



N° d'ordre NNT : 2017LYSEI093

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

L'Institut National des Sciences Appliquées de Lyon

École Doctorale : Électronique, Électrotechnique et Automatique (EDA160)

Spécialité du doctorat : Traitement de l'Image et du Signal

Soutenue publiquement le 23/10/2017, par :

Rémi AGIER

**Recalage de groupes d'images médicales
3D par extraction de points d'intérêt**

Devant le jury composé de :

Jocelyne TROCCAZ	Directrice de Recherche CNRS	Université de Grenoble	Présidente
Adrien BARTOLI	Professeur des Universités	Université Clermont Auvergne	Rapporteur
Xavier PENNEC	Directeur de Recherche INRIA	Université Sophia Antipolis	Rapporteur
Fabrice HEITZ	Professeur des Universités	Université de Strasbourg	Examineur
Sébastien VALETTE	Chargé de recherche CNRS	Université de Lyon	Co-encadrant
Rémy PROST	Professeur des Universités	Université de Lyon	Directeur de thèse

Département FEDORA – INSA Lyon – Écoles doctorales – Quinquennal 2016–2020

SIGLE	ÉCOLE DOCTORALE	NOM ET COORD. RESPONSABLE
CHIMIE	CHIMIE DE LYON www.edchimie-lyon.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3 ^e étage secretariat@edchimie-lyon.fr INSA : R. GOURDON	M. Stéphane DANIELE Institut de recherches sur la catalyse et l'environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 Avenue Albert EINSTEIN 69 626 Villeurbanne CEDEX directeur@edchimie-lyon.fr
E.E.A	ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE edeea.ec-lyon.fr Sec. : M.C. HAVGOUDOUKIAN ecole-doctorale.eea@ec-lyon.fr	M. Gérard SCORLETTI École Centrale de Lyon 36 Avenue Guy DE COLLONGUE 69 134 Écully ☎ 04.72.18.60.97 📠 04.78.43.37.17 gerard.scorletti@ec-lyon.fr
E2M2	ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION e2m2.universite-lyon.fr Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 ☎ 04.72.44.83.62 INSA : H. CHARLES secretariat.e2m2@univ-lyon1.fr	M. Fabrice CORDEY CNRS UMR 5276 Lab. de géologie de Lyon Université Claude Bernard Lyon 1 Bât. Géode 2 Rue Raphaël DUBOIS 69 622 Villeurbanne CEDEX ☎ 06.07.53.89.13 cordey@univ-lyon1.fr
EDISS	INTERDISCIPLINAIRE SCIENCES-SANTÉ www.ediss-lyon.fr Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 ☎ 04.72.44.83.62 INSA : M. LAGARDE secretariat.ediss@univ-lyon1.fr	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 Avenue Jean CAPELLE INSA de Lyon 69 621 Villeurbanne ☎ 04.72.68.49.09 📠 04.72.68.49.16 emmanuelle.canet@univ-lyon1.fr
INFOMATHS	INFORMATIQUE ET MATHÉMATIQUES infomaths.univ-lyon1.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3 ^e étage ☎ 04.72.43.80.46 📠 04.72.43.16.87 infomaths@univ-lyon1.fr	M. Luca ZAMBONI Bât. Braconnier 43 Boulevard du 11 novembre 1918 69 622 Villeurbanne CEDEX ☎ 04.26.23.45.52 zamboni@maths.univ-lyon1.fr
MATÉRIAUX	MATÉRIAUX DE LYON ed34.universite-lyon.fr Sec. : Marion COMBE ☎ 04.72.43.71.70 📠 04.72.43.87.12 Bât. Direction ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIÈRE INSA de Lyon MATEIS Bât. Saint-Exupéry 7 Avenue Jean CAPELLE 69 621 Villeurbanne CEDEX ☎ 04.72.43.71.70 📠 04.72.43.85.28 ed.materiaux@insa-lyon.fr
MEGA	MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE mega.universite-lyon.fr Sec. : Marion COMBE ☎ 04.72.43.71.70 📠 04.72.43.87.12 Bât. Direction mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis Avenue Jean CAPELLE 69 621 Villeurbanne CEDEX ☎ 04.72.43.71.70 📠 04.72.43.72.37 philippe.boisse@insa-lyon.fr
ScSo	SCSO* recherche.univ-lyon2.fr/scso/ Sec. : Viviane POLSINELLI Brigitte DUBOIS INSA : J.Y. TOUSSAINT ☎ 04.78.69.72.76 viviane.polsinelli@univ-lyon2.fr	M. Christian MONTES Université Lyon 2 86 Rue Pasteur 69 365 Lyon CEDEX 07 christian.montes@univ-lyon2.fr

* ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie.

Résumé

Les imageurs des hôpitaux produisent de plus en plus d'images 3D et il y a un nombre croissant d'études de cohortes. Afin d'ouvrir la voie à des méthodes utilisant de larges bases de données, il est nécessaire de développer des approches permettant de rendre ces bases cohérentes en recalant les images. Les principales méthodes actuelles de recalage de groupes utilisent des données denses (voxels) et sélectionnent une référence pour mettre en correspondance l'ensemble des images.

Nous proposons une approche de recalage par groupes, sans image de référence, en utilisant seulement des points d'intérêt (Surf3D), applicable à des bases de plusieurs centaines d'images médicales. Nous formulons un problème global fondé sur l'appariement de points d'intérêt. La variabilité inter-individu étant grande, le taux de faux positifs (paires aberrantes) peut être très important (70%). Une attention particulière est portée sur l'élimination des appariements erronés. Une première contribution effectue le recalage rigide de groupes d'images. Nous calculons les recalages de toutes les paires d'images. En s'appuyant sur le graphe complet de ces recalages, nous formulons le problème global en utilisant l'opérateur laplacien. Des expérimentations avec 400 images scanner CT 3D hétérogènes illustrent la robustesse de notre méthode et sa vitesse d'exécution. Une seconde contribution calcule le recalage déformable de groupes d'images. Nous utilisons des demi-transformations, paramétrées par des pyramides de B-splines, entre chaque image et un espace commun. Des comparaisons sur un jeu de données de référence montrent que notre approche est compétitive avec la littérature tout en étant beaucoup plus rapide.

Ces résultats montrent le potentiel des approches basées sur des points d'intérêt pour la mise en correspondance de grandes bases d'images. Nous illustrons les perspectives de notre approche par deux applications : la segmentation multi-atlas et l'anthropologie.

Mots clés Recalage 3D, Recalage de populations, Recalage déformable, Points d'intérêt, Big data

Abstract

The ever-increasing amount of medical images stored in hospitals offers a great opportunity for big data analysis. In order to pave the way for huge image groups screening, we need to develop methods able to make images databases consistent by group registering those images. Currently, group registration methods generally use dense, voxel-based, representations for images and often pick a reference to register images.

We propose a group registration framework, without reference image, by using only interest points (Surf3D), able to register hundreds of medical images. We formulate a global problem based on interest point matching. The inter-patient variability is high, and the outliers ratio can be large (70%). We pay a particular attention on inhibiting outliers contribution. Our first contribution is a two-step rigid groupwise registration. In the first step, we compute the pairwise rigid registration of each image pair. In a second step, a complete graph of those registrations allows us to formulate a global problem using the laplacian operator. We show experimental results for groups of up to 400 CT-scanner 3D heterogeneous images highlighting the robustness and speed of our approach. In our second contribution, we compute a non-rigid groupwise registration. Our approach involves half-transforms, parametrized by a b-spline pyramid, between each image and a common space. A reference dataset shows that our algorithm provides competitive results while being much faster than previous methods.

Those results show the potential of our interest point based registration method for huge datasets of 3D medical images. We also provide to promising perspectives : multi-atlas based segmentation and anthropology.

Keywords 3D registration, Population registration, Non-rigid registration, Interest points, Big data

Table des matières

1	Introduction	1
1.1	Contexte	1
1.2	Outil : points d'intérêt	2
1.3	Organisation de la thèse	3
2	Etat de l'art : mise en correspondance d'images	5
2.1	Tour d'horizon de la mise en correspondance d'images	7
2.1.1	Création de panoramas	7
2.1.2	Fusions d'images	8
2.1.3	De la 2D à la 3D	8
2.2	Recalage pour l'imagerie médicale	10
2.2.1	Intra-patient et inter-patients	10
2.2.2	Monomodal et multimodal	11
2.3	Modélisation	13
2.4	Transformations	13
2.4.1	Transformation globale	14
2.4.2	Transformations géométriques inspirées de modèles physiques	14
2.4.3	Transformations géométriques inspirées de la théorie de l'interpolation	15
2.5	Critères de qualité de mise en correspondance	21
2.5.1	Critères basés voxels	21
2.5.2	Critères géométriques	22
2.6	Optimisation	23
2.6.1	Descente de gradient	23
2.6.2	Méthode du gradient conjugué	24
2.6.3	Méthode de Newton	25
2.6.4	Méthode "quasi-Newton"	25
2.6.5	Méthode de Gauss-Newton	26
2.6.6	Méthode de Levenberg-Marquardt	26
2.6.7	Méthode du gradient stochastique	26
2.6.8	Recherche linéaire	27
2.6.9	Méthode des moindres carrés pondérés itératifs	28
2.6.10	Critères d'arrêt	28
2.6.11	Méthode "RANSAC"	29
2.7	Recalage d'ensemble	30
2.7.1	Recalage de paires et recalage d'ensembles	30
2.7.2	Applications médicales	30
2.7.3	Représentation par graphe	30
2.8	Extraction de caractéristiques : les points d'intérêt	33
2.8.1	Détection des points d'intérêt	34

2.8.2	Description des points d'intérêt	37
2.8.3	Exemple d'application : l'ajustement de faisceaux	43
2.8.4	Points d'intérêt 3D	45
3	Contribution 1 : mise en correspondance rigide rapide	47
3.1	Introduction	49
3.2	Challenges	50
3.2.1	Application médicale	50
3.2.2	Points d'intérêt	50
3.2.3	Mise en correspondance de grands groupes	50
3.3	SURF 3D	51
3.3.1	Images intégrales 3D	51
3.3.2	Approximation de la hessienne 3D	51
3.3.3	Choix des descripteurs	52
3.3.4	A priori liés à l'imagerie médicale	53
3.4	Mise en correspondance rigide de paires : FPR	53
3.4.1	Introduction	53
3.4.2	Données	54
3.4.3	Extraction des points d'intérêt	54
3.4.4	Appariement des points d'intérêt	54
3.4.5	Transformations	57
3.4.6	Estimation de la transformation	57
3.4.7	RANSAC	58
3.4.8	Conclusion	60
3.4.9	Limites du recalage par paire	61
3.5	Mise en correspondance rigide de grands groupes d'images médicales	62
3.5.1	Une nouvelle représentation par graphe	62
3.5.2	Mise en correspondance des paires	62
3.5.3	Modèle de déformation	63
3.5.4	Formalisation du problème	64
3.5.5	Élagage du graphe et connexité	66
3.5.6	Extension à une translation et un facteur d'échelle	68
3.6	Algorithme proposé et résultats expérimentaux	70
3.6.1	Fast Groupwise Registration : FGR	70
3.6.2	Masquage des yeux	72
3.7	Conclusion	73
4	Contribution 2 : mise en correspondance déformable de grands groupes	75
4.1	Introduction	77
4.2	Challenges	77
4.2.1	Quantité de données	77
4.2.2	Répétabilité et appariements aberrants	78
4.2.3	Complexité calculatoire	78
4.3	Contributions	79
4.3.1	Recalage sans référence par points	79
4.3.2	Pondération-EM	79
4.3.3	Méthode d'optimisation efficace	80
4.4	Appariement des points d'intérêt	80
4.4.1	Méthode	80
4.4.2	Discussion	80

4.5	Demi-transformées guidées par des points d'intérêt	81
4.5.1	Introduction	81
4.5.2	Données d'entrée	81
4.5.3	Choix des demi-transformées	82
4.5.4	Approche pyramidale	84
4.5.5	Attache aux données	85
4.6	Statistiques robustes	86
4.6.1	M-estimateur de Tukey	86
4.6.2	Pondération-EM	87
4.6.3	Test avec des données synthétiques	91
4.6.4	Test avec des données réelles	92
4.6.5	Comparaison	93
4.7	Optimisation	94
4.7.1	Notation matricielle	95
4.7.2	Moindres carrés pondérés itératifs	97
4.8	Régularisation	98
4.8.1	Introduction	98
4.8.2	Protocole expérimental : données synthétiques	99
4.8.3	Régularisation du déplacement des points de contrôle	103
4.8.4	Régularisation sur les couples appariés	104
4.8.5	Contrainte inter-images sur les points de contrôle	105
4.9	Conclusion	108
5	Applications et évaluations	109
5.1	Introduction	110
5.1.1	Méthode	110
5.2	Jeux de données	111
5.2.1	Base de données VISCERAL	111
5.3	Comparaison avec ANTs	114
5.3.1	Introduction	114
5.3.2	Mesure d'erreur sur les points de repère anatomique	114
5.3.3	Procédure de ANTs	114
5.3.4	Résultats : jeu de données A	114
5.3.5	Résultats : jeu de données B	120
5.3.6	Résultats : jeu de données $A \cup C$	123
5.3.7	Résultats : jeu de données $A+B+C$	124
5.3.8	Comportement de FGDR pour différents jeux de données	127
5.4	Conclusion	129
6	Perspectives	131
6.1	Introduction	132
6.2	Vers la mise en correspondance de 10 000 images	132
6.2.1	Introduction	132
6.2.2	Gradient stochastique	132
6.2.3	Observation	132
6.2.4	Méthode	132
6.2.5	Résultats préliminaires	133
6.2.6	Distribution et parallélisme	133
6.3	Application à la segmentation par atlas	134
6.3.1	Principe	134

6.3.2	Méthode	134
6.3.3	Résultats préliminaires	135
6.3.4	Perspectives	135
6.4	Application à l'anthropologie	135
6.4.1	Méthode	136
6.4.2	Résultats préliminaires	137
6.4.3	Analyse de la qualité de la mise en correspondance	137
6.4.4	Conclusion	138
7	Conclusion	139
	Liste des figures	141
	Liste des tableaux	143
	Bibliographie	145
	Liste des travaux et communications	153

Chapitre 1

Introduction

1.1 Contexte

Chaque année, les hôpitaux produisent des centaines de milliers d’images stockés dans leurs PACS (“Picture Archiving and Communication System”). Pour autant, ces images ne sont plus exploitées une fois l’examen effectué. Ces images pourraient bénéficier à de nombreuses applications. Nous pouvons imaginer l’analyse systématique de toutes ces images, par un ensemble de tests automatiques permettant le diagnostic préventif de diverses maladies. Mais avant ce scénario, il convient de faire l’état des lieux du traitement d’images médicales de masses aujourd’hui. Nous allons constater, au travers des différents défis listés ci-dessous, qu’il existe un grand nombre de problèmes à résoudre avant de parvenir à traiter ces bases d’images. La plupart des études de groupes utilisent quelques dizaines, voire centaines d’images. Elles nécessitent des protocoles précis en maîtrisant les caractéristiques des individus concernés. Le chemin est donc long entre ces études et la prise en charge de milliers d’individus, quelles que soient les modalités d’imagerie et la morphologie de ces individus.

Cohérence

Le principal défi adressé par ce manuscrit est la cohérence des images provenant de bases de données génériques. Les images médicales sont stockées en format DICOM (“Digital Imaging and Communications in Medicine”) qui enregistre les conditions de leur acquisition. Malgré le fait que ce format possède des informations spatiales, nous pouvons constater figure 1.1 que les images ne sont pas cohérentes, chaque système d’imagerie créant son propre système de coordonnées.

Au delà du simple positionnement, il est nécessaire d’avoir des bases de données cohérentes pour pouvoir pratiquer des diagnostics automatiques. Les bases de données d’images médicales pour la



FIGURE 1.1 – Un jeu de données contenant 400 images affichées sur une grille. Nous pouvons observer l’incohérence de la base d’images. Pour une base cohérente, les images représentant les mêmes parties devraient être à la même hauteur.

recherche sont généralement triées manuellement, nettoyées, recalées, en ne conservant que les images qui conviennent à l'application envisagée.

Un verrou à l'utilisation de larges bases de données dans le cadre d'études médicales à grande échelle est le manque d'outils permettant de rendre les grands jeux de données "corps entier" cohérents de manière automatisée. [XU et collab., 2016].

Variabilité

Un des challenges liés à la volonté de mettre en correspondance toutes les images provenant de n'importe quel individu est la variabilité morphologique et anatomique. Contrairement aux études scientifiques qui s'attachent à une seule typologie, nous voulons pouvoir inclure des hommes ou des femmes, des enfants ou des adultes, des individus sains ou malades.

Hétérogénéité

L'imagerie par rayon X est un système d'acquisition ionisant présentant des risques de cancer induits pour le patient. En conséquence, les zones imagées sont généralement restreintes aux zones d'intérêt de la pathologie recherchée. Contrairement aux approches actuelles, il est nécessaire de trouver des solutions permettant de mettre en correspondance des images contenant des zones imagées différentes. De plus, les modalités et les protocoles d'acquisition ne sont pas les mêmes et les appareils sont différents. Dans notre objectif de traitement de toutes les images, il est nécessaire de développer des méthodes afin de gérer ses informations partielles et hétérogènes, grâce à des algorithmes robustes.

Quantité de données

La plupart des expérimentations scientifiques actuelles [HILL et collab., 2001; WU et collab., 2011, 2012; ZÖLLEI et collab., 2005] de mise en correspondance de groupes d'images se limitent à des ensembles de 20 à 100 images. Dépasser cette limite est nécessaire pour traiter les images issues d'un PACS. Nous verrons plus particulièrement les utilisations possibles des points d'intérêt, dans cet objectif.

Références

La plupart des méthodes de mises en correspondance de groupes d'images s'appuient sur l'utilisation d'une référence [HILL et collab., 2001; WU et collab., 2011; ZÖLLEI et collab., 2005]. Cette image de référence, sélectionnée parmi les images ou construite (par exemple en calculant la moyenne des images) est utilisée comme "pivot". Cette image de référence permet ensuite de mettre en correspondance les images en liant chaque image avec ce pivot. Dans le cadre de l'utilisation de bases d'images variables et hétérogènes, l'utilisation d'une référence est contraignante. Le choix d'un pivot (image ou moyenne) n'a pas beaucoup de sens si la base contient des parties imagées différentes ou si elle est constituée d'hommes, de femmes, d'enfants ou d'adultes. Faire la moyenne d'une femme et d'un homme n'est évidemment pas très judicieux, tout comme choisir l'un ou l'autre comme référence.

1.2 Outil : points d'intérêt

Les points d'intérêt ont pour vocation la description des images en se limitant à certains points particuliers et en décrivant leurs voisinages (voir figure 1.2). Nous formulons des problèmes globaux fondés sur l'appariement de ces points d'intérêt entre les images. L'utilisation de points d'intérêt dans le cadre de la mise en correspondance d'images médicales nécessite de lever des verrous scientifiques spécifiques.

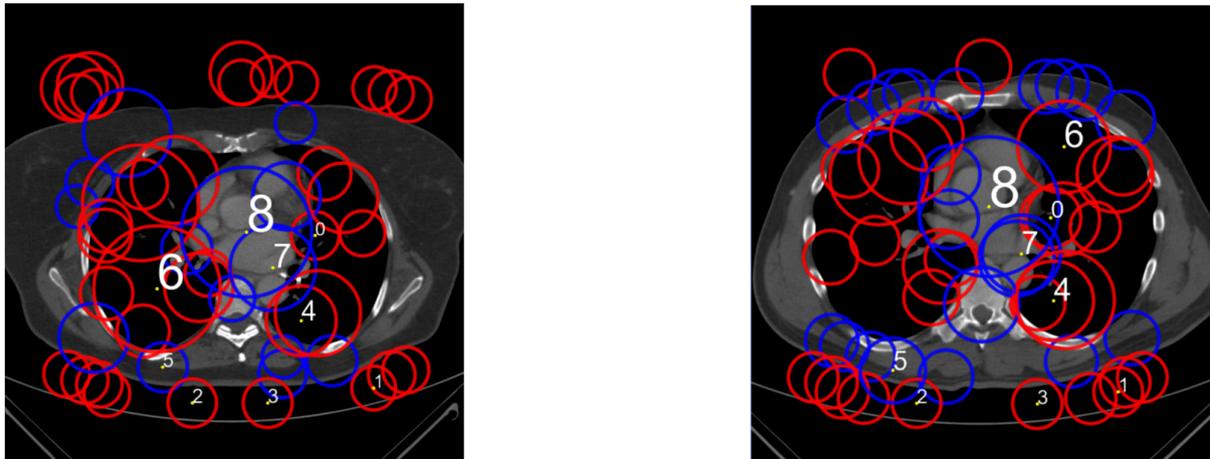


FIGURE 1.2 – Exemple de points d’intérêt dans des images médicales 2D. Les cercles rouges et bleus correspondent aux points d’intérêt détectés. La taille des cercles correspond à l’échelle des points d’intérêt. Les numéros correspondent aux points appariés. Nous pouvons observer par exemple la paire n°8 qui décrit la même zone et permettrait de guider une mise en correspondance. Nous pouvons aussi observer un exemple d’appariement aberrant avec la paire n°6.

Répétabilité inter-patients

Les points d’intérêt ont été initialement développés pour mettre en correspondance des scènes identiques. Dans le cadre d’images médicales, l’objectif consiste à mettre en correspondance des individus différents. Les structures anatomiques ne sont pas parfaitement identiques entre les individus et il est nécessaire de trouver le compromis entre la gestion de la variabilité et la robustesse de la méthode.

La gestion de la variabilité implique de mettre en correspondance des zones ressemblantes mais pas nécessairement identiques (exemple : paire n°8 de la figure 1.2). La principale conséquence du relâchement du critère de ressemblance est l’appariement de points différents (exemple : paire n°6 de la figure 1.2). Il est donc nécessaire de développer des méthodes permettant la gestion de ses appariements aberrants (faux positifs). Des méthodes de sacs de mots visuels [PHILBIN et collab., 2007], proposent des appariements sous forme d’histogramme d’occurrences afin de mettre en relation de grands groupes d’images. Malgré une problématique proche, ces méthodes ne constituent pas des listes de paires, permettant de faire du recalage d’image.

Problème global

L’appariement des points d’intérêt étant un processus léger, il est possible de mettre en correspondance les points entre toutes les images. Cela permet de formuler des problèmes globaux liant l’ensemble des images. Il est donc nécessaire de développer des méthodes d’optimisation permettant de résoudre ces problèmes globaux.

1.3 Organisation de la thèse

Le manuscrit est organisé comme suit : le chapitre 2 présente l’état de l’art des méthodes de mise en correspondance d’images médicales, portant un intérêt particulier à la mise en correspondance de groupes d’images et à l’utilisation des points d’intérêt. Les cinq chapitres suivants présentent les contributions de cette thèse. Le chapitre 3 présente l’utilisation des points d’intérêt 3D dans le cadre de la mise en correspondance de paires d’images médicales. Cette méthode est ensuite utilisée pour développer une méthode de recalage rigide de groupes d’images. Le chapitre 4 présente notre méthode de recalage

déformable de groupes et le chapitre 5 évalue cette méthode sur divers jeux de données. Dans le chapitre 6, nous abordons des perspectives prometteuses : optimisation stochastique, segmentation multi-atlas et anthropologie. Le chapitre 7 conclut le manuscrit.

Chapitre 2

Etat de l'art : mise en correspondance d'images

Sommaire

2.1	Tour d'horizon de la mise en correspondance d'images	7
2.1.1	Création de panoramas	7
2.1.2	Fusions d'images	8
2.1.3	De la 2D à la 3D	8
2.2	Recalage pour l'imagerie médicale	10
2.2.1	Intra-patient et inter-patients	10
2.2.2	Monomodal et multimodal	11
2.3	Modélisation	13
2.4	Transformations	13
2.4.1	Transformation globale	14
2.4.2	Transformations géométriques inspirées de modèles physiques	14
2.4.3	Transformations géométriques inspirées de la théorie de l'interpolation	15
	Splines	15
	Fonctions de base radiales	19
2.5	Critères de qualité de mise en correspondance	21
2.5.1	Critères basés voxels	21
	Méthodes basées "intensité"	21
	Méthodes basées "théorie de l'information"	21
2.5.2	Critères géométriques	22
	Extraction des points caractéristiques	22
	Appariement des points caractéristiques	22
	Estimation de la transformation	22
	Estimation conjointes : appariement et transformation	22
	Méthodes sans mise en correspondance	23
2.6	Optimisation	23
2.6.1	Descente de gradient	23
2.6.2	Méthode du gradient conjugué	24
2.6.3	Méthode de Newton	25
2.6.4	Méthode "quasi-Newton"	25
2.6.5	Méthode de Gauss-Newton	26
2.6.6	Méthode de Levenberg-Marquardt	26
2.6.7	Méthode du gradient stochastique	26

2.6.8	Recherche linéaire	27
2.6.9	Méthode des moindres carrés pondérés itératifs	28
2.6.10	Critères d'arrêt	28
2.6.11	Méthode "RANSAC"	29
2.7	Recalage d'ensemble	30
2.7.1	Recalage de paires et recalage d'ensembles	30
2.7.2	Applications médicales	30
2.7.3	Représentation par graphe	30
	Recalage exhaustif (graphe complet)	30
	Recalage en étoile	31
	Recalage en arbre couvrant minimal	33
2.8	Extraction de caractéristiques : les points d'intérêt	33
2.8.1	Détection des points d'intérêt	34
	Détecteur de Harris	35
	Détecteur SIFT	36
	Détecteur SURF	36
	Invariance en rotation	37
2.8.2	Description des points d'intérêt	37
	Descripteur SIFT	37
	Descripteur SURF	38
	Mise en correspondance	38
	Statistiques robustes	40
2.8.3	Exemple d'application : l'ajustement de faisceaux	43
	Optimisation globale	43
	Utilisation de statistiques robustes, gestion des points aberrants	44
2.8.4	Points d'intérêt 3D	45
	Invariance en rotation	45
	Points particuliers	45
	Temps de calcul	46

2.1 Tour d'horizon de la mise en correspondance d'images

Le recalage d'images, appelé aussi mise en correspondance d'images, est utilisé depuis de nombreuses années pour diverses applications. Avec l'essor de la photographie numérique et l'utilisation d'autres capteurs capables de produire des images (webcam, radar, lidar, imagerie satellitaire), son utilisation s'est encore élargie. Nous allons voir différentes applications de mise en correspondance d'images :

- création de panoramas
- fusion d'images
- reconstruction 3D

Nous nous intéresserons ensuite au sujet principal de ce manuscrit : la mise en correspondance d'images médicales.

2.1.1 Création de panoramas

La création de panoramas est certainement l'utilisation grand public la plus connue dont un exemple est illustré figure 2.1. Les appareils photos numériques et les smartphones se proposent de plus en plus de produire des clichés dits "grand-angle" par assemblage de plusieurs images. Au delà de cet exemple, la création de panoramas consiste à prendre de multiples photographies, soit en acquérant successivement plusieurs images soit en utilisant plusieurs capteurs.

Avec les images créées possédant un certain taux de recouvrement, des algorithmes peuvent reconstruire une image panoramique en détectant ces zones de recouvrement puis en calculant le décalage entre les images.

Des techniques avancées proposent de choisir le mode de projection ou le mode de fusion des pixels à l'intérieur des zones de recouvrement (voir HDR).

L'utilisation de la création de panorama a enregistré un regain d'intérêt avec l'arrivée des photos et vidéos dites immersives, à 360°. Il y a de plus en plus de demandes de création d'images sphériques, de nouveaux systèmes d'acquisition munis de plusieurs capteurs photographiques utilisent des algorithmes issus de la création de panoramas.

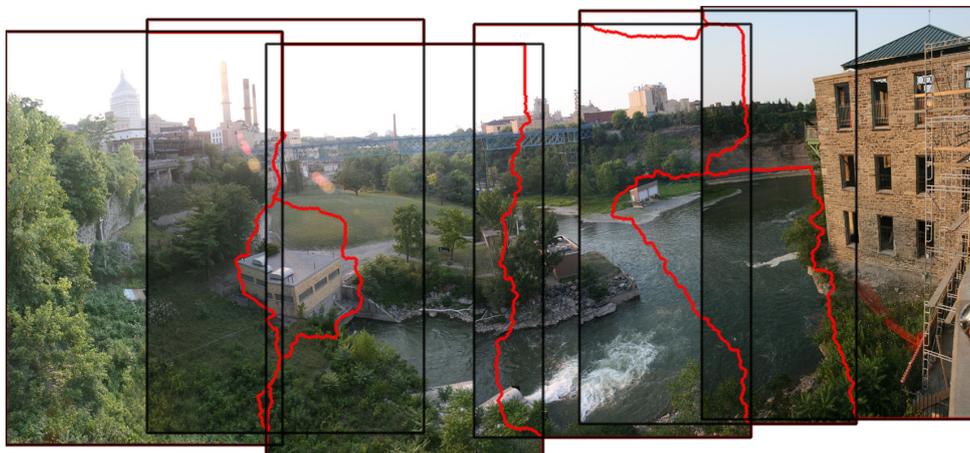


FIGURE 2.1 – Vue panoramique créée à partir de plusieurs images mises en correspondance. Auteur : Noso1

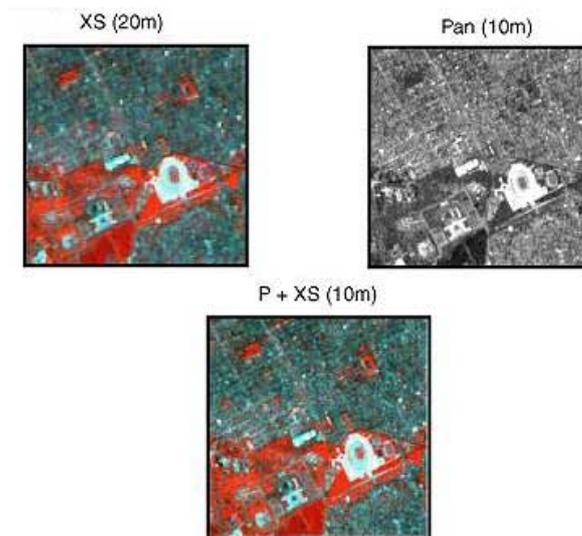


FIGURE 2.2 – Fusion de modalités sur des images satellites : Multispectrale (XS, basse résolution) et panchromatique (Pan, haute résolution). Source : <http://eoeu.belspo.be>

2.1.2 Fusions d'images

La fusion d'images fait partie de la discipline scientifique de la fusion de données. Elle consiste à créer des images agrégeant les données de plusieurs sources hétérogènes.

Un des exemples historique, est la restitution cartographique satellitaire, voir figure 2.2. Elle peut être le résultat de la fusion d'images provenant de divers capteurs à différentes résolutions, comme par exemple :

- des données optiques dans le spectre visible,
- des données images infrarouges,
- des données images radars,
- des données images hyper-spectrales.

Ces données peuvent provenir de capteurs d'un même satellite ou de plusieurs satellites. Chacune des situations induit des algorithmes différents : correction d'angle de vue (orthophotographie [RAU et collab. \[2002\]](#)), recalage des positions inter satellites (voir [HARTLEY et ZISSERMAN \[2003\]](#)).

Un autre exemple, grand public, de fusion d'images est la création d'images possédant une grande plage dynamique ("High Dynamical Range" : HDR) par assemblage d'images possédant des expositions différentes, voir figure 2.3. Les images étant prises successivement par le même appareil, une étape de recalage entre les images peut être nécessaire afin de corriger le mouvement.

2.1.3 De la 2D à la 3D

Les exemples précédents sont tous des exemples en deux dimensions, néanmoins d'un point de vue fonctionnel, la mise en correspondance d'images ne se limite pas à la 2D. L'imagerie médicale est le principal exemple utilisant de nombreuses images acquises en trois dimensions, et comme nous le verrons tout au long de ce manuscrit, la mise en correspondance d'images s'y applique aussi.

D'autres exemples font le lien entre l'image 2D, et le monde 3D dont elle est la projection. Nous pouvons citer le suivi d'objets faisant la mise en correspondance entre le modèle 3D de l'objet et sa projection [[WUEST et collab., 2007](#)], voir figure 2.4.



FIGURE 2.3 – Vue panoramique créée à partir de plusieurs images mises en correspondance. À gauche, les images sources, capturées avec des expositions différentes. À droite, la synthèse par fusion de ces images. Auteur : DaveWilsonPhotography



FIGURE 2.4 – Reconnaissance et suivi d'objets basés sur un modèle 3D. Nous pouvons observer le modèle 3D en fil de fer incrusté dans l'image. Crédit : [WUEST et collab. \[2007\]](#).

Il existe aussi l'exercice inverse, appelé Ajustement de Faisceaux (AF, "Bundle Adjustment", [TRIGGS et collab. \[2000\]](#)) consistant à reprojeter un ensemble d'images 2D dans une scène 3D, en vue de reconstruire un modèle tridimensionnel de la scène. Nous détaillerons cet exemple plus précisément dans un chapitre dédié, car il possède de nombreux éléments pouvant constituer une analogie avec l'approche développée dans ce manuscrit.

2.2 Recalage pour l'imagerie médicale

Le recalage d'images fait partie des principaux défis de l'analyse d'images médicales, soit en tâche principale, soit comme prétraitement avant l'application d'autres algorithmes. Le recalage sera encore plus indispensable avec l'utilisation croissante d'approches par apprentissage, nécessitant de grandes bases d'images cohérentes.

La mise en correspondance d'images a déjà de nombreuses applications en imagerie médicale, comme la reconstruction d'images ([HUANG et collab. \[2008\]](#)), les études longitudinales ([SCAHILL et collab. \[2003\]](#)), la planification en radiothérapie ([KEALL et collab. \[2005\]](#)) ou encore des applications plus récentes avec par exemple l'émergence de la réalité augmentée [[SU et collab., 2009](#)].

Le recalage d'images consiste à établir la correspondance spatiale entre deux images différentes. La distinction peut être faite entre la mise en correspondance globale et la mise en correspondance déformable (ou locale). La mise en correspondance globale consiste à appliquer une transformation avec un très faible nombre de variables. Par opposition, le recalage déformable consiste à associer les deux images à travers un champ de déformation non linéaire, ou variant spatialement, possédant généralement un grand nombre de degrés de liberté. Plusieurs articles de synthèse traitent ce sujet [BROWN \[1992\]](#); [MAINTZ et VIERGEVER \[1998\]](#); [SOTIRAS et collab. \[2013\]](#).

La mise en correspondance d'images est un problème dit mal posé, considéré sous l'angle d'un problème inverse (voir [FISCHER et MODERSITZKI \[2008\]](#)). Il en résulte qu'il ne peut pas y avoir de solution générale et que chaque discipline, chaque champ d'application, a développé ses propres méthodes, selon les a priori ou prémisses propres à son application.

Les différentes approches de mise en correspondances d'images diffèrent principalement par trois éléments : le critère de mise en correspondance, le modèle de déformation et la stratégie d'optimisation. Les deux premiers critères modélisent directement la forme de la fonction objectif. En effet, dans le cas d'un champ de déformation paramétré par un faible nombre de variables, une régularisation directement dépendante du champ de déformation est choisie. Le modèle de déformation possédera intrinsèquement les différentes propriétés, connaissances a priori, souhaitées.

La stratégie d'optimisation dépend naturellement, elle aussi, des critères de mise en correspondance et du modèle de déformation. Elle peut aussi porter intrinsèquement une régularisation. Dans le cas d'un modèle déformable dit "riche", avec de nombreuses variables, une régularisation explicite peut s'avérer nécessaire.

2.2.1 Intra-patient et inter-patients

La mise en correspondance d'images peut porter sur la mise en lien d'images provenant d'un même patient (exemple figure [2.5 droite](#)) ou de patients différents (exemple figure [2.5 gauche](#)).

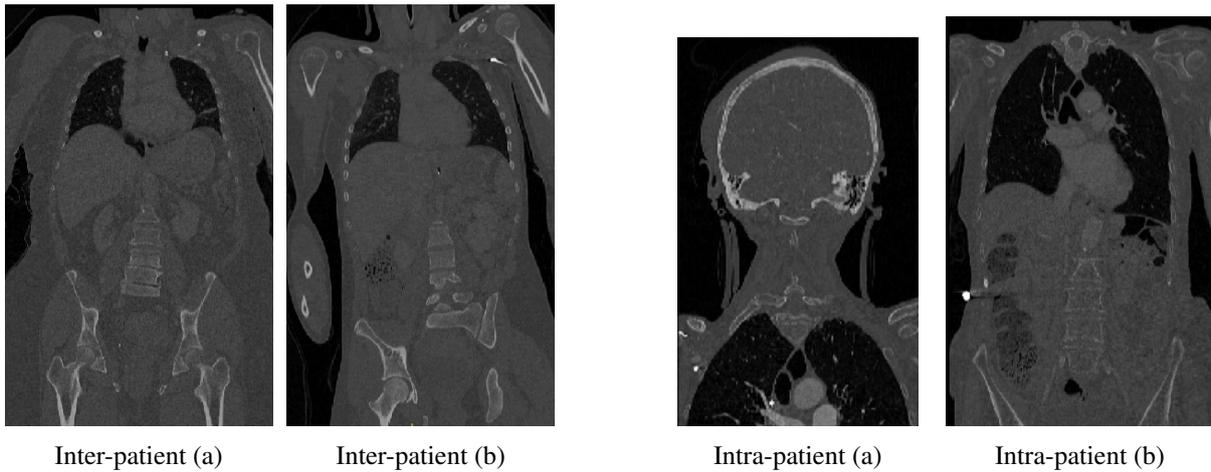


FIGURE 2.5 – À gauche, un exemple de deux coupes d’images tomographiques 3D provenant de patients différents. À droite, exemple d’images provenant du même patient, mais où la partie imagée diffère. Origine des images : Base VISCERAL.

Parmi les applications de mise en correspondance d’images intra-patient nous pouvons citer :

- Le suivi longitudinal,
- La planification en radiothérapie.

Parmi les applications inter-patients :

- Analyse de cohortes,
- Analyse de populations,
- Diagnostic comparatif.

Dans le cadre de la mise en correspondance d’images provenant d’un même patient, des a priori issus du monde physique, comme des lois (ex : la conservation de la masse), des modèles (ex : le modèle de corps élastique), ou certaines contraintes (ex : la préservation de la topologie) sont souhaitables. Lorsque ces images proviennent de patients différents, ces contraintes ne permettent parfois pas de décrire convenablement les déformations, selon la variabilité inter-patients.

2.2.2 Monomodal et multimodal

La mise en correspondance d’images peut concerner des images provenant de la même modalité d’acquisition (2.6 gauche) IRM T1, T2, diffusion, rayons X, etc... mais aussi de modalités d’acquisition différentes (2.6 droite) IRM T1 & T2, IRM & rayons X, etc... Des méthodes différentes sont nécessaires selon les modalités, principalement en ce qui concerne le critère de similarité.

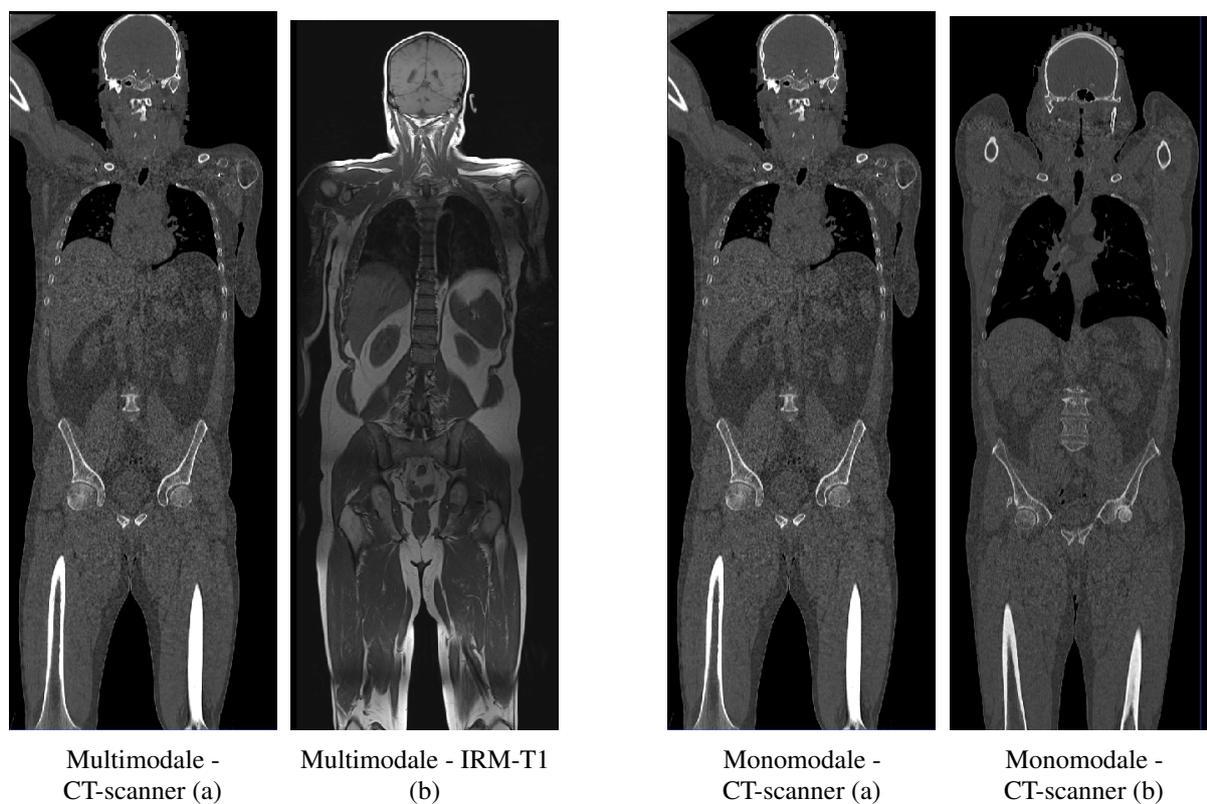


FIGURE 2.6 – À gauche, un exemple de deux coupes d'un même patient provenant d'imageurs différents (CT-scanner à gauche, IRM à droite). À droite, exemple d'images provenant du même patient, mais où la partie imagée diffère. Origine des images : Base VISCERAL.

2.3 Modélisation

La mise en correspondance d'images est utilisée pour associer deux ou plusieurs images. Les approches par "mise en correspondance de paires" consistent à mettre en correspondance deux images. D'autres méthodes, appelées "recalage de groupes", "recalage de cohortes" ou encore "recalage de populations" associent plus de deux images.

Dans le cas du recalage de paires d'images, une image est généralement considérée comme l'image source ou image mouvante, tandis que l'autre image est appelée l'image cible ou image fixe. Dans cette section, nous noterons S l'image source tandis que l'image cible sera notée T . Les deux images sont définies dans le domaine image Ω et sont mises en relation à travers une transformation $\tau : \Omega \mapsto \Omega$. En imagerie médicale, et dans notre application, les images sont généralement 3D, donc $\Omega = \mathbb{R}^3$.

La mise en correspondance d'images consiste à estimer la transformation qui optimise (minimise ou maximise) une énergie issue d'une fonction objectif de la forme :

$$f(T, S, \tau) = \mathcal{M}(T, \tau(S)) + \mathcal{R}(\tau) \quad (2.1)$$

La fonction $f(T, S, \tau)$ comprend deux termes :

- Le terme \mathcal{M} rend compte du degré d'alignement entre l'image cible T et l'image source S déformée. Dans ce manuscrit, nous ferons référence à ce terme comme le critère de mise en correspondance, le critère de similarité ou la mesure de distance, en fonction des méthodes.
- Le terme \mathcal{R} est un terme de régularisation qui a pour but de contraindre la solution afin qu'elle possède certaines propriétés. Elle permet d'introduire des a priori spécifiques à l'application visée et à restreindre la solution, le problème étant mal posé.

La transformation τ permet de faire correspondre n'importe quel point $p \in \Omega$ de l'image source à une position $\tau(p) \in \Omega$ dans l'image cible en ajoutant un champ de déplacement u .

$$\tau(p) = p + u(p) \quad (2.2)$$

Le recalage d'images consiste à optimiser la fonction $f(T, S, \tau)$ en la minimisant ou maximisant selon le critère de mise en correspondance choisi.

2.4 Transformations

Le modèle de déformation est très important dans le processus de mise en correspondance d'images. De son choix dépend directement le choix de la régularisation, la richesse de la déformation et la difficulté algorithmique.

Le nombre de degrés de liberté du modèle, peut être très faible, comme dans le cadre d'un recalage rigide, à plusieurs millions dans le cadre d'une estimation de déplacement par voxel. La complexité algorithmique dépend directement de ce choix. Un compromis apparaît entre la richesse du modèle et le temps de calcul.

Dans le cadre de la mise en correspondance d'images médicales, plusieurs a priori ont été introduits afin que le modèle de déformation possède des propriétés prédéfinies. Ces a priori peuvent être mis en relation directe avec le terme de régularisation \mathcal{R} de l'équation 2.1 et doivent contraindre la solution afin de rendre le problème mieux posé. Ces a priori peuvent aussi être liés à l'application sous-jacente du processus de mise en correspondance et être souhaitables lors de l'utilisation ultérieure des champs de déformations. Ces propriétés incluent, sans s'y limiter, l'inversibilité, la symétrie et le difféomorphisme. Certaines propriétés sont inhérentes au modèle de déformation choisi. Dans d'autres cas, des stratégies explicites sont mises en place pour les respecter.

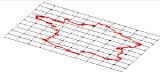
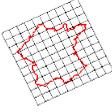
Groupe	Degrés de liberté	Distorsion	Propriétés invariantes
Affine	12		Parallélisme des plans, ratios des volumes, centroïdes.
Similitude	7		Ratios des longueurs.
Euclidienne (rigide)	6		Volumes, longueurs.

TABLEAU 2.1 – Classification des transformations globales : les différentes transformations utilisées pour les mises en correspondance globales. Source partielle : [HARTLEY et ZISSERMAN \[2003\]](#)

- **Inversibilité** : Dans le cadre d'applications médicales, la capacité à produire des transformations inversibles est fondamentale. Le champ de déformation associe à chaque point de l'image source un point de l'image cible. Afin de construire l'image source déformée, il est nécessaire d'associer à chaque position de l'image cible une position de l'image source, en utilisant l'inverse τ^{-1} de la transformation estimée. Cette propriété peut être satisfaite en estimant simultanément le champ de déformation et son inverse. Le critère de similarité rend compte, dans ce cas-ci, de l'alignement de l'image source après application du champ de déformation dans l'espace cible et de l'alignement de l'image cible déformé par l'inverse de ce champ dans l'espace source.
- **Symétrie** : La majorité des algorithmes de mise en correspondance sont asymétriques. Cela signifie qu'inter-changer les images sources et cibles a pour résultat des déformations différentes. Ce choix introduit donc un biais dans la suite du processus et doit être pris en compte. Les approches ne pénalisent généralement pas directement l'asymétrie, mais emploient l'une des stratégies suivantes lorsque le respect de la symétrie est souhaité. La première solution consiste à employer une fonction objectif et un modèle de déformation intrinsèquement symétrique. La deuxième solution consiste à estimer deux transformations au lieu d'une seule. Chaque transformation, respectivement τ_a et τ_b , projette une image dans un espace commun. La transformation permettant de passer d'une image a à une image b est ensuite calculée en inversant une transformation et en la composant avec l'autre : $\tau_{a \rightarrow b} = \tau_b^{-1} \circ \tau_a$.
- **difféomorphisme** : Une transformation est un difféomorphisme si son inverse et elle-même sont différentiables. Un difféomorphisme préserve aussi la topologie. C'est cette contrainte qui apparaît le plus souvent comme souhaitable dans les différentes approches de mise en correspondance. Le difféomorphisme est une synthèse intéressante des différents a priori souhaitables.

2.4.1 Transformation globale

Le recalage global (voir [ASHBURNER et FRISTON \[2007\]](#)) consiste à trouver une transformation τ linéaire ou globale entre les images source et cible. Cette transformation globale peut être entre autres une application euclidienne, affine ou projective (voir table 2.1). Le recalage rigide peut être utilisé pour l'assemblage de mosaïques, la fusion de modalités ou encore comme étape d'initialisation pour une mise en correspondance déformable.

2.4.2 Transformations géométriques inspirées de modèles physiques

Une catégorie d'approches s'inspire de modèles physiques pour définir le champ de déformation. Dans ce cas, un champ de déformation dense est généralement défini, par voxel. Ces choix de contraintes physiques permettent d'ajouter une régularisation au champ et parfois d'utiliser des schémas d'optimisation

particuliers. Nous allons détailler les cinq modèles les plus utilisés : modélisation par corps élastique, modélisation par écoulement de fluides visqueux, modèles de diffusions, recalage par courbure et flot de difféomorphisme.

- **Modèle de corps élastique** : L'image déformée est modélisée par un corps élastique (voir [BAJCSY et KOVAČIČ \[1989\]](#); [BROIT \[1981\]](#); [DAVATZIKOS \[1997\]](#)). L'équation aux dérivés partielles de Navier-Cauchy permet de décrire cette déformation :

$$\mu \nabla^2 \mathbf{u} + (\mu + \lambda) \nabla(\nabla \cdot \mathbf{u}) + \mathbf{F} = 0 \quad (2.3)$$

où $\mathbf{F}(\mathbf{u})$ est le champ de force qui dirige la mise en correspondance en fonction du critère de similitude, μ est la rigidité du matériau et λ est le premier coefficient de Lamé.

- **Modélisation par écoulement de fluides visqueux** : L'image déformée est modélisée par un fluide visqueux (voir [BRO-NIELSEN et GRAMKOW \[1996\]](#); [CHRISTENSEN et collab. \[1996\]](#)). Le principe consiste à définir le champ de déformation par l'intégrale du champ de vitesse, à travers l'équation suivante : $\mathbf{v}(\mathbf{x};t) = \partial_t \mathbf{u}(\mathbf{x};t) + (\nabla \mathbf{u}(\mathbf{x};t) \mathbf{v}(\mathbf{x};t))$. La transformation est guidée par l'équation de Navier-Stokes :

$$\mu \nabla^2 \mathbf{v} + (\mu + \lambda) \nabla(\nabla \cdot \mathbf{v}) + \mathbf{F} = 0 \quad (2.4)$$

- **Modèle de diffusion** : La déformation est modélisée par l'équation de diffusion suivante (voir [FISHER et MODERSITZKI \[2002\]](#) pour une version optimisée) :

$$\Delta \mathbf{u} + \mathbf{F} = 0 \quad (2.5)$$

Cette équation décrit une régularisation et elle n'est généralement pas employée directement dans la fonction objectif car elle peut être remplacée de manière efficace par une convolution avec un noyau Gaussien (voir [CACHIER et PENNEC \[2000\]](#)). Elle a donnée lieu aux approches dites par "Démons" (voir [THIRION \[1998\]](#)).

- **Autres modèles** : Deux autres méthodes inspirées d'analogies avec la physique sont le recalage par courbure, voir [FISCHER et MODERSITZKI \[2003\]](#); [GLOCKER et collab. \[2009\]](#), qui proposent d'exploiter l'équation $\Delta^2 \mathbf{u} + \mathbf{F} = 0$ et le flot de difféomorphisme, où la déformation est modélisée en intégrant la vitesse issue des équations de transport de Lagrange (voir [CHRISTENSEN et collab. \[1997\]](#); [DUPUIS et collab. \[1998\]](#)).

2.4.3 Transformations géométriques inspirées de la théorie de l'interpolation

Splines

Les splines désignent génériquement les modèles de fonctions paramétriques. Historiquement, une spline, appelée "cerce", "latte ou "baguette" est un objet physique servant à tracer des courbes. Il était beaucoup utilisé par les ingénieurs et les architectes (figure 2.7). Il était fait d'un morceau de bois que l'on courbait en le fixant en certains points. Cette section s'inspire des travaux de thèse [BRUNET \[2010\]](#) en ce qui concerne le formalisme des B-splines.

Définition mathématique Une courbe spline [[DE BOOR, 1978](#)] est une fonction polynomiale par morceaux, notée $S : [a,b] \rightarrow \mathbb{R}$ définie sur un intervalle $[a,b]$ divisé en sous intervalles $[t_{i-1}, t_i]$ tel que

$$a = t_0 < t_1 < \dots < t_{k-1} < t_k = b \quad (2.6)$$

Les t_i sont appelés les points de contrôle. Pour chaque intervalle, on définit un polynôme

$$P_i : [t_{i-1}, t_i] \rightarrow \mathbb{R} \quad (2.7)$$

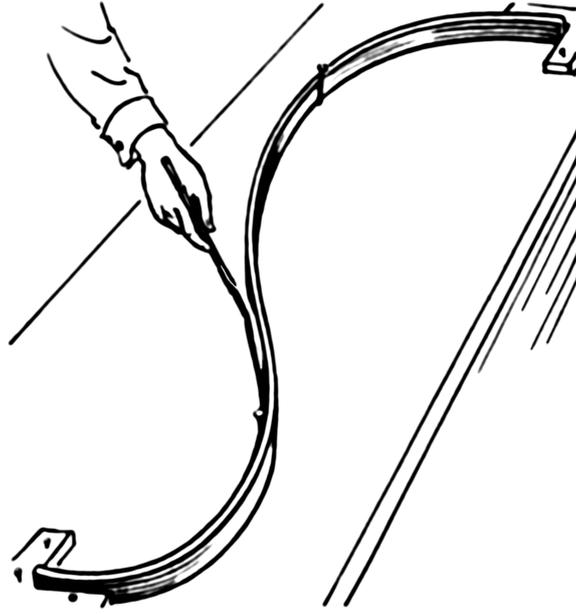


FIGURE 2.7 – Un cerce représentant historiquement une spline. Crédit : Pearson Scott Foresman

Une spline possédant k intervalles est donc définie par :

$$\begin{aligned} S(t) &= P_1(t), t_0 \leq t < t_1, \\ S(t) &= P_2(t), t_1 \leq t < t_2, \\ &\vdots \\ S(t) &= P_k(t), t_{k-1} \leq t < t_k, \end{aligned} \tag{2.8}$$

Le degré d'une spline est défini comme le degré du polynôme $P_i(t)$ ayant le degré le plus élevé.

La dérivabilité d'une spline est limitée par les jointures entre les intervalles, les polynômes ayant une dérivabilité infinie.

En vérifiant la propriété :

$$P_i^{(j)}(t_i) = P_{i+1}^{(j)}(t_i) \quad \forall j \in [0, n], i \in [1, k-1] \tag{2.9}$$

on dit que la spline est de continuité n , notée C_n .

B-splines La fonction de base B-spline [EILERS et MARX, 1996] de degré n et possédant les points de contrôle $t_i < \dots < t_{i+n+1}$ peut être définie récursivement avec la relation suivante, appelée formule de récursivité de Cox de Boor :

$$b_j^0(t) := \begin{cases} 1 & \text{si } t_j \leq t < t_{j+1} \\ 0 & \text{sinon} \end{cases} \tag{2.10}$$

$$b_j^n(t) := \frac{t-t_j}{t_{j+n}-t_j} b_j^{n-1}(t) + \frac{t_{j+n+1}-t}{t_{j+n+1}-t_{j+1}} b_{j+1}^{n-1}(t) \tag{2.11}$$

Le support d'une B-spline est fini. ce qui signifie qu'elle a une valeur nulle en dehors de son domaine de définition naturel :

$$B_j^n(t) = 0 \quad \text{si } t \notin [t_i, t_{i+n+1}] \tag{2.12}$$

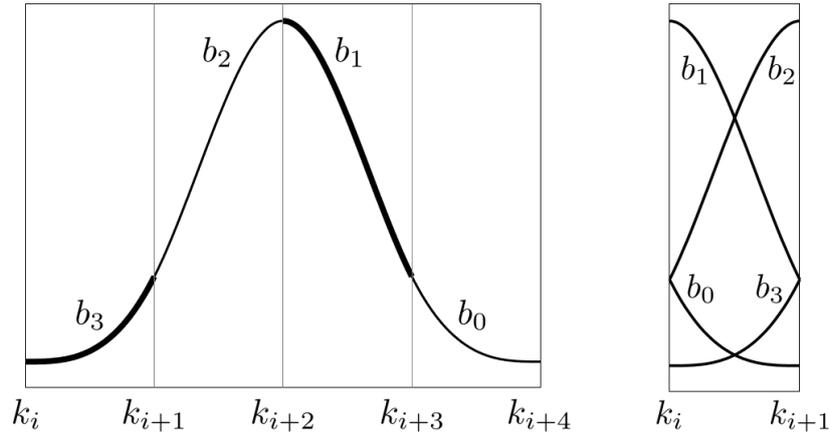


FIGURE 2.8 – Une B-spline cubique uniforme est composée de 4 sections, chacune d'entre elles étant un polynôme de degré 3. (Thèse BRUNET [2010])

Comme une B-spline de degré n couvre $n + 1$ intervalles, $k - n$ B-Splines indépendantes sont définies sur l'intervalle $[a, b]$. Afin d'avoir un ensemble complet, il faut définir $2n$ points de contrôle supplémentaires. Le nouvel ensemble des points de contrôle est donc $[t_{-n}, t_{k+n}]$.

Une B-spline $S^n(t)$ s'écrit de manière unique comme la combinaison linéaire de $k + n$ fonctions de bases $B_j^n(t)$ pour $j \in [-n, k - 1]$:

$$S^n(t) = \sum_{j=-n}^{k-1} d_j B_j^n(t) \quad (2.13)$$

Les splines peuvent s'écrire comme des combinaisons linéaires de fonctions de bases B-splines. Pour un ensemble de points de contrôle donné, la B-spline est unique, le "B" signifiant "Base".

L'ajustement d'une B-spline aux données consiste à estimer les paramètres d_i .

Nota : Une propriété intéressante des fonctions de base des B-splines est qu'elles forment une partition de l'unité :

$$\sum_{j=-n}^{k-1} B_j^n(t) = 1 \quad \forall t \in [t_0, t_n] \quad (2.14)$$

B-Spline cubique uniforme Une B-Spline cubique uniforme (UCBS : "Uniform Cubic B-splines") est un cas particulier de B-spline, particulièrement intéressant pour sa simplicité et son efficacité. Comme son nom l'indique, c'est une B-Spline de degré 3. Elle est uniforme, ce qui signifie que ses points de contrôle sont équidistants (figure 2.8 et 2.9). Dans ce cas, les fonctions de base des B-splines sont des copies décalées les une des autres, afin de simplifier l'écriture, nous noterons B_j^3 avec B_j :

$$B_j(t) = B_j(t - (j + i)s) \quad \forall (i, j) \in [-n, k - 1]^2 \quad (2.15)$$

L'utilisation d'UCBS peut être simplifiée par l'emploi d'une abscisse normalisée. Cela consiste à faire coïncider le premier point de contrôle avec le zéro de l'abscisse, et à avoir une largeur d'intervalle unitaire. L'abscisse normalisée $v(t)$ est définie de la manière suivante :

$$v(t) = \frac{t - t_0}{s} \quad (2.16)$$

où s est la largeur d'un intervalle ($s = k_{j+1} - k_j \quad \forall i \in [-d, n - 2]$).

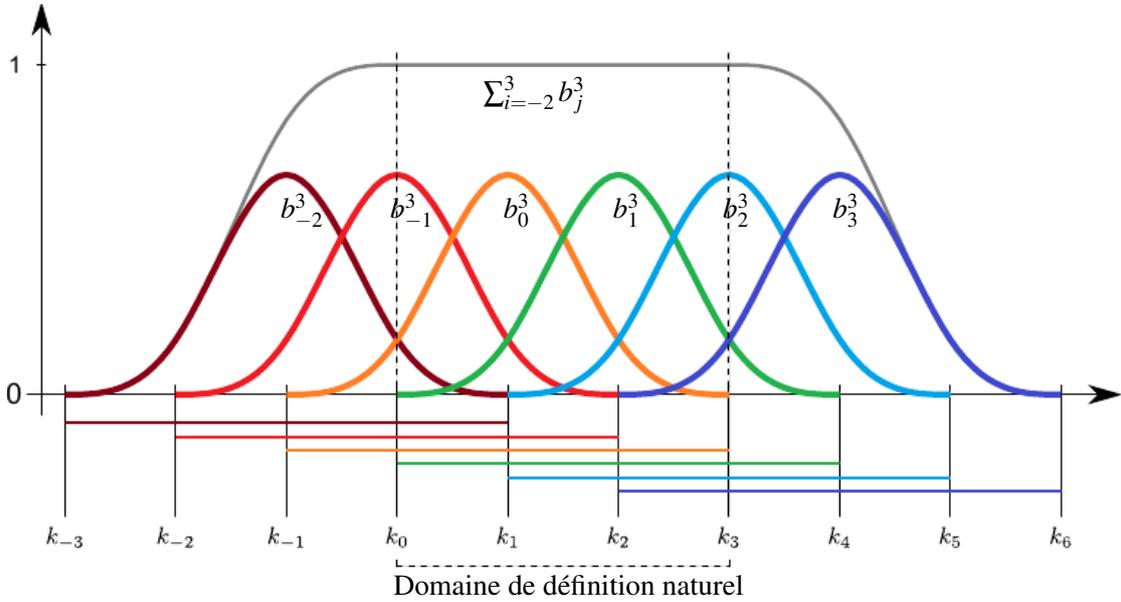


FIGURE 2.9 – Les fonctions de base des B-splines cubiques uniformes. Pour tout point du domaine naturel de définition, il y a toujours 4 fonctions de base non-nulles. Source : Thèse BRUNET [2010]

On peut ainsi définir l'intervalle dans lequel repose un point t : $\iota(t)$ est une fonction de $[t_{-3}, t_{k+3}]$ vers $[-3, k+2]$:

$$\iota(t) := \begin{cases} \lfloor v(t) \rfloor & \text{si } t \in [t_{-3}, t_{k+3}[\\ k+2 & \text{si } t = t_{k+3} \end{cases} \quad (2.17)$$

Avec ces deux fonctions, nous pouvons définir l'abscisse normalisée $o(t)$ de t :

$$o(t) = v(t) - \iota(t) \quad (2.18)$$

Il correspond aux coordonnées du point t dans le référentiel local où la limite inférieure correspond à l'intervalle coïncident à $t = 0$ et possède une largeur de 1.

Les fonctions de bases des B-splines cubiques uniformes peuvent être explicitées en utilisant la règle de récurrence. Nous considérons ici les coordonnées normalisées (si ce n'est pas le cas, il faut remplacer t par $o(t)$ dans les formules suivantes) :

$$B_j(t) := \begin{cases} b_3(t) = \frac{1}{6}t^3 & \text{si } t \in [t_i, t_{i+1}[\\ b_2(t) = \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1) & \text{si } t \in [t_{i+1}, t_{i+2}[\\ b_1(t) = \frac{1}{6}(3t^3 - 9t^2 + 4) & \text{si } t \in [t_{i+2}, t_{i+3}[\\ b_0(t) = \frac{1}{6}(-t^3 + 3t^2 - 3t + 1) & \text{si } t \in [t_{i+3}, t_{i+4}[\\ 0 & \text{sinon} \end{cases} \quad (2.19)$$

Initialement, les B-splines sont des fonctions définies de $\mathbb{R} \rightarrow \mathbb{R}$. Dans le cadre de la mise en correspondance d'images médicales, il est souhaitable d'interpoler de $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Nous verrons dans la partie suivante comment faire cela, à travers l'utilisation du produit tensoriel de B-splines, et de B-splines vectorielles.

Produit tensoriels de B-Splines (en dimension 3) Les produits tensoriels de B-Splines (TP : Tensor Product) permettent de généraliser l'interpolation par B-spline à un nombre de dimensions quelconque. Afin d'alléger l'écriture et de rester dans le cadre du recalage d'images médicales, nous allons aborder ici uniquement le développement en 3 dimensions. L'approche se généralisant pour un nombre quelconque de dimensions.

En dimension 3, l'ensemble des points de contrôle consiste en une grille possédant k_x , k_y et k_z points de contrôle respectivement dans les directions x , y , z . Nous pouvons définir la fonction de base de B-splines multivariées dans l'espace $B_{i,j,k}^n : \mathbf{t} \in \mathbb{R}^3 \rightarrow \mathbb{R}$ comme le produit tensoriel des fonctions de base monovariées, définit par :

$$B_{i,j,k}^n(\mathbf{t}) = B_i^n(t_x)B_j^n(t_y)B_k^n(t_z) \quad (2.20)$$

De cette manière nous pouvons définir le produit tensoriel de B-spline en dimension 3 de la manière suivante :

$$S^n(\mathbf{t}) = \sum_{i=-n}^{k_x-1} \sum_{j=-n}^{k_y-1} \sum_{k=-n}^{k_z-1} d_{i,j,k} B_{i,j,k}^n(t) \quad (2.21)$$

En 1D, l'utilisation de Splines cubique implique que 4 points de contrôle contribuent à l'interpolation d'un point. En 3D, l'utilisation de splines cubiques implique de $4 \times 4 \times 4 = 64$ points de contrôle influencent l'interpolation d'un point. C'est donc un sous ensemble de la grille, un cube, contenant 64 points de contrôle.

Champ de déformation issu de TP-UCBS On peut facilement construire un champ vectoriel à partir d'un champ de TP-UCBS scalaire. Il suffit de changer les poids des points de contrôle par des vecteurs. Dans ce cas là, la TP-UCBS est définie comme $S : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Il suffit de remplacer les coefficients de B-splines scalaires $d_{j,j,k} \in \mathbb{R}$ par des coefficients de splines vecteurs : $\mathbf{d}_{j,j,k} \in \mathbb{R}^3$. De cette manière l'interpolation devient :

$$S^n(\mathbf{t}) = \sum_{i=-n}^{k_x-1} \sum_{j=-n}^{k_y-1} \sum_{k=-n}^{k_z-1} \mathbf{d}_{i,j,k} B_{i,j,k}^n(t) \quad (2.22)$$

Dans le cadre d'une mise en correspondance d'images médicales volumiques, c'est cette interpolation qui va être utilisée, car c'est un déplacement dans l'espace, ce qui correspond bien à une déformation. La notation \mathbf{d} prend tout son sens car les TP-UCBS vont interpoler les déplacements des points des contrôle $\mathbf{t}_{i,j,k}$. Chaque déplacement de l'espace est portée par $4 \times 4 \times 4 = 64$ points de contrôle, ce qui permet, avec les propriétés des UCBS, d'avoir une déformation lisse et adaptable, tout en limitant la complexité de calcul.

Fonctions de base radiales

Cas général Une autre stratégie classique d'interpolation est l'utilisation de fonctions de base radiales (RBF : "Radial Basis Function", voir [ZAGORCHEV et GOSHTASBY \[2006\]](#)). La valeur à un point d'interpolation \mathbf{x} est fonction de sa distance aux N échantillons connus \mathbf{p}_i , à travers l'équation suivante :

$$f(\mathbf{t}) = \sum_{i=1}^N d_i \phi(\|\mathbf{t} - \mathbf{p}_i\|) \quad (2.23)$$

L'estimation du champ de déformation consiste à estimer les poids d_i . La fonction $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ est nommée fonction de base de la RBF. Contrairement aux splines qui doivent utiliser un produit tensoriel pour être généralisées à n dimensions, les RBFs sont naturellement conçues pour être utilisées dans un espace à plusieurs dimensions, grâce au choix de la norme utilisée. De la même manière que les B-splines, nous pouvons facilement étendre les RBF afin de pouvoir estimer un champ vectoriel. Dans ce cas, l'équation 2.23 devient :

$$\mathbf{f}(t) = \sum_{i=1}^N \mathbf{d}_i \phi(\|t - \mathbf{p}_i\|) \quad (2.24)$$

où $\mathbf{d}_i \in \mathbb{R}^3$.

Thin Plate Spline Les Thin Plate Splines (TPS) sont l'application la plus courante des RBFs pour la mise en correspondance (voir BOOKSTEIN [1989, 1991]). Elles sont une extension naturelle du cas bivarié des B-splines cubiques [WAHBA, 1990]. Cela provient du fait que ces deux modèles minimisent le problème variationnel suivant :

$$\min_{f: \mathbb{R}^n \rightarrow \mathbb{R}} B[f] \quad (2.25)$$

où $n = 1$ dans le cadre des b-splines cubique et $n \geq 2$ dans le cadre des TPS. La fonctionnelle B correspond à l'énergie de courbure. pour $n = 2$ elle est définie comme :

$$B[f] = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 dx dy \quad (2.26)$$

Dans le cadre des TPS, cette minimisation est effectuée avec une RBF ayant la fonction de base suivante :

$$\phi(r) = \frac{r^2}{\sigma} \ln \left(\frac{r}{\sigma} \right) \quad (2.27)$$

Fonction de base gaussienne L'utilisation de fonctions de base gaussiennes, aussi appelée noyaux gaussiens (voir ARAD et collab. [1994]) sont une autre application des RBFs. Dans ce cadre, la fonction de base ϕ est définie de la manière suivante :

$$\phi(r) = e^{-(\sigma r)^2} \quad (2.28)$$

Voir figure 2.10.

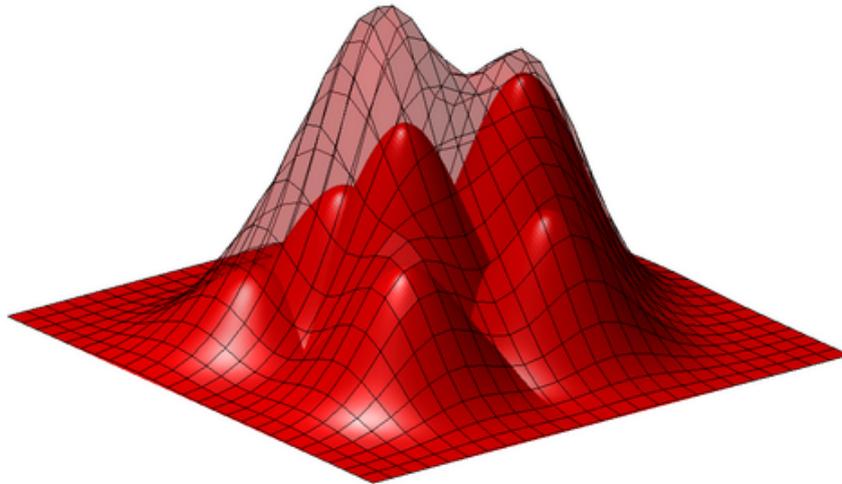


FIGURE 2.10 – Exemple d’approximation par fonctions de base radiales gaussienne. La surface pleine représente les fonctions de bases radiales et la surface transparente représente la fonction approximée. Source : http://www.it.uu.se/research/scientific_computing/project/rbf

2.5 Critères de qualité de mise en correspondance

Nous pouvons distinguer deux types d’approches de mise en correspondance d’images selon la manière dont elles exploitent l’information issue des images pour guider le processus de recalage, les critères dits “géométriques”, utilisant des points de repères et les critères “basés voxels”.

2.5.1 Critères basés voxels

Méthodes basées ”intensité“

Le choix d’une méthode parmi celles basées sur l’intensité des voxels dépend principalement des a priori sur les acquisitions.

- Si les deux images partagent les mêmes structures anatomiques et que celles-ci correspondent aux mêmes valeurs d’intensités, la somme du carré des différences (norme L_2) et la somme de la valeur absolue de la différence (norme L_1) peuvent être utilisées [ULYSSES et CONCI, 2010].
- Si les deux images n’ont pas les mêmes amplitudes d’intensités, mais qu’il existe une relation affine, le critère optimal est la corrélation croisée [ROCHE et collab., 1999].

Méthodes basées ”théorie de l’information“

Dans le cadre d’une mise en correspondance multimodale, ou d’images, n’ayant pas de relation affine entre leur valeur d’intensité, les mesures précédentes ne peuvent pas être utilisées. C’est cette conclusion qui a poussé la création de critères de similarité de plus haut niveau. La minimisation de l’information mutuelle, aussi appelée entropie jointe, est une des approches les plus utilisées en recalage [MAES et collab., 1997; WELLS et collab., 1996]. La mesure d’information mutuelle $I(A,B)$ entre une image A et une image B est définie par :

$$I(A,B) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)p(b)} \quad (2.29)$$

où $p(a)$ et $p(b)$ sont les distributions des valeurs d’intensités respectivement des images A et B, $p(a,b)$ est la distribution jointe.

2.5.2 Critères géométriques

Les méthodes dites géométriques utilisent des mises en correspondance entre des points de repère pour guider le processus de recalage. Les points de repère sont des endroits caractéristiques de l'image par le fait qu'ils ont une signification anatomique ou une signification mathématique (extrema de champs issus de l'image).

Avec l'utilisation de points de repère, une partie de la complexité de la tâche de mise en correspondance est transférée à l'extraction des points. Effectivement, une fois les points de repère extraits et mis en correspondance, la tâche d'estimation du champ de déformation est en partie simplifiée par le fait qu'elle traite directement une information géométrique.

La mise en correspondance par critères géométriques consiste généralement en 3 étapes :

1. extraction de points caractéristiques,
2. appariements des points provenant des deux images,
3. estimation d'une transformation à partir de l'ensemble des appariements.

Extraction des points caractéristiques

Une des tâches inhérentes à l'utilisation des points de repère pour le recalage d'images est l'extraction des points caractéristiques [BERG et collab., 2005; MIKOLAJCZYK et SCHMID, 2005]. La section 2.8 décrit les méthodes d'extraction dans le cadre des points d'intérêt.

Appariement des points caractéristiques

L'appariement des points caractéristiques peut être effectué en considérant les points soit indépendamment, soit tel un nuage.

Les points d'intérêt, décrits dans la section 2.8 associent à chaque point un vecteur de caractéristiques permettant de mesurer la similarité entre deux points.

Des méthodes utilisent des contraintes de plus haut niveau pour permettre des relations plus riches que des relations de paires. Afin d'utiliser les nuages de points de manière unifiée, elles modélisent chaque nuage de points comme un graphe et proposent des méthodes de mise en correspondance de graphes. Nous pouvons citer par exemple la méthode de "mise en correspondance spectrale" par DUCHENNE et collab. [2011] et "Dual décomposition based methods" par KOMODAKIS et collab. [2011].

Estimation de la transformation

L'analyse procrustéenne est une méthode populaire de recalage d'images par points caractéristiques appariés (voir COOTES et collab. [2010]; GOODALL [1991]). Cette méthode d'analyse de forme consiste à minimiser la distance des moindres carrés entre des points de repères, généralement pour estimer une transformation globale. RANSAC (RANdom Sample Consensus) FISCHLER et BOLLES [1981] est une autre approche permettant de gérer une mise en correspondance imparfaite.

Estimation conjointes : appariement et transformation

Certaines méthodes se proposent d'utiliser une estimation conjointe, à la fois de l'appariement des points caractéristiques et de la transformation. Ceci est généralement effectué itérativement, en estimant alternativement l'estimation et l'appariement.

Iterative Closest Point (ICP) proposé par [BESL et MCKAY \[1992\]](#), utilise ce concept d'alternance. La correspondance consiste à faire correspondre le voisin le plus proche, au sens géométrique. À partir de cette estimation, une étape d'estimation de la transformation est calculée en minimisant les distances issues de cette estimation de correspondance de points. Le processus est répété itérativement jusqu'à convergence. [RUSINKIEWICZ et LEVOY \[2001\]](#) propose une synthèse des différentes variantes de l'algorithme.

Méthodes sans mise en correspondance

D'autres méthodes ne faisant pas intervenir de mise en correspondance explicite utilisent une autre manière de représenter l'information géométrique. La principale méthode consiste à représenter la distribution des points de repère comme une distribution de probabilité et de minimiser la distance entre les distributions de l'image source et de l'image cible. Dans ce cadre, les modèles à mélanges Gaussiens (Gaussian Mixture Models) sont utilisés [JIAN et VEMURI \[2011\]](#). Nous pouvons aussi citer les méthodes basés sur les mesures et les courants [[DURRLEMAN et collab., 2008](#); [VAILLANT et GLAUNÈS, 2005](#)].

2.6 Optimisation

En fonction du modèle de déformation et du critère de qualité de mise en correspondance choisis, il est nécessaire d'utiliser des algorithmes d'optimisation afin d'ajuster les paramètres du modèle de déformation au critère de qualité. Nous allons voir ici les différentes méthodes utilisées dans le cadre de la mise en correspondance d'images.

Nous considérerons l'optimisation comme une procédure permettant d'atteindre un jeu de paramètres \mathbf{x} optimal, notée $\hat{\mathbf{x}}$. Elle est utilisée quand aucune forme explicite du problème à résoudre n'existe, ou que celle-ci n'est pas souhaitable (empreinte mémoire, temps de calcul, robustesse), ce qui est généralement le cas du recalage d'images.

L'optimisation continue s'applique à des problèmes dont les variables prennent des valeurs réelles et dont la fonction objectif est différentiable. Certaines approches de recalage d'images peuvent entrer dans le cadre de ces problèmes. C'est une méthode itérative, où chaque itération a pour but de s'approcher de l'optimal, en se déplaçant vers un nouveau point dans l'espace des paramètres $\{\mathbf{x}\}$. C'est la règle de mise à jour qui va définir le type d'approche utilisée.

L'optimisation continue se propose d'utiliser une règle de mise à jour de la solution de la forme :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}(\mathbf{x}^{(k)}) \quad (2.30)$$

où $\mathbf{x}^{(k)}$ est le vecteur de paramètres de la transformation à l'itération k . \mathbf{d} définit le déplacement du vecteur de paramètres dans l'espace de recherche. La notation complète de \mathbf{d} devrait être $\mathbf{d}(\mathcal{M}(\mathbf{x}_k) + \mathcal{R}(\mathbf{x}_k))$ dans le cas de la mise en correspondance d'images, que nous simplifierons pour des raisons de notations en écrivant $\mathbf{d}(\mathbf{x})$ et en définissant $f(\mathbf{x}) = \mathcal{M}(\mathbf{x}_k) + \mathcal{R}(\mathbf{x}_k)$.

2.6.1 Descente de gradient

La méthode de descente de gradient, aussi appelée algorithme du gradient, est une méthode qui optimise la fonction objectif en suivant la direction de la pente la plus forte, ou dit autrement, de gradient négatif le plus élevé (figure 2.31). Elle était utilisée par Cauchy pour résoudre des systèmes linéaires.

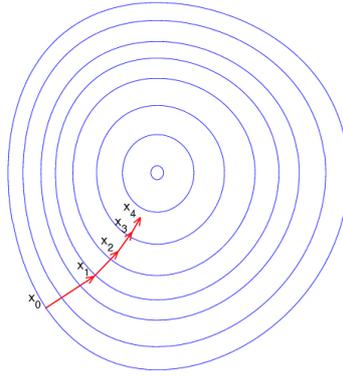


FIGURE 2.11 – Exemple de descente de gradient sur une fonction 2D de type “bol” avec une vue en plan avec courbes de niveaux. Source : https://commons.wikimedia.org/wiki/File:Gradient_descent.png

Dans ce cadre, la fonction \mathbf{d} est définie par l'équation suivante :

$$\mathbf{d}^{(k)}(\mathbf{x}) = -\alpha^{(k)}\nabla f(\mathbf{x}) \quad (2.31)$$

Plusieurs stratégies ont été étudiées (voir KLEIN et collab. [2007]) en ce qui concerne le facteur α :

- Constant : α est fixe pour l'ensemble des itérations.
- Décroissant : α est multiplié par un facteur inférieur à l'unité à chaque itération, la suite des $\{\alpha^{(0)}, \alpha^{(1)}, \dots, \alpha^{(n-1)}, \alpha^{(n)}\}$ est une suite géométrique de raison inférieure à 1.
- Issu d'une optimisation : α est calculé par minimisation de la fonction objectif le long de l'axe du gradient, par recherche linéaire (voir algorithme 1).

Algorithme 1 : Descente de gradient avec recherche linéaire

Entrée : $f : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonctionnelle à minimiser

$\mathbf{x}^{(0)}$ une solution initiale

Sortie : $\hat{\mathbf{x}}$ un minimum (local) de la fonction f

- 1 $k \leftarrow 0$
 - 2 **tant que** critère d'arrêt **faire**
 - 3 $\alpha^{(k)} \leftarrow \operatorname{argmin}_{\alpha} f(\mathbf{x}^{(k)} - \alpha\nabla f(\mathbf{x}_t))$
 - 4 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \alpha^{(k)}\nabla f(\mathbf{x}_t)$
 - 5 **fin**
-

2.6.2 Méthode du gradient conjugué

La méthode du gradient conjugué est aussi une méthode du premier ordre. Elle se propose d'utiliser la connaissance des itérations précédentes et d'ajouter la contrainte que la direction de descente $\mathbf{d}^{(k)}$ soit conjuguée avec la précédente direction d'optimisation $\mathbf{d}^{(k-1)}$. La nouvelle direction est donc définie par

$$\mathbf{d}^{(k)}(\mathbf{x}) = f\left(\nabla f(\mathbf{x}), \mathbf{d}^{(k-1)}\right) \quad (2.32)$$

où f est généralement définie comme la combinaison linéaire de ces deux paramètres :

$$\mathbf{d}^{(k)}(\mathbf{x}) = \nabla f(\mathbf{x}) + \beta^{(k)}\mathbf{d}^{(k-1)} \quad (2.33)$$

Plusieurs approches existent pour définir le facteur $\beta^{(k)}$. Une synthèse des différentes méthodes de descente de gradient conjugué est proposée par HAGER et ZHANG [2006].

2.6.3 Méthode de Newton

La méthode de Newton consiste à exploiter l'information de second ordre du modèle, dans notre cas, la dérivée seconde de la fonction objectif. En plus d'utiliser le vecteur de gradient $\nabla f(\mathbf{x})$, elle exploite la matrice hessienne $\mathbf{H}(f)$ du problème. Les éléments de la matrice hessienne se définissent comme suit :

$$h_{i,j} = \frac{\partial^2 f(x^{(k)})}{\partial x_i^{(k)} \partial x_j^{(k)}} \quad (2.34)$$

où $x_i^{(k)}$ et $x_j^{(k)}$ sont respectivement les i -ème et j -ème composantes du vecteur de paramètres \mathbf{x} à l'itération k .

La méthode de Newton se propose de faire l'approximation à l'ordre 2 de la fonction objectif, et de se déplacer à son minimum. La fonction de mise à jour est donc de la forme suivante :

$$\mathbf{d}^{(k)}(\mathbf{x}^{(k)}) = -\alpha \mathbf{H}^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (2.35)$$

Cette équation donnant le minimum du développement de Taylor à l'ordre deux de la fonction objectif, $\alpha = 1$ est généralement choisi, pour se déplacer directement à ce minimum, mais des valeurs inférieures peuvent être utilisées pour des questions de stabilité, voir algorithme 2.

La méthode de Newton est une méthode qui nécessite peu d'itérations, sa convergence asymptotique est quadratique. Néanmoins, le coût prohibitif de calcul de la hessienne et son instabilité fait qu'elle est rarement utilisée telle quelle.

Algorithme 2 : Algorithme de Newton

Entrée : $f : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonctionnelle à minimiser

$\mathbf{x}^{(0)}$ une solution initiale

Sortie : $\hat{\mathbf{x}}$ un minimum (local) de la fonction f

1 $k \leftarrow 0$

2 **tant que** critère d'arrêt **faire**

3 $\mathbf{d}^{(k)} \leftarrow -\mathbf{H}^{-1} \nabla f(\mathbf{x}^{(k)})$

4 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$

5 **fin**

6 $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(k+1)}$

2.6.4 Méthode “quasi-Newton”

Le coût de calcul de la méthode de Newton étant prohibitif, plusieurs algorithmes utilisent des approximations de la Hessienne en vue d'approcher une convergence quadratique sans le coût de calcul explicite de la hessienne (voir PRESS et collab. [1982], en la remplaçant par une matrice \mathbf{B}_k plus facile à calculer.

La principale approche utilisant une approximation de la hessienne est l'algorithme dit Broyden-Fletcher-Goldfarb-Shanno (BFGS, SHANNO [1970]).

2.6.5 Méthode de Gauss-Newton

La méthode de Gauss-Newton [WEDDERBURN, 1974] se limite aux fonctions objectifs s'écrivant comme une somme de termes au carré. C'est donc une méthode de résolution de problèmes de moindres carrés non linéaires. Ceci est particulièrement intéressant car c'est généralement sous cette forme que se présentent les fonctions objectifs liées à la mise en correspondance d'images (voir ASHBURNER et FRISTON [2011]) :

$$f(\mathbf{x}) = \sum_{i=1}^m (f_i(\mathbf{x}))^2 \quad (2.36)$$

La hessienne est approximée en ignorant les dérivées supérieures à l'ordre 1 avec $\hat{\mathbf{H}} = 2\mathbf{J}^T\mathbf{J}$ où \mathbf{J} est la jacobienne de la fonction objectif. La direction de descente est de la forme :

$$\mathbf{d}^{(k)} = - \left(\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)}) \right)^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (2.37)$$

où \mathbf{J} est la jacobienne de la fonction objectif, ces éléments étant définis par :

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial \mathbf{x}_1} & \dots & \frac{\partial f_1}{\partial \mathbf{x}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial \mathbf{x}_1} & \dots & \frac{\partial f_m}{\partial \mathbf{x}_n} \end{pmatrix} \quad (2.38)$$

2.6.6 Méthode de Levenberg-Marquardt

L'algorithme de Levenberg-Marquardt [MORÉ, 1978] fait partie de la famille des algorithmes dits des méthodes de Newton amorties. Les méthodes de Newton amorties sont la classe des méthodes pondérant l'influence de la Hessienne pour pallier le manque de stabilité lié à son inversion, elles sont de la forme :

$$\mathbf{d}^{(k)} = - (\mathbf{H} + \lambda \mathbf{W})^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (2.39)$$

où \mathbf{W} est une matrice de pondération définie positive et λ un facteur de pondération positif. Lorsque l'on prend l'identité comme matrice de pondération $\mathbf{W} = \mathbf{I}$, la valeur de λ fait varier continument l'algorithme entre une descente de gradient ($\lambda \rightarrow \infty$) et la méthode de Newton ($\lambda = 0$).

L'algorithme de Levenberg-Marquardt propose une stratégie de gestion de λ afin d'accélérer la vitesse de convergence en essayant de garantir la stabilité (voir KYBIC et UNSER [2003] pour une application de recalage médical).

2.6.7 Méthode du gradient stochastique

Les approches précédentes couvrent des méthodes déterministes. Pour des raisons de stabilité ou de temps de calcul, des méthodes probabilistes et stochastiques ont vu le jour.

Le terme "Gradient Stochastique" désigne deux types d'algorithmes. Le premier se propose d'utiliser comme règle de mise à jour du vecteur de paramètres l'équation suivante :

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \hat{\mathbf{g}}_t(\mathbf{x}_t) \quad (2.40)$$

où $\hat{\mathbf{g}}_t$ est une fonction stochastique ou probabiliste. KLEIN et collab. [2007] effectuent une comparaison de trois approches issues de l'approximation stochastique du gradient.

Le deuxième type de descente de gradient stochastique consiste à n'utiliser qu'une fraction des observations à chaque instant [SUTSKEVER et collab. \[2013\]](#). Dans le cadre de la mise en correspondance d'images médicales, le calcul même du gradient peut avoir un coût prohibitif, sachant que les données et l'espace de recherche ont une grande dimension. Lorsque la fonction objectif s'écrit comme la somme de plusieurs éléments, le gradient s'écrit comme la somme du gradient de chacun des éléments. L'approche consiste donc à tirer un échantillon aléatoire de l'ensemble des éléments, et à calculer le gradient partiel à chaque itération :

$$\hat{\nabla}_k f(\mathbf{x}^{(k)}) = \nabla f_i(\mathbf{x}^{(k)}) \quad (2.41)$$

où la distribution des tirages i est uniforme.

Il existe plusieurs méthodes de descente de gradient stochastique, la plus connue consiste à ne tirer qu'un seul échantillon à chaque itération. Cela a pour conséquence d'avoir des itérations possédant des directions d'optimisation éloignées de la direction menant à l'extremum recherché, mais permet d'être très rapide. Le grand nombre d'itérations contrebalance la faible précision.

Une méthode, dite par mini-lots, consiste à prendre plus d'un échantillon pour calculer le gradient sur une fraction de l'ensemble des échantillons. Dans ce cas-ci, il faut arriver à trouver le compromis entre coût du calcul et qualité du gradient :

$$\hat{\nabla}_k f(\mathbf{x}^{(k)}) = \sum_{i \in \{J_k\}} \nabla f_i(\mathbf{x}^{(k)}) \quad (2.42)$$

Différentes méthodes utilisent le ou les gradients précédemment calculés pour améliorer la direction de l'itération courante. Nous pouvons citer par exemple, la méthode du "momentum" qui propose, à travers une analogie avec la notion d'inertie en physique, de réinjecter une fraction du gradient précédant dans le gradient courant. Cette formulation conduit à une moyenne exponentielle :

$$\hat{\nabla}_k f(\mathbf{x}^{(k)}) = \nabla f_i(\mathbf{x}^{(k)}) + \eta \hat{\nabla}_{k-1} f(\mathbf{x}^{(k-1)}) \quad (2.43)$$

où le coefficient η permet de régler la longueur de la mémoire du moyennage ($0 < \eta < 1$).

2.6.8 Recherche linéaire

La recherche linéaire est un cas particulier de l'optimisation continue [GOULD et LEYFFER \[2003\]](#). Elle est souvent utilisée comme sous-algorithme dans les approches précédentes. Elle permet de trouver un minimum dans la direction d'optimisation donnée.

La méthode peut être présentée de manière duale aux méthodes précédentes, où le choix de la direction d'optimisation est générique, voir l'algorithme 3.

Algorithme 3 : Algorithme de recherche linéaire

Entrée : $f : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonctionnelle à minimiser
 $\mathbf{x}^{(0)}$ une solution initiale
Sortie : $\hat{\mathbf{x}}$ un minimum (local) de la fonction f

- 1 $k \leftarrow 0$
- 2 **tant que** CRITERE D'ARRET **faire**
- 3 Calculer \mathbf{p}_k une direction de descente
- 4 $\alpha_k = \min_{\alpha} (f(\mathbf{x}_k + \alpha \mathbf{p}_k))$
- 5 **fin**
- 6 $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(k+1)}$

2.6.9 Méthode des moindres carrés pondérés itératifs

La méthode des moindres carrés pondérés itératifs (IRLS : “Iteratively Reweight Least Square”, [HOLLAND et WELSCH \[1977\]](#)) est utilisée pour résoudre les problèmes d’optimisations ayant pour fonction objectif :

$$\arg \min_{\mathbf{x}} \sum_{i=1}^n \kappa(f_i(\mathbf{x})) \quad (2.44)$$

où $\kappa : \mathbb{R} \rightarrow \mathbb{R}$.

Cette minimisation est faite de manière itérative en résolvant le problème des moindres carrés pondérés de la forme :

$$\mathbf{x}^{(k)} = \arg \min_{\mathbf{x}} \sum_{i=1}^n w_i(\mathbf{x}^{(k-1)}) \left(f_i(\mathbf{x}^{(k-1)}) \right)^2 \quad (2.45)$$

IRLS peut être par exemple être utilisé dans le cadre de la minimisation en norme L_1 ($\kappa = \|\cdot\|_1$), lorsque $f_i(\mathbf{x}) = y_i - \mathbf{A}\mathbf{x}_i$ en prenant comme poids $w_i(\cdot)$:

$$w_i(\mathbf{x}^{(k-1)}) = \frac{1}{|y_i - \mathbf{A}\mathbf{x}_i|} \quad (2.46)$$

La méthode des moindres carrés pondérés itératifs peut aussi être utilisée en statistique robuste (voir [Partie 2.8.2](#)).

2.6.10 Critères d’arrêt

L’un des aspects important des processus itératifs est la formalisation d’un critère d’arrêt. Afin qu’un algorithme puisse être défini et utilisé convenablement, il est nécessaire de définir un critère d’arrêt qui permet l’obtention d’un bon compromis entre précision et temps d’exécution. Nous présentons ici les critères les plus classiques, sachant que le critère d’arrêt d’un algorithme peut être une combinaison logique de ces critères :

- $k < k_{max}$: Critère le plus simple, où on arrête d’itérer après un nombre k_{max} d’itération. Il est généralement combiné avec d’autres critères, permettant de garantir l’exécution en temps fini.
- $\|\mathbf{x}^{i+1} - \mathbf{x}^i\| < \varepsilon$ où par exemple $\varepsilon < 10^{-4}$. Critère permettant l’arrêt de l’algorithme lorsque le changement dans la solution devient très petit. Il ne prend pas en compte l’échelle de la solution, donc on lui préférera souvent le critère suivant.

- $\frac{\|\mathbf{x}^{i+1} - \mathbf{x}^i\|}{\|\mathbf{x}^i\|} < \varepsilon$. Critère le plus utilisé, permettant de s'arrêter lorsque le pas relatif passe sous un certain seuil.

2.6.11 Méthode "RANSAC"

RANSAC est l'abréviation de RANdom SAmple Consensus. C'est une méthode permettant d'estimer les paramètres d'un modèle mathématique en étant robuste à la présence de données aberrantes ("outliers", par opposition aux "inliers").

C'est une méthode itérative non déterministe dans le sens où il n'y a pas de garantie d'optimalité du résultat, mais seulement une certaine probabilité d'optimalité, augmentant avec le nombre d'itérations.

Cette procédure a été proposée par FISCHLER et BOLLES [1981].

Procédure Afin d'utiliser RANSAC, il faut posséder un jeu de données, auquel on essaie d'ajuster un modèle paramétrique ainsi que des intervalles de confiance.

Une itération de la procédure est définie comme suit :

1. Un sous-ensemble minimal des données, pour déterminer un modèle, est tiré au hasard ;
2. Les paramètres du modèle sont ajustés à ce sous-ensemble, par une méthode classique ;
3. L'ensemble des données est confronté à ce modèle, en mesurant la distance des points au modèle. Les données ayant une distance inférieure à l'intervalle de confiance sont considérées comme inliers, les autres comme outliers ;
4. Le modèle choisi est noté avec le nombre d'inliers qu'il possède.

Après un nombre d'itérations fixé à l'avance, le modèle ayant le meilleur taux d'inliers est choisi.

Algorithme 4 : Procédure RANSAC

Entrée :

$P = \{(x, y)\}$: l'ensemble des observations

$f := (\theta, x) \rightarrow y$: Un modèle qui peut être ajusté aux données

Sortie :

$\hat{\theta}$: Un modèle estimé

```

1  $nBest = 0$ 
2 tant que  $i < NOMBRE\_ITERATION$  faire
3    $P_i = TirageSousEnsembleAleatoire(P)$ 
4    $\theta_i = \min_{\theta} \sum_{(x,y) \in P_i} (\|f(\theta, x) - y\|)$ 
5    $nInliers = CalculInliers(\theta_i, P_i)$ 
6   si  $nInliers > nBest$  alors
7      $\hat{\theta} = \theta_i$ 
8      $nBest = nInliers$ 
9   fin
10 fin

```

Variantes Plusieurs variantes sont possibles à partir de cette structure de base :

- **Arrêt prématuré** : si un modèle suffisamment bon est trouvé, c'est-à-dire possédant un taux d'inliers suffisamment élevé, l'algorithme s'arrête et cette estimation des paramètres du modèle est utilisée.

- **Recalcul du modèle** : une étape de raffinement peut avoir lieu après calcul des “inliers”. Dans ce cas, le modèle est recalculé à partir de l'ensemble des points considérés comme “inliers”. Cela permet une plus grande robustesse au bruit, mais au détriment d'un temps de calcul plus élevé.
- **Sous-ensemble non minimal** : classiquement, le nombre de points tirés est le nombre minimal nécessaire à l'estimation du modèle. Un nombre plus grand peut être choisi, ce qui a des conséquences sur la robustesse au bruit, ainsi que sur l'exploration de l'ensemble des modèles possibles.

2.7 Recalage d'ensemble

2.7.1 Recalage de paires et recalage d'ensembles

La mise en correspondance d'images fait généralement référence au recalage entre deux images, suivant le formalisme défini précédemment. Néanmoins, il existe des approches se proposant de mettre en correspondance plus de deux images. Ces méthodes portent plusieurs noms selon leurs domaines d'application :

- “Mise en correspondance pour les études de cohortes” (Cohort analysis, exemple [HILL et collab. \[2001\]](#))
- “Mise en correspondance de groupe” (Groupwise registration, exemple [WU et collab. \[2012\]](#))
- “Mise en correspondance de population” (Population Registration, exemple [ZÖLLEI et collab. \[2005\]](#))

2.7.2 Applications médicales

Les dernières avancées en apprentissage automatisé [MICHALSKI et collab. \[2013\]](#) et en segmentation basée atlas [IGLESIAS et SABUNCU \[2015\]](#), ont un besoin croissant d'outils d'analyse de grandes quantités d'images. Actuellement, ces approches sont limitées par l'hétérogénéité et la petite taille des jeux de données disponibles. Le recalage conjoint de grands jeux de données peut permettre de dépasser cette limitation et d'ouvrir la voie à de nouvelles méthodes basées sur des masses de données.

Le recalage d'ensembles d'images de manière conjointe est aujourd'hui utilisé essentiellement pour l'étude des cerveaux humains. Les applications de recalage de corps entiers restent rares et de nombreux défis y sont encore associés [XU et collab. \[2016\]](#).

2.7.3 Représentation par graphe

Afin de pouvoir traiter les différents formalismes issus de la mise en correspondances de groupes d'images, nous allons traiter les différentes approches en utilisant la théorie des graphes.

Le graphe se définit par :

- **Les sommets**, ils représentent les images.
- **Les arêtes**, elles représentent les liens entre les images. Par lien nous faisons référence à la fois au critère de similarité liant deux images, et à la fois à la transformation nécessaire au recalage.

Recalage exhaustif (graphe complet)

La première approche que l'on pourrait qualifier de recalage de groupe a été proposée par [THOMPSON et collab. \[1997\]](#) pour la détection de structures anatomiques anormales dans des atlas d'images. Cette toute première méthode ne fera pas partie ensuite de ce que nous appellerons recalage de groupe, car elle ne fait qu'itérer des algorithmes de recalage de paires, sans lien entre eux. Nous pouvons aussi difficilement parler de recalage de groupe car si on considère l'utilisation du groupe d'images en lui-même, nous

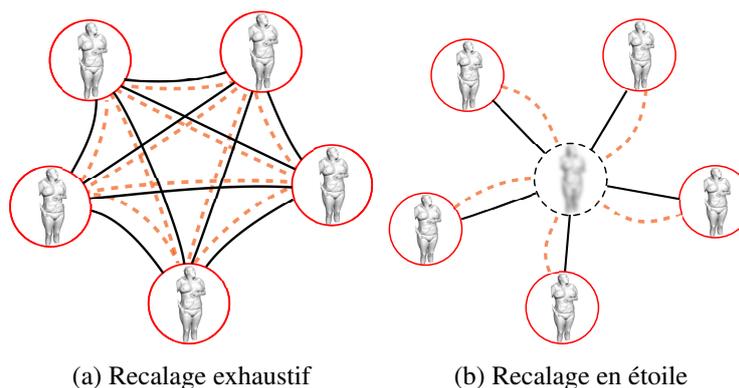


FIGURE 2.12 – Le recalage exhaustif et le recalage en étoile. Les observations et les modèles de déformations sont représentés respectivement par des traits noirs et des traits orange pointillés.

pouvons observer sur la figure 2.12a que le chemin entre deux images n'est pas unique. Cette approche se propose de prendre l'arc direct, mais nous verrons ensuite que le choix d'un "bon" chemin est garant de la robustesse de l'algorithme vis-à-vis de la variabilité.

Le principal intérêt de cette méthode était de fournir une transformation pour n'importe quelle paire d'images. Son principal désavantage est son coût de calcul prohibitif.

Recalage en étoile

Le recalage en étoile 2.12b consiste à créer une image de référence commune et à ensuite utiliser des algorithmes classiques de recalage de paires entre les images du groupe et cette référence, voir algorithme 5.

Les premières approches utilisent une image de référence sélectionnée manuellement par l'utilisateur, ce qui la rend interactive. De plus, sélectionner un individu du groupe comme référence unique introduit un biais statistique dans toutes les études utilisant ce recalage de groupe d'images en prétraitement. C'est principalement pour s'affranchir de biais que les méthodes de groupes ("groupwise") ont été introduites.

Elles consistent à construire itérativement la moyenne de toute la population, puis à utiliser un algorithme classique de recalage de paires entre cette moyenne et chaque image. Ces approches ont, entre autres, été proposées par JOSHI et collab. [2004]; NOBLET et collab. [2012] qui utilisent une moyenne de Fréchet pour construire l'image moyenne.

Du point de vue de la modélisation, ces méthodes peuvent être représentées par un graphe en étoile, avec en son centre l'image de référence. Le résultat est un ensemble de demi-transformations τ_i faisant la correspondance entre chaque image i et l'espace commun contenant la référence. Après recalage, la transformation $\tau_{a \rightarrow b}$ de l'image a vers b peut être calculée en composant deux demi-transformations τ_a et τ_b : $\tau_{a \rightarrow b} = \tau_b^{-1} \circ \tau_a$, voir figure 2.13.

Malgré l'avantage d'être non biaisé, ce type d'approche possède plusieurs inconvénients :

- **Amplitude de la déformation** : L'amplitude de la déformation est limitée par deux aspects. Le premier aspect est la limite en amplitude de l'algorithme de recalage par paires sous-jacent. La deuxième limite est inhérente à la méthode. Étant donné qu'une moyenne est construite, si on la construit à partir d'échantillons trop différents ou mal recalés, la référence résultante sera inutilisable car très floue.

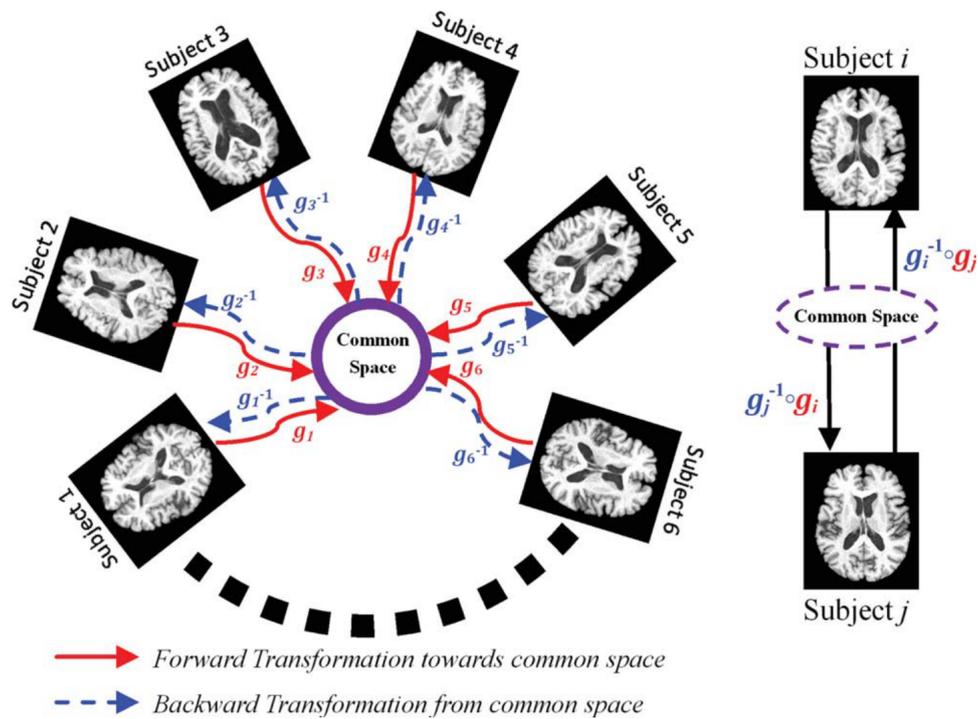


FIGURE 2.13 – Recalage en étoile. À gauche, schéma global : nous pouvons noter les demi-transformations entre les images et la référence. À droite, l'application des demi-transformations pour passer d'une image à l'autre. Crédit : WU et collab. [2012]

Algorithme 5 : Recalage en étoile avec référence par moyenne

Entrée : $\mathcal{I} = \{I_i\}$: un ensemble de n images

Sortie : $\{\tau_i\}$: un ensemble de n demi-transformations

```

1 pour chaque  $I_i \in \mathcal{I}$  faire
2   |  $\hat{I}_i = I_i$ 
3 fin
4 tant que CRITERE D'ARRET faire
5   |  $I_{mean} = \text{mean}(\hat{I}_i)$ 
6   | pour chaque  $I_i \in \mathcal{I}$  faire
7     |  $\tau_i = \text{RecalageParPaire}(I_i, I_{mean})$ 
8     |  $\hat{I}_i = \tau_i \circ I_i$ 
9   | fin
10 fin
    
```

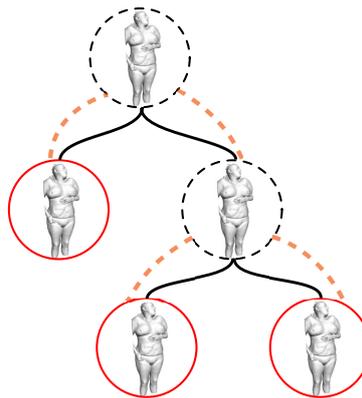


FIGURE 2.14 – Recalage par arbre couvrant minimal

- **Nombre d'images** : Au delà de la moyenne floue due au moyennage d'images différentes, l'utilisation d'un grand nombre d'images résulte aussi en une moyenne floue. Effectivement, moyennner des images résulte en un filtrage passe-bas, qui érode les arêtes et les zones de fort contraste. Or, la plupart des algorithmes par paires utilisent indirectement ces zones de fort contraste dans leur optimisation.
- **Temps de calcul** : Les algorithmes par paires sont déjà des algorithmes coûteux en temps de calculs, l'utilisation de manière itérée de ces algorithmes résulte en des approches encore plus coûteuses.

Recalage en arbre couvrant minimal

Comme mentionné précédemment, une limitation du recalage de groupe en étoile est l'utilisation du recalage de paires d'images et la dernière estimation de la moyenne du groupe. Dans le cadre d'utilisation de grandes bases d'images, l'existence d'importantes différences entre images produit une moyenne de mauvaise qualité, qui induit de mauvais recalages par paires. Ce cercle vicieux est problématique et le recalage dit en "arbre" propose une solution.

La mise en correspondance "en arbre", consiste à créer un arbre couvrant minimal pour lier toutes les images (voir figure 2.14). Cette approche a été proposée par [HAMM et collab. \[2009\]](#) et reprise ensuite par [WU et collab. \[2011\]](#). Elle présente l'avantage de ne pas faire porter toute la variabilité par un seul nœud du graphe, mais au contraire d'utiliser des références multiples.

Le protocole de construction permet aussi de minimiser les différences relatives entre deux sommets d'une arête. Transformer ensuite un point d'une image dans le système de coordonnées d'une autre image revient à suivre le chemin, unique, qui relie les deux images.

2.8 Extraction de caractéristiques : les points d'intérêt

Historiquement, avant que les méthodes basées intensités prennent une importance prépondérante, des méthodes dites de points extrémaux [[PENNEC et collab., 2000](#); [THIRION, 1996](#)] proposaient des approches de recalages d'images par points.

Durant les dernières décennies, les points d'intérêt ont été utilisés avec succès dans de nombreuses applications, comme la reconnaissance d'objets [LOWE \[1999\]](#), la robotique [KARLSSON et collab. \[2005\]](#) ou encore la création de panoramas [ANGUELOV et collab. \[2010\]](#).

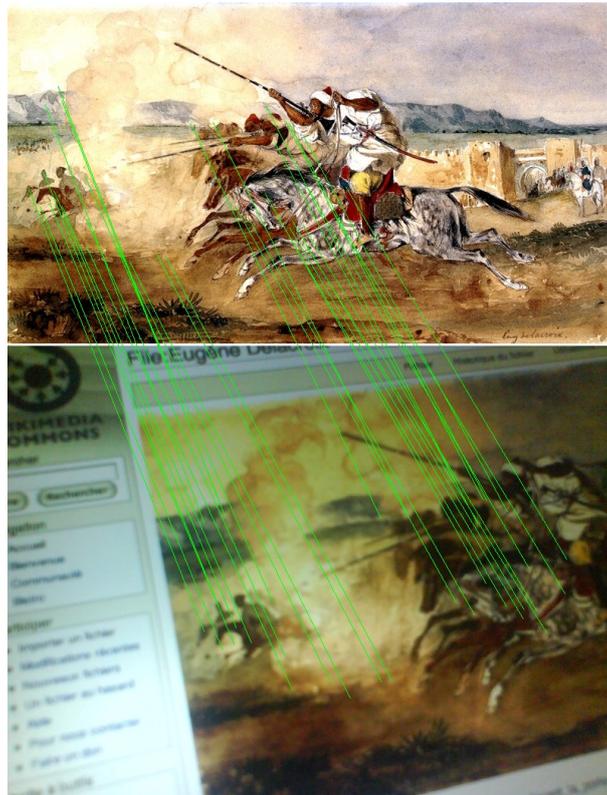


FIGURE 2.15 – Détection et mise en correspondance de points d'intérêt entre deux images, ici avec la description effectué par SIFT. Crédit : [LOWE \[1999\]](#).

L'approche consiste à extraire des points caractéristiques des images de manière à réduire la quantité de données à traiter, principalement pour des tâches temps réel ou des tâches utilisant de grandes quantités de données. Leur utilité dans le cadre de l'imagerie médicale est évaluée dans [LÓPEZ et collab. \[1999\]](#) et de nombreuses applications ont été proposées, comme l'annotation d'images [DATTA et collab. \[2005\]](#) ou encore la segmentation [KÉCHICHIAN et collab. \[2014\]](#).

L'avantage des points d'intérêt pour le recalage d'images est qu'ils fournissent, après avoir été mis en correspondance entre images, une information géométrique de distance. Ceci conduit à un recalage très rapide, contrairement à la mise en correspondance dense [MIKOLAJCZYK et SCHMID \[2005\]](#).

L'extraction de caractéristiques par points d'intérêt comporte deux étapes : la détection des points d'intérêt et la description de ces points par vecteurs de caractéristiques. Nous pouvons noter une troisième étape, celle d'un processus de mise en correspondance des points entre eux, dont un exemple est illustré figure 2.15.

Dans la suite de cette section, nous nous emploierons à décrire SIFT ("scale invariant feature transform", proposé par [LOWE \[1999\]](#)) et SURF ("speeded up robust features", proposé par [BAY et collab. \[2006\]](#)), deux approches combinant un détecteur et un descripteur. De nombreuses autres approches ont été proposées, telles que : BRIEF [CALONDER et collab. \[2010\]](#), ORB [RUBLEE et collab. \[2011\]](#), FREAK [ALAHY et collab. \[2012\]](#).

2.8.1 Détection des points d'intérêt

Les détecteurs de points d'intérêt ont pour but de localiser des points ayant des dérivées (première, seconde, ...) avec des profils particuliers. Il existe principalement deux types de détecteurs :

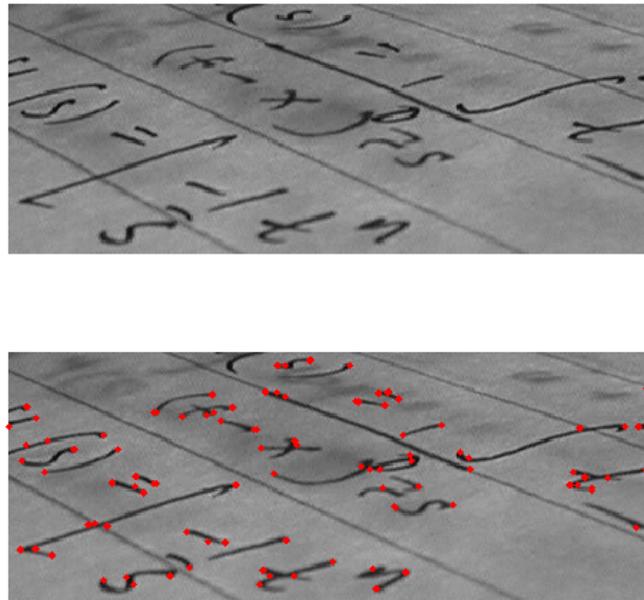


FIGURE 2.16 – Exemple de détection de coins. Source : <http://en.wikipedia.org/wiki/File:Corner.png>

- Les détecteurs de coins extraient les points de l'images étant des extrema du gradient (dérivée première).
- Les détecteurs de "blobs" extraient les points étant des extrema du hessien de l'image (dérivée seconde). Ils se caractérisent par des zones homogènes entourées d'une discontinuité d'intensité de l'image.

Détecteur de Harris

Harris a été le premier à proposer un détecteur de coins [HARRIS et STEPHENS \[1988\]](#). Le détecteur de Harris consiste à construire un tenseur dérivé de l'image 2D de la manière suivante :

$$\mathbf{S} = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (2.47)$$

où I_x et I_y sont les dérivées de l'image respectivement selon les axes x et y . La détection de coins consiste à évaluer les valeurs propres λ_1 et λ_2 de \mathbf{S} .

- Si $\lambda_1 \approx 0$ et $\lambda_2 \approx 0$ le pixel n'a pas de caractéristique d'intérêt.
- Si $\lambda_1 \approx 0$ et λ_2 a une grande valeur, alors c'est une arête.
- Si λ_1 et λ_2 ont de grandes valeurs, alors c'est un coin.

Cette recherche est effectuée dans les deux dimensions de l'espace. Un troisième axe de recherche est utilisé, celui de l'échelle. Il permet de détecter les points d'intérêts de différentes tailles, et d'être ainsi invariant à l'échelle. Cette approche est appelée recherche en espace-échelle ("scale-space", [WITKIN \[1984\]](#)).

La majorité des détecteurs développés ensuite ont pour objectif de trouver des manières plus efficaces d'extraire ces caractéristiques. Nous développerons ici deux détecteurs : SIFT ([LOWE \[1999\]](#)) qui utilise des différences de gaussiennes et SURF ([BAY et collab. \[2006\]](#)) qui utilise des images intégrales.

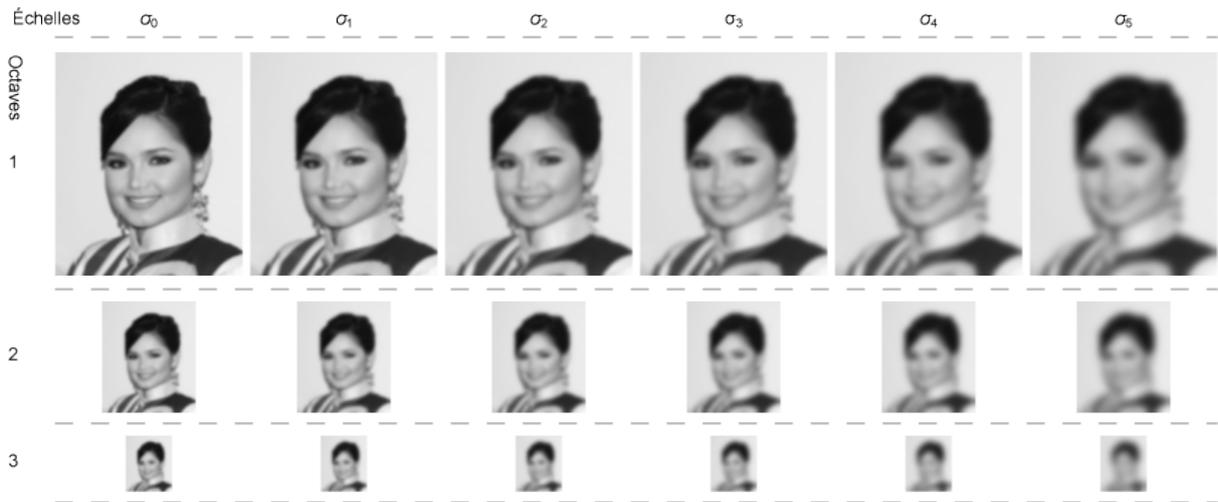


FIGURE 2.17 – SIFT utilise une pyramide de gaussiennes afin d’accélérer le calcul en sous-échantillonnant l’image quand σ le permet sans perte significative d’information. Ici, 3 octaves de 5 gaussiennes. Crédit : [LOWE \[1999\]](#).

Détecteur SIFT

SIFT utilise les extremas de la dérivée seconde de l’image pour la détection. Celle-ci s’effectue dans l’espace-échelle discret. Cet espace comporte trois dimensions : les coordonnées cartésiennes x et y et le facteur d’échelle σ . Une pyramide de gaussienne est visible figure 2.17. Le gradient de facteur d’échelle σ (notée L) est le résultat de la convolution de l’image I par un filtre gaussien G de paramètre σ :

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (2.48)$$

La différence de gaussiennes (Difference of Gaussian : DoG) permet la détection des objets ayant une dimension comprise entre les variances des deux gaussiennes utilisées, le filtre résultant agissant comme un passe-bande :

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (2.49)$$

où k est un paramètre fixe qui règle la bande passante du filtre et qui dépend de la finesse de la discrétisation.

La détection de points d’intérêt par SIFT revient à sélectionner les points qui sont les extrema de $D(x,y,\sigma)$ en espace-échelle.

Détecteur SURF

SURF utilise des images intégrales précalculées pour augmenter la vitesse de détection.

La valeur de chaque pixel de l’image intégrale est la somme de tous les pixels dans le quart de l’espace situé à gauche et au-dessus de lui :

$$S(x,y) = \sum_{i=0}^x \sum_{j=0}^y I(i,j) \quad (2.50)$$

La formulation permet ensuite de calculer la somme d’une zone rectangulaire de taille quelconque de l’image en seulement 4 opérations :

$$\text{SumRect}(x_1,y_1,x_2,y_2) = S(x_2,y_2) - S(x_1,y_2) - S(x_2,y_1) + S(x_1,y_1) \quad (2.51)$$

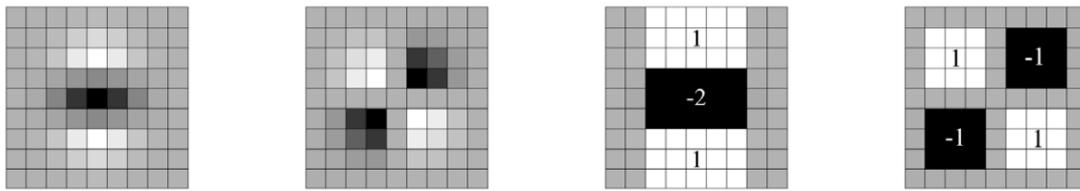


FIGURE 2.18 – De gauche à droite : les dérivées secondes de noyau gaussien (discrétisé et borné) dans les directions y^2 et xy et leurs approximations par bloc. Crédit [BAY et collab.](#) [2006].

L'approche de SURF consiste à approximer par des blocs, les dérivées secondes d'un noyau gaussien pour construire la hessienne de l'image (voir Figure 2.18).

La détection des points d'intérêt consiste à détecter les extrema (en espace-échelle), du déterminant de la hessienne, et de considérer ces points comme caractéristiques.

Invariance en rotation

La plupart des approches produisant des points d'intérêt sont invariantes en rotation. Cette invariance signifie que si une image subit une rotation, le descripteur ne sera pas modifié. Les descripteurs ne sont pas intrinsèquement invariants en rotation. Une solution est d'estimer une rotation, et d'extraire la description des points d'intérêt après avoir appliqué cette rotation. La plupart des méthodes estiment l'axe de plus fort gradient pour construire un nouveau repère et extraire la description dans ce repère.

2.8.2 Description des points d'intérêt

La description emploie un vecteur de caractéristiques possédant une centaine de composantes. Cette description locale a pour objectif de permettre la comparaison des points entre eux et ainsi de créer une mesure de similarité.

Plusieurs propriétés sont recherchées pour la description, et c'est le niveau d'abstraction de celle-ci qui va permettre d'avoir un taux de rappel important ("recall"). Le taux de rappel est la capacité d'un descripteur à réussir à décrire la correspondance entre deux points malgré des modifications des conditions d'acquisition : luminosité, déformation, rotation, point de vue 3D, etc.

Descripteur SIFT

SIFT se propose de construire un vecteur de description basé sur des histogrammes d'orientations.

Chaque région de 16×16 pixels autour du point détecté est subdivisée en 4×4 blocs de 4×4 pixels. Sur chaque bloc est calculé un histogramme des orientations comportant 8 intervalles. En chaque point du bloc, l'orientation et l'amplitude du gradient sont calculées (voir figure 2.19). L'orientation détermine l'intervalle à incrémenter dans l'histogramme. Ceci est fait avec une double pondération : par l'amplitude et par une fenêtre gaussienne centrée sur le point d'intérêt, de paramètre égal à 1,5 fois le facteur d'échelle du point d'intérêt.

Ensuite, les 16 histogrammes à 8 intervalles chacun sont concaténés et normalisés. Dans le but de diminuer la sensibilité du descripteur aux changements de luminosité, les valeurs sont plafonnées à 0,2 et l'histogramme est de nouveau normalisé, pour finalement fournir le descripteur SIFT du point d'intérêt, de dimension 128.

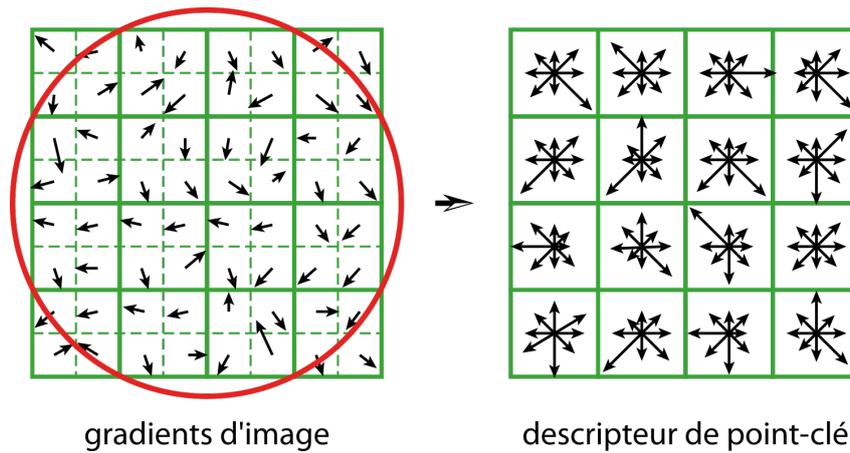


FIGURE 2.19 – Construction du descripteur SIFT par histogrammes de gradients. Crédit : [LOWE \[1999\]](#).

Descripteur SURF

SURF construit quant à lui un vecteur de description utilisant une analyse de textures basée sur les ondelettes de Haar 2D (voir [GRAPS \[1995\]](#) pour une introduction).

Afin d'assurer une extraction rapide, SURF propose de réutiliser l'image intégrale afin de construire le vecteur de description. De la même manière que SIFT, le calcul est effectué dans une fenêtre gaussienne circulaire de paramètre égal à 1,5 fois le facteur d'échelle du point d'intérêt. Il considère une région de 4×4 blocs, eux-mêmes subdivisées en 2×2 blocs où sont calculés les coefficients d'ondelettes de Haar pour les directions x et y (voir figure 2.20 et 2.21).

Chaque composante du vecteur de description représente une caractéristique d'un des 4×4 blocs décrits. Chaque bloc est décrit par 4 éléments : la somme des réponses des ondelettes de Haar dans une direction et la somme de leurs valeurs absolues, et cela pour les deux directions x et y . Figure 2.22, nous pouvons observer la réponse du descripteur pour différents types de textures.

Mise en correspondance

La mise en correspondance constitue une étape obligatoire pour toutes les approches considérées. Elle consiste à trouver les couples de points provenant d'images différentes et décrivant le même objet physique.

Le premier élément constituant l'heuristique de mise en correspondance est une mesure de similarité. Il consiste généralement à calculer la norme de la distance euclidienne entre les deux vecteurs de description

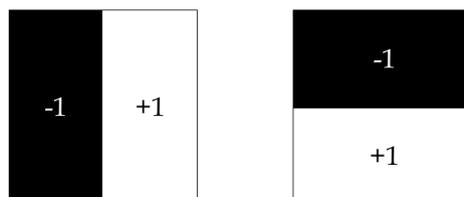


FIGURE 2.20 – Ondelettes de Haar utilisées pour la description de SURF. À gauche l'ondelette de Haar selon la direction x et à droite selon la direction y .

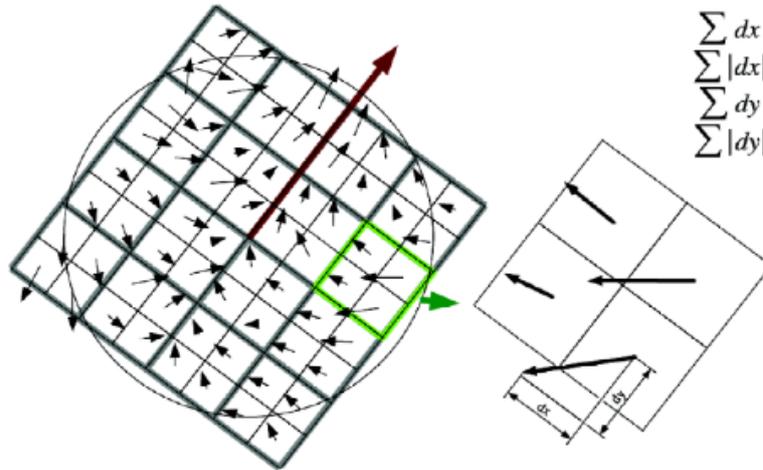


FIGURE 2.21 – Réponses des ondelettes de Haar au voisinage d'un point. Crédit : SURF [BAY et collab.](#) [2006].

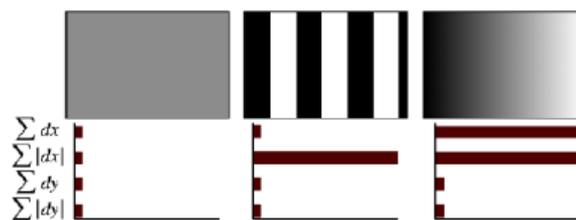


FIGURE 2.22 – Nous pouvons observer ici les exemples de réponses au descripteur par ondelette de Haar pour la direction x . De gauche à droite : une texture unie n'entraînant pas de réponse. Une texture entraînant une réponse de la somme de la valeur absolue des coefficients d'ondelette. Une texture de gradient d'intensité uniforme entraînant une réponse des 2 éléments décrivant la texture selon l'axe x . Crédit : SURF [BAY et collab.](#) [2006].

provenant des candidats et à prendre le candidat le plus proche :

$$d(d_1, d_2) = \sum_{i=1}^N (d_1^i - d_2^i)^2 \quad (2.52)$$

où d_1^i et d_2^i sont les i -ème éléments des vecteurs de description respectivement du point d'intérêt appartenant à l'image 1 et à l'image 2.

Une deuxième heuristique utilisée est le rapport de distance au plus proche voisin dans l'espace des caractéristiques :

$$\frac{d(d_1, d_2)}{d(d_1, d_3)} \quad \text{sachant} \quad d(d_1, d_2) < d(d_1, d_3) < d(d_1, d_i) \forall i \notin \{1, 2, 3\} \quad (2.53)$$

Appliquer à l'équation 2.53 un seuil de décision inférieur à 1 permet d'éliminer les correspondances ambiguës, où il existe un second point ayant quasiment la même distance de correspondance.

Statistiques robustes

La mise en correspondance des points d'intérêt n'est pas parfaite. Parmi les points appariés, il y a des points décrivant des zones qui correspondent effectivement, mais il y a aussi des points qui mettent en relation des zones différentes. Cette ambivalence est due à la fois au fait que deux zones différentes peuvent avoir le même environnement local et aussi à cause du vecteur de caractéristique qui a ses propres limites d'efficacité.

Afin de pallier la présence de couples aberrants, il est généralement fait usage de techniques permettant d'inhiber les couples aberrants, le plus souvent en pondérant les contributions des couples par une valeur unitaire pour les vrais positifs et proche de zéro pour les faux positifs.

M-Estimateur La méthode des moindres carrés peut être vue comme un cas particulier de M-estimateur. Les moindres carrés se définissent comme le minimum de la somme des carrés des résidus, ils peuvent donc être vus comme l'estimation du maximum de vraisemblance, dans le cas où les données possèdent un bruit additif gaussien de variance constante.

En généralisant la méthode des moindres carrés, pour une famille de fonctions de densité de probabilité f paramétrée par \mathbf{x} , un estimateur du maximum de vraisemblance de \mathbf{x} est calculé pour chaque échantillon en maximisant la fonction de vraisemblance dans l'espace des paramètres $\{\mathbf{x}\}$. Quand les observations sont indépendantes et possèdent la même distribution, une estimation $\hat{\mathbf{x}}$ du maximum de vraisemblance vérifie :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \left(\prod_{i=1}^n f(x_i, \mathbf{x}) \right) \quad (2.54)$$

HUBER et collab. [1964] proposent une généralisation de l'estimation du maximum de vraisemblance par la minimisation de :

$$\sum_{i=1}^n \rho(x_i, \mathbf{x}) \quad (2.55)$$

Du choix de la fonction ρ et de sa dérivé Ψ résulte les propriétés du M-estimateur, en matière de biais et d'efficacité, quand les échantillons proviennent réellement d'une distribution de probabilité issue du modèle. Il faut aussi, et c'est ce qui fait l'efficacité d'un M-estimateur, que ces fonctions se comportent convenablement quand l'échantillon est aberrant et ne provient pas de la distribution de probabilité issue du modèle.

La solution

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left(\sum_{i=1}^n \rho(x_i, \mathbf{x}) \right) \quad (2.56)$$

est appelée M-estimateur. Le "M" signifiant "Maximum-Likelihood type". Les estimateurs du maximum de vraisemblance sont un cas particulier des M-Estimeurs.

Pour la plupart des M-estimateurs, il n'existe pas de forme explicite et une approche itérative est nécessaire. Il est possible d'utiliser, par exemple, directement la méthode de Newton, mais généralement, la méthode des moindres carrés pondérés itératifs est utilisée.

Notion de point de rupture Le point de rupture d'un estimateur correspond à la proportion de mauvaises observations (de points aberrants) qu'un estimateur peut contenir tout en fournissant un résultat correct.

Un exemple pour imaginer cette notion est l'utilisation de la moyenne et de la médiane. La moyenne a un point de rupture de 0, étant donné qu'il suffit d'un seul échantillon altéré pour modifier sa valeur. À l'opposé, la médiane est un estimateur qui a un point de rupture de 50% (voir [HUBER \[2011\]](#)). Entre les deux, on peut considérer les moyennes tronquées, qui sont la moyenne de $x\%$ d'éléments centraux, fournissant un point de rupture de $x\%$.

M-estimateur "redescendant" Les M-estimateurs dits "redescendant" ("Redescending M-estimator") sont des M-estimateurs ayant une fonction de pondération revenant à 0 loin de l'origine qui permet de définir des zones de rejet des observations, car au-delà d'une certaine limite, ils rejettent complètement les échantillons aberrants. Ce rejet complet leur permet d'avoir un point de rupture élevé.

En revanche, cette caractéristique a pour conséquence de créer des fonctions objectifs à optimiser non-convexes, voir possédant des minimums locaux. L'utilisation de ces fonctions objectif rend obligatoire l'utilisation de schémas d'optimisation avancés et le fait de garantir une certaine proximité à la solution optimale à chaque itération, afin d'éviter une convergence vers les minima locaux.

Tukey M-estimateur L'estimateur de Tukey [[CRESSIE et HAWKINS, 1980](#)] ("Tukey bisquare", voir figure 2.23), est un M-estimateur de type "redescendant". Il prend comme a priori que la distribution des échantillons est issue d'un mélange de :

- une distribution normale, pour les "bons" points.
- une distribution uniforme, pour les points aberrants.

Il s'écrit sous la forme :

$$\begin{cases} \rho(x) = \frac{c^2}{6} \left(1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right) & \text{si } |x| < c \\ \rho(x) = \frac{c^2}{6} & \text{si } |x| > c \end{cases} \quad (2.57)$$

où c est la variance de la distribution des vrais positifs, généralement déterminée à l'aide de la médiane de la valeur absolue de la distance à la moyenne ("Median Absolute Deviation", MAD).

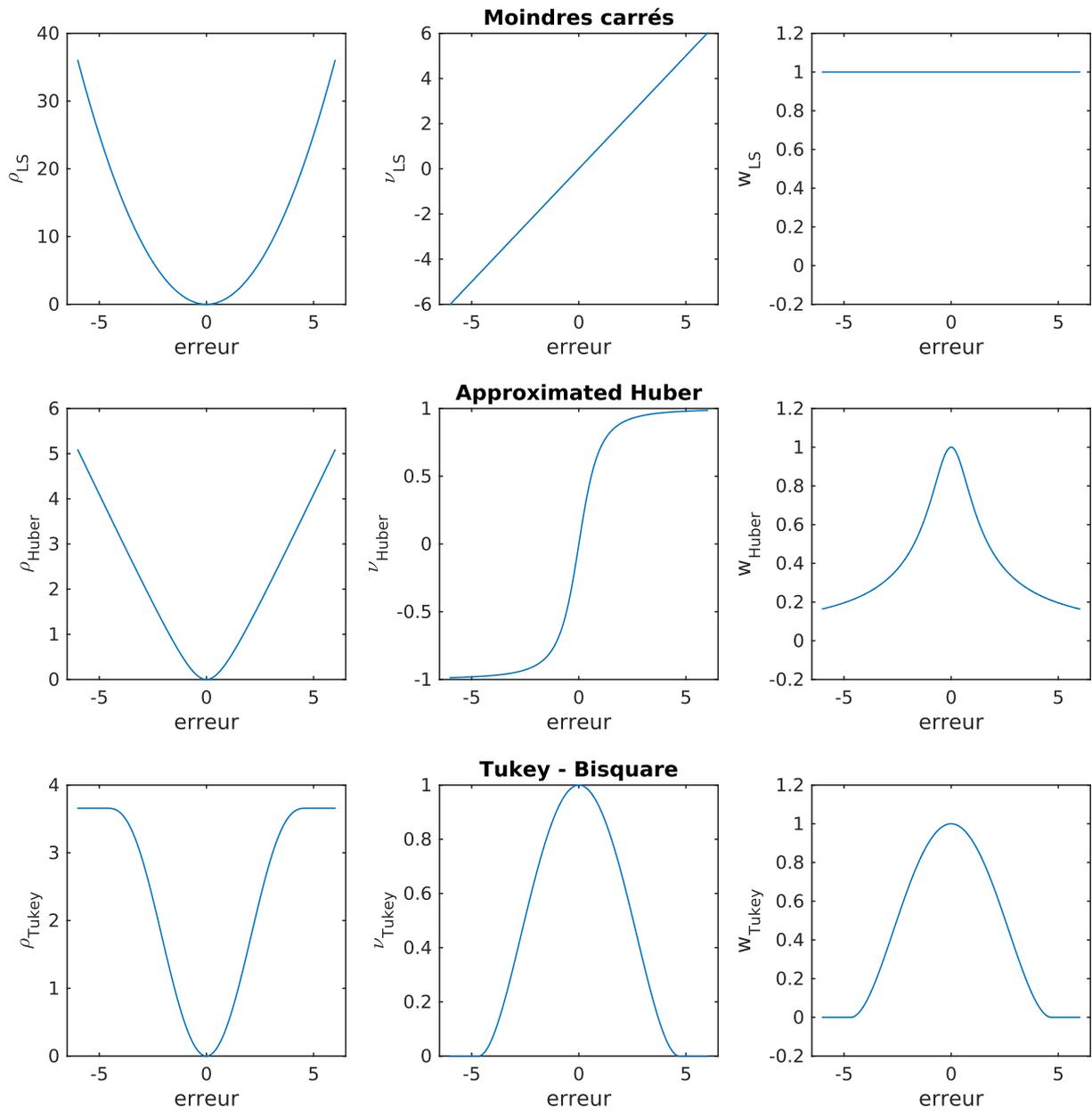


FIGURE 2.23 – Différents M-estimateurs. $\rho(x)$ est la fonction de perte, $\psi(x) = \frac{\partial \rho}{\partial x}$ et $w(x) = \frac{\psi(x)}{x}$ représente la pondération.



FIGURE 2.24 – Exemple de reconstruction 3D par ajustement de faisceaux. En bas, nous pouvons noter les multiples positions des caméras estimées par l’ajustement de faisceaux pour reconstruire le bâtiment. Crédit : [GHERARDI et collab. \[2010\]](#).

2.8.3 Exemple d’application : l’ajustement de faisceaux

Dans le cadre de la vision par ordinateur, les approches par Ajustements de Faisceaux (AF, “Bundle Adjustment”, voir synthèse : [TRIGGS et collab. \[2000\]](#)) sont assez représentatives des avantages des points d’intérêt. L’AF permet de reconstruire une scène 3D unique à partir de multiples vues 2D en rétroprojectant les points d’intérêt extraits des vues 2D dans la scène 3D avec l’utilisation d’une optimisation globale (ici, globale est utilisé dans le sens : générale, large). L’AF permet l’emploi de caméras bas de gamme [KARLSSON et collab. \[2005\]](#) et les dernières implémentations ont permis de reconstruire de grandes parties de villes [FRAHM et collab. \[2010\]](#) (figure 2.24).

Cette classe d’algorithmes comporte de nombreuses similarités avec le sujet de ce manuscrit : l’optimisation globale, qui utilise de manière conjointe de nombreuses images au cours du processus, et des statistiques robustes.

Optimisation globale

L’ajustement de faisceaux utilise une approche par optimisation globale. Il utilise pour fonction objectif l’erreur quadratique moyenne de reprojction afin d’optimiser tous les paramètres du problème de manière conjointe :

- Paramètres des caméras (matrice intrinsèque, calibration : longueur focale, déformation, ...)
- Positions des caméras (6 degrés de liberté par caméra : 3 translations, 3 rotations)
- Positions des points d’intérêt dans l’espace (3 degrés de liberté par point : 3 translations)

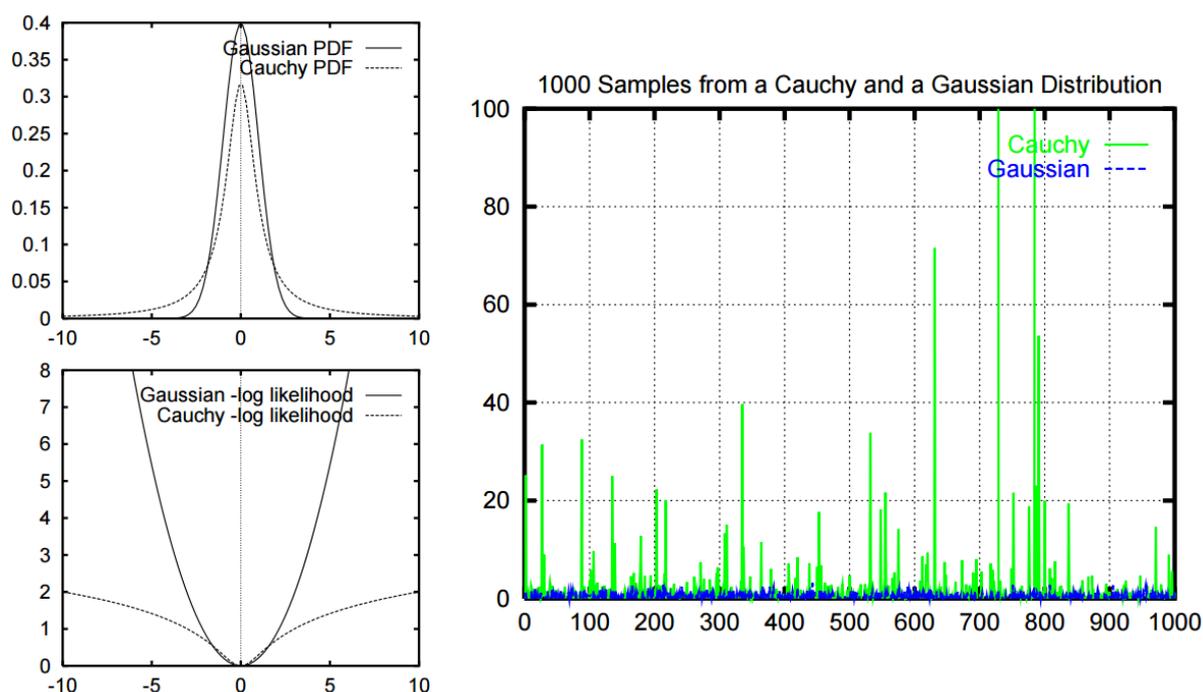


FIGURE 2.25 – Un exemple comparatif de deux distributions : une distribution gaussienne et une distribution à queue lourde : Cauchy ($p(x) = (\pi(1+x^2))^{-1}$). Nous pouvons observer que malgré des densités de probabilités similaires (en haut à gauche), le comportement des deux distributions est assez différents si nous observons leurs fonctions de log-vraisemblance (en bas à gauche). L'augmentation de la probabilité au niveau de la queue de la distribution de Cauchy rend la présence d'outliers beaucoup plus probable (voir figure à droite). Crédit : TRIGGS et collab. [2000].

Utilisation de statistiques robustes, gestion des points aberrants

L'ajustement de faisceaux est basé sur la mise en correspondance des points d'intérêt provenant d'une image avec les points provenant d'autres images projetées dans un espace 3D. Comme toute technique utilisant des points d'intérêt, la mise en correspondance n'est pas parfaite, il y a des correspondances erronées, des faux positifs. Ces faux positifs sont des points ayant été mis en correspondance car possédant des vecteurs de caractéristiques proches mais ne représentant pas le même objet physique. Ces mises en correspondance sont nommées points (de mesure) aberrants ("outliers" en anglais) et doivent être rejetés.

L'approche usuelle d'optimisation globale consiste à utiliser la méthode des moindres carrés, que nous avons déjà vue, peut aussi être vue comme une estimation du maximum de vraisemblance si les mesures d'erreurs sont indépendantes et distribuées selon une loi normale de variance constante.

Afin de pouvoir optimiser les différents paramètres de l'ajustement de faisceaux, les équations issues de la géométrie projective étant non-linéaires, une méthode de moindres carrés non-linéaires est nécessaire. Les moindres carrés (linéaires et non linéaires) sont sensibles aux points aberrants, du fait de la finesse des queues de la distribution gaussienne (voir figure 2.25 pour un exemple). La présence de paires aberrantes lors de la mise en correspondance de points d'intérêt rend impossible l'utilisation des moindres carrés classiques.

En première approximation, la distribution des erreurs de mise en correspondance des points d'intérêt peut être vue comme la somme d'une distribution gaussienne rendant compte des mises en correspondance correctes et d'une distribution uniforme, rendant compte des points aberrants.

L'ajustement de faisceaux fait donc usage des deux méthodes détaillées dans les parties suivantes :

- le M-estimateur qui améliore la robustesse,
- l'utilisation de schémas d'optimisation avancés, garantissant la convergence même si le problème est non-convexe, voir non monotone.

2.8.4 Points d'intérêt 3D

Les points d'intérêt ont été initialement créés pour des images 2D. Afin de les utiliser pour extraire des points d'intérêt d'images médicales 3D, il convient de généraliser le concept.

Plusieurs approches existent pour généraliser les points d'intérêt en trois dimensions, nous nous intéresserons à celles réellement 3D, laissant de côté celles exploitant des maillages surfaciques, comme [DUTAGACI et collab. \[2012\]](#) ou des images 3D à partir de maillages, comme [YU et collab. \[2013\]](#).

Dans un premier temps, le passage de 2D à 3D est plutôt simple, la plupart des concepts se généralisant bien à une dimension supplémentaire. Néanmoins, cette généralisation possède quelques difficultés présentées dans les sections suivantes :

Invariance en rotation

Passer de deux dimensions à trois dimensions fait passer le nombre de variables pour définir l'orientation d'un point de un à trois. Il est nécessaire dans ce cas de construire un repère possédant trois rotations. L'utilisation de gradients pour la seconde et la troisième dimension est assez instable, la norme du gradient restant étant faible. [CHEUNG et HAMARNEH \[2007\]](#) proposent une solution en généralisant à la dimension n en considérant le gradient en coordonnées hyper-sphériques.

Un deuxième effet négatif qui n'existe pas en deux dimensions apparaît en 3D : les axes étant multiples, ils peuvent être permutés avec de faibles changements de valeur du gradient. Ce qui a pour conséquence d'avoir une forte dissimilarité entre des points qui décrivent pourtant la même localité.

Points particuliers

En deux dimensions, il peut y avoir des points "selles", qui se caractérisent par des valeurs propres de la matrice hessienne de signe contraire. Généralement, les méthodes de mise en correspondance des points d'intérêt rejettent de tels points.

En trois dimensions, le nombre de ces points particuliers augmente. La matrice hessienne étant dorénavant une matrice 3×3 , le nombre de configurations où les valeurs propres ne sont pas de même signe passe théoriquement de 50% (2 cas parmi 4) à 75% (6 cas parmi 8).

De plus, dans le cas 2D, un déterminant positif suffit à garantir des valeurs propres de même signe. Dans le cas 3D, ceci n'est pas possible, et le calcul explicite des valeurs propres semble nécessaire. Pour garantir que les valeurs propres soient de même signe, [ALLAIRE et collab. \[2008\]](#) proposent une solution élégante introduisant la somme des mineurs principaux du second ordre.

Soit \mathbf{H} la matrice hessienne en un point (x, y, z) de l'image :

$$\mathbf{H} = \begin{bmatrix} L_{xx} & L_{xy} & L_{xz} \\ L_{xy} & L_{yy} & L_{yz} \\ L_{xz} & L_{yz} & L_{zz} \end{bmatrix} \quad (2.58)$$

Soit $\alpha \geq \beta \geq \gamma$ les valeurs propres de \mathbf{H} . La somme des mineurs principaux du second ordre est définie par :

$$\sum det_2^P(\mathbf{H}) = \beta\gamma + \gamma\alpha + \alpha\beta = L_{yy}L_{zz} - L_{yz}^2 + L_{zz}L_{xx} - L_{xz}^2 + L_{xx}L_{yy} - L_{xy}^2 \quad (2.59)$$

La condition de détection d'un point d'intérêt devient :

$$\sum det_2^P(\mathbf{H}) > 0 \quad \text{et} \quad tr(\mathbf{H})det(\mathbf{H}) > 0 \quad (2.60)$$

Temps de calcul

Le problème du temps de calcul doit être considéré pour le passage 2D à 3D. L'ajout d'une 3ème dimension augmente radicalement la complexité de n'importe quel algorithme. Dans le cadre du traitement d'images si l'on exclut le temps réel, l'extraction de points d'intérêt est un processus suffisamment rapide pour être négligeable face aux autres coûts, lors du passage en 3D ce coût peut devenir prépondérant.

Chapitre 3

Contribution 1 : mise en correspondance rigide rapide

Sommaire

3.1	Introduction	49
3.2	Challenges	50
3.2.1	Application médicale	50
3.2.2	Points d'intérêt	50
3.2.3	Mise en correspondance de grands groupes	50
3.3	SURF 3D	51
3.3.1	Images intégrales 3D	51
3.3.2	Approximation de la hessienne 3D	51
3.3.3	Choix des descripteurs	52
3.3.4	A priori liés à l'imagerie médicale	53
	Orientation	53
	Dimensions réelles	53
3.4	Mise en correspondance rigide de paires : FPR	53
3.4.1	Introduction	53
3.4.2	Données	54
3.4.3	Extraction des points d'intérêt	54
3.4.4	Appariement des points d'intérêt	54
3.4.5	Transformations	57
3.4.6	Estimation de la transformation	57
3.4.7	RANSAC	58
3.4.8	Conclusion	60
3.4.9	Limites du recalage par paire	61
3.5	Mise en correspondance rigide de grands groupes d'images médicales	62
3.5.1	Une nouvelle représentation par graphe	62
3.5.2	Mise en correspondance des paires	62
3.5.3	Modèle de déformation	63
3.5.4	Formalisation du problème	64
	Énoncé du problème	64
	Équation utilisant l'opérateur laplacien	64
3.5.5	Élagage du graphe et connexité	66
	Critère de sélection	66
	Nombre d'arêtes restantes	68

3.5.6	Extension à une translation et un facteur d'échelle	68
3.6	Algorithme proposé et résultats expérimentaux	70
3.6.1	Fast Groupwise Registration : FGR	70
	Données	70
	Algorithme FGR	70
	Résultats	71
3.6.2	Masquage des yeux	72
3.7	Conclusion	73

3.1 Introduction

Vers une représentation compacte : les points d'intérêt L'objectif de notre approche est le traitement de grands jeux d'images médicales. Ce défi nécessite une représentation compacte des données, permettant la mise en correspondance des images. Les points d'intérêt, largement utilisés en vision par ordinateur, répondent à ce besoin. Cependant, l'utilisation des points d'intérêt dans le cadre de l'imagerie médicale implique de dépasser plusieurs verrous scientifiques, qui font l'objet de ce chapitre.

Nous avons décidé d'utiliser une version 3D de SURF ("Speeded Up Robust Feature", [BAY et collab. \[2006\]](#)) dans notre approche de mise en correspondance grands groupes. Un des intérêts de SURF pour le passage à la 3D est son calcul en 2 étapes. La première étape, le calcul de l'image intégrale, a une complexité linéaire en fonction du nombre de voxel. La deuxième étape, la détection et l'extraction des points d'intérêt, fait usage de cette image intégrale. Cette étape, normalement de coût calculatoire quadratique en fonction du nombre de voxels, est linéaire grâce à l'usage d'images intégrales. La figure 3.1 fait la synthèse des différentes étapes de mise en œuvre des points d'intérêt. Ces travaux ont fait l'objet d'une publication [AGIER et collab. \[2016a\]](#) et d'une communication nationale [AGIER et collab. \[2016b\]](#).

Applicabilité des points d'intérêt à l'imagerie médicale Les points d'intérêt sont déjà utilisés dans le cadre d'images médicales [[ALLAIRE et collab., 2008](#); [CHEUNG et HAMARNEH, 2007](#)]. Cependant, les expériences les utilisant sont généralement restreintes au cerveau, avec une variabilité faible. La seconde partie de ce chapitre s'attachera à prouver l'applicabilité des points d'intérêt dans notre cadre d'expérimentation : inter-patient, variabilité forte.

Dans ce chapitre, nous nous appliquerons à développer une méthode robuste, compatible avec les points d'intérêt. L'analyse procrustéenne généralisé est une méthode très proche de notre proposition. Par exemple, [[BARTOLI et collab., 2013](#)] propose une méthode de recalage rigide pour des groupes d'images utilisant des appariements incomplets entre points mais ne permettant pas une utilisation robuste avec les faux positifs inhérents aux points d'intérêt.

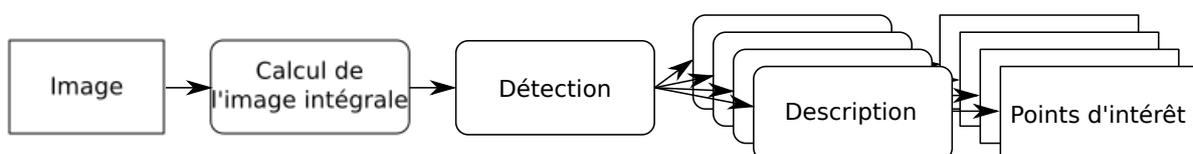


FIGURE 3.1 – Schéma du processus complet de l'extraction de points d'intérêt par SURF 3D.

3.2 Challenges

3.2.1 Application médicale

Description de l’anatomie humaine Les descripteurs des points d’intérêt ont initialement pour objectif de décrire des scènes identiques, vues sous des angles différents. Ils doivent donc être robustes aux changements d’illumination, aux changements de point de vue, etc. Dans le cadre de l’imagerie médicale, les points d’intérêt décrivent l’anatomie humaine, et les variations entre images ne sont pas de même nature. Il convient donc d’évaluer la robustesse face aux variations de l’anatomie.

Appariement : Compromis entre faux positifs et variabilité Nous sommes dans un contexte de forte variabilité. Ceci implique que des zones censées correspondre n’ont pas nécessairement le même descripteur, car elles proviennent de patients différents. Il est donc nécessaire d’être moins exigeant concernant le critère d’appariement des points. La conséquence est une augmentation du nombre de points d’intérêt appariés qui ne devrait pas l’être. Un compromis doit donc être trouvé entre ses deux aspects, tout en utilisant des méthodes permettant une robustesse à la présence de ces appariements aberrants.

3.2.2 Points d’intérêt

Nous pouvons décomposer SURF en plusieurs parties. Certaines d’entre elles, comme le calcul de l’image intégrale ou le parcours des octaves, s’accommodent très bien du changement de dimension et la généralisation est immédiate. Par contre, il subsiste des verrous :

Hessienne 3D Le calcul de hessienne en 3 dimensions à partir d’images intégrales nécessite d’étendre l’approximation utilisée par SURF en 3 dimensions, ce qui demande une certaine attention, afin de respecter les hypothèses de départ (approximation non biaisée, détection des blobs).

Choix du voisinage et de la description Étendre un descripteur à une dimension supplémentaire nécessite des ajustements afin de garder un vecteur cohérent, qui conserve un bon compromis expressivité-taille.

Application médicale Étant donné que nous avons une application précise et bien définie, le recalage monomodale d’images issues de l’imagerie médicale, il est intéressant d’utiliser tous les a priori disponibles (orientations, position, composition), afin d’obtenir les meilleures performances possibles.

3.2.3 Mise en correspondance de grands groupes

Images partielles Notre objectif est de pouvoir traiter des données géométriquement hétérogènes, provenant de sources diverses. Dans ce but, nous voulons être capable de traiter des images contenant des corps entièrement imagés, mais aussi des images dont la zone imagée est plus réduite. Les images partielles sont un problème usuel dans le cadre de la mise en correspondance d’images. La principale difficulté est due au recouvrement partiel qui limite la robustesse des approches existantes.

Difficultés liées à l’utilisation d’une référence Les références sont une difficulté spécifique qui limite l’utilisation d’images incomplètes avec les approches actuelles. En considérant les approches en étoiles (voir Chapitre 2.7.3), il est nécessaire que l’élément central soit une image ayant un recouvrement suffisant avec l’ensemble de la base. Même si le choix d’une image corps entier pourrait sembler fournir une solution satisfaisante, en réalité, ce choix ne résout pas toutes les difficultés. Dans le cadre de larges bases de données contenant aussi bien des adultes que des enfants, des hommes et des femmes, la notion de référence et “d’humain moyen” est discutable.

3.3 SURF 3D

3.3.1 Images intégrales 3D

Notre objectif consiste à étendre SURF de la 2D à la 3D. La première étape est d'étendre le concept d'images intégrales à la 3ème dimension.

Le calcul d'une image intégrale (Section 2.8.1 et BAY et collab. [2006]) se généralise assez facilement de 2 à 3 dimensions. En 3 dimensions, la valeur d'un voxel de l'image intégrale vaut :

$$\mathbf{S}(x, y, z) = \sum_{i=0}^x \sum_{j=0}^y \sum_{k=0}^z I(i, j, k) \quad (3.1)$$

En 3 dimensions, l'opération accélérée par l'usage des images intégrales est le calcul d'un bloc parallélépipédique paramétré par deux points (x_1, y_1, z_1) et (x_2, y_2, z_2) .

Le calcul de l'intégrale d'un bloc à partir d'une image 3D est :

$$\text{Naive3DBox}(x_1, y_1, z_1, x_2, y_2, z_2) = \sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} \sum_{k=z_1}^{z_2} I(i, j, k) \quad (3.2)$$

Le calcul de l'intégrale sur ce bloc à partir d'une image intégrale \mathbf{S} est le suivant :

$$\begin{aligned} \text{Integral3DBox}(x_1, y_1, z_1, x_2, y_2, z_2) = & + S(x_2, y_2, z_2) - S(x_2, y_2, z_1) \\ & - S(x_2, y_1, z_2) - S(x_1, y_2, z_2) \\ & + S(x_1, y_1, z_2) + S(x_1, y_2, z_1) \\ & + S(x_2, y_1, z_1) - S(x_1, y_1, z_1) \end{aligned} \quad (3.3)$$

On peut noter que le calcul d'un bloc reste de complexité constante lors du passage de 2D à 3D, passant de $O(4)$ à $O(8)$, alors que le calcul sans usage d'image intégrale passe de $O(n^2)$ à $O(n^3)$, n étant la dimension du bloc.

3.3.2 Approximation de la hessienne 3D

SURF utilise une approximation de la hessienne par blocs, voir figure 2.18. Cette approximation par blocs introduit un biais dans le calcul du déterminant, les valeurs du masque gaussien approximé par blocs étant biaisées par rapport à celles du masque gaussien non approximé.

Le déterminant du Hessien 2D est le suivant :

$$\det(\mathbf{H}) = L_{xx}L_{yy} - L_{xy}^2 \quad (3.4)$$

où L_{xx}, L_{xy}, L_{yy} sont les éléments de la matrice hessienne.

Les filtres gaussiens sont approximés par un modèle sous forme de blocs. Nous noterons D_{xx} l'approximation de L_{xx} et il en sera de même pour les autres éléments de la matrice hessienne approximée.

Du fait de l'approximation, le poids relatif des éléments diagonaux et des éléments non diagonaux n'est pas respecté, ce qui apporte un biais ω [BAY et collab., 2006] :

$$\omega = \frac{\|L_{xy}\|_F \times \|D_{yy}\|_F}{\|L_{yy}\|_F \times \|D_{xy}\|_F} = 0.912\dots \quad (3.5)$$

où $\|\dots\|_F$ indique la norme de Froebenius.

Cette constante ω est introduite dans le calcul du déterminant :

$$\det(\mathbf{H}_{approx}) = D_{xx}D_{yy} - (\omega D_{xy})^2 \quad (3.6)$$

Comme nous l'avons vu dans la section 2.8.4, l'extension 3D implique l'utilisation de la matrice hessienne (3 × 3) :

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \quad (3.7)$$

et de calculer son déterminant approximé :

$$\det(\mathbf{H}_{approx}) = D_{xx}D_{yy}D_{zz} + 2\omega^3 D_{xy}D_{yz}D_{xz} - \omega^2 D_{xx}D_{yz}D_{yz} - \omega^2 D_{yy}D_{xz}D_{xz} - \omega^2 D_{zz}D_{xy}D_{xy} \quad (3.8)$$

ainsi que l'approximation de la somme des mineurs principaux du second ordre :

$$\sum \det_2^P(\mathbf{H}) = D_{yy}D_{zz} - (\omega D_{yz})^2 + D_{zz}D_{xx} - (\omega D_{xz})^2 + D_{xx}D_{yy} - (\omega D_{xy})^2 \quad (3.9)$$

3.3.3 Choix des descripteurs

SURF 2D utilise un voisinage 4 × 4 d'où il est extrait 4 descriptions (voir section 2.8.2) :

$$4 \times 4 \times \begin{cases} \sum Haar_x \\ \sum |Haar_x| \\ \sum Haar_y \\ \sum |Haar_y| \end{cases} \quad (3.10)$$

Ce qui résulte en un vecteur à 64 dimensions. Si l'on étend ce vecteur de description en 3 dimensions, nous obtenons le vecteur de description suivant :

$$4 \times 4 \times 4 \times \begin{cases} \sum Haar_x \\ \sum |Haar_x| \\ \sum Haar_y \\ \sum |Haar_y| \\ \sum Haar_z \\ \sum |Haar_z| \end{cases} \quad (3.11)$$

soit un vecteur à 384 dimensions. Un espace de caractéristiques aussi grand n'est pas souhaitable, et l'état de l'art au sujet des points d'intérêt montre qu'un bon compromis de la taille du vecteur de caractéristiques est compris entre 50 et 100 dimensions (voir la saturation en l-precision à partir de 50 composantes dans BAY et collab. [2008]).

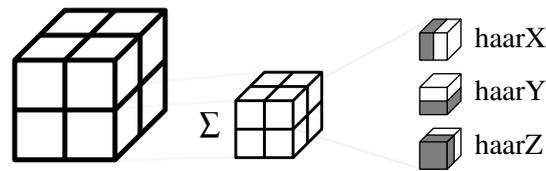


FIGURE 3.2 – Schéma de la localité utilisée pour la description d'un point d'intérêt SURF3D. Le premier cube (gauche) est composé des $2 \times 2 \times 2$ blocs décrits. Le cube central contient une des zones décrites en sommant les réponses des ondelettes de Haar (droite) pour chacun de ses sous-cubes.

En réduisant le voisinage à un cube de deux zones de côté et en gardant la même description des zones, il en résulte le vecteur de caractéristiques suivant :

$$2 \times 2 \times 2 \times \begin{cases} \sum Haar_x \\ \sum |Haar_x| \\ \sum Haar_y \\ \sum |Haar_y| \\ \sum Haar_z \\ \sum |Haar_z| \end{cases} \quad (3.12)$$

Nous utiliserons ce vecteur à 48 dimensions, caractérisant un voisinage plus proche du point (les voisins directs) dont la structure est visible figure 3.2.

3.3.4 A priori liés à l'imagerie médicale

La connaissance a priori des images sur lesquelles va être appliqué SURF peut être exploitée de deux manières.

Orientation

Les patients imagés ayant la même posture, nous pouvons utiliser la version "Upright" de SURF (U-SURF BAY et collab. [2006]), non-invariante en rotation. Cette version nous permet d'éviter de devoir gérer l'ambiguïté en rotation des points d'intérêt qui est d'autant plus complexe en 3D. Le principal désavantage de ce choix provient de la variabilité inter-patient qui peut modifier l'orientation des structures anatomiques et donc induire de petites rotations, même si les patients sont imagés selon le même protocole et donc la même orientation.

Dimensions réelles

Les images médicales possèdent une information concernant la taille réelle des pixels (en mm). Cette information nous permet d'en déduire la taille réelle des points d'intérêt. La phase d'appariement bénéficie de cet a priori par l'ajout d'un critère de similarité de la taille des points d'intérêt.

3.4 Mise en correspondance rigide de paires : FPR

3.4.1 Introduction

"Fast Pairwise Registration" est la première contribution de ce manuscrit. Elle consiste en la mise en correspondance rigide de deux images grâce à l'utilisation de points d'intérêt SURF 3D et l'estimation de la matrice de transformation à l'aide d'un processus RANSAC (voir figure 3.3).

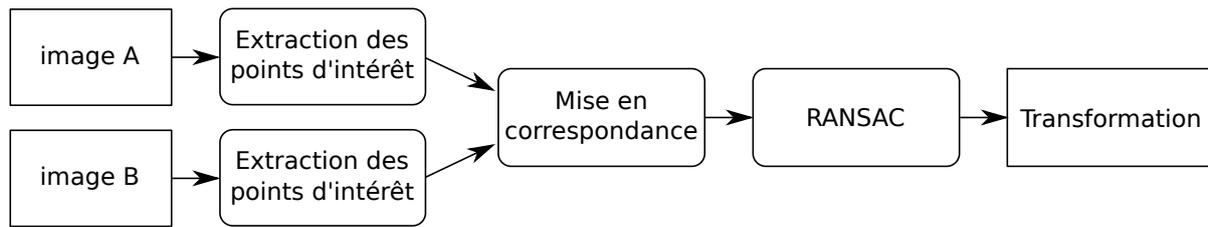


FIGURE 3.3 – Schéma du processus complet de FPR. On peut voir de gauche à droite : les images d’entrées, les différents calculs ainsi qu’en sortie la transformée calculé par FPR.

Cette contribution permet de fournir des éléments de réponse sur la viabilité de l’utilisation de points d’intérêt dans le cadre de mise en correspondance inter-patient. Le principal verrou scientifique concerne l’estimation de la robustesse d’un détecteur et d’un descripteur vis-à-vis de la variabilité humaine sur l’ensemble du corps. Jusqu’à présent, les expérimentations utilisant des points d’intérêt en imagerie médicale se sont principalement concentrées sur le cerveau.

3.4.2 Données

Pour cette application, nous utilisons des images CT-scanner 3D. Les deux images que nous utilisons pour illustrer les résultats sont :

- **(A) Un corps entier** : Un corps imagé quasiment entièrement, de la tête aux tibias. l’image a pour dimension $512 \times 512 \times 316$ avec une taille de voxel de $1mm \times 1mm \times 5mm$. Chaque voxel est encodé en niveau de gris représentant l’absorption en rayon X normalisée (échelle de Hounsfield fenêtrée), codé sur 2 octets, ce qui représente une image de 158Mo.
- **(B) Des jambes** : l’image a pour dimension $512 \times 512 \times 171$ avec une taille de voxel de $1mm \times 1mm \times 5mm$. Chaque voxel est aussi codée sur 2 octets, ce qui représente une image de 86Mo.

3.4.3 Extraction des points d’intérêt

La première étape de FPR est l’extraction des points d’intérêt des 2 images 3D. En plus de la mise en place d’un protocole permettant l’extraction automatique de points de qualité (permettant un bon appariement), il est nécessaire de fournir à l’estimateur de la transformation la meilleure information de position. C’est un compromis entre le nombre de points d’intérêt et la qualité de ces points (capacité de mise en correspondance, probabilité de générer des vrais positifs, localisation précise), qui doit être réglé.

L’extraction des points d’intérêt ne se fait que sur des images isotropes, possédant des voxels cubiques. Un ré-échantillonnage est effectué sur les deux images afin d’avoir une taille de voxel de $1mm \times 1mm \times 1mm$. Le sur-échantillonnage d’un facteur 5 résultant dans l’axe Z n’est pas problématique, car SURF sous-échantillonne d’un facteur 2 pour la résolution la plus fine et le masque gaussien minimal à une taille de 9×9 voxels. Les points d’intérêt recherchés les plus petits ont une taille de $18mm^3$, ce qui est bien plus grand que l’échantillonnage original. La figure 3.4 présente un exemple d’extraction et d’appariement de points d’intérêt de deux images médicales.

L’ensemble des paramètres utilisés pour SURF, identiques pour les deux images, sont : tailles des voxels (“spacing”) : 1mm, seuil d’extraction des points d’intérêt : 7000 (déterminant normalisé de la hessienne, sans unité).

3.4.4 Appariement des points d’intérêt

Nous rappelons que l’appariement a pour but de fournir un ensemble de couples de points d’intérêt dont l’un est issu de l’image A et l’autre de l’image B. Un couple de points d’intérêt est censé décrire

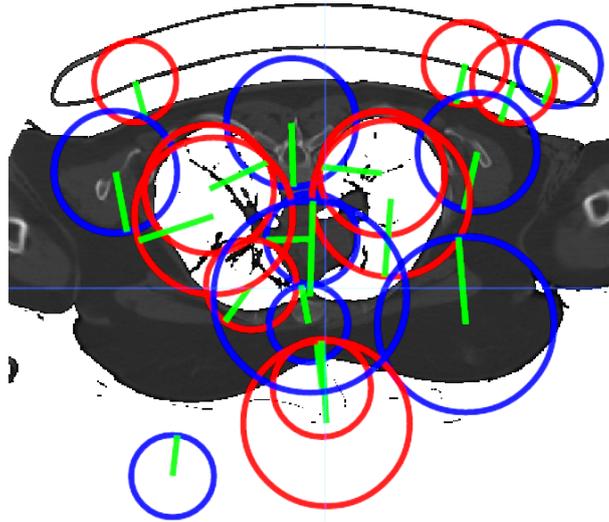


FIGURE 3.4 – Exemple d'extraction de points d'intérêt SURF sur une image médicale. Les points sont extraits de l'image 3D. Nous pouvons voir ici une coupe transverse de cette image 3D au niveau du thorax, avec les poumons visibles. Les cercles rouges représentent les points d'intérêt clairs sur fond sombre et les cercles bleus représentent les points d'intérêt sombres sur fond clair.

la même zone (exemple : un poumon, le sommet d'un os). L'ensemble des couples contient des vrais positifs ("inliers") respectant cette heuristique, mais il possède aussi des points aberrants ("outliers"), décrivant des endroits différents mais qui possèdent la même signature caractéristique, ce qui trompe l'algorithme. L'appariement a pour but de trouver un maximum de vrais positifs, en minimisant le taux de points aberrants.

Le processus d'appariement est le suivant. Pour chaque point de l'image A, nous considérons l'ensemble des points de l'image B satisfaisant les critères suivants (issus de leurs vecteurs de description) :

- **Laplacien de même signe** [BAY et collab., 2006] : les points décrivent des zones de même type : sombres sur fond clair, ou claires sur fond sombre.
- **Similarité d'échelle** : les points possèdent une échelle similaire, avec un rapport d'échelle inférieur à 1.3.

Parmi ce sous ensemble de points de l'images B, le processus consiste à trouver les deux plus proches voisins, au sens de la distance euclidienne dans l'espace des caractéristiques. Des conditions supplémentaires permettent d'être plus sélectif [BAY et collab., 2006; MIKOLAJCZYK et SCHMID, 2005] :

- Le voisin le plus proche, dans l'espace des caractéristiques, est à une distance inférieure à un seuil, ici 0.8, ce qui signifie que les textures locales autour des points sont similaires,
- Le rapport de la distance au second voisin sur la distance au premier voisin est inférieur à un seuil, ici 0.98, ce qui signifie qu'il n'y a pas d'autres points dans la seconde image qui ressemblent au premier.

Alors le couple formé par le point courant de la première image et le plus proche voisin de la seconde image sont ajoutés à l'ensemble final des points appariés.

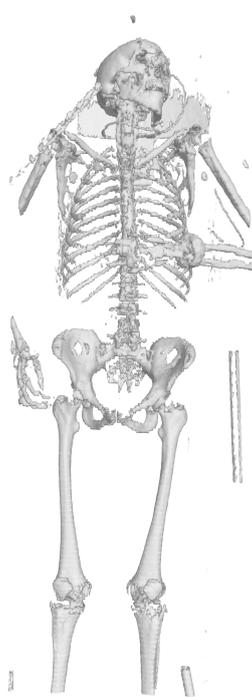
Dans l'expérience illustrée dans la figure 3.5, le nombre de points d'intérêt issus de l'image A est de 5570, tandis que le nombre de points issus de l'image B est de 1408. L'appariement a généré 3062 couples de points.



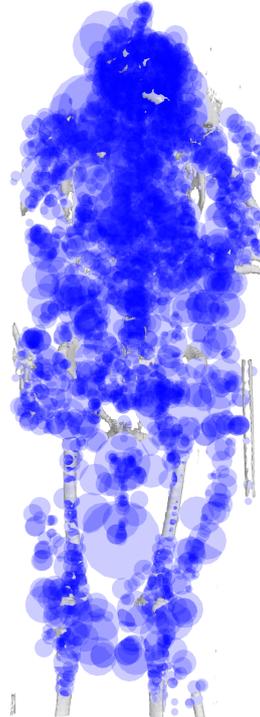
(a) Image (A)



(b) Image (A) + points d'intérêt.
Nombre de points : 5570



(c) Image (B)



(d) Image (B) + points d'intérêt.
Nombre de points : 1408

FIGURE 3.5 – Extraction de points d'intérêt de 2 images CT-scanner 3D. L'extraction du squelette n'est effectuée qu'à des fins d'affichage.

On peut noter une asymétrie dans le traitement de l'image. Si on intervertit les images A et B, l'ensemble des couples appariés n'est pas le même. Ceci est mis en évidence par la présence de plus de couples que de points dans l'image B. Intervertir les 2 images produit néanmoins des performances similaires, malgré la présence de moins de couples. Cette caractéristique d'asymétrie se retrouve dans l'ensemble des méthodes par points d'intérêt. Selon l'application, il pourrait être intéressant de prendre l'intersection ou l'union des ensembles issues de la mise en correspondance de A avec B et de B avec A. Mais dans notre cas, ça ne modifie pas la capacité de RANSAC à trouver ou non un modèle.

3.4.5 Transformations

L'objectif de cet appariement est d'estimer des transformations globales, qui sont décrites par un faible nombre de degrés de liberté et une matrice de transformation.

Dans le cadre de cette expérimentation, nous avons retenu 3 transformations rigides, décrites dans le tableau 3.1.

Nom	Nombre de ddl.	Expression de la transformée
Translation	3	$p \rightarrow p + t$
Translation, Rotation	6	$p \rightarrow R_{\rho, \theta, \phi} \times p + t$
Translation, Rotation, échelle	7	$p \rightarrow s \times R_{\rho, \theta, \phi} \times p + t$

TABLEAU 3.1 – Les différentes transformations globales utilisées dans l'approche FPR. Pour chaque transformation, il est spécifié le nombre de degrés de liberté (DDL), ainsi que la fonction de transformation associée.

3.4.6 Estimation de la transformation

Le processus de mise en correspondance par RANSAC nécessite l'estimation de la transformation à partir d'un ensemble de couples, dans notre cadre, ceux issus de l'appariement des points d'intérêt.

Afin d'estimer les transformations globales décrites dans la partie précédente, nous utilisons la méthode des centroïdes créée par UMEYAMA [1991]. Nous décrivons ici le cas de la transformation possédant sept degrés de liberté : la translation, la rotation et l'échelle, les autres étant des variations avec moins de degrés de liberté.

Soit un ensemble de couples de points d'intérêt $\{x_i, y_i\}$, nous souhaitons minimiser la fonction de coût suivante :

$$e(\mathbf{R}, \mathbf{t}, s)^2 = \frac{1}{n} \sum_{i=1}^n \|y_i - (s\mathbf{R}x_i + \mathbf{t})\|^2 \quad (3.13)$$

Umeyama se propose de travailler avec la décomposition en valeur singulière (SVD) en estimant préalablement la translation afin de travailler sur des ensembles centrés :

$$x'_i = x_i - \bar{x} \quad y'_i = y_i - \bar{y} \quad (3.14)$$

Ce qui nous fournit en premier lieu la translation :

$$\mathbf{t} = \bar{y} - \bar{x} \quad (3.15)$$

Nous avons maintenant la relation $y'_i = s\mathbf{R}x'_i$ que nous pouvons exprimer de manière matricielle :

$$\mathbf{Y}_{3 \times N} = s\mathbf{R}\mathbf{X}_{3 \times N} \quad (3.16)$$

où \mathbf{Y} est la matrice contenant les y'_i et \mathbf{X} la matrice contenant les x'_i .

La méthode proposée par Umeyama consiste à calculer la rotation grâce à la SVD de la matrice de corrélation des observations :

$$\text{svd}(\mathbf{X}'\mathbf{Y}'^T) = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (3.17)$$

$$\mathbf{R} = \mathbf{V}\mathbf{U}^T \quad (3.18)$$

Le facteur d'échelle est donné par :

$$s = \frac{\text{tr}(\mathbf{X}'\mathbf{Y}'^T)}{\text{tr}(\mathbf{X}'\mathbf{X}'^T)} \quad (3.19)$$

3.4.7 RANSAC

Nous utilisons l'algorithme RANSAC, voir partie 2.6.11, pour trouver les paramètres des modèles ajustés aux données en rejetant les échantillons aberrants. Pour cette expérimentation, nous utilisons un seuil de 40mm de distance entre deux points pour définir si un échantillon est considéré comme vrai ou faux positif.

RANSAC ne fournit pas de solution optimale mais une solution dit "PAC" : Probablement Approximativement Correcte. Ce qui signifie qu'il y a une certaine probabilité α d'atteindre une solution avec une distance à la valeur optimale inférieure à ε . Dans le cadre de RANSAC, α est directement lié à la sélection d'un modèle en ne prenant que de bons échantillons. ε est fixé, lié au bruit auxquelles ces échantillons sont sujets.

$$\alpha = 1 - (1 - w^n)^k \quad (3.20)$$

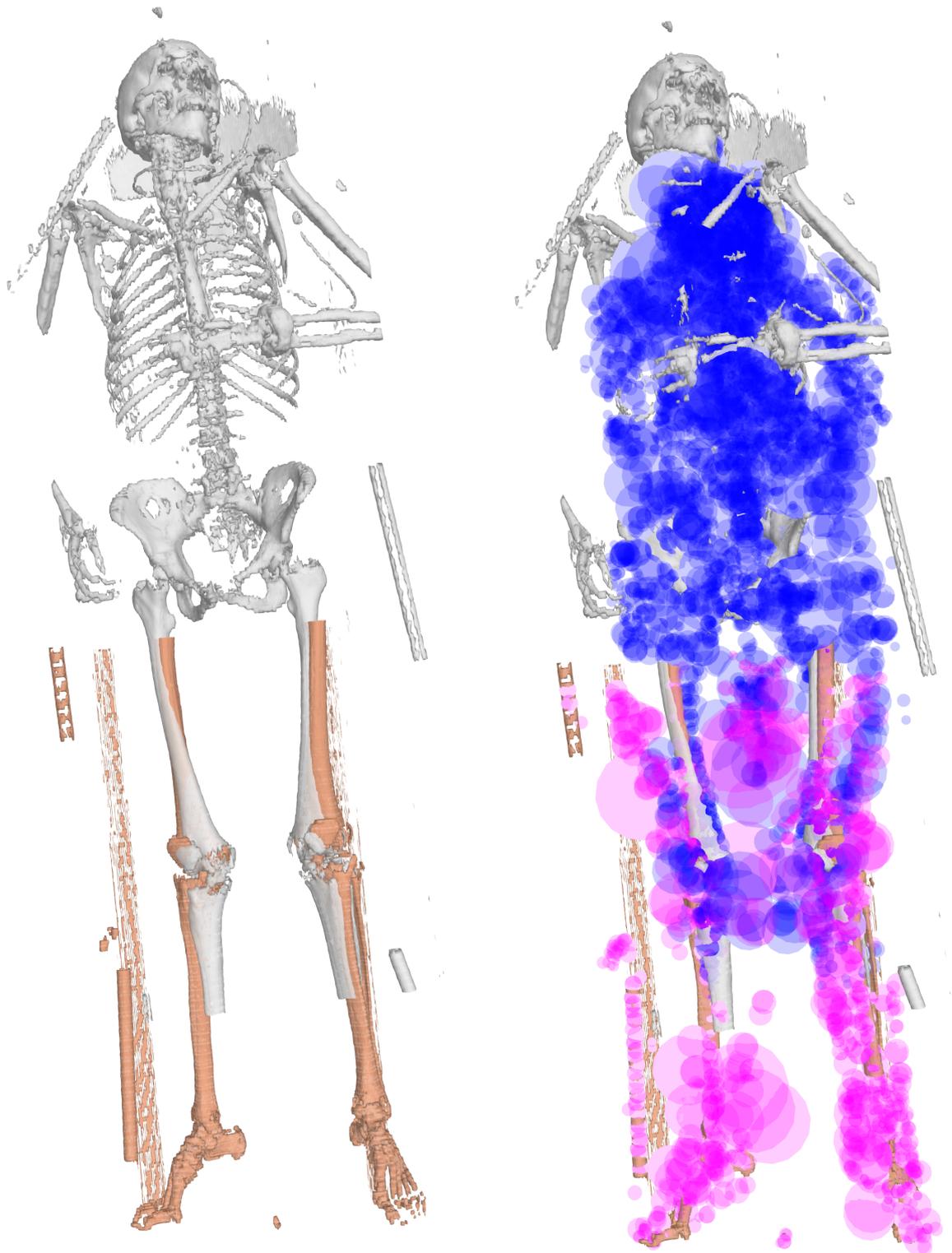
où w est la fraction de bons points, n le nombre de points nécessaires pour l'estimation d'un modèle et k le nombre d'essais. Lors de l'utilisation de RANSAC, w n'est généralement pas connu et k est à déterminer.

Le nombre de tirages (le nombre d'estimations de transformations) de RANSAC diffère donc selon le nombre de degrés de liberté. En effet, entre une transformation possédant 3 degrés de liberté et une en possédant 7, l'espace de recherche n'a pas la même richesse. Dans notre cas, le nombre de points pour estimer la transformation va de 1 dans le cadre d'une translation à 3 pour une transformation translation-rotation-échelle.

Il faut noter qu'en conditions réelles le conditionnement des échantillons doit être satisfaisant : réparti convenablement, ne pas être dégénéré (par des points alignés par exemple). Ce qui conduit à majorer le nombre de tirages nécessaires.

Une faiblesse de l'estimation des modèles incluant une rotation \mathbf{R} est qu'il fournit parfois des anti-rotations (une rotation avec une symétrie) détectées par un déterminant négatif : $\det(\mathbf{R}) = -1$. Ces anti-rotations ne correspondent pas à une réalité physique dans notre expérimentation, et doivent être interdites. Les transformations ainsi dégénérées sont simplement ignorées, seuls les tirages offrant des transformations réalistes sont considérés.

Nous avons choisi $k = 2000$ dans le cadre d'une transformation à 3 degrés de liberté et $k = 10000$ pour les transformations à 6 et 7 degrés de liberté.



(a) Images (A) (fig 3.5a) et (B) (fig 3.5c) recalés.

(b) Images (B)

FIGURE 3.6 – Exemple de mise en correspondance, avec les paires de points d'intérêt. Les points d'intérêt sont extraits de l'image, l'extraction du squelette n'est effectuée qu'à des fins d'affichage.

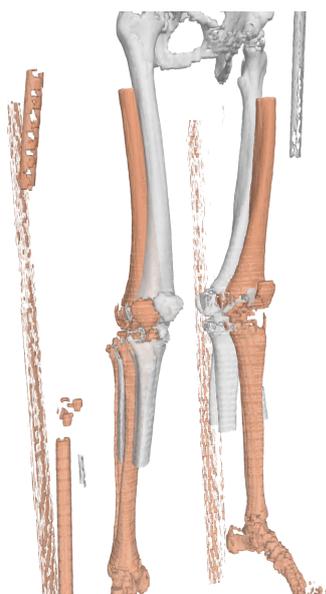


FIGURE 3.7 – Expérimentation FPR avec estimation uniquement de la translation. Nous pouvons observer la bonne estimation de la translation. Néanmoins, nous pouvons constater la limite du modèle, un défaut en rotation limitant la mise en correspondance.

Sur la figure 3.6, nous pouvons observer le recalage des squelettes des images A et B avec une similitude grâce à notre approche. Ce scénario est un scénario difficile pour les approches actuelles de mise en correspondance d'images car les espaces imagés ne sont pas les mêmes, la zone de recouvrement étant relativement petite par rapport à la dimension des images.

Avec $k = 10000$ tirages, RANSAC trouve la transformation (7 ddl) dont le résultat est visible figure 3.6, avec 124 vrais positifs. Un essai avec seulement la translation, visible figure 3.7, ne possède que 119 vrais positifs et nous pouvons observer sur la figure un défaut en rotation.

Nous avons 3062 points appairés initialement ce qui signifie un taux de vrais positifs de 4%, ce qui est extrêmement faible. La première explication est géométrique. Les zones imagées étant différentes, Les vrais positifs ne peuvent être situés que dans la zone de recouvrement. A contrario, l'ensemble des points d'intérêt peut créer des couples aberrants. Une seconde hypothèse est que la variabilité inter-patient dégrade la qualité des descriptions et nécessite l'abaissement des seuils afin d'avoir suffisamment de couples.

3.4.8 Conclusion

Plusieurs expérimentations avec FPR montrent que c'est un moyen efficace de mise en correspondance d'images. Nous pouvons noter que la richesse de la transformation influe sur la qualité du recalage, dans un compromis entre richesse du modèle et limite de l'estimateur.

L'avantage de cette méthode est sa rapidité : environ 1 minute pour extraire les points d'intérêt d'une image et quelques secondes pour appliquer la méthode RANSAC entre deux images. De plus, l'extraction des points d'intérêt ne doit être faite qu'une seule fois, ce qui permet d'économiser du temps de calcul dans le cas d'un scénario calculant de multiples mises en correspondance, et réutilisant donc les points d'intérêt.

Nous avons pu montrer l'efficacité de l'utilisation de points d'intérêt dans le cadre de l'imagerie médicale corps entiers, malgré la variabilité inter-patient. Il faut cependant être attentif au taux de vrais

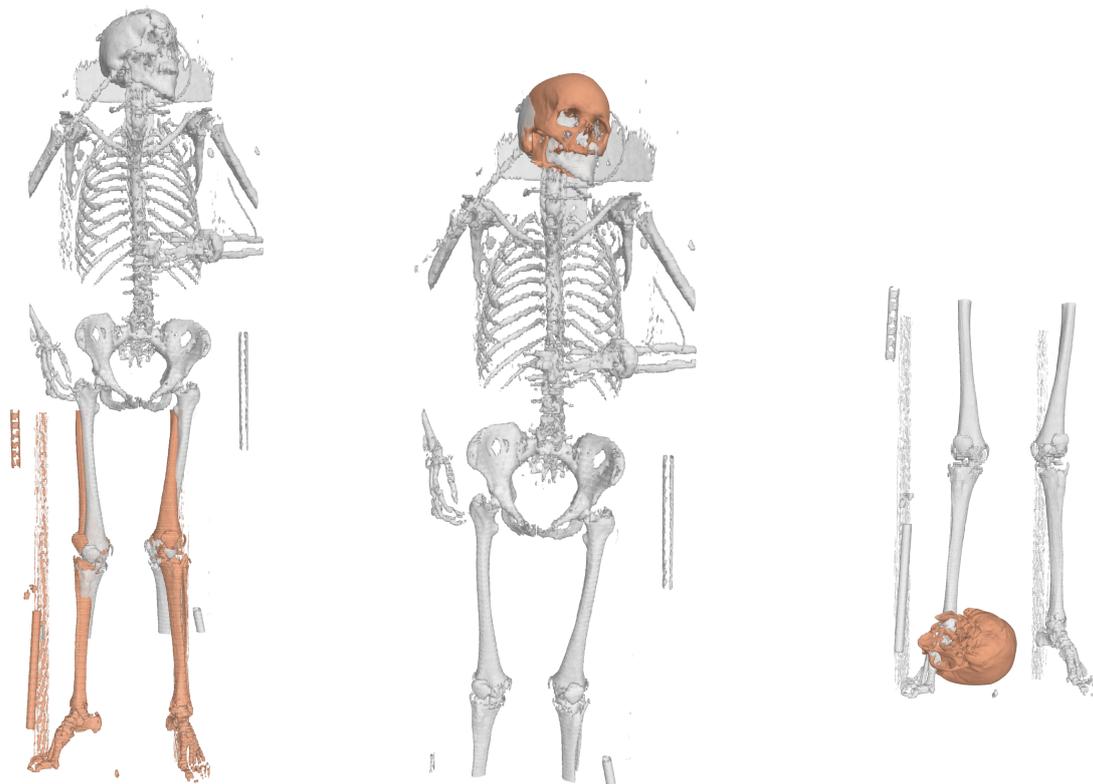


FIGURE 3.8 – Autre exemple de mises en correspondance FPR : On peut constater pour la troisième mise en correspondance, un mauvais recalage dû à l'absence de recouvrement entre les deux images.

positifs qui est extrêmement faible (4%) et peut poser des problèmes de convergence vers la bonne transformation dans le cadre de l'utilisation d'approches actuelles, ayant des hypothèses fortes vis-à-vis de ce taux.

3.4.9 Limites du recalage par paire

Hétérogénéité Dans l'objectif de ces travaux et de ce manuscrit en terme d'hétérogénéité, cette approche permet de mettre en correspondance de manière robuste des images provenant d'individus différents. Cependant, notre objectif est de pouvoir mettre en correspondance des images avec recouvrement partiel. La figure 3.8 gauche et milieu montre que l'on est capable de mettre en correspondance des images dont les parties imagées diffèrent. Mais comme attendu par une approche de recalage de paire, elle n'est pas capable de mettre en correspondance des images n'ayant pas de zones communes de recouvrement, figure 3.8 droite.

Richesse des transformations Cette expérimentation est limitée aux transformations rigides, ce qui peut sembler en deçà des approches déformables actuelles. Cependant, elle ouvre la porte à l'utilisation des points d'intérêt dans d'autres cadres, que nous allons approfondir dans les chapitres suivants.

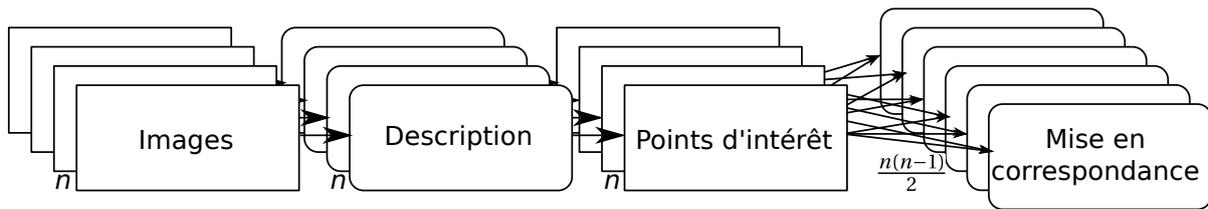


FIGURE 3.9 – Schéma d'extraction et de mise en correspondance des points d'intérêt de n images.

3.5 Mise en correspondance rigide de grands groupes d'images médicales

3.5.1 Une nouvelle représentation par graphe

Afin de répondre au besoin de dépasser le concept d'image de référence, et de pouvoir gérer les images partielles, nous proposons l'introduction d'une nouvelle approche par graphe complet, illustrée figure 3.10, en reprenant nos travaux des méthodes de mise en correspondance du chapitre 2.

Les approches "exhaustives" que nous avons présentées précédemment se proposent aussi de faire le lien entre toutes les images. Vu comme un graphe, il existe plusieurs chemins entre deux images, ce qui laisse une ambiguïté pour le choix de la transformation qui les relie. Nous levons ici cette ambiguïté en garantissant un chemin unique entre les images. Le graphe des transformations liant les images est un arbre recouvrant minimal. Nous avons donc l'avantage des deux mondes : exhaustif et en étoile.

Pour fusionner ces deux approches, il est nécessaire de faire la distinction entre deux types d'arêtes :

- **Les arêtes portant de l'information de recalage**, les données d'entrée, utilisées par le critère de similarité. C'est l'interprétation du graphe complet (exhaustif) de notre approche
- **Les arêtes portant une demi-transformation**, les données de sortie, utilisées pour déplacer les points. C'est l'interprétation par graphe en étoile de notre approche.

Une conséquence de l'utilisation d'un schéma différencié de connectivité est qu'il n'est pas nécessaire de construire une image de référence. Du point de vue de l'optimisation, cela revient à faire la distinction entre les sources d'observations et les variables à optimiser. Nous pouvons définir ce qui est "souhaitable" pour une telle structure d'optimisation :

- **Observations** : il est souhaitable d'avoir un nombre d'observations maximal, afin de bénéficier d'un maximum d'information pour contraindre l'optimisation de la meilleure manière possible vers la solution.
- **Variables** : il est souhaitable d'avoir un nombre de variables minimal, afin d'obtenir un système bien conditionné. Une contrainte intéressante est un graphe ne possédant pas de cycle (ce qui correspondrait à la présence de multiples chemins entre les paires d'images).

L'optimisation consiste à minimiser les différences entre toutes les paires d'images possibles. Le nombre de liens entre n images est égal au nombre d'arêtes dans un graphe complet, soit $\frac{n(n-1)}{2}$. Le nombre de transformations est égale à n , correspondant à une transformation entre chaque image et une référence abstraite. Nous définissons cette référence comme "abstraite" car elle ne contient pas d'image et est uniquement un intermédiaire de calcul.

3.5.2 Mise en correspondance des paires

En préliminaire à notre méthode, nous nous proposons d'utiliser la mise en correspondance par paires FPR décrite dans le chapitre 3 afin de calculer la mise en correspondance entre toutes les paires d'images possibles au sein du jeu de données considéré.

Ce processus se fait en deux temps, synthétisés dans la figure 3.9. La première étape consiste à extraire

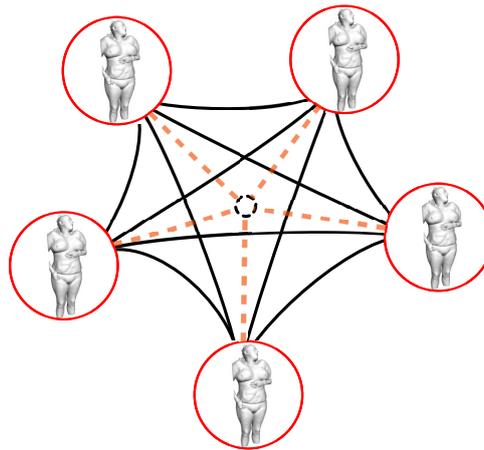


FIGURE 3.10 – Recalage “sans référence” par graphe complet. Les sommets représentent les images. Les arêtes sont de deux types : Les arêtes noires en traits pleins représentent les mises en correspondance de paires qui permettent de calculer les demi-transformations représentées par des pointillés oranges. Le sommet central, le cercle noir pointillé, est un intermédiaire de calcul pour composer les demi-transformations, que nous appelons “référence abstraite” car elle ne contient aucune image.

les points d’intérêt des n images considérées. À partir de cette étape, les images à base de voxels ne sont plus utilisées. La deuxième étape consiste à calculer les $\frac{n(n-1)}{2}$ mises en correspondances possibles entre les n images représentées par leurs points d’intérêt.

Nous utilisons la variante à 3 degrés de liberté de notre méthode FPR afin d’estimer une translation entre 2 images. Ainsi, la transformation suivante peut être estimée :

$$\mathbf{x}_k = \mathbf{x}_j + \mathbf{t}_j^k \quad (3.21)$$

pour l’ensemble des couples (j,k) possibles issus du jeu de données. Il en résulte un ensemble de $\frac{n(n-1)}{2}$ vecteurs \mathbf{t}_j^k décrivant l’ensemble des translations au sein de chaque paire d’images de la base.

3.5.3 Modèle de déformation

Tous les vecteurs de translations \mathbf{t}_j^k ne sont pas cohérents et l’objectif de notre méthode est de fournir une solution afin de proposer un ensemble de vecteurs de translation cohérents. Cette ensemble nous permettra ensuite de pouvoir transformer un point provenant d’une image quelconque dans le repère d’une autre image.

Nous illustrons notre approche avec un exemple de 3 images. Dans ce cas, il y a aussi 3 translations, mais c’est un cas particulier car pour $n > 3$ il y a plus de translations que d’images. Il n’est cependant pas souhaitable de faire un exemple avec plus de 3 images, pour des raisons de lisibilité.

La figure 3.11 fournit un schéma récapitulatif ainsi que les données de départ. Les flèches bleues représentent les recalages observés, tandis que les segments pointillés orange représentent les erreurs liées aux l’observations imparfaites. La matrice \mathbf{Y} représente les translations entre toutes les images, c’est notre vérité terrain. La matrice $\tilde{\mathbf{Y}}$ représente les transformations estimées par le processus RANSAC.

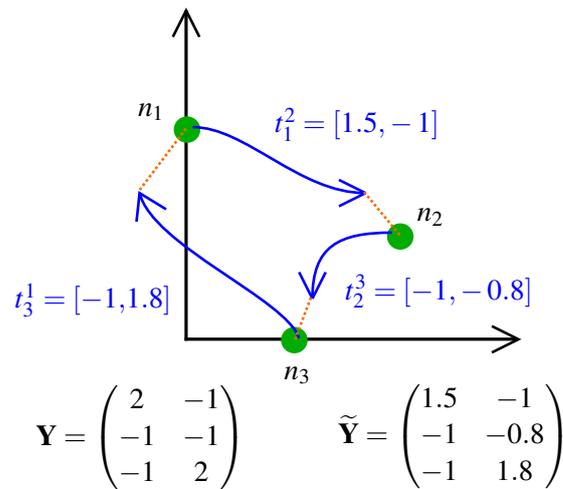


FIGURE 3.11 – Exemple de recalage rigide de groupes avec 3 images (points verts) et 3 translations.

3.5.4 Formalisation du problème

Énoncé du problème

L'objectif de notre approche est de passer d'un ensemble de transformations rigides entre toutes les images du jeu de données à une représentation cohérente, minimale et unique. Le problème revient à positionner dans un repère unique les sous-repères locaux des images.

Afin de déterminer de manière unique les positions des n images, il est nécessaire d'estimer $n - 1$ translations. En réalité, l'objectif de notre méthode est de trouver la position des images dans un repère arbitraire. Cette proposition est équivalente à trouver un vecteur de translation par image, soit n vecteurs au total, entre une origine abstraite et l'origine du repère local des images. Cette origine abstraite peut être la position moyenne des images ou une image que l'on positionne arbitrairement à des coordonnées prédéfinies.

Il faut noter que ce repère est purement arbitraire, d'où sa dénomination "abstrait" et l'approche ne possède pas de référence sous la forme d'une image, ce qui lève l'un des verrous scientifiques.

Équation utilisant l'opérateur laplacien

Nous proposons de résoudre ce problème en utilisant l'opérateur laplacien, voir [COHEN-OR et SORKINE \[2006\]](#), en utilisant les translations issues du graphe complet :

- à chaque nœud n_i correspond une image I_i , avec son repère local x_i inconnu qui va être ajusté au reste du jeu de donnée.
 - à chaque arête e_j^k correspond à la translation t_j^k entre deux images I_j et I_k , ce sont nos observations.
- Le problème consiste donc à trouver l'ensemble des repères locaux x_i en connaissant l'ensemble des translations t_j^k .

En théorie des graphes, une analogie peut être faite en relation avec l'information relative et absolue, basée sur une représentation matricielle du graphe :

$$E^t \cdot P = Y \tag{3.22}$$

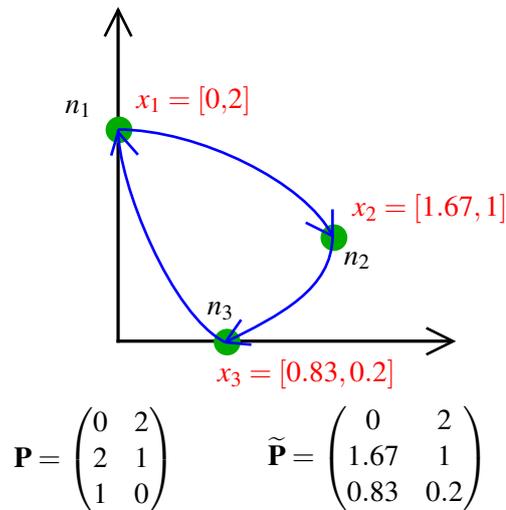


FIGURE 3.12 – Positions reconstruites en utilisant notre approche sur le problème décrit figure 3.11. Les flèches bleues représentent les recalages corrigés entre les images. La matrice \mathbf{P} représente les positions réelles. La matrice $\tilde{\mathbf{P}}$ représente les positions estimées.

où \mathbf{E} est la matrice d’incidence qui exprime la relation entre les sommets, elle possède un 1 et un -1 par ligne. \mathbf{P} est la matrice des positions des repères locaux dans le repère global, il contient l’ensemble des x_i , chaque ligne représentant un sommet, chaque colonne représentant une dimension (voir la figure 3.12). \mathbf{Y} fait référence à la matrice d’observations, chaque ligne contenant la translation portée par une arête e_j^k .

Dans notre cas, l’objectif est de trouver la meilleure estimation $\hat{\mathbf{P}}$ de \mathbf{P} étant donné une matrice d’observation $\tilde{\mathbf{Y}}$ imparfaite. Ce que nous proposons de faire en utilisant l’opérateur laplacien.

Pour un graphe donné, la matrice $\mathbf{L} = \mathbf{E}^t \cdot \mathbf{E}$ est connue sous le nom ‘matrice de l’opérateur laplacien’. Cette matrice est symétrique, singulière (avec une valeur propre nulle) et semi-défini positive.

En multipliant l’équation (3.22) par \mathbf{E} à gauche, la matrice de l’opérateur laplacien apparaît :

$$\mathbf{L}^t \mathbf{P} = \mathbf{E} \mathbf{Y} \tag{3.23}$$

Cette équation ne peut pas être résolue telle quelle parce que la matrice laplacienne $n \times n$ est toujours singulière, de rang $n - 1$. D’un point de vue géométrique, nous essayons de reconstruire des positions absolues en ne connaissant que des positions relatives. Il manque donc une ‘‘ancree’’ [COHEN-OR et SORKINE, 2006], car sans elle, on peut constater que déplacer tous les repères locaux de la même manière ne modifie pas leurs positions relatives. Afin de rendre la matrice laplacienne non singulière, nous fixons les coordonnées absolues d’un sommet du graphe, donc d’une image.

Nous introduisons donc \mathbf{L}^* , la matrice de l’opérateur laplacien augmenté d’une ligne et de la même manière \mathbf{Y}^* .

$$\begin{aligned} \mathbf{L}^* \mathbf{L}^t \mathbf{P} &= \mathbf{L}^* \mathbf{E} \mathbf{Y}^* \\ \mathbf{P} &= (\mathbf{L}^* \mathbf{L}^t)^{-1} \mathbf{L}^* \mathbf{E} \mathbf{Y}^* \end{aligned} \tag{3.24}$$

Nous résolvons l’équation (3.23) en utilisant une minimisation des moindres carrés, équation (3.24). La figure 3.12 illustre la solution associée au problème défini par la figure 3.11 avec une ancree fixée sur le sommet n_1 .

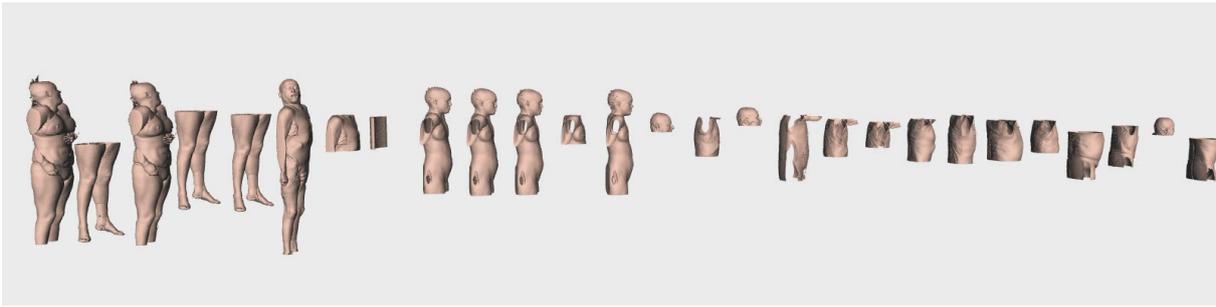


FIGURE 3.13 – Exemple avec 27 images scanner 3D. Les images contiennent différentes parties du corps. Nous avons aligné les images recalées sur un axe horizontal, afin de mettre en évidence le recalage dans l’axe Z vertical. Nous pouvons observer de mauvais recalages pour les parties non thoraciques.

3.5.5 Élagage du graphe et connexité

Problème En appliquant notre méthode sur un jeu de données composé de 27 images 3D obtenues par tomographie à rayon X qui imagent diverses parties du corps, visibles dans la figure 3.13, nous pouvons observer que la mise en correspondance échoue, particulièrement pour les images ne contenant pas le thorax. Ceci est dû à l’interaction entre les différentes parties de notre algorithme :

- **Les points d’intérêt** : L’extraction et surtout l’appariement des points d’intérêt n’est pas parfait, il y a donc de mauvaises mises en correspondance. Dans le cadre de notre approche, nous faisons toutes les mises en correspondance possibles. Si l’on met en correspondance des jambes et une tête, n’ayant donc aucune zone de recouvrement, l’ensemble des couples ne sera pas vide et il ne contiendra que des couples de points aberrants.
- **RANSAC** : Lorsque nous utilisons cet ensemble avec RANSAC, celui-ci trouve forcément une solution. Bien que cette solution possède le maximum de points considérés comme “bons” par l’algorithme, elle peut contenir très peu de couples de points appariés correctement. RANSAC fournit donc toujours une estimation de la transformation, celle-ci étant forcément fausse dans le cadre de l’estimation d’une transformation entre une tête et des jambes.
- **Equation laplacienne** : Comme toutes méthodes des moindres carrés classiques, la résolution des équations laplaciennes n’est pas robuste aux échantillons aberrants. Il suffit d’un mauvais échantillon, dans notre cas un vecteur de translation, pour biaiser la solution trouvée.

La contraction de toutes les images sur une ligne centrale que nous pouvons observer figure 3.13 s’explique par le fait que les mauvais recalages tendent à vouloir réaliser un chevauchement de toutes les images, en les regroupant.

Nécessité de l’élagage : la connexité du graphe Comme décrit au paragraphe précédent, RANSAC fournit toujours une solution. Cependant, il est possible d’utiliser un seuil sur le nombre d’appariements vérifiant le modèle pour déclarer une transformation non valide. Dans le cadre de la mise en correspondance de groupes d’images, cette solution ne garantit pas que le graphe des transformations est connexe. C’est à dire, qu’il existe au moins un chemin liant n’importe quelle paire d’images. Il est donc nécessaire d’effectuer un élagage a posteriori, en garantissant la connexité du graphe.

L’élagage de graphe a donc pour objectif de retirer les arêtes faisant le lien entre des images n’ayant pas de zones de recouvrement, comme par exemple entre une tête et des jambes, voir figure 3.14, tout en garantissant la connexité du graphe.

Critère de sélection

Nous avons testé plusieurs critères de sélection, que l’on pourrait classer en deux catégories. Le premier type est une catégorie d’indicateurs que nous appelons “intrinsèques”, car ils n’utilisent que des

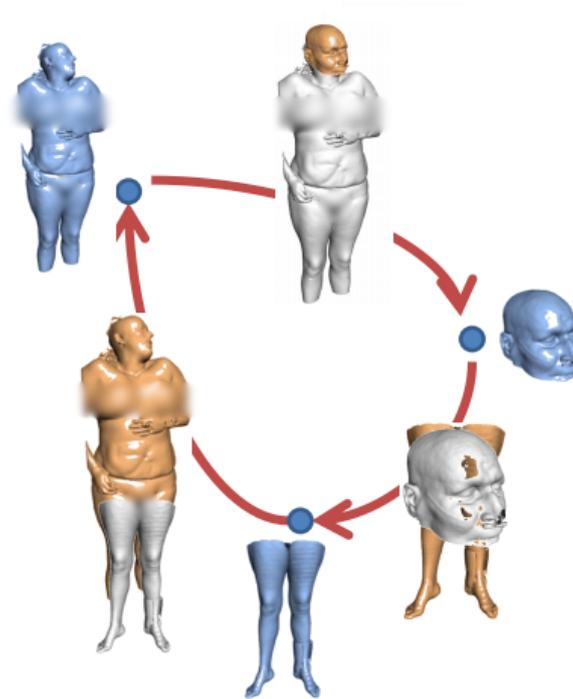


FIGURE 3.14 – Exemple problématique de mise en correspondance avec 3 images hétérogènes : des jambes, une tête et un corps entier. Il y a donc un recalage, entre la tête et les jambes, qui doit être retirée.

informations issues du recalage pour trier les arêtes :

- **Taille de l'intersection des boîtes englobantes (BE)** : Les boîtes englobantes des volumes sont calculées à partir des images. L'intersection des boîtes englobantes des deux images est calculée :

$$BE_1 \cap BE_2 \quad (3.25)$$

L'objectif de ce critère est de rejeter les arêtes liant des images n'ayant pas ou ayant un très faible recouvrement.

- **Taille relative de l'intersection des boîtes englobantes** : Cette fois-ci, c'est la taille relative de l'intersection par rapport à l'union de celle-ci :

$$\frac{BE_1 \cap BE_2}{BE_1 \cup BE_2} \quad (3.26)$$

Cette mesure permet, comme la précédente, de rejeter les arêtes ayant un très faible recouvrement. L'avantage de cette mesure est d'être relative à la taille des images, de la même manière que la mesure de similarité DICE [ZOU et collab., 2004].

- **Vérification de la cohérence des cycles** : Si une image est mal recalée avec une autre, ce recalage doit pouvoir être détecté en prenant un cycle du graphe et en composant les translations, le résultat de cette composition ne sera pas un vecteur nul :

$$\text{exemple : } \left\| t_j^k \circ t_k^l \circ t_l^j \right\| \quad (3.27)$$

L'objectif de cette mesure est de détecter et supprimer les arêtes participants à des erreurs de cohérence de cycles.

Les critères intrinsèques ne permettent pas de détecter les bonnes arêtes. Le critère et le recalage des images interfèrent, ne permettant pas la convergence de l'élagage. Par exemple, le calcul des boîtes englobantes étant effectué après une étape de mise en correspondance, il dépend du recalage. L'algorithme a ainsi déjà divergé quand le critère est utilisé, il retire donc les mauvaises arêtes.

L'inefficacité des critères intrinsèques nous ont conduit à la recherche et l'utilisation de critères "extrinsèques", ne dépendant pas de la mise en correspondance. Ce critère doit donc dépendre des étapes précédentes de notre algorithme, c'est à dire l'extraction de points d'intérêt, l'appariement des points d'intérêt ou RANSAC. Il est caractérisé par :

- **Nombre de couples de points d'intérêt.** Ce nombre, bien que décrivant une similarité entre 2 images, possède des biais qui l'empêchent d'être efficace. Le principal biais est lié à la taille de l'image. Plus l'image est grande, plus le nombre de points d'intérêt détectés est grand et par conséquent le nombre de couples, même si ce sont des couples aberrants. Nous avons observé qu'aucune normalisation par la taille des ensembles, ou directement par le volume des images, n'est efficace.
- **Rapport bons points / points aberrants de RANSAC.** C'est le critère extrinsèque le plus intuitif. Après avoir appliqué RANSAC, on calcule le rapport entre les couples consistants avec la transformation estimée et les couples non cohérents. Ce critère a le défaut de ne pas prendre en compte le nombre de couples. Ceci peut être inadéquat lorsqu'il y a peu de couples parce que ce rapport peut être artificiellement élevé si, par exemple, il y a 5 points respectant le modèle (ce qui est très peu), mais provenant d'un ensemble de 15 couples (typiquement issus de petites images).
- **Nombre de couples de points en accord avec la transformation estimée par RANSAC.** Ce critère est préférable au rapport bons points / points aberrants décrit ci-dessus, car il permet de prendre en compte la taille des ensembles.

Nombre d'arêtes restantes

Il reste à définir le nombre d'arêtes conservées après leur élagage. Le critère retenu étant extrinsèque, le processus n'est plus itératif, et la méthode peut être appliquée en 2 étapes :

1. Suppression des arêtes,
2. Calcul des positions par résolution de l'équation utilisant le laplacien.

Dans ce schéma, le critère d'arrêt consiste à choisir le nombre d'arêtes que nous conservons. La première stratégie que nous avons développée consiste à contrôler les résidus des moindres carrés en retirant arête par arête, ce qui oblige à recalculer la solution à chaque étape. Cependant, nous avons observé que choisir un nombre d'arêtes cible dès le début est aussi efficace, tout en étant plus rapide.

Sachant que le nombre d'arêtes d'un graphe complet à n sommets est $\frac{n(n-1)}{2}$ et que le nombre de positions à déterminer est de n , le problème est surdéterminé et l'est de plus en plus au fur et à mesure que le nombre d'images augmente.

Après plusieurs expérimentations, les résultats montrent qu'un nombre d'arêtes final proportionnel au nombre d'images est satisfaisant. Dans nos expérimentations, nous avons choisi un nombre d'arêtes final égal à $3n$.

Discussion Ce critère d'arrêt ne semble pas optimal. En observant les valences des sommets du graphe lors de nos expérimentations, nous avons remarqué une répartition hétérogène des arêtes, de nombreux sommets n'étant reliés au graphe que par une seule arête.

3.5.6 Extension à une translation et un facteur d'échelle

La méthode présentée ci-dessus est restreinte à la translation. Ceci est dû à la nécessité d'avoir une écriture soustractive du lien entre deux sommets, sous la forme :

$$\mathbf{t}_j^k = \mathbf{x}_k - \mathbf{x}_j \quad (3.28)$$

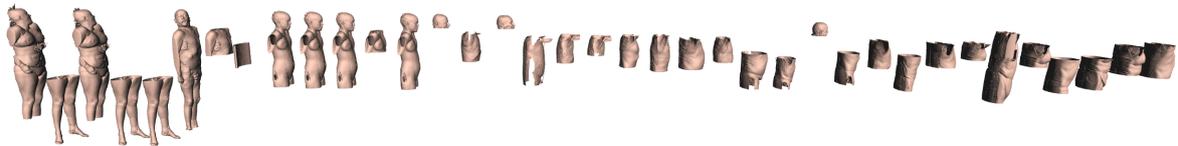


FIGURE 3.15 – Exemple de la figure 3.13, en utilisant l'élagage du graphe. Nous pouvons observer la bonne mise en correspondance des images ayant des positions éloignées du centre géométrique.

où les vecteurs des positions \mathbf{x} sont de la forme (x, y, z) .

Une extension intéressante pour les applications visées est l'estimation d'un facteur d'échelle. Ce facteur d'échelle peut facilement être estimé entre deux images en ajoutant un quatrième degré de liberté à notre méthode FPR. Le facteur d'échelle est un facteur multiplicatif, entre les deux repères locaux des images que l'on peut traduire par :

$$\mathbf{x}_j = s_j^k \mathbf{x}_k \quad (3.29)$$

où s , le facteur d'échelle, est un scalaire.

En prenant le logarithme de cette équation, nous avons un facteur additif :

$$\ln(\mathbf{x}_j) = \ln(s_j^k) + \ln(\mathbf{x}_k) \quad (3.30)$$

soit

$$\ln(s_j^k) = \ln(\mathbf{x}_k) - \ln(\mathbf{x}_j) \quad (3.31)$$

De cette manière, nous pouvons estimer un facteur d'échelle par image, qui, dans ce cas, représente une valeur absolue par rapport à une taille normalisée. Cette taille normalisée revient au même concept que la notion d'ancre utilisée pour la position.

Pour estimer un facteur d'échelle, il suffit donc d'utiliser un vecteur \mathbf{x} de la forme $(x, y, z, \ln(s))$.

3.6 Algorithme proposé et résultats expérimentaux

3.6.1 Fast Groupwise Registration : FGR

"Fast Groupwise Registration" (FGR) est le nom que nous attribuons à la méthode décrite précédemment, elle permet la mise en correspondance d'un ensemble d'images. Dans notre expérimentation, nous utilisons une base de données possédant 400 images. Nous avons déjà observé sur la figure 1.1 présenté en introduction que cette base n'est pas cohérente.

Données

La base de données utilisée est constituée d'images 3D obtenues à partir de tomographies à rayon X. Les images possèdent des dimensions différentes, des résolutions différentes et, le plus important, des parties imagées différentes. La base est composée de 400 images pour une taille totale de 102Go.

Les calculs sont effectués sur une station de travail possédant 24 cœurs et 128Go de mémoire vive. nous utilisons la plateforme DESK [JACINTO et collab., 2012] pour les calculs et l'affichage.

Algorithme FGR

La figure 3.16 présente un schéma récapitulatif de l'algorithme FGR. L'algorithme suit les étapes suivantes pour n images :

1. **Extraction des points d'intérêt**, nous utilisons un seuil sur le déterminant de la hessienne de 8000 sur des images ré-échantillonnées avec une taille de voxel de $1.5mm^3$. (n opérations)
2. **Mises en correspondance** des points d'intérêt, entre toutes les paires possibles d'images au sein de l'ensemble. Nous utilisons une distance maximale au plus proche voisin de 0.3 et un rapport au second de 0.98 (dans l'espace des caractéristiques). ($\frac{n(n-1)}{2}$ opérations)
3. **Estimations des paramètres** de toutes les transformations avec RANSAC. Nous faisons 6000 tirages aléatoires, avec un seuil de décision de distance de $40mm$, afin de trouver la meilleure transformation, et cela pour chaque paire d'images. ($\frac{n(n-1)}{2}$ opérations)

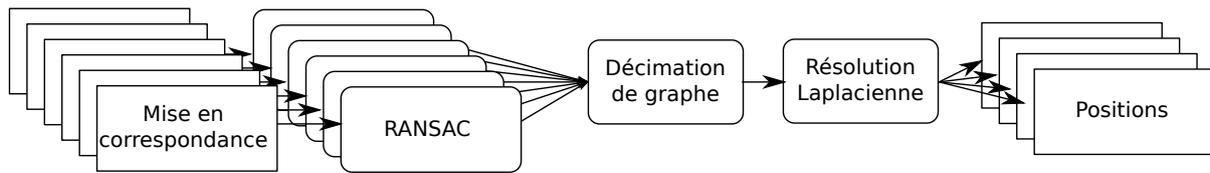


FIGURE 3.16 – Schéma de l'approche FGR en prenant en entrée l'ensemble des mises en correspondance et en sortie l'ensemble des positions recalées.

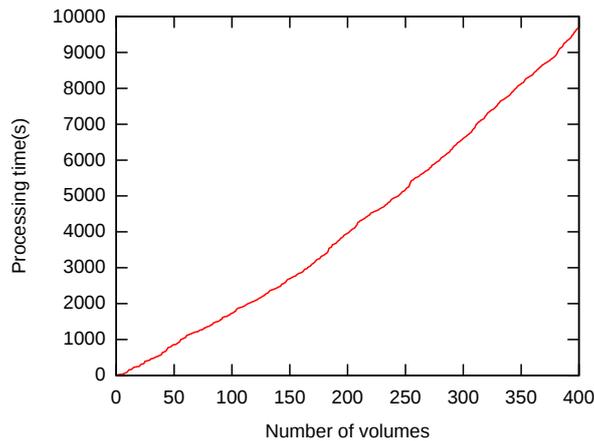


FIGURE 3.17 – Temps de calcul en fonction du nombre d'images.

4. **Élagage du graphe** des mises en correspondance en retirant des arêtes selon le critère du nombre de "inliers" de RANSAC. Comme décrit précédemment, nous conservons $3n$ arêtes, tout en garantissant la connexité du graphe.
5. **Résolution des équations laplaciennes**, en utilisant une ancre sur une image dont nous fixons arbitrairement les coordonnées à $\mathbf{0}$.

Résultats

L'extraction des points d'intérêt SURF montre toute son importance ici, par une représentation très compacte. Une image qui pèse entre 200 et 600Mo peut être réduite à une description SURF de 0.1 à 1 Mo. La complexité de l'extraction des points d'intérêt SURF est linéaire avec le nombre d'images, et l'extraction peut facilement être parallélisée. Ce processus de réduction du volume des données nous permet de passer de 102Go de données images à seulement 80Mo de description pour 400 images.

De plus, SURF n'est exécuté qu'une fois par image, chaque extraction de SURF est donc utilisée n fois. Pour une acquisition corps entier classique, $1.8m \times 0.6m \times 0.6m$, ré-échantillonnée avec une taille de voxels de 1.5mm, approximativement 5000 points d'intérêt sont extraits en environ 6 secondes. Grâce à cette représentation compacte, RANSAC ne met que quelques secondes pour estimer une transformation entre deux images.

À l'origine, les points d'intérêt ont été conçus pour mettre en correspondance deux images représentant la même scène avec des points de vue différents. Le principal critère d'évaluation de la performance de l'utilisation des points d'intérêt est la répétabilité, soit la capacité d'un descripteur à retrouver deux fois le même point dans des images différentes et à le mettre en correspondance. Dans notre cas, la répétabilité est moins significative car nous mettons en relation des images provenant de patients différents, n'ayant pas forcément les mêmes structures anatomiques et présentant une grande variabilité.

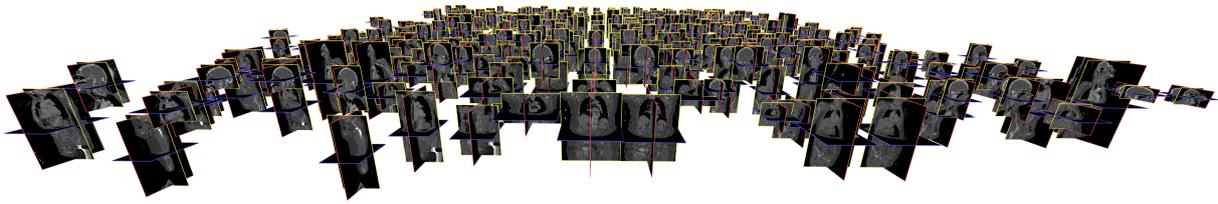


FIGURE 3.18 – La même vue que la figure 1.1 après recalage. Nous pouvons maintenant constater la cohérence des images entre elles.

La figure 3.17 montre le temps d'exécution de notre méthode pour un ensemble dont le nombre d'images varie de 0 à 400. Nous pouvons observer un léger comportement quadratique, même si celui-ci n'est pas encore dominant à cette échelle. L'extraction des points d'intérêt représente le principal coût en temps de calcul.

La figure 1.1 de l'introduction montre l'ensemble des images avant mise en correspondance. La figure 3.18 présente le résultat après application de notre méthode. Nous pouvons observer que l'ensemble des images est bien recalé, en particulier les têtes et les jambes seules. Notre mise en correspondance génère une base de données cohérentes qui peut être utilisée pour d'autres applications.

3.6.2 Masquage des yeux

Afin d'illustrer l'intérêt d'une base cohérente, nous avons développé une application de masquage des yeux de manière automatique. Celle-ci a pour but d'anonymiser les patients, ce qui est contraignant à faire manuellement avec de grandes bases d'images.

Avec la méthode que nous développons ici, la seule action manuelle est de placer une boîte englobante qui délimite la tête dans l'une des images, et une autre au niveau des yeux. La propagation à l'ensemble des autres patients est obtenue automatiquement par le procédé suivant :

1. Positionnement manuel d'une boîte englobante sur la tête dans une image.
2. Application de notre méthode de mise en correspondance de grand groupes.
3. Extraction des images contenant une tête, par propagation de la boîte englobante sur l'ensemble des images, ce qui peut s'apparenter à de la détection d'organe, et qui est effectuée dans l'esprit de la segmentation basée atlas ;
4. Application d'une deuxième passe de notre méthode de mise en correspondance de grands groupes. Pour celle-ci, nous n'utilisons que les points d'intérêt situés au niveau de la tête, ce qui accroît la précision.
5. Projection du pré-positionnement des yeux sur toutes les images contenant des yeux en utilisant les transformations de l'étape précédente.
6. Application d'un bandeau noir à la position décrite.

Le recalage utilisant seulement l'ensemble partiel des points situés au niveau de la tête permet une amélioration de la précision. Conceptuellement, cette mise en correspondance limitée à une zone peut être vue comme une première étape vers le recalage déformable, apparenté au recalage affine par morceaux présentés au chapitre 2. Le résultat est présenté figure 3.19. On peut noter que les yeux sont bien couverts, exception faite des patients dont la tête est tournée. Ceci est directement lié au modèle de déformation choisi, celui-ci n'étant pas assez riche pour décrire une rotation locale.

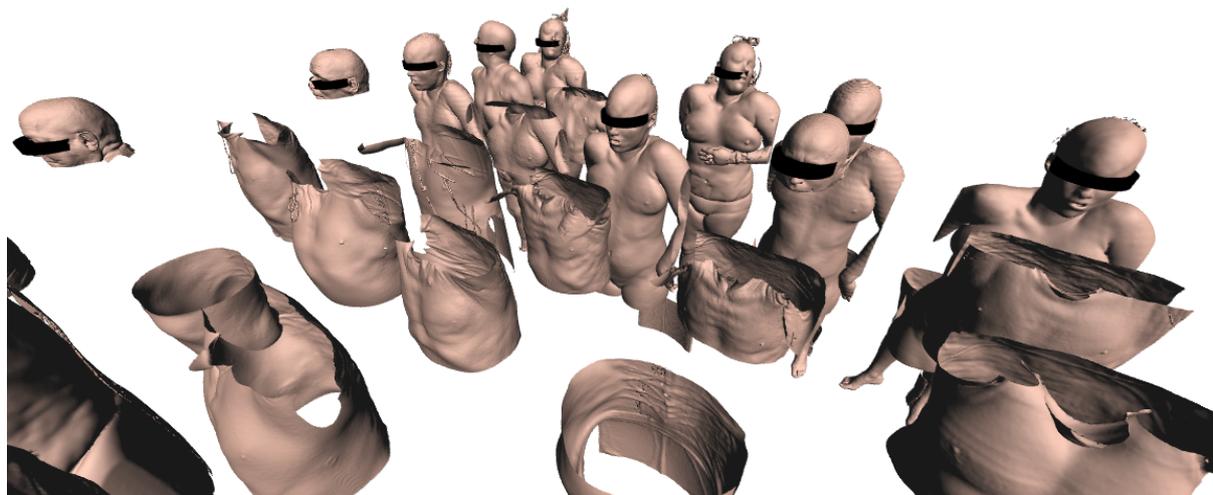


FIGURE 3.19 – Masquage des yeux appliqués à des images CT. Note : les maillages sont extraits uniquement dans le but de visualiser les résultats, seules les images 3D sont utilisées dans la méthode.

3.7 Conclusion

Nous proposons une méthode de recalage permettant la mise en correspondance rigide de grands groupes d'images médicales sans référence. Nous avons montré l'intérêt de cette méthode en matière de rapidité et de robustesse, particulièrement dans le cadre de l'utilisation de jeux de données médicales contenant des ensembles disjoints. Cette approche à faible empreinte mémoire, grâce à l'utilisation de points d'intérêt, permet de gérer de larges jeux de données, de manière très rapide.

Notre approche est un moyen efficace de gestion de grandes bases de données en utilisant un modèle translation-échelle simple. Ce modèle de transformation est une première étape qui permet d'apporter une cohérence à des bases de données d'images, pour des applications dites de "screening".

Chapitre 4

Contribution 2 : mise en correspondance déformable de grands groupes

Sommaire

4.1	Introduction	77
4.2	Challenges	77
4.2.1	Quantité de données	77
4.2.2	Répétabilité et appariements aberrants	78
4.2.3	Complexité calculatoire	78
4.3	Contributions	79
4.3.1	Recalage sans référence par points	79
4.3.2	Pondération-EM	79
4.3.3	Méthode d'optimisation efficace	80
4.4	Appariement des points d'intérêt	80
4.4.1	Méthode	80
4.4.2	Discussion	80
4.5	Demi-transformées guidées par des points d'intérêt	81
4.5.1	Introduction	81
4.5.2	Données d'entrée	81
4.5.3	Choix des demi-transformées	82
	Fonctions de base radiales	82
	B-Splines	83
4.5.4	Approche pyramidale	84
4.5.5	Attache aux données	85
4.6	Statistiques robustes	86
4.6.1	M-estimateur de Tukey	86
4.6.2	Pondération-EM	87
	Paramétrisation par image	89
	Paramétrisation locale	90
	Limite de localité	90
4.6.3	Test avec des données synthétiques	91
4.6.4	Test avec des données réelles	92
4.6.5	Comparaison	93
4.7	Optimisation	94
4.7.1	Notation matricielle	95
4.7.2	Moindres carrés pondérés itératifs	97

	Itération de Gauss-Newton	97
	Descente de gradient	98
4.8	Régularisation	98
4.8.1	Introduction	98
4.8.2	Protocole expérimental : données synthétiques	99
	Génération des points d'intérêt	101
	Génération de champs de déformation	101
	Génération de la liste des couples	101
	Exemple synthétique	102
4.8.3	Régularisation du déplacement des points de contrôle	103
4.8.4	Régularisation sur les couples appairés	104
4.8.5	Contrainte inter-images sur les points de contrôle	105
4.9	Conclusion	108

4.1 Introduction

Nous avons vu au chapitre précédent que les points d'intérêt pouvaient être utilisés à des fins de mise en correspondance d'images médicales possédant une variabilité élevée. Nous avons proposé une méthode permettant le recalage rigide d'un ensemble d'individus. Ce nouveau chapitre a pour objectif d'étendre l'utilisation des points d'intérêt pour la mise en correspondance déformable d'images, en estimant des champs de déformation. Ces travaux font l'objet d'une publication en cours, soumise en janvier 2017 [AGIER et collab. \[2017\]](#).

La mise en correspondance déformable de masses de données est un verrou scientifique [[XU et collab., 2016](#)]. Le recalage rigide permet la détection d'organes et permet de rendre spatialement cohérente une base de données. La mise en correspondance déformable va plus loin et permet un recalage beaucoup plus riche. Après un recalage déformable, il donc est possible de reporter les coordonnées d'un voxel sur un autre. Autrement dit, nous sommes capables de comparer l'environnement local des organes de plusieurs individus. En pointant par exemple la bifurcation de la trachée d'un individu, nous sommes capables de la comparer aux bifurcations des trachées des autres individus, en appliquant les champs de déformation.

Ceci est déjà largement utilisé dans le cadre de la segmentation basée atlas [[IGLESIAS et SABUNCU, 2015](#); [KÉCHICHIAN et collab., 2014](#)], qui permet de reporter des segmentations de plusieurs images (l'atlas) sur une nouvelle image à segmenter. Une fois les segmentations reportées dans le repère cible, un système de vote est utilisé pour construire la segmentation finale. Ce type d'approche utilise le recalage de plusieurs images pour transporter les segmentations, cette opération est effectuée sur de petits ensembles d'images (une vingtaine d'éléments dans l'atlas) en utilisant des recalages par paires.

L'usage croissant de techniques d'apprentissage montre aussi des besoins en bases d'apprentissage et de tests de grande envergure. Pour construire ces bases, des méthodes de recalage déformable permettant de gérer de grands ensembles, supérieurs au millier d'images, sont nécessaires.

4.2 Challenges

Nous proposons une méthode unifiée permettant de mettre en correspondance plusieurs centaines d'images. Cette approche peut être vue comme une approche par graphe complet, de manière similaire à notre contribution du chapitre précédent. Dans un premier temps, cette approche se limite à des images contenant les mêmes zones imagées. Plusieurs verrous scientifiques sont à prendre en considération et doivent être dépassés.

4.2.1 Quantité de données

Notre jeu de données de tests contient, selon les expérimentations de 20 à plus de 100 images, chaque image étant composée d'environ 512^3 voxels, pour un total qui peut représenter plus de 100Go de données. Les approches actuelles traitent généralement beaucoup moins de données, de l'ordre de la vingtaine d'images, avec des tailles d'images plus petites. En plus d'une approche sans référence, il est nécessaire de développer une méthode légère et efficace permettant de ne pas conserver les voxels et d'éviter l'emploi d'un champ de déformation dense.

La méthode décrite au chapitre précédent n'avait besoin que de quelques points d'intérêt pour décrire une transformation car elle est rigide, les autres permettant juste de valider ou d'invalidier l'estimation. Dans le cas d'une mise en correspondance déformable, la richesse en points d'intérêt définit directement la richesse du champ de déformation que l'on pourra estimer.

De plus, contrairement à l'approche précédente, le champ de déformation défini pour chaque image est lui-même générateur d'un grand nombre de données intermédiaires, car nous passons d'un vecteur de quelques composantes à un champ de déformation riche. Il semble évident qu'en terme de quantités de données, nous ne pouvons pas utiliser des champs denses, par voxel, dont la taille est similaire à la taille des images. Les images faisant 512^3 voxels, le champ de déformation d'une image nécessiterait 1.5Go d'espace mémoire.

4.2.2 Répétabilité et appariements aberrants

Nous devons trouver un moyen de gérer la variabilité inter-patients. Dans le cadre de la vision par ordinateur, lorsque les points d'intérêt sont utilisés pour décrire des scènes identiques avec des points de vue différents, nous avons observé que le nombre d'appariements aberrants est de l'ordre de 20%, en dépassant très rarement les 40%. Nous avons observé que dans le cadre de l'imagerie médicale, la quantité d'appariements aberrants est de 30% pour un même patient et plus de 70% pour des patients différents.

Deux raisons sont à l'origine de ce phénomène, toutes deux induites par la variabilité inter-patients :

1. La sélectivité d'appariement des points d'intérêt doit être abaissée, afin de prendre en compte les variations dans la structure d'une même zone chez deux patients différents.
2. La consistance entre les couples ne peut pas être un critère de décision (La consistance implique que si trois points A, B, C provenant d'images différentes sont liés par deux appariements : AB et BC, A et C doivent correspondre).

La conséquence est que le nombre d'appariements aberrants dépasse la capacité de gestion des M-estimateurs, technique classiquement utilisée. Il est donc nécessaire de créer un estimateur plus robuste.

4.2.3 Complexité calculatoire

L'analogie avec l'AF n'est pas parfaite. L'AF utilise un ensemble d'images 2D distribuées dans un espace 3D. Chaque image n'a des zones de recouvrement qu'avec un nombre réduit d'autres images. L'AF rentre dans la classe des problèmes parcimonieux, et le problème d'optimisation se simplifie par la présence de matrice à bandes.

Dans notre cas, nous mettons en correspondance des images 3D dans un espace 3D et la zone de recouvrement entre les images est beaucoup plus grande : toutes les images sont localisées sur le corps humain, mais à travers différents patients. Ceci résulte en un problème dense, illustré par la matrice de la figure 4.1.

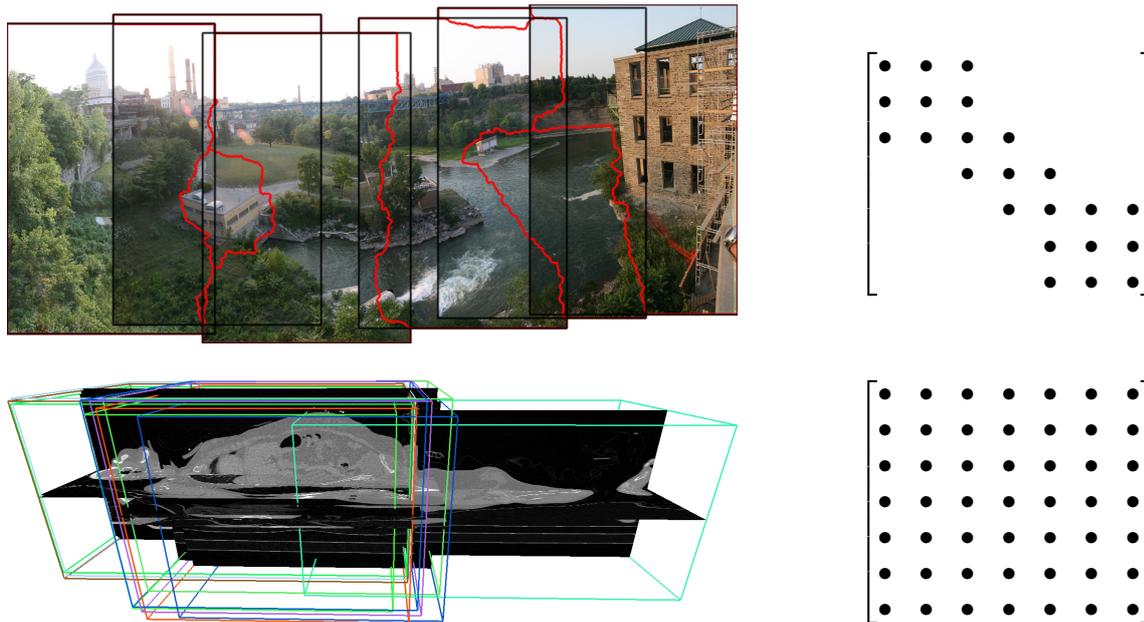


FIGURE 4.1 – Exemple de la parcimonie induite par la reconstruction d’un panorama (en haut, par ajustement de faisceaux), comparé à notre application (en bas). Les matrices de recouvrements sont données à droite. Crédit de l’image supérieure : Noso1 via Wikimedia Commons.

4.3 Contributions

Afin d’apporter une solution pour résoudre ces problèmes, nous proposons les trois contributions ci-dessous.

4.3.1 Recalage sans référence par points

Nous n’utilisons que les points d’intérêt issus des images, sans jamais réutiliser les images complètes, les voxels (données d’entrée). Nous utilisons toujours un repère commun abstrait et des demi-transformées comme pour notre approche FGR. Cette approche est une mise en correspondance de groupes avec un graphe complet, mais sans référence : l’optimisation est directement dirigée par les mises en correspondance inter-images. Nous faisons usage d’un graphe en étoile uniquement pour les demi-transformations entre chaque image et un espace commun abstrait. Les demi-transformations sont modélisées par des pyramides compositionnelles de splines et nous utilisons un graphe qui met en relation toutes les images (un graphe complet) pour l’optimisation. Cette méthode compacte permet de mettre en correspondance 100 images soit un total supérieur à 100Go avec une empreinte mémoire inférieure à 10Go.

4.3.2 Pondération-EM

Afin d’être capable de gérer un taux d’appariements aberrants supérieur à 50%, nous avons développé un algorithme d’espérance-maximisation qui permet d’estimer explicitement les paramètres des distributions des appariements qui sont de vrais positifs ou de faux positifs (aberrants). De plus, les résultats expérimentaux montrent que notre approche par pondération EM a une meilleure sélectivité vis-à-vis des appariements aberrants que les M-estimateurs. Dans notre expérimentation, seule la pondération EM permet la convergence de notre méthode.

4.3.3 Méthode d'optimisation efficace

Le problème d'optimisation globale que nous posons est un problème de minimisation d'une fonction d'énergie dense et un problème de grande dimension. Nous avons développé une méthode séparable et rapide dédiée à notre problème. Avec cet algorithme, nous sommes capable de mettre en correspondance 20 images 3D en 75 minutes avec notre plateforme de test (24-cœurs, 32Go RAM).

4.4 Appariement des points d'intérêt

L'appariement des points d'intérêt est la première étape avant toute optimisation. Contrairement à notre approche rigide, nous n'estimons pas un modèle entre chaque paire d'images. L'optimisation globale utilise directement la liste des couples de points appairés entre toutes les paires d'images.

Cet appariement est fait de manière naïve, en comparant tous les points entre eux. Elle relève donc d'une complexité quadratique qui est non négligeable dans le temps d'exécution de notre approche.

4.4.1 Méthode

Considérons une paire d'images A et B , et un point d'intérêt p_A de l'image A . L'appariement consiste à trouver les deux plus proches voisins (p_B^1 et p_B^2) de p_A dans l'images B et à décider, avec les techniques décrites au chapitre précédent, si p_A et p_B^1 sont considérés comme appairés.

Il peut être vu comme un problème de plus proches voisins dans l'espace des caractéristiques. Malheureusement, aucune structure de données et approche algorithmique n'a permis d'accélérer la recherche. Ceci est principalement dû au fait que la recherche se fait dans l'espace de caractéristiques qui possède 48 dimensions dans le cas des points d'intérêt SURF 3D que nous avons développés.

La recherche de plus proches voisins en grandes dimensions est un problème connu, qui reste un verrou mathématique.

Nous avons utilisé les méthodes suivantes.

- **Méthode naïve** : La méthode naïve, aussi appelée force brute, consiste à ne pas utiliser de structure particulière et à simplement itérer sur l'ensemble des candidats.
- **Kd-tree BENTLEY [1975]** : une k -grille régulière possède k^d cellules, et dans l'application des plus proches voisins, il convient d'itérer sur au moins toutes les grilles voisines, soit 2^d grilles, soit $2^{48} = 10^{14}$ à mettre en relation des 10^3 à 10^5 points d'intérêt sur lesquelles nous itérons avec la méthode naïve.
- **Approximate Nearest Neighbour** : Les méthodes de plus proches voisins approximatifs permettent aussi une correspondance exacte.
- **Ball-tree OMOHUNDRO [1989]** : Les ball-tree ont été développés en réponse à la limite des kd-tree en grande dimension. Ils utilisent l'inégalité triangulaire $|x + y| < |x| + |y|$ afin de répartir les points dans des hypersphères et permettre une recherche rapide du plus proche voisin.

Les expérimentations ont été faite en Python, avec la bibliothèque "scikit-learn", sur des données synthétiques. Dans notre cas, l'approche naïve est plus rapide.

4.4.2 Discussion

À l'échelle actuelle, nous avons quelques centaines d'images et une dizaine de milliers de points par image, ce qui représente $(10^3 \times 10^4)^2 = 10^{14}$ comparaisons. Une implémentation optimisée, parallélisée

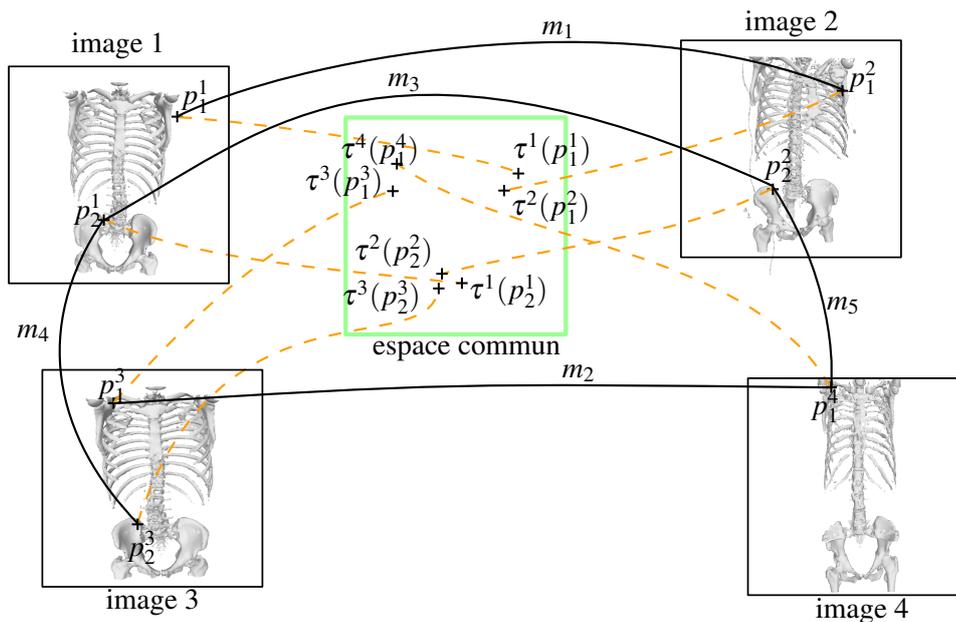


FIGURE 4.2 – Exemple avec 4 images. Pour des raisons de lisibilité, seul le squelette est affiché. Les appariements sont tracés en noir et la transformation des points en pointillés orange.

sur l'ensemble des 48 cœurs de notre plateforme de test et utilisant des instructions SIMD nécessite 12 heures, ce qui est convenable.

En comparaison, l'extraction des points d'intérêt, qui a une complexité linéaire en fonction du nombre d'images, nécessite 3 heures pour mille images. C'est donc à cet ordre de grandeur que le temps nécessaire à l'optimisation devient prépondérant.

4.5 Demi-transformées guidées par des points d'intérêt

4.5.1 Introduction

Afin de répondre aux challenges listés précédemment, il est nécessaire de développer une méthode permettant de s'affranchir des données images denses, ainsi que d'avoir une représentation compacte des champs de déformations.

Afin d'obtenir une représentation minimale et dépasser les limites de l'utilisation d'une référence, nous utilisons un schéma de recalage sans référence par graphe complet. Afin d'obtenir une méthode rapide, nous utilisons exclusivement pour information d'entrée la liste des appariements entre les points d'intérêt de toutes les images (d'où la dénomination par graphe complet).

Cette information est utilisée indirectement en la projetant dans un espace commun qui ne sert que d'intermédiaire mathématique, cette technique nous permet d'avoir une représentation compacte et minimale des champs de déformations, qui sont représentés en utilisant des interpolateurs par splines ou par RBF.

4.5.2 Données d'entrée

Nous considérons un ensemble \mathcal{I} de n images 3D défini dans un espace Ω de \mathbb{R}^3 . Chaque image i est associée à un ensemble de points d'intérêt $\mathcal{P}^i = \{p_a^i\}$. Pour des raisons de simplicité, nous définissons

$\mathcal{P} = \bigcup \mathcal{P}^i$ comme l'ensemble de tous les points d'intérêt. L'algorithme utilise des coordonnées réelles (en millimètres) et est invariant aux conditions d'échantillonnage.

Comme décrit précédemment, nous calculons l'ensemble $\mathcal{M} = \{(p_a^i, p_b^j)\}$ des points d'intérêt appairés, pour toutes les paires d'images (i, j) .

4.5.3 Choix des demi-transformées

L'ensemble $\mathcal{T} = \{\tau^i\}$ est l'ensemble des demi-transformées entre l'espace local de chaque image et un espace commun abstrait. Il permet de transformer n'importe quel point p grâce à l'application de la demi-transformée :

$$\tau^i : p \in \Omega \mapsto \tau^i(p) \in \Omega \quad (4.1)$$

Nous avons le choix d'utiliser deux types d'interpolateurs pour le champ de déformation : des fonctions de base radiales (RBF) ou des B-Splines cubiques (TP-UCBS). Dans les deux cas, la fonction τ^i peut s'écrire grâce à un ensemble \mathcal{B} de points de contrôle sous la forme suivante :

$$\tau^i : p \mapsto p + \mathbf{b}^i(p) \mathbf{x}^i \quad (4.2)$$

où le $j^{\text{ème}}$ élément de $\mathbf{b}^i(p)$ est l'évaluation de la $j^{\text{ème}}$ fonction de base (Spline ou Radiale) au point p dans l'image i . La matrice \mathbf{x}^i est la matrice des coefficients représentant le déplacement des k points de contrôle \mathbf{d} (une colonne par coordonnée), ce sont les inconnues que l'on cherche à calculer.

$$\mathbf{x}^i = \begin{pmatrix} d_x^{i,1} & d_y^{i,1} & d_z^{i,1} \\ \vdots & \vdots & \vdots \\ d_x^{i,|\mathcal{B}|} & d_y^{i,|\mathcal{B}|} & d_z^{i,|\mathcal{B}|} \end{pmatrix}_{|\mathcal{B}| \times 3} \quad \text{et } \mathbf{b}(p) = (b^1(p) \quad \dots \quad b^{|\mathcal{B}|}(p))_{1 \times |\mathcal{B}|} \quad (4.3)$$

Nous utilisons des grilles rectilinéaires anisotropes. Afin d'obtenir une grille la plus régulière possible, nous prenons un nombre d'intervalles permettant d'avoir une distance entre points de contrôle la plus constante possible. Dans les faits, La grille a généralement les dimensions $n \times n \times 2n$ pour des patients imagés en corps entiers (images plus grandes selon l'axe Z).

Fonctions de base radiales

Dans le cadre des Fonctions de Bases Radiales (RBF) ZAGORCHEV et GOSHTASBY [2006], décrites dans la partie 2.4.3, les éléments $b^i(p)$ du vecteur $\mathbf{b}(p)$ sont définis par :

$$b^i(p) = \phi(\|\mathbf{p} - \mathbf{p}_c^i\|) \quad (4.4)$$

où \mathbf{p}_c^i est la position du $i^{\text{ème}}$ point de contrôle.

Dans notre méthode, nous utilisons des fonctions de bases gaussiennes :

$$\phi(r) = e^{-(\sigma r)^2} \quad (4.5)$$

où σ est le paramètre de variance de la gaussienne qui est à définir.

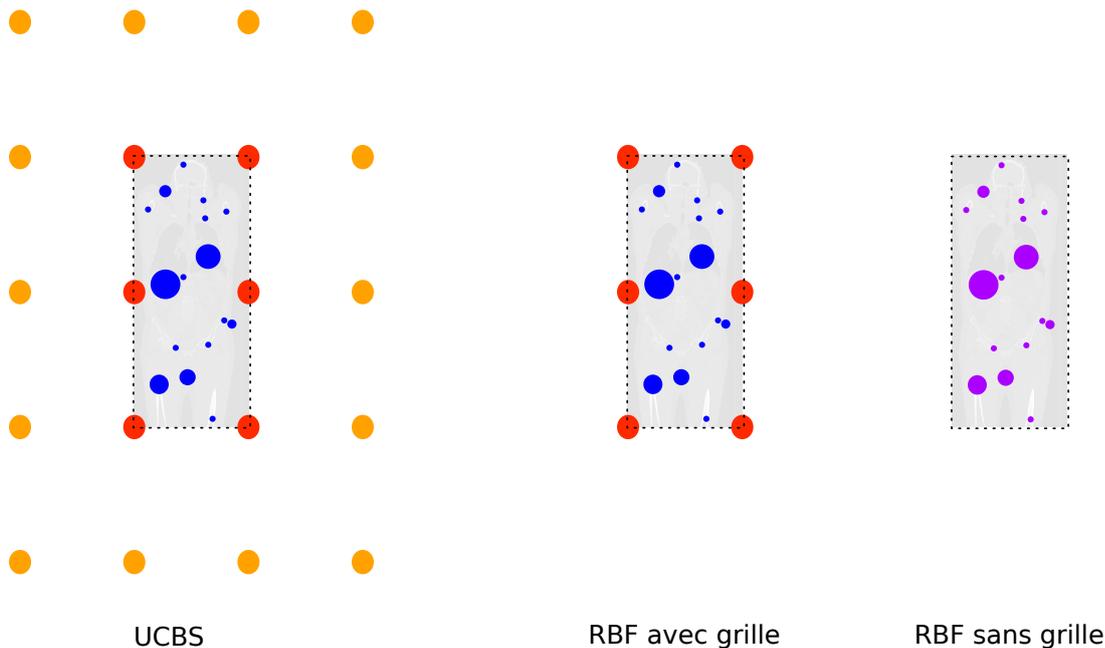


FIGURE 4.3 – Différentes structures des champs de déformation. Nous pouvons observer la grille anisotrope possédant plus d’intervalles dans l’axe Z. Les points d’intérêt sont en bleu. Les points de contrôle en rouge et orange. Les points orange représentent les points nécessaires à l’utilisation d’un domaine de définition naturel. Dans le cas des RBF sans grilles, les points violets représentent à la fois les points de contrôle et d’intérêt, ceux-ci étant confondus.

Les fonctions de base radiales ne sont pas nécessairement placées sur une grille (approche non-séparable), comme sur la figure 4.3 centrale. Lors des premières expérimentations que nous avons effectuées, chaque point d’intérêt portait une RBF, comme sur la figure 4.3 de droite. Cette disposition libre permet d’adapter la précision du champ de déformation vis-à-vis la quantité locale de points d’intérêt. Cependant, l’utilisation du critère de régularisation inter-images sur les points de contrôle (voir partie 4.8.5) nous a fait privilégier la solution par grille.

B-Splines

Les B-Splines cubiques [EILERS et MARX, 1996], décrites dans la partie 2.4.3, sont définies par le produit tensoriel de B-splines afin de les utiliser en 3D :

$$b^i(p) = B_i^n(p_x)B_j^n(p_y)B_k^n(p_z) \quad (4.6)$$

$B_i^n(p)$ est défini, en coordonnées normalisées, par :

$$B_j(t) := \begin{cases} b_3(t) = \frac{1}{6}t^3 & \text{si } t \in [t_i, t_{i+1}[\\ b_2(t) = \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1) & \text{si } t \in [t_{i+1}, t_{i+2}[\\ b_1(t) = \frac{1}{6}(3t^3 - 9t^2 + 4) & \text{si } t \in [t_{i+2}, t_{i+3}[\\ b_0(t) = \frac{1}{6}(-t^3 + 3t^2 - 3t + 1) & \text{si } t \in [t_{i+3}, t_{i+4}[\\ 0 & \text{sinon} \end{cases} \quad (4.7)$$

Nous utilisons le domaine de définition naturel des splines. Il est donc nécessaire de définir 3 points de contrôle en plus du nombre d’intervalles et cela pour chaque dimension (Voir figure 4.3 gauche).

4.5.4 Approche pyramidale

L'utilisation d'une approche pyramidale est classique dans le cadre de la mise en correspondance déformable d'images. Dans notre méthode, elle est motivée par l'utilisation de la pondération EM et par certaines particularités (approximation des espaces d'arrivée par les espaces de départ, difféomorphisme).

L'utilisation de la pondération EM nécessite la conservation d'une proximité avec la solution car le critère de sélection de l'EM est basé sur la distance, il est nécessaire que les vrais positifs apportent une erreur proche de zéro, afin de les discriminer des faux positifs ayant une erreur de distance supérieure. Afin de pouvoir discriminer les vrais des faux positifs, il est nécessaire que l'intervalle entre les points de contrôle, qui peut être relié à la fréquence spatiale maximale que les champs de déformation peuvent estimer, soit cohérent avec les erreurs des vrais positifs.

Une contrainte inter-images sur les points de contrôle utilise l'approximation entre l'espace d'arrivée et de départ des déformations. Afin que cette approximation reste valide, il est nécessaire que le rapport entre la déformation d'un point de contrôle et l'intervalle entre deux points de contrôle reste petit. [CHOI et LEE \[2000\]](#) montrent que conserver ce faible rapport permet d'approcher un champ de déformation difféomorphe. Cette hypothèse de petites déformations se traduit d'abord par l'application d'un recalage rigide avant d'utiliser l'approche déformable.

Application d'un recalage rigide Il est nécessaire d'effectuer une mise en correspondance rigide avant d'appliquer notre méthode déformable. Celle-ci est rendu nécessaire par l'hypothèse de proximité à la solution et l'incohérence des données brutes. Pour satisfaire cette hypothèse, nous utilisons l'approche FGR développée dans le chapitre 3 qui nous permet d'estimer un premier ensemble de champs de déformations rigides $\mathcal{T}_{(r)} = \{\tau_{(r)}^i\}$. Ces champs de déformations sont appliqués aux points avant notre méthode déformable, ce qui permet d'avoir un ensemble cohérent de points, avec une erreur faible.

Compositions des champs De la même manière que pour l'approche rigide, l'approche pyramidale utilisée est une approche par composition. De cette manière, les champs des différents niveaux de la pyramide sont composés entre eux et avec le recalage rigide afin d'avoir le champ de déformation total :

$$\tau^i = \tau_{(n)}^i \circ \dots \circ \tau_{(1)}^i \circ \tau_{(0)}^i \circ \tau_{(r)}^i \quad (4.8)$$

Schéma de raffinement À chaque niveau de la pyramide, nous divisons par deux la distance entre les points de contrôle. Ce qui est équivalent à doubler la fréquence spatiale maximale que le champs de déformation peut capturer.

Dans le cadre des fonctions de bases radiales, ce schéma consiste à commencer par un schéma possédant 2^d points de contrôle, d étant le nombre de dimensions du problème et de doubler le nombre de points de contrôle dans chaque direction à chaque nouveau niveau de la pyramide. De manière générale, le nombre de points de contrôle est de $(2n)^d$ au niveau n .

Dans le cadre des UCBS, nous utilisons le domaine naturel de définition et non le domaine étendu. L'utilisation du domaine naturel a pour conséquence d'avoir un nombre de points de contrôle égal aux nombres d'intervalles plus 3, voir figure 4.4. Le nombre de points de contrôle est défini pour le niveau n par : $(2n + 3)^d$.

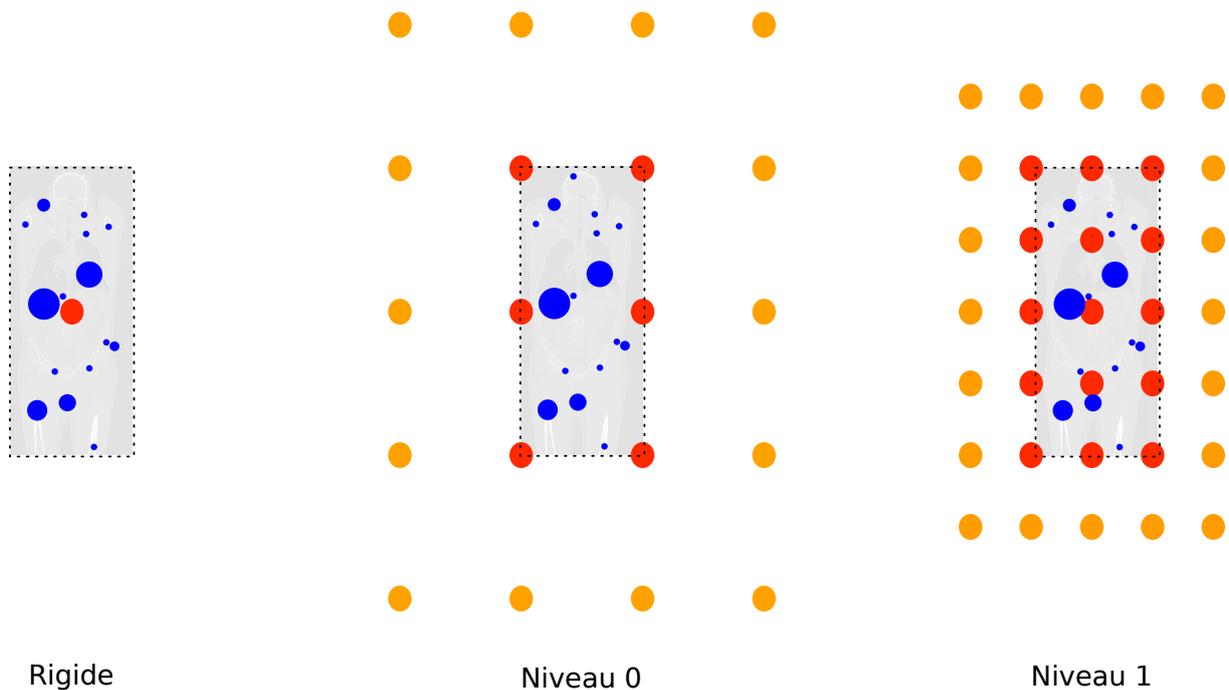


FIGURE 4.4 – Exemple d’une pyramide de points de contrôle (rouges). A gauche, le recalage rigide, qui peut être vu comme un champ contrôlé par un seul point. Ensuite les niveau 0 et 1, où les points de contrôle UCBS nécessaires à l’utilisation d’un domaine de définition naturel sont représentés en orange.

4.5.5 Attache aux données

Le critère utilisé pour optimiser les champs de déformation est la distance intra-paire, transformée vers l’espace commun (voir figure 4.11) :

$$d(p_a^i, p_b^j) = \|\tau^i(p_a^i) - \tau^j(p_b^j)\| \quad (4.9)$$

Cette distance est la seule mesure effectuée dans l’espace commun abstrait.

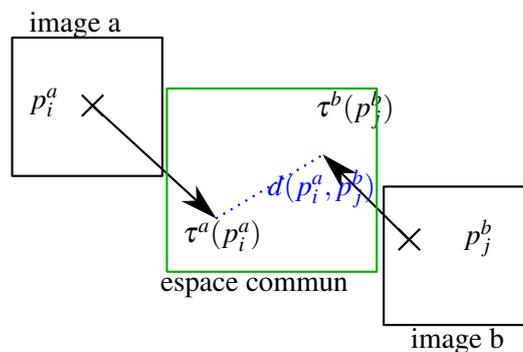


FIGURE 4.5 – Schéma illustrant l’erreur de distance $d(p_a^i, p_b^j)$ entre deux points d’intérêt transformés dans l’espace commun $\tau^i(p_a^i)$ et $\tau^j(p_b^j)$.

4.6 Statistiques robustes

L'utilisation de points d'intérêt implique l'utilisation de statistiques robustes. L'estimation des demi-transformées entre toutes les images est effectuée grâce à la liste des couples correspondant aux points appariés entre toutes les images. Comme décrit précédemment, cet ensemble contient des couples aberrants qui ne décrivent pas des zones de réelle correspondance. Afin d'avoir une bonne estimation de la déformation, il ne faut pas que les faux positifs ('correspondances aberrantes') viennent dégrader la performance des vrais positifs ("vrais correspondances").

L'idée générale pour permettre de discriminer les faux positifs des vrais est d'observer la distance entre les points appariés. Les vrais positifs décrivent les mêmes zones, ils doivent donc être proches. Les faux positifs peuvent être considérés comme des points appariés au hasard, ils ont donc une erreur de distance plus grande. Les approches de statistiques robustes ont donc pour objectif de fournir un critère de sélection des vrais positifs basé sur l'erreur courante.

La distance $d(p_a^i, p_b^j)$ est calculée entre les points d'intérêt appariés. Comme nous l'avons vu, cet appariement est imparfait et conduit à de nombreux faux positifs. Afin que l'estimation des demi-transformées ne soit pas faussée par la contribution des appariements aberrants, il est nécessaire d'attribuer une pondération $w(p_a^i, p_b^j)$ aux observations. De manière idéale, cette pondération doit être unitaire $w = 1$ pour les vrais positifs et nulle $w = 0$ pour les faux positifs. L'utilisation de statistiques robustes permet d'estimer ce terme permettant de séparer les vrais positifs des faux.

La particularité de notre approche est que la quantité d'appariements aberrants est très élevée, supérieure à 50% et avoisinant souvent les 80%. Les méthodes classiques de statistiques robustes, tel que les M-Estimeurs, ne permettent pas de gérer autant de faux positifs. Nous avons développé une approche spécifique à notre problématique permettant de gérer un taux élevé de mauvais appariements.

Discussion Nous avons choisi le parti pris de ne pas réutiliser les informations caractérisant les points d'intérêt, de la même manière que les approches classiques d'AF. Cela permet de séparer le processus, en considérant que les informations caractérisant les points d'intérêt ne sont utilisées que lors de l'appariement. Cette indépendance permet d'avoir une optimisation purement géométrique pouvant utiliser n'importe quelle source de points. Cependant, le développement d'une pondération alliant toutes les informations (distance, caractéristique) pourrait permettre une classification vrais/faux positifs plus fine.

4.6.1 M-estimateur de Tukey

Comme décrit au chapitre 2, Le M-estimateur de Tukey [HUBER, 2011] est un M-estimateur permettant d'inhiber totalement l'influence des appariements aberrants. Le critère qui permet de distinguer les bons échantillons des échantillons aberrants se base sur leurs distances. À une itération donnée de l'optimisation, nous faisons l'hypothèse que les vrais positifs ont une distance $\|\tau^i(p_a^i) - \tau^j(p_b^j)\|$ inférieure aux faux positifs (qui mettent en relation des points "au hasard" dans les images).

La méthode du M-estimateur de Tukey considère l'a priori selon lequel la distribution des échantillons est issue d'un mélange de distributions :

- une distribution normale, pour les vrais positifs (les bons appariements, décrivant des régions similaires correspondant réellement).
- une distribution uniforme, pour les faux positifs (les appariements aberrants).

La fonction de pondération utilisée est :

$$\begin{cases} \rho(x) = \frac{c^2}{6} \left(1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right) & \text{si } |x| < c \\ \rho(x) = \frac{c^2}{6} & \text{si } |x| > c \end{cases} \quad (4.10)$$

Le seuil d'exclusion des appariements aberrants c est défini par une estimation de la variance des vrais positifs. Cette variance est calculée grâce à l'estimateur robuste MAD (pour "Median Absolute Deviation"). La robustesse du M-estimateur de Tukey repose sur la robustesse de l'estimateur MAD qui repose lui même sur la robustesse de la médiane. La médiane ayant un point de rupture de 50% (voir section 2.8.2, HOAGLIN et collab. [1983]), Le M-estimateur de Tukey permet de gérer un maximum de 50% d'échantillons aberrants, ce qui est trop peu pour notre application.

Une première idée a été d'utiliser, non pas la médiane, mais un quantile inférieur pour estimer la variance de la distribution normale. Les bons appariements, qui ont une distance inférieure, permettent une estimation non-biaisée par la présence d'échantillons aberrants. Malheureusement cette solution n'offre pas de performances satisfaisantes, probablement à cause des hypothèses sur les distributions, éloignées de la réalité, comme nous le verrons dans la partie suivante.

4.6.2 Pondération-EM

La recherche d'une nouvelle méthode de pondération est motivée par la quantité de faux positifs de notre application, ainsi que des distributions particulières.

En effet, les distributions sous-jacentes suivent les lois suivantes :

- **les bons appariements** : Les vrais positifs peuvent être modélisés par la norme d'un vecteur gaussien de \mathcal{R}^3 . Ils suivent une distribution de χ à 3 degrés de liberté, connue aussi sous le nom de distribution de Maxwell.
- **les mauvais appariements** : Les faux positifs peuvent être modélisés comme une distance entre deux vecteurs de \mathcal{R}^3 dont les composantes suivent une distribution uniforme. Nous proposons d'approximer cette distribution en utilisant aussi une distribution de Maxwell.

La fonction de densité de probabilité en fonction de la distance peut s'exprimer de la manière suivante :

$$\begin{aligned} P(d) &= r.f(d, s_1) + (1 - r)f(d, s_2) \\ f(d, s) &= \sqrt{\frac{2}{\pi}} \frac{d^2}{s^2} e^{-\frac{d^2}{2s^2}} \end{aligned} \quad (4.11)$$

où f est la fonction de densité de probabilité de Maxwell, s_1 et s_2 sont respectivement les paramètres d'échelle des vrais et faux positifs des deux lois de Maxwell. $0 < r < 1$ est le paramètre du taux de vrais positifs.

Au lieu d'estimer la variance des vrais positifs et de faire l'hypothèse que les faux positifs sont distribués uniformément, nous estimons directement les paramètres $\theta = (s_1, s_2, r)$ de la loi $P(d)$ en utilisant un algorithme d'espérance-maximisation, MOON [1996] voir l'algorithme 6. Comme nous estimons explicitement les paramètres des lois de Maxwell au lieu d'utiliser l'estimateur MAD, nous pouvons gérer plus de 50% de faux positifs.

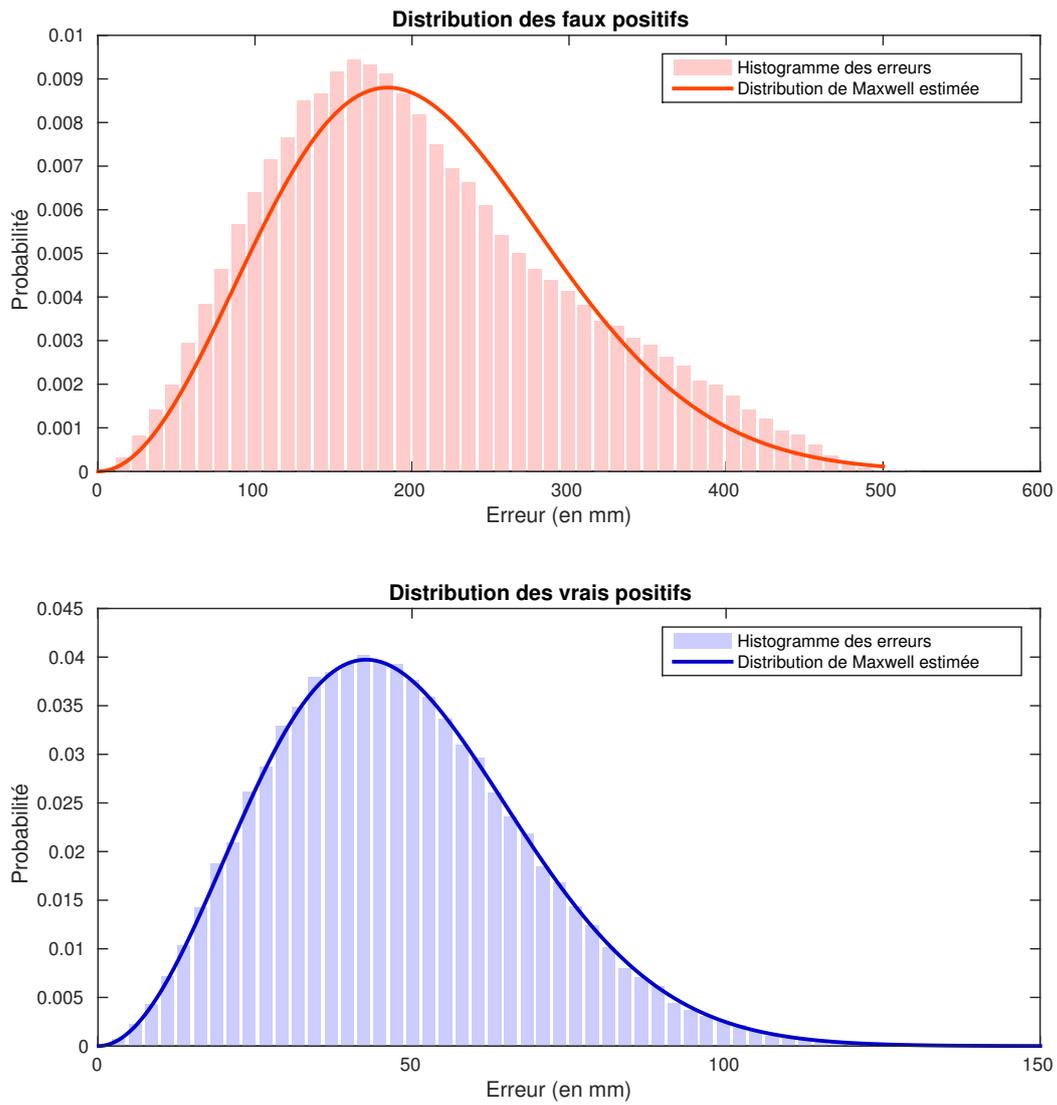


FIGURE 4.6 – Distribution synthétique des vrais (en bas) et faux (en haut) positifs. L'histogramme correspond aux distributions générées, tandis que les courbes pleines correspondent aux lois de probabilité estimées, issues d'une distribution de Maxwell.

Algorithme 6 : Estimation EM

Entrée : Un ensemble d'échantillons $\{x_i\}$
Sortie : Un jeu de paramètre : $\theta = (s_1, s_2, r)$

- 1 $(r, s^1, s^2) \leftarrow \text{some_initialization_values}$
- 2 **répéter**
- 3 **pour** chaque échantillon **faire**
- 4 $p_i^1 \leftarrow rf(x_i, s_1)$
- 5 $p_i^2 \leftarrow (1 - r)f(x_i, s_2)$
- 6 $t_i \leftarrow \frac{p_i^1}{p_i^1 + p_i^2}$
- 7 **fin**
- 8 $s_1 \leftarrow \frac{\Gamma(3/2) \sum_i t_i x_i}{\sqrt{2}\Gamma(2) \sum_i t_i}$
- 9 $s_2 \leftarrow \frac{\Gamma(3/2) \sum_i (1-t_i) x_i}{\sqrt{2}\Gamma(2) \sum_i (1-t_i)}$
- 10 $r \leftarrow \frac{\sum_i t_i}{\sum_i}$
- 11 **jusqu'à** convergence

En utilisant le théorème de Bayes, nous pouvons inférer la probabilité d'appartenance des vrais positifs sachant l'erreur (la distance) à la classe V , avec les paramètres des distributions :

$$P(V|d, \theta^i) = \frac{r^i f(d, s_1^i)}{r^i f(d, s_1^i) + (1 - r^i) f(d, s_2^i)} \quad (4.12)$$

Cette probabilité peut être utilisée comme une fonction de pondération pour inhiber l'influence des faux positifs, à la manière des M-estimateurs.

Note : Nous utilisons cette méthode de pondération uniquement dans le cadre d'images en 3D. Néanmoins, cette méthode peut facilement s'appliquer en 2 dimensions en utilisant des distributions de Rayleigh.

Paramétrisation par image

Afin d'obtenir une estimation plus fine, il est souhaitable d'avoir une estimation des paramètres des deux distributions par image. La variabilité des différentes images et le comportement pendant les itérations d'optimisation font que les images peuvent avoir des paramètres de distributions différents. L'estimation des paramètres par image consiste à estimer un jeu de paramètres $\theta^i = (s_1^i, s_2^i, r^i)$ pour chaque image i , en prenant en compte uniquement les appariements des points faisant intervenir cette image.

Chaque appariement de points (p_a^i, p_b^j) lie deux points provenant de deux images. Deux probabilités peuvent être calculées pour chaque appariement $P(V|(p_a^i, p_b^j), \theta^i)$ et $P(V|(p_a^i, p_b^j), \theta^j)$, en utilisant, respectivement, les images i et j pour contexte statistique.

Comme un critère symétrique est nécessaire, nous prenons le minimum de ces deux valeurs pour chaque appariement :

$$w(p_a^i, p_b^j) = \min[P(V|d(p_a^i, p_b^j), \theta^i), P(V|d(p_a^i, p_b^j), \theta^j)] \quad (4.13)$$

Le minimum est une bonne heuristique, car il augmente la sélectivité vis-à-vis des faux positifs. Nous avons expérimenté l'utilisation d'autres moyens d'agrégation, comme par exemple la moyenne, mais il est plus efficace de rejeter quelques vrais positifs plutôt que d'accepter des faux positifs.

Paramétrisation locale

Nous avons vu, dans la partie précédente, comment générer des statistiques propres à chaque image. Cette caractérisation par image permet de faire face à l'hétérogénéité due à la variabilité inter-patients, mais la mise en correspondance déformable entre des images très différentes induit aussi une hétérogénéité spatiale au sein même d'un patient. Plus précisément, lors de l'optimisation, certaines zones peuvent être bien recalées tandis que d'autres ne le sont pas encore. Cette différence résulte en des statistiques différentes, et une adaptivité spatiale est nécessaire.

Afin de créer une estimation locale des vrais et des faux positifs, nous utilisons les mêmes splines que pour l'estimation de la déformation. Au lieu de calculer les paramètres θ^i pour l'image entière, nous calculons une estimation par points de contrôle θ_j^i , en ne prenant que les points à l'intérieur du support de la fonction de base associée. Plus précisément, nous pondérons l'ensemble des mises en correspondance d'une image par la fonction de base associée au point de contrôle considéré.

Algorithme 7 : Estimation EM pondérée

Entrée : Un ensemble des couples appariés $\{(x_i, w_i^k)\}$
Sortie : Un jeu de paramètre : $\theta^k = (s_1^k, s_2^k, r^k)$

- 1 $(r, s^1, s^2) \leftarrow \text{some_initialization_values}$
- 2 **répéter**
- 3 **pour** chaque échantillon **faire**
- 4 $p_i^1 \leftarrow r^k f(x_i, s_1^k)$
- 5 $p_i^2 \leftarrow (1 - r^k) f(x_i, s_2^k)$
- 6 $t_i \leftarrow \frac{p_i^1}{p_i^1 + p_i^2}$
- 7 **fin**
- 8 $s_1^k \leftarrow \frac{\Gamma(3/2)}{\sqrt{2}\Gamma(2)} \frac{\sum_i t_i x_i w_i}{\sum_i t_i w_i}$
- 9 $s_2^k \leftarrow \frac{\Gamma(3/2)}{\sqrt{2}\Gamma(2)} \frac{\sum_i (1-t_i) x_i w_i}{\sum_i (1-t_i) w_i}$
- 10 $r^k \leftarrow \frac{\sum_i t_i w_i}{\sum_i w_i}$
- 11 **jusqu'à convergence**

Grâce à cette estimation, nous pouvons utiliser l'estimation EM en un point spécifique p afin d'avoir l'interpolation locale $[s_1^1, s_1^2, r_i]$ de la distribution en utilisant les TP-UCBS.

$$\theta(p) : [s^1, s^2, r]_p = \sum_{\forall k \in \mathcal{K}} \beta^k(p) \theta^k = \mathbf{b}(p) \boldsymbol{\theta} \quad \forall i \in \mathcal{I} \quad (4.14)$$

où chaque ligne de θ est le vecteur θ^k du $k^{\text{ième}}$ point de contrôle.

Limite de localité

Grâce à la paramétrisation locale, nous sommes passés d'une représentation globale à une estimation des distributions de probabilités locale à chaque image et à chaque position spatiale. Chaque image et chaque localité peut avoir un comportement distinct. Prendre ces comportements en compte de manière différenciée permet de modéliser plus finement le comportement des appariements.

Nous pouvons aussi interpréter cela comme suit : les distributions des erreurs globales sont des distributions marginalisées sur toutes les images et toutes les localités.

Localité	Nombres d'échantillons
Globale	$(c.n)^2$
Par image	$(c.n)$
Locale et par image	$\frac{c.n}{V}$
Locale et par paire	$\frac{c}{V}$

TABEAU 4.1 – Nombre d'échantillons disponibles en fonction de la localité, n est le nombre d'images. V représente le nombre de localités par images, c est un coefficient constant.

Cependant ces distributions locales sont encore des distributions marginales. Afin d'être le plus spécifique que possible, il aurait fallu créer des distributions, non pas pour chaque image, mais pour chaque paire d'images. En effet, les échantillons utilisés pour calculer les distributions locales sont les appariements entre les points d'une image particulière et d'une localité spécifique avec les points correspondants sur l'ensemble des autres images. De la même manière que le comportement d'une image peut être différente d'une autre, le comportement d'une paire d'images $A \leftrightarrow B$ peut différer d'une autre paire d'images $A \leftrightarrow C$.

Ces distributions spécifiques pour chaque paire ne peuvent pas être utilisées. Ceci est dû au nombre d'échantillons disponibles pour estimer les distributions. Les distributions spécifiques à chaque image et locales bénéficient d'un nombre d'échantillons proportionnel au nombre d'images n . Dans cette situation il suffit d'avoir un nombre d'images n suffisant pour pouvoir estimer les distributions.

Pour les distributions spécifiques par paires, le nombre d'échantillons est indépendant du nombre d'images et dépend directement de la taille des localités V des images.

4.6.3 Test avec des données synthétiques

Afin de confirmer la capacité de notre méthode à estimer les distributions approximées, nous avons conçu un jeu de données de synthèse, nous permettant de maîtriser les distributions sources.

En ce qui concerne les vrais positifs, ils sont modélisés comme des vecteurs gaussiens (des vecteurs dont les composantes sont issues de distributions gaussiennes) :

- **nombres de points** : 100 000
- **Ecart-type des composantes du vecteur gaussien** : 30 mm
- **Dimensions** : 3

Les faux positifs, les appariements aberrants, sont modélisés par la différence de deux vecteurs issues de tirages aléatoires uniformes au sein d'une boîte englobante. Le rapport final de points aberrants est de 50% :

- **nombres de points** : 100 000
- **Dimensions de la boîte englobante** : $350\text{mm} \times 350\text{mm} \times 600\text{mm}$
- **Dimensions** : 3

La recherche des paramètres s_1, s_2 et r est effectuée grâce à l'algorithme 6 d'espérance-maximisation EM en prenant pour valeurs initiales $s_1 = 30, s_2 = 300, r = 0.5$. Le critère d'arrêt de l'algorithme consiste à obtenir une valeur de mise à jour relative inférieure à 10^{-3} pour les 3 paramètres : $\frac{\|\theta^n - \theta^{n-1}\|_\infty}{\|\theta^n\|_\infty} < 10^{-3}$.

L'algorithme d'espérance-maximisation converge en 12 itérations (avec le critère d'arrêt $\frac{\|\theta^n - \theta^{n-1}\|_\infty}{\|\theta^n\|_\infty} < 10^{-3}$). Les résultats sont visibles figure 4.7. Cette figure montre que les estimations des deux distributions sont de bonne qualité. L'approximation des échantillons aberrants par une distribution de Maxwell introduit très peu de déformation sur la forme de la distribution.

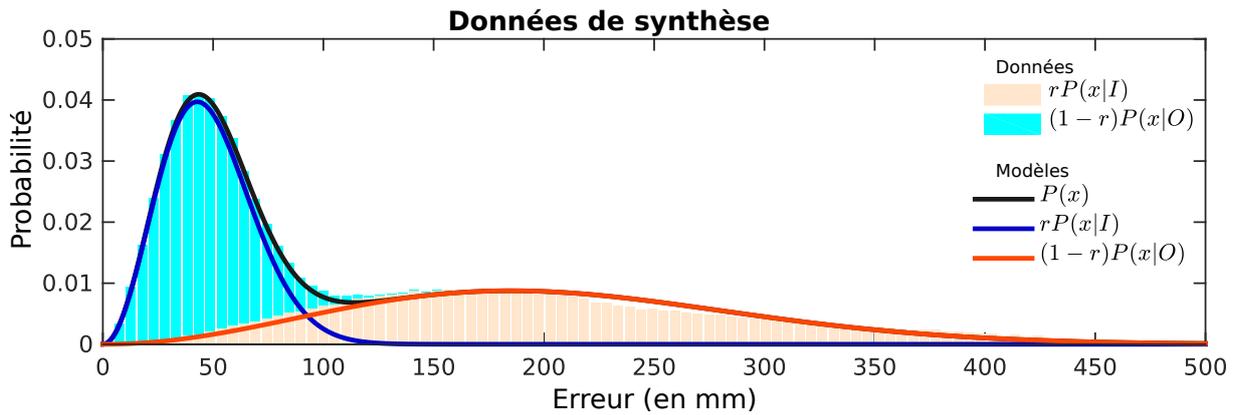


FIGURE 4.7 – Distribution des distances. Données synthétiques, distribution des vrais positifs en bleu clair et des faux positifs en orange. Les courbes bleues foncées et rouges représentent leurs approximations avec des distributions de Maxwell

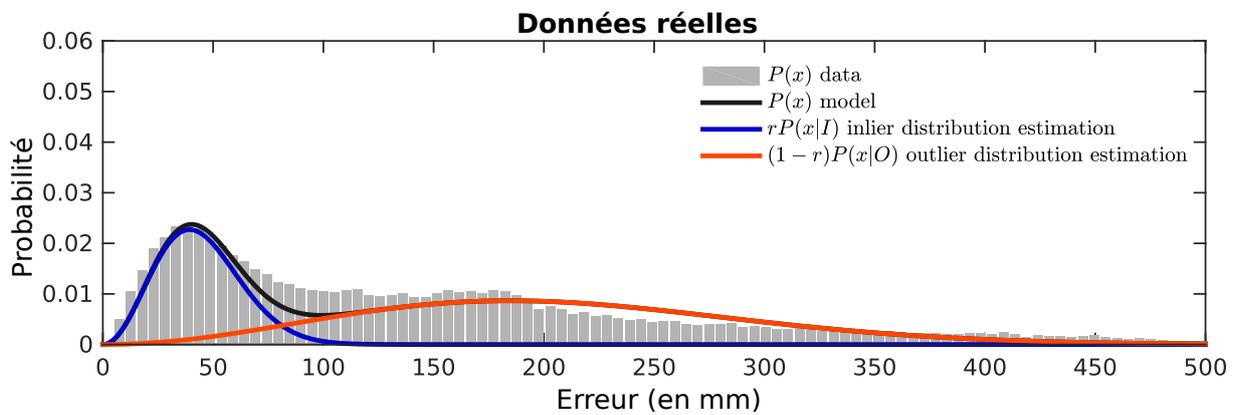


FIGURE 4.8 – Exemple avec des données réelles, avec plus de 50% d'échantillons aberrants. L'histogramme en gris est calculé avec une image du jeu de données VISCERAL après mise en correspondance. Les courbes bleues foncées et rouges représentent, respectivement, les distributions estimées des vrais positifs et des échantillons aberrants, avec des distributions de Maxwell.

4.6.4 Test avec des données réelles

Les données utilisées pour cette expérience sont issues du jeu de données VISCERAL [LANGS et collab., 2013]. Il contient 20 images thoraciques CT-scanner 3D avec produit de contraste. L'erreur de distance finale après optimisation est utilisée. L'expérience consiste à estimer les paramètres locaux de la zone du poumon droit en utilisant la variante locale de la pondération EM.

Grâce aux résultats visibles dans la figure 4.8, nous pouvons observer que les données réelles sont assez différentes des données de synthèse (figure 4.7). La plus grande proportion d'échantillons aberrants (65% estimés) contribue à rendre moins visible les deux modes observables dans l'expérience avec données de synthèse.

C'est la distribution des échantillons aberrants qui semble la moins bien approximée par une distribution de Maxwell. Plusieurs hypothèses peuvent être faites : la première est que nous considérons que la distribution des faux positifs provient de la différence de deux vecteurs tirés uniformément au sein de la boîte englobante. En réalité, la distribution des points d'intérêt au sein de l'image n'est pas uniforme et dépend de plusieurs facteurs complexes à modéliser : la densité des points d'intérêt dépend des structures anatomiques d'une part et de la détection en espace-échelle d'autre part. Ceci résulte en une distribution favorisant les courtes distances, car les points d'intérêt sont plus concentrés au centre de l'image.

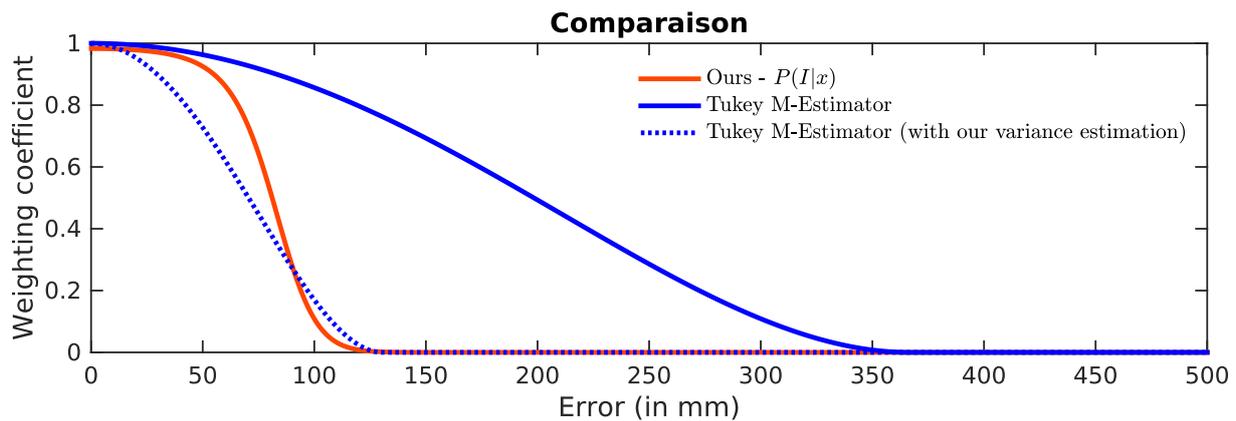


FIGURE 4.9 – Comparaison des fonctions de pondération avec les mêmes données réelles. En bleu : l’estimation incorrecte de la variance biaise le M-Estimateur qui inclut beaucoup d’échantillons aberrants. En pointillés bleu : le M-Estimateur avec la variance rectifiée. En rouge : notre pondération EM qui possède une meilleure sélectivité.

		M-estimateur		M-estimateur*		Pondération EM	
		Estimation		Estimation		Estimation	
		V	F	V	F	V	F
Réalité	V	99%	1%	91%	9%	97%	3%
	F	61%	39%	14%	86%	9%	91%

TABLEAU 4.2 – Matrices de confusion : M-estimateur à gauche, M-Estimateur rectifié (*, au centre) et pondération EM à droite. Ces matrices de confusions sont issues des données réelles.

Malgré ce phénomène, la pondération EM parvient à estimer des paramètres de distributions qui sont visuellement cohérents, en faisant correctement la distinctions entre les vrais et les faux positifs.

4.6.5 Comparaison

Afin de juger la pertinence de notre méthode, nous proposons une comparaison avec le M-estimateur de Tukey en utilisant les données réelles. Étant donné que la proportion d’échantillons aberrants dépasse le point de rupture du M-Estimateur de Tukey, nous proposons d’introduire un troisième scénario où la mauvaise variance estimée par l’estimateur MAD est remplacée par une variance optimale, ajustée manuellement.

Une comparaison de ces trois scénarios est disponible figure 4.9 et leurs matrices de confusion respectives dans la table 4.2. Tout d’abord nous pouvons voir que le M-estimateur de Tukey non rectifié (courbe bleue pleine) fournit une pondération significativement non-nulle pour une grande partie des échantillons aberrants. Ceci est dû, comme expliqué précédemment, à la mauvaise estimation de la variance des vrais positifs. La courbe de pondération issue du M-estimateur de Tukey rectifié fournit une pondération satisfaisante. Elle permet d’observer que notre méthode de pondération EM propose une meilleure sélectivité que le M-estimateur de Tukey : la pondération EM conduit à une séparation plus nette en vrais et faux positifs.

Les courbes ROC, de la figure 4.10 permettent aussi d’observer les qualités des différents classifieurs.

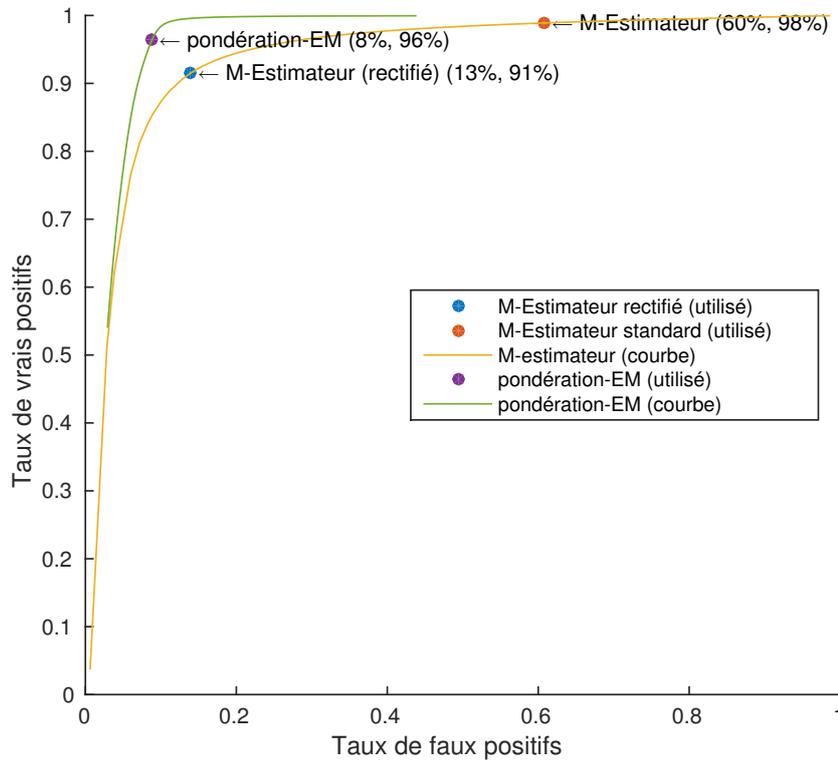


FIGURE 4.10 – Courbe ROC : M-Estimateur et Pondération EM.

La courbe du M-estimateur est créée en remplaçant la variance σ par $\alpha\sigma$ et en faisant varier α . Le point orange représente le M-Estimateur classique, en prenant $\alpha = 1$. Le point bleu représente la position optimale du M-Estimateur, nommé “M-estimateur rectifié”, en optimisant α pour avoir le meilleur compromis. Dans cette expérience, $\alpha = 0.2$.

Dans le cadre de la pondération EM, les distributions étant directement estimées, il n’y a pas de paramètre à faire varier. Afin de pouvoir tracer une courbe ROC, nous avons utilisé une fonction de transfert suivant une loi de puissance : $f(x) : x \rightarrow x^\alpha$ permettant d’ajuster le compromis entre taux de vrais positifs et taux de faux positifs. Dans nos expérimentations, nous n’utilisons pas cette fonction de transfert, donc $\alpha = 1$, qui est représentée par le point violet.

4.7 Optimisation

Nous proposons de trouver le meilleur ensemble de transformations \mathcal{T}^* en minimisant, pour chaque point d’intérêt, la distance quadratique moyenne (4.9) à ses points appairés, modérée par la pondération-EM (4.13).

La fonction objectif représentant notre problème $\mathbf{E}(\mathcal{T})$ est dépendante des demi-transformées choisies et doit avoir une valeur minimale lorsque les points de tous les vrais couples coïncident dans l’espace commun.

$$\mathbf{E}(\mathcal{T}) = \sum_{i=1}^n \sum_{p_a^i \in \mathcal{P}^i} \left(\frac{1}{|\mathcal{N}(p_a^i)|} \sum_{p_b^j \in \mathcal{N}(p_a^i)} w(p_a^i, p_b^j) d(p_a^i, p_b^j)^2 \right) \quad (4.15)$$

où $\mathcal{N}(p_a^i)$ est l'ensemble des points appariés à p_a^i et $|\mathcal{N}(p_a^i)|$ son cardinal.

Elle correspond à la somme quadratique, sur tous les couples \mathcal{P}^i de toutes les images i , des erreurs de distances $\|\tau^i(p_a^i) - \tau^j(p_b^j)\|^2$ intra-couple, pondéré par $w(p_a^i, p_b^j)$ déterminant si le couple (p_a^i, p_b^j) est un vrai ou faux positif.

Le terme $\mathcal{N}(p_a^i)$ correspond au nombre de couples auxquelles le points p_a^i appartient et permet d'avoir une contribution identique pour tous les points d'une image, indépendamment du nombre de couples auxquels ils appartiennent.

$$\mathcal{T}^* = \min_{\mathcal{T}} \mathbf{E}(\mathcal{T}) \quad \text{s.t.} \quad \sum_{i=1}^n \mathbf{x}^i = \mathbf{0} \quad (4.16)$$

L'équation $\sum_{i=1}^n \mathbf{x}^i = \mathbf{0}$ permet une minimisation avec contrainte justifiée dans la section 4.8.5.

4.7.1 Notation matricielle

Afin de simplifier les notations des parties suivantes, nous allons adopter une notation matricielle de notre problème.

Nous pouvons réécrire l'équation (4.15) sous la forme suivante :

$$\mathbf{E}(\mathcal{T}) = \sum_{i=1}^n \sum_{p_a^i \in \mathcal{P}^i} \sum_{p_b^j \in \mathcal{N}(p_a^i)} \left(\sqrt{\frac{w(p_a^i, p_b^j)}{|\mathcal{N}(p_a^i)|}} d(p_a^i, p_b^j) \right)^2 \quad (4.17)$$

En parcourant l'ensemble \mathcal{M} des appariements, la fonction objectif (4.17) peut être réécrite :

$$\mathbf{E}(\mathcal{T}) = \sum_{(p_a^i, p_b^j) \in \mathcal{M}} \left(s(p_a^i, p_b^j) d(p_a^i, p_b^j) \right)^2 \quad (4.18)$$

où $s(p_a^i, p_b^j)$ est défini par :

$$s(p_a^i, p_b^j) = \sqrt{w(p_a^i, p_b^j) \left(\frac{1}{|\mathcal{N}(p_a^i)|} + \frac{1}{|\mathcal{N}(p_b^j)|} \right)} \quad (4.19)$$

L'équation (4.18) peut être exprimée sous forme matricielle :

$$\begin{aligned} \mathbf{A} &= \mathbf{S}\mathbf{M}(\mathbf{P} + \mathbf{B}\mathbf{X}) \\ \mathbf{E}(\mathbf{X}) &= \text{tr}(\mathbf{A}^t \mathbf{A}) \end{aligned} \quad (4.20)$$

où \mathbf{S} est la matrice diagonale $|\mathcal{M}| \times |\mathcal{M}|$ contenant tous les $s(p_a^i, p_b^j)$.

\mathbf{P} est la matrice $|\mathcal{P}| \times 3$ contenant les positions des points d'intérêt.

\mathbf{M} est la matrice parcimonieuse $|\mathcal{M}| \times |\mathcal{P}|$ contenant les appariements. Chaque ligne de \mathbf{M} représente un couple de points appariés, elle contient exactement deux valeurs non-nulles : un 1 et un -1 dans les colonnes correspondantes au deux points d'intérêt considérés. La matrice \mathbf{M} peut être vue comme la matrice d'incidence du graphe orienté des appariements où les sommets sont des points d'intérêt et

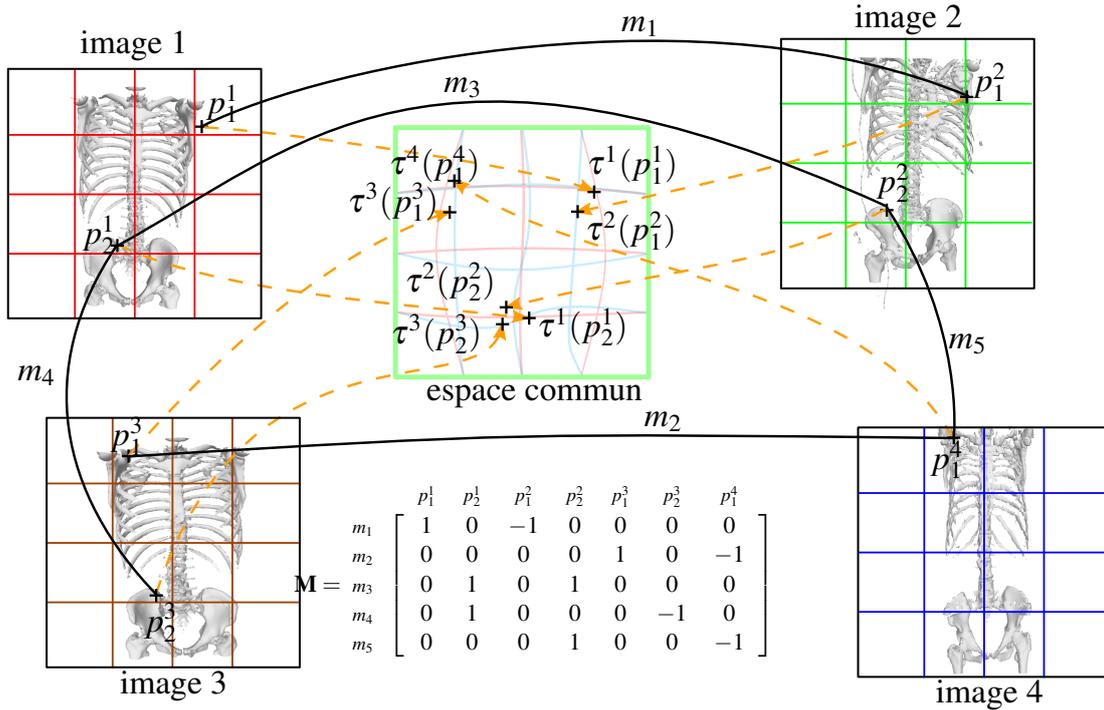


FIGURE 4.11 – Exemple avec 4 images. Pour des raisons de lisibilité, seul le squelette est affiché. En surimpression sur chaque image i : la grille associée aux splines utilisée pour représenter la transformée déformable τ^i . Au milieu : l'espace commun, qui est utilisé pour mesurer les distances intra-paires. de même que pour la figure 4.2, les appariements sont tracés en noir et la transformation des points en pointillés orange. Dans cet exemple simple, 7 points d'intérêt ont été extraits : $\mathcal{P} = \{p_1^1, p_2^1, p_1^2, p_2^2, p_1^3, p_2^3, p_1^4\}$, et 5 paires ont été trouvées : $\mathcal{M} = \{(p_1^1, p_1^2), (p_1^1, p_1^4), (p_2^1, p_2^2), (p_2^1, p_2^3), (p_2^2, p_1^4)\}$. La paire (p_2^2, p_1^4) est un faux positif. \mathbf{M} est la matrice des appariements.

les arêtes les appariements avec une orientation arbitraire. Un exemple de matrice \mathbf{M} est donné dans la figure 4.11.

\mathbf{B} est la matrice $|\mathcal{P}| \times n|\mathcal{B}|$ des fonctions de base. \mathbf{C} 'est une matrice rectangulaire diagonale par bloc, où chaque bloc non-nul \mathbf{B}^i est la matrice $|\mathcal{P}^i| \times |\mathcal{B}|$ de fonctions de base pour l'image i (la concaténation de tous les $\mathbf{b}(p_a^i)$) :

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{B}^n \end{bmatrix}_{|\mathcal{P}| \times n|\mathcal{B}|} \quad \text{avec } \mathbf{B}^a = \begin{bmatrix} \mathbf{b}(p_a^1) \\ \vdots \\ \mathbf{b}(p_a^{|\mathcal{P}^a|}) \end{bmatrix}_{|\mathcal{P}^a| \times |\mathcal{B}|} \quad (4.21)$$

\mathbf{X} est la matrice qui contient tous les déplacements des points de contrôle \mathbf{x}^i de chaque image i , i.e. les inconnues :

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^n \end{bmatrix}_{|\mathcal{M}| \times 3} \quad (4.22)$$

4.7.2 Moindres carrés pondérés itératifs

Nous utilisons les “moindres carrés pondérés itératifs” [HOLLAND et WELSCH, 1977] IRLS afin de prendre en compte la pondération EM. IRLS permet de simplifier le problème en considérant constante la pondération lors de l’optimisation et en séparant le processus en 2 étapes :

1. Estimation des distributions et des pondérations associées :

$$\mathbf{S} = f(\hat{\mathbf{X}}, \mathbf{M}, \mathbf{B}) \quad (4.23)$$

2. Minimisation des distances intra-paires.

$$\hat{\mathbf{X}} = \min_{\mathbf{X}} \mathbf{E}(\mathbf{X}) \quad (4.24)$$

Afin de résoudre le problème de minimisation de l’étape 2, nous avons essayé plusieurs méthodes à l’ordre 1 et 2.

Itération de Gauss-Newton

Méthode La méthode de Gauss-Newton [WEDDERBURN, 1974] est une méthode d’optimisation pour les fonctions s’écrivant sous la forme (voir section 2.6.5) :

$$f(\mathbf{x}) = \sum_{k=1}^m (f_k(\mathbf{x}))^2 \quad (4.25)$$

où, dans notre cas, $f_k(\mathbf{x})$ est de la forme :

$$\begin{aligned} f_k(\mathbf{x}) &= \sqrt{\frac{w(p_a^i, p_b^j)}{|\mathcal{N}(p_a^i)|}} d(p_a^i, p_b^j) \\ &= \sqrt{\frac{w(p_a^i, p_b^j)}{|\mathcal{N}(p_a^i)|}} \|\tau^i(p_a^i) - \tau^j(p_b^j)\| \\ &= \sqrt{\frac{w(p_a^i, p_b^j)}{|\mathcal{N}(p_a^i)|}} \left\| p_a^i + \mathbf{b}^i(p_a^i)\mathbf{x}^i - \left(p_b^j + \mathbf{b}^j(p_b^j)\mathbf{x}^j \right) \right\| \end{aligned} \quad (4.26)$$

La jacobienne de $f(\mathbf{x})$ est :

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial \mathbf{x}_1} & \cdots & \frac{\partial f_1}{\partial \mathbf{x}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial \mathbf{x}_1} & \cdots & \frac{\partial f_m}{\partial \mathbf{x}_n} \end{pmatrix}_{|\mathcal{M}| \times |\mathcal{B}|n} \quad (4.27)$$

La méthode de Gauss-Newton se propose d’approximer la hessienne $\mathbf{H} = \frac{d^2 f}{d\mathbf{x}^2}(\mathbf{x})$ en ignorant les dérivées supérieures à l’ordre 1 par $\hat{\mathbf{H}} = 2\mathbf{J}^T \mathbf{J}$. Dans notre cas, la hessienne est de dimensions $|\mathcal{B}|n \times |\mathcal{B}|n$.

Une itération de Gauss-Newton est définie par :

$$\mathbf{d}^{(k)} = -\alpha \mathbf{H}^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (4.28)$$

où le gradient $\nabla f(\mathbf{x}^{(k)})$ s’écrit :

$$\nabla f(\mathbf{x}^{(k)}) = \frac{df}{d\mathbf{x}}(\mathbf{x}) = \begin{bmatrix} \nabla_{\mathbf{x}^1} \\ \vdots \\ \nabla_{\mathbf{x}^n} \end{bmatrix}_{|\mathcal{B}|n \times 3} \quad (4.29)$$

En notation matricielle :

$$\nabla f(\mathbf{x}^{(k)}) = \nabla \mathbf{E} = \nabla \text{tr}(\mathbf{A}\mathbf{A}^T) = 2\mathbf{B}'\mathbf{M}'\mathbf{S}^2\mathbf{M}(\mathbf{P} + \mathbf{B}\mathbf{X}) \quad (4.30)$$

Résultat La forme de la fonction de coût, équation 4.15, est une forme quadratique. La résolution par la méthode de Gauss-Newton permet donc une convergence rapide. Il est cependant nécessaire de ralentir la vitesse de convergence en fixant $\alpha < 1$ afin de réduire les pas et d'obtenir un algorithme IRLS stable et à une faible mise à jour des pondérations $w(p_a^i, p_b^j)$.

Afin de trouver le meilleur compromis entre vitesse et stabilité, nous avons aussi mis en œuvre des itérations de Levenberg-Marquart et de Gauss-Newton amorti. Ces schémas d'optimisation permettent de réduire encore le nombre d'itérations, tout en préservant la stabilité.

Limitation Les méthodes du second ordre nécessitent de construire la matrice hessienne du problème. L'utilisation de la matrice hessienne pose problème car c'est une matrice dense (avec aucune valeur nulle) de dimensions $|\mathcal{B}|n \times |\mathcal{B}|n$. L'utilisation de cette matrice limite la résolution du problème à des jeux de données de l'ordre de quelques dizaines d'images.

Afin de pouvoir gérer des jeux de données contenant un nombre d'images supérieur à la centaine, il est nécessaire d'utiliser des méthodes ne mettant pas en œuvre cette matrice du second ordre. Cette limitation ouvre la voie aux méthodes du premier ordre, nécessitant plus d'itérations, mais possédant des temps d'itération courts. Elles possèdent aussi une complexité linéaire, contrairement à la complexité au moins quadratique des méthodes du second ordre.

Descente de gradient

Comme nous l'avons décrit dans la partie précédente, il est nécessaire d'utiliser une méthode du premier ordre afin d'avoir une complexité linéaire et une empreinte mémoire permettant l'usage d'un grand nombre d'images.

Une itération de descente de gradient permet de mettre à jour le vecteur \mathbf{X} avec une nouvelle valeur :

$$\mathbf{X}_k = \mathbf{X}_{k-1} - \alpha \nabla \mathbf{E} \quad (4.31)$$

où α est un coefficient positif inférieur à 1.

Calcul du gradient Le gradient de la fonction d'énergie calculé à partir de la forme matricielle s'écrit :

$$\nabla \mathbf{E} = 2\mathbf{B}'\mathbf{M}'\mathbf{S}^2\mathbf{M}(\mathbf{P} + \mathbf{B}\mathbf{X}) \quad (4.32)$$

De cette manière, une étape de descente de gradient s'écrit :

$$\begin{aligned} \mathbf{X}_k &= \mathbf{X}_{k-1} - \alpha \nabla \mathbf{E} \\ &= \mathbf{X}_{k-1} - \alpha 2\mathbf{B}'\mathbf{M}'\mathbf{S}_k^2\mathbf{M}(\mathbf{P} + \mathbf{B}\mathbf{X}_{k-1}) \end{aligned} \quad (4.33)$$

4.8 Régularisation

4.8.1 Introduction

De même que dans le cas rigide, seulement $n - 1$ demi-transformations sont nécessaires pour mettre en correspondance n images. Dans notre approche, nous avons n transformations τ^i inconnues. Cette

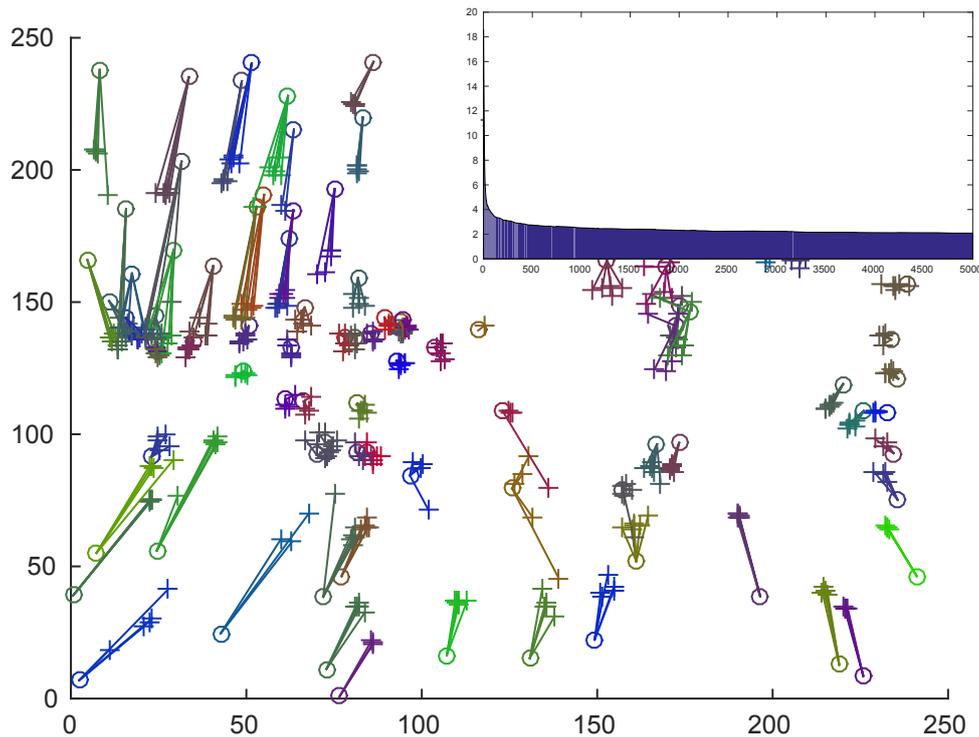


FIGURE 4.12 – Exemple de dérive des champs de déformations. Tous les champs se déforment de la même façon, ce qui n’augmente pas la distance-erreur, mais produit une dérive.

différence du nombre d’inconnues a pour conséquence de rendre le problème sous-déterminé : la convergence n’est pas garantie et les transformations peuvent être sujettes à une dérive globale, comme dans la figure 4.12.

Localement, pour un appariement entre deux points d’intérêt $\{p_a^i, p_b^j\}$, il y a une infinité de transformations $\{\tau^i(p_a^i), \tau^j(p_b^j)\}$ qui minimisent la distance $d(p_a^i, p_b^j)$, voir figure 4.13.

La régularisation a pour but de limiter la dérive globale. Plusieurs solutions sont possibles, chacune possédant des avantages et des inconvénients. En effet, la régularisation peut prendre la forme d’une contrainte douce sur les points de contrôle ou sur les couples appariés. Elle peut aussi prendre la forme d’une contrainte dure sur les points de contrôle.

Dans les parties suivantes, nous détaillerons les différentes solutions testées, avec leurs avantages et leurs inconvénients.

4.8.2 Protocole expérimental : données synthétiques

Afin de pouvoir expérimenter et vérifier facilement les performances de notre approche, nous avons développé un générateur de données de synthèse paramétrable permettant de générer les données nécessaires. Nous l’utilisons pour générer des scénarios en deux dimensions.

La génération comporte 4 étapes, schématisées avec la figure 4.14 : la génération de points d’intérêt, la génération de champs de déformations et la génération des mises en correspondance.

Ces expériences ont pour objectif de tester l’optimisation et l’utilisation des champs de déformation. Elles prennent donc en entrée des listes de points appariés (contenant des vrais et des faux positifs). Ici

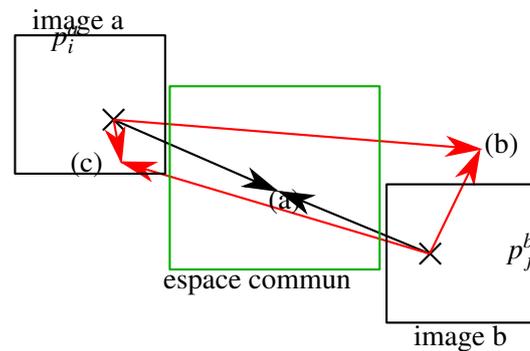


FIGURE 4.13 – Exemple de solutions de la minimisation. Plusieurs solutions sont possibles du fait de la sous-détermination du problème, la solution (a) est souhaitable tandis que les solutions (b) et (c) correspondent à une dérivation.

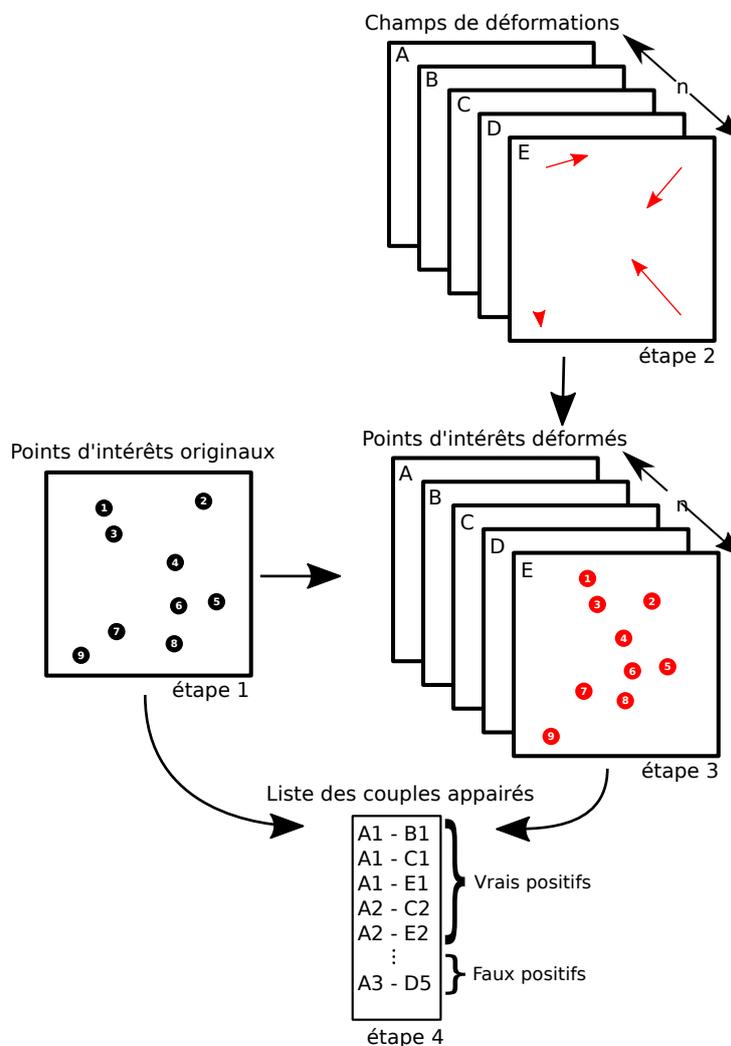


FIGURE 4.14 – Schéma de génération des données synthétiques. Etape 1 : un ensemble de points est généré. Etape 2 : un champ de déformation pour chaque image est généré. Etape 3 : les points sont projetés dans chaque image, en interpolant le champ de déformation. Etape 4 : la liste des appariements est créée, comportant à la fois des vrais et de faux positifs.

nous ne testons pas la détection, la description et l'appariement des points d'intérêt.

Génération des points d'intérêt

La génération des points d'intérêt consiste à créer une liste de points. Dans nos expérimentations de synthèse nous générons 200 points.

Ils sont créés dans un espace commun, ils peuvent être vus comme issus d'un patient moyen à partir desquelles seront générées toutes les images, par déformation.

La liste des paramètres pour cette étape est la suivante :

- Nombre de points
- Nombre d'images

Génération de champs de déformation

La deuxième étape consiste à générer un champ de déformation par image. Celui-ci a pour but de rendre compte de la variabilité entre les patients. Nous faisons ici l'hypothèse qu'il existe des champs de déformation permettant de lier tous les patients.

Les champs de déformation sont générés par tirage aléatoire de vecteurs placés sur une grille. Un champ de déformation dense est ensuite obtenu par interpolation de ce champ. Dans nos expérimentations, nous générons des grilles de tailles 2×2 ou 3×3 .

Afin de rendre compte de la répétabilité de la détection des points d'intérêt, une probabilité de présence des points dans chaque image est utilisée. Si le point est présent dans l'image, nous le déformons grâce au champ de déformation calculé précédemment. Ce sont ces coordonnées déformées auxquelles notre méthode à ensuite accès avec pour objectif de retrouver les coordonnées initiales.

La liste des paramètres pour cette étape est la suivante :

- Résolution de la grille,
- Amplitude de la déformation : échelle de la loi uniforme sur les vecteurs de la grille,
- Probabilité de présence d'un point.

Génération de la liste des couples

La dernière étape consiste à générer une liste de couples. Celle-ci contient des couples de référence de points dans les différentes images. Elle contient à la fois des vrais positifs : des points ayant le même point source (avant déformation) et des faux positifs : des couples aberrants composés de points n'ayant pas le même point source. Afin de rendre compte d'un processus d'appariement le plus réaliste que possible, nous associons deux probabilités à cette opération : une probabilité liée aux vrais positifs et une probabilité liée aux faux positifs.

La liste des paramètres pour cette étape est la suivante :

- Probabilité d'appariement de vrais positifs : Pour chaque couple possible, nous associons une probabilité uniforme de détection.
- Probabilité d'appariement de faux positifs : Nous tirons des points au hasard afin d'atteindre le taux cible d'appariements aberrants.
- Probabilité de présence d'un point dans une image.

La liste des paramètres est représentée dans le tableau [4.3](#).

Nom	Valeur
Nombre de points modèles	100
Nombre d'images	4
Résolution de la grille	2×2
Amplitude de déformation (% taille d'image)	10%
Probabilité d'apparition	0.8
Probabilité d'appariement	0.8
Fraction de faux positifs	0.2

TABLEAU 4.3 – Valeur des paramètres utilisés.

Exemple synthétique

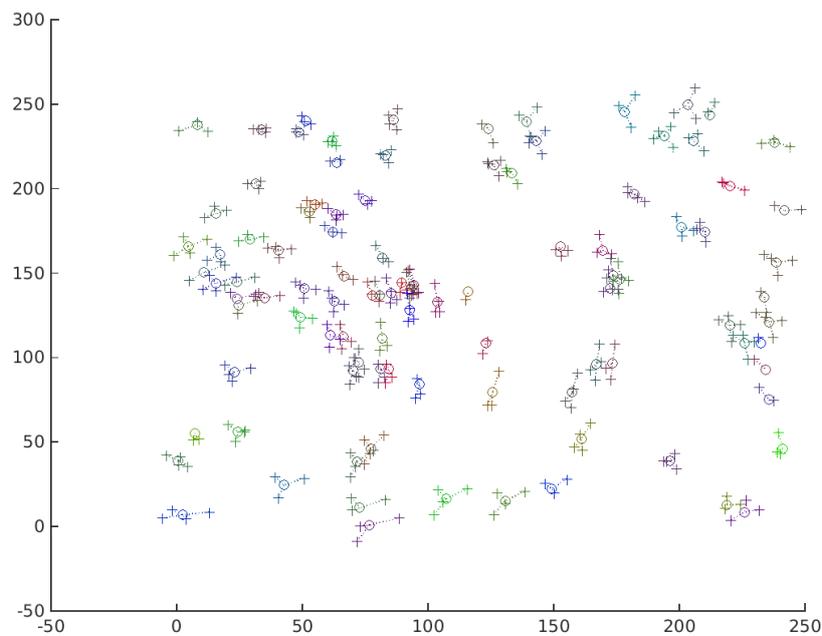


FIGURE 4.15 – Exemple de données de synthèse. Nous pouvons observer les points du modèle (marqueurs circulaires), et leurs apparitions dans les acquisitions (marqueurs en croix).

L'exemple présenté dans la figure 4.15 est issue de la réalisation suivante :

Nom	Valeur
Nombre de points par image :	80
Nombre de mise en correspondance :	961
dont faux positifs :	193

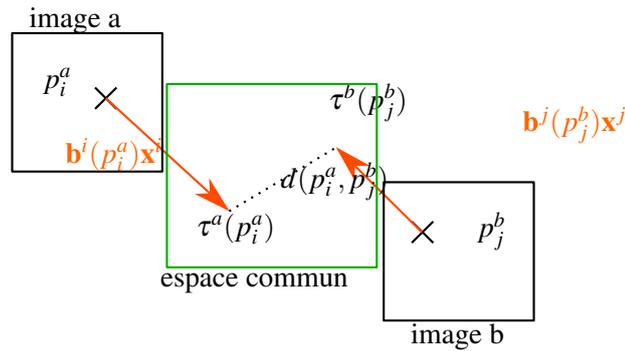


FIGURE 4.16 – Exemple de régularisation du déplacement de deux points de contrôle p_i^a et p_j^b . La régularisation dépend de la norme des vecteurs $\mathbf{b}^i(p_i^a)\mathbf{x}^i$ et $\mathbf{b}^j(p_j^b)\mathbf{x}^j$.

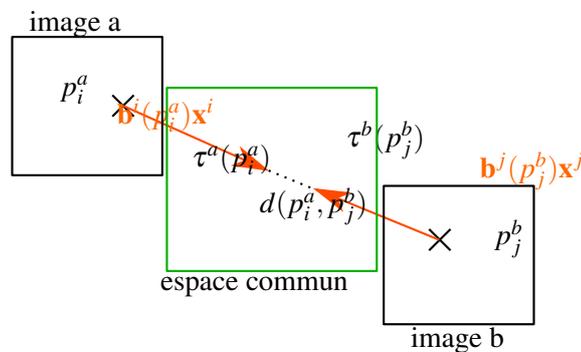


FIGURE 4.17 – Résultat de régularisation de deux points de contrôle. Les points déformés ne coïncident pas du fait de la présence de la régularisation.

4.8.3 Régularisation du déplacement des points de contrôle

La régularisation du déplacement des points de contrôle est la solution la plus simple. Elle consiste à pénaliser les grandes déformations. Dans ce cas, un terme de régularisation $\mathcal{R}(\mathbf{x})$ prend la forme suivante :

$$\mathcal{R}(\mathbf{x}) = \sum_{p \in \mathcal{P}} \|\mathbf{b}^i(p)\mathbf{x}^i\|^2 \quad (4.34)$$

Avantage : Ce terme de régularisation est très simple à mettre en œuvre, et s'intègre bien dans l'optimisation. De plus, il n'apporte aucune contrainte sur le choix des champs de déformations.

Inconvénient : La principale limitation de cette régularisation est qu'elle conduit à un compromis dans l'estimation du déplacement du point entre le terme de distance-erreur et le terme de régularisation. Ce compromis a pour conséquence de converger vers une solution avec un champ de déformation trop faible, les points déformés ne se rejoignent pas, laissant une certaine distance dépendante du poids de la régularisation, voir figure 4.17.

À l'aide de la figure 4.18, nous pouvons observer que cette régularisation permet effectivement d'empêcher la dérive observée dans le cas sans régularisation.

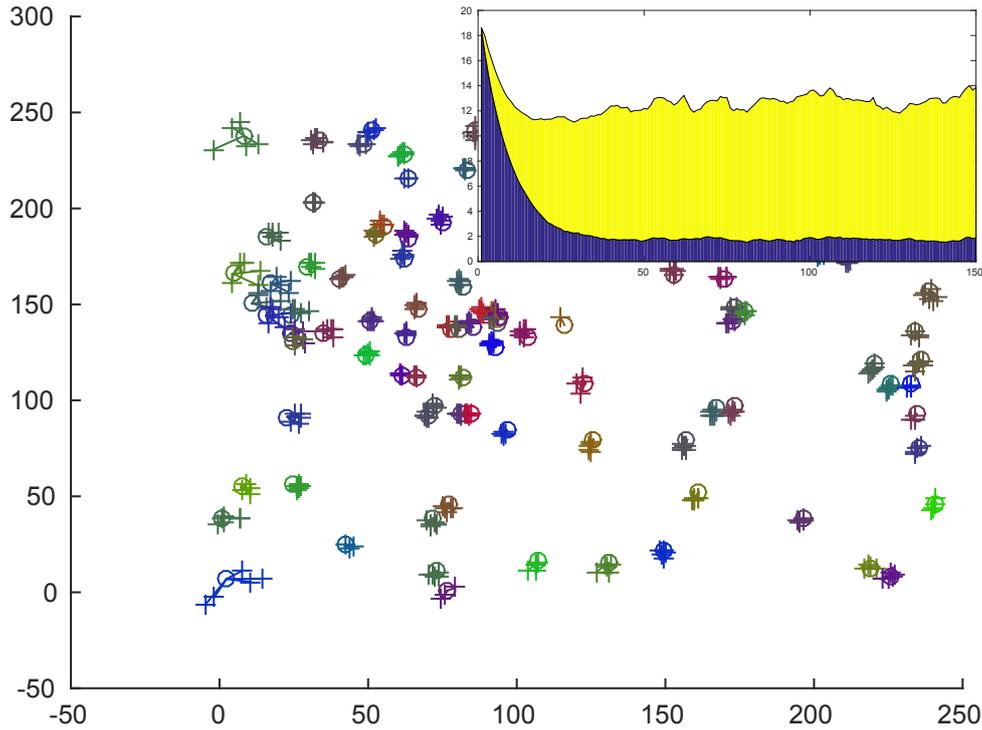


FIGURE 4.18 – Positions finales et courbes de convergence avec une régularisation du déplacement des points de contrôle. Nous pouvons observer qu’il reste une partie des déformations non capturée par l’estimation du champ de déformation. Grâce aux courbes de convergence cumulées, nous pouvons observer un compromis entre les deux termes : la régularisation (en jaune) et l’erreur de distance (en bleue).

En observant les courbes de convergence, on peut conclure que l’optimisation consiste en un compromis entre l’énergie de distance entre les points et l’énergie pénalisant les points de contrôle. La conséquence est une sous-estimation du champ de déformation qui conduit à conserver une erreur liée à la distance.

En portant attention à la position finale des points après optimisation, sur la figure 4.18, nous pouvons observer que le comportement diffère selon la localité.

4.8.4 Régularisation sur les couples appariés

Une autre régularisation consiste à se baser sur la source de la dérive globale des couples appariés, et à pénaliser directement leur dérive. La position optimale où les points d’intérêt appariés se rejoignent est au milieu des deux positions initiales. C’est le cas lorsqu’il n’y a aucune dérive.

Le terme de régularisation suivant se propose de pénaliser directement la dérive, en calculant le vecteur de dérive $\mathbf{b}^i(p_i^a)\mathbf{x}^i - \mathbf{b}^j(p_j^b)\mathbf{x}^j$ pour chaque appariement.

$$R = \sum_{p \in \mathcal{P}} \|\mathbf{b}^i(p_i^a)\mathbf{x}^i - \mathbf{b}^j(p_j^b)\mathbf{x}^j\|^2 \quad (4.35)$$

Avantage : Ce terme de régularisation semble optimal dans le cas de deux images. Comme le précédent, il n’induit aucune contrainte sur le choix du champ de déformation.

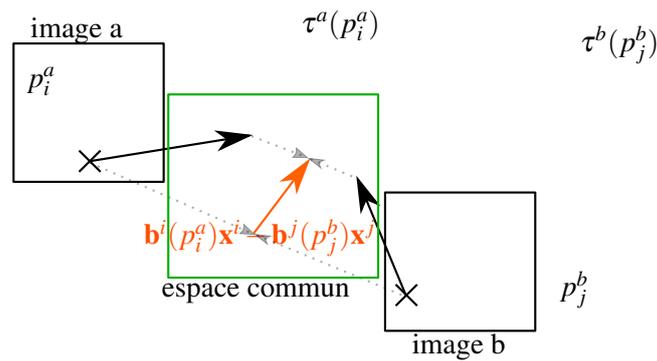


FIGURE 4.19 – Exemple de régularisation sur un couple apparié.

Inconvénient : Nous pouvons observer que dans le cas de points liés au travers de plus de deux images, voir figure 4.20, cette solution résulte en des champs trop faibles, dus à l’absence de solution exacte. Il est par ailleurs impossible de généraliser le calcul exact à plus de deux points d’intérêt, car la présence de faux positifs fait qu’il n’est pas possible de détecter des sous-graphes connexes indépendants.

Cette solution n’est pas idéale et elle conduit à une dépendance entre le terme de distance et le terme de régularisation, comme dans le cas de la régularisation sur les points de contrôle (figure 4.19). Les courbes de convergences de la figure 4.21 nous montrent ce compromis, l’erreur finale n’est pas nulle et les deux termes se stabilisent à des valeurs significativement non-nulles.

C’est l’appariement de plus de deux points qui a pour conséquence cette limitation en précision et une sous estimation des champs de déformations. Figure 4.22, nous pouvons observer que, dans cette expérience, le nombre d’appariements le plus fréquent est 3. Le cas idéal pour cette régularisation (un seul appariement), est minoritaire. Le cas décrit dans la figure 4.20 arrive donc régulièrement.

4.8.5 Contrainte inter-images sur les points de contrôle

La dernière approche que nous avons testée est une contrainte dure et non pas une régularisation. Elle consiste à utiliser la même répartition des points de contrôle pour toutes les images et à contraindre la moyenne des déplacements d’un point de contrôle à travers toutes les images vers une valeur nulle, voir figure 4.23. Cette approche a déjà été proposée par [WU et collab. \[2012\]](#).

$$\sum_{i=1}^n \mathbf{x}^i = \mathbf{0} \quad (4.36)$$

Avantage : Le principal avantage est qu’il supprime le degré de liberté supplémentaire grâce à une nouvelle équation, en fournissant une contrainte n’ayant pas d’effet sur les autres données du problème.

Inconvénient : Le premier inconvénient est de contraindre les points de contrôle à avoir la même position pour toutes les images, et ceci peut poser un problème quand les parties imagées ne sont pas les mêmes. De plus, la moyenne nulle est faite sur des points de contrôle ne reposant pas dans le même espace, on approxime donc l’espace commun par l’espace propre à chaque image. Ceci rajoute l’hypothèse nécessaire de petites déformations à l’optimisation qui existe déjà pour les splines.

L’ajout d’une contrainte sur la moyenne des points de contrôle à travers les images a pour objectif de fixer la déformation moyenne à 0. D’un point de vue applicatif, nous pouvons assimiler cette contrainte à forcer l’espace intermédiaire (qui peut être vu comme un humain moyen) à rester centré.

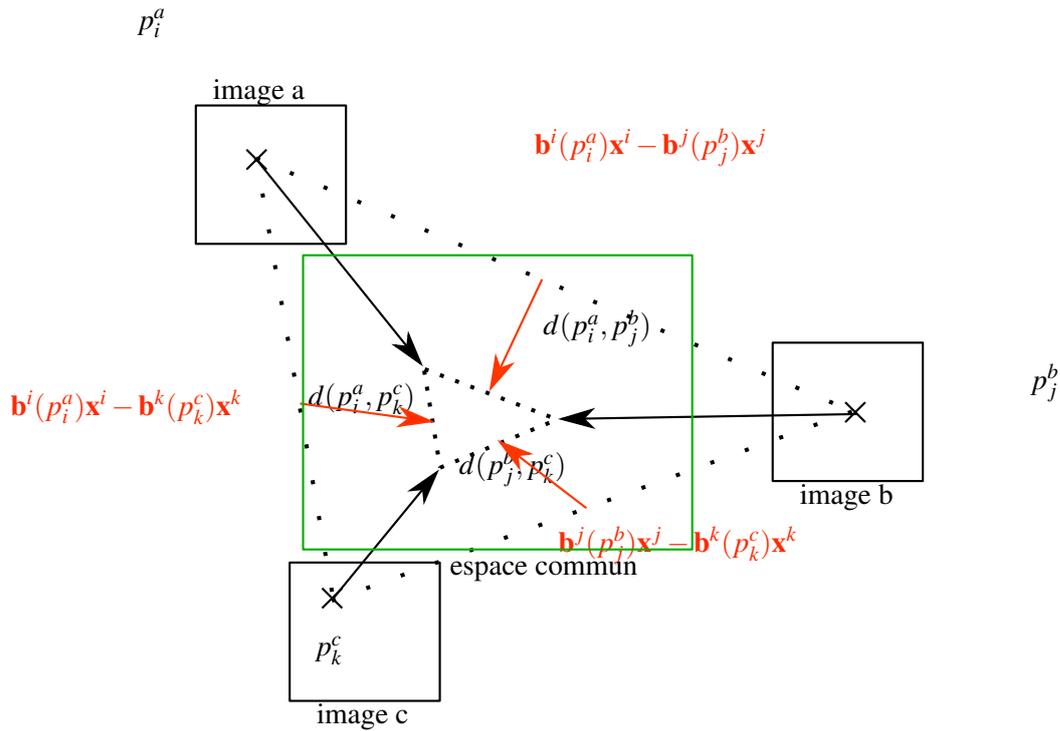


FIGURE 4.20 – Résultat de l’optimisation sur les couples appairés avec trois points d’intérêt p_i^a , p_j^b , p_k^c et 3 couples appairés (p_i^a, p_j^b) , (p_i^a, p_k^c) et (p_j^b, p_k^c) . Nous pouvons observer l’existence d’une énergie de régularisation non-nulle et de déformations trop faibles.

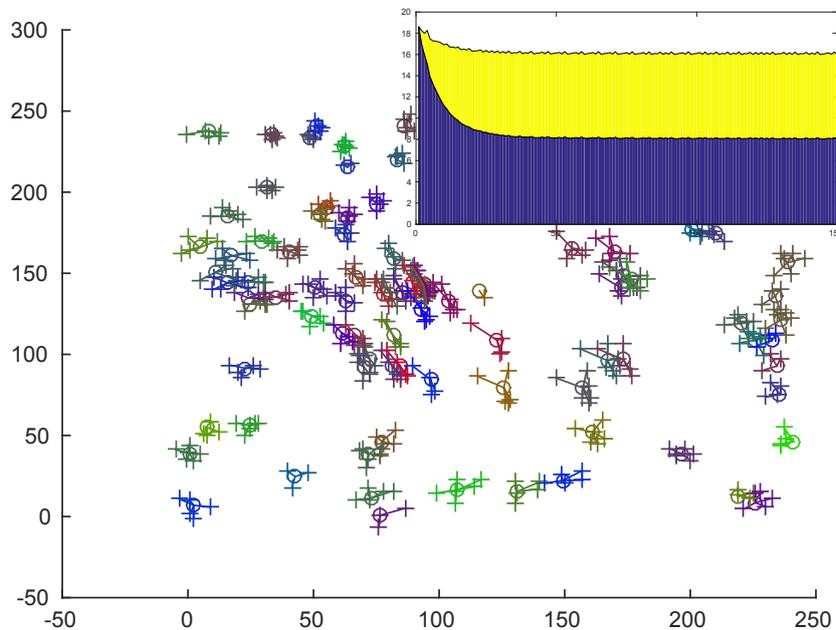


FIGURE 4.21 – Positions finales et courbes de convergence avec une régularisation sur les couples appairés. Nous pouvons observer qu’il reste une partie des déformations non capturée par l’estimation du champs de déformation. Grâce aux courbes cumulées de convergences, nous pouvons observer un compromis entre les deux termes : la régularisation (en jaune) et l’erreur de distance (en bleue).

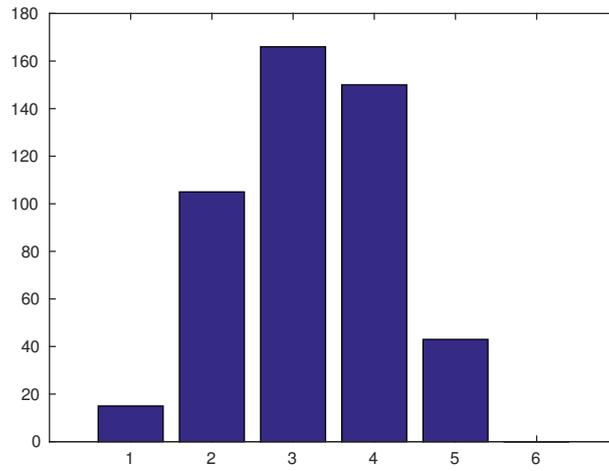


FIGURE 4.22 – Histogramme du nombre moyen d'appariement pour un point d'intérêt.

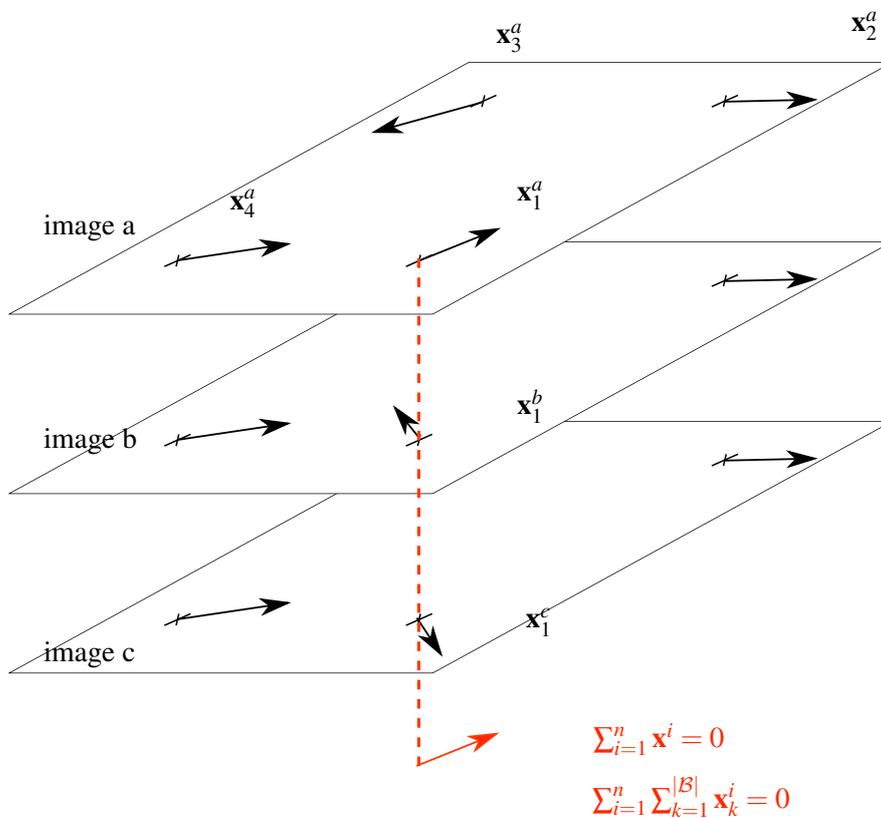


FIGURE 4.23 – Exemple de contrainte inter-images sur les points de contrôle.

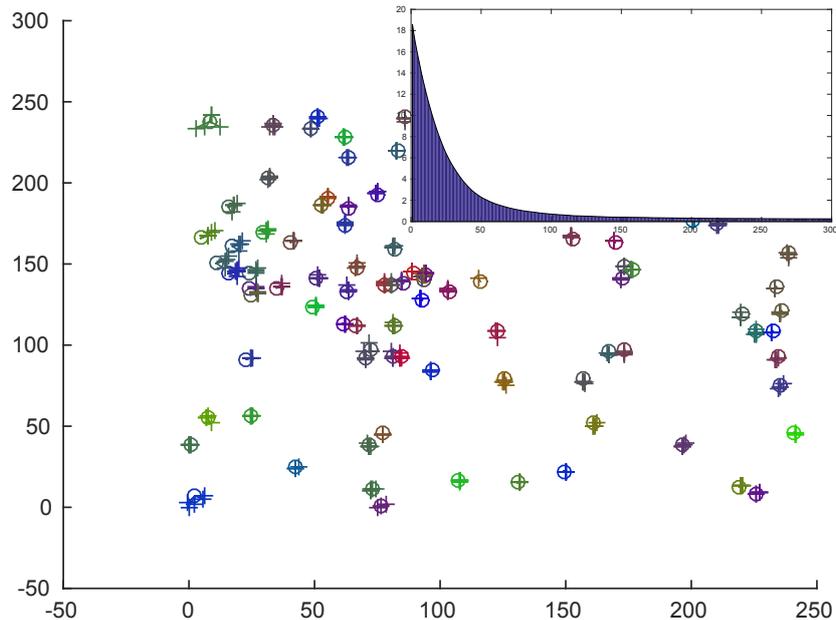


FIGURE 4.24 – Positions finales et courbes de convergence avec une contrainte inter-images sur les points de contrôle. Nous pouvons observer des positions finales qui possèdent une erreur de distance inférieure aux précédentes solutions. Nous pouvons observer sur la courbe représentant la fonction énergie que l'erreur tend vers une valeur proche de 0. Note : La contrainte n'étant pas une régularisation au sein de la fonction objectif, il n'y a pas de terme de régularisation.

Cette régularisation a deux spécificités vis-à-vis des autres approches évaluées ci-dessus. La première est un découplage entre le terme de la fonction objectif lié à la distance entre les points d'intérêt appairés et la suppression de ce degré de liberté surnuméraire. Ce découplage permet d'éviter un compromis entre les deux termes de la fonction d'énergie : l'erreur-distance et la régularisation. La deuxième spécificité amplifie cette indépendance de ces deux aspects de l'optimisation. En effet, nous utilisons une contrainte dure en retirant la moyenne de déformation à chaque point de contrôle après chaque pas d'optimisation. L'application d'une contrainte dure permet de découpler de manière forte les deux aspects, à travers une méthode en deux étapes alternées, la figure 4.24 montre la bonne convergence en utilisant cette contrainte.

4.9 Conclusion

Nous avons montré les avantages d'utiliser des mises en correspondance de groupes utilisant un graphe complet. Ceci permet une robustesse dans le cadre de jeux de données d'imagerie médicale hétérogènes qui peuvent contenir des zones imagées différentes. Notre approche parcimonieuse, utilisant uniquement des points d'intérêt et une interpolation par spline sur les intervalles d'une grille, permet de dépasser les limites actuelles du nombre d'images pouvant être recalées ensemble, en un temps raisonnable.

Chapitre 5

Applications et évaluations

Sommaire

5.1	Introduction	110
5.1.1	Méthode	110
	Processus	110
	Extraction des points d'intérêt	110
	Recalage rigide	110
	Mise en correspondance déformable	111
5.2	Jeux de données	111
5.2.1	Base de données VISCERAL	111
	Images	111
	Points de repère anatomique	112
5.3	Comparaison avec ANTs	114
5.3.1	Introduction	114
5.3.2	Mesure d'erreur sur les points de repère anatomique	114
5.3.3	Procédure de ANTs	114
5.3.4	Résultats : jeu de données A	114
	ANTs - Corrélation croisée	116
	ANTs - Information mutuelle	116
	ANTs - Information mutuelle - sous-échantillonnage d'un facteur 2	116
	ANTs - Information mutuelle - sous-échantillonnage d'un facteur 4	117
	Notre approche FGDR	117
5.3.5	Résultats : jeu de données B	120
5.3.6	Résultats : jeu de données $A \cup C$	123
5.3.7	Résultats : jeu de données $A+B+C$	124
5.3.8	Comportement de FGDR pour différents jeux de données	127
5.4	Conclusion	129

5.1 Introduction

Afin de mettre en application la méthode décrite au chapitre précédent, nous avons créé l’approche “Fast Deformable Groupwise Registration” (FGDR). Elle consiste à mettre en correspondance des images médicales issues de larges bases de données de manière déformable.

Dans les expériences qui vont suivre nous attacherons de l’importance à la fois à la comparaison de la précision de notre approche avec une approche de référence (ANTs, [AVANTS et collab. \[2008\]](#)) et aussi au comportement de notre algorithme en fonction de la dimension de la base d’images.

5.1.1 Méthode

Afin de respecter les conditions de proximité à la solution de notre méthode, une première étape de mise en correspondance rigide par groupe est effectuée à l’aide de notre approche FDR. La deuxième étape consiste à utiliser notre approche déformable FGDR sur ces images recalées.

Avant ces deux étapes de mise en correspondance rigide puis déformable, nous extrayons les points d’intérêt Surf3D. Nous pourrions noter que les deux méthodes FGR et FGDR utilisent la même source de points d’intérêt, ce qui supprime rapidement l’utilisation des images complètes.

Processus

Le processus complet nécessite les étapes suivantes :

1. Extraction des points d’intérêt (chapitre 3.3) :
 - Détection des points d’intérêt
 - Caractérisation des points d’intérêt
2. Recalage rigide (chapitre 3.5) :
 - Calcul de la transformation de chaque paire par RANSAC
 - Élagage du graphe
 - Recalage rigide de groupes
3. Mise en correspondance déformable (chapitre 4) :
 - Calcul des appariements de tous les points
 - Calcul du recalage déformable

Extraction des points d’intérêt

Afin d’extraire les points d’intérêt de manière générique et de pouvoir les utiliser dans les différentes étapes de notre approche, nous extrayons tous les maxima en espace-échelle pour la détection des blobs (seuil de détection à 0). Cette extraction exhaustive permet d’appliquer ensuite un seuil spécifique à chaque application (rigide et déformable).

L’extraction est effectuée sur des images rééchantillonnées de manière à avoir des voxels isotropes de $0.75mm^3$.

Recalage rigide

Nous utilisons les mêmes paramètres que décrit dans le chapitre 3 qui, pour mémoire, sont :

1. **Appariements** des points d’intérêt, entre toutes les paires possibles d’images au sein de l’ensemble. Nous utilisons une distance maximale au plus proche voisin de 0.3 et un rapport au second de 0.98.

2. **Estimation des paramètres** de toutes les transformations avec RANSAC. Nous faisons 6000 tirages aléatoires par paire, avec un seuil de décision de distance de $40mm$, afin de trouver la meilleure transformation pour chaque paire d'images. Un seuil de $40mm$ peut paraître élevé, mais il convient de rappeler que c'est un recalage rigide. Il est donc nécessaire de choisir un seuil qui permet d'accepter les couples de points correspondant mais éloignés à cause d'une déformation non prise en compte par le modèle.
3. **Élagage du graphe** de mises en correspondance en retirant des arêtes selon le critère du nombre de vrais positifs défini par RANSAC. Nous conservons $3n$ arêtes, tout en garantissant la connexité du graphe.
4. **Résolution des équations laplacienne**, en utilisant pour ancre une image dont nous fixons arbitrairement les coordonnées à 0 .

Mise en correspondance déformable

Pour chaque image, nous gardons jusqu'à 8000 points d'intérêt s'ils ont une réponse de détection supérieure à 2000.

L'appariement est effectué entre tous les points de toutes les images avec une distance inférieure à 0.22 (sans dimension) et un rapport de distance au second de 0.95 maximum.

La pyramide de splines utilisée contient 3 niveaux, avec une taille de grille de, respectivement, $200mm^3$, $100mm^3$ et $50mm^2$. L'algorithme d'optimisation utilise des pas de descente de gradient sous sa forme optimisée, en ralentissant la vitesse de convergence avec α , le pas de descente, ayant pour valeur, par niveau de pyramide, 0.1, 0.05, 0.025. Pour chaque niveau de pyramide 200 itérations de descente de gradient sont effectuées.

Afin d'améliorer la vitesse d'exécution, l'estimation des paramètres locaux des distributions de la pondération EM est effectuée toute les 10 itérations.

Les expérimentations sont mises en œuvre avec un ordinateur possédant 24 cœurs (48 cœurs logiques avec l'hyperthreading, à 2.3GHz) et 128Go de RAM.

5.2 Jeux de données

5.2.1 Base de données VISCERAL

Afin d'évaluer notre approche, nous utilisons la base de données VISCERAL [[LANGS et collab., 2013](#)].

Images

Cette base se compose de 3 groupes (voir figure 5.1) :

- **Groupe A** : Ce groupe est composé de 20 images provenant de scanner CT 3D avec un agent de contraste. Pour ces images, la partie imagée englobe le thorax et l'abdomen. Les dimensions de ces images sont environ de $500 \times 500 \times 400$ voxels de dimensions $0.7 \times 0.7 \times 1.5mm$. L'ensemble des images nécessite 7.6Go d'espace de stockage.
- **Groupe B** : Ce groupe se compose de 20 images provenant de scanner CT (sans agent de contraste). L'ensemble du corps est imagé. Les dimensions sont environ de $500 \times 500 \times 900$ voxels de dimensions $0.8 \times 0.8 \times 1.5mm$. L'ensemble des images du groupe B nécessite 18Go d'espace de stockage.
- **Groupe C** : Ce groupe se compose de 63 images. Il possède les mêmes caractéristiques que le groupe A. Le consortium VISCERAL fait référence à ces images comme "Silver Corpus" car il

comporte des repères anatomiques reportés automatiquement. L'ensemble des images du groupe C nécessite 31.5Go d'espace de stockage.

Points de repère anatomique

La base de données VISCERAL contient aussi des points de repère anatomiques. Ces points de repère sont positionnés sur les images et annotés manuellement par des spécialistes. Seuls les groupes A et B sont annotés, contrairement au groupe C qui est seulement annoté par report automatique d'annotation, d'où l'appellation "Silver", par comparaison au "Gold standard" des groupes A et B. Nous n'utiliserons pas les repère du groupe C.

- **Groupe A** : Les images du groupe A sont annotées avec 45 points de repère anatomiques. Les points de repère ne sont pas présents dans toutes les images, chaque image contenant en moyenne entre 41 et 42 points.
- **Groupe B** : Les images du groupe B sont annotées avec 53 points de repère anatomiques. Chaque image en contient entre 52 et 53 en moyenne.

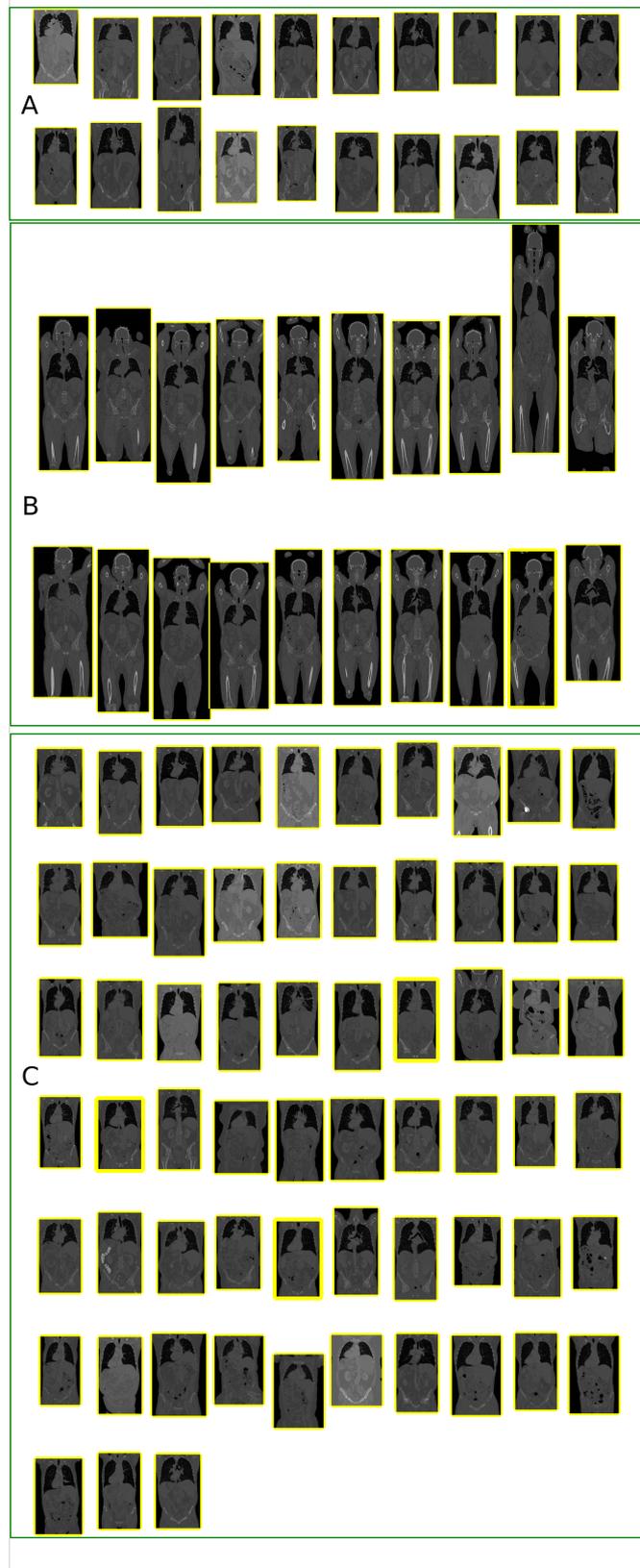


FIGURE 5.1 – Illustration des jeux de données A, B et C

5.3 Comparaison avec ANTs

5.3.1 Introduction

À notre connaissance, notre méthode est la première qui utilise seulement des points d'intérêt pour faire de la mise en correspondance de groupes. Afin de nous comparer à d'autres travaux, nous avons utilisé des méthodes de recalages utilisant les voxels. Un des logiciels de référence dans ce domaine, dont le code source est disponible, est ANTs (Advanced Normalization Tools, [AVANTS et collab. \[2008\]](#)). Cette famille de méthodes permet la mise en correspondance de groupes d'images en utilisant des algorithmes standards.

ANTs a d'abord été conçu pour la mise en correspondance des structures du cerveau. Cette objectif initial nous montre une fois de plus que la mise en correspondance d'images corps entier est un challenge encore non résolu.

La comparaison avec ANTs consiste à confronter notre méthode avec la mise en correspondance de groupes par approche en étoile. Cette approche est définie par le script *antsMultivariateTemplateConstruction2.sh* fourni avec le logiciel ANTs. Ce script propose une mise en correspondance de groupes en étoile grâce à la génération d'une image de référence à partir de l'ensemble des images.

5.3.2 Mesure d'erreur sur les points de repère anatomique

Grâce à la présence des points de repère placés manuellement par des experts, l'utilisation de ce jeu de données permet l'évaluation objective de la précision des deux méthodes. Ces points de repère anatomiques permettent une comparaison fiable et proche de la réalité du terrain, même s'ils n'évaluent la précision qu'à un nombre restreint de positions.

La mesure d'erreur E_r est définie par :

$$E_r = \sum_{a \in \mathcal{I}} \sum_{i \in \mathcal{L}} \|\tau^i(\mathbf{I}_i^a) - \bar{\mathbf{I}}_i\| \quad (5.1)$$

où \mathbf{I}_i^a est le vecteur des coordonnées du point de repère i provenant de l'image a déformée dans l'espace commun.

$\bar{\mathbf{I}}_i$ est le vecteur de la position moyenne pour toutes les images du point repère i dans l'espace commun :

$$\bar{\mathbf{I}}_i = \frac{1}{n} \sum_{a \in \mathcal{I}} \tau^i(\mathbf{I}_i^a) \quad (5.2)$$

5.3.3 Procédure de ANTs

Le script par défaut de ANTs exécute 5 itérations de mise en correspondance par paire d'images entre chaque image et la référence. Le premier est un recalage rigide tandis que les autres sont des recalages déformables. La corrélation croisée est le critère de similarité utilisé par défaut par ANTs. Les 4 étapes de mise en correspondance déformable par paires se font dans une logique d'approche pyramidale : A chaque itération de recalage de paire, une pyramide de raffinement est utilisée. La figure 5.2 propose un schéma synthétique de l'approche.

5.3.4 Résultats : jeu de données A

Grâce aux 20 images qu'il contient, le jeu de données A permet un test de mise en correspondance de groupes "classiques" car ressemblant par le nombre d'images aux jeux de données utilisés dans les publications liées aux méthodes de l'état de l'art.

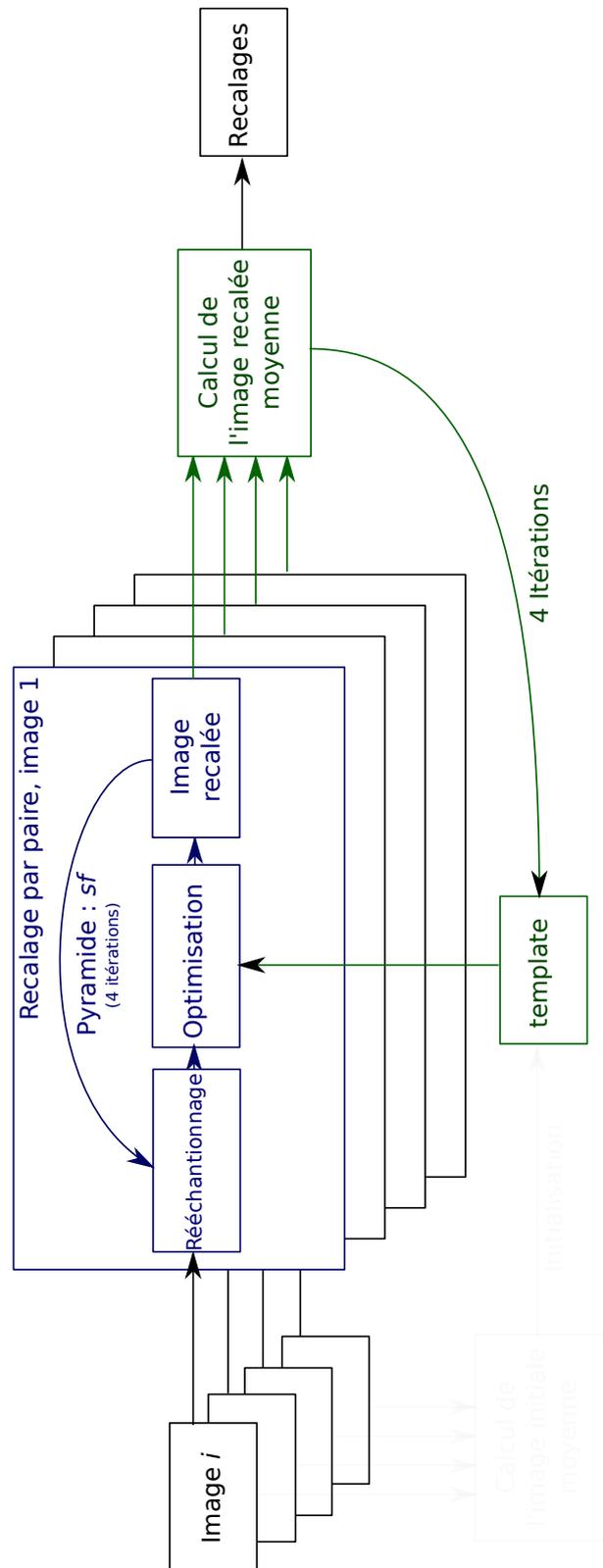


FIGURE 5.2 – Schéma bloc de la procédure de recalage par groupe de ANTs. L'image moyenne est initialisée par la moyenne des images d'entrée.

Nous nous attacherons dans les prochaines parties à détailler le comportement de la méthode ANTs avec différents paramètres et notre méthode, afin de pouvoir explorer en détail les comportements des deux approches. L'analyse se concentrera sur l'évaluation de l'effet du recalage sur les points de repère et une observation visuelle de la qualité de la mise en correspondance.

ANTs - Corrélation croisée

La première expérimentation consiste en la mise en correspondance des 20 images du jeu de données A de VISCERAL en appliquant pour ANTs les valeurs par défaut qui utilisent la corrélation croisée de son approche de mise en correspondance de groupes.

Résultat ANTs nécessite 327 heures pour mettre en correspondance toutes les images. Le résultat est une distance moyenne des points de repère de $14.0mm$ et une distance maximale de $152mm$.

Si l'on regarde plus en détail les résultats en observant l'image moyenne de la figure 5.3a, nous pouvons voir que les résultats ne sont pas probants. L'enveloppe est mal recalée (vue sagittale), tout comme l'ensemble des organes mous. Seuls les os (bassin, colonne vertébrale) semblent être recalés un peu plus convenablement. Le plancher bas des poumons (vue coronale), permet de voir la large erreur commise dans l'axe Z. De plus, nous avons observé que le recalage de l'une des images, visible figure 5.3b, a divergé.

En retirant l'image concernée et en recalculant la mise en correspondance sur les 19 images restantes, l'algorithme parvient à converger (en 317h), mais ne fournit pas un résultat sensiblement meilleur.

ANTs - Information mutuelle

Un deuxième paramétrage de ANTs consiste à utiliser l'information mutuelle comme métrique de similarité au lieu de la corrélation croisée. Ce critère permet de capturer des dépendances non linéaires et est sensiblement plus rapide.

Résultat En changeant la métrique de similarité entre les images de la corrélation croisée par l'information mutuelle, ANTs est beaucoup plus rapide : les 20 images sont mise en correspondance en 62.5 heures. Cependant, l'erreur moyenne augmente à $15.5mm$, avec une erreur maximale de $150mm$. Une image "fantôme" sur le bas des poumons nous indique qu'une des images est mal recalée au moins à cet endroit.

ANTs - Information mutuelle - sous-échantillonnage d'un facteur 2

Afin d'observer l'influence du sous échantillonnage sur la vitesse et la précision de ANTs, nous avons dans ce cas sous-échantillonné les images d'un facteur 2. Cela correspond à appliquer au paramètre de l'algorithme "shrinking factor" (sf) une valeur de $12 \times 6 \times 4 \times 2$.

Résultat De manière évidente, ANTs est plus rapide avec ce sous-échantillonnage : la mise en correspondance des 20 images nécessite 13.8 heures. En comparant les figures 5.3c et 5.3d, nous pouvons observer que le "fantôme" sur les bas des poumons n'est pas visible, ce qui indique que le sous-échantillonnage augmente la robustesse. L'erreur moyenne est de $11.3mm$, ce qui appuie le fait que ANTs bénéficie de la robustesse apportée par le sous-échantillonnage.

ANTs - Information mutuelle - sous-échantillonnage d'un facteur 4

Afin de trouver les limites de ANTs en terme de vitesse et de précision, nous utilisons un sous-échantillonnage de 4 ($sf = 20 \times 12 \times 6 \times 4$) associé à une mesure de similarité par information mutuelle.

Résultat Avec ces paramètres, ANTs calcule la mise en correspondance des 20 images en seulement 9.4 heures. Cette rapidité se fait au détriment de la précision, avec une erreur moyenne de $13.8mm$. Un sous-échantillonnage d'un facteur 2 est donc optimal en terme de précision. L'algorithme reste robuste, mais le sous-échantillonnage diminue la précision de manière significative.

Notre approche FGDR

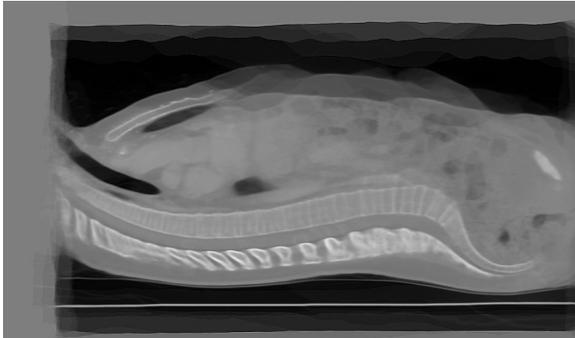
Notre approche a réussi à mettre en correspondance les 20 images du jeu de données A, tout en étant beaucoup plus rapide, en seulement 1h20. FGDR met en correspondance les images avec une erreur finale sur les points de repère de seulement $10.5mm$, ce qui est significativement meilleur que la méthode ANTs la plus précise (MI-S2 : $11.3mm$, p-value égale à 0.031 avec le test t de Welch unilatéral).

Résultat La figure 5.4 permet d'observer les résultats en détails. Sur ce graphique, nous avons séparé les points de repère en 5 régions anatomiques : 1-épaules (points de repère 1 à 4, 2-vertèbres (5 à 23), 3-torse (24 à 32), 4-reins (33 à 38) et 5-hanches (39 à 45). Les courbes des erreurs de notre approche suivent la même tendance que la courbe des erreurs de ANTs. Notre approche est compétitive avec ANTs.

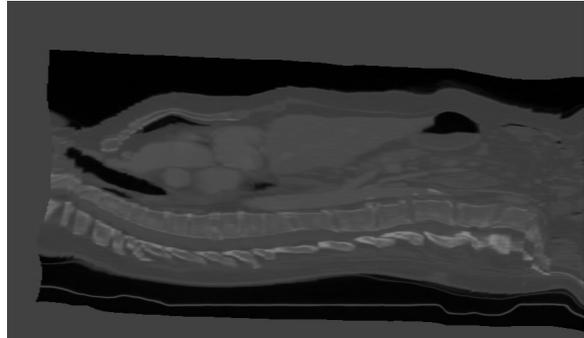
Images moyennes Il est intéressant de comparer les images moyennes de ANTs, figure 5.3d, et notre approche, figure 5.3e. Les images moyennes issues de ANTs semblent plus nettes, les bords sont plus précis et avec des transitions plus floues dans les zones de fort contraste que celles de notre approche FGDR. Cependant, si l'on observe attentivement les images, nous pouvons observer que notre approche est plus précise. Par exemple si nous observons l'enveloppe des individus ou les zones contenant les organes mous. Ce comportement peut s'interpréter par la différence d'approche entre les deux méthodes. ANTs utilise directement l'intensité des images, particulièrement les différences de contraste, pour diriger son optimisation. Les zones de fort contraste sont donc naturellement bien recalées. Quant à notre méthode, elle n'utilise pas cette information, mais des points d'intérêt, ces zones ne sont donc pas mieux recalées que les autres. Notre méthode a un comportement plus uniforme (écart-type de 7.9 au lieu de 8.3). Nous pouvons aussi noter que les deux méthodes ont des difficultés pour mettre en correspondance les organes mous. Cette limitation est due à l'utilisation de rayons X, qui sont peu absorbés par les tissus mous.

La figure 5.5 illustre les courbes de convergence de notre méthode. La courbe bleue représente l'énergie liée à la fonction objectif. Afin d'avoir une mesure homogène à une longueur, nous prenons sa racine carré. Cette formulation revient à observer la racine carré de la moyenne des carrés des distances entre tous les points appariés (restreinte aux vrais positifs estimés). La courbe noire décrit l'erreur sur les points de repère. La courbe jaune décrit l'erreur sur les points de repère en utilisant le M-estimateur à la place de notre pondération EM. Nous pouvons observer dans ce cas que l'erreur sur les points de repère n'est pas stable et remonte. La pyramide de splines est raffinée aux itérations 200 et 400. Le comportement en marche d'escalier de la courbe bleue est dû à l'estimation des paramètres des distributions des vrais et faux positifs une fois toutes les 10 itérations.

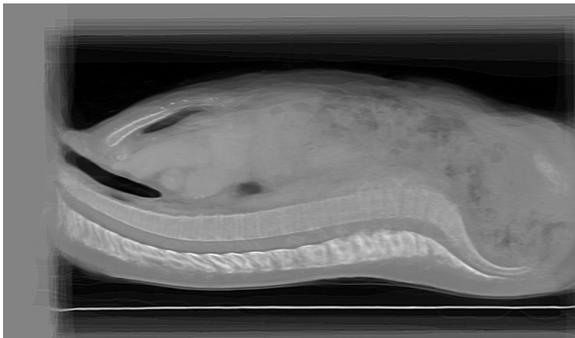
En observant la courbe bleue, on peut noter que les étapes de raffinement sont effectuées alors qu'il n'y a pas convergence vers une valeur constante. Nous avons observé que l'erreur réelle (représentée par les points de repère) se stabilise avant l'erreur de la fonction objectif des points d'intérêt. Ce phénomène peut s'apparenter à ce que l'on observe en théorie de l'apprentissage entre les erreurs des ensembles



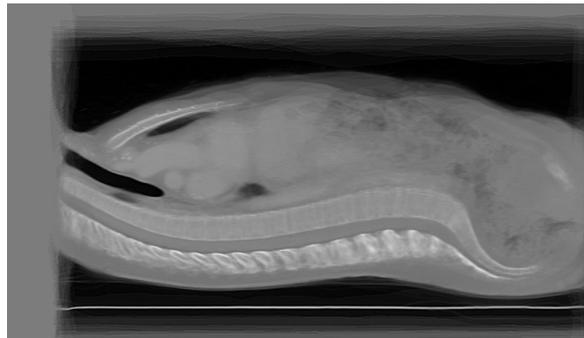
(a) Image moyenne issue de ANTs, utilisant le critère de corrélation-croisée.



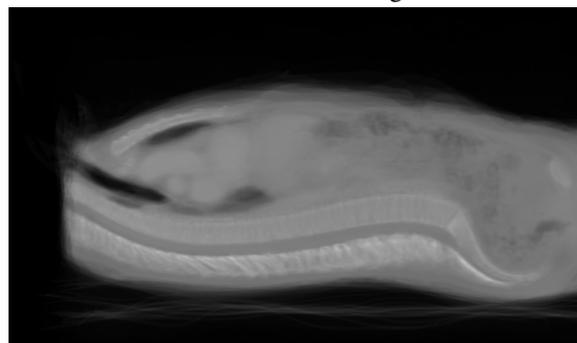
(b) Une des images transformée dans l'espace commun grâce à ANTs (Corrélation Croisée). L'algorithme n'a pas convergé vers une bonne position. Nous pouvons observer de larges déformations ne correspondant pas à l'image moyenne de la figure 5.3a.



(c) Image moyenne issue de ANTs, utilisant le critère d'information mutuelle.



(d) Image moyenne issue de ANTs, utilisant le critère d'information mutuelle avec un facteur de sous échantillonnage de 2.



(e) Image moyenne issue de notre approche FGDR.

FIGURE 5.3 – Résultats des différentes approches pour les 20 images du jeu de données A.

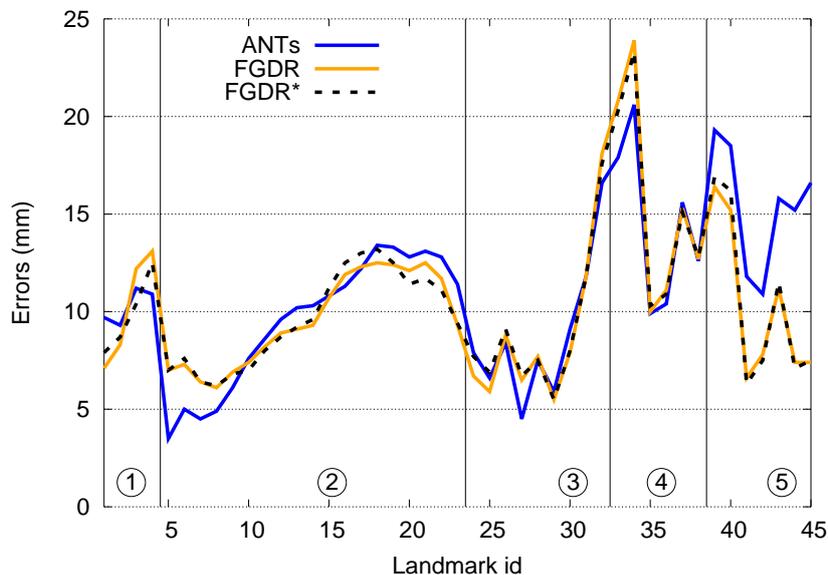


FIGURE 5.4 – Comparaison entre notre approche FGDR et ANTs. La Distance est en millimètre. Les points de repère sont séparés en 5 régions (1 : Epaules, 2 : Vertèbres, 3 : Torse, 4 : Reins, 5 : Hanches)

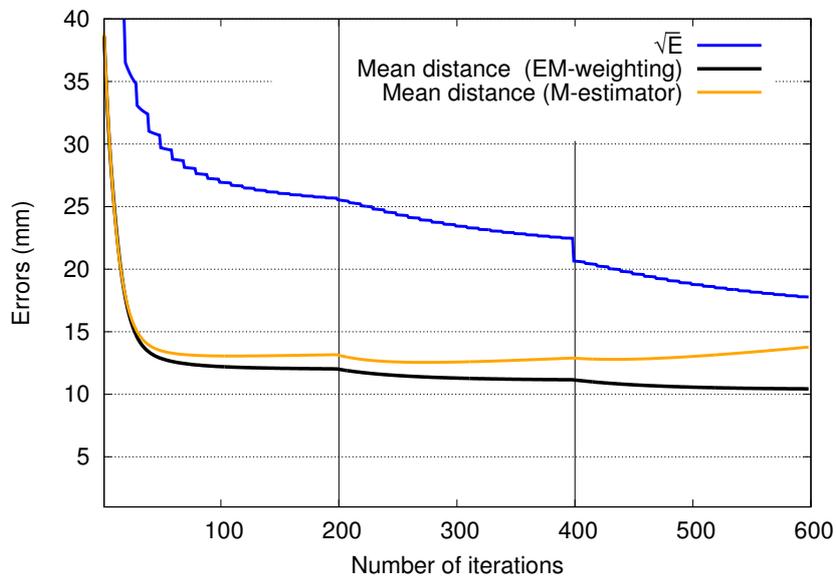


FIGURE 5.5 – Convergence du critère de minimisation et distance moyenne lors de la mise en correspondance des groupes A et C. Les distances sont en millimètres. La courbe bleu représente la racine carrée du critère défini par l'équation (4.15). Les courbes noires et oranges représentent l'erreur de distance moyenne globale des points de repère en utilisant, respectivement, la pondération EM et le M-estimateur de Tukey.

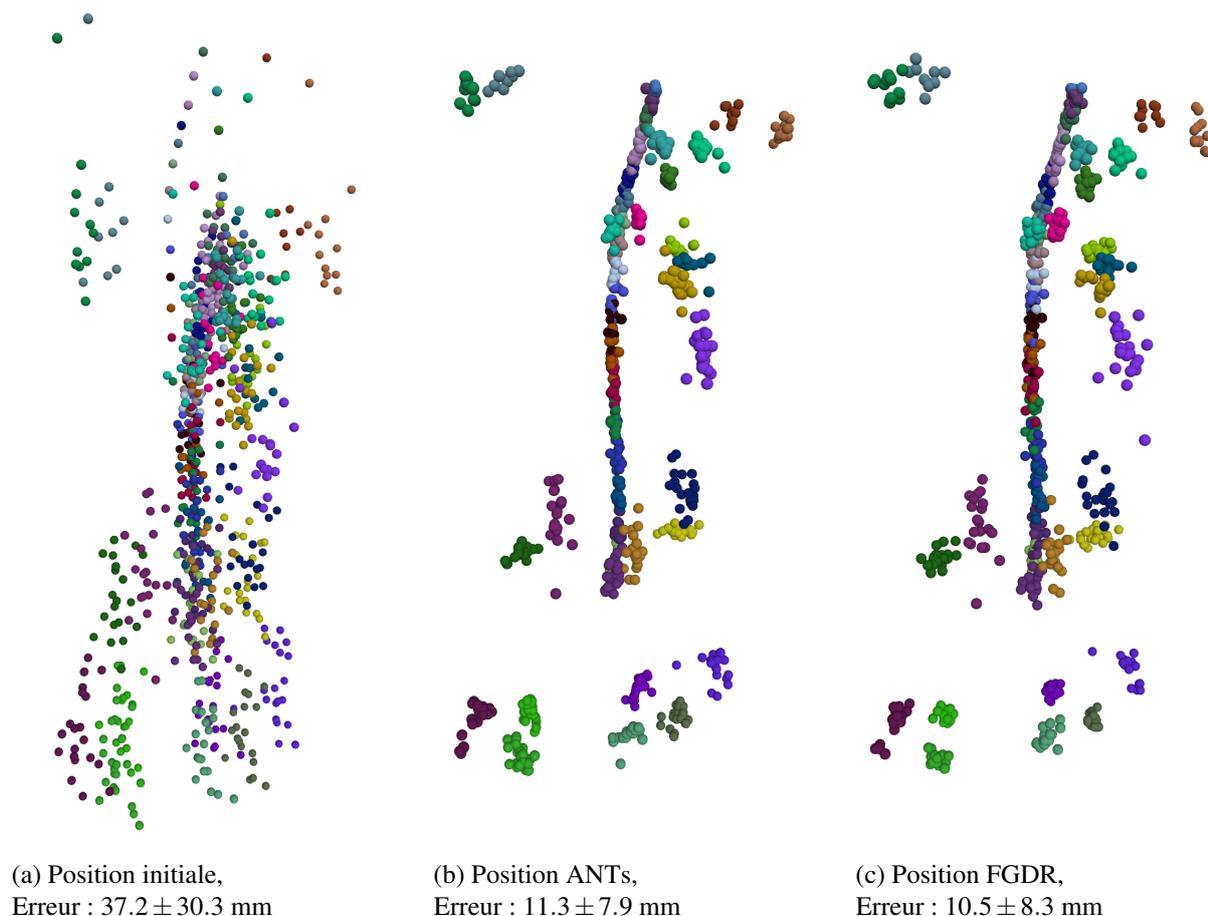


FIGURE 5.6 – Points de repère du groupe A affichés dans le même repère. Chaque couleur représente un point de repère (présent dans plusieurs images). De gauche à droite : Les positions initiales, avant recalage ; les positions finales estimées en appliquant les transformations issues de ANTs ; Les positions estimées après application de notre méthode FGDR.

d'apprentissage et de test. La notion d'arrêt prématuré est très utilisée dans ce cadre, pour éviter le sur-apprentissage [CARUANA et collab., 2000; PRECHELT, 1998]. Cependant, il pourrait sembler judicieux de définir un critère d'arrêt adaptatif.

Points de repère Les positions finales des points de repère visibles figure 5.6 permettent d'observer le comportement des deux mises en correspondances pour certains points de l'image. À gauche, nous pouvons observer les positions initiales des points d'intérêt, affichés dans un espace commun. Au centre, les points de repère sont déformés en appliquant les champs de transformation de ANTs (Information Mutuelle, $sf=2$). À droite, les résultats avec notre approche FGDR. Nous pouvons tout d'abord constater que l'approche FGDR est significativement meilleure en ce qui concerne l'erreur moyenne. En observant plus finement les dispositions des points, nous pouvons voir que les deux méthodes sont équivalentes de par leurs comportements : le recalage est meilleur au niveau du bassin (en bas) et au niveau de la colonne vertébrale. Les deux méthodes tirent partie du fort contraste de ces zones.

5.3.5 Résultats : jeu de données B

Le groupe B contient des images corps entier. Celui-ci contient des images ayant une plus grande variabilité et permet de vérifier la robustesse des méthodes.

Note : Un individu imagé du jeu de données B est particulièrement grand par rapport aux autres. Cette spécificité nous permet d'avoir un cas précis d'observation de la robustesse des méthodes.

De plus, de même que le jeu de données A, le jeu de données B contient des points de repère anatomiques permettant la comparaison quantitative de la précision des approches.

ANTs En prenant les paramètres optimaux de ANTs, à savoir : critère de similarité par information mutuelle et un sous-échantillonnage de 2, ANTs nécessite 34h pour mettre en correspondance les 20 images corps entier du jeu de données B, dont l'image moyenne est visible figure 5.8b. L'erreur sur les points de contrôle est relativement grande, 17.9mm. La figure 5.7 compare les mises en correspondance avec le jeu de données B. Nous avons extrait les squelettes de chaque volume et nous les avons superposés dans l'espace commun. La figure du milieu montre le résultat de la mise en correspondance de ANTs et nous permet de confirmer que la mise en correspondance n'est pas de bonne qualité : une image a le bassin décalé vers le bas, tandis que les autres images se superposent avec peu de précision.

ANTs manque de robustesse pour traiter ce jeu de données. Plusieurs raisons peuvent expliquer les mauvaises performances pour ce jeu de données. Tout d'abord, les paramètres optimaux ont été définis pour un autre jeu de données. Pour celui-ci, ils sont peut-être différents. Les paramètres n'étant pas généralisables, cette nécessité de les régler manuellement constitue une première faiblesse de ANTs, car il faudrait ajuster les paramètres spécifiquement pour ce jeu de données, ce qui l'empêche d'être une méthode entièrement automatique. Une deuxième source de différence avec le jeu de données A est la taille des images, ainsi que l'amplitude des déformations. Les corps étant imagés entièrement, les déplacements peuvent être plus grands et ANTs semble limité dans sa capacité à estimer de grands déplacements. Ce qui expliquerait la position du bassin qui est trop basse.

FGDR En utilisant les mêmes paramètres que pour les autres expériences, notre approche réussit à mettre en correspondance les 20 images corps entier du jeu de données B avec une bonne précision. L'erreur sur les points de repère est de 10.1mm, ce qui peut être considéré comme faible vues les images en question : une erreur de 10.1mm pour des individus mesurant 1.3m de haut (soit une erreur relative de 0.8%).

Nous pouvons observer que l'erreur maximale se concentre sur les bras des individus, ce qui est un défi vu la variabilité des positions des différents individus (un des individus a les bras le long du corps). Cependant, cette difficulté ne vient pas interférer avec les autres parties du corps et l'individu plus grand que les autres est bien mis en correspondance avec le reste du groupe.

Grâce à l'utilisation de la pondération EM, nous avons accès à une estimation locale des paramètres des distributions des vrais et faux positifs. Les couleurs appliquées sur l'image de la figure 5.7 correspondent à l'estimation de l'erreur locale sur les points d'intérêt. C'est la variance de la distribution des vrais positifs qui nous permet d'avoir une estimation de l'erreur de mise en correspondance. Cet indicateur nous montre que le bassin est une zone avec une faible variance de distance au sein des vrais positifs et donc de faible erreur de recalage. Nous pouvons observer cela visuellement par la superposition de bonne qualité des bassins des différents individus. A contrario, l'erreur est plus grande au niveau du torse. Une explication de l'origine de cette différence de précision entre le bassin et le torse découle du processus d'extraction des points d'intérêt. Le bassin est une zone riche, tant en contraste (les os sont très contrastés en imagerie scanner) qu'au niveau de la forme (le bassin est un os ayant une géométrie complexe) qui favorise la détection de points de qualité. Le torse est une zone plus difficile pour l'extraction de points d'intérêt, avec des organes mous, et des structures répétitives similaires (vertèbres, côtes).

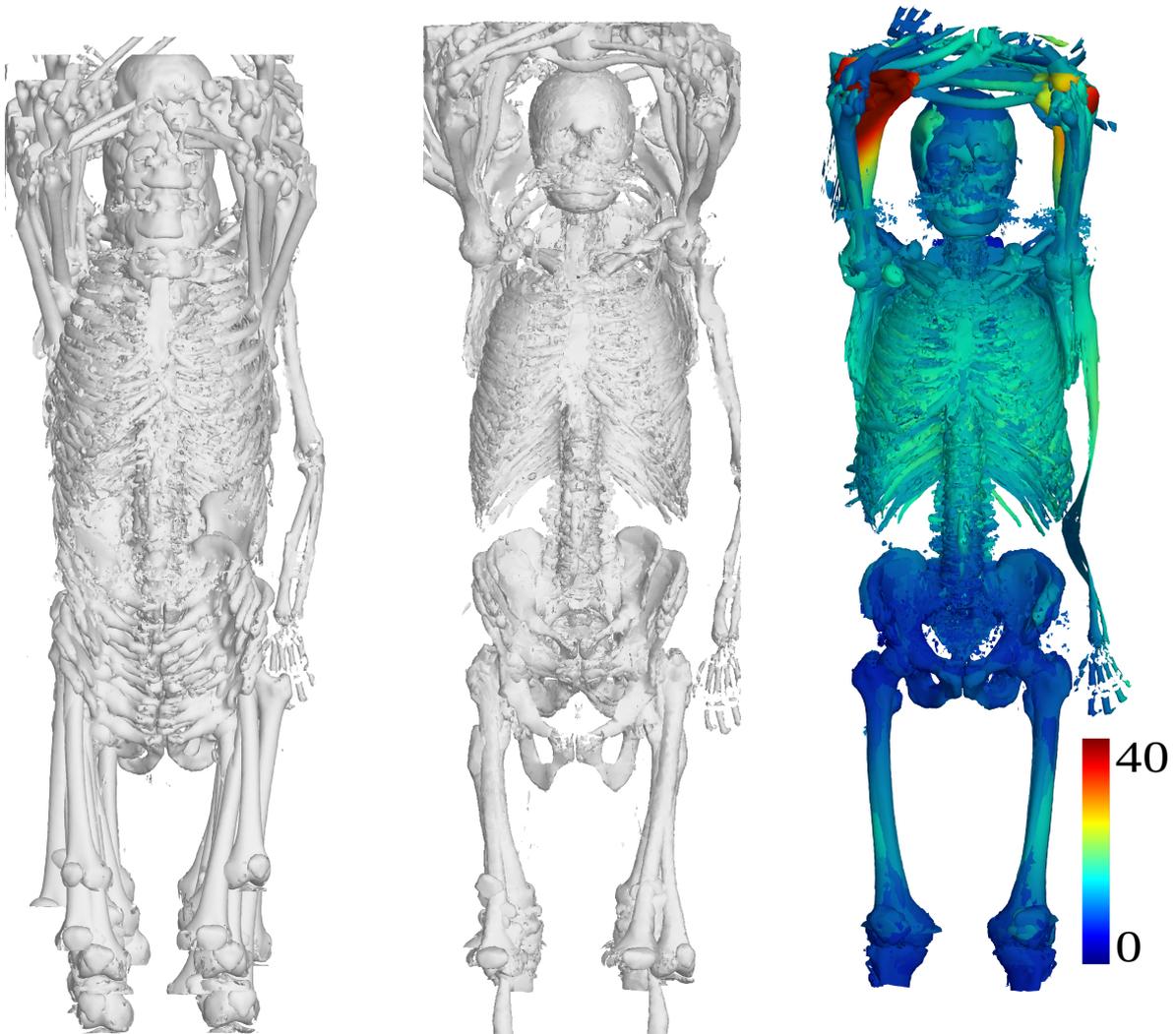
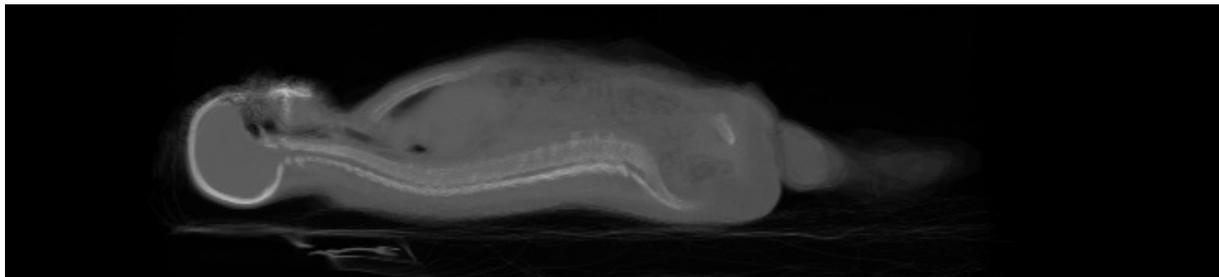


FIGURE 5.7 – Squelettes extraits lors de la mise en correspondance des 20 images corps entier du groupe B. À gauche, la superpositions des squelettes avant la mise en correspondance. Au centre, après mise en correspondance grâce à la méthode ANTs et à droite grâce à notre méthode BFGR. La superposition des squelettes indique que, visuellement, notre méthode est plus précise. L'échelle de couleur utilisée pour notre méthode correspond au paramètre d'échelle issue de la distribution des vrais positifs, cela permet une estimation de la précision du recalage.



(a) Image moyenne issue de notre approche FGDR.



(b) Image moyenne issue de ANTs, utilisant le critère d'information mutuelle avec un facteur de sous échantillonnage de 2.

FIGURE 5.8 – Résultats des différentes approches pour les 20 images corps entiers du groupe B

5.3.6 Résultats : jeu de données $A \cup C$

Afin de pouvoir tester la capacité des approches à traiter plus de données, nous avons assemblé les jeux de données A et C afin de créer un ensemble d'images contenant 83 images.

Les images issues de C ne contiennent pas de points de repère anatomiques placés par des médecins, mais les repère des images de A permettant d'avoir une évaluation de la précision sur une fraction des images. Pour l'évaluation, ce sont donc 83 images qui sont mises en correspondance, mais seulement 20 d'entre elles portent des points de repère et contribuent à la mesure de l'erreur de mise en correspondance.

ANTs Avec un jeu de données contenant 83 images, ANTs atteint ses limites en taille du jeu de données. Pour effectuer les 5 itérations de son schéma de mise en correspondance, ANTs nécessite 82.8 heures de calculs. L'erreur finale sur les points d'intérêt est de $44.7mm$, ce qui montre bien qu'il ne parvient pas à recalcr les images.

Plusieurs hypothèses peuvent être posées afin d'expliquer cette limitation. La première est l'utilisation d'image moyenne. Cette expérience semble confirmer que l'utilisation d'image moyenne pour de grandes bases de données est problématique. En observant l'image moyenne, figure 5.9a, nous pouvons observer qu'elle est très chaotique. Le moyennage d'un grand nombre d'images non parfaitement superposées diminue le contraste. De manière générale, plus le nombre d'images augmente, plus la qualité de l'image moyenne est bonne si le recalage est précis, et on peut constater ici qu'elle est sûrement insuffisante pour avoir une métrique d'information mutuelle efficace.

FGDR La mise en correspondance des 83 images des jeux de données A et C est effectuée de manière satisfaisante par notre approche, figure 5.9b. L'erreur finale sur les points de repère des images du jeu de données A est de $10.4mm$ ce qui est significativement meilleur que ANTs (p-value : 0.017). Le rapport

estimé de faux positifs est de 70% pour cette expérience, ce qui justifie la nécessité de notre approche par pondération EM.

Comparaison pour des bases de données de 20 et 83 images Nous faisons l’hypothèse, dans cette recherche, d’une amélioration de la précision lorsque le nombre d’image croît. Le nombre d’appariement augmente quadratiquement par rapport au nombre d’images. Selon nous, cette différence de croissance peut améliorer la précision pour plusieurs raisons. La première est une meilleure capacité à rejeter les faux positifs tout en ayant assez de vrais couples appairés. La deuxième est liée au nombre d’appariements pour chaque point d’intérêt, qui augmente avec le nombre d’images et contraint mieux la position finale du point dans l’espace commun. La troisième raison est liée aux données utilisées. Les images proviennent d’individus différents et possédant une grande variabilité. En augmentant la taille de la base, nous augmentons le nombre d’images similaires, permettant de bonnes mises en correspondance. D’un point de vue de la théorie des graphes, nous avons un graphe complet plus grand, ce qui augmente le nombre de chemins entre chaque paire d’images. Dit autrement, la probabilité d’avoir un chemin de proche en proche similaire augmente.

Si l’on porte attention aux résultats de notre approche avec 20 et 83 images, des erreurs respectivement de 10.5mm et 10.4mm, nous pouvons observer une diminution de l’erreur. Cependant, cette différence n’est pas significative (p-value = 0.38) et il semble nécessaire de mener d’autres expérimentations, avec plus d’images, pour confirmer ou infirmer nos hypothèses.

5.3.7 Résultats : jeu de données A+B+C

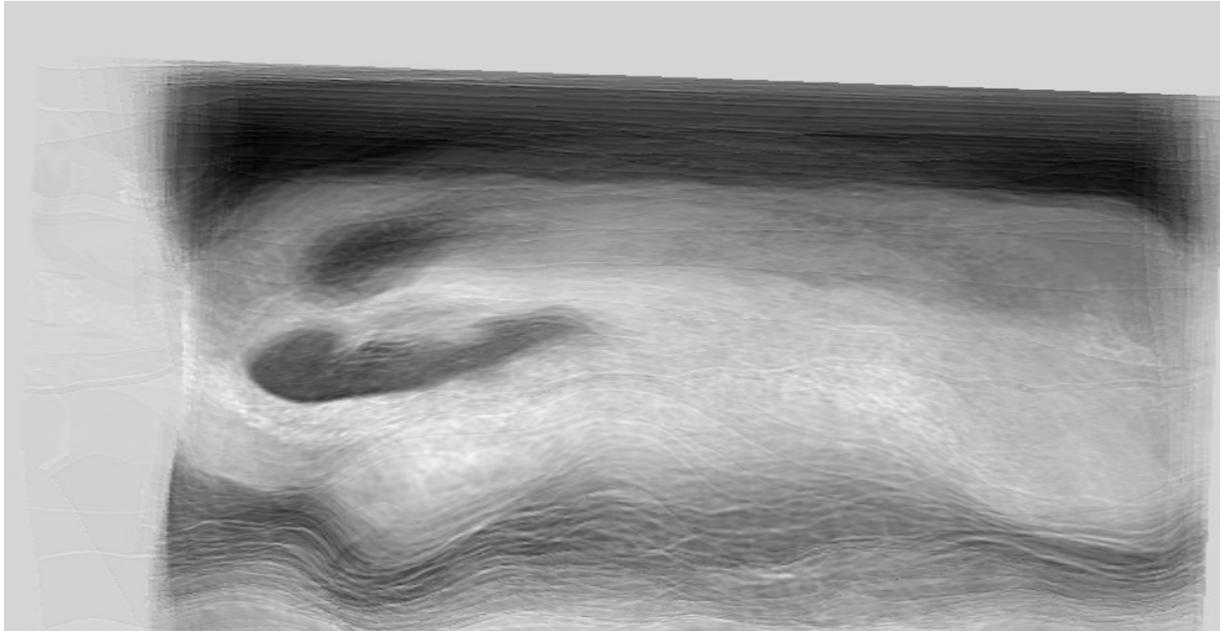
Un dernier jeu de données a été créé à partir des ensembles $A \cup B \cup C$. Ce dernier contient 103 images et permet de démontrer la capacité des algorithmes à dépasser la centaine d’images.

Étant donné que les images de l’ensemble B contiennent des images corps entier, nous avons extrait la partie thoracique et abdominale de ses images afin d’avoir un ensemble cohérent ayant les mêmes zones imagées.

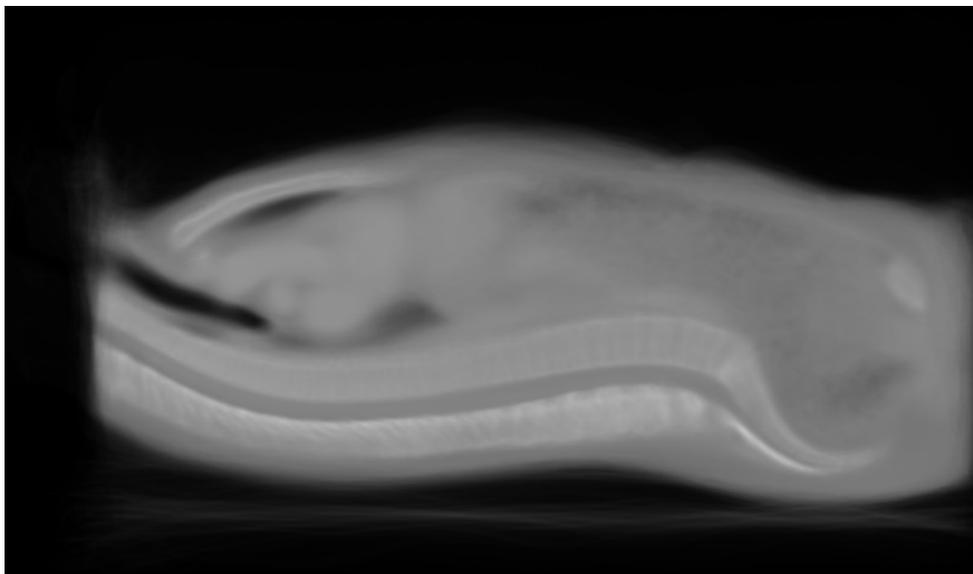
ANTs Nous n’avons pas utilisé ANTs pour cette expérience, celui-ci ayant déjà échoué avec 83 images. De plus les ressources nécessaires (espace mémoire, temps de calcul disponible) pour effectuer la mise en correspondance de 103 images n’étaient pas disponibles.

FGDR Notre approche réussit à mettre en correspondance les 103 images des jeux de données A, B et C recalés conjointement. L’erreur finale est de 11.1mm ce qui est du même ordre de grandeur que les autres expériences (voir figure 5.10 et le tableau 5.1).

Cette expérience ne permet pas de confirmer l’hypothèse d’une amélioration de la précision avec le nombre d’images, l’erreur finale étant supérieure aux expériences précédentes avec 20 et 83 images. Cette augmentation peut s’expliquer par la légère différence de modalité : avec agent de contraste (groupe A et C) et sans agent de contraste (groupe B) et une moindre qualité des images du jeu de données B ajouté. Cette différence implique deux aspects : nous mettons en correspondance des images provenant de protocoles d’acquisition légèrement différents et nous ajoutons des images ayant une qualité moindre. Pour ce dernier point, il faut savoir que le produit de contraste permet comme son nom l’indique, d’augmenter le contraste des organes dit “mous” bien irrigués. Cette augmentation de contraste peut directement bénéficier à la détection et la description des points d’intérêt pour ces zones.



(a) Image moyenne issue de ANTs, utilisant le critère d'information mutuelle sur les 83 images des groupes A+C réunis avec un facteur de sous échantillonnage de 4.



(b) Image moyenne issue de notre approche BFGR avec les 83 images des groupes A+C réunis.

FIGURE 5.9 – Images moyennes des différentes approches pour les 20 images du groupe A+C

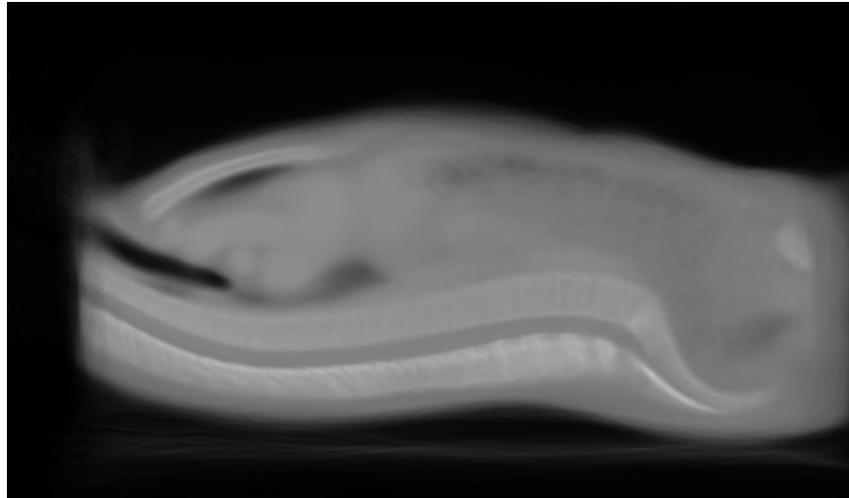


FIGURE 5.10 – Image moyenne issue de notre approche BFGR avec les 103 images des groupes AUBUC réussies. Les images corps entiers (groupe B) sont coupées afin de posséder la même zone imagée que les autres images (groupes A et C).

Jeu	n	Algorithme	Max d	Moyenne d	Temps
A	20	Notre approche (BFGR)	71.3	10.5 ± 8.3	1.3
A	20	ANTs-IM-S4	152.7	14.0 ± 17.0	9.7
A	20	ANTs-IM-S2	70.6	11.3 ± 7.9	13.8
A	20	ANTs-IM	150.2	15.5 ± 16.7	62.5
A	20	ANTs	153.0	15.4 ± 17.0	327.0
B	20	Notre approche (BFGR)	66.8	10.1 ± 8.7	2.3
B	20	ANTs-IM-S2	231	17.9 ± 27.3	36.0
AC	83	Notre approche (BFGR)	69.7	10.4 ± 8.0	18.5
AC	83	ANTs-IM-S2	198.0	44.7 ± 31.6	82.8
ABC	103	Notre approche (BFGR)	69.7	11.1 ± 8.6	46.0

TABLEAU 5.1 – Comparaison entre notre approche et ANTs avec les différents jeux de données de VISCERAL pour différents scénarios. n est le nombre d’images mises en correspondance. Les distances sont en millimètres, les temps en heures. Pour l’algorithme ANTs, IM signifie Information Mutuelle, S2 et S4 font référence au sous-échantillonnage respectivement $sf=2$ et $sf=4$.

Jeux	n	$ \mathcal{M} /n$	I (h)	II (h)	III (h)	total (h)
A	20	41k	0.1	0.5	0.7	1.3
B	20	65k	0.3	0.5	1.5	2.3
AC	83	178k	0.1	8.0	10.4	18.5
ABC	103	272k	0.3	12.0	33.7	46.0

TABLEAU 5.2 – Nombre d’appariements par image et temps de calcul de notre algorithme. n est le nombre d’images mis en correspondance. $|\mathcal{M}|/n$ est le nombre d’appariements par image. Les colonnes I, II et III indiquent, respectivement, le temps passé (en heures) pour, respectivement, l’extraction des points d’intérêt, l’appariement et l’optimisation.

5.3.8 Comportement de FGDR pour différents jeux de données

En portant attention aux différentes expérimentations effectuées avec notre approche FGDR et en détaillant les caractéristiques de ces expériences, nous pouvons observer plusieurs phénomènes intéressants.

Temps d’exécution Nous pouvons observer au sein de la table 5.2 les temps d’exécution des différentes parties, respectivement : I l’extraction des points d’intérêt, II l’appariement des points d’intérêt, III l’optimisation. Ces phases peuvent se classer en deux parties. La première concerne l’étape d’extraction des points d’intérêt, qui est linéaire en fonction du nombre n d’images et les phases d’appariement et d’optimisation qui sont quadratiques. L’aspect quadratique de l’appariement est immédiate, tandis que celle de l’optimisation est issue de l’observation.

Nombre de mises en correspondance par image Un autre aspect intéressant de cette synthèse est le nombre moyen de points d’intérêt par image ($|\mathcal{M}|/n$). Celui-ci est bien quadratique en fonction du nombre d’images. Pour rappel, il y a environ 8000 points d’intérêt par image. Chaque point d’intérêt est apparié en moyenne à 34 images pour la dernière expérience (soit 33% des appariements possibles). Ce nombre nous montre à la fois que nous avons un réseau de liens denses entre toutes les images et que nous sommes dans un cas où l’appariement est très fréquent, conduisant à un nombre élevé de faux positifs.

Empreinte mémoire L’empreinte mémoire de notre approche n’excède jamais 10Go, tandis que nous avons pu observer des pics supérieurs à 65Go avec ANTs en pleine résolution, et 34Go avec un sous-échantillonnage de facteur 2 ($sf = 2$).

Landmark	ANTs (20)	FGDR (20)	FGDR (83)
1 : Clavicle L	9.7 ±4.6	7.1 ±2.4	7.6 ±2.3
2 : Clavicle R	9.3 ±4.4	8.3 ±3.3	8.7 ±3.2
3 : Tubercul. L	11.2 ±5.4	12.2 ±6.4	10.5 ±5.3
4 : Tubercul. R	10.9 ±5.5	13.1 ±4.4	12.1 ±3.6
5 : C6	3.5 ±2.6	7.0 ±1.7	7.2 ±2.7
6 : C7	5.0 ±3.3	7.3 ±3.8	7.7 ±4.2
7 : Th1	4.5 ±3.0	6.4 ±4.4	6.4 ±4.6
8 : Th2	4.9 ±3.4	6.1 ±4.3	6.2 ±4.2
9 : Th3	6.1 ±4.2	6.9 ±4.4	6.8 ±4.1
10 : Th4	7.6 ±4.3	7.4 ±4.8	7.1 ±4.3
11 : Th5	8.6 ±3.6	8.2 ±4.9	8.0 ±4.3
12 : Th6	9.6 ±3.4	8.9 ±6.0	8.7 ±5.1
13 : Th7	10.2 ±3.4	9.1 ±7.0	9.2 ±5.9
14 : Th8	10.3 ±3.9	9.3 ±7.8	9.5 ±7.2
15 : Th9	10.8 ±4.9	10.7 ±9.0	11.0 ±8.6
16 : Th10	11.3 ±5.4	11.9 ±10.0	12.0 ±9.8
17 : Th11	12.2 ±5.9	12.3 ±10.1	12.4 ±10.0
18 : Th12	13.4 ±6.2	12.5 ±9.5	12.5 ±9.3
19 : L1	13.3 ±6.6	12.4 ±8.8	12.0 ±8.2
20 : L2	12.8 ±6.6	12.1 ±8.2	11.2 ±7.7
21 : L3	13.1 ±7.1	12.5 ±7.2	11.6 ±6.9
22 : L4	12.8 ±7.2	11.7 ±5.8	11.1 ±6.7
23 : L5	11.4 ±6.7	9.3 ±6.3	9.3 ±6.2
24 : Sternoclav. L	7.9 ±5.5	6.7 ±4.1	7.3 ±4.4
25 : Sternoclav. R	6.6 ±3.9	5.9 ±2.7	6.3 ±2.9
26 : Aortic arch	8.4 ±3.2	8.8 ±3.5	9.1 ±3.9
27 : Trachea bif.	4.5 ±3.8	6.5 ±2.3	6.7 ±3.1
28 : Bronchus L	7.5 ±5.5	7.7 ±3.7	7.6 ±4.3
29 : Bronchus R	5.9 ±4.0	5.5 ±2.8	5.7 ±3.0
30 : Coronaria	9.1 ±6.4	7.9 ±3.6	8.1 ±3.2
31 : Aortic valve	11.8 ±7.2	11.8 ±4.5	11.9 ±6.4
32 : Xyphoideus	16.6 ±7.8	18.1 ±12.0	17.0 ±10.4
33 : Renal pelvis L	17.9 ±9.2	20.8 ±11.0	20.1 ±9.3
34 : Renal pelvis R	20.6 ±15.1	23.9 ±15.4	23.0 ±15.7
35 : Crista iliaca L	9.9 ±6.0	10.0 ±5.6	10.4 ±5.9
36 : Crista iliaca R	10.4 ±8.8	11.1 ±7.9	10.9 ±7.5
37 : Aorta bif.	15.6 ±10.5	15.3 ±12.4	14.9 ±11.0
38 : VCI bif.	12.6 ±6.5	12.7 ±9.5	12.8 ±9.0
39 : Troch. maj. L	19.3 ±10.8	16.4 ±9.3	17.0 ±8.9
40 : Troch. maj. R	18.5 ±9.1	15.2 ±8.3	16.1 ±8.1
41 : Ischiadicum L	11.8 ±6.4	6.6 ±3.3	6.4 ±3.5
42 : Ischiadicum R	10.9 ±6.3	7.8 ±5.0	7.4 ±5.2
43 : Symphysis	15.8 ±9.9	11.2 ±8.1	11.4 ±8.0
44 : Troch. min. L	15.2 ±7.7	7.4 ±3.1	7.2 ±3.3
45 : Troch. min. R	16.6 ±5.8	7.4 ±3.9	7.5 ±4.1
Mean	11.3 ±7.9	10.5 ±8.3	10.4 ±8.0
Maximum	70.6	71.3	69.7

TABLEAU 5.3 – Comparaison entre notre approche et ANTs : distances moyennes des points repère anatomiques placés par des médecins (en mm).

5.4 Conclusion

Dans ce chapitre, nous avons pu montrer que notre approche est capable de mettre en correspondance des cas réels de bases de données d'images 3D scanner CT, en un temps raisonnable et avec une faible empreinte mémoire. Les résultats expérimentaux montrent la robustesse de notre approche FGDR face à la variabilité des images et une comparaison avec ANTs montre que nous arrivons à avoir une meilleure précision sur des ensembles de points de repère anatomique placés par des médecins, avec un temps de calcul inférieur.

Notre méthode est générique et peut s'appliquer à tout ensemble de points d'intérêt et n'est pas limitée à l'interpolation par des B-splines. Un travail de recherche de points d'intérêt spécifiquement conçus pour les images médicales pourrait être intéressant. L'utilisation de caractéristiques issues de la théorie de l'apprentissage serait être une piste prometteuse.

Dans le cadre de cette thèse, nous portons plutôt notre attention sur la validation de notre méthode face aux deux défis principaux que nous nous sommes fixés, à savoir la capacité à mettre en correspondance des images possédant une grande variabilité et la capacité à traiter de grandes bases de données. Le prochain chapitre s'attachera à approfondir ces deux points pour notre approche FGDR.

Chapitre 6

Perspectives

Sommaire

6.1	Introduction	132
6.2	Vers la mise en correspondance de 10 000 images	132
6.2.1	Introduction	132
6.2.2	Gradient stochastique	132
6.2.3	Observation	132
6.2.4	Méthode	132
6.2.5	Résultats préliminaires	133
6.2.6	Distribution et parallélisme	133
6.3	Application à la segmentation par atlas	134
6.3.1	Principe	134
6.3.2	Méthode	134
6.3.3	Résultats préliminaires	135
6.3.4	Perspectives	135
6.4	Application à l'anthropologie	135
6.4.1	Méthode	136
6.4.2	Résultats préliminaires	137
6.4.3	Analyse de la qualité de la mise en correspondance	137
6.4.4	Conclusion	138

6.1 Introduction

Nous avons montré au chapitre précédent que notre approche de mise en correspondance de groupes d'images permet de recalibrer une centaine d'images de manière déformable avec une précision permettant d'envisager de multiples applications. Nous abordons ici les perspectives qu'ouvrent les résultats de ces travaux.

Si l'on poursuit notre objectif initial de traiter les données provenant du PACS d'un hôpital, il est nécessaire de pouvoir passer à l'échelle : de 100 à plus de 10 000 images. Pour cela, il est nécessaire de développer des méthodes permettant de dépasser les limites actuelles de l'algorithme.

Dans un second temps, nous présentons ici deux applications de notre méthode, utilisées d'une part dans le cadre de la segmentation multi-atlas [IGLESIAS et SABUNCU \[2015\]](#); [KÉCHICHIAN et collab. \[2017\]](#) et d'autre part par des médecins légistes dans le cadre de l'anthropologie.

6.2 Vers la mise en correspondance de 10 000 images

6.2.1 Introduction

Notre approche nécessite actuellement quelques heures pour mettre en correspondance une centaine d'images 3D. Sachant que sa complexité dominante est quadratique, il n'est pas possible de mettre en correspondance 10 000 images directement. Afin de pouvoir traiter plusieurs dizaines de milliers d'images, il est donc nécessaire de trouver une solution pour paralléliser notre méthode et de rendre sa complexité linéaire. Les techniques d'apprentissage artificiel sont soumises aux mêmes contraintes et ont trouvé une solution en développant l'utilisation du gradient stochastique. Nous allons développer les perspectives d'utilisation de cette technique dans notre méthode de mise en correspondance.

6.2.2 Gradient stochastique

Le gradient stochastique est une méthode largement utilisée dans les approches pour l'apprentissage [\[BOTTOU, 2010\]](#). Cette méthode consiste à effectuer des pas d'optimisation en n'utilisant qu'une partie des données disponibles, choisie aléatoirement. Les méthodes d'apprentissage actuelles n'utilisent généralement qu'un exemple à la fois, en ajoutant un terme d'inertie (momentum, [SUTSKEVER et collab. \[2013\]](#)), ou des petits sous-ensembles de données ("batch learning"). Cela permet aussi de paralléliser le calcul [\[TSITSIKLIS et collab., 1986\]](#).

6.2.3 Observation

Si nous observons le comportement de notre approche FGDR avec un grand nombre d'images (>100), nous constatons que chaque point d'intérêt est apparié avec de nombreux points appartenant aux autres images. L'optimisation consiste donc à modifier les valeurs des champs de déformation afin de déplacer chaque point d'intérêt vers la moyenne des points (centre de gravité) avec lesquels il est apparié. Pour une itération donnée, avec un grand nombre d'images, cette moyenne tend à se stabiliser à une certaine valeur même si nous ajoutons ou supprimons un couple. Il ne semble pas nécessaire d'utiliser toutes les images pour construire ces moyennes pour une itération donnée.

6.2.4 Méthode

La méthode du gradient stochastique appliquée à notre problème est un peu particulière. Dans notre cas, nous ne pouvons pas optimiser en utilisant une seule image à la fois car notre fonction objectif lie intrinsèquement des paires d'images. Par contre nous pouvons utiliser des approches par sous-ensembles.

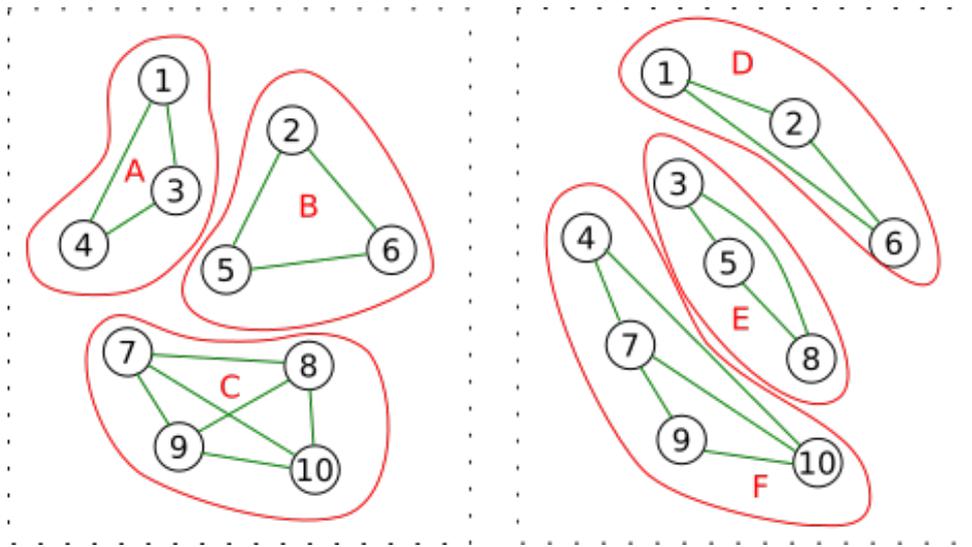


FIGURE 6.1 – Deux exemples d’itération (carrés pointillés) de l’optimisation pour 10 images par sous-ensembles (en rouge). Dans chaque cas, nous pouvons voir une 3-partition aléatoire de l’ensemble des images. Seuls les points appariés appartenant au même sous-ensemble d’images sont utilisés (en vert).

Comme nous pouvons en observer un exemple figure 6.1, l’approche consiste, dans notre cas, à partitionner l’ensemble des images en des sous-ensembles disjoints de N_b images. Pour une itération donnée, chaque sous-ensemble est considéré individuellement, ce qui signifie que seuls les couples de points appariés entre deux images d’un même sous-ensemble sont utilisés. À chaque itération un nouveau partitionnement est tiré aléatoirement.

Soit n le nombre d’images. Le nombre de paires d’images de notre approche déterministe en fonction de ce nombre d’images est :

$$n_{\text{paires}} = \frac{n(n-1)}{2} \quad (6.1)$$

Ce qui correspond à une complexité quadratique. Dans le cadre de l’approche par gradient stochastique, le nombre de paires d’images utilisées au sein d’une itération est :

$$n_{\text{paires}} = \frac{n}{N_b} \times \frac{N_b(N_b-1)}{2} = \frac{n(N_b-1)}{2} \quad (6.2)$$

où N_b est le nombre, constant, de sous-ensembles d’images.

En utilisant une approche stochastique, une itération de l’algorithme devient de complexité linéaire. Cependant, le nombre d’itération pour arriver à convergence est supérieur à la méthode déterministe, ce qui conduit à une complexité globale entre le linéaire et le quadratique.

6.2.5 Résultats préliminaires

Les résultats préliminaires montrent que le nombre d’images nécessaire au sein d’un sous-ensemble est très faible. Nous avons effectué des expérimentations avec $N_b = 5$ et les 83 images issues de la base de données VISCERAL. Nous observons des précisions similaires à l’approche exhaustive, tout en nécessitant 10 fois moins de temps.

6.2.6 Distribution et parallélisme

En plus de sa rapidité, cette implémentation permet de distribuer le calcul sur des grappes d’ordinateurs. À chaque itération, l’optimisation de chaque sous-ensemble peut être distribuée sur une machine spécifique.

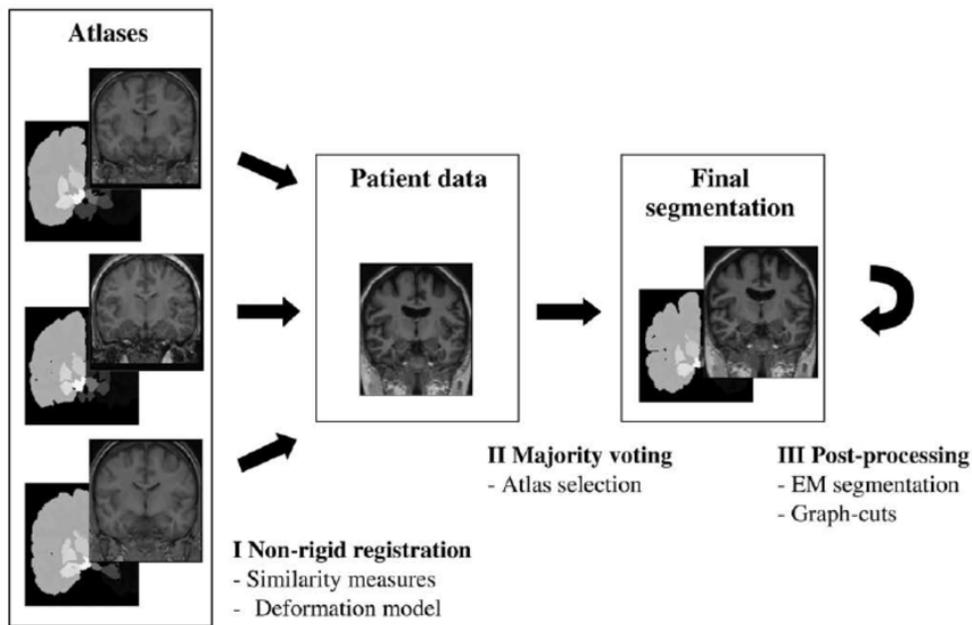


FIGURE 6.2 – Étapes de la segmentation multi-atlas : I. Un recalage déformable est effectué entre les images de la base et l’image cible. Ce recalage permet de transférer les segmentations dans l’espace cible. II. Les atlas sont agrégés afin d’avoir un atlas final. III. Des post-traitements basés sur l’intensité des images améliorent la segmentation finale. Crédit : LÖTJÖNEN et collab. [2010]

6.3 Application à la segmentation par atlas

6.3.1 Principe

La segmentation multi-atlas, introduit par ROHLFING et collab. [2004] et dont IGLESIAS et SABUNCU [2015] ont fait une synthèse, consiste à utiliser des segmentations effectuées manuellement sur un ensemble d’images (les atlas), pour ensuite les reporter sur l’image que l’on souhaite segmenter. Différentes méthodes sont utilisées afin d’agréger les différentes segmentations et des approches de post-traitement permettent d’affiner la segmentation en utilisant les données d’intensité de l’image à segmenter (bord, gradient, etc.). La figure 6.2 fournit un schéma synthétique de ces approches.

6.3.2 Méthode

Usuellement, les approches multi-atlas utilisent des mises en correspondance par paires entre les atlas et l’image à segmenter. Étant donné la présence d’un ensemble d’images (les atlas) la mise en correspondance de groupe et plus particulièrement notre méthode FGDR semble bien adaptée à ce cadre. Quelques approches récentes se proposent de lier le recalage et la segmentation comme par exemple MARTIN et collab. [2008].

Le premier avantage d’un recalage de groupe pour la segmentation multi-atlas est l’utilisation conjointe de toutes les images disponibles, non seulement pour la segmentation, mais aussi pour la mise en correspondance, qui permet d’augmenter la robustesse. Notre approche permettrait une amélioration de la rapidité des approches utilisant de multiples recalages par paires.

La segmentation multi-atlas nous permet aussi d’avoir un autre cadre d’évaluation des performances de notre méthode FGDR, en utilisant des segmentations de référence au lieu des repères anatomiques.

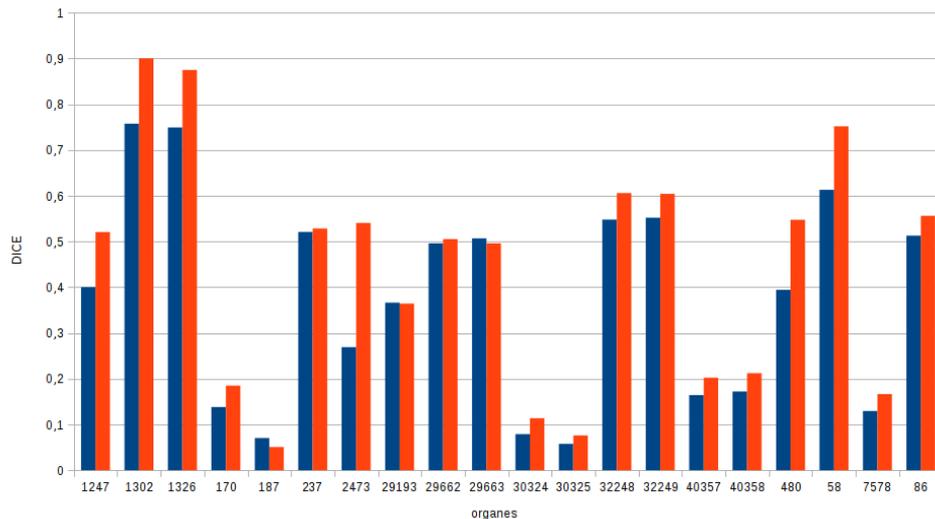


FIGURE 6.3 – Résultat de l’étape de recalage déformable dans le cadre de la segmentation multi-atlas. Nous observons le critère DICE moyen par organe transformé vers l’espace cible, dans un contexte de validation croisée. En bleu, les résultats de la méthode localement rigide et en rouge les résultats de notre méthode.

6.3.3 Résultats préliminaires

Nous évaluons ici notre méthode FGDR en utilisant les organes segmentés de la base de données VISCERAL. Nous effectuons une validation croisée en mettant en correspondance l’ensemble des images. Pour chaque image, nous transformons les segmentations des autres images (les atlas) vers cette image cible. Ensuite, nous calculons différentes métriques de similarités (le coefficient de DICE $\frac{2|X \cap Y|}{|X| + |Y|}$, distance moyenne) entre chacun des atlas et l’image cible dont nous connaissons la vérité terrain.

La figure 6.3 présente les critères DICE moyens pour chaque organe, en faisant la moyenne pour chaque atlas et pour chaque scénario. La comparaison est effectuée entre une méthode de recalage localement rigide par paires (barres bleues) et notre méthode FGDR (barres rouges). Nous pouvons observer au minimum une équivalence et une amélioration pour la majorité des organes. Des observations plus spécifiques montrent que la méthode bénéficie principalement aux petits organes.

6.3.4 Perspectives

D’autres résultats, prenant en compte la chaîne de traitement complète de segmentation multi-atlas est nécessaire pour valider ces résultats. Cependant, ces premiers résultats prouvent la capacité de notre méthode dans son utilisation pour la segmentation. En ce qui concerne la segmentation multi-atlas, la structure de notre mise en correspondance de groupe ouvre de nouvelles possibilités. Son approche par demi-transformations déforme les atlas en deux étapes. Cela permet de transformer les atlas dans l’espace commun, de calculer le vote par majorité (ou les cartes probabilistes dans le cadre d’approche bayésienne, exemple KÉCHICHIAN et collab. [2017]) dans cet espace. Ensuite, lors de la segmentation d’une nouvelle image, il suffit de déformer le résultat de cette étape dans l’image cible, ce qui est très rapide. En utilisant cette méthode, FGDR augmenterait la vitesse non seulement de la mise en correspondance, mais aussi du reste du processus de segmentation multi-atlas.

6.4 Application à l’anthropologie

L’un des défis de l’anthropologie est la détermination de la silhouette d’identité à partir d’un corps inconnu et dont l’état ne permet pas l’identification visuelle. La silhouette d’identité consiste en différentes

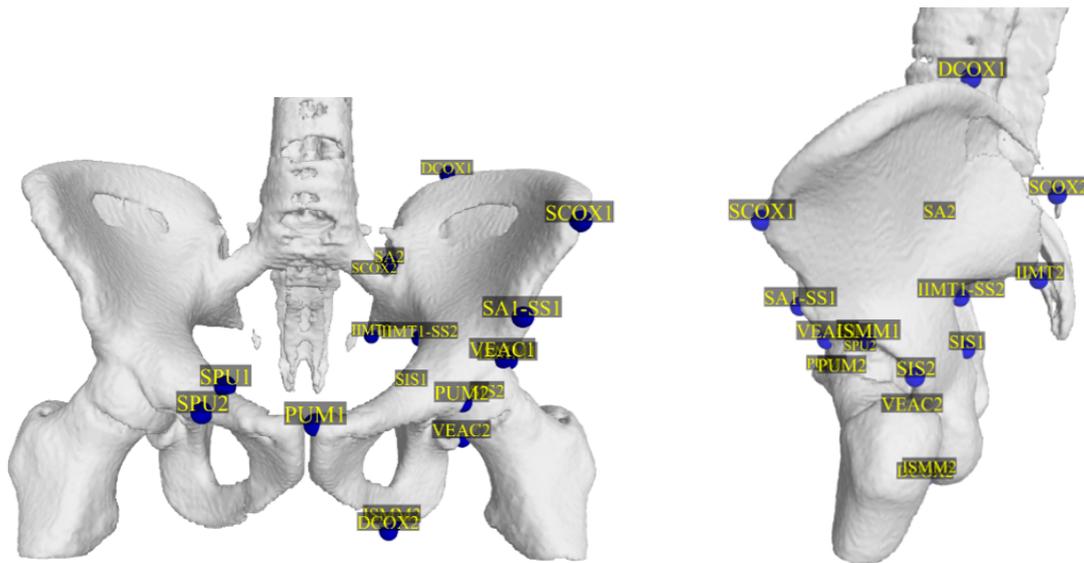


FIGURE 6.4 – Image moyenne issue de la mise en correspondance de groupe FGDR. Les repères anatomiques (points bleus), issus de l’approche DSP, sont placés manuellement sur l’image moyenne, et sont ensuite reportés automatiquement sur chacune des images.

informations : le genre, l’âge et l’origine biogéographique. Les méthodes actuelles d’anthropologie utilisent des mesures sur certains os caractéristiques (bassin, crâne, etc.) afin de déterminer la silhouette d’identité. Ces méthodes sont complexes à mettre en œuvre et n’utilisent que quelques points de mesures (généralement issues d’a priori biologiques).

L’utilisation de notre méthode de mise en correspondance déformable de masse pourrait permettre de dépasser ses limites grâce à l’automatisation, en utilisant de larges échantillons et une analyse exhaustive de l’anatomie afin de trouver plus de critères discriminants que ceux utilisés actuellement.

Afin de vérifier la faisabilité de notre idée, nous avons créé une méthode de détection du genre de manière automatique en utilisant uniquement une image scanner 3D et sans intervention supplémentaire de l’utilisateur. Pour cela, nous utilisons une méthode de la littérature anthropologique : DSP (“Diagnose Sexuelle Probabiliste” [MURAIL et collab. \[2005\]](#)), en automatisant la partie positionnement de repères et des mesures avec de l’imagerie. Ces travaux ont fait l’objet d’une communication internationale [FRANCHI et collab. \[2017\]](#).

6.4.1 Méthode

L’objectif est de déterminer le sexe d’individus dont nous possédons des images scanner 3D du bassin, à l’aide de notre méthode FGDR. Notre expérience de faisabilité de la détermination du genre consiste à recalculer un groupe d’images contenant des images d’hommes et de femmes. Un spécialiste place un ensemble de repères anatomiques discriminants sur l’image moyenne (voir figure 6.4). Ces repères sont issus de la méthode DSP qui se propose de déterminer le genre d’un individu à partir d’un maximum de dix mesures de distances entre repères anatomiques.

Grâce à la connaissance des demi-transformations liant l’image moyenne et les images des individus, nous pouvons reporter les positions des repères pour chaque individu. Nous calculons grâce à ces repères

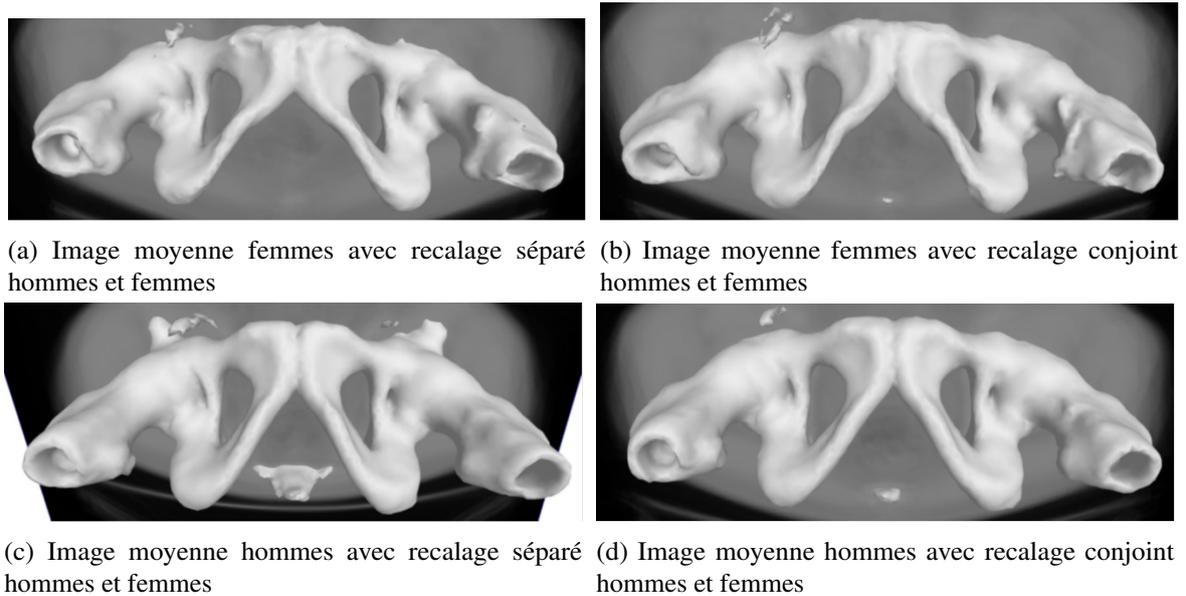


FIGURE 6.5 – Images moyennes dont le bassin est extrait. En comparant les images en haut à gauche et en bas à gauche, on peut observer la différence de symphyse pubienne entre les hommes et les femmes. En comparant les paires d’images gauches et droites, nous pouvons observer la variation due à la différence de genre (gauche) et la variabilité non capturée par notre algorithme (droite).

les distances nécessaires à la méthode DSP pour déterminer le genre de chaque individu. Nous pouvons finalement comparer les résultats fournis par la méthode DSP et la vérité terrain.

6.4.2 Résultats préliminaires

En testant notre approche sur 83 images, nous avons déterminé correctement 62% des échantillons, 37% des genres n’ont pas été déterminés (pas de différence statistiquement significative) avec seulement 1% d’erreur.

6.4.3 Analyse de la qualité de la mise en correspondance

L’objectif de cette approche, au-delà de fournir un outil de détermination du sexe, permet de vérifier la capacité de notre mise en correspondance à fournir des recalages suffisamment précis. La figure 6.5 illustre cette analyse. L’image en bas à gauche est l’image moyenne des hommes recalés ensemble, tandis que celle en haut à gauche est la moyenne des femmes. Ces images mettent en évidence un critère discriminant du sexe connu des anthropologues : la symphyse pubienne (les femmes ont un bassin statistiquement plus large que les hommes). Cette différence est visible à l’aide du triangle que nous observons au milieu de chaque image, le triangle du haut est plus ouvert que le triangle du bas.

Les images de droite correspondent à un recalage général (hommes et femmes ensemble) en calculant ensuite les images moyennes en prenant uniquement les femmes (en haut à droite) et les hommes (en bas à droite). Si le recalage était parfait, ses images seraient identiques. La comparaison des paires gauches et droites permet d’analyser la variance entre les hommes et les femmes, ainsi que la variance non capturée par notre méthode.

En conclusion, sans être parfaite, notre méthode permet de mettre en avant les différences hommes/femmes de manière significative, mais conserve cependant une erreur.

6.4.4 Conclusion

Ces résultats préliminaires prouvent que les champs d'applications potentielles sont larges. Des expériences avec des bases d'images plus grandes permettraient d'explorer l'utilisation de repères anatomiques découverts automatiquement. Ces résultats préliminaires ont fait l'objet d'un exposé oral lors du 50ème congrès international francophone de médecine légale [FRANCHI et collab. \[2017\]](#).

Chapitre 7

Conclusion

L'exploitation de larges bases de données d'images médicales présente un réel intérêt pour de nombreuses applications. La capacité à traiter des groupes d'images hétérogènes possédant une grande variabilité, est nécessaire pour développer des applications "big-data" en imagerie médicale.

La problématique de cette thèse consistait à développer une méthode rapide de mise en correspondance de grands groupes capable d'apporter une cohérence spatiale à ces bases de données hétérogènes. Nous avons développé des méthodes de mise en correspondance de groupes qui utilisent des points d'intérêt et qui sont applicables sur des bases de plusieurs centaines d'images. Ces méthodes utilisent des formulations de problèmes globaux, nécessitant uniquement des appariements de points d'intérêt. Cela permet de maîtriser à la fois, le volume de données, la vitesse et la robustesse des approches. Une particularité de ces problèmes globaux est l'attention qui doit être portée à la gestion des mauvais appariements de points d'intérêt.

Notre travail s'articule autour de plusieurs contributions nécessaires à l'élaboration d'un protocole complet de mise en correspondance de grands groupes. Dans notre première contribution, nous formulons un problème global de recalage de groupes, s'appuyant sur l'opérateur laplacien associé au graphe complet des recalages de paires (chapitre 3). Cette première contribution nous permet de rendre cohérente de larges bases d'images 3D en leur appliquant des recalages rigides, préliminaire indispensable à des approches plus précises. Notre deuxième contribution calcule le recalage déformable de ces groupes d'images (chapitre 4). Nous utilisons des demi-transformations composées de pyramides de splines entre chaque image et un espace commun. Ces demi-transformations sont estimées grâce aux appariements entre points d'intérêt provenant des images, tout en inhibant les mauvais appariements.

Nous avons évalué l'efficacité de notre approche à l'aide de données de référence. Des points de repères anatomiques placés par des experts nous ont permis de qualifier la précision de notre méthode tandis que des bases de données de grandes tailles nous ont permis de vérifier sa robustesse et sa capacité à traiter un grand nombre d'images. Deux applications, la détermination de la silhouette d'identité en anthropologie et la segmentation multi-atlas, illustrent la pertinence de notre proposition.

La mise en correspondance de grandes bases d'images permet d'envisager des méthodes d'apprentissage à des fins de diagnostics et de prévention. Ces méthodes nécessitent cependant une évaluation éthique, quant à l'utilisation de données personnelles médicales de plusieurs milliers de personnes. L'avantage de notre méthode qui utilise des points d'intérêt est de ne pas exploiter directement les images, mais seulement des points d'intérêt, ce qui garantit l'anonymat et la préservation des données personnelles.

Cette thèse ouvre la voie à la création et à l'utilisation de larges bases d'images médicales contenant plus d'un millier d'images.

Liste des figures

1.1	Jeu de données contenant 400 images.	1
1.2	Exemples de points d'intérêt	3
2.1	Une mosaïque d'images	7
2.2	Fusion de modalités sur des images satellites	8
2.3	Assemblage d'images pour un rendu HDR	9
2.4	Reconnaissance et suivi d'objets basés sur un modèle 3D	9
2.5	Recalages intra- et inter-patients	11
2.6	Recalages multimodale et monomodale	12
2.7	Splines (historique)	16
2.8	Fonctions de base des B-splines cubiques uniformes	17
2.9	Fonctions de base des B-splines cubiques uniformes	18
2.10	Exemple d'approximation par fonctions de base radiales gaussienne.	21
2.11	Exemple de descente de Gradient.	24
2.12	Le recalage exhaustif et le recalage en étoile	31
2.13	Recalage en étoile	32
2.14	Recalage par arbre couvrant minimal	33
2.15	Détection et mise en correspondance de points d'intérêt	34
2.16	Exemple de détection de coins	35
2.17	Pyramide de gaussienne.	36
2.18	Approximation par bloc des dérivées secondes d'un noyau gaussien.	37
2.19	Description SIFT par histogrammes d'orientations	38
2.20	Ondelettes de Haar	38
2.21	Réponses des ondelettes de Haar au voisinage d'un point	39
2.22	Réponse du descripteur à différentes textures.	39
2.23	Différents M-estimateurs	42
2.24	Exemple reconstruction 3D par ajustement de faisceaux.	43
2.25	Différence entre une distribution Gaussienne et une distribution à queue lourde.	44
3.1	Schéma du processus complet de l'extraction de points d'intérêt par SURF 3D.	49
3.2	Schéma de la localité utilisée pour la description d'un point d'intérêt SURF3D.	53
3.3	Schéma du processus complet de FPR	54
3.4	Exemple d'extraction de points d'intérêt SURF sur une image médicale.	55
3.5	Extraction des points d'intérêt de 2 images 3D.	56
3.6	Exemple de mise en correspondance, avec paires de points d'intérêt.	59
3.7	Expérimentation FPR avec estimation uniquement de la translation	60
3.8	Autre exemple de mises en correspondance FPR	61
3.9	Schéma d'extraction et d'appariement des points d'intérêt	62
3.10	Recalage "sans référence" par graphe complet	63
3.11	Exemple de recalage rigide de groupes avec 3 images	64

3.12	Positions reconstruites en utilisant notre approche sur le problème décrit figure 3.11.	65
3.13	Exemple avec 27 images 3D obtenues par tomographie à rayon X.	66
3.14	Exemple problématique de recalage avec 3 images hétérogènes.	67
3.15	Exemple de la figure 3.13, en utilisant l'élagage du graphe.	69
3.16	Schéma de l'approche FGR	71
3.17	Temps de calcul en fonction du nombre d'images.	71
3.18	La même vue que la figure 1.1 après recalage.	72
3.19	Masquage des yeux appliqués à des images CT	73
4.1	Exemple de parcimonie induite par l'ajustement de faisceaux.	79
4.2	Exemple avec 4 images.	81
4.3	Différentes structures des champs de déformation.	83
4.4	Exemple d'une pyramide de points de contrôle.	85
4.5	Schéma illustrant l'erreur	85
4.6	Distribution synthétique des vrais et faux positifs	88
4.7	Distribution des distances. Données synthétiques	92
4.8	Exemple avec des données réelles	92
4.9	Comparaison des fonctions de pondération avec les mêmes données réelles	93
4.10	Courbe ROC : M-Estimateur et Pondération EM	94
4.11	Exemple avec 4 images.	96
4.12	Exemple de dérive.	99
4.13	Exemple de minimisation	100
4.14	Schéma de génération des données synthétiques	100
4.15	Exemple de données de synthèse.	102
4.16	Exemple de régularisation sur deux points de contrôle.	103
4.17	Résultat de régularisation de deux points de contrôle.	103
4.18	Positions finales avec une régularisation du déplacement des points de contrôle	104
4.19	Exemple de régularisation sur un couple apparié.	105
4.20	Résultat de l'optimisation sur les couples appariés avec trois points d'intérêt.	106
4.21	Positions finales avec une régularisation sur les couples appariés	106
4.22	Histogramme du nombre moyen d'appariements pour un point d'intérêt	107
4.23	Exemple de contrainte inter-images sur les points de contrôle.	107
4.24	Positions finales avec une contrainte inter-images sur les points de contrôle	108
5.1	Illustration des jeux de données A, B et C	113
5.2	Schéma bloc de la procédure de recalage par groupe de ANTs	115
5.3	Résultats des différentes approches pour les 20 images du jeu de données A.	118
5.4	Comparaison entre notre approche FGDR et ANTs	119
5.5	Convergence du critère de minimisation et distance moyenne	119
5.6	Points de repère du groupe A affichés dans le même repère	120
5.7	Squelettes extraits lors de la mise en correspondance des 20 images du groupe B.	122
5.8	Résultats des différentes approches pour les 20 images corps entiers du groupe B	123
5.9	Images moyennes des différentes approches pour les 20 images du groupe A+C	125
5.10	Image moyenne issue de notre approche BFGR avec les 103 images des groupes A+B+C	126
6.1	Deux exemples d'itérations de l'optimisation par sous-ensembles	133
6.2	Étapes de la segmentation multi-atlas	134
6.3	Résultat de l'étape de recalage déformable dans le cadre de la segmentation multi-atlas.	135
6.4	Image moyenne et repères anatomiques	136
6.5	Images moyennes dont le bassin est extrait	137

Liste des tableaux

2.1	Classification des transformations globales	14
3.1	Les différentes transformations globales utilisées dans l'approche FPR	57
4.1	Nombre d'échantillons disponibles en fonction de la localité.	91
4.2	Matrices de confusion	93
4.3	Valeur des paramètres utilisés.	102
5.1	Comparaison entre notre approche et ANTs	126
5.2	Nombre d'appariements par image et temps de calcul de notre algorithme	127
5.3	Comparaison entre notre approche et ANTs : distances moyennes des points repère anatomiques placés par des médecins (en mm).	128

Bibliographie

- AGIER, R., S. VALETTE, L. FANTON, P. CROISILLE et R. PROST. 2016a, «Hubless 3d medical image bundle registration», dans VISAPP 2016 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 3, p. 265–272. [49](#)
- AGIER, R., S. VALETTE, L. FANTON, P. CROISILLE et R. PROST. 2016b, «Recalage massif d’images 3d», dans GDR Madics, workshop ATLAS. [49](#)
- AGIER, R., S. VALETTE, R. KÉCHICHAN et R. PROST. 2017, «Hubless 3d medical image bundle registration», soumis à IEEE Transactions on Image Processing. [77](#)
- ALAHY, A., R. ORTIZ et P. VANDERGHEYNST. 2012, «Freak : Fast retina keypoint», dans Computer vision and pattern recognition (CVPR), 2012 IEEE conference on, Ieee, p. 510–517. [34](#)
- ALLAIRE, S., J. J. KIM, S. L. BREEN, D. A. JAFFRAY et V. PEKAR. 2008, «Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis», dans Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on, IEEE, p. 1–8. [45](#), [49](#)
- ANGUELOV, D., C. DULONG, D. FILIP, C. FRUEH, S. LAFON, R. LYON, A. OGALE, L. VINCENT et J. WEAVER. 2010, «Google street view : Capturing the world at street level», Computer, vol. 43. [33](#)
- ARAD, N., N. DYN, D. REISFELD et Y. YESHURUN. 1994, «Image warping by radial basis functions : Application to facial expressions», CVGIP : Graphical models and image processing, vol. 56, n° 2, p. 161–172. [20](#)
- ASHBURNER, J. et K. J. FRISTON. 2007, «Rigid body registration», Statistical parametric mapping : The analysis of functional brain images, p. 49–62. [14](#)
- ASHBURNER, J. et K. J. FRISTON. 2011, «Diffeomorphic registration using geodesic shooting and gauss–newton optimisation», NeuroImage, vol. 55, n° 3, p. 954–967. [26](#)
- AVANTS, B., C. EPSTEIN, M. GROSSMAN et J. GEE. 2008, «Symmetric diffeomorphic image registration with cross-correlation : Evaluating automated labeling of elderly and neurodegenerative brain», Medical Image Analysis, vol. 12, n° 1, p. 26 – 41, ISSN 1361-8415. [110](#), [114](#)
- BAJCSY, R. et S. KOVAČIČ. 1989, «Multiresolution elastic matching», Computer vision, graphics, and image processing, vol. 46, n° 1, p. 1–21. [15](#)
- BARTOLI, A., D. PIZARRO et M. LOOG. 2013, «Stratified generalized procrustes analysis», International journal of computer vision, vol. 101, n° 2, p. 227–253. [49](#)
- BAY, H., A. ESS, T. TUYTELAARS et L. VAN GOOL. 2008, «Speeded-up robust features (surf)», Computer vision and image understanding, vol. 110, n° 3, p. 346–359. [52](#)

- BAY, H., T. TUYTELAARS et L. VAN GOOL. 2006, «Surf : Speeded up robust features», dans Computer vision–ECCV 2006, Springer, p. 404–417. [34](#), [35](#), [37](#), [39](#), [49](#), [51](#), [52](#), [53](#), [55](#)
- BENTLEY, J. L. 1975, «Multidimensional binary search trees used for associative searching», Communications of the ACM, vol. 18, n° 9, p. 509–517. [80](#)
- BERG, A. C., T. L. BERG et J. MALIK. 2005, «Shape matching and object recognition using low distortion correspondences», dans Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, IEEE, p. 26–33. [22](#)
- BESL, P. J. et N. D. MCKAY. 1992, «Method for registration of 3-d shapes», dans Robotics-DL tentative, International Society for Optics and Photonics, p. 586–606. [23](#)
- BOOKSTEIN, F. L. 1989, «Principal warps : Thin-plate splines and the decomposition of deformations», IEEE Transactions on pattern analysis and machine intelligence, vol. 11, n° 6, p. 567–585. [20](#)
- BOOKSTEIN, F. L. 1991, «Thin-plate splines and the atlas problem for biomedical images», dans Biennial International Conference on Information Processing in Medical Imaging, Springer, p. 326–342. [20](#)
- BOTTOU, L. 2010, «Large-scale machine learning with stochastic gradient descent», dans Proceedings of COMPSTAT'2010, Springer, p. 177–186. [132](#)
- BRO-NIELSEN, M. et C. GRAMKOW. 1996, «Fast fluid registration of medical images», dans Visualization in Biomedical Computing, Springer, p. 265–276. [15](#)
- BROIT, C. 1981, «Optimal registration of deformed images», . [15](#)
- BROWN, L. G. 1992, «A survey of image registration techniques», ACM Comput. Surv., vol. 24, n° 4, doi :10.1145/146370.146374, p. 325–376, ISSN 0360-0300. [10](#)
- BRUNET, F. 2010, Contributions to Parametric Image Registration and 3D Surface Reconstruction, thèse de doctorat, Université d’Auvergne, Technische Universitat Munchen. [15](#), [17](#), [18](#)
- CACHIER, P. et X. PENNEC. 2000, «3d non-rigid registration by gradient descent on a gaussian-windowed similarity measure using convolutions», dans Mathematical Methods in Biomedical Image Analysis, IEEE, p. 182–189. [15](#)
- CALONDER, M., V. LEPETIT, C. STRECHA et P. FUA. 2010, «Brief : Binary robust independent elementary features», Computer Vision–ECCV 2010, p. 778–792. [34](#)
- CARUANA, R., S. LAWRENCE et L. GILES. 2000, «Overfitting in neural nets : Backpropagation, conjugate gradient, and early stopping», dans NIPS, p. 402–408. [120](#)
- CHEUNG, W. et G. HAMARNEH. 2007, «N-sift : N-dimensional scale invariant feature transform for matching medical images», dans Biomedical Imaging : From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on, IEEE, p. 720–723. [45](#), [49](#)
- CHOI, Y. et S. LEE. 2000, «Injectivity conditions of 2d and 3d uniform cubic b-spline functions», Graphical models, vol. 62, n° 6, p. 411–427. [84](#)
- CHRISTENSEN, G. E., S. C. JOSHI et M. I. MILLER. 1997, «Volumetric transformation of brain anatomy», IEEE transactions on medical imaging, vol. 16, n° 6, p. 864–877. [15](#)
- CHRISTENSEN, G. E., R. D. RABBITT et M. I. MILLER. 1996, «Deformable templates using large deformation kinematics», IEEE transactions on image processing, vol. 5, n° 10, p. 1435–1447. [15](#)

- COHEN-OR, D. et O. SORKINE. 2006, «Encoding meshes in differential coordinates», dans SCCG06) Proceedings of the 22nd Spring Conference on Computer Graphics. ACM, New York, Citeseer. 64, 65
- COOTES, T. F., C. J. TWINING, V. S. PETROVIC, K. O. BABALOLA et C. J. TAYLOR. 2010, «Computing accurate correspondences across groups of images», IEEE transactions on pattern analysis and machine intelligence, vol. 32, n° 11, p. 1994–2005. 22
- CRESSIE, N. et D. M. HAWKINS. 1980, «Robust estimation of the variogram : I», Journal of the International Association for Mathematical Geology, vol. 12, n° 2, p. 115–125. 41
- DATTA, R., J. LI et J. Z. WANG. 2005, «Content-based image retrieval : approaches and trends of the new age», dans Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, ACM, p. 253–262. 34
- DAVATZIKOS, C. 1997, «Spatial transformation and registration of brain images using elastically deformable models», Computer Vision and Image Understanding, vol. 66, n° 2, p. 207–222. 15
- DE BOOR, C. 1978, A practical guide to splines, vol. 27, Springer-Verlag New York. 15
- DUCHENNE, O., F. BACH, I.-S. KWEON et J. PONCE. 2011, «A tensor-based algorithm for high-order graph matching», IEEE transactions on pattern analysis and machine intelligence, vol. 33, n° 12, p. 2383–2395. 22
- DUPUIS, P., U. GRENANDER et M. I. MILLER. 1998, «Variational problems on flows of diffeomorphisms for image matching», Quarterly of applied mathematics, p. 587–600. 15
- DURRELMAN, S., X. PENNEC, A. TROUVÉ, P. THOMPSON et N. AYACHE. 2008, «Inferring brain variability from diffeomorphic deformations of currents : an integrative approach», Medical image analysis, vol. 12, n° 5, p. 626–637. 23
- DUTAGACI, H., C. P. CHEUNG et A. GODIL. 2012, «Evaluation of 3d interest point detection techniques via human-generated ground truth», The Visual Computer, vol. 28, n° 9, p. 901–917. 45
- EILERS, P. H. et B. D. MARX. 1996, «Flexible smoothing with b-splines and penalties», Statistical science, p. 89–102. 16, 83
- FISCHER, B. et J. MODERSITZKI. 2003, «Curvature based image registration», Journal of Mathematical Imaging and Vision, vol. 18, n° 1, p. 81–85. 15
- FISCHER, B. et J. MODERSITZKI. 2008, «Ill-posed medicine-an introduction to image registration», Inverse Problems, vol. 24, n° 3, p. 034 008. 10
- FISCHLER, M. A. et R. C. BOLLES. 1981, «Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography», Communications of the ACM, vol. 24, n° 6, p. 381–395. 22, 29
- FISHER, B. et J. MODERSITZKI. 2002, «Fast diffusion registration, ams contemporary mathematics : inverse problems'», Image Analysis and Medical Imaging, vol. 313, p. 117–129. 15
- FRAHM, J.-M., P. FITE-GEORGEL, D. GALLUP, T. JOHNSON, R. RAGURAM, C. WU, Y.-H. JEN, E. DUNN, B. CLIPP, S. LAZEBNIK et collab.. 2010, «Building rome on a cloudless day», dans Computer Vision–ECCV 2010, Springer, p. 368–381. 43
- FRANCHI, A., R. AGIER, R. PROST, R. KÉCHICHAN, S. VALETTE et L. FANTON. 2017, «Anatomie computationnelle : Quelles perspectives en anthropologie médico-légale ?», dans 50ème congrès international francophone de Médecine Légale. 136, 138

- GHERARDI, R., M. FARENZENA et A. FUSIELLO. 2010, «Improving the efficiency of hierarchical structure-and-motion», dans Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, p. 1594–1600. [43](#)
- GLOCKER, B., N. KOMODAKIS, N. PARAGIOS et N. NAVAB. 2009, «Approximated curvature penalty in non-rigid registration using pairwise mrfs», dans International Symposium on Visual Computing, Springer, p. 1101–1109. [15](#)
- GOODALL, C. 1991, «Procrustes methods in the statistical analysis of shape», Journal of the Royal Statistical Society. Series B (Methodological), p. 285–339. [22](#)
- GOULD, N. I. et S. LEYFFER. 2003, «An introduction to algorithms for nonlinear optimization», dans Frontiers in numerical analysis, Springer, p. 109–197. [27](#)
- GRAPS, A. 1995, «An introduction to wavelets», IEEE computational science and engineering, vol. 2, n° 2, p. 50–61. [38](#)
- HAGER, W. W. et H. ZHANG. 2006, «A survey of nonlinear conjugate gradient methods», Pacific journal of Optimization, vol. 2, n° 1, p. 35–58. [25](#)
- HAMM, J., C. DAVATZIKOS et R. VERMA. 2009, «Efficient large deformation registration via geodesics on a learned manifold of images», dans International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, p. 680–687. [33](#)
- HARRIS, C. et M. STEPHENS. 1988, «A combined corner and edge detector.», dans Alvey vision conference, vol. 15, Manchester, UK, p. 50. [35](#)
- HARTLEY, R. et A. ZISSERMAN. 2003, Multiple view geometry in computer vision, Cambridge university press. [8](#), [14](#)
- HILL, D. L., P. G. BATCHELOR, M. HOLDEN et D. J. HAWKES. 2001, «Medical image registration», Physics in medicine and biology, vol. 46, n° 3, p. R1. [2](#), [30](#)
- HOAGLIN, D. C., F. MOSTELLER et J. W. TUKEY. 1983, Understanding robust and exploratory data analysis, vol. 3, Wiley New York. [87](#)
- HOLLAND, P. W. et R. E. WELSCH. 1977, «Robust regression using iteratively reweighted least-squares», Communications in Statistics-theory and Methods, vol. 6, n° 9, p. 813–827. [28](#), [97](#)
- HUANG, B., W. WANG, M. BATES et X. ZHUANG. 2008, «Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy», Science, vol. 319, n° 5864, p. 810–813. [10](#)
- HUBER, P. J. 2011, Robust statistics, Springer. [41](#), [86](#)
- HUBER, P. J. et collab.. 1964, «Robust estimation of a location parameter», The Annals of Mathematical Statistics, vol. 35, n° 1, p. 73–101. [40](#)
- IGLESIAS, J. E. et M. R. SABUNCU. 2015, «Multi-atlas segmentation of biomedical images : A survey», Medical Image Analysis, vol. 24, n° 1, p. 205 – 219, ISSN 1361-8415. [30](#), [77](#), [132](#), [134](#)
- JACINTO, H., R. KÉCHICHIAN, M. DESVIGNES, R. PROST et S. VALETTE. 2012, «A web interface for 3d visualization and interactive segmentation of medical images», dans 17th International Conference on 3D Web Technology (Web 3D 2012), Los-Angeles, USA, p. 51–58. URL <http://hal.archives-ouvertes.fr/hal-00732335>. [70](#)

- JIAN, B. et B. C. VEMURI. 2011, «Robust point set registration using gaussian mixture models», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, n° 8, p. 1633–1645. [23](#)
- JOSHI, S., B. DAVIS, M. JOMIER et G. GERIG. 2004, «Unbiased diffeomorphic atlas construction for computational anatomy», NeuroImage, vol. 23, p. S151–S160. [31](#)
- KARLSSON, N., E. DI BERNARDO, J. OSTROWSKI, L. GONCALVES, P. PIRJANIAN et M. E. MUNICH. 2005, «The vslam algorithm for robust localization and mapping», dans Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, IEEE, p. 24–29. [33](#), [43](#)
- KEALL, P. J., S. JOSHI, S. S. VEDAM, J. V. SIEBERS, V. R. KINI et R. MOHAN. 2005, «Four-dimensional radiotherapy planning for dmlc-based respiratory motion tracking», Medical physics, vol. 32, n° 4, p. 942–951. [10](#)
- KÉCHICHIAN, R., S. VALETTE et M. DESVIGNES. 2017, «Automatic multiorgan segmentation using hierarchically registered probabilistic atlases», dans Cloud-Based Benchmarking of Medical Image Analysis, Springer, p. 185–201. [132](#), [135](#)
- KÉCHICHIAN, R., S. VALETTE, M. SDIKA et M. DESVIGNES. 2014, «Automatic 3d multiorgan segmentation via clustering and graph cut using spatial relations and hierarchically-registered atlases», dans Medical Computer Vision : Algorithms for Big Data, édité par B. Menze, G. Langs, A. Montillo, M. Kelm, H. Müller, S. Zhang, W. Cai et D. Metaxas, Lecture Notes in Computer Science, Springer International Publishing, Springer International Publishing, ISBN 978-3-319-13971-5, p. 201–209. [34](#), [77](#)
- KLEIN, S., M. STARING et J. P. PLUIM. 2007, «Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines», IEEE transactions on image processing, vol. 16, n° 12, p. 2879–2890. [24](#), [26](#)
- KOMODAKIS, N., N. PARAGIOS et G. TZIRITAS. 2011, «Mrf energy minimization and beyond via dual decomposition», IEEE transactions on pattern analysis and machine intelligence, vol. 33, n° 3, p. 531–552. [22](#)
- KYBIC, J. et M. UNSER. 2003, «Fast parametric elastic image registration», IEEE transactions on image processing, vol. 12, n° 11, p. 1427–1442. [26](#)
- LANGS, G., A. HANBURY, B. MENZE et H. MÜLLER. 2013, «Visceral : Towards large data in medical imaging – challenges and directions», dans Proceedings of the Third MICCAI International Conference on Medical Content-Based Retrieval for Clinical Decision Support, MCBR-CDS’12, Springer-Verlag, Berlin, Heidelberg, ISBN 978-3-642-36677-2, p. 92–98, doi :10.1007/978-3-642-36678-9_9. URL http://dx.doi.org/10.1007/978-3-642-36678-9_9. [92](#), [111](#)
- LÓPEZ, A. M., F. LUMBRERAS, J. SERRAT et J. J. VILLANUEVA. 1999, «Evaluation of methods for ridge and valley detection», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, n° 4, p. 327–335. [34](#)
- LÖTJÖNEN, J. M., R. WOLZ, J. R. KOIKKALAINEN, L. THURFJELL, G. WALDEMAR, H. SOININEN, D. RUECKERT, A. D. N. INITIATIVE et collab.. 2010, «Fast and robust multi-atlas segmentation of brain magnetic resonance images», Neuroimage, vol. 49, n° 3, p. 2352–2365. [134](#)
- LOWE, D. G. 1999, «Object recognition from local scale-invariant features», dans Computer vision, 1999. The proceedings of the seventh IEEE international conference on, vol. 2, Ieee, p. 1150–1157. [33](#), [34](#), [35](#), [36](#), [38](#)

- MAES, F., A. COLLIGNON, D. VANDERMEULEN, G. MARCHAL et P. SUETENS. 1997, «Multimodality image registration by maximization of mutual information», IEEE transactions on medical imaging, vol. 16, n° 2, p. 187–198. [21](#)
- MAINTZ, J. A. et M. A. VIERGEVER. 1998, «A survey of medical image registration», Medical image analysis, vol. 2, n° 1, p. 1–36. [10](#)
- MARTIN, S., V. DAANEN et J. TROCCAZ. 2008, «Atlas-based prostate segmentation using an hybrid registration», International Journal of Computer Assisted Radiology and Surgery, vol. 3, n° 6, p. 485–492. [134](#)
- MICHALSKI, R. S., J. G. CARBONELL et T. M. MITCHELL. 2013, Machine learning : An artificial intelligence approach, Springer Science & Business Media. [30](#)
- MIKOLAJCZYK, K. et C. SCHMID. 2005, «A performance evaluation of local descriptors», IEEE transactions on pattern analysis and machine intelligence, vol. 27, n° 10, p. 1615–1630. [22](#), [34](#), [55](#)
- MOON, T. K. 1996, «The expectation-maximization algorithm», IEEE Signal processing magazine, vol. 13, n° 6, p. 47–60. [87](#)
- MORÉ, J. J. 1978, «The levenberg-marquardt algorithm : implementation and theory», dans Numerical analysis, Springer, p. 105–116. [26](#)
- MURAIL, P., J. BRUZEK, F. HOUËT et E. CUNHA. 2005, «Dsp : a tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements», Bulletins et Mémoires de la Société d'Anthropologie de Paris, , n° 17 (3-4), p. 167–176. [136](#)
- NOBLET, V., C. HEINRICH, F. HEITZ et J.-P. ARMSPACH. 2012, «An efficient incremental strategy for constrained groupwise registration based on symmetric pairwise registration», Pattern Recognition Letters, vol. 33, n° 3, p. 283–290. [31](#)
- OMOHUNDRO, S. M. 1989, «Five balltree construction algorithms», cahier de recherche. [80](#)
- PENNEC, X., N. AYACHE et J.-P. THIRION. 2000, «Landmark-based registration using features identified through differential geometry», . [33](#)
- PHILBIN, J., O. CHUM, M. ISARD, J. SIVIC et A. ZISSERMAN. 2007, «Object retrieval with large vocabularies and fast spatial matching», dans Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, p. 1–8. [3](#)
- PRECHELT, L. 1998, «Automatic early stopping using cross validation : quantifying the criteria», Neural Networks, vol. 11, n° 4, p. 761–767. [120](#)
- PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING et B. P. FLANNERY. 1982, Numerical recipes in C, vol. 2, Cambridge Univ Press. [25](#)
- RAU, J., N.-Y. CHEN et L.-C. CHEN. 2002, «True orthophoto generation of built-up areas using multi-view images», Photogrammetric Engineering and Remote Sensing, vol. 68, n° 6, p. 581–588. [8](#)
- ROCHE, A., G. MALANDAIN et N. AYACHE. 1999, Unifying maximum likelihood approaches in medical image registration, thèse de doctorat, Inria. [21](#)
- ROHLFING, T., R. BRANDT, R. MENZEL et C. R. MAURER. 2004, «Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains», NeuroImage, vol. 21, n° 4, p. 1428–1442. [134](#)

- RUBLEE, E., V. RABAUD, K. KONOLIGE et G. BRADSKI. 2011, «Orb : An efficient alternative to sift or surf», dans Computer Vision (ICCV), 2011 IEEE international conference on, IEEE, p. 2564–2571. [34](#)
- RUSINKIEWICZ, S. et M. LEVOY. 2001, «Efficient variants of the icp algorithm», dans 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, IEEE, p. 145–152. [23](#)
- SCAHILL, R. I., C. FROST, R. JENKINS, J. L. WHITWELL, M. N. ROSSOR et N. C. FOX. 2003, «A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging», Archives of neurology, vol. 60, n° 7, p. 989–994. [10](#)
- SHANNO, D. F. 1970, «Conditioning of quasi-newton methods for function minimization», Mathematics of computation, vol. 24, n° 111, p. 647–656. [25](#)
- SOTIRAS, A., C. DAVATZIKOS et N. PARAGIOS. 2013, «Deformable medical image registration : A survey», Medical Imaging, IEEE Transactions on, vol. 32, n° 7, p. 1153–1190. [10](#)
- SU, L.-M., B. P. VAGVOLGYI, R. AGARWAL, C. E. REILEY, R. H. TAYLOR et G. D. HAGER. 2009, «Augmented reality during robot-assisted laparoscopic partial nephrectomy : toward real-time 3d-ct to stereoscopic video registration», Urology, vol. 73, n° 4, p. 896–900. [10](#)
- SUTSKEVER, I., J. MARTENS, G. DAHL et G. HINTON. 2013, «On the importance of initialization and momentum in deep learning», dans International conference on machine learning, p. 1139–1147. [27](#), [132](#)
- THIRION, J.-P. 1996, «The extremal mesh and the understanding of 3d surfaces», International Journal of Computer Vision, vol. 19, n° 2, p. 115–128. [33](#)
- THIRION, J.-P. 1998, «Image matching as a diffusion process : an analogy with maxwell’s demons», Medical image analysis, vol. 2, n° 3, p. 243–260. [15](#)
- THOMPSON, P. M., D. MACDONALD, M. S. MEGA, C. J. HOLMES, A. C. EVANS et A. W. TOGA. 1997, «Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces», Journal of computer assisted tomography, vol. 21, n° 4, p. 567–581. [30](#)
- TRIGGS, B., P. F. MCLAUCHLAN, R. I. HARTLEY et A. W. FITZGIBBON. 2000, «Bundle adjustment – a modern synthesis», dans Vision algorithms : theory and practice, Springer, p. 298–372. [10](#), [43](#), [44](#)
- TSITSIKLIS, J., D. BERTSEKAS et M. ATHANS. 1986, «Distributed asynchronous deterministic and stochastic gradient optimization algorithms», IEEE transactions on automatic control, vol. 31, n° 9, p. 803–812. [132](#)
- ULYSSES, J. et A. CONCI. 2010, «Measuring similarity in medical registration», dans IWSSIP 17th International Conference on Systems, Signals and Image Processing. [21](#)
- UMEYAMA, S. 1991, «Least-squares estimation of transformation parameters between two point patterns», IEEE Transactions on pattern analysis and machine intelligence, vol. 13, n° 4, p. 376–380. [57](#)
- VAILLANT, M. et J. GLAUNÈS. 2005, «Surface matching via currents», dans Information processing in medical imaging, Springer, p. 1–5. [23](#)
- WAHBA, G. 1990, Spline models for observational data, SIAM. [20](#)
- WEDDERBURN, R. W. 1974, «Quasi-likelihood functions, generalized linear models, and the gauss—newton method», Biometrika, vol. 61, n° 3, p. 439–447. [26](#), [97](#)

- WELLS, W. M., P. VIOLA, H. ATSUMI, S. NAKAJIMA et R. KIKINIS. 1996, «Multi-modal volume registration by maximization of mutual information», Medical image analysis, vol. 1, n° 1, p. 35–51. [21](#)
- WITKIN, A. 1984, «Scale-space filtering : A new approach to multi-scale description», dans Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84., vol. 9, IEEE, p. 150–153. [35](#)
- WU, G., H. JIA, Q. WANG et D. SHEN. 2011, «Sharpmean : groupwise registration guided by sharp mean image and tree-based registration», NeuroImage, vol. 56, n° 4, p. 1968–1981. [2](#), [33](#)
- WU, G., Q. WANG, H. JIA et D. SHEN. 2012, «Feature-based groupwise registration by hierarchical anatomical correspondence detection», Human brain mapping, vol. 33, n° 2, p. 253–271. [2](#), [30](#), [32](#), [105](#)
- WUEST, H., D. STRICKER et J. HERDER. 2007, «Tracking of industrial objects by using cad models», Journal of Virtual Reality and Broadcasting, vol. 4, n° 1, p. 1–9. [8](#), [9](#)
- XU, Z., C. P. LEE, M. P. HEINRICH, M. MODAT, D. RUECKERT, S. OURSELIN, R. G. ABRAMSON et B. A. LANDMAN. 2016, «Evaluation of six registration methods for the human abdomen on clinically acquired ct», IEEE Transactions on Biomedical Engineering, vol. 63, n° 8, doi :10.1109/TBME.2016.2574816, p. 1563–1572, ISSN 0018-9294. [2](#), [30](#), [77](#)
- YU, T.-H., O. J. WOODFORD et R. CIPOLLA. 2013, «A performance evaluation of volumetric 3d interest point detectors», International journal of computer vision, p. 1–18. [45](#)
- ZAGORCHEV, L. et A. GOSHTASBY. 2006, «A comparative study of transformation functions for nonrigid image registration», IEEE transactions on image processing, vol. 15, n° 3, p. 529–538. [19](#), [82](#)
- ZÖLLEI, L., E. LEARNED-MILLER, E. GRIMSON et W. WELLS. 2005, «Efficient population registration of 3d data», dans International Workshop on Computer Vision for Biomedical Image Applications, Springer, p. 291–301. [2](#), [30](#)
- ZOU, K. H., S. K. WARFIELD, A. BHARATHA, C. M. TEMPANY, M. R. KAUS, S. J. HAKER, W. M. WELLS, F. A. JOLESZ et R. KIKINIS. 2004, «Statistical validation of image segmentation quality based on a spatial overlap index 1 : Scientific reports», Academic radiology, vol. 11, n° 2, p. 178–189. [67](#)

Liste des travaux et communications

Revue avec comité de lecture

AGIER, R., S. VALETTE, R. KÉCHICHAN R. PROST. 2017, «Hubless 3d medical image bundle registration», soumis le 20 janvier 2017 à IEEE Transactions on Image Processing.

Congrès international

AGIER, R., S. VALETTE, L. FANTON, P. CROISILLE R. PROST. 2016a, «Hubless 3d medical image bundle registration», VISAPP 2016 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 3, p. 265–272.

Communication internationale

FRANCHI, A., R. AGIER, R. PROST, R. KÉCHICHAN, S. VALETTE L. FANTON. 2017, «Anatomie computationnelle : Quelles perspectives en anthropologie médico-légale ?», 50ème congrès international francophone de Médecine Légale.

Communication nationale

AGIER, R., S. VALETTE, L. FANTON, P. CROISILLE R. PROST. 2016b, «Recalage massif d'images 3d», GDR Madics, workshop ATLAS.



FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

NOM : AGIER

DATE de SOUTENANCE : 23/10/2017

(avec précision du nom de jeune fille, le cas échéant)

Prénom : Rémi

TITRE : Recalage de groupes d'images médicales 3D par extraction de points d'intérêt.

NATURE : Doctorat

Numéro d'ordre : 2017LYSEI093

École doctorale : EEA

Spécialité : Traitement de l'Image et du Signal

RESUME :

Les imageurs des hôpitaux produisent de plus en plus d'images 3D et il y a un nombre croissant d'études de cohortes. Afin d'ouvrir la voie à des méthodes utilisant de larges bases de données, il est nécessaire de développer des approches permettant de rendre ces bases cohérentes en recalant les images. Les principales méthodes actuelles de recalage de groupes utilisent des données denses (voxels) et sélectionnent une référence pour mettre en correspondance l'ensemble des images.

Nous proposons une approche de recalage par groupes, sans image de référence, en utilisant seulement des points d'intérêt (Surf3D), applicable à des bases de plusieurs centaines d'images médicales. Nous formulons un problème global fondé sur l'appariement de points d'intérêt. La variabilité inter-individu étant grande, le taux de faux positifs (paires aberrantes) peut être très important (70%). Une attention particulière est portée sur l'élimination des appariements erronés. Une première contribution effectue le recalage rigide de groupes d'images. La première étape calcule les recalages de toutes les paires d'images. En s'appuyant sur le graphe complet de ces recalages, nous formulons le problème global en utilisant l'opérateur laplacien. Des expérimentations avec 400 images scanner CT 3D hétérogènes illustrent la robustesse de notre méthode et sa vitesse d'exécution. Une seconde contribution calcule le recalage déformable de groupes d'images. Nous utilisons des demi-transformations, paramétrées par des pyramides de B-splines, entre chaque image et un espace commun. Des comparaisons sur un jeu de données de référence montrent que notre approche est compétitive avec la littérature tout en étant beaucoup plus rapide.

Ces résultats montrent le potentiel des approches basées sur des points d'intérêt pour la mise en correspondance de grandes bases d'images. Nous illustrons les perspectives de notre approche par deux applications : la segmentation multi-atlas et l'anthropologie.

MOTS-CLES :

Recalage 3D, Recalage de populations, Recalage déformable, Points d'intérêt, Big data

Laboratoire(s) de recherche :

CREATIS - INSA Lyon

Directeur de thèse :

Rémy PROST

Président du jury :

Composition du jury :

Adrien BARTOLI

Fabrice HEITZ

Xavier PENNEC

Jocelyne TROCCAZ

Rémy PROST

Sébastien VALETTE