, the "ground truth" does not need to be extremely accurate.

We then performed a number of automatic registrations by initializing the algorithm with random perturbations from the "ground truth" position: a rotation vector Δr with random direction and constant magnitude $\|\Delta r\| = 15$ degrees, and a translation vector Δt with random direction and constant magnitude $\|\Delta t\| = 20$ mm. These values hopefully correspond to a large upper bound of the manual registration errors that can be made by a non-expert. For each random transformation, registrations were alternatively performed using six different registration criteria:

- CR : the monovariate correlation ratio using the MR image as a template. We used the "standard" version corresponding to a squared intensity distance and a piecewise constant fitting function with 256 intervals (see chapter 3, section 3.3).
- CR* : the same as CR but with the MR gradient norm image as a template.
- MI : mutual information as implemented in (Maes et al., 1997), combining the US image and the MR image.
- MI* : mutual information combining the US image and the MR gradient norm image.
- BCR_s : the bivariate generalized correlation ratio as described in section 6.2, but using the squared intensity distance ($\rho(x) = x^2$) instead of the Geman-McClure ρ-function.
- BCR_g : the bivariate generalized correlation ratio using the Geman-McClure ρ-function. This is the similarity measure we used in the accuracy study (see section 6.4).

We performed 200 registrations for each registration criterion and for each MR/US pair (yielding a total of $200 \times 6 \times 3 = 3600$ registrations).

Doing this, we try to answer the question: "if we had used a different initialization, would the algorithm have found the same result?" In other words, we want to characterize: (1) the *robustness*, i.e. the ability of the algorithm to find the "same" minimum of the registration criterion from different initializations, and (2) the *precision*, i.e. the residual variation of the solution when convergence to the same minimum is declared. There is no doubt that the precision is strongly related to the interpolation method, which is here PV interpolation for each criterion.

This has not to be confused with the *accuracy* evaluation described in section 6.4, where we characterized the variability of the registration result with respect to the variability of the data. Here, we aim at estimating the variability of the registration result with respect to the variability of the initial transformation parameters, no matter this result corresponds to a good registration or not.

6.5.2 Computing the robustness and precision

In order to quantify both robustness and precision, we compute for each set of registrations a mean transformation. If the distance of a transformation to this mean is less than a given threshold, the algorithm is declared to be successful and this transformation is used to compute the repeatability (i.e. the variance with respect to the mean transformation), otherwise it is considered as a failure. More precisely, we used a robust version of the Fréchet mean transformation estimation presented in (Pennec, 1998) : using the same robust version of the left invariant distance between transformations as in section 6.4.2, the Fréchet mean rigid transformation is defined as:

$$\bar{T} = \arg\min_{T} \left(\sum_i \mathrm{dist}^2(T_i, T) \right)$$

This minimization is performed using a gradient descent described in (Pennec, 1998). However, this algorithm only gives us a local minimum around the starting point. Thus, to obtain the global minimum, we repeated this minimization for every transformation T_i as the starting point and we kept the best one.

Within this framework, the success rate is defined as the proportion of transformations that have a distance to $\bar{T}$ less than χ^2, and the precision values are computed on successful transformations using:

$$\sigma^2_{rot} = \frac{1}{N_{sucess}} \sum_i \|\bar{r}^{(-1)} \circ r_i\|^2 \qquad \text{and} \qquad \sigma^2_{trans} = \frac{1}{N_{sucess}} \sum_i \|\bar{t} - t_i\|^2$$

By the way, we should note that computing the Fréchet mean transformation maximizes the success rate since a success always account for less than a failure in the Fréchet minimization.

6.5.3 Results and discussion

The results of the robustness analysis are given in table 6.2. Experiments were reproduced with different χ^2 values (from 10 to 40) without significative differences.

Firstly, we should notice that the magnitude of the initial transformation perturbation corresponds to an upper bound of the acceptable initialization. This explains why no algorithm converges in 100 % of the cases.

The second observation is that no measure based only on the MR intensity or only on the MR gradient norm is robust for every dataset. The CR measure using intensity provides consistent results only for the patient images, while MI using intensity is unstable except for the baby images. When using the gradient norm information, these measures are useless for the patient images, although stable for the phantom and baby images. This suggests that combining both the intensity and gradient norm information is crucial in terms of robustness.

In all cases, the BCR measures (using either the square intensity distance or the Geman-McClure distance) perform at the best level with respect to precision, and yield acceptable success rates. It seems that BCR_s is a bit more robust than BCR_g: their success rates are comparable for the phantom and baby datasets, but mostly in favour of the former for the patient dataset. This may sound self-contradictory since the advantage of using the Geman-McClure function is precisely to achieve some robustness properties. However, one should not confuse the robustness to initialization with the robustness involved in the use of an S-estimator, which acts as reducing the sensitivity of the registration measure to intensity artifacts (see section 6.2.3).

A general observation is that quadratic versions of the correlation ratio seem to yield a wider

	Criterion	Success rate	Precision for successes		Visual inspection
			rot (deg)	trs (mm)	
Phantom dataset	MI	39 %	0.40	0.27	poor
	MI*	80 %	0.32	0.17	fair
	CR	52 %	0.43	0.25	poor
	CR*	71 %	0.25	0.15	fair
	BCR_s	76 %	0.08	0.04	ok
	BCR_g	76 %	0.14	0.09	ok
Baby dataset	MI	94 %	0.03	0.02	ok
	MI*	85 %	0.05	0.02	ok
	CR	14 %	0.32	0.10	poor
	CR*	79 %	0.02	0.01	ok
	BCR_s	68 %	0.12	0.01	ok
	BCR_g	71 %	0.02	0.01	ok
Patient dataset	MI	29 %	0.53	0.25	ok
	MI*	0 %	-	-	-
	CR	90 %	0.45	0.17	ok
	CR*	0 %	-	-	-
	BCR_s	85 %	0.39	0.11	ok
	BCR_g	55 %	0.43	0.08	ok

TAB. 6.2 – *Robustness results for $\sigma_r = 0.2$ degrees, $\sigma_t = 0.1$ mm and $\chi^2 = 18$.*

attraction basin than those based on a robust intensity distance. However, we also observed from practical experience that registration results using $\mathrm{BCR_g}$ seems to be a little more accurate than when using $\mathrm{BCR_s}$. Although further quantitative evaluation is needed, we believe at present that it is safe to use $\mathrm{BCR_g}$ provided that a fine initialization has been achieved.

Of course, these figures do not tell us anything about the ability of the different measures to find a *good* registration. They only provide information regarding the consistency of the results with respect to the initial transformation parameters. For each set of experiments, we inspected the registration result corresponding to the mean transformation as computed in section 6.5.2. Registration was declared 'ok', 'fair', or 'poor', respectively if there was no obvious misalignment, if a slight misalignment could be seen, or if the result was clearly wrong.

From this subjective evaluation (also reported in table 6.2), we observe that there is a correlation between the accuracy and robustness performances of the different registration criteria, which was not obvious *a priori*. Here again, the two bivariate measures are the only ones to provide satisfying results across experiments.

6.6 Conclusion

We have presented a new automated method to rigidly register 3D US with MR images. It is based on a bivariate version of the generalized correlation ratio that allows to better take into account the differential nature of US images. We have related our similarity measure to a rough dependence model of the MR and US signals.

Testings were performed on several phantom and clinical data, and accuracy was evaluated using an original method that does not require the knowledge of a ground truth. We estimated the worst registration errors (errors at the US corners) to be of the order of 1 millimeter.

In our experiments, we used US images stored in Cartesian format. This obviously does not help registration owing to the fact that:

1. intensities of voxels distant from the probe are unreliable,
2. resampling on a cubic lattice artificially increases the concentration of such voxels, consequently increasing their influence on the registration criterion.

We believe that using polar US images as inputs to our registration algorithm could significantly improve both its accuracy and robustness.

Further developments also include non-rigid registration, in order to correct for false distance artefacts in US, as well as for tissue deformations due to brain shift and operative manipulations. The next chapter is a preliminary work in this direction.

Chapitre 7

Une adaptation multimodale de l'algorithme des «démons»

7.1 Introduction

Au chapitre 3, nous avons justifié le rapport de corrélation en tant que mesure de similarité appropriée au recalage multimodal d'images. Les applications que nous en avons présenté par la suite concernaient exclusivement des problèmes de recalage rigide. Et pour cause : l'implémentation décrite au chapitre 3 repose sur la méthode de Powell, une procédure d'optimisation qui n'est adaptée en pratique qu'à des espaces de recherche de «petite» dimension. Le travail que nous présentons ici est parti d'une question somme toute inévitable : le rapport de corrélation est-il applicable au recalage multimodal non-rigide ?

Il est difficile de donner une signification quantitative à la notion de «petite» dimension. De fait, la méthode de Powell permet d'explorer des espaces de recherche plus grands que celui des transformations rigides. Nous avons implémenté une version affine du recalage par rapport de corrélation qui fonctionne convenablement et dans des temps de calcul raisonnables. Il est probable que l'on puisse aller encore un peu plus loin avec des classes de transformations paramétriques du type de celles qui ont été évoquées au chapitre 2, paragraphe 2.4.5.

Mais ce que nous entendons habituellement par «transformation non-rigide» est synonyme d'expressions telles que : champ de déformation, transformation élastique, transformation libre, etc. Il s'agit implicitement de transformations dont le nombre de degrés de liberté est extrêmement élevé ou, plus fréquemment, de transformations non-paramétriques. Dans ce cas, il ne fait aucun doute que la méthode de Powell est inutilisable. La procédure d'optimisation n'est cependant pas seule en cause. Nous avons déjà montré au chapitre 2, paragraphe 2.4.6, qu'optimiser une mesure de similarité sur un espace de recherche «trop grand» est un problème mal posé. Il est alors nécessaire de formuler certains *a priori* concernant la transformation recherchée, autrement dit : régulariser le problème.

Bien entendu, la question ne se pose pas uniquement à propos du rapport de corrélation. Plusieurs auteurs ont déjà cherché à étendre l'information mutuelle au recalage multimodal non-rigide (Maintz et al., 1998; Gaens et al., 1998; Hellier et Barillot, 2000). Tous font appel à des techniques d'appariement de régions dont la philosophie consiste à mener la recherche dans des espaces imbriqués de transformations constantes par morceaux (éventuellement, rigides ou affines par morceaux). Cette approche est également retenue par (Lau et al., 1999) mais couplée avec le rapport de corrélation, un choix que nous ne pouvons que cautionner.

Les techniques d'appariemment de région offrent un cadre naturel pour raffiner progressivement l'estimation de la transformation et contrôler ainsi sa régularité. Néanmoins, comme nous l'avons signalé au chapitre 2, paragraphe 2.5.3.2, il est difficile de savoir en quoi elles résolvent effectivement un problème d'optimisation.

7.1.1 Principe général de la méthode

Notre méthode se fonde sur une approche différente. Nous avons montré au chapitre 3 que la maximisation du rapport de corrélation peut être mise en œuvre au moyen d'un schéma itératif alternant entre une étape de correction d'intensité et une étape de recalage. Étant données deux images I (source) et J (cible), ce schéma de résolution correspond à une minimisation alternée de la fonctionnelle :

$$\tilde{S}(T,f) = \frac{1}{n^{\downarrow}\mathrm{Var}(I^{\downarrow})} \sum_k \left(i_k - f(j_k^{\downarrow}) \right)^2, \tag{7.1}$$

dont le minimum à f fixé définit précisément le rapport de corrélation :

$$1 - \eta^2_{I|J}(T) \equiv \min_f \tilde{S}(T,f)$$

Dans le contexte du recalage rigide/affine, la minimisation par rapport à T peut être accomplie par une méthode standard comme celle de Powell. C'est en premier lieu cette étape qui pose problème lorsque l'on envisage un recalage non-rigide. Or on remarque que si f est considérée comme fixée, le critère $\tilde{S}(T,f)$ s'apparente à une somme de différences d'intensité au carré (SDC), à la division près par $n^{\downarrow}\mathrm{Var}(I^{\downarrow})$. Rappelons que la raison d'être de cette normalisation est de garantir la robustesse de la mesure au recouvrement partiel entre les images.

En somme, minimiser $\tilde{S}(T,f)$ en fonction de T (à f fixée) revient à résoudre un problème de recalage monomodal, le rôle de la fonction f étant justement de rendre comparables les intensités des deux images. Cette remarque nous a servi de point de départ. Il se trouve que le projet Epidaure dispose d'un excellent algorithme de recalage monomodal non-rigide, dit algorithme des démons, dont le développement a été initié par (Thirion, 1995; Thirion, 1998) et poursuivi au cours des dernières années par quelques collègues (Guimond, 1999; Cachier et Pennec, 2000; Cachier et Rey, 2000).

Mieux encore, plusieurs auteurs (Bro-Nielsen et Gramkow, 1996; Cachier et al., 1999; Pennec et al., 1999; Guimond et al., 2001) s'accordent à considérer l'algorithme des démons comme une méthode de minimisation de la SDC pénalisée au sens de Tikhonov :

$$\min_T \left[\sum_k \left(i_k - j_k^{\downarrow} \right)^2 + \lambda U(T) \right], \tag{7.2}$$

où λ est une constante positive et $U(T)$ un stabilisateur destiné à dissuader le comportement «oscillatoire» de la transformation. Ainsi, l'algorithme des démons peut remplacer la méthode de Powell dans notre schéma de minimisation alternée. Ceci correspond à redéfinir légèrement la fonctionnelle à minimiser (7.1) en ignorant les termes de normalisation et en ajoutant un potentiel de régularisation :

$$\min_{T,f} \left[\sum_k \left(i_k - f(j_k^{\downarrow}) \right)^2 + \lambda U(T) \right]$$

Le fait de supprimer la quantité $n^{\downarrow}\mathrm{Var}(I^{\downarrow})$ est légitime. En pratique, le recalage non-rigide fait suite à un recalage affine dont le rôle est de compenser grossièrement la composante «macroscopique» du déplacement relatif des images. Le recalage non-rigide intervient comme une étape de raffinement. Il s'agit alors d'estimer des déformations résiduelles, plus chaotiques mais de faible amplitude. Dans ce contexte, nous pouvons ignorer l'influence des termes de normalisation dans la mesure où la zone de recouvrement n'est plus susceptible de varier notablement.

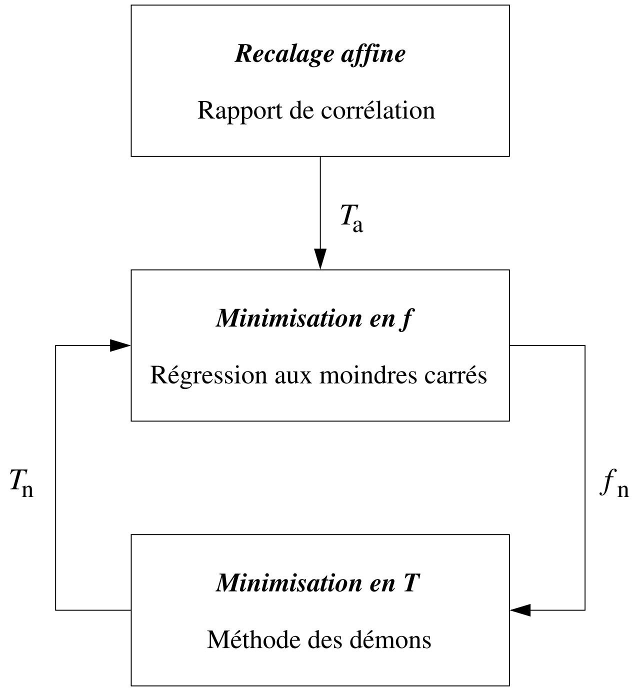


FIG. 7.1 – *Schéma théorique de l'algorithme de recalage non-rigide fondé sur la maximisation du rapport de corrélation.*

Nous aboutissons finalement au schéma de résolution représenté par la figure 7.1. Il s'agit d'une base théorique à laquelle nous avons apporté un certain nombre d'adaptations pratiques motivées par les observations suivantes :

1. La méthode des démons n'est pas *strictement* équivalente à la formulation de Tikhonov correspondant à l'équation (7.2). L'une et l'autre souffrent de défauts que nous discuterons au paragraphe 7.2. Il apparaîtra judicieux de combiner leurs avantages respectifs.
2. La minimisation du critère à T fixé équivaut à une régression aux moindres carrés qui s'avère insuffisamment robuste dans le contexte du recalage non-rigide. Nous proposons plusieurs idées pour améliorer cette étape au paragraphe 7.3.

7.1.2 Plan du chapitre

Les paragraphes 7.2 et 7.3 détaillent les méthodes que nous utilisons en pratique pour estimer respectivement T à f fixé et f à T fixé. Nous avons choisi cet ordre de présentation car il nous semble plus facile de cerner les difficultés liées à la normalisation des intensités une fois que la procédure de mise en correspondance spatiale est clarifiée. Il doit cependant être clair que chaque itération de l'algorithme commence par l'étape de correction d'intensité.

Le paragraphe 7.4 décrit une implémentation non-rigide de l'information mutuelle fondée sur les aspects théoriques abordés au paragraphe 7.2. Enfin, nous présentons et discutons aux paragraphes 7.5 et 7.6 plusieurs résultats de recalage multimodal non-rigide entre images cérébrales de sujets différents.

7.2 Estimation de la transformation spatiale

Dans ce paragraphe, la fonction de correction f est considérée comme fixée et nous cherchons à améliorer l'estimée de la tranformation spatiale T. Pour simplifier les notations, nous assimilerons l'image corrigée $f(J)$ à l'image J elle-même. Par ailleurs, nous supposons avoir estimé au préalable la composante affine T_a de la transformation, de sorte que seul le champ de déformations résiduelles est encore amené à évoluer :

$$\mathbf{u}(\mathbf{x}) \equiv T(\mathbf{x}) - T_a(\mathbf{x})$$

7.2.1 Minimisation de la SDC pénalisée

Tel qu'il est formulé par l'équation (7.2), notre problème consiste à minimiser un critère de la forme :

$$S(T) + \lambda U(T), \tag{7.3}$$

où $S(T)$ est la SDC calculée entre l'image I et l'image J (corrigée) et $U(T)$ est un stabilisateur que nous supposerons quadratique et séparable. Intuitivement, nous cherchons des déformations résiduelles aussi peu chaotiques que possible. Nous pouvons dans ce but utiliser un stabilisateur de la forme :

$$U(T) = V(T - T_a) = \frac{1}{2} \sum_{k=1}^{3} \int_{\mathbb{R}^3} \frac{|\hat{T}^i(\mathbf{s}) - \hat{T}_a^i(\mathbf{s})|^2}{\hat{G}(\mathbf{s})} \, d\mathbf{s} \tag{7.4}$$

où $\hat{T}^i(\mathbf{s})$ (resp. $\hat{T}_a^i(\mathbf{s})$) désigne la transformée de Fourier de la i-ème application coordonnée de T (resp. T_a) tandis que $\hat{G}(\mathbf{s})$ est la transformée de Fourier d'un filtre passe-bas $G : \mathbb{R}^3 \to \mathbb{R}$. Un choix simple pour G est la fonction gaussienne :

$$G_\sigma(\mathbf{x}) = e^{-\frac{\|\mathbf{x}\|^2}{\sigma^2}},$$

dont on note qu'elle fait intervenir un paramètre σ qu'il s'agira de régler convenablement.

Au paragraphe 2.4.6 du chapitre 2, nous avons établi la solution générale d'un problème de minimisation du type (7.3) :

$$T(\mathbf{x}) = T_0(\mathbf{x}) + \sum_k G_\sigma(\mathbf{x} - \mathbf{x}_k)\, \mathbf{c}_k \qquad \text{avec} \quad \mathbf{c}_k \equiv -\frac{1}{\lambda}\frac{\partial S}{\partial T_k}(T) \tag{7.5}$$

Dans cette formule, T_0 est une certaine transformation appartenant au noyau du stabilisateur, c'est-à-dire vérifiant :

$$\frac{\partial U}{\partial T}(T_0) = 0$$

Dans notre cas particulier, on peut vérifier que le seul candidat possible est la transformation T_a. Nous avons donc nécessairement : $T_0 = T_a$.

Il faut bien comprendre par ailleurs que l'équation (7.5) définit T de manière implicite puisque les vecteurs $\mathbf{c}_k$ dépendent de T. Chaque $\mathbf{c}_k$ est le gradient de la SDC par rapport au déplacement individuel $T_k \equiv T(\mathbf{x}_k)$ du voxel $\mathbf{x}_k$. On constate que le calcul de ces vecteurs suppose que l'image J soit différentiable, ce qui n'a en principe aucun sens puisqu'une image est une entité discrète. Rien ne nous interdit cependant de plonger l'image J dans un espace continu au moyen d'une méthode quelconque d'interpolation (voir le chapitre 2, paragraphe 2.3.2). A condition de choisir une méthode suffisamment régulière, nous obtenons :

$$\frac{\partial S}{\partial T_k} = -2(i_k - j_k^\downarrow)\, \nabla J(T(\mathbf{x}_k))$$

En posant $\nabla J_k^\downarrow \equiv \nabla J(T(\mathbf{x}_k))$, nous concluons que la transformation qui minimise la SDC pénalisée est solution de l'équation implicite suivante :

$$T(\mathbf{x}) = T_a(\mathbf{x}) + \frac{2}{\lambda}\sum_k G_\sigma(\mathbf{x} - \mathbf{x}_k)\, (i_k - j_k^\downarrow)\nabla J_k^\downarrow \tag{7.6}$$

Cette expression correspond assez bien à ce que l'on attend. Elle fait clairement apparaître la déformation résiduelle comme un terme correctif ajouté à la composante affine. Le champ des déformations résiduelles n'est autre que la convolution par une gaussienne d'un champ de «forces» dont les points d'applications sont les voxels de l'image source :

$$\mathbf{u}(\mathbf{x}) = G_\sigma \otimes \mathbf{f}(\mathbf{x},T) \qquad \text{avec} \quad \mathbf{f}(\mathbf{x},T) \equiv \delta(\mathbf{x} - \mathbf{x}_k)\, \frac{2}{\lambda}\, (i_k - j_k^\downarrow)\nabla J_k^\downarrow \tag{7.7}$$

Ces forces sont «créées» localement par l'existence d'un écart d'intensité entre le voxel $\mathbf{x}_k$ de l'image I et son point coïncidant dans l'image J. La force est en outre proportionnelle au gradient de J calculé au point coïncidant. On comprend ainsi que ce sont essentiellement les contours de l'image cible qui vont gouverner le recalage.

L'équation (7.6) définit T sous la forme implicite $T = F(T)$ avec $F(T) = T_a(\mathbf{x}) + G_\sigma \otimes \mathbf{f}(\mathbf{x},T)$. Autrement dit, T est un point fixe de la fonction F, ce qui suggère une méthode de résolution itérative du type :

$$T_{n+1}(\mathbf{x}) = T_a(\mathbf{x}) + G_\sigma \otimes \mathbf{f}(\mathbf{x},T_n)$$

ou, de manière équivalente :

$$\mathbf{u}_{n+1}(\mathbf{x}) = G_\sigma \otimes \mathbf{f}(\mathbf{x},T_n) \tag{7.8}$$

Ce schéma itératif est illustré par la figure 7.2. Nous allons voir maintenant qu'il est étroitement lié à l'algorithme des démons.

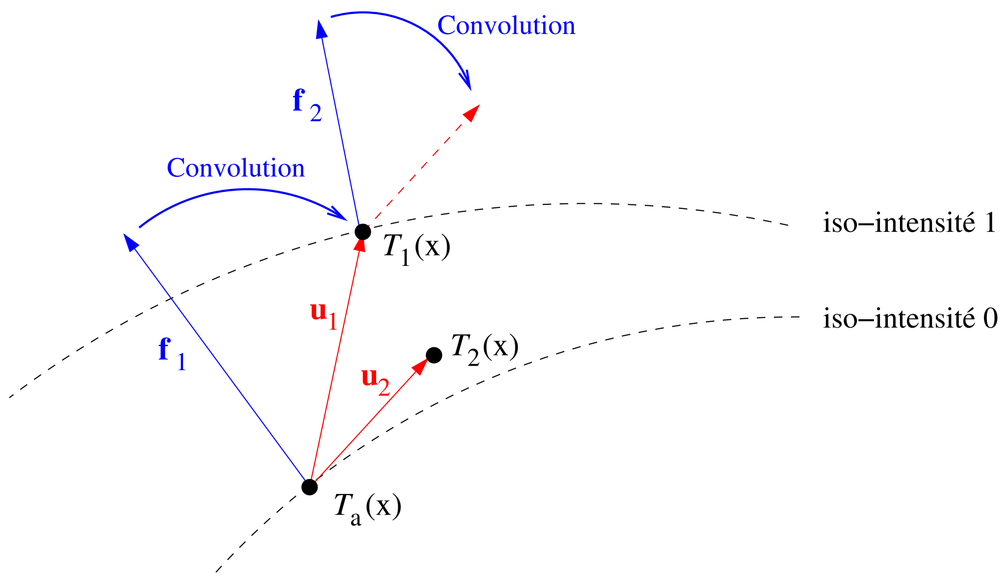


FIG. 7.2 – *Schéma itératif correspondant à la minimisation de la SDC pénalisée.*

7.2.2 L'algorithme des démons

En se fondant sur des considérations de physique statistique[1] sans rapport évident avec la minimisation de la SDC, (Thirion, 1995; Thirion, 1998) a proposé un schéma itératif qui s'avère similaire à l'équation (7.8) :

$$\mathbf{u}_{n+1}(\mathbf{x}) = G_\sigma \otimes [\mathbf{u}_n(\mathbf{x}) + \mathbf{f}(\mathbf{x},T_n)] \tag{7.9}$$

La comparaison des figures 7.2 et 7.3 permet de bien comprendre en quoi ces schémas diffèrent. A la première itération, le déplacement d'un point $\mathbf{x}$ est le même dans les deux cas, mais ce n'est pas vrai aux itérations suivantes. Dans le premier schéma, le point repart systématiquement de sa position initiale, $T_a(\mathbf{x})$. Tout se passe en fait comme si la force s'appliquait toujours au même

1. L'appelation démons provient d'une analogie avec les démons de Maxwell, paradoxe resté célèbre dans l'histoire de la thermodynamique.

endroit tout en variant au cours du temps. On conçoit que cet effet de rappel conduit à limiter fortement l'amplitude des déplacements.

Dans le deuxième schéma, au contraire, un point garde en mémoire le chemin qu'il a parcouru au fil des itérations. On peut faire une analogie avec une particule évoluant sous l'effet de chocs successifs. Un point pourra ainsi s'éloigner davantage de sa position initiale. En pratique, nous avons une nette tendance à préférer le schéma des démons car il autorise plus de déformations résiduelles, ce qui s'avère souhaitable en particulier dans le contexte du recalage inter-patients.

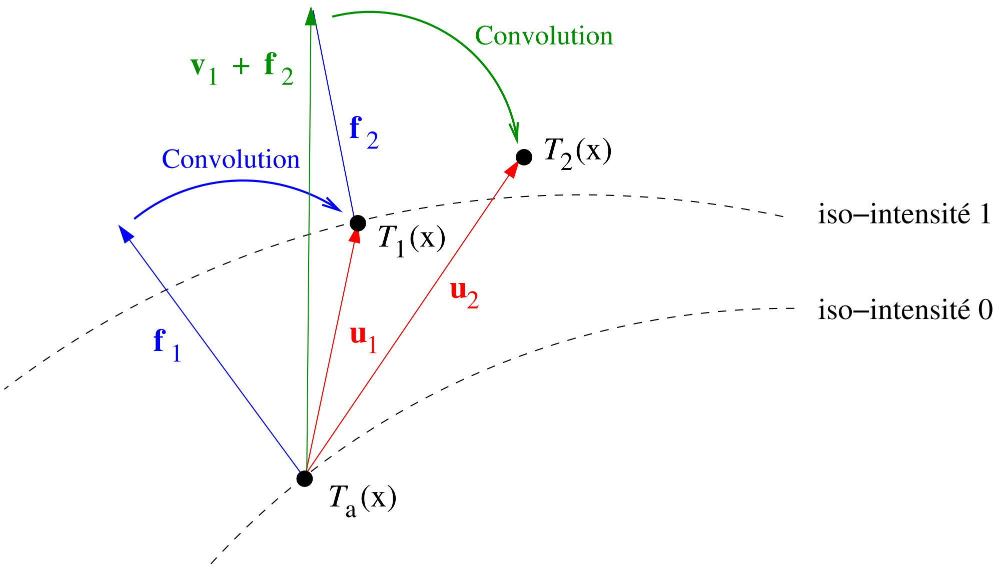


FIG. 7.3 – *Schéma itératif des démons.*

Nous avons raisonné jusqu'ici comme si les champs de forces utilisés par les deux schémas étaient les mêmes. Thirion propose en fait des forces sensiblement différentes de celles qui correspondent à la minimisation de la SDC pénalisée :

$$\mathbf{f}(\mathbf{x},T) \equiv \delta(\mathbf{x} - \mathbf{x}_k)\, \frac{(i_k - j_k^{\downarrow})\nabla I_k}{\|\nabla I_k\|^2 + (i_k - j_k^{\downarrow})^2}, \tag{7.10}$$

à comparer avec l'équation (7.7).

La principale similitude entre ces forces est qu'elles dépendent de l'écart d'intensité $(i_k - j_k^{\downarrow})$. Toutes deux font également intervenir un gradient d'image mais, et c'est là une première différence importante, les forces proposées par Thirion utilisent le gradient de l'image source ∇I_k et non celui de l'image cible $\nabla J_k^{\downarrow}$ comme le suggère l'équation (7.7). Lorsque les images sont correctement recalées, ces deux gradients doivent être approximativement égaux. Nous avons cependant observé expérimentalement que l'utilisation de $\nabla J_k^{\downarrow}$ accélère sensiblement la convergence de l'algorithme.

Sur la base de ces observations et de l'argument théorique lié à la minimisation de la SDC pénalisée, nous avons modifié les forces initialement proposées par Thirion en remplaçant simplement

∇I_k par $\nabla J_k^{\downarrow}$:

$$\mathbf{f}(\mathbf{x},T) \equiv \delta(\mathbf{x} - \mathbf{x}_k) \frac{(i_k - j_k^{\downarrow})\nabla J_k^{\downarrow}}{\|\nabla J_k^{\downarrow}\|^2 + (i_k - j_k^{\downarrow})^2} \tag{7.11}$$

On voit alors que l'équation (7.7) peut être rendue formellement équivalente à l'équation (7.11) à condition de poser :

$$\lambda = 2 \left[\|\nabla J_k^{\downarrow}\|^2 + (i_k - j_k^{\downarrow})^2 \right]$$

Cette équivalence est cependant très artificielle. Dans la formulation du problème de minimisation (7.3), λ était considérée comme une constante. Ici, λ dépend de la transformation et n'est plus uniforme dans l'image source. Les forces de l'équation (7.11) ne découlent donc pas d'une minimisation d'énergie du type (7.3). En pratique, le choix d'une valeur convenable de λ s'avère cependant problématique en termes de stabilité numérique. C'est pourquoi nous avons choisi de conserver l'approche du «λ variable» héritée de Thirion.

Au final, notre méthode est une variante légère de l'algorithme original des démons. Celui-ci nous a servi d'inspiration pour adapter la stratégie fondée sur la minimisation de la SDC pénalisée. Si nous gagnons en performance, nous perdons aussi la formulation énergétique qui faisait la base théorique de notre approche. Il est en effet difficile de relier l'algorithme des démons à une minimisation d'énergie.

Nos collègues (Cachier et al., 1999; Pennec et al., 1999) ont étudié cette question sous un angle légèrement différent du nôtre. Ils établissent notamment que le «λ variable» de Thirion est une approximation naturelle du hessien de la SDC et ils relient ainsi l'agorithme des démons à une descente de gradient du 2ème ordre sur la SDC. Toutefois, ce formalisme ne permet pas de retrouver l'ensemble du schéma itératif comme résultant de la minimisation d'une seule et même énergie. La convolution gaussienne correspond à la minimisation d'une fonction objectif concurrente de la SDC, si bien que les dernières publications de (Cachier et Pennec, 2000) tendent à faire un parallèle avec la théorie des jeux plutôt qu'avec une formulation énergétique de type Tikhonov.

7.2.3 Déformations fluides

Nous ouvrons ici une petite parenthèse afin de mieux situer l'algorithme des démons dans l'ordre des méthodes de recalage de type flux optique (Horn et Schunck, 1981; Barron et al., 1994; Lester, 1999). Revenons un instant sur les schémas itératifs proposés par les équations (7.8) et (7.9) :

- *Minimisation de la SDC pénalisée :* $\mathbf{u}_{n+1} = G_\sigma \otimes \mathbf{f}_n$
- *Démons :* $\mathbf{u}_{n+1} - G_\sigma \otimes \mathbf{u}_n = G_\sigma \otimes \mathbf{f}_n$

En ignorant le fait que les forces utilisées par ces deux schémas ne sont pas nécessairement les mêmes, nous observons que la seule différence formelle réside dans l'existence du terme $G_\sigma \otimes \mathbf{u}_n$ dans la formule des démons. Dans le premier schéma, le champ de force (régularisé) agit directement sur le champ de déplacements alors que dans le second, il agit sur le champ différentiel $\mathbf{u}_{n+1} - G_\sigma \otimes \mathbf{u}_n$.

On peut se demander ce que représente ce champ. S'il ne faisait pas intervenir l'opérateur de convolution, on pourrait l'interpréter comme un champ de vitesses :

$$\mathbf{v}_{n+1} = \mathbf{u}_{n+1} - \mathbf{u}_n$$

Cette remarque suggère un troisième schéma itératif du type :

$$\mathbf{u}_{n+1} - \mathbf{u}_n = G_\sigma \otimes \mathbf{f}_n \tag{7.12}$$

Autrement dit, au lieu d'influencer directement le champ de déplacements comme c'est le cas dans le schéma correspondant à la SDC, les forces agissent sur le champ de vitesses. C'est en substance l'idée proposée par (Christensen, 1994; Christensen et al., 1996) dans sa formulation du recalage d'images par déformations fluides. La figure 7.4 permet de visualiser la façon dont les déplacements sont estimés dans un schéma de ce type. Comme dans le schéma des démons (voir la figure 7.3), un point garde la mémoire de sa trajectoire. La différence avec les démons concerne l'ordre dans lequel l'addition des déplacements et la convolution sont effectuées. Dans le schéma des démons, la convolution intervient *après* que le nouveau déplacement a été ajouté au précédent ; c'est le contraire dans le cas des déformations fluides.

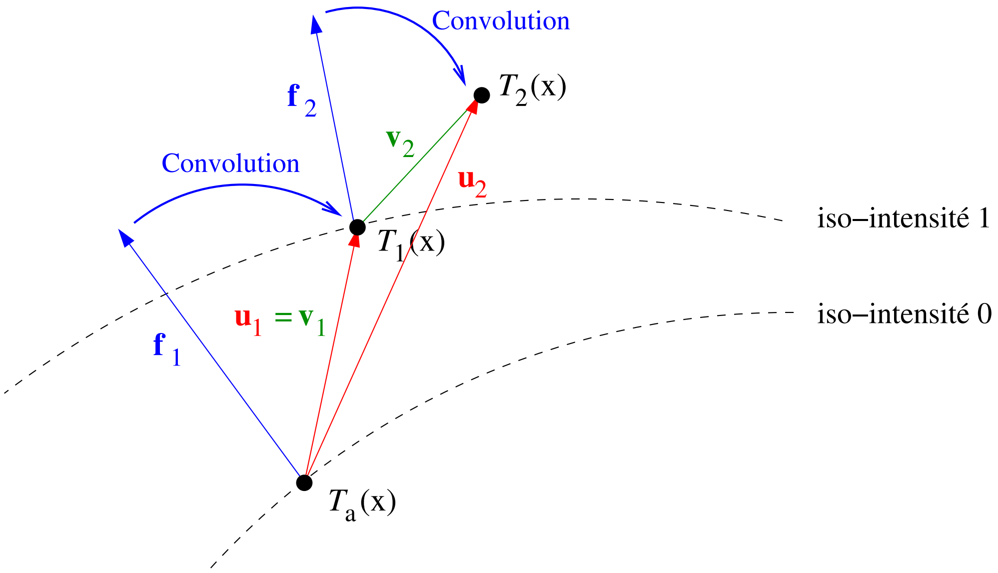


FIG. 7.4 – *Schéma itératif correspondant à des déformations fluides.*

L'algorithme des démons apparaît ainsi comme un cas intermédiaire entre une approche dynamique comme celle des déformations fluides et une approche énergétique comme celle de la SDC pénalisée. Il est d'ailleurs intéressant de remarquer que Christensen a étudié une formulation «élastique» analogue à la minimisation de la SDC pénalisée avant de proposer son recalage fluide. C'est en constatant la capacité de cette dernière méthode à déformer davantage les images qu'il a finalement abandonné la première. Il semble qu'en pratique, le schéma correspondant à des déformations fluides converge difficilement, ce qui nous incite à lui préférer le schéma des démons.

7.3 Estimation de la transformation d'intensité

Dans ce paragraphe, la transformation spatiale T est maintenant fixée et nous cherchons à estimer la fonction d'intensité f qui réalise le meilleur ajustement de l'image I par l'image J. De façon plus générale, il s'agit de corriger J de manière à ce qu'elle «ressemble» le plus possible à I. Nous verrons plus loin qu'une simple correction mono-fonctionnelle n'est pas toujours suffisante.

En pratique, la correction d'intensité est régulièrement mise à jour après un nombre fixe d'itérations du schéma des démons. Dans la version luxueuse du programme, ce nombre vaut 1, ce qui signifie que la correction d'intensité est réestimée à chaque fois que le champ de déplacements évolue. Il est probable qu'une cadence aussi élevée ne soit pas nécessaire, ce qui permettrait de gagner en temps de calcul.

Nous définissons l'ensemble $\mathcal{C}^{\downarrow}$ des couples d'intensité induits par la transformation T :

$$\mathcal{C}^{\downarrow} = \left\{ \left(i_k, j_k^{\downarrow}\right),\ 1 \leq k \leq n \right\},$$

où $n = \mathrm{Card}\Omega_I$ est le nombre de voxels de l'image source. Les valeurs $j_k^{\downarrow}$ sont calculées en pratique par interpolation trilinéaire et éventuellement extrapolées à 0 lorsque un voxel $\mathbf{x}_k$ est transformé en dehors du domaine délimité par la grille cible Ω_J.

7.3.1 Monofunctional dependence assumption

Our goal is to model the transformation that characterizes the mapping from voxel intensities in J to those in I, knowing that some elements of $\mathcal{C}^{\downarrow}$ are erroneous, i.e. that would not be present if the images were perfectly matched. If we can assume a monofunctional dependence of the intensities of I with regards to those of J as well as additive stationary Gaussian white noise ε on the intensity values of I, then we can adopt the model

$$i_k = f(j_k^{\downarrow}) + \varepsilon_k, \tag{7.13}$$

where f is an unknown function to be estimated. This is exactly the model which leads to the correlation ratio as the measure to be maximized for registration. In that approach, for a given transformation, one seeks the function that best describes I in terms of J. It is shown that, in a maximum likelihood context, the intensity function $\hat{f}$ that best approximates f is a least squares (LS) fit of I in terms of J.

Here the major difference is that we seek a high-dimensional geometrical transformation. As opposed to affine registration where the transformation is governed by the majority of good matches, we have seen in section 7.2 that using the elastic registration model, displacements are found using mainly local information (i.e. gradients, local averages, etc.). Hence, we can not expect good displacements in one structure to correct for bad ones in another; we have to make certain each voxel is moved properly during each iteration. For this, since the geometrical transformation is found using intensity similarity, the most precise intensity transformation is required. Consequently, instead of performing a standard least squares fit, we have opted for a robust estimator which will remove outlying elements of $\mathcal{C}^{\downarrow}$ during the estimation of the intensity transformation.

To estimate f we use the least trimmed squares (LTS) method followed by a binary reweighted least squares (RLS) estimation (Rousseeuw et Leroy, 1987). The combination of these two methods provides a very robust regression technique with outliers detection, while ensuring that a maximum of pertinent points are used for the final estimation.

Least trimmed squares For our particular problem, we will constrain the unknown function f to be a polynomial function with degree p:

$$f(j) = \theta_0 + \theta_1 j + \theta_2 j^2 + \cdots + \theta_p j^p,$$

where we need to estimate the polynomial coefficients $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_p]$. A regression estimator will provide a $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \ldots, \hat{\theta}_p]$ which can be used to predict the value of i_k from $j_k^{\downarrow}$.

Given a trial function f, let us define the residual errors :

$$r_k \equiv i_k - f(j_k^{\downarrow}),$$

where we omit the arrow notation for clarity. The method to obtain $\hat{\boldsymbol{\theta}}$ which is consistent with the correlation ratio measure is to minimize the sum of squared residual errors,

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{n} r_k^2,$$

which leads to the standard LS solution. It is found by solving a linear system using a singular value decomposition (SVD). The LS estimation is known to be very sensitive to outliers and thus it is expected to provide a poor estimate of the monofunctional mapping from J to I. The LTS method solves this problem by minimizing the same sum on a subset of all residual errors, thus rejecting large ones corresponding to outliers,

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{h} r_{k:q}^2,$$

where $r_{k:q}^2$ is the q^{th} smallest value of the set $\{r_1^2, \ldots, r_n^2\}$. In some way, this corresponds to a standard LS on the h values that best approximate the function we are looking for. However these values are not determined since the function is unknown. Essentially, h/n represents the percentage of "good" points in $\mathcal{C}^{\downarrow}$ and must be at least 50%. A lesser value would allow to estimate parameters that model a minority of points which could then all be outliers. The value of h will vary according to the modalities used during registration. Assigning actual values to h is postponed to section 7.6.

Our method for LTS minimization is a simple iterative technique. First, we randomly pick h points from $\mathcal{C}^{\downarrow}$. We then iterate between calculating $\hat{\boldsymbol{\theta}}$ using the standard LS technique on the selected points and updating the h closest points from $\mathcal{C}^{\downarrow}$. This process is carried out until convergence, usually requiring less than 5 iterations and is guaranteed to find at least a local minimum of the LTS criterion (Rousseeuw et Van Driessen, 1999).

Reweighted least squares As discussed in (Rousseeuw et Leroy, 1987), the LTS method is very robust, but it tends to provide an estimate $\hat{\boldsymbol{\theta}}$ that is notably less accurate than the one we would obtain with a standard LS in the absence of outliers. The solution may be refined by considering all the points that relate well to the LTS estimate, not only a fixed fraction of them. An efficient technique to achieve this is the so-called RLS regression (Rousseeuw et Leroy, 1987), which minimizes the sum of squared residuals over all the points that do not stand "too far" from the LTS estimate,

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{n} w_k r_k, \quad \text{where} \quad w_k = \begin{cases} 1 & \text{if } r_k \leq 3\hat{\sigma}, \\ 0 & \text{otherwise}, \end{cases}$$

where $\hat{\sigma}$ is a scale parameter which actually estimates the standard deviation of the Gaussian noise termed ε that we introduced in equation (7.13). Such an estimate can be computed directly from the final value of the LTS criterion,

$$\hat{\sigma} = \sqrt{\frac{K}{h} \sum_{k=1}^{h} r_{k:q}^2}, \qquad \text{with} \quad \frac{1}{K} = \int_{-\alpha}^{+\alpha} x^2 G_1(x)\, dx, \tag{7.14}$$

where $G_1(x)$ is the Gaussian distribution $N(0{,}1)$ and α is the $(0.5 + c/2n)^{\text{th}}$ quantile of $G_1(x)$. In equation (7.14), K is a normalization factor introduced because the LTS criterion is not a consistent estimator of σ when the r_k are distributed like $N(0{,}\sigma^2)$, except when $h = n$.

7.3.2 Bifunctional dependence assumption

Functional dependence as expressed in equation (7.13) implicitly assumes that two structures having similar intensity ranges in J should also have similar intensity ranges in I. With some combinations of multimodal images, this is a crude approximation. For example, ventricles and bones generally give similar response values in a MR T1 weighted image while they appear with very distinct values in a CT scan. Conversely, white and black matter are well contrasted in a T1 image while corresponding to similar intensities in a CT.

To circumvent this difficulty, we have developed a strategy that enables the mapping of an intensity value in J to not only one, but two possible intensity values in I. This method is a natural extension of the previous method. Instead of computing a single function that maps the intensities of J to those of I, two functions are estimated and the mapping becomes a weighted sum of these two functions.

We start with the assumption that if a point has an intensity j in J, the corresponding point in I has an intensity i that is normally distributed around two possible values depending on j, $f_{\boldsymbol{\theta}}(j)$ and $f_{\boldsymbol{\psi}}(j)$. In statistical terms, this means that, given j, i is drawn from a mixture of Gaussian distribution,

$$P(i|j) = \pi_1(j)\, N\left(f_1(j),\sigma^2\right) + \pi_2(s)\, N\left(f_2(j),\sigma^2\right), \tag{7.15}$$

where $\pi_1(j)$ and $\pi_2(j) = 1 - \pi_1(j)$ are mixing proportions that depend on the intensity in the target image, and σ^2 represents the variance of the noise in the source image. Consistently with the functional case, we will restrict ourselves to polynomial intensity functions, i.e.

$$\begin{cases} f_{\boldsymbol{\theta}}(j) &= \theta_0 + \theta_1 j + \theta_2 j^2 + \cdots + \theta_p j^p \\ f_{\boldsymbol{\psi}}(j) &= \psi_0 + \psi_1 j + \psi_2 j^2 + \cdots + \psi_p j^p \end{cases}$$

An intuitive way to interpret this modelling is to state that for any voxel, there is a binary "selector" variable $\epsilon = \{1{,}2\}$ that would tell us, if it was observed, which of the two functions $f_{\boldsymbol{\theta}}$ or $f_{\boldsymbol{\psi}}$ actually serves to map j to i. Without knowledge of ϵ, the best intensity correction to apply to J in the minimum variance sense (Papoulis, 1991) is seen to be a weighted sum of the two functions,

$$f(i{,}j) = P(\epsilon = 1|i{,}j)\, f_{\boldsymbol{\theta}}(j) + P(\epsilon = 2|i{,}j)\, f_{\boldsymbol{\psi}}(j), \tag{7.16}$$

in which the weights correspond to the probability that the point be mapped according to either the first or the second function. We see that the intensity correction is now a function of both j and i. Applying Bayes' law, we find that for $\epsilon = \{1{,}2\}$:

$$P(\epsilon|i{,}j) = \frac{P(\epsilon|j)P(i|\epsilon{,}j)}{P(i|j)},$$

and thus, using the fact that $P(\epsilon|j) = \pi_\epsilon(j)$ and $P(i|\epsilon{,}j) = G_\sigma\big(i - f_\epsilon(j)\big)$, the weights are determined by

$$P(\epsilon|i{,}j) = \frac{\pi_\epsilon(j)\, G_\sigma\big(i - f_\epsilon(j)\big)}{\pi_1(j)\, G_\sigma\big(i - f_{\boldsymbol{\theta}}(j)\big) + \pi_2(j)\, G_\sigma\big(i - f_{\boldsymbol{\psi}}(j)\big)}, \tag{7.17}$$

where it should be clear from the context that $f_\epsilon \equiv f_{\boldsymbol{\theta}}$ if $\epsilon = 1$, and $f_\epsilon \equiv f_{\boldsymbol{\psi}}$ if $\epsilon = 2$.

In order to estimate the parameters of this model, we employ an ad hoc strategy that proceeds as follows. First, $\boldsymbol{\theta}$ is estimated using the LTS/RLS method described in section 7.3.1. The points not used to compute $\boldsymbol{\theta}$, in a number between 0 and $n - h$, are used to estimate $\boldsymbol{\psi}$ still using the same method. Note that if this number is less than $10 \times p$, p being the polynomial degree, functional dependence is assumed and we fall back to the monofunctional assumption.

This provides a natural estimation of the "selector" variable for each voxel: the n_1 points that were used to build $f_{\boldsymbol{\theta}}$ are likely to correspond to $\epsilon = 1$, while the n_2 points used to build f_ψ are likely to correspond to $\epsilon = 2$. Finally, the points that are rejected while estimating $\boldsymbol{\psi}$ are considered as bad intensity matches. A natural estimator for the variance σ^2 is then:

$$\hat{\sigma}^2 = \frac{n_1}{n_1 + n_2}\hat{\sigma}_1^2 + \frac{n_2}{n_1 + n_2}\hat{\sigma}_2^2,$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the variances found respectively for $f_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\psi}}$ during the RLS regression (see section 7.3.1.). Similarly, the mixing proportions are computed according to:

$$\hat{\pi}_\epsilon(j) = \frac{n_\epsilon(j)}{n_1(j) + n_2(j)}, \qquad \epsilon = \{1{,}2\},$$

in which $n_\epsilon(j)$ is the number of voxels having an intensity j and used to build the function f_ϵ. Notice that in the case where $n_1(j) = n_2(j) = 0$ (i.e. no voxel corresponding to the intensity class j

has been taken into account in the computation of $f_{\boldsymbol{\theta}}$ or $f_{\boldsymbol{\psi}}$), then we arbitrarily set the mixing proportions to $\hat{\pi}_1(j) = \hat{\pi}_2(j) = 0.5$.

The intensity correction of J can now be performed by reinjecting the estimated parameters in equations (7.17) and (7.16).

7.4 Information mutuelle

La méthode de mise en correspondance que nous avons détaillée aux paragraphes 7.2 et 7.3 trouve son origine dans le principe d'une minimisation alternée du rapport de corrélation. Nous proposons ici une généralisation de cette approche à un autre critère de vraisemblance, très proche de l'information mutuelle définie par (Viola, 1995; Viola et Wells, 1997).

Au chapitre 5, nous avons défini la log-vraisemblance négative d'un couple d'images (I,J) en fonction d'une transformation spatiale T et d'une densité de probabilité bivariée P :

$$L(T,P) = -\sum_{\mathbf{x}_k \in \Omega_I^\downarrow} \log \frac{P(i_k|j_k^\downarrow)}{P(i_k)} \; - \sum_{\mathbf{y}_\ell \in \Omega_J} \log P(j_\ell) \; - \sum_{\mathbf{x}_k \in \Omega_I} \log P(i_k),$$

où Ω_I et Ω_J sont les grilles respectives des images et $\Omega_I^\downarrow$ est le sous-ensemble de Ω_I défini par les voxels $\mathbf{x}_k$ ayant un correspondant dans Ω_J. De façon cohérente avec le paragraphe 7.3, nous supposerons que tous les voxels ont un correspondant, de sorte que $\Omega_I^\downarrow = \Omega_I$. En pratique, on se ramène à cette hypothèse par interpolation/extrapolation de l'image J.

Comme pour le rapport de corrélation, nous pouvons envisager une minimisation alternée de la fonctionnelle $L(T,P)$. Certains résultats obtenus au chapitre 5 nous sont alors utiles :

1. A P fixée, minimiser $L(T,P)$ équivaut à maximiser $n^\downarrow \mathcal{I}(T,P)$, où $n^\downarrow \equiv \mathrm{Card}\Omega_I^\downarrow = n$ est ici une constante et $\mathcal{I}(T,P)$ est l'information mutuelle empirique des images définie par la distribution P (voir le paragraphe 5.3.4) :

$$\mathcal{I}(T,P) = \frac{1}{n} \sum_k \log \frac{P(i_k|j_k^\downarrow)}{P(i_k)}$$

2. A T fixée, si P est recherchée comme une combinaison linéaire de noyaux du type $\phi(i - x)\psi(j - y)$, alors la distribution $\hat{P}$ qui réalise le minimum de $L(T,P)$ peut être approchée par la méthode de Parzen (voir le paragraphe 5.4.3):

$$\hat{P}(i,j) = \frac{1}{n} \sum_k \phi(i - i_k)\psi(j - j_k^\downarrow)$$

L'intérêt de ce critère de vraisemblance est qu'il repose sur des hypothèses plus générales que celle d'une dépendance mono- ou bi-fonctionnelle entre les intensités. Ainsi, et contrairement à l'approche présentée au paragraphe 7.3, il ne s'agit pas de corriger l'une des images pour qu'elle ressemble à l'autre. Le modèle de dépendance est exprimé ici d'une façon purement statistique, via la distribution conjointe P.

Pour ce qui est de la minimisation par rapport à T (à P fixée), le critère d'information mutuelle va simplement changer l'expression des forces intervenant dans le schéma d'estimation incrémentale des déplacements. En procédant par analogie avec le paragraphe 7.2, nous pouvons commencer par envisager la minimisation de la log-vraisemblance négative pénalisée. D'après l'équation (7.5), il est clair que nous aboutissons alors à un schéma itératif du type :

$$\mathbf{u}_{n+1}(\mathbf{x}) = G_\sigma \otimes \mathbf{f}(\mathbf{x},T_n), \tag{7.18}$$

où le champ de force est relié au gradient de L par rapport à T. En posant : $\gamma(i,j) \equiv \dfrac{\partial}{\partial j} \log P(i|j)$, nous avons :

$$\mathbf{f}(\mathbf{x},T) = -\frac{1}{\lambda}\frac{\partial L}{\partial T} = \frac{1}{\lambda}\sum_k \delta(\mathbf{x}-\mathbf{x}_k)\,\gamma(i_k,j_k^{\downarrow})\,\nabla J_k^{\downarrow},$$

la constante $\lambda > 0$ étant le paramètre contrôlant l'influence relative de la log-vraisemblance et du stabilisateur $U(T)$ que nous choisissons bien entendu comme dans l'équation (7.4).

La fonction $\gamma(i_k,j_k^{\downarrow})$ joue vis-à-vis de la log-vraisemblance le même rôle fondamental que $(i_k - j_k^{\downarrow})$ vis-à-vis de la SDC. Son signe détermine localement le *sens* de la force, la *direction* étant comme précédemment imposée par celle du gradient $\nabla J_k^{\downarrow}$. On peut voir $\gamma(i_k,j_k^{\downarrow})$ comme une généralisation «multimodale» de la notion de polarité introduite par (Thirion, 1998). Son calcul explicite fait intervenir l'expression de la distribution P courante, c'est-à-dire l'estimée de Parzen associée à la transformation T^* considérée au début de l'étape de minimisation par rapport à T. Un petit calcul permet d'obtenir :

$$\gamma(i,j) = \frac{\sum_k \phi(i-i_k)\psi'(j-j_k^*)}{\sum_k \phi(i-i_k)\psi(j-j_k^*)} - \frac{\sum_k \psi'(j-j_k^*)}{\sum_k \psi(j-j_k^*)},$$

où, par définition, $j_k^* \equiv J(T^*(\mathbf{x}_k))$. De façon à permettre une comparaison objective entre l'information mutuelle et la méthode décrite aux paragraphes 7.2 et 7.3, nous avons remplacé comme précédemment le schéma itératif correspondant à l'équation (7.18) par un schéma du type démons :

$$\mathbf{u}_{n+1}(\mathbf{x}) = G_\sigma \otimes [\mathbf{u}_n(\mathbf{x}) + \mathbf{f}(\mathbf{x},T_n)]$$

En revanche, contrairement à ce que nous avons fait au paragraphe 7.2, nous n'avons pas adapté l'expression du champ de forces car nous n'avons pas vu de moyen évident de le faire. La méthode dépend donc du choix de la constante $\lambda > 0$, qui s'avère délicat en pratique. Nous présentons au paragraphe 7.6.3 les résultats de recalage non-rigide que nous avons obtenus avec cette implémentation de l'information mutuelle.

7.5 Data

Most of the data used in the following experiments were obtained from BrainWeb[2] (Cocosco et al., 1997). This tool uses an atlas (Collins et al., 1998) with a resolution of $1\times1\times1\text{mm}^3$ comprising ten segmented regions (see table 7.1) from which T1, T2 and PD images can be generated (Kwan

2. `http://www.bic.mni.mcgill.ca/brainweb/`

et al., 1996). We use the discrete phantom atlas and three simulated MR images, one of each modality, with the same resolution as the atlas, 5% noise and no intensity non-uniformity. These three images are named $T1_a$, $T2_a$, PD_a, respectively. Since they are generated from the same atlas, they represent the same underlying anatomy and are all perfectly matched.

Tissue type	Intensity	Tissue type	Intensity
Background	0	Muscle / Skin	140
CSF	28	Skin	168
Gray Matter	56	Skull	196
White Matter	84	Glial Matter	224
Fat	112	Connective	252

TAB. 7.1 – *Intensity values corresponding to the different tissue types present in the discrete phantom atlas.*

We also use a T1 MR image and a CT image, both from different subjects and having a resolution of $1 \times 1 \times 1\text{mm}^3$. Both these images were affinely registered with the atlas using the correlation ratio method. The two images are called T1 and CT respectively.

The images all respect the neurological convention, i.e. on axial slices, the patient's left is on the left side of the image.

7.6 Results and discussion

In the following section we present non-rigid registration results involving images obtained from several different kinds of modalities. We wish to make clear to the reader that all registrations were performed in 3D, *not* slice by slice.

First, we show a typical example where monofunctional dependence can be assumed: registration of an atlas with an MR image. Then, more practical examples are shown where images from different modalities are registered and where bifunctional dependence may be assumed. We also provide registration results obtained from our implementation of mutual information maximization as described in section 7.4.

We made use of a three-level implementation, namely 4mm, 2mm and 1mm per voxel. Displacement fields at one level are initialized from the result of the previous level. The initial displacement field $\mathbf{u}_0$ is set to zero. 128 iterations are performed at 4mm/voxel, 32 at 2mm/voxel and 8 at 1mm/voxel. These are twice the number of iterations used for registration of monomodal images using the conventional demons algorithm. We believe that making use of a stopping criterion, such as the difference of the SSD values between iterations, would probably improve the results shown below but this aspect has not been investigated. It should be in future implementations.

The Gaussian filter G_σ used to smooth the displacement field has a standard deviation of 1 voxel regardless of the resolution. This models stronger constraints on the deformation field at the beginning of the registration process to correct for gross displacements, and weaker constraints near the end when fine displacements are sought. The resampling process makes use of trilinear interpolation, except in the case of the atlas where nearest-neighbor interpolation is used.

Computation time to obtain the following results is around 60 minutes on a 450MHz PC with 500MB of RAM (10 minutes at 4mm, 20 minutes at 2mm and 30 minutes at 1mm). Most of this time ($\approx 85\%$) is devoted to the intensity correction part, which has not been optimized in this first version of our program. The other 15% is taken by the standard registration code which is stable and well optimized.

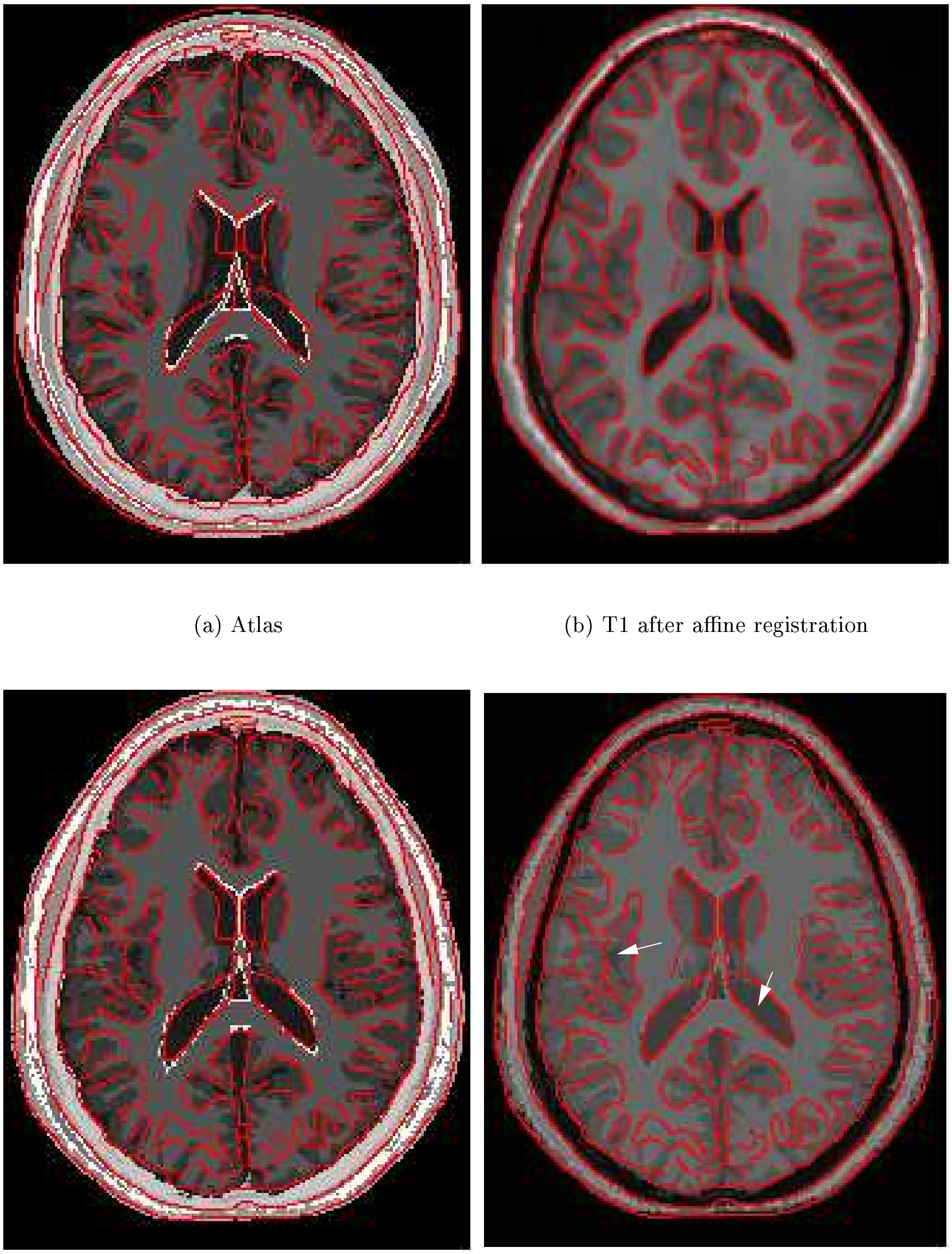
7.6.1 Monofunctional dependence

We present here the result of registering the atlas with T1. Since the atlas can be used to generate realistic MR images, it is safe to assume a functional dependence between the intensity of the atlas and those of T1. Also, since T1 and the atlas are well aligned due to the affine registration, we have roughly estimated that the number of points already well matched, i.e. the number of good points in $\mathcal{C}^{\downarrow}$, are at least $0.80 \times n$, to which we have set the value of h. Since 10 classes are present in the atlas, the polynomial degree chosen was set to 9. Having a polynomial of this degree will accommodate any permutation of the structure intensities in the atlas. This is true because it is always possible to fit a polynomial of degree 9 to match 10 points exactly.

The result of registration is presented in figures 7.5, 7.6, and 7.7, respectively with axial and sagittal views. From left to right, top to bottom, the first pictures (a) show slices of the atlas. The pictures (b) are the corresponding slices of T1 after affine registration. The pictures (c) and (d) show the deformed atlas after non-rigid registration, respectively without and with intensity correction. We have resampled the T1 in terms of the inverse transformation found by affine registration, while the atlas has undergone the residual deformation field. This choice was motivated mainly by esthetical considerations. In both cases (affine and non-affine registration), however, it should be made clear that the atlas was chosen as the target image.

As can be seen, large morphometric differences have been corrected. Still, the matching is not perfect which may be observed by comparing the shape of several structures between T1 and the deformed atlas, e.g. the ventricles. Registration imperfections are reflected in the intensity corrected image, where one may notice that the CSF intensity is slightly brighter than that in T1 (as can be seen in the ventricles and around the cortex, see arrows in figures 7.5, 7.6, and 7.7). This problem can also be observed by looking at the intensity transformation function presented in figure 7.8. The intensity level corresponding to the CSF is overestimated due to an overlap of the CSF in the atlas with the gray and white matter in T1, especially around the cortical area which is known to present large variations between subjects.

This is probably an inherent limitation of quadraric stabilizers when used in the context of inter-subject registration. The strong smoothness constraints imposed by the Gaussian regularization (or related linear regularization techniques) may prevent the assessment of large and uneven displacements required to match the most variable anatomical structures, e.g. the cortex. A non-linear



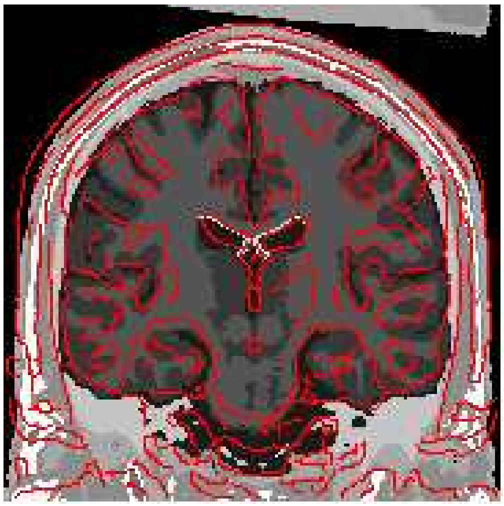
(a) Atlas

(b) T1 after affine registration

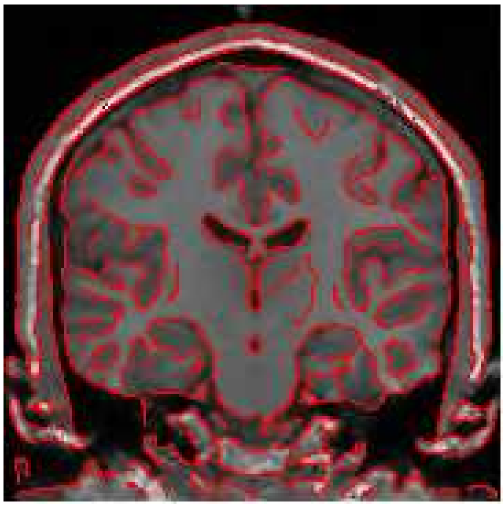
(c) Atlas without intensity correction after non-registration

(d) Atlas with intensity correction after non-rigid registration

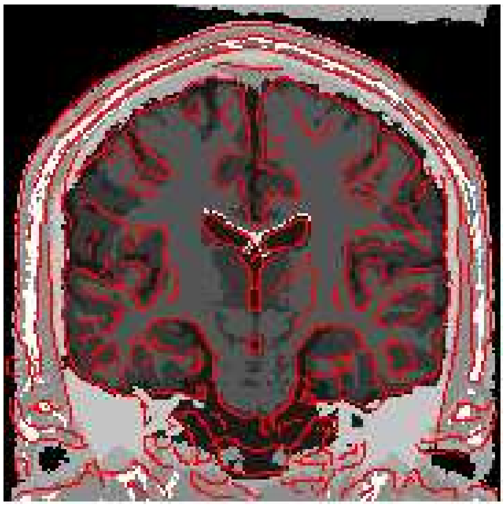
FIG. 7.5 – *Corresponding axial views of the atlas to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration. Arrows point to examples of bad intensity correction.*



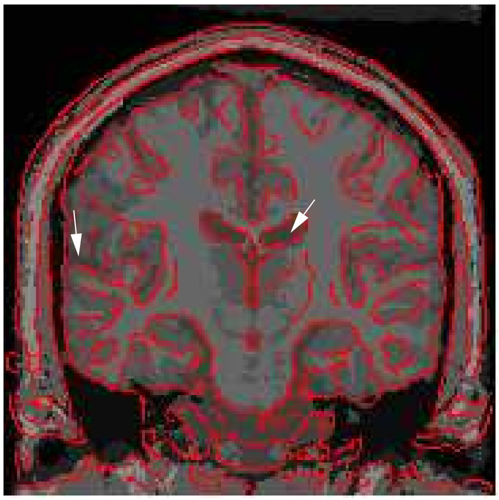
(a) Atlas



(b) T1 after affine registration

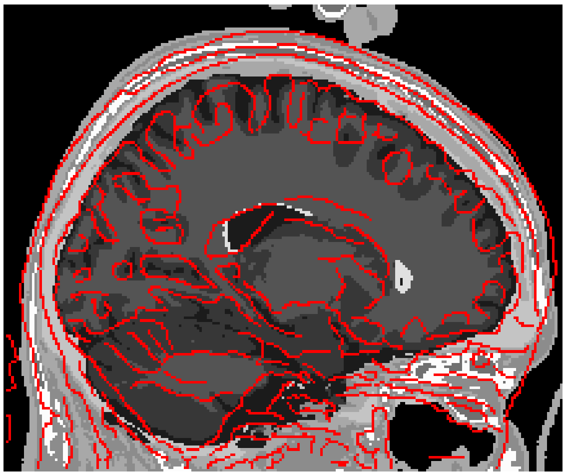


(c) Atlas without intensity correction after non-rigid registration



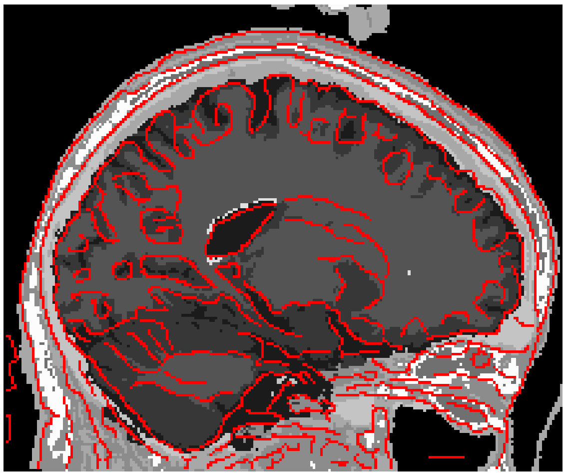
(d) Atlas with intensity correction after non-rigid registration

FIG. 7.6 – *Corresponding coronal views of the atlas to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration. Arrows point to examples of bad intensity correction.*

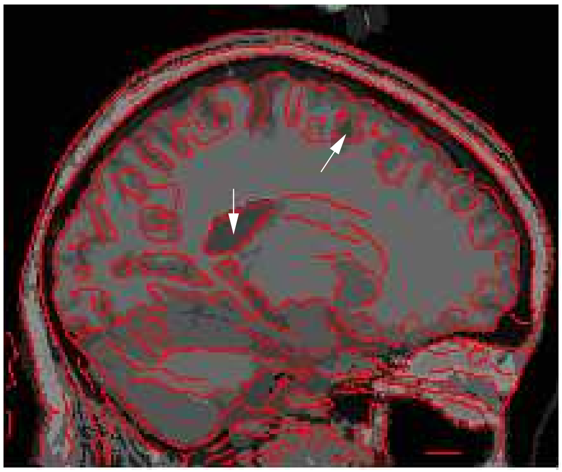


(a) Atlas

(b) T1 after affine registration



(c) Atlas without intensity correction after non-rigid registration



(d) Atlas with intensity correction after non-rigid registration

FIG. 7.7 – *Corresponding sagittal views of the atlas to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration. Arrows point to examples of bad intensity correction.*

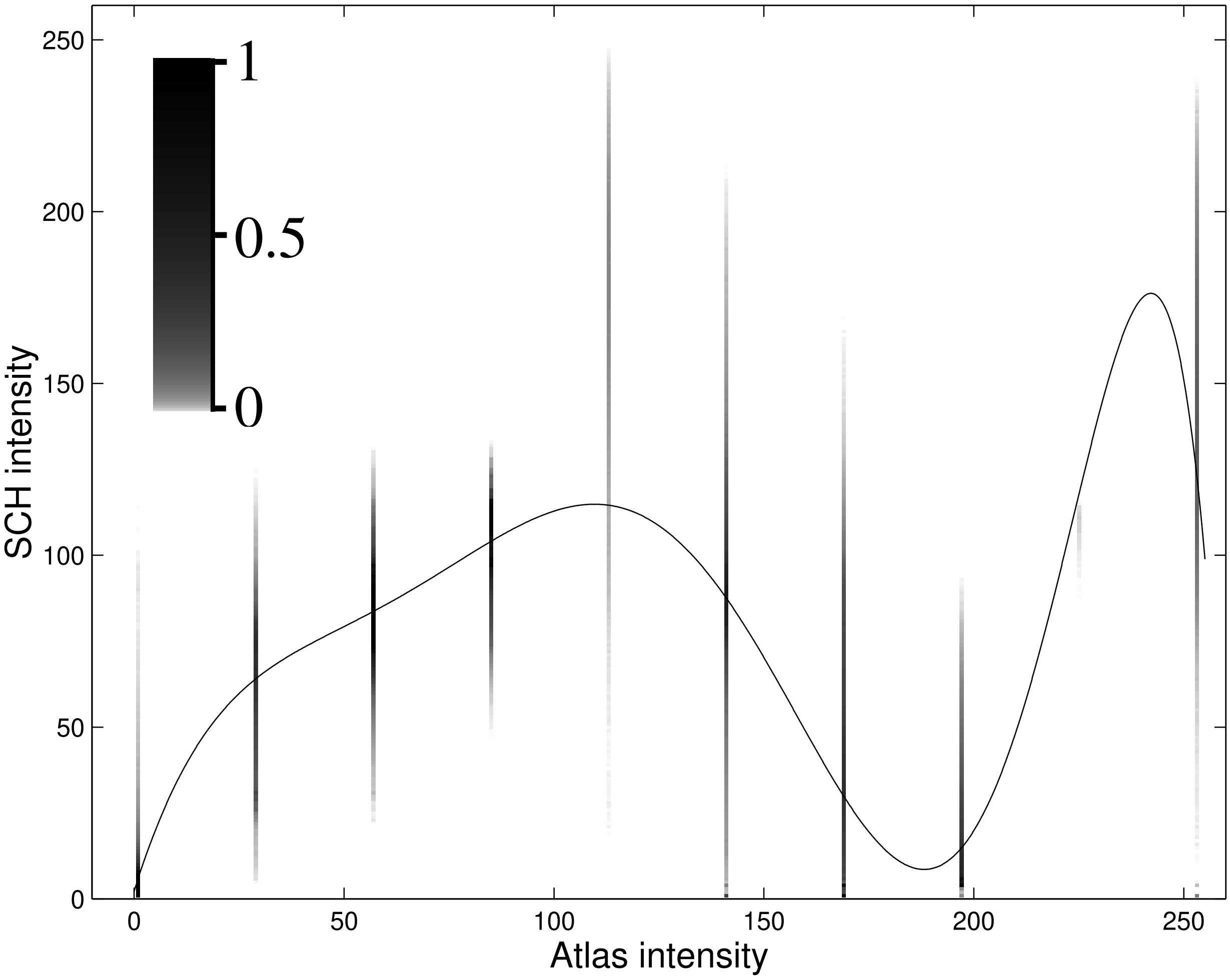


FIG. 7.8 – *Intensity transformation found by registering the atlas with T1 and assuming mono-functional dependence. The function f is overlaid on the joint histogram of the two images after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. The tissue types corresponding to the different atlas intensities are presented in table 7.1.*

regularization strategy such as that presented in (Hellier et al., 1999; Hellier et Barillot, 2000) would probably allow for stronger discontinuities in the displacement field.

To verify this assumption, we registered $T1_a$ with T1 and deformed $T1_a$ using the resulting displacement field. This provided an image T1' that closely resembles T1 and for which we are assured that the shape variation from the atlas can be assessed by our algorithm. We then registered the atlas with T1'. This result is presented in figures 7.9, 7.10, and 7.11. As can be observed, the CSF intensity value is now well corrected. By looking at the intensity transformation shown in figure 7.12, we also notice that each atlas structure has corresponding intensity ranges in T1 that are less extended than those of figure 7.8, reflecting a better match between the images.

Another difference between figures 7.8 and 7.12 is the difference in the intensity mapping of the glial matter (intensity 224 in the atlas). This structure forms a small envelope around the ventricles and is represented by so few points that they are considered as outliers. This could be corrected by considering more points during the intensity transformation. In fact, h should be increased at each iteration to reflect that, during registration, J gradually aligns with I and more points in $\mathcal{C}^{\downarrow}$ can be considered as well matched.

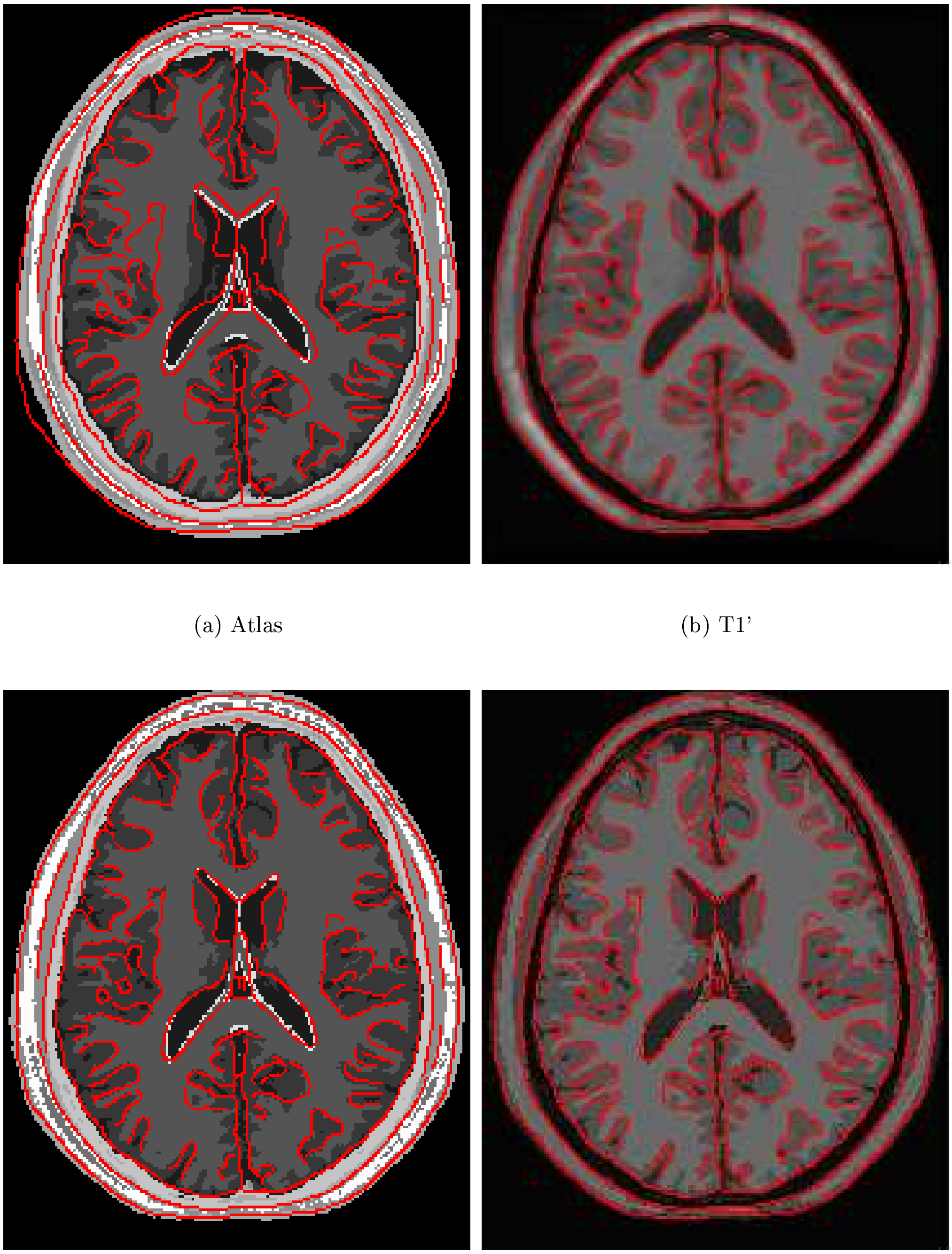
7.6.2 Bifunctional dependence

When registering images from different modalities, monofunctional dependence may not necessarily be assumed. Here, we applied the method described in section 7.3.2 where two polynomial functions of degree 12 are estimated. This number was set arbitrarily to a relatively high value to enable important intensity transformations.

T1/CT warping Figures 7.13, 7.14, and 7.15 present the result of registering $T1_a$ with CT. Using these last two modalities, most intensities of $T1_a$ should be mapped to gray and only the skull, representing a small portion of the image data, should be mapped to white. After affine registration almost all voxels are well matched, i.e. the number of good points in $\mathcal{C}^{\downarrow}$ is large. Hence, in this particular case, we have chosen a high value for h set to $0.90 \times n$.

As we can see in figures 7.13 to 7.15, the skull, shown in black in the MR image and in white in the CT scan, is well registered and the intensity transformation adequate. Figure 7.16 (a) presents the joint histogram of the two images after registration. This histogram is color-coded and ranges from red representing high point densities to blue depicting low point densities. Figure 7.16 (b) presents the functions f_1 and f_2 found during the registration process. The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity j is proportional to the value of the corresponding $\pi_\epsilon(j)$. The gray values represent the joint histogram after registration. As can be observed on these graphs, the polynomials found fit well with the high density clusters of the joint histogram. Still, some points need to be addressed.

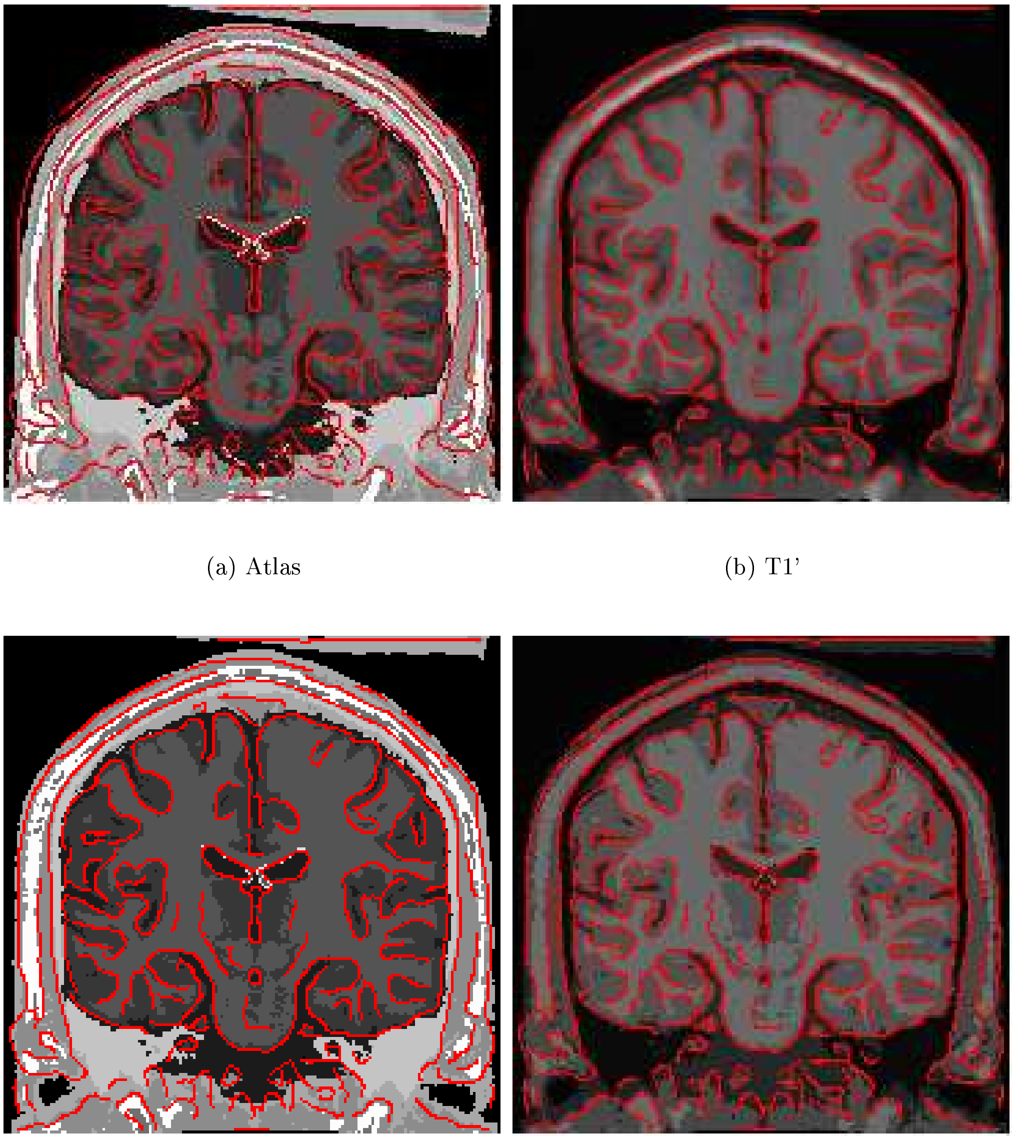
We can observe that due to the restricted polynomial degree, f_1, (shown in red) oscillates around the CT gray value instead of fitting a straight line. This is reflected in the intensity corrected $T1_a$ image, where the underlying anatomy can still be observed by small intensity variations inside the



(a) Atlas (b) T1'

(c) Atlas without intensity correction after non-rigid registration (d) Atlas with intensity correction after non-rigid registration

FIG. 7.9 – *Corresponding axial views of the atlas to T1' registration result. Contours were obtained using a Canny-Deriche edge detector on the T1' image (b) and overlaid on the other images to better assess the quality of registration.*



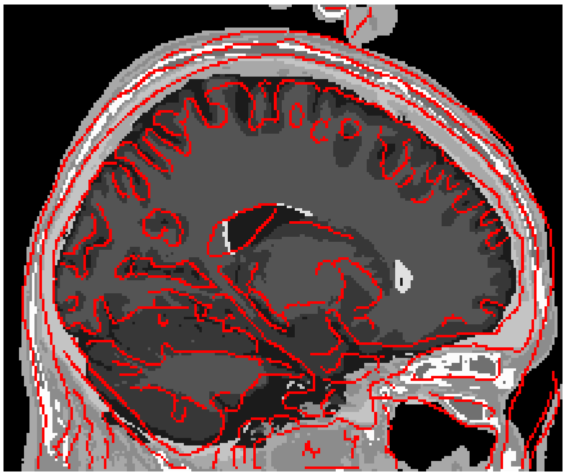
(a) Atlas

(b) T1'

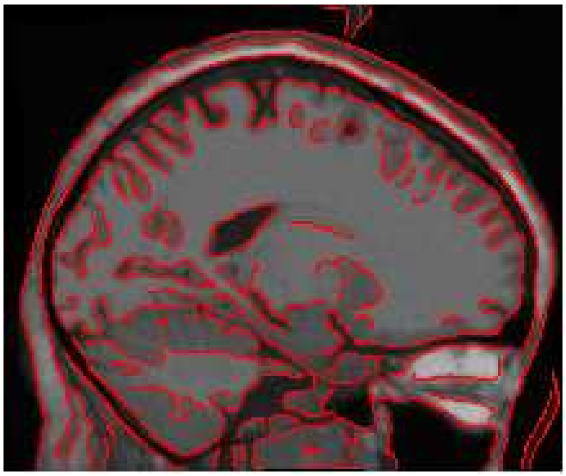
(c) Atlas without intensity correction after non-rigid registration

(d) Atlas with intensity correction after non-rigid registration

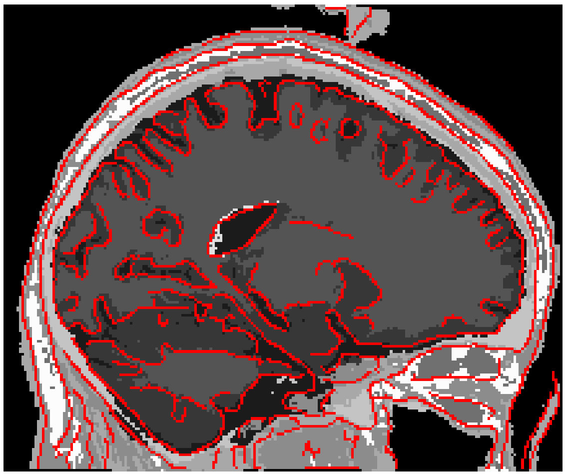
FIG. 7.10 – *Corresponding coronal views of the atlas to T1' registration result. Contours were obtained using a Canny-Deriche edge detector on the T1' image (b) and overlaid on the other images to better assess the quality of registration.*



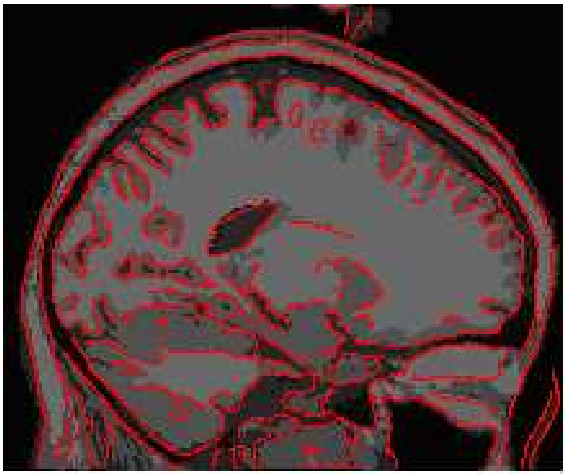
(a) Atlas



(b) T1'



(c) Atlas without intensity correction after non-rigid registration



(d) Atlas with intensity correction after non-rigid registration

FIG. 7.11 – *Corresponding sagittal views of the atlas to T1' registration result. Contours were obtained using a Canny-Deriche edge detector on the T1' image (b) and overlaid on the other images to better assess the quality of registration.*

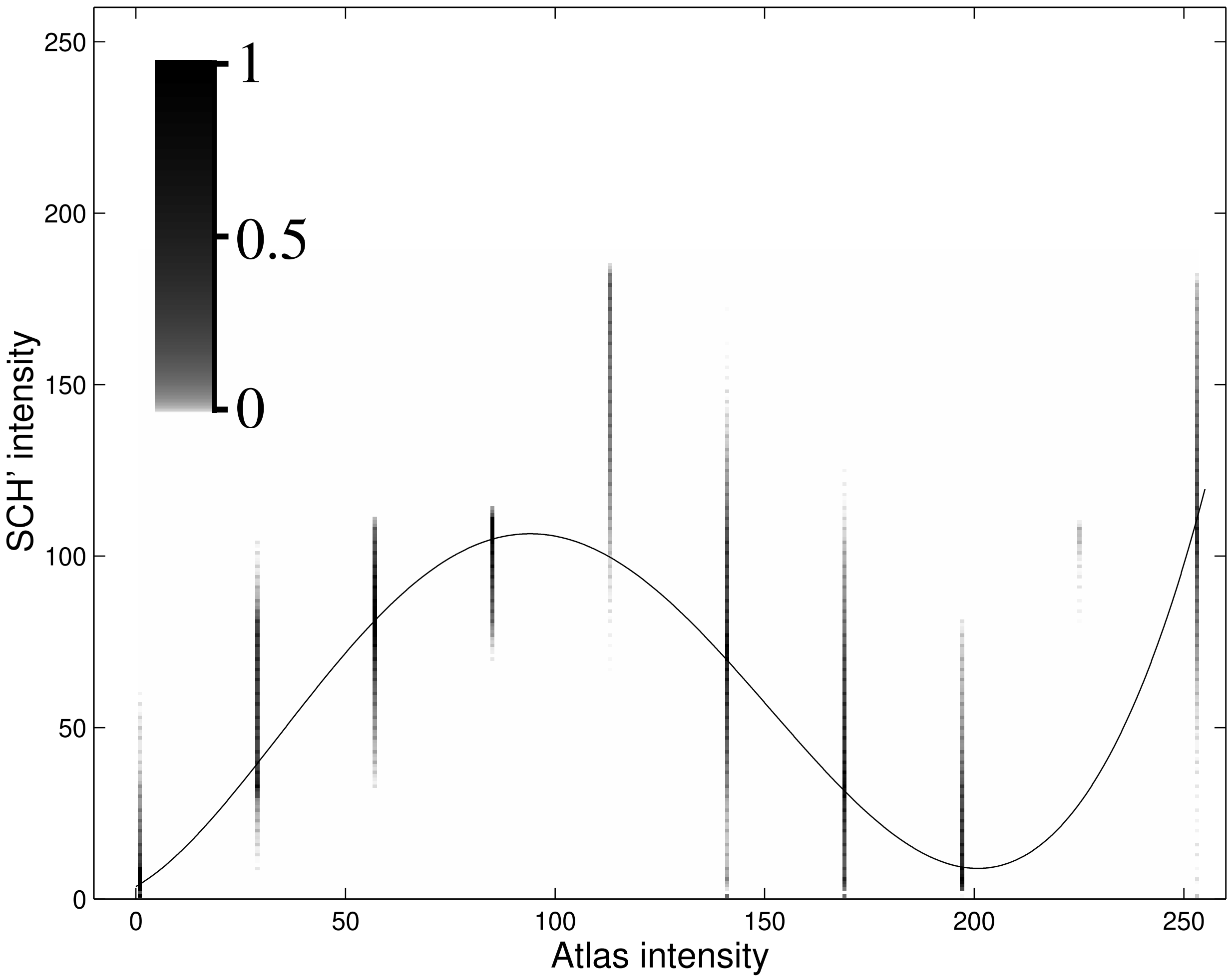
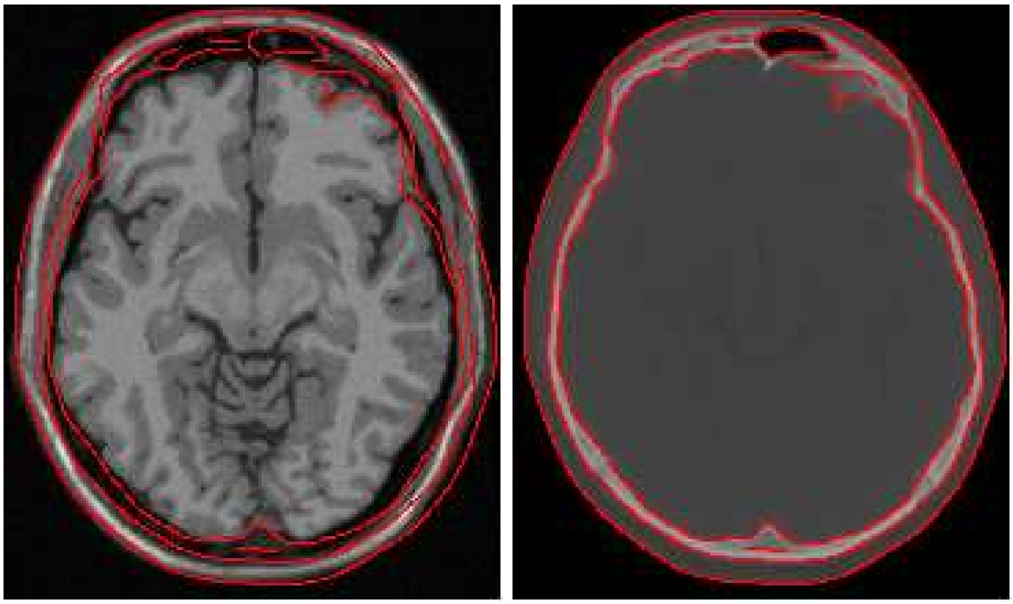
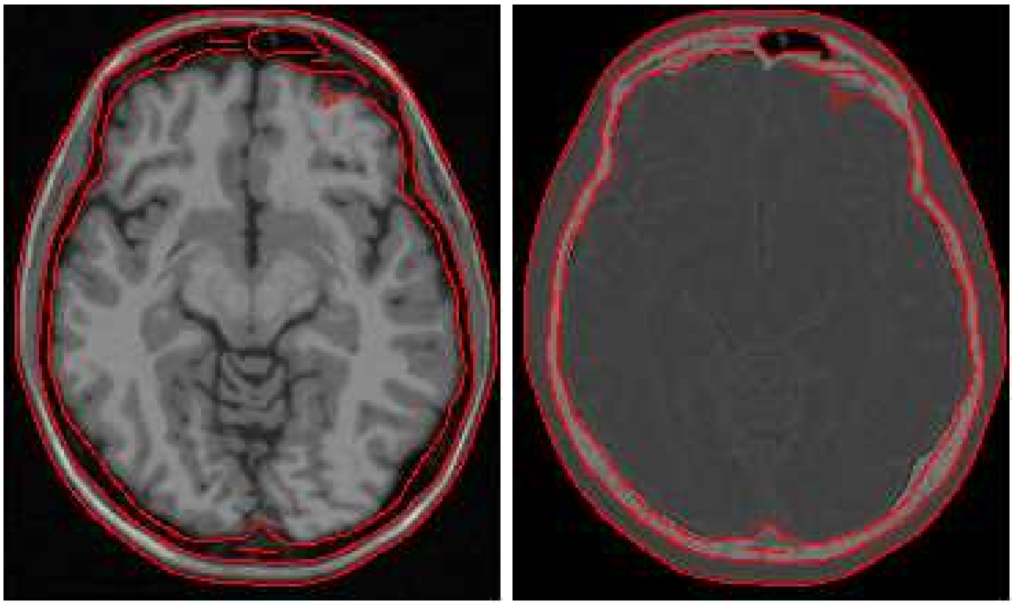


FIG. 7.12 – *Intensity transformation found by registering the atlas with T1' and assuming mono-functional dependence. The function f is overlaid on the joint histogram of the two images after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. The tissue types corresponding to the different atlas intensities are presented in table 7.1.*

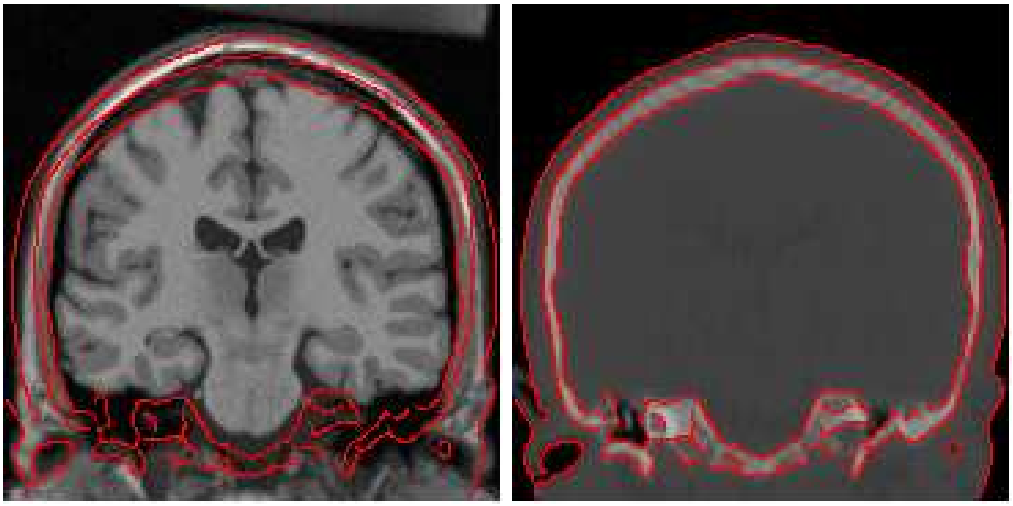


(a) $T1_a$ (b) CT after affine registration

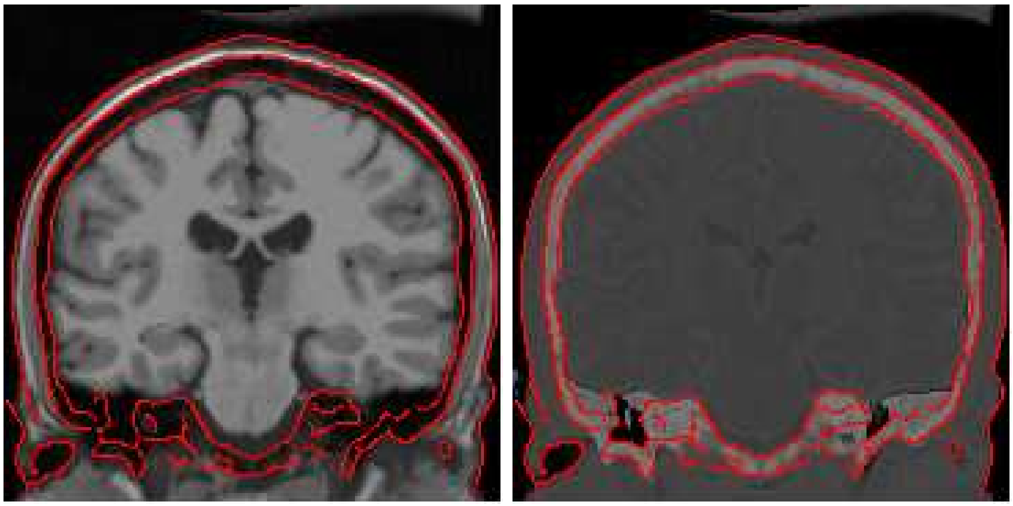


(c) $T1_a$ without intensity correction after non-rigid registration (d) $T1_a$ with intensity correction after non-rigid registration

FIG. 7.13 – *Corresponding axial views of $T1_a$ to CT registration result. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*

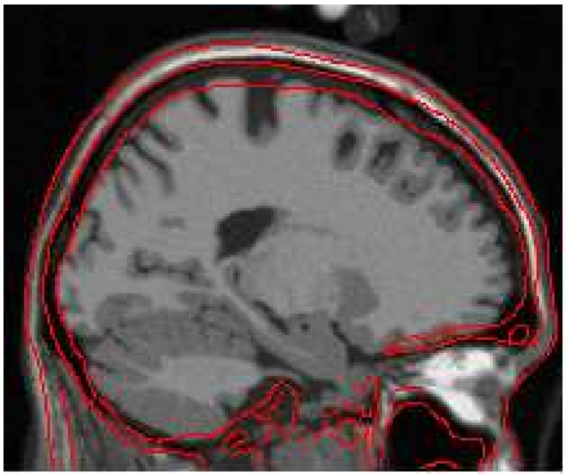


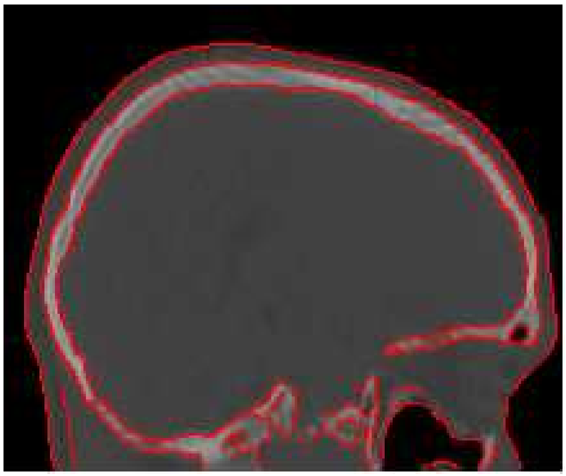
(a) $T1_a$ (b) CT after affine registration



(c) $T1_a$ without intensity correction after non-rigid registration (d) $T1_a$ with intensity correction after non-rigid registration

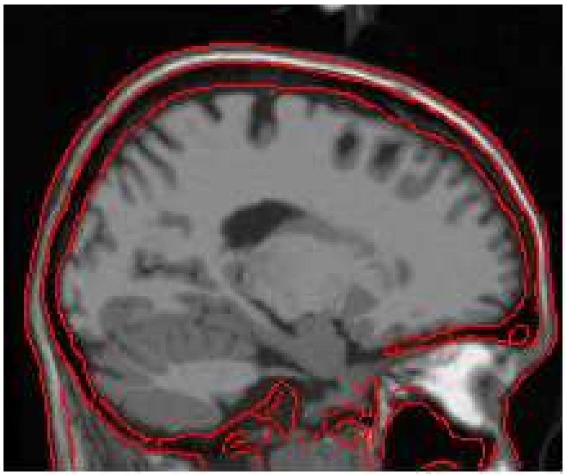
FIG. 7.14 – *Corresponding coronal views of $T1_a$ to CT registration result. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*

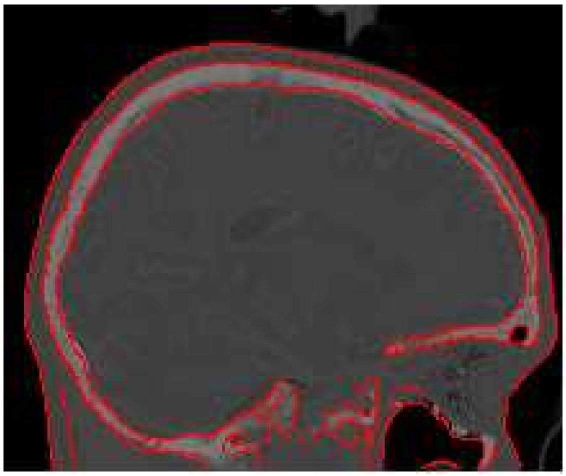




(a) $T1_a$

(b) CT after affine registration

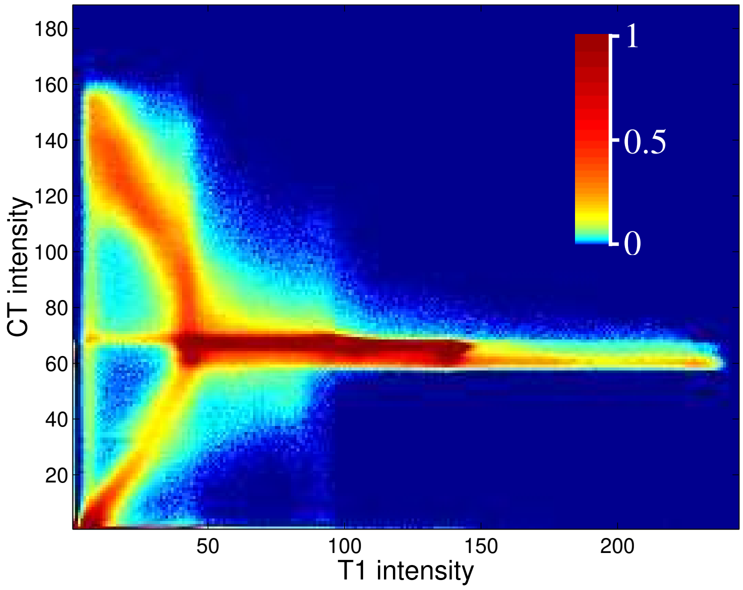




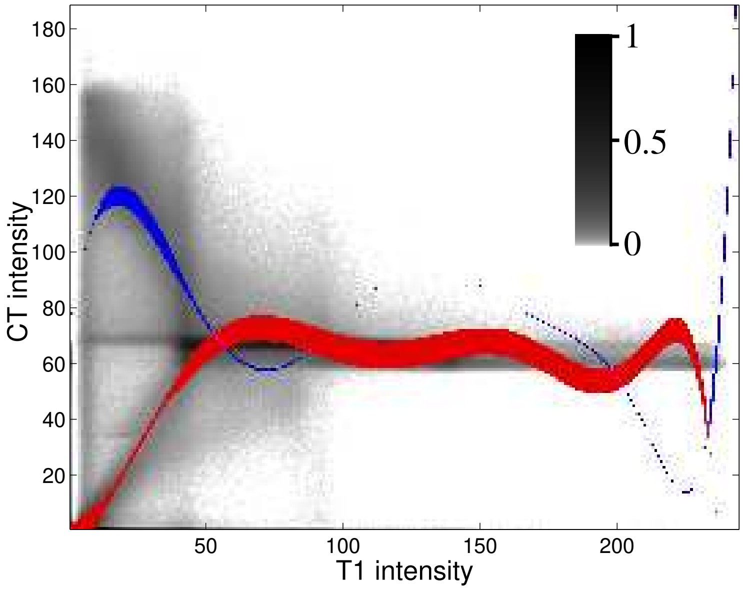
(c) $T1_a$ without intensity correction after non-rigid registration

(d) $T1_a$ with intensity correction after non-rigid registration

FIG. 7.15 – *Corresponding sagittal views of $T1_a$ to CT registration result. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*



(a) Joint histogram values range from red representing high point densities to blue depicting low point densities.



(b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value j in $\mathrm{T1_a}$ corresponds to the value of the corresponding proportion $\pi_\epsilon(j)$.

FIG. 7.16 – *Intensity transformation found when registering $T1_{\mathrm{a}}$ with CT and assuming bifunctional dependence.*

skull. This artifact has insubstantial consequences during the registration process since the difference between most of the voxels intensities is zero, resulting in null displacements. The displacements driving the deformation will be those of the skull and the skin contours, and will be propagated in the rest of the image with the Gaussian filtering of the displacement field.

We also notice that f_2 (shown in blue), which is mainly responsible for the mapping of the skull, does not properly model the cluster it represents for intensities smaller than 5. The mapping for these intensities is slightly underestimated. This may have two causes. First, as in the previous case, it might be due to the restricted polynomial degree. Second, we can notice that some of the background values in $\mathrm{T1_a}$ that have an intensity close to 0 are mapped to gray values in the CT which correspond to soft tissues. This means that some of the background in $\mathrm{T1_a}$ is matched with the skin in the CT. This has the effect of "pulling" f_2 closer to the small cluster positioned around 2.65. If the underestimation of f_2 arises because of the second reason, letting the algorithm iterate longer might provide a better result.

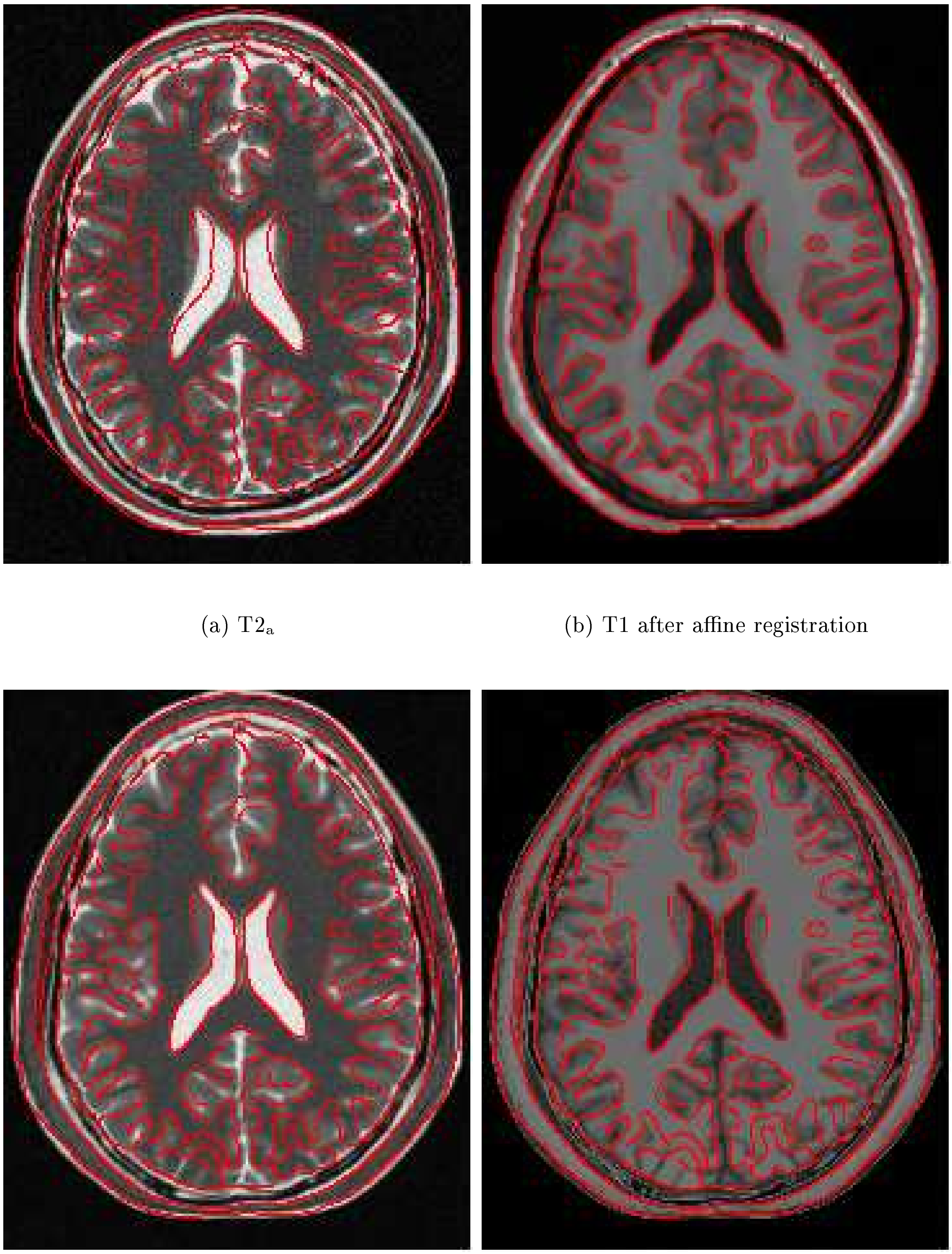
T1/T2 and T1/PD warping Figures 7.17 to 7.24 present the result of registering $\mathrm{T2_a}$ and $\mathrm{PD_a}$ respectively with T1. Figures 7.20 and 7.24 show the corresponding intensity transformations. For these experiments, h was set to $0.60 \times n$, a value we have found to be effective for these types of modalities after affine registration.

One observation that can be made by looking at the graphs in figures 7.20 and 7.24 is that the estimated functions f_1 and f_2 are quite similar in both cases. This suggests that assuming a monofunctional dependence would be relevant. However, the results we obtained when registering T2 with T1, and PD with T1, using the monofunctional model were less convincing than when using the bifunctional model (Guimond, 1999; Guimond et al., 2001).

This may be explained by a closer look at our bifunctional intensity modeling. Equation 7.15 reflects the assumption that if an anatomical point has an intensity j in J, the corresponding point has an intensity i in I that is distributed normally around two possible values depending on j. But it makes no assumption about how the intensities in J are distributed. This models the intensities of J without noise, which may not necessarily be well justified, but enables the use of linear regression to estimate the intensity transformation.

The effect of noise in J is reflected in the joint histograms by enlarging clusters along the x axis. This, added to bad matches and partial volume effect, creates many outliers in $\mathcal{C}^{\downarrow}$ and makes the assessment of the true intensity transformation more difficult and more resistant to our robust regression technique. Preprocessing of J using for example anisotropic diffusion may narrow the clusters and provide better results.

Adding the estimation of a second function in the bifunctional model helps counter the effect of noise on J. For example, the CSF in the $\mathrm{PD_a}$ image has intensity values ranging from about 200 to 240 and gray matter from about 175 to 210. In T1, these ranges are about 30 to 70 and 55 to 80 respectively. As can be seen in figure 7.24, f_1 models well the gray matter cluster but fails to reflect the CSF transformation. Estimating the second polynomial f_2 solves this problem by considering



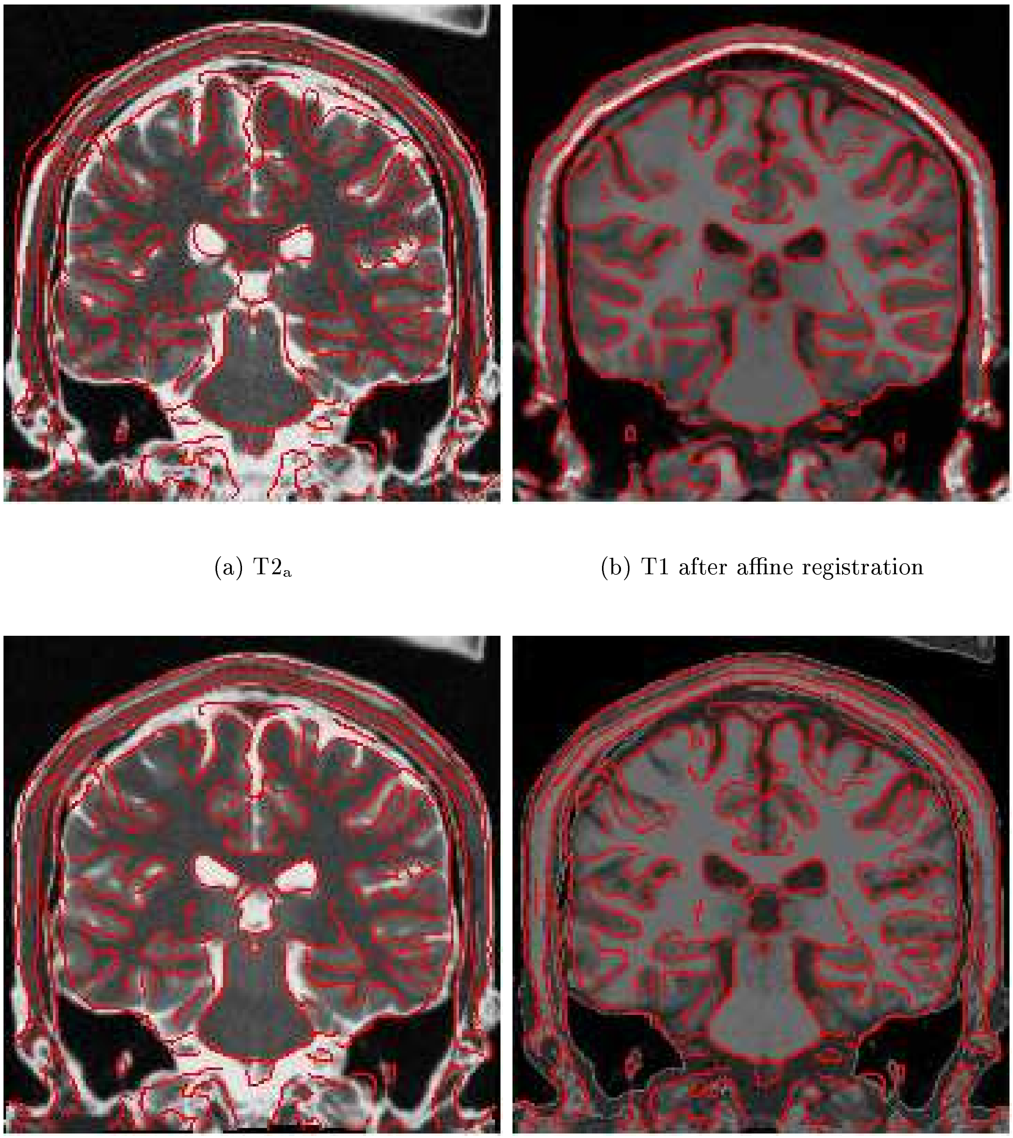
(a) $T2_a$

(b) T1 after affine registration

(c) $T2_a$ without intensity correction after non-rigid registration

(d) $T2_a$ with intensity correction after non-rigid registration

FIG. 7.17 – *Corresponding axial views of $T2_a$ to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*



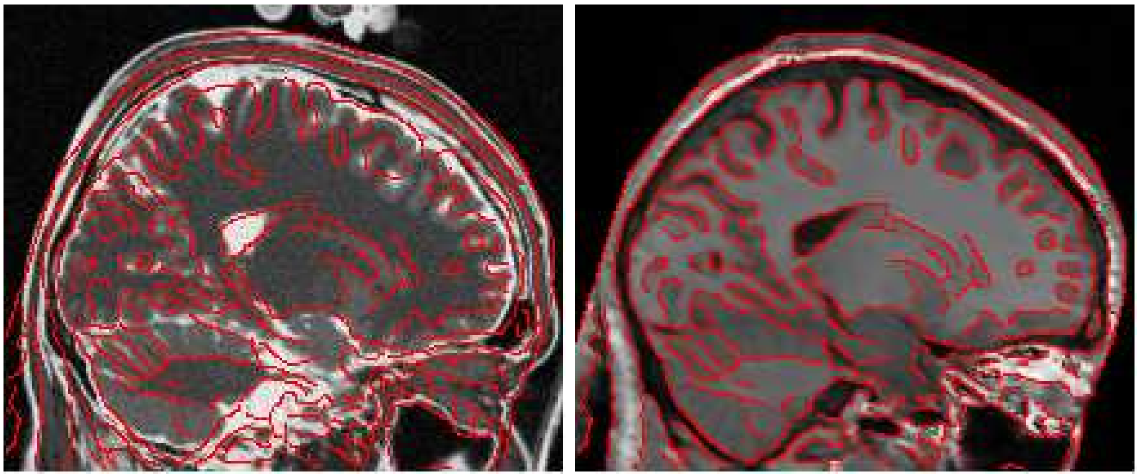
(a) $T2_a$

(b) T1 after affine registration

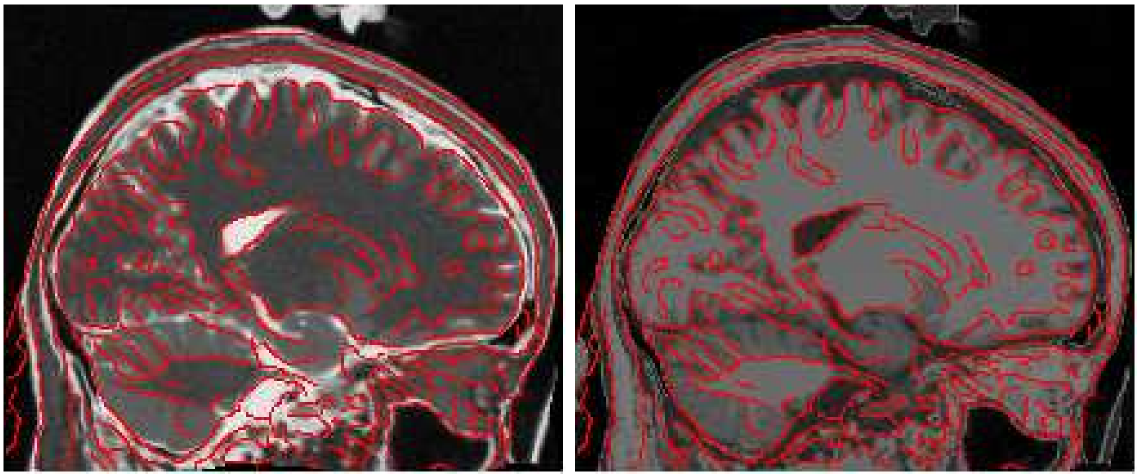
(c) $T2_a$ without intensity correction after non-rigid registration

(d) $T2_a$ with intensity correction after non-rigid registration

FIG. 7.18 – *Corresponding coronal views of $T2_a$ to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*

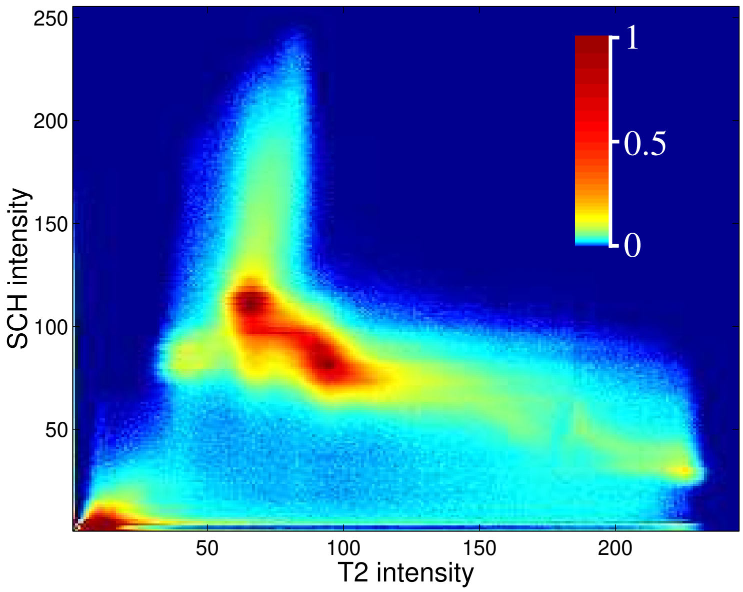


(a) $T2_a$ (b) T1 after affine registration

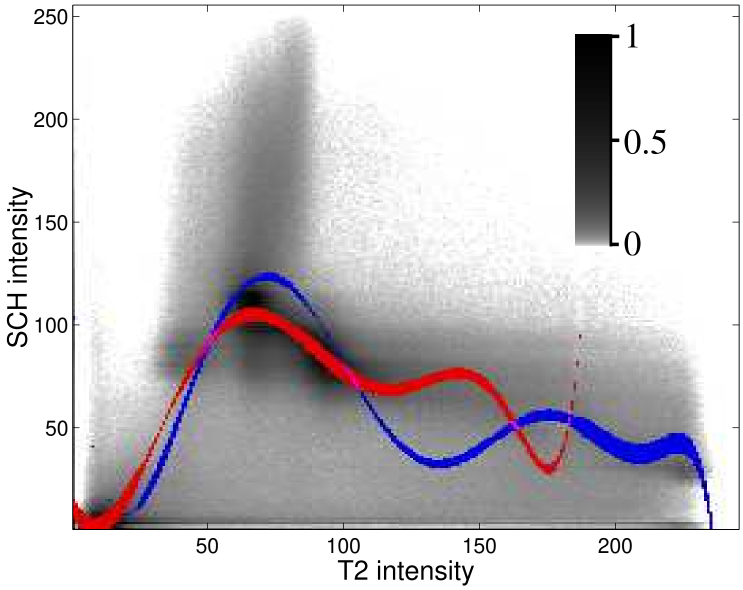


(c) $T2_a$ without intensity correction after non-rigid registration (d) $T2_a$ with intensity correction after non-rigid registration

FIG. 7.19 – *Corresponding sagittal views of $T2_a$ to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*

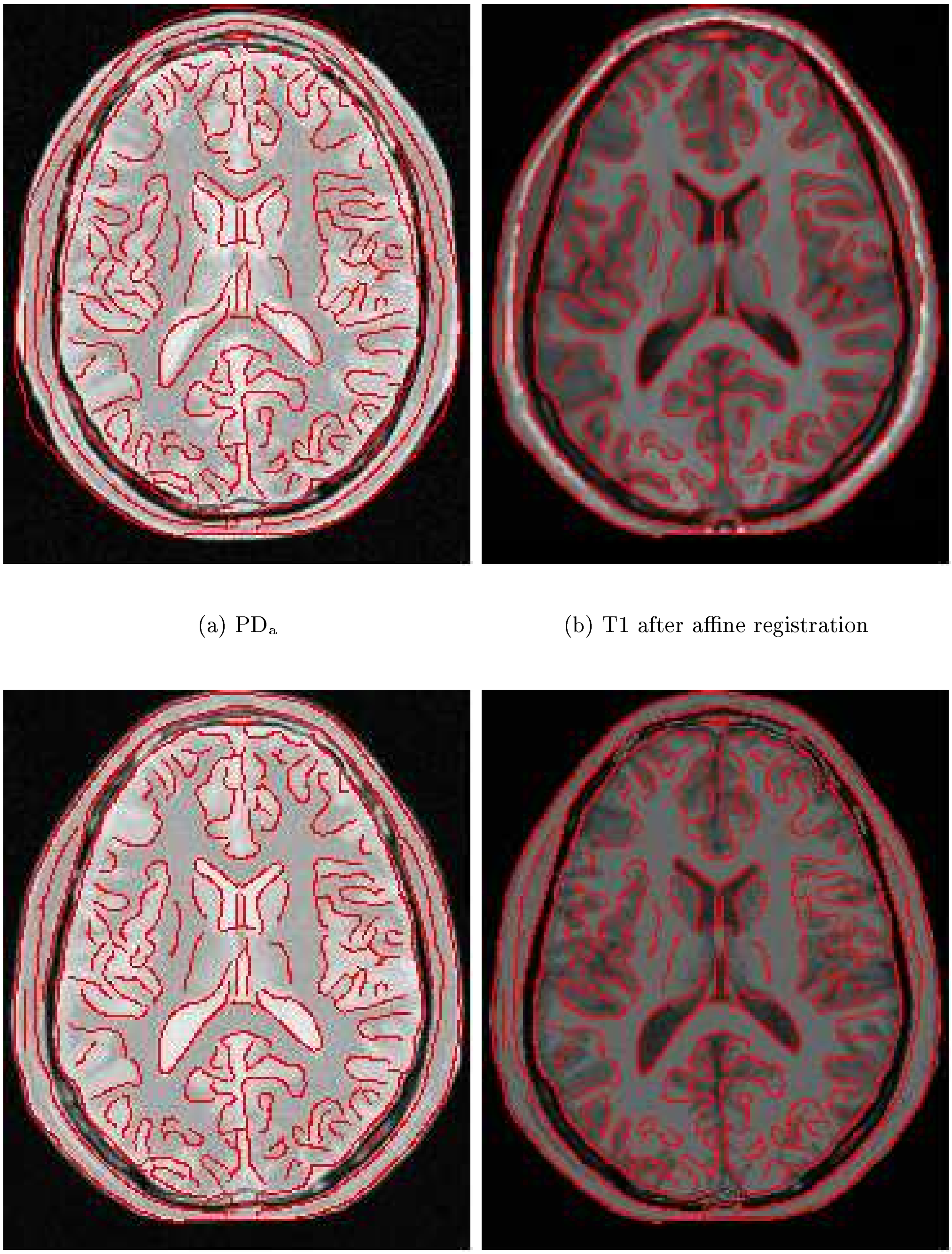


(a) Joint histogram values range from red representing high point densities to blue depicting low point densities.



(b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value j in $T2_a$ corresponds to the value of the corresponding proportion $\pi_\epsilon(j)$.

FIG. 7.20 – *Intensity transformation found when registering $T2_a$ with T1 and assuming bifunctional dependence.*



(a) PD_a

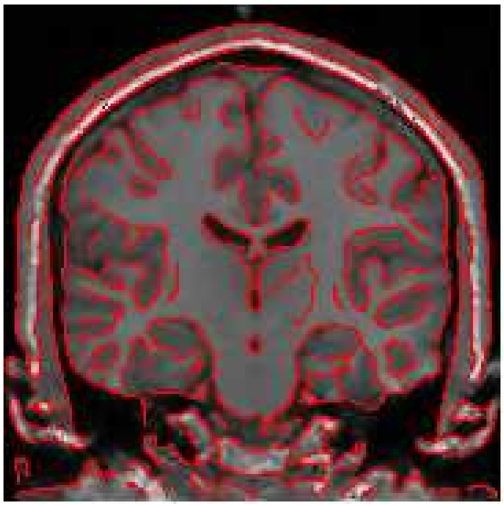
(b) T1 after affine registration

(c) PD_a without intensity correction after non-rigid registration

(d) PD_a with intensity correction after non-rigid registration

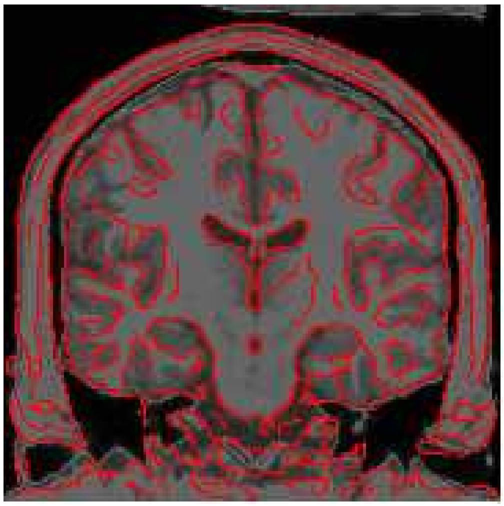
FIG. 7.21 – *Corresponding axial views of PD_a to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*

(a) PD_a



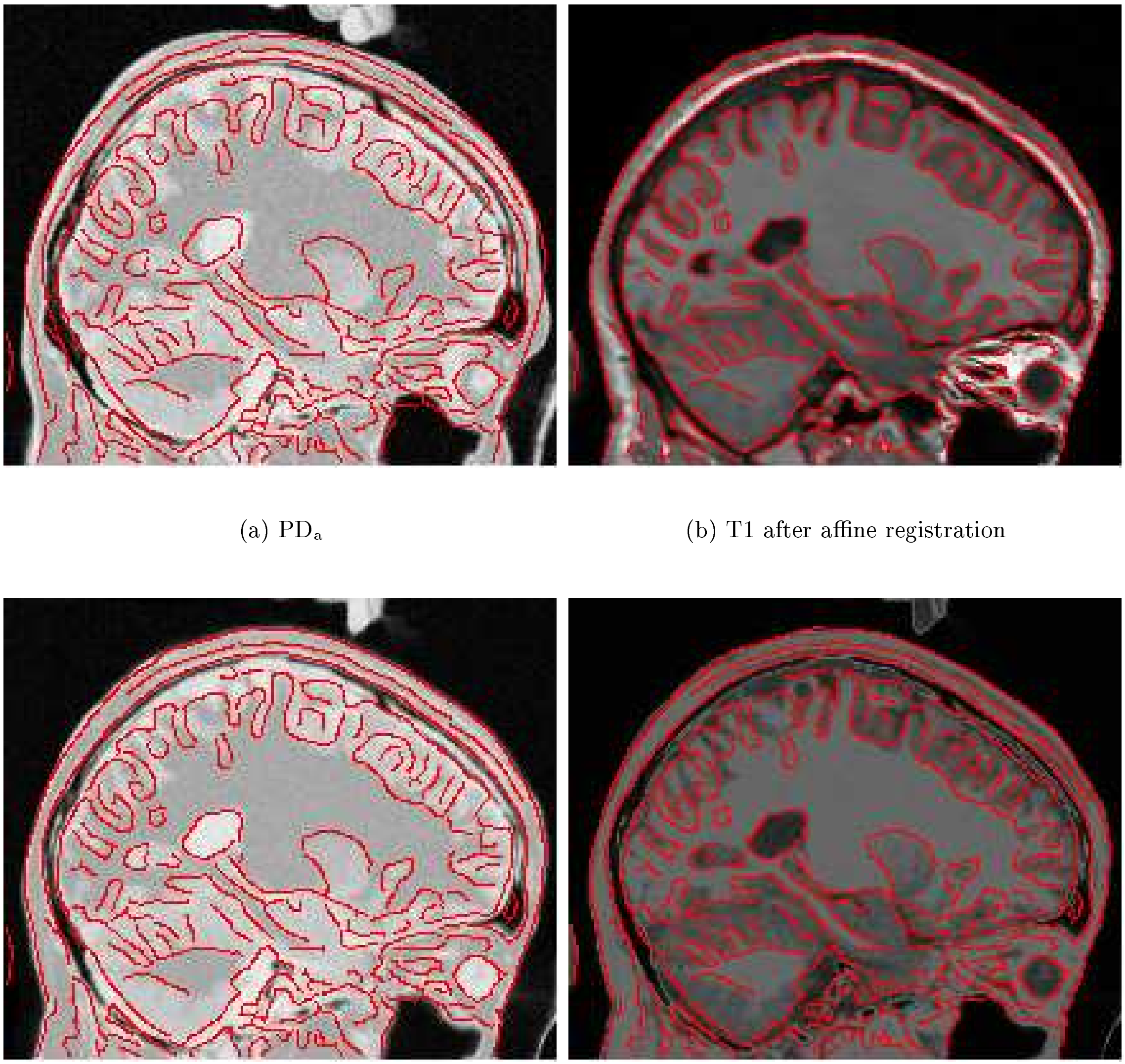
(b) T1 after affine registration

(c) PD_a without intensity correction after non-rigid registration



(d) PD_a with intensity correction after non-rigid registration

FIG. 7.22 – *Corresponding coronal views of PD_a to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*



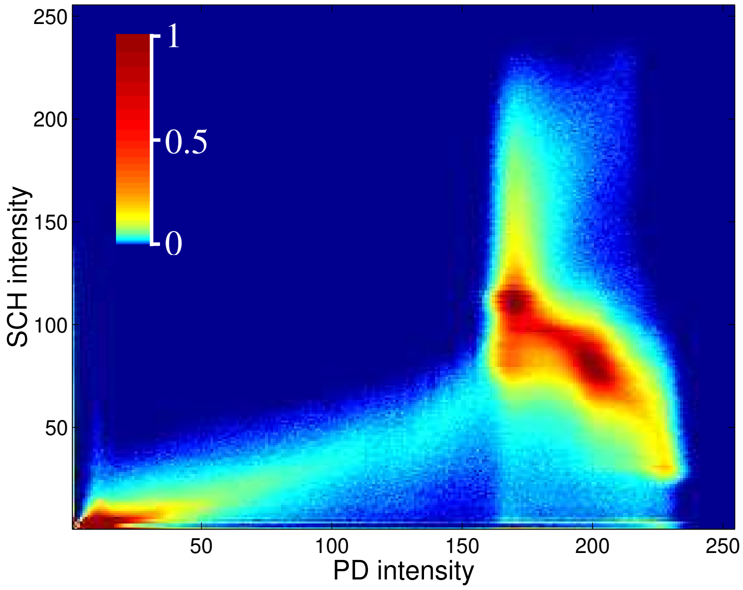
(a) PD_a

(b) T1 after affine registration

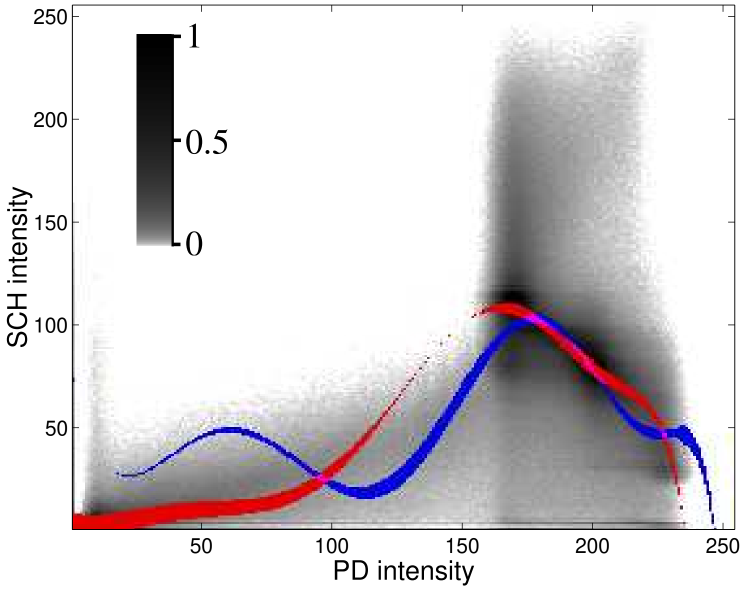
(c) PD_a without intensity correction after non-rigid registration

(d) PD_a with intensity correction after non-rigid registration

FIG. 7.23 – *Corresponding sagittal views of PD_a to T1 registration result. Contours were obtained using a Canny-Deriche edge detector on the T1 image (b) and overlaid on the other images to better assess the quality of registration.*



(a) Joint histogram values range from red representing high point densities to blue depicting low point densities.



(b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value j in $T2_a$ corresponds to the value of the corresponding proportion $\pi_\epsilon(j)$.

FIG. 7.24 – *Intensity transformation found when registering PD_a with T1 and assuming bifunctional dependence.*

the CSF cluster.

7.6.3 Mutual Information

We present in figures 7.25, 7.26, and 7.27 the result of registering $T1_a$ with CT using the mutual information method described in section 7.4. A typical difference between using our bifunctional method instead of our mutual information implementation can be appreciated by comparing figures 7.13 and 7.25. As can be seen, the contours of corresponding structures do not match after registration using mutual information. The head contours seem to be attracted by the image borders, which means that the driving forces have misleading directions in this region.

Many registrations were performed using this implementation of mutual information with several values for the control parameter λ and several Parzen window sizes. We could not obtain convincing results; in general, images deformed very little unless the λ parameter was set to a low value, in which case the displacements looked random. This outcome might be due to several reasons. First of all, it is likely that choosing a uniform λ is not the best achievable strategy to control the amplitude of the local forces. This may result in a bad conditioning of the force field, which may in turn hamper the estimation of the displacements. We also believe that Parzen windowing provides too little information on how to match intensities. As a consequence, the direction of the mutual information gradient, from which the local driving force is derived, might be unreliable.

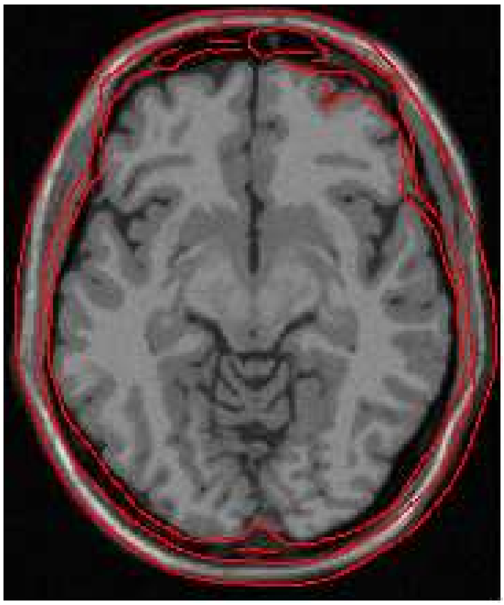
Although it would be irrelevant to conclude that mutual information is not well-suited for non-rigid registration, a partial explanation may be given. What mutual information does is to measure the similarity of the images from the joint probability density function (pdf) of their intensities. As discussed in chapter 5, section 5.3.4, there is actually a hidden estimation problem when using mutual information. We have shown that Parzen windowing closely relates to a maximum likelihood estimate of the joint pdf when it is not assumed to have a specified number of modes.

In the context of rigid/affine registration, where no more than 12 geometrical parameters need to be estimated, such an enormous search space is usually affordable. However, with non-rigid registration, the geometrical transformation is granted many degrees of freedom and maximizing mutual information might then become an under-constrained problem. While Parzen windowing is a non-parametric approach, our method explicitly restricts the intensity space using polynomial functions.

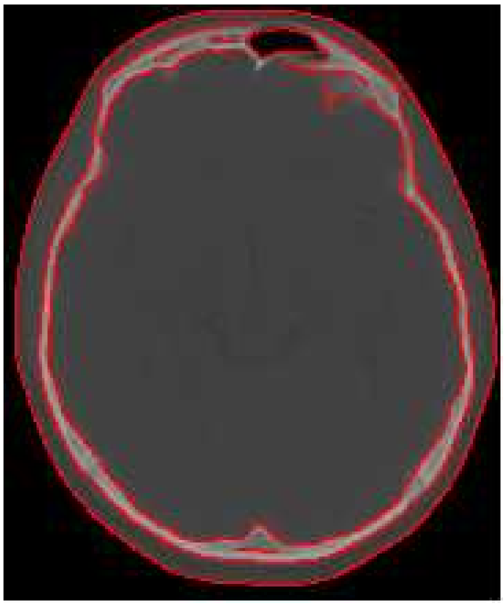
7.6.4 Displacement field comparison

Since the atlas, $T1_a$, $T2_a$ and PD_a have all been registered with T1, it is relevant to compare some statistics of the resulting displacement fields to assess if our algorithm provides consistent results across modalities.

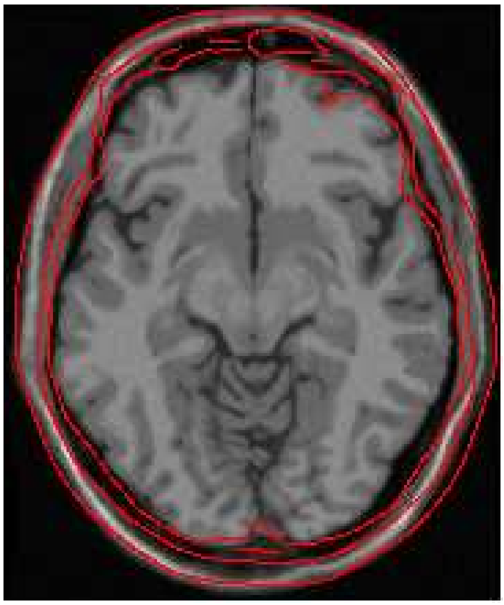
We computed statistics regarding the difference between any two of these displacement fields. The length of the vectors of the resulting difference fields were calculated. Each cell of table 7.2 presents, for each combination of displacement fields, the median length, the average length with the corresponding standard deviation and the maximum length of the difference field.



(a) $T1_a$

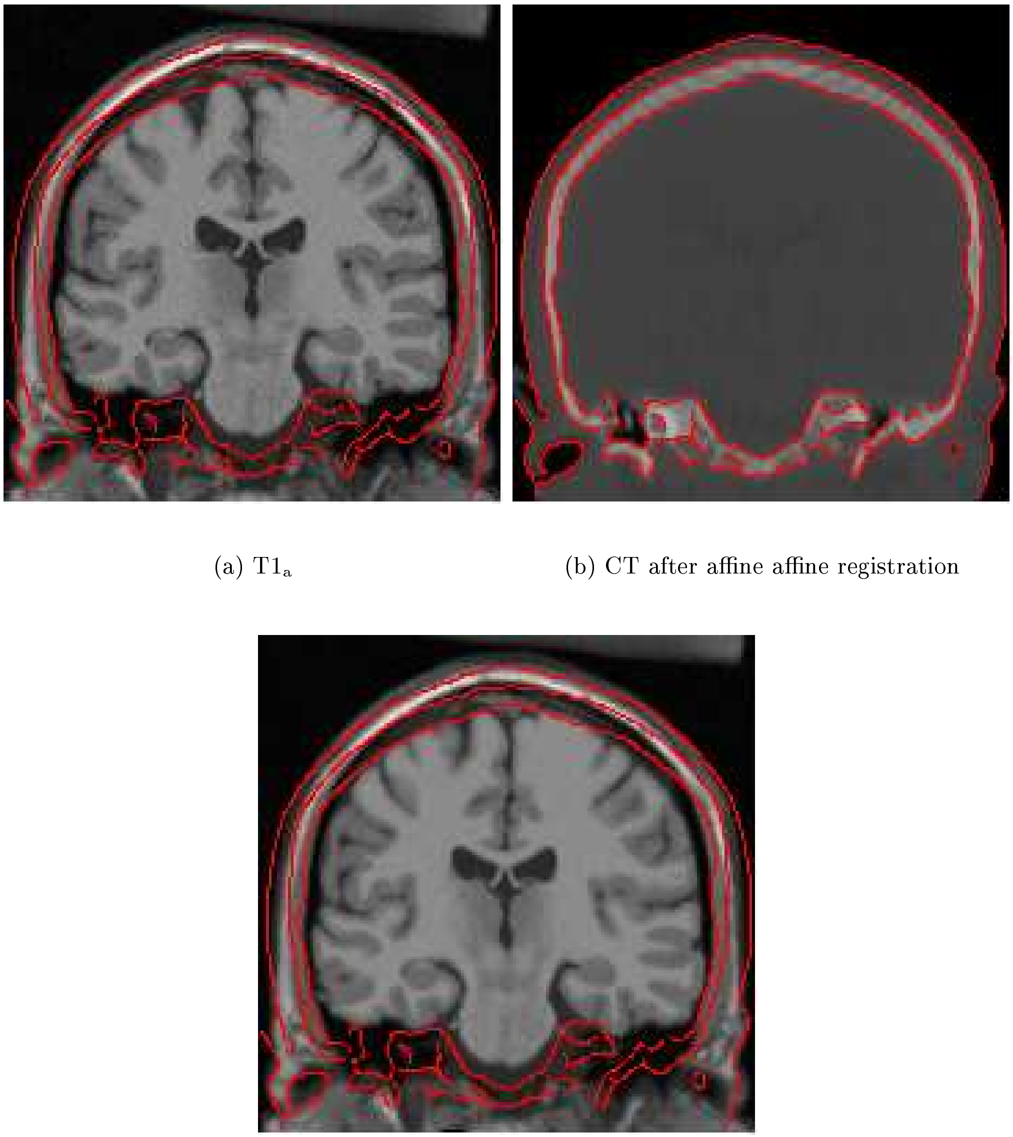


(b) CT after affine affine registration



(c) $T1_a$ after non-rigid registration

FIG. 7.25 – *Corresponding axial views of the $T1_a$ to CT registration result using mutual information. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*

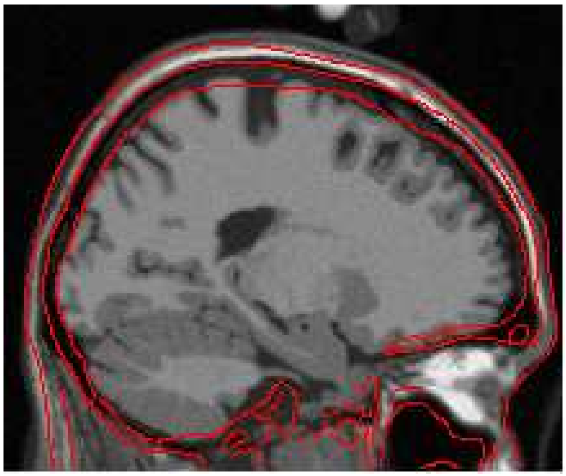


(a) $T1_a$

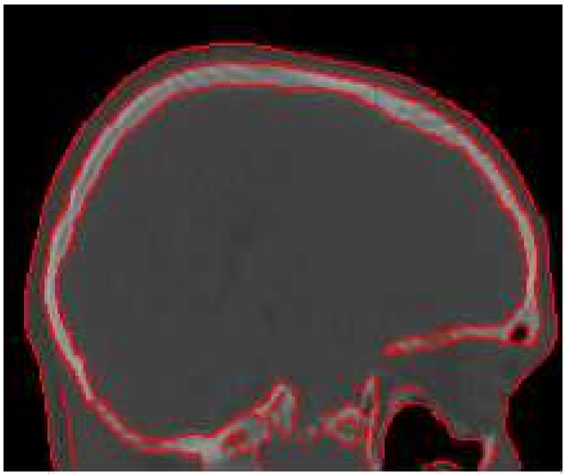
(b) CT after affine affine registration

(c) $T1_a$ after non-rigid registration

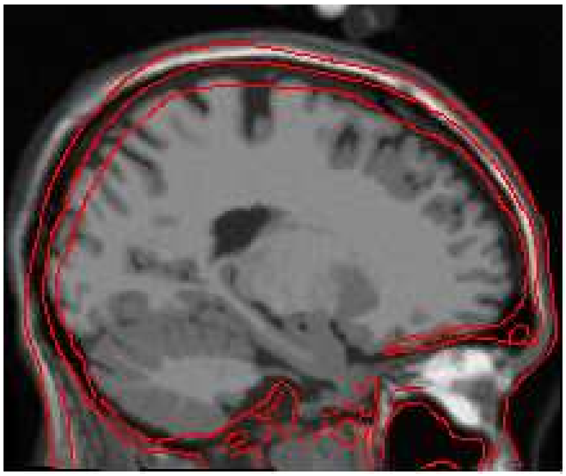
FIG. 7.26 – *Corresponding coronal views of the $T1_a$ to CT registration result using mutual information. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*



(a) $T1_a$



(b) CT after affine affine registration



(c) $T1_a$ after non-rigid registration

FIG. 7.27 – *Corresponding sagittal views of the $T1_a$ to CT registration result using mutual information. Contours were obtained using a Canny-Deriche edge detector on the CT image (b) and overlaid on the other images to better assess the quality of registration.*

Difference (mm)	Atlas-T1 $T1_a$-T1	Atlas-T1 $T2_a$-T1	Atlas-T1 PD_a-T1	$T1_a$-T1 $T2_a$-T1	$T1_a$-T1 PD_a-T1	$T2_a$-T1 PD_a-T1
median	1.46	1.13	1.67	1.00	1.01	1.32
average	1.58	1.23	1.76	1.18	1.16	1.40
std. dev.	0.84	0.63	0.79	0.78	0.71	0.68
maximum	6.99	5.14	7.10	7.17	8.08	6.86

TAB. 7.2 – *Statistics regarding the displacements difference between each type of registration. Each cell presents the median length, the average length with the corresponding standard deviation and the maximum length. All measures are in millimeters.*

The two largest average differences are 1.76mm and 1.58mm and were found when registering the atlas with $T1_a$ and PD_a respectively. This may be explained by the intensity correction bias for the CSF that would tend to attenuate displacements and produce larger errors, a problem invoked in section 7.6.1. Aside from these, the average error length varies between 0.97mm and 1.40mm and the median error is between 0.85mm and 1.32mm, with the largest errors occurring in the cortex area. These are values in the range of the image resolution of 1.0mm. Note also that all the standard deviations are below this value.

Also, we observe that the results obtained when registering images from different modalities (Atlas-T1, $T2_a$-T1, and PD_a-T1) seem to be consistent with the monomodal registration result ($T1_a$-T1), in which no intensity correction was performed. This suggests that the intensity correction may not cause a sensible degradation of the registration when compared to the monomodal case. We point out, however, that these are global measures that are presented to provide an idea of the differences between the displacement fields. They do not provide a validation of the method, but do show a certain coherence between the different results we obtained.

7.7 Conclusion

Nous avons abordé dans ce chapitre le problème du recalage multimodal et non-rigide. La méthode que nous avons développée est fondée sur un schéma itératif alternant le calcul du champ de déplacements et l'estimation d'une transformation d'intensité visant à corriger l'une des images de manière à autoriser une comparaison directe des deux ensembles d'intensités.

La premiére étape est mise en œuvre par une variante de l'algorithme des démons dû à (Thirion, 1995; Thirion, 1998). Nous avons discuté du lien qui unit cette technique à d'autres approches de la famille du flux optique : la minimisation de la SDC pénalisée et le recalage fluide proposé par (Christensen, 1994; Christensen et al., 1996).

Nous avons décrit deux stratégies possibles pour l'étape de correction d'intensité. L'une, fondée sur un modèle de dépendance mono-fonctionnelle, consiste en une technique classique de régression linéaire robuste. L'autre est fondée sur un modèle bi-fonctionnel et combine de manière ad-hoc deux régressions linéaires du même type. Pour l'une et l'autre de ces stratégies, nous avons constaté que la robustesse de l'estimation est un aspect essentiel dans la mesure où des erreurs de correction d'intensité peuvent entraîner de mauvais appariements géométriques qui seront difficilement corrigés par la suite.

Nous avons appliqué notre méthode à la mise en correspondance inter-sujets d'images 3D du cerveau. L'utilisation de données simulées a permis une comparaison objective des champs de déplacements obtenus par recalage multimodal et par recalage monomodal. Nos résultats indiquent une bonne cohérence, ce qui tend à prouver que la correction d'intensité a peu d'influence sur les performances du recalage. Ces résultats sont cependant préliminaires et une étude plus poussée est nécessaire à leur confirmation.

La validation reste un problème épineux concernant les méthodes de recalage non-rigide, d'autant plus lorsqu'elles sont appliquées à la mise en correspondance de sujets différents. Une approche classique est de repérer des structures équivalentes dans chacune des images à recaler, d'appliquer la transformation estimée à la segmentation de l'image source, et de comparer finalement cette segmentation «propagée» à la segmentation de l'image cible. Une telle étude a été menée par (Dawant et al., 1998; Dawant et al., 1999) pour l'algorithme des démons monomodal et aboutit à des conclusions globalement positives.

Enfin, nous avons implémenté une version non-rigide du recalage par information mutuelle fondée sur certains principes communs à la méthode précédente. Celle-ci n'a pas donné de résultats probants. La raison pourrait en être que cette deuxième approche procède d'hypothèses trop faibles concernant la relation entre les intensités des images. Néanmoins, elle peut facilement être adaptée à des modèles de dépendance plus contraints, ce qui laisse espérer de futures améliorations.

Chapitre 8

Conclusion

8.1 Bilan

Les travaux présentés dans ce mémoire portent sur les méthodes iconiques de recalage d'images médicales et, plus particulièrement, sur le choix de la mesure de similarité. Nous sommes partis d'un problème spécifique, le recalage multimodal ultrason/IRM, pour lequel nos expériences ont indiqué qu'aucune mesure de similarité de l'état de l'art n'était adaptée (voir le chapitre 6, paragraphe 6.5). Nous avons abordé la recherche d'une nouvelle mesure de similarité à travers deux formulations mathématiques distinctes.

La première formulation, décrite au chapitre 3, se fonde sur une hypothèse *heuristique* de dépendance fonctionnelle entre les intensités des images à recaler. Nous avons défini une mesure générique de liaison fonctionnelle basée sur une distance entre les intensités (plus généralement, un M-estimateur d'échelle) et normalisée de façon à tenir compte des effets de recouvrement partiel inhérents à toute procédure de recalage. Cette mesure, asymétrique par nature, est appelée *rapport de corrélation généralisé* car une de ses instanciations particulières s'identifie à la notion probabiliste connue de rapport de corrélation. La maximisation d'un rapport de corrélation généralisé revient à coupler un recalage monomodal avec un problème classique de régression linéaire.

Au chapitre 4, nous avons testé expérimentalement différentes mesures de similarité dans le cadre du recalage scanner/IRM et TEP/IRM d'images provenant de la base de données «Vanderbilt». Nos résultats indiquent que les performances de certaines versions du rapport de corrélation généralisé sont comparables à celles de l'information mutuelle, ce qui les situe au niveau des meilleures méthodes évaluées dans (West et al., 1997).

Le chapitre 5 propose une seconde formulation du recalage, fondée sur le principe du maximum de vraisemblance et généralisant en partie celle du chapitre 3. Cette approche consiste à identifier les paramètres d'un modèle de dépendance inter-images que l'on s'est donné *a priori*. Les paramètres inconnus caractérisent non seulement la relation géométrique entre les images mais aussi la relation entre leurs intensités. Ces derniers paramètres sont considérés comme auxiliaires bien que leur estimation fasse partie intégrante du processus de recalage. Nous avons montré que le rapport de corrélation quadratique ainsi que l'information mutuelle au sens de (Viola, 1995) sont en fait, à quelques approximations raisonnables près, des critères de vraisemblance particuliers.

De façon beaucoup plus générale, le recalage par maximum de vraisemblance permet de repenser le concept de mesure de similarité et de le situer dans la perspective d'une modélisation explicite des procédés d'imagerie. Bon nombre de mesures de similarité usuelles peuvent être vues dans ce contexte. Du fait qu'elles peuvent le plus souvent être appréhendées dans le formalisme d'un histogramme conjoint, on peut les relier à des hypothèses simplificatrices dont les principales sont les suivantes :

- l'indépendance spatiale de la scène anatomique commune aux images. Cette hypothèse revient à n'incorporer aucune connaissance *a priori* concernant la structure géométrique des tissus représentés.
- l'indépendance spatiale du bruit affectant les images. Cette hypothèse est discutable par exemple dans le cas du scintillement (*speckle*) présent dans les images ultrasonores.
- la stationnarité du bruit, hypothèse qui revient à négliger les inhomogénéités telles que le biais du champ magnétique en IRM ou les effets d'ombrage en imagerie ultrasonore.

Moyennant certaines simplifications concernant la physique d'acquisition des images ultrasonores et IRM, nous avons justifié au chapitre 6 un modèle de dépendance ultrason/IRM du type considéré au chapitre 5. Ce modèle exprime fondamentalement une correspondance fonctionnelle entre, d'une part, l'intensité de l'IRM et son gradient projeté sur la ligne d'acquisition ultrasonore et, d'autre part, l'intensité de l'image ultrasonore. Nous en déduisons que la mesure correspondant (approximativement) à un recalage par maximum de vraisemblance est un rapport de corrélation bivarié, fondé sur la métrique quadratique dans la mesure où nous faisons l'hypothèse que le scintillement suit une loi gaussienne. Nous justifions ensuite le passage à une métrique robuste pour pallier aux insuffisances du modèle de dépendance (indépendance spatiale du speckle, omission des effets d'ombrage,...).

En pratique, nous calculons le rapport de corrélation bivarié avec la norme du gradient de l'IRM et non, comme la théorie le suggère, avec la projection du gradient sur la ligne d'acquisition ultrasonore, cette dernière étant inconnue à l'avance. Le rapport de corrélation bivarié a été testé dans le cadre du recalage *rigide* d'images ultrasonores et IRM et il a fait l'objet d'une étude préliminaire de validation croisée utilisant un certain nombre de données fantômes et cliniques (voir le chapitre 6, paragraphes 6.3 à 6.5). Les résultats indiquent que les pires erreurs de recalage sont de l'ordre de la résolution spatiale de l'IRM. Cependant, il apparaît qu'une initialisation relativement précise est nécessaire (pour donner un ordre de grandeur, le décalage initial entre les images ne doit pas excéder 10 mm et 10 deg).

Enfin, dans la perspective d'étendre le rapport de corrélation bivarié au cas du recalage non-rigide, nous avons présenté au chapitre 7 une étude de faisabilité du recalage multimodal non-rigide. L'idée de départ était de minimiser une énergie composée d'un rapport de corrélation (monovarié) et d'un stabilisateur quadratique destiné à régulariser la transformation spatiale. Compte tenu des difficultés particulières que la recherche d'une transformation non-rigide suscite, nous avons été amenés à modifier cette formulation de manière ad-hoc. L'algorithme que nous proposons ne minimise pas strictement une énergie comme c'est le cas dans une formulation rigide/affine ; on peut le voir comme une adaptation mutimodale de l'algorithme des «démons» (Thirion, 1998).

Les résultats que nous avons obtenus en recalant différentes modalités d'images cérébrales (atlas labélisé, IRM, scanner) sont visuellement probants. En outre, l'utilisation de volumes IRM simulés par la méthode développée au Montreal Neurological Institute a permis de montrer que la procédure de recalage multimodal non-rigide fournit des champs de déformations cohérents avec la procédure monomodale bien établie des «démons».

8.2 Perspectives

8.2.1 Recalage non-rigide ultrason/IRM

Le recalage non-rigide ultrason/IRM est nécessaire pour corriger les distorsions géométriques dans l'image ultrasonore. Ces distorsions, que nous avons considérées comme négligeables dans notre étude, sont notamment liées au fait que la formation d'une image ultrasonore se fonde sur l'hypothèse simplificatrice que l'onde acoustique est à célérité constante dans le milieu traversé. Par ailleurs, lorsque des images ultrasonores sont acquises au cours d'une opération du cerveau, ce qui était le contexte de notre étude, le recalage non-rigide permet de compenser les déformations dues aux variations de pression du liquide céphalo-rachidien et aux interactions avec le chirurgien.

Les résultats de recalage multimodal non-rigide que nous avons obtenus (voir le chapitre 7) laissent espérer que la méthode de recalage rigide ultrason/IRM présentée au chapitre 6 pourra être étendue au cas non-rigide. Une adaptation immédiate, que nous n'avons cependant pas encore testée, consisterait à généraliser la correction d'intensité robuste fondée sur une hypothèse de dépendance mono-fonctionnelle (voir le paragraphe 7, paragraphe 7.3.1), de manière à rechercher une fonction de deux variables : l'intensité de l'IRM et le gradient de l'IRM projeté sur la ligne d'acquisition ultrasonore. Dans la mesure où le recalage non-rigide interviendrait après un recalage rigide, la position et l'orientation de la sonde ultrasonore serait connue dans le repère IRM, ce qui permettrait de calculer effectivemment le gradient projeté et d'éviter l'approximation consistant à utiliser la norme du gradient.

8.2.2 Des algorithmes plus robustes

Parmi les limitations pratiques des algorithmes que nous avons testés, la robustesse vis-à-vis de l'initialisation nous semble être l'aspect sur lequel d'importantes améliorations sont encore nécessaires dans la perspective d'une automatisation complète du recalage d'images. Plus encore que le problème de la *précision*, le problème de la robustesse s'est posé de manière particulièrement sensible dans le cas du recalage ultrason/IRM. Il est commun à de très nombreux autres problèmes de recalage.

Le recalage par maximum de vraisemblance fournit un cadre théorique pour définir des mesures de similarité mieux contraintes et, dans le même temps, plus spécifiques aux applications visées. Ce type d'approche, combinée à une optimisation robuste, pourrait permettre d'améliorer considérablement la robustesse des algorithmes actuels.

Des modèles d'acquisition plus réalistes La méthodologie du recalage par maximum de vraisemblance pourrait être appliquée à des modèles d'acquisition d'images plus réalistes. La problématique qui en découle est la prise en compte des artefacts ne vérifiant pas les hypothèse de stationnarité et d'indépendance spatiale rappelées ci-dessus. En effet, le modèle générique que nous avons proposé au chapitre 5 ne tient pas compte de ce type d'artefacts. Mais il doit être clair que le *principe* du recalage par maximum de vraisemblance est très général et qu'il n'impose aucune restriction aux modèles de dépendance inter-images que l'on est susceptible de considérer.

Pour reprendre le fil rouge de ce mémoire, nous imaginons aisément le bénéfice que le recalage ultrason/IRM pourrait tirer d'un modèle de scintillement dans l'image ultrasonore qui soit spatialement corrélé, de la prise en compte du biais dans l'image IRM, ... D'une certaine manière, nous avons tenu compte de ces artefacts à travers l'emploi d'une métrique robuste pour comparer les intensités (voir le chapitre 6, paragraphe 6.2.3). Cette approche revient en quelque sorte à rejeter l'information fournie par les images dès lors qu'elle n'est pas conforme au modèle utilisé. De ce point de vue, il semble évident qu'une modélisation plus fine conduira inéluctablement à une meilleure exploitation de l'information et, par là-même, à une mesure de similarité plus efficace.

Prise en compte de corrélations spatiales Le modèle de dépendance inter-images sur lequel repose l'estimation par maximum de vraisemblance fait intervenir non seulement des processus d'acquisition mais aussi une représentation de la scène anatomique commune aux images (ou supposée telle). Comme nous l'avons vu au chapitre 5 et en particulier au paragraphe 5.4.1, le recalage par maximum de vraisemblance conduit à une estimation implicite de cette scène commune. On peut donc dire que l'estimation par maximum de vraisemblance offre un cadre naturel pour résoudre *simultanément* le problème du recalage et de la segmentation, l'un aidant l'autre et réciproquement. Au fond, cela n'a rien d'étonnant car l'objet du recalage est fondamentalement d'apparier des structures anatomiques équivalentes.

On a bien compris qu'il n'est pas nécessaire de réaliser une segmentation parfaite pour obtenir un recalage satisfaisant. Toutefois, il est raisonnable de penser que le recalage est d'autant plus aisé que la segmentation est précise. Dans cette optique, on souhaiterait décrire la scène anatomique comme un champ spatialement corrélé ; typiquement, un champ de Gibbs-Markov. Bien que nous ayons montré la possibilité d'une telle approche au chapitre 5, il reste sans doute un travail important à faire pour en concevoir une implémentation raisonnable en temps de calcul.

Les modèles markoviens traduisent une certaine connaissance *a priori* concernant la géométrie des tissus anatomiques *via* une contrainte de continuité spatiale. On pourrait également envisager des contraintes géométriques plus explicites à condition d'adopter un modèle de scène qui soit lui-même paramétré de manière géométrique. Un tel modèle pourrait, par exemple, être obtenu à partir d'une description par modèles déformables des interfaces entre structures. Son caractère aléatoire serait hérité de notions de statistique de formes. Bien que beaucoup plus délicate d'un point de vue mathématique, une telle approche nous semble prometteuse et pourrait conduire, à terme, à remettre en question la dichotomie usuelle entre méthodes iconiques et méthodes géométriques de recalage.

Optimisation L'optimisation est un sujet que nous avons trop peu abordé au cours de ce mémoire, ayant préféré nous concentrer sur le *critère* à optimiser. Il n'en reste pas moins que c'est un aspect essentiel des méthodes iconiques de recalage et qu'il en conditionne directement la robustesse vis-à-vis de l'initialisation. Disons les choses naïvement : l'optimisation serait un faux problème si on était capable d'exhiber des mesures de similarité qui soient systématiquement convexes en tant que fonctions des paramètres de transformation spatiale. Malheureusement, nous n'avons pas trouvé de telles mesures et, pour tout dire, nous doutons sérieusement de leur existence-même – mais peut-être la question mérite-t-elle d'être posée.

Au chapitre 2, paragraphe 2.5.3, nous avons brièvement exposé certaines méthodes d'optimisation qui exploitent la structure mathématique particulière de la mesure de similarité à optimiser. Il s'agissait en l'occurrence de mesures simples de type SDC pouvant être calculées à partir d'une image de différence. Les méthodes dont nous avons parlé sont l'algorithme ICP 4D proposé par (Feldmar et al., 1997) ainsi que les algorithmes d'appariement de régions (*block-matching*).

Dans l'équipe Epidaure, nous avons commencé une étude visant à comparer les méthodes d'optimisation conventionnelles du type de Powell avec des techniques d'appariement de régions. Nos premiers résultats, dont certains ont été publiés dans (Ourselin et al., 2000a), tendent à montrer la supériorité de ces dernières en termes de robustesse à l'initialisation. Il nous semble donc que la philosophie d'une optimisation *spécifique* est particulièrement recommandée. Il reste à généraliser les approches d'appariement de régions ou ICP 4D à des mesures de similarité plus complexes telles que, par exemple, le rapport de corrélation généralisé ou d'autres critères de vraisemblance...

Annexes

Annexe A

Couples de variables aléatoires et conditionnement

Cette annexe donne quelques éléments sur l'étude des couples de variables aléatoires de carré sommable et replace la notion de rapport de corrélation dans son contexte probabiliste originel. Pour de plus amples détails, nous conseillons l'ouvrage de (Saporta, 1990). La dernière section, A.7, fournit une interprétation probabiliste de la mesure de similarité proposée par (Woods et al., 1993).

A.1 L^2 space

Given a probability space $(\Omega,\mathcal{T},Pr)$, L^2 is defined as the space of square integrable real-valued random variables [1], that is the variables which verify:

$$E(X^2) = \int_\Omega X^2 \, dPr < +\infty,$$

where E denotes the expectation operator. One shows that L^2 is a Hilbert space with respect to the dot product $\langle X,Y \rangle = E(XY)$. Thus, the corresponding norm is the second-order moment of a variable:

$$\|X\| = \sqrt{E(X^2)}. \tag{A.1}$$

The L^2 norm is closely related to the classical notions of expectation, variance and standard deviation. Equation (A.1) can be rewritten:

$$\|X - E(X)\| = \sqrt{\mathrm{Var}(X)} = \mathrm{StdDev}(X).$$

Due to its Hilbertian structure, L^2 has interesting geometric properties. One can define a notion of orthogonality between two variables:

$$X \perp Y \iff E(XY) = 0.$$

1. In fact, L^2 is the quotient space of square integrable variables with respect to the "almost everywhere" equivalence relationship. This distinction has no importance here.

The meaning of orthogonality in L^2 relates with the notion of independence, but in a less restrictive way. Recall that two variables X and Y are said to be independent if their joint probability density function (pdf) is equal to the product of their marginal pdf's, that is $p(x,y) = p(x)p(y)$. One easily shows that $E(XY) = E(X)E(Y)$ for two such variables. Thus:

$$(X,Y) \text{ independent } \Rightarrow \; X - E(X) \;\perp\; Y - E(Y).$$

However, the converse is false. Orthogonality in L^2 is a weaker constraint than independence. It may be seen as a notion of *average* independence.

In a general way, the angle α between two variables X and Y is defined thanks to a basic property of dot products:

$$\langle X,Y \rangle = \|X\| \|Y\| \cos\alpha. \tag{A.2}$$

A.2 Expectation

L^2 contains the one-dimensional space Δ of deterministic variables, i.e. variables which are constant on the state space Ω. Given a variable X, its expectation is:

$$E(X) = \int_{\mathbb{R}} x\, p(x)\, dx = \langle X,1 \rangle.$$

Therefore, $E(X)$ is nothing but the orthogonal projection of X onto Δ. In the sense of the L^2 norm, it is the constant variable which best approximates X (classical notion of mean).

$$E(X) = \arg\min_{C \in \Delta} \|X - C\|^2.$$

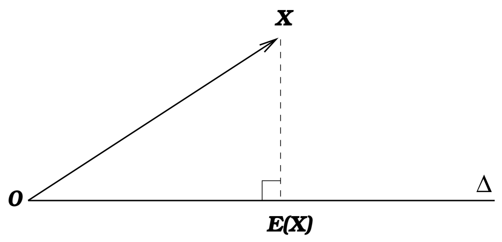


FIG. A.1 – *Geometric interpretation of the expectation operator as an orthogonal projection.*

A.3 Correlation coefficient

A fast method to evaluate (approximately) the degree of dependence between two variables is to compute their correlation coefficient. Given two variables X and Y, this is defined as:

$$\rho^2(X,Y) = \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{[E(XY) - E(X)E(Y)]^2}{[E(X^2) - E(X)^2]\,[E(Y^2) - E(Y)^2]}.$$

From a geometric point of view, we write it:

$$\rho^2(X,Y) = \frac{\langle\, X - E(X),Y - E(Y)\,\rangle^2}{\|X - E(X)\|^2\|Y - E(Y)\|^2}.$$

Using equation (A.2), the correlation coefficient between X and Y can be interpreted in a geometric way. Let α denote the angle between $X - E(X)$ and $Y - E(Y)$. We have:

$$\rho^2(X,Y) = \cos^2\alpha.$$

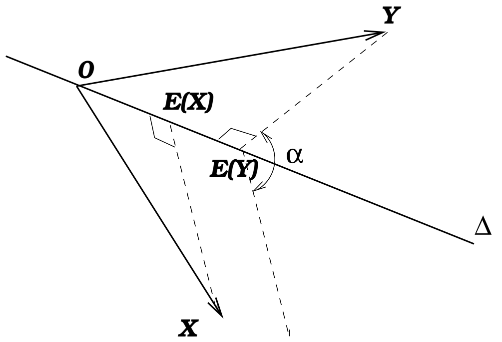


FIG. A.2 – *Geometric interpretation of the correlation coefficient. The constant (or deterministic) direction is denoted by* Δ*.*

We see that $\rho^2(X,Y)$ is larger as the angle α is small. It reaches 1 if $X - E(X)$ and $Y - E(Y)$ are colinear. This is to say that the correlation coefficient measures the *linear* dependence between two variables. As we want to take into account general functions between X and Y, possibly non-linear and non-monotonic, this is not a good measure of *functional* dependence.

A.4 Conditional expectation

Evaluating the functional dependence between two variables comes down to an interpolation problem with no constraint. Suppose we want to estimate a variable Y in terms of another variable X. A natural approach would be: (1) find the function $\phi^*(X)$ that best fits Y among all possible functions of X; (2) quantify the quality of the estimate $\phi^*(X)$ with respect to Y. The notion of conditional expectation provides a straightforward method for such an evaluation, without having to test every possible function of X.

If X and Y are not independent, knowing an outcome $X = x$ should provide some new information about Y. Any outcome $X = x$ induces a conditional pdf for Y, that is:

$$p(y|x) \equiv \begin{cases} \dfrac{p(x,y)}{p(x)} & \text{if } p(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, the corresponding a posteriori expectation of Y is:

$$\phi^*(x) = E(Y|X=x) = \int y\, p(y|x)\, dy.$$

To any possible outcome of X corresponds an a posteriori expected value for Y. Doing so, we define a function of X, which is the conditional expectation of Y in terms of X:

$$E(Y|X) = \phi^*(X).$$

Notice that $E(Y|X)$ is *also* a random variable. It is easy to verify that it is an unbiased estimate, i.e., $E\left[E(Y|X)\right] = E(Y)$. The major interest of the conditional expectation results from the following theorem:

Theorem A.1 *$E(Y|X)$ is the optimal approximator in the sense of the L^2 norm, that is:*

$$\phi^* = \arg\min_{\phi}\ \|Y - \phi(X)\|.$$

Proof. Let H_x be the space of functions which are square integrable with respect to the density measure $p(x)$:

$$H_x = \left\{ \phi : \mathbb{R} \to \mathbb{R} \ \middle|\ \int_{\mathbb{R}} \phi^2(x)\, p(x)\, dx \ < +\infty \right\}.$$

It is an Hilbert space when provided with the dot product $\langle \phi,\psi \rangle = \int_{\mathbb{R}} \phi(x)\, \psi(x)\, p(x)\, dx$. We want to minimize the following criterion on H_x:

$$J(\phi) = \int_{\mathbb{R}}\int_{\mathbb{R}} \left[y - \phi(x)\right]^2\ p(x,y)\, dxdy. \tag{A.3}$$

Clearly, J is convex. Therefore, minimizing equation (A.3) is equivalent to finding ϕ^* such that $\nabla J\left(\phi^*\right) = 0$. We express the directional derivative of J with respect to ϕ:

$$\begin{aligned}
\frac{dJ}{d\phi}(\phi,f) &= -2\int_{\mathbb{R}}\int_{\mathbb{R}} \left[y - \phi(x)\right]\ f(x)\, p(x,y)\, dxdy, \\
&= -2\int_{\mathbb{R}} \left[\left(\int_{\mathbb{R}} y\, p(y|x)\, dy \right) - \phi(x) \right]\ f(x)\, p(x)\, dx, \\
&= 2\left\langle \phi - \int_{\mathbb{R}} y\, p(y|x)\, dy\, ,\, f \right\rangle_{H_x} \\
\Longrightarrow &\quad \nabla J\left(\phi\right) = 2\left(\phi - \int_{\mathbb{R}} y\, p(y|x)\, dy \right)
\end{aligned}$$

The gradient is zero iff $\phi(x) = \int_{\mathbb{R}} y\, p(y|x)\, dy = E(Y|x)$. □

A.5 Total variance theorem

Let us now give a geometric interpretation of the conditional expectation. We consider the sub-space L^2_x of the variables that are functions of X (provided that they remain in L^2):

$$L^2_x = L^2 \cap \{\phi(X) \mid \phi : \mathbb{R} \to \mathbb{R}\}.$$

Notice that every constant variable is a (constant) function of X, so that $\Delta \subset L^2_x$. Since the conditional expectation $E(Y|X)$ minimizes the distance between Y and L^2_x, $E(Y|X)$ is the orthogonal projection of Y onto L^2_x (see figure A.3). This is due to the Hilbertian structure of L^2. This simple

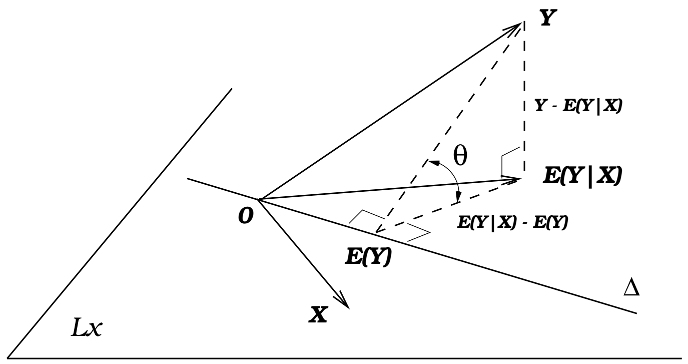


FIG. A.3 – *Geometric interpretation of the conditional expectation. It is the orthogonal projection onto L_x^2.*

geometrical property allows us to compute the distance between Y and L_x^2. Indeed, $Y - E(Y|X)$ is orthogonal to any vector of L_x^2 by definition of the orthogonal projection. Notably,

$$Y - E(Y|X) \;\perp\; E(Y|X) - E(Y).$$

Therefore, the triangle whose vertices are Y, $E(Y)$, and $E(Y|X)$ is right-angled in $E(Y|X)$. Applying the Pythagorean theorem, we retrieve a result known as the total variance theorem:

$$\begin{aligned}
\mathrm{Var}(Y) &= \|Y - E(Y)\|^2, \\
&= \|E(Y|X) - E\left[E(Y|X)\right]\|^2 + \|Y - E(Y|X)\|^2, \\
&= \|E(Y|X) - E(Y)\|^2 + \int_{\mathbb{R}}\int_{\mathbb{R}} \left[y - \phi^*(x)\right]^2\, p(x,y)\, dx dy, \qquad \text{(A.4)} \\
&= \mathrm{Var}\left[E(Y|X)\right] + \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left[y - \phi^*(x)\right]^2\, p(y|x)\, dy\right) p(x)\, dx. \qquad \text{(A.5)}
\end{aligned}$$

Step (A.4) relies on the fact that $E\left[E(Y|X)\right] = E(Y)$. Let us denote:

$\mathrm{Var}(Y|X=x) \equiv \int_{\mathbb{R}} \left[y - \phi^*(x)\right]^2\, p(y|x)\, dy,$

E_X, the operator defined by: $\forall \psi : \mathbb{R} \to \mathbb{R},\;\; E_X(\psi) = \int_{\mathbb{R}} \psi(x)\, p(x)\, dx.$

Then, equation (A.5) can be rewritten in a more compact way:

$$\mathrm{Var}(Y) = \mathrm{Var}\left[E(Y|X)\right] + E_X\left[\mathrm{Var}(Y|X=x)\right].$$

This may be seen as an energy conservation equation. The variance of Y is decomposed as a sum of two energy terms:

- $\mathrm{Var}\left[E(Y|X)\right]$ is the variance of the conditional expectation $E(Y|X)$. It measures the part of Y which is predicted by X.
- Conversely, the term $E_X\left[\mathrm{Var}(Y|X=x)\right]$, which is called the conditional variance, represents the squared distance of Y to the space L_x^2. It measures the part of Y which is functionally independent of X.

A.6 Correlation ratio

We now can design a measure of functional dependence between X and Y. Accounting for the interpretation of the total variance theorem in terms of energy, it seems natural to compare the "explained" energy of Y with its total energy. This leads to the definition of the correlation ratio:

$$\eta^2(Y|X) = \frac{\mathrm{Var}\,[E(Y|X)]}{\mathrm{Var}(Y)}.$$

The correlation ratio also has a simple geometric interpretation. Let θ denote the angle between $Y - E(Y)$ and the space L^2_x. By definition, θ is also the angle between $Y - E(Y)$ and $E(Y|X) - E(Y)$ (see figure A.3). We have:

$$\cos^2\theta = \frac{\mathrm{Var}\,[E(Y|X)]}{\mathrm{Var}(Y)}, \qquad \sin^2\theta = \frac{E_X\,[\mathrm{Var}(Y|X=x)]}{\mathrm{Var}(Y)}.$$

Then: $\eta^2(Y|X) = \cos^2\theta = 1 - \sin^2\theta$.

Unlike the correlation coefficient which measures the *linear* dependence between two variables, the correlation ratio measures the *functional* dependence. It takes on values between 0 and 1. A value near 1 indicates a high functional dependence, while a value near 0 indicates a low functional dependence. The two extreme cases are:

$$\begin{aligned}
\eta^2(Y|X) = 1 &\iff \exists\phi, \quad Y = \phi(X),\\
\eta^2(Y|X) = 0 &\iff E(Y|X) = \mathrm{cte} = E(Y).
\end{aligned}$$

By nature, the correlation ratio is asymmetric since the two variables fundamentally do not play the same role in the functional relationship. Thus, in general $\eta^2(Y|X) \neq \eta^2(X|Y)$.

Some additional properties of the correlation ratio are summarized below:

1. $\eta^2(Y|X) = \rho^2\,[E(Y|X),Y]$
2. $\eta^2(Y|X) \geq \rho^2(X,Y)$
3. If (X,Y) is Gaussian, $\eta^2(Y|X) = \eta^2(X|Y) = \rho^2(X,Y)$
4. $\eta^2(Y|X) = \rho^2(X,Y)$ iff $E(Y|X)$ is linear[2] with respect to X, i.e.

$$\exists(\alpha,\beta) \quad E(Y|X) = \alpha + \beta X$$

A.7 Woods criterion

In this framework, we can interpret the heuristic criterion devised by (Woods et al., 1993) as another measure of functional dependence. According to the previous notations, the Woods criterion may be written as follows:

$$W(Y|X) = E_X\left(\frac{\sqrt{\mathrm{Var}(Y_x)}}{E(Y_x)}\right), \tag{A.6}$$

2. To be accurate, one should rather say *affine* instead of *linear*.

where Y_x is the conditional variable induced on Y by the outcome $X = x$.

The dispersion measure $\sqrt{\mathrm{Var}(Y_x)}/E(Y_x)$ is a normalized standard deviation. It has the property of being invariant to multiplicative changes in the Y variable. Equation (A.6) should be compared to the following depiction of the correlation ratio:

$$1 - \eta^2(Y|X) = \frac{E_X\left[\mathrm{Var}(Y_x)\right]}{\mathrm{Var}(Y)}. \tag{A.7}$$

Though different, equation (A.6) and equation (A.7) express the same basic idea. There are two main differences. First, the correlation ratio sums up variances while the Woods criterion sums up normalized standard deviations. Second, normalization is achieved in the correlation ratio via a global division by $\mathrm{Var}(Y)$. In the Woods criterion, every term of the sum is divided by a mean.

A geometric interpretation of the Woods criterion may be given by noticing that the ratio in equation (A.6) corresponds to a tangent (see figure A.4):

$$\begin{aligned} \frac{\sqrt{\mathrm{Var}(Y_x)}}{E(Y_x)} &= \tan\theta_x, \\ W(Y|X) &= E_X(\tan\theta_x). \end{aligned}$$

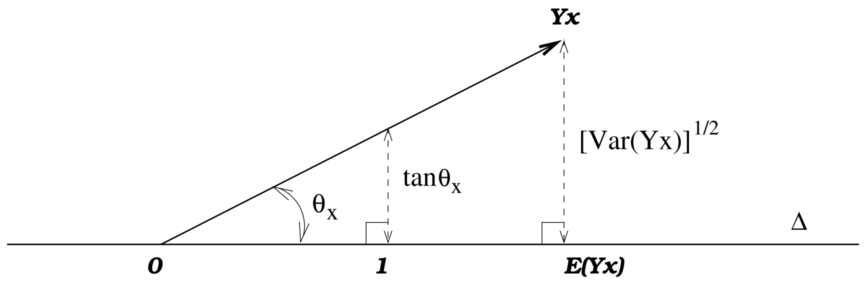


FIG. A.4 – *Geometric interpretation of the Woods criterion. It is the average tangent between the conditional variables Y_x and the constant direction Δ.*

One may wonder if the Woods criterion is equivalent to the correlation ratio. The following theorem shows that the answer is no.

Theorem A.2 *The Woods criterion and the correlation ratio are not equivalent in the sense that there exists no monotonic mapping f, such that,*

$$\forall (X,Y) \in L^2, \quad \eta^2(Y|X) = f\left[W(Y|X)\right].$$

Proof. We give a counter-example. Let X and Y be independent L^2 variables with expectation 1 and variance 1. Such variables exist. As X and Y are independent, $\eta^2(Y|X) = 0$. Moreover, it is easy to check that $W(Y|X) = 1$. Let now $Z = Y + 1$. Z is independent from X and we have: $E(Z) = 2$, $\mathrm{Var}(Z) = 1$. Thus, $\eta^2(Z|X) = 0$ and $W(Z|X) = 1/2$, which completes the proof. $\square$

Annexe B

Régression linéaire avec interpolation par volume partiel

B.1 Least squares parametric fitting

At a fixed transformation T, consider the problem of minimizing the criterion :

$$\tilde{S}(T,f) = \frac{\sum_{k,\ell} w^{\downarrow}_{k\ell}\,[i_k - f(j_\ell)]^2}{n^{\downarrow}\mathrm{Var}(I^{\downarrow})}, \tag{B.1}$$

among the set of functions f that are linear combinations of specified basis functions ϕ_p. The unknown f is thus parameterized by a finite set of coefficients $\mathbf{c} = (c_1,c_2,\ldots)^t$:

$$f(j) = \sum_p c_p \phi_p(j).$$

A polynomial fitting corresponds to the special case where the basis functions are chosen as monoms, i.e., $\phi_p(j) = j^p$ if J is a real-valued image and, more generally, functions of the form $\phi_p(j) = (j^1)^{k_1}(j^2)^{k_2}\ldots(j^d)^{k_d}$ if J is a d-dimensional vectorial image with components $(j^1,j^2,\ldots,j^d)$. Also, since the denominator in equation (B.1) is independent of f, our problem amounts to solving for the coefficients that minimize the numerator:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \sum_{k,\ell} w^{\downarrow}_{k\ell} \left[i_k - \sum_p c_p \phi_p(j_\ell) \right]^2,$$

which is nothing but a weighted least squares linear regression problem. Since the criterion is quadratic in $\mathbf{c}$, a necessary and sufficient condition for $\mathbf{c}$ to be a solution is $\nabla_{\mathbf{c}}\tilde{S} = 0$, yielding:

$$\forall q, \quad \sum_{k,\ell} w^{\downarrow}_{k\ell}\phi_q(j_\ell) \left[\sum_p c_p\phi_p(j_\ell) - i_k \right] = 0, \tag{B.2}$$

or equivalently, by rewriting this expression in algebraic form,

$$\mathbf{Ac} = \mathbf{b}, \tag{B.3}$$

where $\mathbf{A}$ and $\mathbf{b}$ are respectively the square matrix and the column vector defined by:

$$\begin{cases} a_{pq} &= \sum_{k,\ell} w^{\downarrow}_{k\ell}\phi_p(j_\ell)\phi_q(j_\ell) \\ b_p &= \sum_{k,\ell} w^{\downarrow}_{k\ell} i_k \phi_p(j_\ell) \end{cases}$$

Provided that $\mathbf{A}$ is not singular, the solution is unique and it is given by $\hat{\mathbf{c}} = \mathbf{A}^{-1}\mathbf{b}$. Owing to the fact that $\mathbf{A}$ is symmetric, an efficient method to compute $\mathbf{A}^{-1}$ which has proven to be numerically stable is the singular value decomposition (SVD). In case $\mathbf{A}$ is singular (unlikely to occur in practice), there is no longer a unique solution but the SVD will still provide one since it actually computes the pseudo-inverse $\mathbf{A}^{\dagger}$ so that the equality $\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{I}$ holds in every circumstance.

B.2 Robust parametric fitting

As an extension of the previous least squares regression, consider now the more general regression problem defined by the minimization of the cost function :

$$\tilde{S}(T,f) = \frac{S_0^2}{Kn^{\downarrow}\mathrm{Var}(I^{\downarrow})} \sum_{k,\ell} w^{\downarrow}_{k\ell}\, \rho\left(\frac{i_k - f(j_\ell)}{S_0}\right), \tag{B.4}$$

where ρ is the objective function associated with a given M-estimator, while S_0 and K are known constants. Without loss of generality, we may assume that $S_0 = 1$, which amounts to the variable change $\rho(x) \leftarrow \rho(x/S_0)$. Here again, the denominator is seen to play no role in the minimization with respect to f (at fixed T), so that our problem boils down to finding:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \sum_{k,\ell} w^{\downarrow}_{k\ell}\rho\left(i_k - \sum_p c_p\phi_p(j_\ell)\right),$$

whenever f is parameterized in the same fashion as in section B.1. Unless ρ is chosen as $\rho(x) = x^2$, as was previously the case, the criterion is generally not quadratic with respect to $\mathbf{c}$. Therefore, determining an analytical solution is hopeless[1]. In practice, one seeks a solution that fulfills the necessary condition $\nabla_{\mathbf{c}}\tilde{S} = 0$, yielding:

$$\forall q, \quad \sum_{k,\ell} w^{\downarrow}_{k\ell} W(r_{k\ell})\phi_q(j_\ell)\left[\sum_p c_p\phi_p(j_\ell) - i_k\right] = 0, \tag{B.5}$$

where $r_{k\ell}$ are the residuals defined by $r_{k\ell} = \sum_p c_p\phi_p(j_\ell) - i_k$, and W is the weighting function associated with ρ, i.e., by definition:

$$W(x) = \frac{\rho'(x)}{x}$$

The reason for writing equation (B.5) in this manner is to give a flavour of the commonly adopted resolution strategy. Ignoring the fact that the residuals $r_{k\ell}$ depend on $\mathbf{c}$, equation (B.5)

1. Also notice that proofs of uniqueness are subject to the additional assumption that ρ is strictly convex, a requirement that is not mqet, e.g., by the Geman-McClure function.

becomes equivalent to equation (B.2) up to a simple reweighting: the original weights $w^{\downarrow}_{k\ell}$ are replaced with $w^{\downarrow}_{k\ell}W(r_{k\ell})$, which incorporate the non-quadratic nature of the ρ-function. Then, solving equation (B.5) amounts to solving a linear system, which still may be achieved by means of the SVD method. However, since the residuals actually depend on $\mathbf{c}$, a single iteration is generally not sufficient to provide an acceptable solution. The process has to be iterated until convergence, updating the weights at each step in terms of the residuals corresponding to the current estimate of $\mathbf{c}$. As shown in (Rousseeuw et Leroy, 1987), this iterative weighted least squares procedure is guaranteed to converge to a solution of equation (B.5). The initial estimate for $\mathbf{c}$ is practically chosen as the result of the standard least squares regression.

Estimation of S_0

To ensure that this robust regression does a good job, it is important that the constant S_0 be tuned adequately. A too large value would let the fitting function be spoiled by outliers, whereas a too small value would act as rejecting too many inliers, and again making the resulting fit unreliable. Ideally, S_0 should be a robust scale estimate (Rousseeuw et Croux, 1992): roughly speaking, an estimate of the typical error of fit. Thus, finding an optimal value for S_0 would require the fitting function to be known; but the regression method needs S_0 to work.

The common trade-off is to use the standard least squares solution as a rough initial guess of the fitting function and then compute S_0 from the corresponding residuals using some robust scale estimator. A simple choice is the median of absolute deviations,

$$\mathrm{MAD} = 1.4826\,\mathrm{median}\,\{r_1,r_2,\ldots,r_n\}\,,$$

where the normalization factor 1.4826 ensures that the MAD be an unbiased estimate of the residuals' standard deviation when they are normally distributed. Here, our problem is in fact slightly more complex since the residuals are weighted. We use a straightforward generalization of the MAD estimator based on a weighted median (Brownrigg, 1984):

$$S_0 = 1.4826\,\mathrm{wmedian}\left\{r_{k\ell};w^{\downarrow}_{k\ell}\right\}.$$

Bibliographie

ALPERT, N. M., BERDICHEVSKY, D., LEVIN, Z., MORRIS, E. D., et FISCHMAN, A. J. (1996). « Improved Methods for Image Registration ». *NeuroImage*, 3:10–18.

ALPERT, N. M., BRADSHAW, J. F., KENNEDY, D., et CORREIA, J. A. (1990). « The Principal Axes Transformation - A Method for Image Registration ». *Journal of Nuclear Medicine*, 31(10):1717–1722.

AMIT, Y. et KONG, A. (1996). « Graphical templates for model registration ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:225–236.

ARDEKANI, B. A., BRAUN, M., HUTTON, B. F., KANNO, I., et ILIDA, H. (1995). « A Fully Automatic Multimodality Image Registration Algorithm ». *Journal of Computer Assisted Tomography*, 19(4):615–623.

ASHBURNER, J. et FRISTON, K. J. (1997). « Multimodal Image Coregistration and Partitioning - a Unified Framework ». *NeuroImage*, 6(3):209–217.

AYACHE, N. (1989). *Vision stéréoscopique et perception multisensorielle*. InterEditions, Paris.

AYACHE, N. (1998). « L'analyse des images médicales, état de l'art et perspectives ». *Annales de l'Institut Pasteur*, 9:13–21. Also INRIA Research Report N. 3364.

BAJCSY, R. et KOVAČIČ, S. (1989). « Multi-resolution elastic matching ». *Computer Vision, Graphics, and Image Processing*, 46(1):1–21.

BANSAL, R., STAIB, L. H., CHEN, Z., RANGARAJAN, A., KNISELY, J., NATH, R., et DUNCAN, J. S. (1998). « A Novel Approach for the Registration of 2D Portal and 3D CT Images for Treatment Setup Verification in Radiotherapy ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 1075–1086, Cambridge, MA, USA. Springer Verlag.

BARBER, D. C. (1992). « Registration of Low Resolution Medical Images ». *Physics in Medecine and Biology*, 37(7):1485–1498.

BARNEA, D. et SILVERMAN, H. (1972). « A Class of Algorithms for Fast Digital Image Registration ». *IEEE Trans. Comput.*, 21(2):179–186.

BARRON, J. L., FLEET, D. J., et BEAUCHEMIN, S. S. (1994). « Performance of Optical Flow Techniques ». *International Journal of Computer Vision*, 12(1):43–77.

BESAG, J. (1986). « On the statistical analysis of dirty pictures ». *Journal of the Royal Statistical Society*, 48:259–302.

BESL, P. et MCKAY, N. (1992). « A method for registration of 3-D shapes ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256.

BLAHUT, R. E. (1987). *Principles and Practice of Information Theory*. Addison-Wesley Pub. Comp.

BLAKE, A. et ZISSERMAN, A. (1987). *Visual Reconstruction*. MIT Press.

BORGEFORS, G. (1988). « Hierarchical chamfer matching: a parametric edge matching algorithm ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:849–865.

BRO-NIELSEN, M. (1997). « Rigid Registration of CT, MR and Cryosection Images Using a GLCM Framework ». Dans *First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery (CVRMed-MRCAS'97)*, Lecture Notes in Computer Science, pages 171–180. Springer Verlag.

BRO-NIELSEN, M. et GRAMKOW, C. (1996). « Fast Fluid Registration of Medical Images ». Dans HÖHNE, K. H. et KIKINIS, R., éditeurs, *4th International Conference on Visualisation in Biomedical Computing (VBC'96)*, volume 1131 de *Lecture Notes in Computer Science*, pages 267–276, Hamburg, Germany. Springer Verlag.

BROIT, C. (1981). « *Optimal registration of deformed images* ». Thèse de Doctorat, University of Pennsylvania, Philadelphia.

BROWN, L. G. (1992). « A Survey of Image Registration Techniques ». *ACM Computing Surveys*, 24(4):325–376.

BROWNRIGG, D. (1984). « The Weighted Median Filter ». *Communications of the Association of Computing Machinery*, 27(8):807–818.

BUZUG, T. M. et WEESE, J. (1998). « Voxel-Based Similarity Measures for Medical Image Registration in Radiological Diagnosis and Image Guided Surgery ». *Journal of Computing and Information Technology*, pages 165–179.

CACHIER, P. et PENNEC, X. (2000). « 3D Non-Rigid Registration by Gradient Descent on a Gaussian Weighted Similarity Measure Using Convolutions ». Dans *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'00)*, pages 182–189, Hilton Head, USA. IEEE.

CACHIER, P., PENNEC, X., et AYACHE, N. (1999). « Fast Non Rigid Matching by Gradient Descent: Study and Improvements of the "Demons" Algorithm ». Rapport technique 3706, INRIA.

CACHIER, P. et REY, D. (2000). « Symmetrization of the Non-Rigid Registration Problem Using Inversion-Invariant Energies: Application to Multiple Sclerosis ». Dans DELP, S., DIGIOIA, A., et JARAMAZ, B., éditeurs, *Third International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'00)*, volume 1935 de *Lecture Notes in Computer Science*, pages 472–481, Pittsburgh. Springer Verlag.

CELEUX, G. et GOVAERT, G. (1991). « Comparison of the mixture and the classification maximum likelihood in cluster analysis ». Rapport technique 1517, INRIA.

CHRISTENSEN, G. E. (1994). « *Deformable shape models for anatomy* ». Thèse de Doctorat, Washington University.

Christensen, G. E., Rabbitt, R. D., et Miller, M. I. (1996). « Deformable Templates Using Large Deformation Kinematics ». *IEEE Transactions on Medical Imaging*, 5(10):1435–1447.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., et Evans, A. C. (1997). « BrainWeb: Online Interface to a 3D MRI Simulated Brain Database ». *NeuroImage, Proceedings of the Third International Conference on Functional Mapping of the Human Brain (HBM'97, Copenhagen, Danmark, May 19–23 1997)*, 5(4):S425.

Cohen, L. D. (1996). « Auxiliary variables and two-step iterative algorithms in computer vision problems ». *Journal of Mathematical Imaging and Vision*, 6(1):59–83.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., et Marchal, G. (1995a). « Automated multi-modality image registration based on information theory ». Dans Bizais, Y. et Barillot, C., éditeurs, *Proc. XIVth international conference on information processing in medical imaging (IPMI'95)*, pages 263–274. Kluwer Academic Publishers.

Collignon, A., Vandermeulen, D., Suetens, P., et Marchal, G. (1995b). « 3D multi-modality medical image registration using feature space clustering ». Dans Ayache, N., éditeur, *First Conference on Computer Vision, Virtual Reality and Robotics in Medicine (CVRMed'95)*, volume 905 de *Lecture Notes in Computer Science*, pages 195–204. Springer Verlag.

Collins, D. L. et Evans, A. C. (1997). « ANIMAL: Validation and Applications of Nonlinear Registration-Based Segmentation ». *International Journal of Pattern Recognition and Artificial Intelligence*, 11(8):1271–1294.

Collins, D. L., Neelin, P., Peters, T. M., et Evans, A. C. (1994). « Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space ». *Journal of Computer Assisted Tomography*, 18:192–205.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., et Evans, A. C. (1998). « Design and Construction of a Realistic Digital Brain Phantom ». *IEEE Transactions on Medical Imaging*, 17(3):463–468.

Costa, W. L. S., Haynor, D. R., Haralick, R. M., Lwellen, T. K., et Graham, M. M. (1993). « A Maximum-Likelihood Approach to PET Emission/Attenuation Image Registration ». Dans *IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 2, pages 1139–1143, San Francisco. IEEE.

Couvreur, C. (1996). « The EM Algorithm: A Guided Tour ». Dans *Proc. 2d IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing*, Prague, Czech Republik.

Cover, T. M. et Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons, inc.

Davis, M. H., Khotanzad, A., Flamig, D. P., et Harms, S. E. (1997). « A Physics-Based Coordinate Transformation for 3-D Image Matching ». *IEEE Transactions on Medical Imaging*, 16(3):317–328.

Dawant, B. M., Hartmann, S. L., et Gadamsetty, S. (1999). « Brain Atlas Deformation in the Presence of Large Space-occupying Tumors ». Dans *Second International Conference on*

Medical Image Computing And Computer-Assisted Intervention (MICCAI'99), volume 1679 de *Lecture Notes in Computer Science*, pages 589–596. Springer Verlag.

DAWANT, B. M., THIRION, J.-P., MAES, F., VANDERMEULEN, D., et DEMAEREL, P. (1998). « Automatic 3D Segmentation of Internal Structures of the Head in MR Images Using a Combination of Similarity and Free Form Transformations ». Dans HANSON, K. M., éditeur, *Medical Imaging 1998: Image Processing (MI'98)*, volume 3338 de *SPIE Proceedings*, pages 545–554, San Diego. SPIE Press.

DECLERCK, J. (1997). « *Étude de la dynamique cardiaque par analyse d'images tridimensionnelles* ». Thèse de Doctorat, Université de Nice-Sophia Antipolis.

DEMPSTER, A. P., LAIRD, N. M., et RUBIN, D. B. (1977). « Maximum likelihood from incomplete data via the EM algorithm ». *Journal of the Royal Statistical Society*, 39:1–38.

DERICHE, R. (1992). « Recursively Implementing The Gaussian and Its Derivatives ». Dans *Proc. Second International Conference On Image Processing*, pages 263–267, Singapore.

DUCHON, J. (1977). « Spline minimizing rotation-invariant semi-norms in Sobolev spaces ». Dans *Constructive Theory of Functions of Several Variables*, Lecture Notes in Mathematics, pages 85–100, Berlin-Heidelberg. Springer Verlag.

DUDA, R. O. et HART, P. E. (1973). *Pattern classification and scene analysis.* Jonh Wiley & Sons.

DUNCAN, J. et AYACHE, N. (2000). « Medical Image Analysis: Progress over two decades and the challenges ahead ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85–106.

DUTT, V. et GRENLEAF, J. F. (1996). « Adaptive Speckle Reduction Filter for Log-Compressed B-Scan Images ». *IEEE Transactions on Medical Imaging*, 15(6):802–813.

EGGERT, D., LORUSSO, A., et FISHER, R. (1997). « Estimating 3D Rigid Body transformations: A Comparison of Four Major Algorithms ». *Machine Vision Applications, Special Issue on Performance Characteristics of Vision Algorithms*, 9(5/6):272–290.

ERBE, H., KRIETE, A., JÖDICKE, A., DEINSBERGER, W., et BÖKER, D.-K. (1996). « 3D-Ultrasonography and Image Matching For Detection of Brain Shift During Intracranial Surgery ». Dans *Applications of Computer Vision in Medical Image Processing*, volume 1124 de *Excerpta Medica – Int. Congress Series*, pages 225–230. Elsevier.

FAUGERAS, O. (1993). *Three-Dimensional Computer Vision.* The MIT Press.

FELDMAR, J. (1994). « *Recalage rigide, non rigide et projectif d'images médicales tridimensionnelles* ». Thèse de Doctorat, École Polytechnique.

FELDMAR, J., DECLERCK, J., MALANDAIN, G., et AYACHE, N. (1997). « Extension of the ICP Algorithm to Non-Rigid Intensity-Based Registration of 3D Volumes ». *Computer Vision and Image Understanding*, 66(2):193–206.

FERNÁNDEZ-VIDAL, S., BARDINET, E., MALANDAIN, G., DAMAS, S., et de la BLANCA CAPILLA, N. (2000). « Object Representation and Comparison Inferred from Its Medial Axis ». Dans *Inter-*

national Conference on Pattern Recognition (ICPR'00), volume 1, pages 712–715, Barcelona, Spain.

FOLEY, T. et NIELSON, G. (1989). « *Mathematical Methods in Computer Aided Geometric Design* », Chapitre Knot selection for parametric spline interpolation, pages 261–271. Academic Press.

FRISTON, K., ASHBURNER, J., POLINE, J.-B., FRITH, C., HEATHER, J., et FRACKOWIAK, R. (1995). « Spatial registration and normalisation of images ». *Human Brain Mapping*, 3(3):165–189.

GAENS, T., MAES, F., VANDERMEULEN, D., et SUETENS, P. (1998). « Non-rigid Multimodal Image Registration Using Mutual Information ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 1099–1106. Springer Verlag.

GEE, J., REIVICJ, M., et BAJCSY, R. (1993). « Elasticity deforming 3d atlas to match anatomical brain images ». *Journal of Computer Assisted Tomography*, 17(2):225–236.

GERIG, G., KÜBLER, O., KIKINIS, R., et JOLESZ, F. A. (1992). « Nonlinear Anisotropic Filtering of MRI Data ». *IEEE Transactions on Medical Imaging*, 11(2):221–231.

GIROSI, F., JONES, M., et POGGIO, T. (1995). « Regularization Theory and Neural Networks Architectures ». *Neural Computation*, 7:219–269.

GUILLEMAUD, R. et BRADY, M. (1997). « Estimating the Bias Field of MR Images ». *IEEE Transactions on Medical Imaging*, 16(3):238–251.

GUIMOND, A. (1999). « *Construction et utilisation de modèles anatomiques* ». Thèse de Doctorat, Université de Montréal / Université de Nice - Sophia Antipolis.

GUIMOND, A., ROCHE, A., AYACHE, N., et MEUNIER, J. (2001). « Multimodal Brain Warping Using the Demons Algorithm and Adaptative Intensity Corrections ». *IEEE Transactions on Medical Imaging*, 20(1).

HATA, N., SUZUKI, M., DOHI, T., ISEKI, H., TAKAKURA, K., et HASHIMOTO, D. (1994). « Registration of Ultrasound echography for Intraoperative Use: A Newly Developed Multiproperty Method ». Dans *Visualization in Biomedical Computing*, volume 2359 de *SPIE Proceedings*, pages 252–259, Rochester, MN, USA. SPIE Press.

HAYTON, P. M., BRADY, M., SMITH, S. M., et MOORE, N. (1999). « A non-rigid registration algorithm for dynamic breast MR images ». *Artificial Intelligence*, 114(1-2):125–156.

HELLIER, P. (2000). « *Recalage non rigide en imagerie médicale : méthodes et validation* ». Thèse de Doctorat, Université Rennes I.

HELLIER, P. et BARILLOT, C. (2000). « Multimodal non-rigid warping for correction of distortions in functional MRI ». Dans *Third International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'00)*, volume 1935 de *Lecture Notes in Computer Science*, pages 512–520, Pittsburgh. Springer Verlag.

HELLIER, P., BARILLOT, C., MÉMIN, E., et PÉREZ, P. (1999). « Medical Image Registration with Roust Multigrid Techniques ». Dans *Second International Conference on Medical Image*

Computing And Computer-Assisted Intervention (MICCAI'99), volume 1679 de *Lecture Notes in Computer Science*, pages 680–687, Cambridge, UK. Springer Verlag.

HILL, D., STUDHOLME, C., et HAWKES, D. (1994). « Voxel similarity measures for automated image registration ». Dans *Visualization in Biomedical Computing*, volume 2359 de *SPIE Proceedings*, pages 205–216. SPIE Press.

HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

HORN, B. et SCHUNCK, B. (1981). « Determining Optical Flow ». *Artificial Intelligence*, 17:185–203.

IONESCU, G., LAVALLÉE, S., et DEMONGEOT, J. (1999). « Automated Registration of Ultrasound with CT Images: Application to Computer Assisted Prostate Radiotherapy and Orthopedics ». Dans *Second International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'99)*, volume 1679 de *Lecture Notes in Computer Science*, pages 768–777, Cambridge, UK. Springer Verlag.

JAIN, A. K. (1981). « Image Data Compression: A Review ». *Proceedings of the IEEE*, 69:349–389.

JENKINSON, M. et SMITH, S. (1999). « An investigation of the robustness of registration methods ». Dans PERNUS, F., KOVAČIČ, S., STIEHL, H., et VIERGEVER, M., éditeurs, *International Workshop on Biomedical Image Registration (WBIR'99)*, pages 200–210.

JENKINSON, M. et SMITH, S. (2000). « Global Optimisation for Robust Affine Registration ». *NeuroImage*. Submitted.

KAPUR, T. (1999). « *Model based three dimensional Medical Image Segmentation* ». Thèse de Doctorat, Massachusetts Institute of Technology.

KRISSIAN, K., MALANDAIN, G., et AYACHE, N. (1996). « Directional Anisotropic Diffusion applied to segmentation of vessels in 3D images ». Rapport technique 3064, INRIA.

KRUGGEL, F. et BARTENSTEIN, P. (1995). « Automatical registration of brain volume datasets ». Dans BIZAIS, Y. et BARILLOT, C., éditeurs, *Proc. XIVth international conference on information processing in medical imaging (IPMI'95)*, pages 389–390. Kluwer Academic Publishers.

KWAN, R. K.-S., EVANS, A. C., et PIKE, G. B. (1996). « An Extensible MRI Simulator for Post-Processing Evaluation ». Dans HÖHNE, K. H. et KIKINIS, R., éditeurs, *4th International Conference on Visualisation in Biomedical Computing (VBC'96)*, volume 1131 de *Lecture Notes in Computer Science*, pages 135–140, Hamburg, Germany. Springer Verlag.

KYBIC, J., THÉVENAZ, P., et UNSER, M. (1999). « Multiresolution spline warping for EPI registration ». Dans *Wavelet Applications in Signal and Image Processing*, volume 3813, pages 571–579, Denver. SPIE Press.

LAU, Y., BRAUN, M., et HUTTON, B. (1999). « Non-rigid 3d image registration using regionally constrained matching and the correlation ratio ». Dans PERNUS, F., KOVAČIČ, S., STIEHL, H., et VIERGEVER, M., éditeurs, *International Workshop on Biomedical Image Registration (WBIR'99)*, pages 137–148.

LAVALLÉE, S. (1995). Registration for Computer Integrated Surgery: methodology, state of the art.

Dans Taylor, R., Lavallée, S., Burdea, G., et Moesges, R., éditeurs, *Computer Integrated Surgery*, pages 77–97. MIT Press.

Lemaréchal, C. (1989). *Méthodes numériques d'optimisation*. INRIA.

Lester, H. (1999). « *Non-linear registration of medical images* ». Thèse de Doctorat, University of London, UK.

Leventon, M. E. et Grimson, W. E. L. (1998). « Multi-modal Volume Registration Using Joint Intensity Distributions ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 1057–1066, Cambridge, MA, USA. Springer Verlag.

Li, S. Z. (1995). *Markov random field modeling in computer vision*. Springer, Tokyo, New-York, Berlin.

Loupas, T., McDicken, W., et Allan, P. (1989). « An Adaptive Weighted Median Filter for Speckle Suppression in Medical Ultrasonic Images ». *IEEE Transactions on Circuits Systems*, 36(1):802–813.

Maes, F. (1998). « *Segmentation and Registration of Multimodal Medical Images: from Theory, Implementation and Validation to a Useful Tool in Clinical Practice* ». Thèse de Doctorat, Katholieke Univeristeit Leuven, Belgium.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., et Suetens, P. (1997). « Multimodality Image Registration by Maximization of Mutual Information ». *IEEE Transactions on Medical Imaging*, 16(2):187–198.

Maes, F., Vandermeulen, D., et Suetens, P. (1999). « Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information ». *Medical Image Analysis*, 3(4):373–386.

Maintz, J. B. A., Meijering, E. H. W., et Viergever, M. A. (1998). « General Multimodal Elastic Registration Based on Mutual Information ». Dans Hanson, K. M., éditeur, *Medical Imaging 1998: Image Processing (MI'98)*, volume 3338 de *SPIE Proceedings*, pages 144–154, Bellingham, WA, USA. SPIE Press.

Maintz, J. B. A. et Viergever, M. A. (1998). « A survey of medical image registration ». *Medical Image Analysis*, 2(1):1–36.

Malandain, G., Fernández-Vidal, S., et Rocchisani, J. (1994). « Improving registration of 3-D images using a mechanical based method ». Dans *Proc. ECCV'94*, Lecture Notes in Computer Science, pages 131–136. Springer Verlag.

Mangin, J., Frouin, V., Bloch, I., Régis, J., et López-Krahe, J. (1995). « From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations ». *Journal of Mathematical Imaging and Vision*, 5(4):297–318.

Mangin, J.-F. (1995). « *Mise en correspondance d'images médicales 3D multi-modalités multi-individus pour la corrélation anatomo-fonctionnelle* ». Thèse de Doctorat, École Nationale Supérieure des Télécommunications, Paris.

McGillem, C. D. et Svedlow, M. (1976). « Image Registration Error Variance as a Measure of Overlay Quality ». *IEEE Transactions on Geoscience Electronics*, 4(1):44–49.

Meer, P., Stewart, C. V., et Tyler, D. E. (2000). « Robust Computer Vision: An Interdisciplinary Challenge ». *Computer Vision and Image Understanding*, 78(1):1–7.

Meijering, E. H. W., Niessen, W. J., Pluim, J. P. W., et Viergever, M. A. (1999). « Quantitative Comparison of Sinc-Approximating Kernels for Medical Image Interpolation ». Dans *Second International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'99)*, volume 1679 de *Lecture Notes in Computer Science*, pages 210–217, Cambridge, UK. Springer Verlag.

Mémin, E. et Pérez, P. (1998). « Optical flow estimation and object-based segmentation with robust techniques ». *IEEE Transactions on Image Processing*, 7(5):703–719.

Meyer, C. R., Boes, J. L., Kim, B., et Bland, P. H. (1998). « Evaluation of Control Point Selection in Automatic Mutual Information Driven, 3D Warping ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 944–951, Cambridge, MA, USA. Springer Verlag.

Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., et Wahl, R. L. (1997). « Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate warped geometric deformations ». *Medical Image Analysis*, 1(3):195–206.

Montagnat, J. (1999). « *Modèles déformables pour la segmentation et la modélisation d'images médicales 3D et 4D* ». Thèse de Doctorat, Université de Nice-Sophia Antipolis.

Moreau, N. (1995). *Techniques de compression des signaux.* Masson, Paris Milan Barcelone.

Mort, M. S. et Srinath, M. D. (1988). « Maximum Likelihood Image Registration with Subpixel Accuracy ». Dans *Applications of digital Image Processing XI*, volume 974 de *SPIE Proceedings*, pages 38–45, San Diego. SPIE Press.

Mulet-Parada, M. (2000). « *Intensity independent feature extraction and tracking in echocardiographic sequences* ». Thèse de Doctorat, University of Oxford, UK.

Neal, R. M. et Hinton, G. E. (1998). « A view of the EM algorithm that justifies incremental, sparse, and other variants ». Dans Jordan, M., éditeur, *Learning in Graphical Models*, pages 355–368. Kluwer Academic Publishers.

Nikou, C. (1999). « *Contributions au recalage d'images médicales multimodales: approches par fonctions de similarité robustes et modèles déformables sous contraintes statistiques* ». Thèse de Doctorat, Université Louis Pasteur - Strasbourg I.

Nikou, C., Heitz, F., et Armspach, J. (1999). « Robust voxel similarity metrics for the registration of dissimilar single and multimodal images ». *Pattern Recognition*, 32(8):1351–1368.

Nikou, C., Heitz, F., Armspach, J.-P., et Namer, I.-J. (1998). « Single and multimodal subvoxel registration of dissimilar medical images using robust similarity measures ». Dans *Medical*

Imaging 1998: Image Processing (MI'98), volume 3338 de *SPIE Proceedings*, pages 167–178. SPIE Press.

Ourselin, S., Roche, A., Prima, S., et Ayache, N. (2000a). « Block Matching: a General Framework to Improve Robustness of Rigid Registration of Medical Images ». Dans Delp, S., DiGioia, A., et Jaramaz, B., éditeurs, *Third International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI'00)*, volume 1935 de *Lecture Notes in Computer Science*, pages 557–566, Pittsburgh, USA. Springer Verlag.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., et Ayache, N. (2000b). « Reconstructing a 3D Structure from Serial Histological Sections ». *Image and Vision Computing*, 19(1-2):25–31.

Pagoulatos, N., Edwards, W., Haynor, D., et Kim, Y. (1999). « Interactive 3-D Registration of Ultrasound and Magnetic Resonance Images Based on a Magnetic Position Sensor ». *IEEE Transactions on Information Technology in Biomedicine*, 3(4):278–288.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., third edition.

Pelizzari, C., Chen, G., Spelbring, D., Weichselbaum, R., et Chen, C. (1989). « Accurate three-dimensional registration of CT,PET and/or MR images of the brain ». *Journal of Computer Assisted Tomography*, 13(20).

Pennec, X. (1996). « *L'incertitude dans les problèmes de reconnaissance et de recalage. Application en imagerie médicale et biologie moléculaire* ». Thèse de Doctorat, École Polytechnique.

Pennec, X. (1998). « Computing the mean of geometric features - Application to the mean rotation ». Rapport technique 3371, INRIA.

Pennec, X., Ayache, N., et Thirion, J. (2000). Chap. 31: Landmark-based registration using features identified through differential geometry. Dans Bankman, I., éditeur, *Handbook of Medical Imaging*, pages 499–513. Academic Press.

Pennec, X., Cachier, P., et Ayache, N. (1999). « Understanding the "Demon's Algorithm": 3D Non-Rigid registration by Gradient Descent ». Dans *Second International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'99)*, volume 1679 de *Lecture Notes in Computer Science*, pages 597–605. Springer Verlag.

Pennec, X., Guttmann, C., et Thirion, J. (1998). « Feature-based Registration of Medical Images: Estimation and Validation of the Pose Accuracy ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, numéro 1496 dans Lecture Notes in Computer Science, pages 1107–1114, Cambridge, MA, USA. Springer Verlag.

Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L. G., et Hawkes, D. J. (1998). « A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration ». Dans *First International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 1153–1161, Cambridge, MA, USA. Springer Verlag.

Perona, P. et Malik, J. (1990). « Scale-Space and edge detection using anisotropic diffusion ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

Pluim, J. P. W., Maintz, J. B. A., et Viergever, M. A. (1999). « Mutual information matching and interpolation artefacts ». Dans Hanson, K. M., éditeur, *Medical Imaging: Image Processing*, volume 3661 de *SPIE Proceedings*, Bellingham, WA. SPIE Press.

Pluim, J. P. W., Maintz, J. B. A., et Viergever, M. A. (2000a). « Image registration by maximization of combined mutual information and gradient information ». *IEEE Transactions on Medical Imaging*, 19(8):809–814.

Pluim, J. P. W., Maintz, J. B. A., et Viergever, M. A. (2000b). « Interpolation artefacts in mutual information based image registration ». *Computer Vision and Image Understanding*, 77(2):211–232.

Press, W. H., Flannery, B. P., Teukolsky, S. A., et Vetterling, W. T. (1992). *Numerical Recipes in C*. Cambrige University Press, 2nd edition.

Prima, S. (2000). « Maximum likelihood estimation of the bias field in MR images: investigating different models for the imaging process ». Rapport technique, INRIA. In press.

Prima, S., Ourselin, S., et Ayache, N. (2000). « Computation of the Mid-Sagittal Plane in 3D Images of the Brain ». Dans Vernon, D., éditeur, *Sixth European Conference on Computer Vision, ECCV'2000*, volume 1842-3 de *Lecture Notes in Computer Science*, pages 685–701, Dublin, Ireland. Springer Verlag.

Rajapakse, J. C., Kruggel, F., et von Cramon, D. Y. (1998). « Segmentation of MR Images with Intensity Inhomogeneities ». *Image and Vision Computing*, 16(10):165–180.

Rangarajan, A., Chui, H., et Bookstein, F. (1997). « The Softassign Procrustes Matching Algorithm ». Dans *Proc. XVth international conference on information processing in medical imaging (IPMI'97)*, volume 3 de *Lecture Notes in Computer Science*, pages 29–42. Springer Verlag.

Rangarajan, A., Chui, H., et Duncan, J. (1999). « Rigid Point Feature Registration Using Mutual Information ». *Medical Image Analysis*, 3(4):425–440.

Régis, J., Mangin, J.-F., Frouin, V., Sastre, F., Peragut, J. C., et Samson, Y. (1995). « Generic model for the localization of the cerebral cortex and preoperative multimodal integration in epilepsy surgery ». *Stereotactic and Functional Neurosurgery*, 65:72–80.

Rey, D., Subsol, G., Delingette, H., et Ayache, N. (1999). « Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images: Application to Multiple Sclerosis ». Dans *Proc. XVIth international conference on information processing in medical imaging (IPMI'99)*, pages 154–167, Visegrád, Hungary. Springer Verlag.

Roche, A., Guimond, A., Ayache, N., et Meunier, J. (2000a). « Multimodal Elastic Matching of Brain Images ». Dans *Computer Vision - ECCV 2000*, volume 1843 de *Lecture Notes in Computer Science*, pages 511–527, Dublin, Eire. Springer Verlag.

Roche, A., Malandain, G., et Ayache, N. (2000b). « Unifying Maximum Likelihood Approaches in Medical Image Registration ». *International Journal of Imaging Systems and Technology*, 11:71–80.

Roche, A., Malandain, G., Ayache, N., et Pennec, X. (1998a). « Multimodal Image Registration by Maximization of the Correlation Ratio ». Rapport technique 3378, INRIA.

Roche, A., Malandain, G., Ayache, N., et Prima, S. (1999). « Towards a Better Comprehension of Similarity Measures used in Medical Image Registration ». Dans *Second International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI'99)*, volume 1679 de *Lecture Notes in Computer Science*, pages 555–566, Cambridge, UK. Springer Verlag.

Roche, A., Malandain, G., Pennec, X., et Ayache, N. (1998b). « The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration ». Dans *First International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI'98)*, volume 1496 de *Lecture Notes in Computer Science*, pages 1115–1124, Cambridge, MA, USA. Springer Verlag.

Roche, A., Pennec, X., Malandain, G., Ayache, N., et Ourselin, S. (2000c). « Generalized Correlation Ratio for Rigid Registration of 3D Ultrasound with MR Images ». Rapport technique 3980, INRIA.

Roche, A., Pennec, X., Rudolph, M., Auer, D. P., Malandain, G., Ourselin, S., Auer, L. M., et Ayache, N. (2000d). « Generalized Correlation Ratio for Rigid Registration of 3D Ultrasound with MR Images ». Dans Delp, S., DiGioia, A., et Jaramaz, B., éditeurs, *Third International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI'00)*, volume 1935 de *Lecture Notes in Computer Science*, pages 567–577, Pittsburgh, USA. Springer Verlag.

Rohling, R. N. (1998). « *3D freehand ultrasound: reconstruction and spatial compounding* ». Thèse de Doctorat, University of Cambridge, UK.

Rousseeuw, P. et Croux, C. (1992). « Explicit Scale Estimators with High Breakdown Point ». Dans Dodge, Y., éditeur, *L1-Statistical Analysis and Related Methods*, pages 77–92, Amsterdam, Netherlands.

Rousseeuw, P. et Van Driessen, K. (1999). « Computing LTS Regression for Large Data Sets ». Rapport technique, Statistics Group, University of Antwerp. submitted.

Rousseeuw, P. J. et Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley Series In Probability And Mathematical Statistics, first edition.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., et Hawkes, D. J. (1999). « Non-rigid Registration using Free-Form Deformations: Application to Breast MR Image ». *IEEE Transactions on Medical Imaging*, 18(8):712–721.

Saporta, G. (1990). *Probabilités, analyse des données et statistique*. Editions Technip, Paris.

Sarrut, D. (2000). « *Recalage multimodal et plate-forme d'imagerie médicale à accès distant* ». Thèse de Doctorat, Université Lumière Lyon 2.

SARRUT, D. et FESCHET, F. (1999). « The Partial Intensity Difference Interpolation ». Dans ARABNIA, H. R., éditeur, *International Conference on Imaging Science, Systems and Technologiy*, Las Vegas. CSREA Press.

SARRUT, D. et MIGUET, S. (1999). « Similarity Measures for Image Registration ». Dans *European Workshop on Content-Based Multimedia Indexing*, pages 263–270, Toulouse, France. IHMPT–IRIT.

STUDHOLME, C., HILL, D. L. G., et HAWKES, D. J. (1995). « Multiresolution Voxel Similarity Measures for MR-PET Registration ». Dans BIZAIS, Y. et BARILLOT, C., éditeurs, *Proc. XIVth international conference on information processing in medical imaging (IPMI'95)*, pages 287–298. Kluwer Academic Publishers.

STUDHOLME, C., HILL, D. L. G., et HAWKES, D. J. (1996). « Automated 3-D registration of MR and CT images of the head ». *Medical Image Analysis*, 1(2):163–175.

STUDHOLME, C., HILL, D. L. G., et HAWKES, D. J. (1998). « An overlap invariant entropy measure of 3D medical image alignment ». *Pattern Recognition*, 1(32):71–86.

SUBSOL, G., THIRION, J. P., et AYACHE, N. (1998). « A General Scheme for Automatically Building 3D Morphometric Anatomical Atlases: application to a Skull Atlas ». *Medical Image Analysis*, 2(1).

SVEDLOW, M., MCGILLEM, C., et ANUTA, P. (1978). « Image Registration: Similarity Measure and Processing Method Comparisons ». *AeroSys*, 14(1):141–149.

TALAIRACH, J. et TOURNOUX, P. (1993). *Referentially oriented cerebral MRI anatomy.* Georg Thieme Verlag, New-York.

TERZOPOULOS, D. (1986). « Regularization of Inverse Visual Problems Involving Discontinuities ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):413–424.

THÉVENAZ, P. et UNSER, M. (1998). « An efficient mutual information optimizer for multiresolution image registration ». Dans *International Conference on Image Processing*, volume 1, pages 833–837, Chicago. IEEE.

THIRION, J. (1995). « Fast non-rigid matching of 3-D medical images ». Dans *Medical Robotics and Computer Aided Surgery*, Lecture Notes in Computer Science, pages 47–54. Springer Verlag.

THIRION, J. P. (1996). « New Feature Points Based on Geometric Invariants for 3D Image Registration ». *International Journal of Computer Vision*, 18(2):121–137.

THIRION, J.-P. (1998). « Image matching as a diffusion process: an analogy with Maxwell's demons ». *Medical Image Analysis*, 2(3):243–260.

TITTERINGTON, D. M., SMITH, A. F. M., et MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons.

TOGA, A. W. (1999). *Brain Warping.* Academic Press.

UNSER, M. (1999). « Splines : a perfect fit for signal/image processing ». *IEEE Signal Processing Magazine*, 16(6):22–38.

UNSER, M., ALDROUBI, A., et EDEN, M. (1993a). « B-spline signal processing: Part I-theory ». *IEEE Transactions on Signal Processing*, 30(2):821–833.

UNSER, M., ALDROUBI, A., et EDEN, M. (1993b). « B-spline signal processing: Part II-efficient design and applications ». *IEEE Transactions on Signal Processing*, 30(2):834–848.

van den ELSEN, P., MAINTZ, J., et VIERGEVER, M. (1996). « Comparison of edge-based and ridge-based registration of CT and MR brain images ». *Medical Image Analysis*, 1(2):151–161.

van den ELSEN, P., POL, E., et VIERGEVER, M. (1993). « Medical Image Matching - A Review with Classification ». *IEEE Engineering in Medicine and Biology*, 12(4):26–39.

van den ELSEN, P. A., POL, E.-J. D., SUMANAWEERA, T. S., HEMLER, P. F., NAPEL, S., et ADLER, J. R. (1994). « Grey value correlation techniques for automatic matching of CT and MR brain and spine images ». Dans *Visualization in Biomededical Computing*, volume 2359 de *SPIE Proceedings*, pages 227–237. SPIE Press.

VAN LEEMPUT, K., MAES, F., VANDERMEULEN, D., et SUETENS, P. (1999). « Automated model-based bias field correction of MR images of the brain ». *IEEE Transactions on Medical Imaging*, 18(10):885–896.

VENOT, A., LEBRUCHEC, J., et ROUCAYROL, J. (1984). « A New Class of Similarity Measures for Robust Image Registration ». *Computer Vision, Graphics, and Image Processing*, 28:176–184.

VIOLA, P. (1995). « *Alignment by Maximization of Mutual Information* ». Thèse de Doctorat, Massachusetts Institute of Technology.

VIOLA, P. et WELLS, W. M. (1997). « Alignment by Maximization of Mutual Information ». *International Journal of Computer Vision*, 24(2):137–154.

WAHBA, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

WELLS, W. M., GRIMSON, W. E. L., KIKINIS, R., et JOLESZ, F. A. (1996a). « Adaptive Segmentation of MRI Data ». *IEEE Transactions on Medical Imaging*, 15(4):429–442.

WELLS, W. M., VIOLA, P., ATSUMI, H., et NAKAJIMA, S. (1996b). « Multi-modal volume registration by maximization of mutual information ». *Medical Image Analysis*, 1(1):35–51.

WEST, J. et AL. (1997). « Comparison and evaluation of retrospective intermodality brain image registration techniques ». *Journal of Computer Assisted Tomography*, 21:554–566.

WOODS, R. P., CHERRY, S. R., et MAZZIOTTA, J. C. (1992). « Rapid Automated Algorithm for Aligning and Reslicing PET Images ». *Journal of Computer Assisted Tomography*, 16(4):620–633.

WOODS, R. P., GRAFTON, S. T., HOLMES, C. J., CHERRY, S. R., et MAZZIOTTA, J. C. (1998a). « Automated image registration: I. General methods and intrasubject, intramodality validation ». *Journal of Computer Assisted Tomography*, 22:141–154.

WOODS, R. P., GRAFTON, S. T., WATSON, J. D. G., SICOTTE, N. L., et MAZZIOTTA, J. C. (1998b). « Automated image registration: II. Intersubject validation of linear and nonlinear models ». *Journal of Computer Assisted Tomography*, 22:155–165.

WOODS, R. P., MAZZIOTTA, J. C., et CHERRY, S. R. (1993). « MRI-PET Registration with Automated Algorithm ». *Journal of Computer Assisted Tomography*, 17(4):536–546.

ZHANG, Y., BRADY, M., et SMITH, S. (2000). « A hidden Markov random field model for segmentation of brain MR images ». Dans *Medical Imaging*, volume 3979 de *SPIE Proceedings*, pages 1126–1137, San Diego. SPIE Press.

ZHANG, Z. (1993). « Le problème de la mise en correspondance : l'état de l'art ». Rapport technique 2146, INRIA.

ZHANG, Z. (1994). « Iterative Point Matching for Registration of Free-Form Curves and Surfaces ». *International Journal of Computer Vision*, 13(2):119–152.

Recalage d'images médicales par inférence statistique

Alexis Roche

INRIA - projet Epidaure, B.P. 93, 06902 Sophia-Antipolis Cédex, France.

Le recalage est un problème classique en vision par ordinateur qui intervient notamment dans de nombreuses tâches d'analyse des images médicales. Le principe général d'un algorithme de recalage est d'optimiser un critère mesurant la correspondance entre deux images sur un espace prédéterminé de transformations spatiales. Le choix du critère, appelé *mesure de similarité*, conditionne de façon déterminante la précision et la robustesse de l'algorithme. Il existe aujourd'hui un dictionnaire de mesures de similarité dans lequel le concepteur d'un programme choisit, en se fiant le plus souvent à son intuition et à son expérience, celle qui est la mieux adaptée à son problème particulier. Afin de rendre plus objectif un tel choix, cette thèse propose une méthodologie pour construire des mesures de similarité en fonction de modèles probabilistes de formation d'images.

En considérant d'abord un modèle simple de liaison fonctionnelle entre les intensités des images, nous définissons une nouvelle classe de mesures de similarité liée à la notion de rapport de corrélation. Nous montrons expérimentalement que cette classe est adaptée au recalage rigide multimodal d'images acquises par résonance magnétique (IRM), scanner et scintigraphie. La méthode du rapport de corrélation est ensuite étendue au cas du recalage multimodal non-rigide au prix de quelques adaptations pratiques. Enfin, nous formulons le recalage d'images comme un problème général d'estimation par maximum de vraisemblance, ce qui nous permet de prendre en compte des modèles de dépendance inter-images plus complexes que des modèles fonctionnels. Cette approche est appliquée au recalage rigide d'images ultrasonores 3D et IRM.

Mots clés : *recalage, mesures de similarité, images médicales, modèles probabilistes, estimation, optimisation, rapport de corrélation, information mutuelle.*

Medical Image Registration through Statistical Inference

Registration is a classical problem in computer vision which is essential in many tasks of medical image analysis. The general principle of a registration algorithm is to optimize a criterion measuring the match between two images among a predetermined set of spatial transformations. Choosing the criterion (the *similarity measure*) is determinant for both the accuracy and the robustness of the algorithm. There exists currently a catalog of similarity measures in which a developer may select the one which is the most appropriate to a given problem, based on his intuition and experience. In order to make this choice more objective, we propose in this thesis to derive similarity measures from probabilistic image acquisition models.

Starting with a simple model of functional dependence between the image intensities, we define a new class of similarity measures that closely relate to the concept of correlation ratio. We demonstrate experimentally that these measures are adapted to several multimodal rigid registration problems involving images acquired by Magnetic Resonance (MRI), Computed Tomography, and Scintigraphy. The correlation ratio method is then extended to the multimodal non-rigid case using a few practical adaptations. We finally formalize image registration as a general maximum likelihood estimation problem, which enables us to take into account image dependence models that are more complex than functional models. This approach is applied to rigid registration of 3D Ultrasound images and MRI.

Keywords: *registration, similarity measures, medical images, probabilistic models, estimation, optimization, correlation ratio, mutual information.*