"Dense deformation estimation for pairwise and multi-subjects registration"

De Craene, Mathieu

ABSTRACT

Medical imaging is nowadays a vital component of a large number of clinical applications. For comparing images of the same patient (sometimes acquired using different modalities) or for comparing different patients, the images need to be aligned. When images from the same patient or from a collection of patients are acquired using different modalities, their alignment is necessary. Registration is the process of finding the best alignment between a pair or a collection of images. The main topic of this thesis is the conception and the application of intensity-based techniques, both for pairwise and multi-subjects registration. In the context of pairwise registration, we have investigated the use of a stochastic gradient descent technique (SPSA) for optimizing the mutual information metric. For non-rigid registration, the use of volumetric tetrahedral meshes has been implemented as a deformation model in collaboration with A. du Bois d'Aische. The targeted application for our algorithm is the tracking of anatomical changes between pre-operative and intra-operative images in brain, prostate and liver surgery. A second method, equivalent to optical flow but developed for multi-modal images is also described and applied to the problem of atlas to pathological brain registration. In the context of multi-subjects registration, we developed an unbiased atlas generation technique in the Expectation-Maximization framework. At each iteration, the method estimates a reference for the registration problem by performing an average giving more weight to consistent experts (E step). ...

CITE THIS VERSION

De Craene, Mathieu. *Dense deformation estimation for pairwise and multi-subjects registration.* Prom. : Warfield, Simon ; Macq, Benoit http://hdl.handle.net/2078.1/5028

Université catholique de Louvain
Faculté des Sciences Appliquées
LABORATOIRE DE TÉLÉCOMMUNICATIONS
ET
TÉLÉDÉTECTION

B - 1348 Louvain-la-Neuve
Belgique

Dense Deformation Field Estimation for Pairwise and Multi-subjects Registration

Mathieu De Craene

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences Appliquées

Examination Committee:

Benoit MACQ (UCL/TELE)
Simon WARFIELD (Harvard Medical School, USA)
Vincent GREGOIRE (UCL/IMRE)
Jean-Philippe THIRAN (EPFL, Switzerland)
Christophe DE VLEESCHOUWER (UCL/TELE)
Luc VANDENDORPE (UCL/TELE) - *President*

August 2005

Acknowledgments

Research is often seen as a lonely quest, spending years of scribbling unintelligible equations or debugging messy programs ... and to a certain extent, this might be the truth. But all along these three years, I discovered that research offered the opportunity of sharing ideas and make collaborative work with people having the same interests all over the world. My first acknowledgments go therefore to Benoît Macq who introduced me to a stimulating work environment favorable to the development of enriching relationships both on human and professional point of views.

I was lucky enough to make a good part of my research work in close collaboration with Aloys du Bois d'Aische. Our stay in Boston confirmed and reinforced our friendship. The time shared together was the opportunity of many debates on research as well as many other topics. These debates lasted and last for days since we are both equally obstinated.

Most of the ideas developed in this thesis have been suggested or encouraged by Simon Warfield, head of the Computational Radiology Laboratory, BWH, Boston. His many advises about algorithmic or implementation issues have been vital to the acomplishment of this work.

I would like to thank all the colleagues I met during my thesis, at the SPL-CRL (Boston) and TELE (Louvain-la-Neuve) labs for the wonderful atmosphere in these labs (sorry but the list is far too long). I shared my office in TELE with Vincent, Daniella, Patrice, Cedric, Annabelle, Sandrine, Monica, Jacek (a great jazz musican and an inspired senior), Damien (another waffels eater), Aloys, Pedro... I hope they do not mind the frequent interuptions in their work for chatting or having coffee breaks. Many thanks to the team of Jean-Philippe Thiran (Valérie Duay, Torsten Butz, Edu Solanas, Meritxell Bach) at EPFL for having me introduced during my master thesis to the field of medical images processing and for the friendship and collaborations that have resulted from this stay in Lausanne.

Finally, I would like to express all my gratitude to my family, and more specifically to my brother Thomas, for showing me every day the real path to happiness.

Contents

Acknowledgments **i**

List of abbreviations **vii**

List of figures **ix**

1 Introduction **1**
- 1.1 Context of the Work 1
- 1.2 Targeted applications 2
  - 1.2.1 Pairwise Registration 2
  - 1.2.2 Registration of a Collection of Subjects 3
- 1.3 Classification of Registration Algorithms 3
- 1.4 Contributions 5
- 1.5 Organization of the Thesis 6

2 Mutual Information for Automatic Alignment of Multi-modal Images **9**
- 2.1 Introduction 9
- 2.2 A Brief History 9
- 2.3 Definition of Mutual Information 11
- 2.4 Implementations 13
  - 2.4.1 Viola and Wells Implementation 13
  - 2.4.2 Mattes Implementation 14
  - 2.4.3 Transformation Jacobian and Metric Flow 16
- 2.5 Discussion about the Use of MI Compared to Other Similarity Metrics 17

3 Maximizing MI : a Variational approach for Free-form Transformations **19**
- 3.1 Variational Principles for Similarity Measures 19
- 3.2 Application to Mutual Information 22
- 3.3 Application : MR-CT Registration using MI Flow 23

3.4 Application : Atlas-Based Segmentation of Pathological Brain MR Images 25
  3.4.1 Introduction 25
  3.4.2 Algorithm Description 26
  3.4.3 Results 28

4 Maximizing Mutual Information for Prior Transformation Models 33
4.1 Registration Pipeline 33
4.2 Transformation Models 33
  4.2.1 Rigid and Affine Transformation Models 33
  4.2.2 Finite Elements as a Non-Rigid Transformation Model 35
4.3 Optimization Strategies 41
  4.3.1 Gradient Descent 41
  4.3.2 Simultaneous Perturbation Stochastic Approximation for Gradient Descent 41
  4.3.3 Genetic Optimization 44
4.4 Application : Mutual Information based Rigid Registration using SPSA 46
  4.4.1 Implementation 46
  4.4.2 CT-MR Liver Registration 46
  4.4.3 Brain MR T1/T2 Registration and DTI/MR Registration 46
4.5 Non-rigid FEM Registration : Preliminary Experiments . . . 47
  4.5.1 Synthetic Luminance Distortion 47
  4.5.2 Synthetic Transformation and Landmark Validation . 48
  4.5.3 Intra-Operative Case 50
4.6 Application : Non-rigid Registration of Pre-Op. and Intra-Op. Images 54
  4.6.1 Background and Related Work 54
  4.6.2 Registration Strategy 55
  4.6.3 Prostate Biopsy 58
  4.6.4 RF Ablation of Liver Tumors 62
  4.6.5 Brain Shift Estimation 65
4.7 Application : Simultaneous Registration and Bias Correction 71
  4.7.1 Introduction 71
  4.7.2 Marginal Bias Correction 72
  4.7.3 Joint Registration and Bias Correction 74
  4.7.4 Results 77

5 Multi-subject Registration for Unbiased Probabilistic Atlas Generation 83
5.1 Existing Probabilistic Atlas Construction Methods 83
5.2 The EM algorithm 85
5.2.1 Optimal Bound 86
5.3 The STAPLE Algorithm 87
5.3.1 E Step 87
5.3.2 M Step 88
5.4 Multi-subjects Registration Algorithm 89
5.4.1 E Step 90
5.4.2 M Step 90
5.4.3 Multi-Subjects Rigid Alignment 92
5.4.4 Multi-Subjects Non-Rigid Alignment 92
5.5 Experiments 93
5.5.1 Rigid Alignment 93
5.5.2 Non-Rigid Alignment 94
5.5.3 Deformations Principal Components Analysis 94

6 Conclusion and Perspectives 103
6.1 Achievements and Contributions 103
6.2 Perspectives and Future Works 104
6.3 Related Publications 105

Bibliography 107

List of abbreviations

CT	Computed Tomography
EM	Expectation Maximization
DTI	Diffusion Tensor Imaging
FEM	Finite Element Method
MI	Mutual Information
MRI	Magnetic Resonance
NE	Normalized Entropy
PCA	Principal Component Analysis
PDF	Probability Density Function
RF	Radio Frequency
SPGR	SPoiled Gradient Recalled
SPSA	Simultaneous Perturbation Stochastic Approximation

List of Figures

1.1 Classification of registration algorithms. 8

2.1 Joint histogram example. 10

3.1 Stationarity of a cost functional. 20
3.2 MR-CT articulated and free-form registration for patient 1. . 24
3.3 MR-CT articulated and free-form registration for patient 2. . 24
3.4 Mutual information and optical flows comparison. 29
3.5 Segmentation results obtained after alignment of the atlas on 3 patients with large tumors 30
3.6 Segmentation results obtained after alignment of the atlas on 3 patients with large tumors 31
3.7 3D rendering of tumor, ventricles, thalamus and central nuclei after atlas alignment on 3 patients with large tumors . . 31

4.1 Registration pipeline. 34
4.2 Tetrahedral mesh generation from a regular cube lattice . . . 40
4.3 Starting from a lattice of two interlaced cubic grids, a regular mesh of tetrahedrons can be generated. 40
4.4 SPSA optimization of mutual information for rigid registration of CT-MR liver images. 47
4.5 Rigid registration of pre-operative MR T1 image on a intra-operative T2 image of the same subject. 48
4.6 Registration of the baseline component of a pre-operative DTI image on a T1 MRI acquired in the operating room. . . 49
4.7 Synthetic luminance distortion for evaluating MI robustness: 2D experiment. 51
4.8 Synthetic transformation and landmark validation: 2D experiment. 52
4.9 Intra-operative brain shift case : 2D experiment. 53
4.10 Non-rigid registration flowchart for a finite elements deformation model with SPSA optimization. 57
4.11 Bias correction. 59

4.12 Bias field inducing a divergence of the registration process . 60
4.13 Rigid and non-rigid registration of a 1.5 pre-operative image of the prostate to 0.5-T intra-operative image(a). 61
4.14 Non-rigid registration of MR liver images. 63
4.15 Non-rigid registration of MR liver images. 64
4.16 Non-rigid registration of brain shift MR images. 66
4.17 Non-rigid registration of brain shift MR images. 67
4.18 Non-rigid transformation inversion. 68
4.19 Non-rigid brain shift displacement field. 69
4.20 Brain shift : difference images before and after registration. . 70
4.21 Algorithmic flow chart 78
4.22 MI/NE comparison for features selection. 79
4.23 Optimal number of bins joint selection using Normalized Entropy. 80
4.24 Optimal bias correction using Normalized Entropy. 80
4.25 Joint histograms before and after alignment. 81
4.26 Contours of the preoperative image reported on the intra-operative (ODIN) image. 82

5.1 Sum of the 80 segmentations (four labels) after rigid and scale alignment and after 2 iterations of non-rigid warping. . 96
5.2 Weighted sum. 97
5.3 Rigid and scale alignment of typical subjects (sagittal slice). . 98
5.4 Non-rigid alignment of typical subjects (sagittal slice). 99
5.5 Rigid and scale alignment of typical subjects (coronal slice). 100
5.6 Non-rigid alignment of typical subjects (coronal slice). 101
5.7 Principal deformation modes. 102

Introduction

1

1.1 Context of the Work

During the past decades, medical imaging has become a vital component of a large number of clinical applications. One of them is the track of anatomical features for the planning and assessment of surgical and radiotherapeutic procedures.

Since information gained from two or several image modalities acquired in the track of anatomical features is usually of a complementary nature, a proper *integration* of useful data obtained from the separated images is often desired. The preliminary step in this integration process is to bring the acquired images into spatial alignment. The term *registration* refers to this alignment procedure.

The first registration techniques - still in use in clinical practice - performed registration by manual adjustment of rotations and translations. The practitioner proceeds by translating the contours of one image onto a second. The main drawback of manual registration is the lack of reproducibility and therefore the intra and extra observers errors that result from experience or external conditions.

Over time, automated *rigid registration algorithms* have been developed, by minimization of a mean square error for mono-modal images and by matching corresponding boundary surfaces extracted from each modality for multi-modal images. As many applications require to estimate more complex transformations, automated *non-rigid registration algorithms* have been investigated by constraining the transformation with landmarks (see e.g. Bookstein [1989]) in each modality and then projecting the transformations on basis functions such as thin-plate splines.

Registration algorithms have also been studied in other domains than medical imaging. In video images processing, such algorithms can detect movement between two consecutive frames. One of these techniques, called optical flow [Thirion 1998] has been applied for matching 3D medical images.

In 1995, mutual information came up as a powerful measure to measure the global similarity between multi-modal medical images [Maes and Collignon 1997, Viola and Wells 1995]. After rigid registration (see Collignon et al. [1995a], Wells et al. [1996]), mutual information has been successfully used for non-rigid matching of multi-modal volumes (see Rohde et al. [2003], Ruekert et al. [1999]) and for a general survey, see Pluim et al. [2003]). Other multi-modal criterions such as correlation ratio or normalized entropy have also been investigated (see Roche et al. [1998]). The volumes registered were, until this point, smooth tissues, such as brain, breast and the methods to register these were quite slow. New methods for surgery appeared, registering pre-segmented surfaces together and inferring the displacement found on the surface to the whole volume through a linear elastic model implemented using a finite element method (see e.g. Ferrant et al. [2001], Papademetris et al. [2002]). These algorithms brought forward the possibility of including various elasticities within the mechanical model.

1.2 Targeted applications

1.2.1 Pairwise Registration

During surgery, registration is required to fuse pre-operative and intra-operative data in the Operating Room (OR).

- In neuro-surgery, (non-rigid) brain deformations after craniotomy need to be estimated to register pre-operative segmentations or landmarks with the OR referential space.

- After Radio frequency ablation of liver tumors, the overlap between the initial tumor and the necrosis after surgery has to be estimated. For computing this overlap, the pre-procedural and post-procedural images have to be aligned. Non-rigid deformations between these images are induced by different breath hold patterns.

- MR guided prostate biopsy is another surgical application of registration algorithms : tumor foci and internal prostate structures can be better localized and identified on the pre-operative 1.5-T MR image than in the 0.5-T used for surgical navigation. Fusion of these two modalities brings the advantages of MR guided navigation and an accurate prostate visualization.

1.2.2 Registration of a Collection of Subjects

Bringing a collection of subjects in a common space of coordinates enables to build statistical atlases. Such atlases are an important representation for characterizing anatomy and anatomical variation. They are also used in identifying structural differences between populations, such as between healthy newborn infants and infants with white matter injury. Furthermore, statistical atlases provide prior probability models useful as constraints to robustify and speed up segmentation and registration algorithms. For example, given prior information about the main modes of shape variation, a non-rigid registration algorithm may optimize initially upon these main modes of variation instead of considering all parameters of the deformation model simultaneously. The use of statistical atlases has also been investigated for improving the accuracy of automatic segmentations algorithms and to robustify them against random noise and artifacts [Warfield et al. 2000].

1.3 Classification of Registration Algorithms

This section attempts to classify existing algorithms using *voxel-based similarity metrics* for automatic registration. These methods distinguish themselves from other registration techniques by the fact that they operate directly on the image grey values, without reducing first the data to a set of segmented structures to be matched or without any identification by the user of key points in the modalities to match.

A quick glance to the medical images registration litterature (see Maintz and Viergever [1998], Pluim et al. [2003] for a survey) shows two separated algorithmic classes :

- registration algorithms using a prior transformation model. The most common models are rigid or affine. More sophisticated models have been proposed for non-rigid registration (B-Splines by Rueckert et al. [2003]). Having a limited number of parameters to optimize raise the possibility to optimize the similarity metric with a wide range of optimization methods : gradient descent, Brent-Powell [Brent 1973], stochastic methods [Goldberg 1989, Spall 1998], quasi-Newton methods [Nocedal and Wright 2000], etc...
- registration algorithms addressing the estimation of a dense deformation field, considering independently the 3 components of the

displacement at each voxel. Variational methods, estimating incrementally a dense deformation field bringing the two images to the closest local minimum of the metric, are the only option for addressing such a huge dimension of the optimization space. The theory beyond these algorithms is related to functional analysis and Euler-Lagrange equations. A short description is given in Chapter 3. The reader is invited to refer to Hermosillo et al. [2001] for more details about the mathematical background.

Figure 1.1 illustrates this classification. Double arrows indicate links between algorithmic entities constituting an integrated registration solution.

The first class of algorithm makes use of three algorithmic objects :

- The transformation object which takes as input a set of parameters characterizing the transformation. It transforms coordinates in the fixed image domain to coordinates in the moving image domain.
- The metric which provides a score (cost function) for a set of transformation parameters
- The optimizer, searching for the set of transformation parameters yielding the best metric score

In the case of a dense deformation field (second class of algorithms), an incremental dense displacement field is estimated at each iteration. The similarity metric has to be differentiable for computing at each point the infinitesimal variation of the metric due to a variation in the three directions of space of the current displacement at this point. For similarity metrics based on statistics and taking the images in their globality, a preliminary step is often the construction of a joint probability density and its derivatives. The various implementations of mutual information in the literature differ over the way of estimating this probability density.

Simple arrows in Figure 1.1 show a classification between similarity metrics. Registration algorithms have been mainly focused on *pairwise registration* either for registering an atlas on a subject or to register two images of the same subjects. A wide set of metrics have been proposed [Maintz and Viergever 1998] : Sum of Squared Differences [Christensen et al. 1998, Unser et al. 1995] (only valid for the same modality with properly normalized intensities), Normalized Cross-Correlation [Maintz et al. 1996, Studholme et al. 1995b] (which allows for linear relationship between the intensities of the two images), mutual information [Maes and

Collignon 1997, Viola and Wells 1995], minimization of variance of intensity ratios [Hill et al. 1993b, Zuo et al. 1996].

More recently, an important research effort has been dedicated to registration techniques able to align a collection of subjects in the same space of coordinates. An overview of these existing methods is given in Chapter 5, Section 5.1.

1.4 Contributions

In the light of this global introduction, we summarize in this section the main contributions of this thesis.

In the context of intra-subject registration the contributions of this thesis are

- The development in collaboration with Aloys du Bois d'Aische of a FEM non-rigid registration method. A classical gradient descent scheme has been used as optimization method. This algorithm has been published in Media [du Bois d'Aische et al. 2004]. Our contribution has been to integrate Finite Elements as a non-rigid registration model in the ITK design.

- The use of a stochastic optimization method, the Simultaneous Perturbation Stochastic Approximation (SPSA) for maximizing the mutual information metric between pre-operative and intra-operative images. The SPSA method has been widely used in many fields where high dimensional optimization problems must be addressed (adaptive optics, atmospheric and planetary modeling, cardiological data analysis, . . .).[1] We chose to investigate the use of SPSA in our work because of the noisy nature of mutual information. Our code of the SPSA has been included in the Insight registration and segmentation toolkit. The experiments described in Chapter 4 (Sections 4.4 and 4.6) show that SPSA is suited for capturing rigid as well as non-rigid deformations. The decreasing in the amplitude of stochastic perturbations is quite similar to simulated annealing techniques : a large part of the search domain is explored at the begining of the process before restraining the search in a closer part of the domain.

---

[1] A complete list of references is available on this website : `http://www.jhuapl.edu/SPSA/Pages/References-List_Ref.htm`

- The implementation of a mutual information flow for estimating a dense deformation field between multi-modal images has also been submitted for inclusion in the Insight Segmentation and Registration toolkit. It mixes the Mattes et al. [2003] implementation of mutual information and the flow concept introduced by Hermosillo et al. [2001]. We applied this algorithm for capturing non-rigid deformations between MR and CT image in the head and neck area (Chapter 3, Section 3.3) and for atlas to pathological brain registration (Chapter 3, Section 3.4).

In the context of multi-subject registration for building probabilistic atlases, the main difficulty is to address the matching of a large data base of subjects. A straight extension of classical image similarity metrics would require to estimate histograms whose dimensionality is related to the number of subjects in the database. Our contribution consists in first estimating a central tendency represented by tissue probabilities at each voxel. All the subjects are then matched on this central tendency by maximizing mutual information between the probabilistic atlas and each subject. The computation of the joint histogram is slightly different than in the classical case since one of the images is a probability map [De Craene et al. 2005; 2004b]. Our strategy iteratively builds an unbiased reference giving more weight to consistent subjects in the database. All subjects are then aligned on this rereference to increase the global consistency of the database. Our method avoids to introduce any bias in the selection of a subject as a refence. The weigthing scheme giving more importance to consistent subjects avoids to work with a fuzzy reference that a simple average would produce.

1.5 Organization of the Thesis

Figure 1.1 illustrates how each chapter of this thesis fits into the short classification of registration algorithms made in Section 1.3.

Review of existing methods *Chapter 2* presents an overview of existing methods for performing automatic registration by maximizing mutual information. Different implementations and algorithms are summarized in this chapter. We also show the link between a classical gradient descent and the variational flow introduced in Chapter 3.

Pairwise registration *Chapters 3 and 4* present two different registration strategies following the transformation model used in the registration procedure. In *Chapter 3*, the deformation is represented by a displacement vector at each voxel. The dimension of the search space is so big that a variational solution is the best alternative for reaching the nearest localoptimum. In *Chapter 4*, a prior model of the transformation (rigid or non-rigid) enables to keep to the optimization problem a reasonably low complexity. Gradient descent, stochastic gradient descent and genetic optimization schemes are shortly described.

Multi-subject registration *Chapter 5* presents a novel algorithm for bringing a collection of subjects into the same space of coordinates. This algorithm is currently designed for aligning a collection of segmentations. Experiments on a database of 80 brain segmentations are presented.

Conclusion and perspectives Overall conclusions of the thesis are given in *Chapter 6* along with a review of our contributions. We also point out perspectives for further research and collaborations started or continued all along this thesis work.

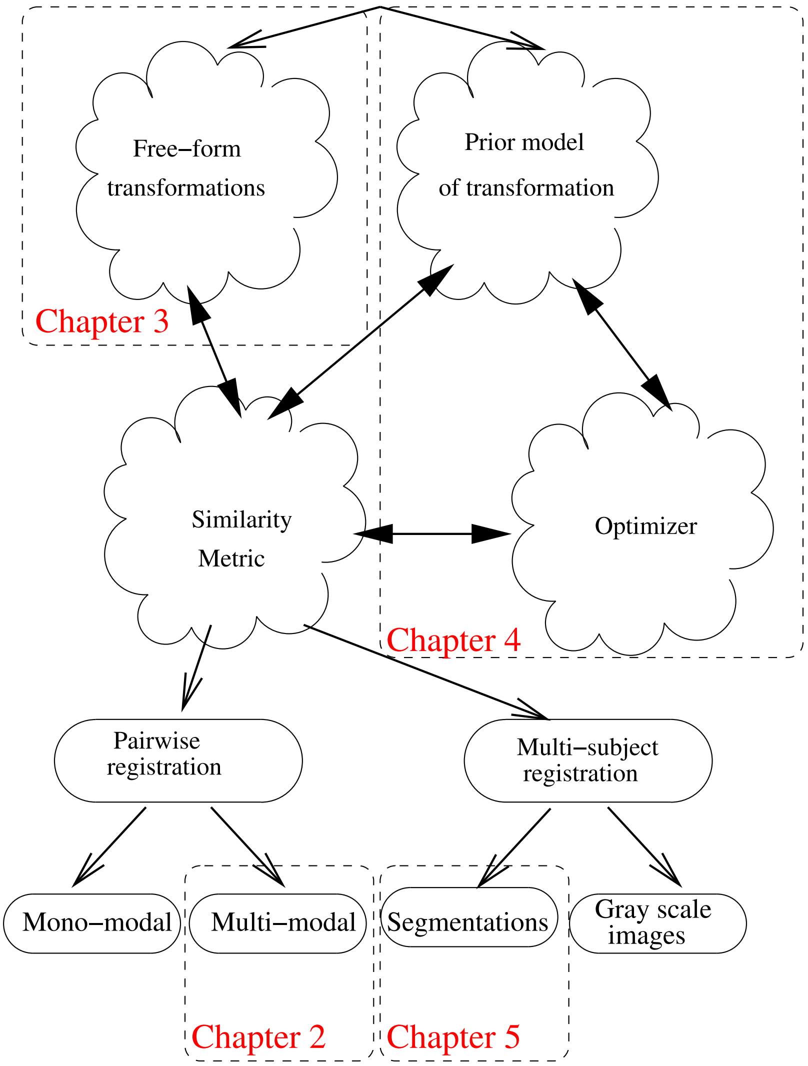


Figure 1.1: *Classification of registration algorithms.*
This figure shows how the different chapters of this thesis fit into the existing registration algorithms. Double arrows indicate collaborations between constitutive parts of a registration algorithm. Single arrows indicate a classification of volumetric similarity metrics depending on the nature of the information to be matched.

Mutual Information for Automatic Alignment of Multi-modal Images

2

In this chapter, we introduce the mutual information metric as a similarity measure for automatic images alginment. We present different implementations described in the litterature. We show the link between a first order derivative for a given transformation model and the concept of flow which is developped further in Chapter 3.

2.1 Introduction

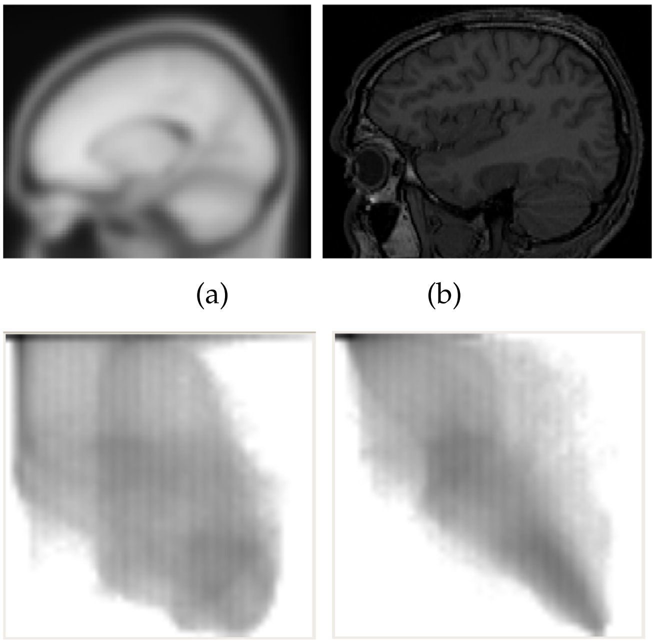
The reserach which lead to the wide use of mutual information in registration starts in the 1990's when the problem of matching multi-modal images like MR-CT images raised the need of multi-modal similarity measures. An excellent review of of the litterature related to mutual information based registration of medical images can be found in [Pluim et al. 2003]. In this section, only two different implementations of mutual information are presented :

- the one by Viola and Wells [1996; 1995] which uses two set of random samples for estimating by cross-validation the expectation of a random variable whose probability density is modeled by Parzen windowing
- the one by Mattes et al. [2003] which mixes joint histograms and continous kernel basis functions in order to estimate the probability densities.

2.2 A Brief History

Woods et al. [1992; 1993] first introduced the assumption that regions with similar tissues (and hence similar grey values) in one image would correspond to homogeneous regions in the other image. Therefore, the ratio of grey levels at corresponding points in a certain regions in either images should vary little. The variance of this ratio is thus minimized to achieve registration.

Hill et al. [1993a] introduced the use of a feature space or *joint histogram* measuring if the images are aligned or not. This joint histogram is built from pairs of joint intensities in the fixed and moving images for the current alignment. The idea is intuitively that a right alignment between the images will create clusters in the joint histogram. An example of joint histogram before and after rigid alignment is given in Figure 2.1. In this case, a MR image (Figure 2.1(b)) has to be aligned with the SPM atlas (Figure 2.1(a)) which is obtained by averaging a collection of aligned subjects. The joint histogram after alignment (Figure 2.1(d)) shows two picks for the background and the skull whereas the initial histogram (Figure 2.1(c))



(a) (b)

(c) (d)

Figure 2.1: *Joint histogram example.*
The images to be aligned are shown in (a) (SPM atlas) and (b). Sub-figures (c) and (d) show the joint histogram before and after alignment.

is more dispersed and shows associations between the background in the first image and a wide range of gray values in the second image.

A wide family of similarity measures based on the construction of a joint histogram has emerged. Hill [Hill et al. 1994; 1996] has proposed the third order moment of the joint histogram, which measures the skewness of a distribution. Collignon et al. [1995a;b] and Studholme et al. [1995a] suggested both to use joint entropy as a measure to perform automatic registration. Mutual information appeared then quickly, simultaneously introduced by Maes and Collignon [1997] and Viola and Wells [1996; 1995]. In a few years, it became one the most investigated measure for medical images registration.

2.3 Definition of Mutual Information

For measuring mutual information, two features are extracted from the images to align. The most common choice of feature is the signal intensity at each voxel of the two images. Each of these features, considered separately can be modelized by a random variable (RV) whose probability density function must be estimated from a large set of realizations (since each voxel is a realization of the RV). For measuring similarity between these random variables, a new random variable is built from the co-occurences of image features in the fixed and moving images. When another alignment is considered, the overlap between grey level is changed and the probability density function of the joint random variable is modified.

Mutual information actually measures a balance between the Shannon entropies of the marginal and joint entropies of the image features. If we call F the feature extracted from the fixed image, M the feature extracted from the moving image and (M, F) the co-occurrence of these two random variable for the current alignment, mutual information is defined as

$$I(F, M) = H(F) + H(M) - H(F, M) \tag{2.1}$$

where $H(\cdot)$ is the marginal entropy [1]. Note that in Equation (2.1), both the moving image entropy and the joint fixed-moving images entropy are changing when another transformation is defined. Equation (2.1) has thus

---

[1] By definition, $H(A) = -\mathrm{E}\{\log(p(A))\}$ if p is the probability density function of the RV A

to be rewritten this way

$$\begin{aligned} I(F, M) &= H(F) + H(M_T) - H(F, M_T) && (2.2) \\ &= H(M_T) - H(M_T|F) && (2.3) \end{aligned}$$

where the T indices shows the dependency to the transformation.

Interpretations of MI Equation (2.3) shows that mutual information actually measures the difference between *a priori* and *a posteriori* entropies. For optimization considerations, the difference between the conditional $H(M_T|F)$ and joint $H(F, M_T)$ entropies is equal to a constant $H(F)$. The minimization of the second term in Equation (2.3) is therefore identical to the minimization of the joint entropy which tends to make the joint histogram as compact as possible. However, a transformation sending the whole moving image outside the fixed image domain would result in a single peak whose entropy is minimal. Mutual information attempts to get far from this undesired solution by adding to the cost function the marginal (prior knowledge) of the moving image for the current transformation. Indeed, there would not be any gain in reducing the entropy of a deterministic random variable.

Another interpretation of mutual information is the Kullback-Leibler divergence. The Kullback-Leibler divergence between two random variables A and B is defined as

$$E_A \left\{ \log \left(\frac{f_A}{f_B} \right) \right\} \tag{2.4}$$

From Equation (2.2) mutual information can be developed as

$$E_{F,M_T} \left\{ \log \left(\frac{p_{F,M_T}}{p_F \cdot p_{M_T}} \right) \right\} \tag{2.5}$$

which shows that mutual information measures the divergence between the joint probability for the current alignment and the worst case scenario of independent variables where the joint probability equals the product of marginal densities. This observations confirms that mutual information measures the statistical dependency of random variables.

Estimation of Probability Densities In practice, the probability density functions of F, M_T and F, M_T are estimated from the realizations of these RV at each voxel of the fixed image domain. If X stands for the fixed image domain and if $x_i \in X$ is a specific voxel, one can observe in x_i the following realizations of F and M_T[2]

$$F \rightarrow f(x_i) \tag{2.6}$$

$$M_T \rightarrow m(T(x_i)) = m(x_i + u_i) \tag{2.7}$$

In Equation (2.7), the transformation of the voxel x_i can be written using $T(x_i)$ or by introducing the displacement u at this voxel

$$T(x_i) = x_i + u_i$$

The key point of all mutual information based registration algorithms is the estimation of a joint probability density function from the marginal and joint realizations (2.6) and (2.7). In the next section, we show how the choice of a specific PDF estimation technique influences the computation of the first derivative, which can be used in a classical gradient descent scheme.

2.4 Implementations

2.4.1 Viola and Wells Implementation

Viola introduced (see Viola and Wells [1996; 1995]) the use of Parzen Windows estimation technique and its application to entropy estimation. If a kernel function g_ψ if defined as

$$g_\psi(z) = (2\pi)^{\frac{-N}{2}} \exp(-\frac{1}{2} z^t \psi^{-1} z) \tag{2.8}$$

(where ψ is a scalar if z is a scalar random variable and ψ is a covariance matrix of size N in case of z is a N dimensional random vector), the estimated probability density will be

$$p(z) = \frac{1}{N_A} \sum_{z_n \in A} g_\psi(z - z_j) \tag{2.9}$$

where A is a sample containing N_A realizations $\{z_j\}$ of the RV Z. To avoid the computation of a continuous integral in the probability domain, the expectation of a random variable is estimated using a sample B

$$E_Z(f(z)) = \frac{1}{N_B} \sum_{z_i \in B} f(z_i) \tag{2.10}$$

---

[2] with f and m standing for the image intensity function of the fixed and the moving images

In the following, we suppose that the sample A is used to estimate the probability density function p and the sample B is used to estimate the expectation operator in $H(Z) = -\mathrm{E}\{\log(p(Z))\}$.

MI First Derivative

Using these definitions, and assuming that the covariance matrix ψ to estimate the fixed-moving images gray levels distribution is diagonal ($\psi_{fm}^{-1} = DIAG(\psi_f^{-1}, \psi_m^{-1})$), Viola shows (in [Viola and Wells 1996]) that the derivative of the Mutual Information metric regarding a parameter T of the transformation can be written as

$$\frac{dI}{dT} = \frac{1}{N_B} \sum_{x_i \in B} \sum_{x_j \in A} (m_i - m_j)(W_m(m_i, m_j) - W_{fm}(fm_i, fm_j)) \psi_m^{-1} \frac{d}{dT}(m_i - m_j) \tag{2.11}$$

where

- m_i(resp. m_j) is the moving image luminance function evaluated at $x_i + u_i$(resp. $x_j + u_j$) : $m_i = m(x_i + u_i)$ and $m_j = m(x_j + u_j)$
- $fm_i = (f_i, m_i)^t$ where $f_i = f(x_i)$
- W_m and W_{fm} are weight factors defined by

$$W_m(m_i, m_j) = \frac{g_{\psi_m}(m_i - m_j)}{\sum_{x_k \in A} g_{\psi_m}(m_i - m_k)} \tag{2.12}$$

$$W_{fm}(fm_i, fm_j) = \frac{g_{\psi_{fm}}(fm_i - fm_j)}{\sum_{x_k \in A} g_{\psi_{fm}}(fm_i - fm_k)} \tag{2.13}$$

In equation (2.11), it remains to evaluate for each transformation model

$$\frac{d}{dT}(m(x_i + u_i)) = (\nabla_x m(x_i + u_i))^t \cdot \frac{du_i}{dT} \tag{2.14}$$

where T is a given parameter.

2.4.2 Mattes Implementation

Mattes[Mattes et al. 2003] has introduced the use of a joint histogram smoothed by 3^{rd} BSpline kernels. This way, the variations of the joint histogram for small perturbations of the transformation parameters get smoother than with box kernels. This is an essential property for first derivative based optimization methods.

In this implementation, the joint probability of moving and fixed image features is estimated from the realizations (2.6) and (2.7) by

$$\begin{aligned} p^{f,m}(i_1,i_2) = \frac{1}{N}\sum_i \beta^1 \left(i_1 - \frac{f(\boldsymbol{x_i}) - f_{min}}{\Delta b_F}\right) \\ \cdot\, \beta^3\left(i_2 - \frac{m(\boldsymbol{x}_i + \boldsymbol{u}_i) - m_{min}}{\Delta b_M}\right) \end{aligned} \tag{2.15}$$

MI First Derivative

For computing the first derivative of mutual information in this case, Equation (2.2) is expanded by injecting the definitions of marginal and joint entropies

$$\begin{aligned} MI = & \sum_{i_1,i_2} p^{f,m}(i_1,i_2)\log(p^{f,m}(i_1,i_2)) \\ & - \sum_{i_2} p^m(i_2)\log(p^m(i_2)) - \sum_{i_1} p^f(i_1)\log(p^f(i_1)) \end{aligned} \tag{2.16}$$

The derivative of Equation (2.16) regarding a parameter T of the transformation is [3] (remind that only p^m and p^{fm} are depending on T)

$$\begin{aligned} MI_{,T} = & \sum_{i_1,i_2} p^{f,m}_{,T}(i_1,i_2)\cdot(\log(p^{f,m}(i_1,i_2))+1) \\ & - \sum_{i_2} p^m_{,T}(i_2)\cdot(\log(p^m(i_2))+1) \\ = & \sum_{i_1,i_2} p^{f,m}_{,T}(i_1,i_2)\cdot\log(p^{f,m}(i_1,i_2)) - \sum_{i_2} p^m_{,T}(i_2)\cdot\log(p^m(i_2)) \end{aligned} \tag{2.17}$$

since $\sum_{i_1,i_2} p^{f,m}_{,T}(i_1,i_2) = 1_{,T} = 0$ (the same property can be used for the marginal moving image density). Moreover, it can be observed that

$$p^m_{,T}(i_2) = \left[\sum_{i_1} p^{f,m}(i_1,i_2)\right]_{,T} = \sum_{i_1} p^{f,m}_{,T}(i_1,i_2) \tag{2.18}$$

---

[3] a short notation of the partial derivative is used : $f_{,x}$ denotes the partial derivative of f regarding x

By injecting Equation (2.18) in Equation (2.17), the following expression of mutual information derivative is obtained

$$MI_{,T} = \sum_{i_1,i_2} p^{f,m}_{,T}(i_1,i_2) \cdot \log\left(\frac{p^{f,m}(i_1,i_2)}{p^m(i_2)}\right) \tag{2.19}$$

Equation (2.19) shows that derivative of mutual information is a weighted sum of the contributions of each bin. The derivative of the joint PDF bins can be easily obtained from Equation (2.15)

$$\begin{aligned} p^{f,m}_{\boldsymbol{u}}(i_1,i_2) = \frac{1}{N}\sum_i \beta^1 &\left(i_1 - \frac{f(\boldsymbol{x_i}) - f_{min}}{\Delta b_F}\right) \\ &\cdot \beta^3_{,i_2}\left(i_2 - \frac{m(\boldsymbol{x}_i + \boldsymbol{u}_i) - m_{min}}{\Delta b_M}\right) \\ &\nabla m(\boldsymbol{x}_i + \boldsymbol{u}_i) \cdot \frac{d\boldsymbol{u}_i}{dT} \end{aligned} \tag{2.20}$$

2.4.3 Transformation Jacobian and Metric Flow

A common characteristic, which can be observed from both Mattes and Viola's implementations, is the presence of a

$$\frac{d\boldsymbol{u}_i}{dT} \tag{2.21}$$

factor in the derivative expression

In the ITK [Ibanez and Schroeder 2005] design, this factor is called "transformation Jacobian" and the metric delegates to another object the computation of this factor to keep the design modular.

Another interesting consideration is to point the fact that with these two implementations, the mutual information can be rewritten as

$$\text{Similarity metric derivative} = \sum_{\text{Set of samples}} \textbf{flow}(\text{sample}) \cdot \frac{d\boldsymbol{u}_{\text{sample}}}{dT} \tag{2.22}$$

where the flow vector is a complex expression which can be identified from Eqs. (2.11) and (2.14) for Viola and Eqs. (2.20) and (2.19) for Mattes.

If we consider the simpler case of a mean square error metric for instance, the metric definition looks like

$$MSE = \sum_i \left(m(\boldsymbol{x_i} + \boldsymbol{u_i}) - f(\boldsymbol{x_i})\right)^2 \tag{2.23}$$

and the derivative is thus

$$MSE_{,T} = \sum_i 2 \cdot (m(\boldsymbol{x_i} + \boldsymbol{u_i}) - f(\boldsymbol{x_i})) \cdot \nabla m(\boldsymbol{x_i} + \boldsymbol{u_i}) \cdot \frac{d\boldsymbol{u}_i}{dT} \tag{2.24}$$

which brings back the $\frac{d\boldsymbol{u}_i}{dT}$ term.

By identifying this derivative with Equation (2.22), we obtain

$$\textbf{flow}(\text{sample}) = 2 \cdot (m(\boldsymbol{x_i} + \boldsymbol{u_i}) - f(\boldsymbol{x_i})) \cdot \nabla m(\boldsymbol{x_i} + \boldsymbol{u_i}) \tag{2.25}$$

which is fairly identical to the numerator of the Optical Flow as first applied to medical images registration by Thirion [1998].

The concept of flow for image similarity metric will be developed further in the next chapter. Both the flow vector of a similarity metric and its first derivative for a prior transformation model are strongly related concepts since they are related to the first variation of the metric.

2.5 Discussion about the Use of MI Compared to Other Similarity Metrics

Numerous registration measures have been proposed over the years for performing automatic registration. For an extensive list of voxel based similarity measures, the reader can refer to [Maintz and Viergever 1998], page 7. Of all these measures, mutual information has been one of the most intensively researched. Such an interest is a logical consequence of two advantages of mutual information

- Mutual information is an automatic measure, it uses only the image intensities without requiring any segmentation or landmarks. However, as seen in Chapter 4, a pre-processing step is sometimes required to correct intensity inhomogeneities.
- Mutual information is one of the few intensity based measures that is well suited to registration of multi-modal images. Unlike measures based on correlation of grey values or differences of grey values, mutual information does not assume a linear relationship between the grey values in the images.

Several independent studies have shown the suitability of mutual information as a registration measure for multi-modal medical images [Maes and Collignon 1997, Meyer et al. 1997, Viola and Wells 1996; 1995].

One of the best illustration of mutual information performances can be found in the Retrospective Registration Evaluation Project (RREP), an international study comparing the accuracy of sixteen registration methods against a screw marker gold standard [West et al. 1997].

Despite these many successes, some applications have been reported where mutual information performs poorly [Penney et al. 1999, Studholme et al. 1999, Thevenaz and Unser 1997]. In these cases, the mutual information registration function is often ill-defined and contains a lot of local maxima. This can occur for instance when the images are of low resolution, or when the region of overlap is small.

To improve the behavior of mutual information in these problematic cases, several methods have been suggested, such as multi-resolution methods [Thevenaz and Unser 1997], using higher-order mutual information [Rueckert et al. 2000] or by including spatial information into the registration measure [Pluim et al. 2000].

Another alternative is the use of normalized mutual information as an overlap invariant measure [Studholme et al. 1999]. Butz [Butz et al. 2002, Butz and Thiran 2002] has shown that normalized mutual information makes possible to select features automatically during the registration process when mutual information always chooses for features of maximal entropy.

Maximizing Mutual Information for Free-form Transformations : a Variational Approach

3

The mathematical developments in this chapter are mainly taken from [Hermosillo et al. 2001]. The reader should refer to this reference for a complete description of the variational framework applied to multi-modal images matching. It will be shown in Chapter 5 how this framework is used in our probabilistic atlas estimation procedure.

3.1 Variational Principles for Similarity Measures

If no assumption is made regarding the nature of the displacement field to recover, the registration problem can be seen as the optimization of a cost functional $\mathcal{I}[\boldsymbol{u}]$ where $\boldsymbol{u}$ is the 2D or 3D displacement field.

The principles behind a variational analysis is to look for a displacement field $\boldsymbol{u}$ making the cost functional *stationary* i.e. a displacement field vanishing the first variation of the functional

$$\left.\frac{\partial \mathcal{I}(\boldsymbol{u} + \epsilon \boldsymbol{h})}{\partial \epsilon}\right|_{\epsilon=0} = 0, \quad \forall \boldsymbol{h} \tag{3.1}$$

where $\boldsymbol{h}$ is an arbitrary function belonging to the same space of continuous and differentiable functions as $\boldsymbol{u}$. A common requirement for $\boldsymbol{h}$ is to restrict its space to functions equal to 0 on the border of the spatial domain Ω. In Equation (3.1), ϵ is a scalar factor multiplying the perturbation function $\boldsymbol{h}$. Figure 3.1 illustrates the definition of $\boldsymbol{u}$, $\boldsymbol{h}$ and ϵ for a 1D domain on the $(0, 1)$ interval.

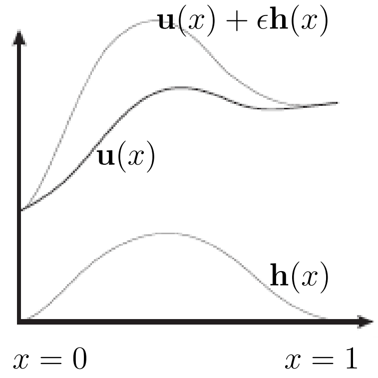


Figure 3.1: *Stationarity of a cost functional.*
A perturbation function $\boldsymbol{h}$ is added to the current displacement field $\boldsymbol{u}$ with a scaling parameter ϵ. The purpose is to solve for the displacement field $\boldsymbol{u}$ such that the derivative of the cost functional regarding ϵ tends to 0 for $\epsilon \to 0$.

Generally speaking, the cost functional $\mathcal{I}$, is made of two terms, a matching measure and a regularization term. The regularization term is usually depending on the displacement field spatial derivatives $\frac{\partial u_i}{\partial x_j}$ and constrains the spatial variations. If we define the Jacobian matrix of the displacement field

$$(D\boldsymbol{u}(x))_{ij} = \frac{\partial u_i}{\partial x_j} \tag{3.2}$$

a general form for the cost functional is

$$\begin{aligned}\mathcal{I}(\boldsymbol{u}) &= \mathcal{J}(\boldsymbol{u}) + \mathcal{R}(D\boldsymbol{u}) &(3.3)\\ &= \int_\Omega \phi(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}), D\boldsymbol{u}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} &(3.4)\end{aligned}$$

For computing the derivative in Equation (3.1), a development of $\mathcal{I}(\boldsymbol{u}+\epsilon\boldsymbol{h})$

is used [1]

$$\begin{aligned}
\left.\frac{\partial \mathcal{I}(\boldsymbol{u}+\epsilon \boldsymbol{h})}{\partial \epsilon}\right|_{\epsilon=0} &= \int_\Omega \mathrm{d}\boldsymbol{x}(\phi_{,\boldsymbol{u}}\boldsymbol{h} + \phi_{,\boldsymbol{u}_{,x}}\boldsymbol{h}_{,x} + \phi_{,\boldsymbol{u}_{,y}}\boldsymbol{h}_{,y} + \phi_{,\boldsymbol{u}_{,z}}\boldsymbol{h}_{,z}) && (3.5)\\
&= \int_\Omega \mathrm{d}\boldsymbol{x}(\phi_{,\boldsymbol{u}} - \frac{\partial \phi_{,\boldsymbol{u}_{,x}}}{\partial x} - \frac{\partial \phi_{,\boldsymbol{u}_{,y}}}{\partial y} - \frac{\partial \phi_{,\boldsymbol{u}_{,z}}}{\partial z})\boldsymbol{h}\\
&+ \underbrace{\int_\Omega \mathrm{d}\boldsymbol{x}\left(\frac{\partial(\phi_{,\boldsymbol{u}_{,x}}\boldsymbol{h})}{\partial x} + \frac{\partial(\phi_{,\boldsymbol{u}_{,y}}\boldsymbol{h})}{\partial y} + \frac{\partial(\phi_{,\boldsymbol{u}_{,z}}\boldsymbol{h})}{\partial z}\right)}_{(*)} && (3.6)
\end{aligned}$$

The transition from Equation (3.5) to Equation (3.6) is obtained after integration by parts. The second term $(*)$ of Equation (3.6) can be rewritten using the divergence theorem

$$(*) = \int_{\partial\Omega} \mathrm{d}s\, \boldsymbol{h}\left(\phi_{,\boldsymbol{u}_{,x}}\boldsymbol{n}_x + \phi_{,\boldsymbol{u}_{,y}}\boldsymbol{n}_y + \phi_{,\boldsymbol{u}_{,z}}\boldsymbol{n}_z\right) = 0 \quad \text{since}\, \boldsymbol{h} = 0\, \text{on}\, \partial\Omega \tag{3.7}$$

Since the second term of Equation (3.6) is equal to zero, Equation (3.1) is equivalent in our case to

$$\left.\frac{\partial \mathcal{I}(\boldsymbol{u}+\epsilon \boldsymbol{h})}{\partial \epsilon}\right|_{\epsilon=0} = \int_\Omega \mathrm{d}\boldsymbol{x}\, \boldsymbol{h}(\phi_{,u} - \mathbf{div}(\phi_{,Du})) = 0 \quad \forall \boldsymbol{h} \tag{3.8}$$

Equation (3.8) is equivalent to a scalar product

$$< \boldsymbol{h}, \phi_{,u} - \mathbf{div}(\phi_{,Du}) >= 0 \quad \forall \boldsymbol{h} \tag{3.9}$$

Since Equation (3.9) must hold $\forall h$, the displacement field u making the cost functional stationary must satisfy the following condition on Ω

$$\nabla_{\boldsymbol{u}}\mathcal{I} \overset{\text{def}}{=} \underbrace{\phi_{,u}}_{\text{matching term}} - \underbrace{\mathbf{div}(\phi_{,Du})}_{\text{regularization term}} = 0 \quad \forall \boldsymbol{x} \in \Omega \tag{3.10}$$

The most common way of solving Equation (3.10) is a gradient descent scheme computing at each iteration an incremental displacement field that is then added to the current displacement field : $u_{t+1} = u_t + \Delta u_t$ with

$$\Delta u_t = -\eta \cdot \nabla_u \mathcal{I} \tag{3.11}$$

where η is a learning rate also referred to as time step in the ITK [Ibanez and Schroeder 2005] implementation.

---

[1] As in the previous chapter, a short notation of the partial derivative is used : $f_{,x}$ denotes the partial derivative of f regarding x

3.2 Application to Mutual Information

In this section, we apply the derivation scheme explained for the general case of a cost functional in Section 3.1 to the particular case of the mutual information metric.

As seen in the previous chapter, mutual information belongs to an important family of histogram based similarity metrics. These metrics rely on the computation of a joint density probability of gray levels in the pair of images for the current alignment. For a given transformation, each voxel of the fixed image domain yields a pair of intensities in the fixed and the moving image

$$(f(\boldsymbol{x}), m(T(\boldsymbol{x}))) \tag{3.12}$$

where f (resp. m) is the signal intensity function in the fixed (resp. moving) image. All these pairs of joint intensities enable the computation of a joint probability density (either by Parzen windowing [Viola and Wells 1995]or by histogram binning [Maes and Collignon 1997]).

If $p^{f,m}$ denotes the joint probability density between the fixed and the moving image and p^f (p^m) denotes the marginal density of the fixed (moving) image signal intensities, mutual information is computed for a displacement field $u(x)$ following

$$\begin{aligned} MI_u &= -H(I_1, I_2) + H(I_1) + H(I_2) && (3.13) \\ &= -\sum_{i_1} p^f(i_1)\log(p^f(i_1)) - \sum_{i_2} p_u^m(i_2)\log(p_u^m(i_2)) \\ &\quad + \sum_{i_1,i_2} p_u^{f,m}(i_1,i_2)\log(p_u^{f,m}(i_1,i_2)) && (3.14) \end{aligned}$$

A variational flow of this metric is obtained as described in Section 3.1 by adding a continuous perturbation $\epsilon\boldsymbol{h}(\boldsymbol{x})$ to the current displacement field $\boldsymbol{u}(\boldsymbol{x})$. The metric derivative regarding the ϵ scalar parameter is then computed (Equation (3.1)) for $\epsilon = 0$:

$$\left.\frac{\partial MI_{\boldsymbol{u}+\epsilon\boldsymbol{h}}}{\partial\epsilon}\right|_{\epsilon=0} = \sum_{i_1,i_2} \left.\frac{\partial p^{f,m}_{\boldsymbol{u}+\epsilon\boldsymbol{h}}(i_1,i_2)}{\epsilon}\right|_{\epsilon=0} \cdot \log\left(\frac{p^{f,m}_{\boldsymbol{u}+\epsilon\boldsymbol{h}}(i_1,i_2)}{p^{m}_{\boldsymbol{u}+\epsilon\boldsymbol{h}}(i_2)}\right) \tag{3.15}$$

where the i_1 and i_2 indices discretize the range of gray levels in the fixed and moving image into equal bins. The joint probability distribution must be estimated using continuous interpolation functions in order to get a well defined derivative. Possible interpolation schemes are trilinear [Fransens et al. 2004] or B-Spline [Mattes et al. 2003].

Mattes's implementation (see Chapter 2, Section 2.4.2) uses a separable kernel obtained by composing a "box car" function in the fixed image dimension and a 3rd order B-Spline in the moving image dimension (the joint PDF derivative needs only to be computed in this dimension). If we replace the discrete sum in Equation (2.15) by a continuous integral, this equation can be rewritten as

$$p_{\boldsymbol{u}}^{f,m}(i_1,i_2) = \frac{1}{|\Omega|}\int_{\Omega} \beta^1(f(\boldsymbol{x})-i_1)\beta^3(m(\boldsymbol{x}+\boldsymbol{u})-i_2)\mathrm{d}\boldsymbol{x} \tag{3.16}$$

Injecting the definition of p^{fm} (Eq.3.16) in Eq.3.15 yields

$$\left.\frac{\partial MI_{\boldsymbol{u}+\epsilon\boldsymbol{h}}}{\partial\epsilon}\right|_{\epsilon=0} = \int_{\Omega} \kappa_{\boldsymbol{u}}^{f,m}(\boldsymbol{x})\nabla m(\boldsymbol{x}+\boldsymbol{u})\boldsymbol{h}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \overset{\mathrm{def}}{=} < \nabla_{\boldsymbol{u}} MI, \boldsymbol{h}(\boldsymbol{x}) > \tag{3.17}$$

where $\kappa_{\boldsymbol{u}}^{f,m}$ is defined by

$$\kappa_{\boldsymbol{u}}^{f,m}(\boldsymbol{x}) = \frac{1}{|\Omega|}\sum_{i_1,i_2} \log\left(\frac{p_{\boldsymbol{u}}^{f,m}(i_1,i_2)}{p_{\boldsymbol{u}}^{m}(i_2)}\right)\beta^1(f(\boldsymbol{x})-i_1)\beta^{3'}(m(\boldsymbol{x}+\boldsymbol{u})-i_2) \tag{3.18}$$

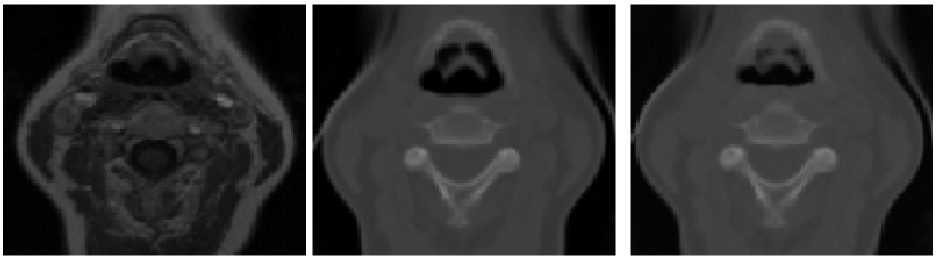
Equation 3.17 shows that the derivative regarding the ϵ parameter is the scalar product between two functions : $\boldsymbol{h}(\boldsymbol{x})$ (the perturbation function) and the MI metric flow defined by

$$\nabla_{\boldsymbol{u}} MI = \kappa_{\boldsymbol{u}}^{f,m}(\boldsymbol{x})\nabla m(\boldsymbol{x}+\boldsymbol{u}) \tag{3.19}$$

3.3 Application : MR-CT Registration using MI Flow

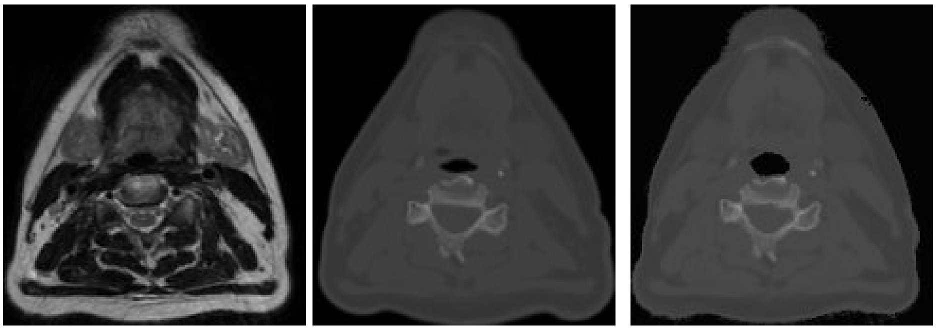
The data set for this experiment is composed of two patients. For each patient, a CT scan (size of 512x512x70) and a MR image (size of 256x256x50) were acquired. The patients moved between both acquisitions which results in a bending of the spine. The transformation between both images is recovered in a first step using an articulated transformation model (this algorithm has been developed by du Bois d'Aische et al. [2005], du Bois d'Aische et al. [2005]). In a second time, our non-rigid MI flow registration algorithm has been applied to recover the non-rigid deformations induced by the spine displacement. The results are shown in Figure 3.2 for the first patient and in figure Figure 3.3 for the second patient after articulated and free-form alignment.

The MR-CT results have been validated by an oncologist. Landmarks have been selected in the images. The landmarks have been placed everywhere in the neck region, at muscle crossings, in the vessels and near the spinal column. The mean error after the articulated registration decreases from 3.89 to 2.22 mm and the maximum error decreases from 7.9 to 4.1 mm. After the mutual information based flow, the mean error decreases to 1.61 mm and the maximum error decreases to 3.5 mm.



(A) (B) (C)

Figure 3.2: *MR-CT articulated and free-form registration for patient 1.*
Image A presents the target MR image, image B presents the result of the articulated registration of the CT to the MR, and image C presents the result of the mutual information flow-based registration



(A) (B) (C)

Figure 3.3: *MR-CT articulated and free-form registration for patient 2.*
Image A presents the target MR image, image B presents the result of the articulated registration of the CT to the MR, and image C presents the result of the mutual information flow-based registration

3.4 Application : Atlas-Based Segmentation of Pathological Brain MR Images by Combining a Mutual Information Flow and a Radial Lesion Growth Model

3.4.1 Introduction

Various approaches have been proposed in the literature for matching pathological brains with anatomical atlases. Kyriacou and Davatzikos [1999] presented a bio-mechanical based model simulating a contraction of the tumor region. A "normal" patient is registered on the atlas. A tumor growing algorithm is then applied to the registered atlas. This phase is driven by anatomical features (segmented structures) to estimate the localization of the tumor and the level of strain. An opposite strategy for atlas to pathological brain registration has been introduced by Dawant et al. [1999]. In their approach a synthetic lesion (seed) is introduced in the atlas. Thiron's [Thirion 1998] demons algorithm is then run on the seeded atlas. However, to obtain a precise tumor growth, the seed can not be too small compared to the tumor size. This requires to mask important anatomical information in the atlas which tends to produce segmentation errors in case of big tumors. For this reason, Bach Cuadra et al. [2004] suggested to introduce an hybrid non-registration model splitting the image into two regions. Inside the tumor area, a model of lesion growth is applied which assumes a radial growth from a single voxel seed until tumor contour. Outside the tumor area, demons algorithm is applied and a total regular deformation field is ensured by using an adaptive Gaussian filter. Recently, two approaches [Duay et al. 2004, Stefanescu et al. 2004] solved the problem of atlas to pathological brain registration by locally adapting different elasticities of the transformation rather than modelling the deformation induced by the tumor.

Optical flow has been widely used in the context of atlas-based segmentation. This technique estimates iteratively a dense deformation field at each point of the fixed image. The classical optical flow formulation [Thirion 1998] relies on the assumption that the intensity distributions are identical in the fixed and moving images. Even if both images are from the same modality (MR in our case), a contrast product is often used in clinical practise to better appreciate the anatomical limits of tumors. Since statistical similarity measures are more robust to variations of contrast in both images, variational approaches estimating the flow of such metrics have been investigated. Hermosillo has introduced in [Hermosillo et al. 2001] a mathematical framework for computing a variational flow from various similarity metrics (mean square error, mutual information, correlation ratio). The flow computed from global similarity measures incorporates local information (moving image gradient at this point) as well as global information (marginal and joint probability distributions of signal intensities). Our implementation differs from Hermosillo's by the use of B-Spline kernel functions (instead of Parzen windowing) to estimate the joint probability distribution [Mattes et al. 2003]. A continuous estimate of the the joint probability distribution is a critical point for computing infinitesimal variations of the metric.

The approach we present here is based on the works of [Bach Cuadra et al. 2004]. However, an important change is introduced trying to overcome their main limitation. Their approach presented excessive deformability of some areas such as *meninges* and *sinuses*. The assumption of constant intensity (needed to be fulfilled by demons algorithm) between image structures is violated in these areas due to the presence of a contrast agent in the patient images. Instead of using a demons algorithm which is based on the least squares minimization criteria, an optical flow method that minimizes the mutual information is proposed. That makes the deformation more robust in regions where a contrast product is present.

3.4.2 Algorithm Description

Mutual Information Flow Registration for Normal Subjects

The number of incremental estimations of the 3D dense displacement field has to be fixed by the user. At each iteration, the joint histogram is estimated by random sampling through the image domain. The mutual information flow is then computed at each voxel using Eq.3.19.

The implementation of the mutual information flow has been integrated in the Insight Toolkit[Ibanez and Schroeder 2005] as a subclass of the `PDEDeformableRegistrationFunction` class.

Atlas to Pathological Brain Non-rigid Registration

Our approach to brain atlas deformation in the presence of space occupying tumors is based on Bach's MLG algorithm [Bach Cuadra et al. 2004] but differs from the fact that a flow optimizing mutual information is used instead of the demons algorithm. The overall process can be summarized in 4 steps:

1. An affine transformation [Cuisenaire et al. 1996] is applied to the atlas image in order to globally match the patient.
2. The lesion is segmented using the Adaptive Template Moderated Spatially Varying Statistical Classification (ATM SVC) algorithm [Warfield et al. 2000].
3. The atlas is manually seeded by an expert with a single-voxel placed on the estimated origin of the patient's lesion.
4. The non-linear registration based on a mutual information flow and a MLG tumor growth model is performed in order to deform the seeded atlas on the target patient.

The non-rigid deformation method we propose distinguishes between those two areas fixed from the lesion segmentation. Outside the lesion, the *optical flow* force as defined in Eq. 3.19 is applied. Inside the lesion, the tumor growth model assumes a *radial* growth of the tumor from the tumor seed.

The displacement vector computed at every voxel using either the MI flow force or the tumor growth model is regularized by an adaptive Gaussian filter to avoid discontinuities around the lesion borders.

Three areas are considered in the smoothing process: inside the lesion area, close to the lesion (within 10 mm of the tumor) where large deformations occur, and the rest of the brain. Smoothing is not necessary inside the lesion because the vector field induced by the model of lesion growth is highly regular and the continuity is ensured. So, $\sigma = 0$[2] inside the lesion area. In the region close to the tumor (including the tumor contour) there are large deformations due to the tumor growth. Then, it is necessary to allow large elasticity, i.e. σ should have a small value, typically 0.5 mm. In the rest of the brain, deformations are smaller, due primarily to inter-patient anatomical variability. So, a larger σ proves to be better, as it simulates a more rigid transformation.

---

[2] where σ^2 is the variance of the gaussian filter

3.4.3 Results

Data Set

The patient images have been retrieved from the Surgical Planning Laboratory (SPL) of the Harvard Medical School & NSG Brain Tumor Database[3]. They consist in volumes of 128 coronal slices of 256×256 pixels and $0.9375 \times 0.9375 \times 1.5\ mm^3$ of voxel size and all of them have a meningioma. No brain edema was observed on the data set. Notice that all the patient images have been acquired using a contrast agent. The digital atlas used in this work also comes from the SPL [Kikinis, R. 1996]. It is made of a large number of anatomical structures manually drawn on the MR of a single normal subject scanned in 160 coronal slices with $0.9375 \times 0.9375 \times 1.5 mm^3$ voxel size.

Segmentation of Brain Structures

The SPL digital atlas MR image has been deformed on 3 patients using the algorithm described in Sec 3.4.2.

Fig. 3.5 and 3.6 show the atlas contours of the ventricles, the central nuclei and the thalamus projected after transformation on the patient images. For segmenting these structures, no significant difference has been observed between optical and mutual information flow. However, the two flows behave differently in the presence of different contrasts in both images. For instance, around the mid-sagittal plane, the classical optical flow is disturbed by a high difference in signal intensities (the mid-sagittal plane is bright in the patient image and dark in the atlas). This is illustrated on Fig.3.4: Fig.3.4(a) shows a coronal slice of the target patient image, the deformed atlas image is shown in (b) using optical flow and in (c) using MI flow. The dotted line shows that an important distortion appears around the mid-sagittal plane when using the optical flow deformation.

---

[3] http://spl.bwh.harvard.edu:8000/˜warfield/tumorbase/

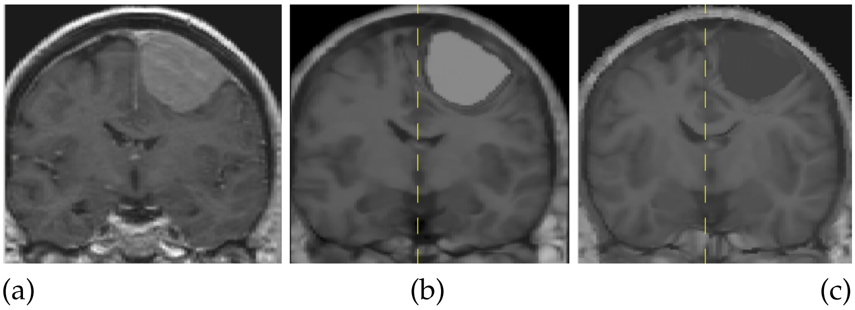

(a) (b) (c)

Figure 3.4: *Mutual information and optical flows comparison.*
The patient target image is shown in (a). Sub-figure (b)and (c) show registration results using optical (b) and MI(c) flow. Mutual information is more robust to the contrast difference along the doted line than the Demon's algorithm.

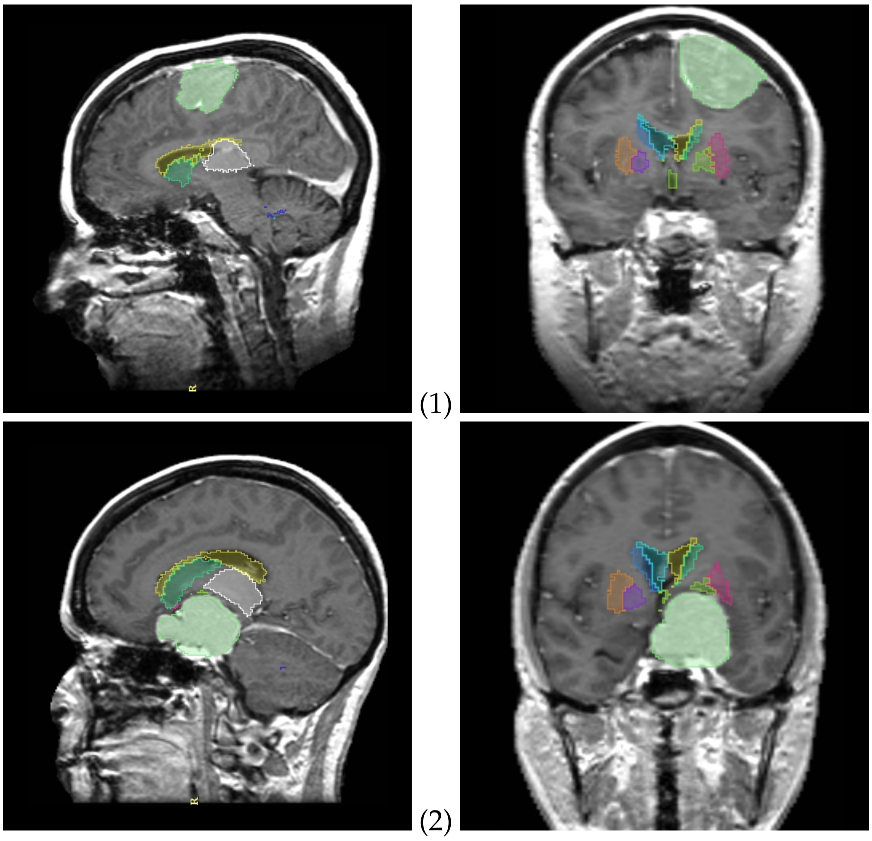
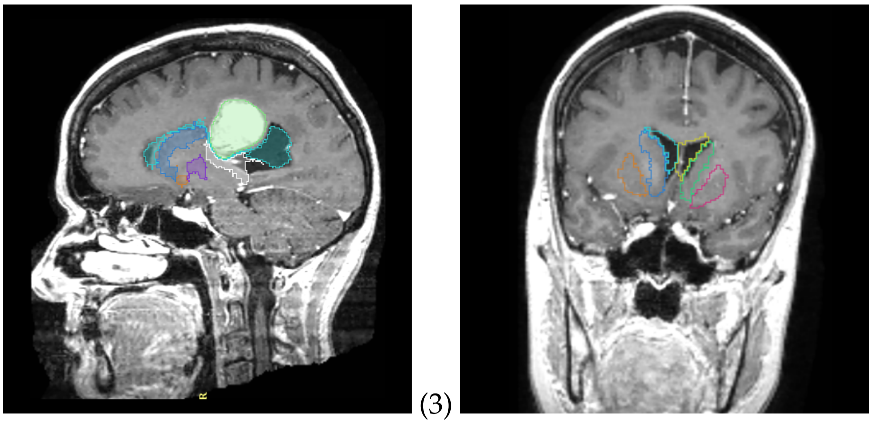
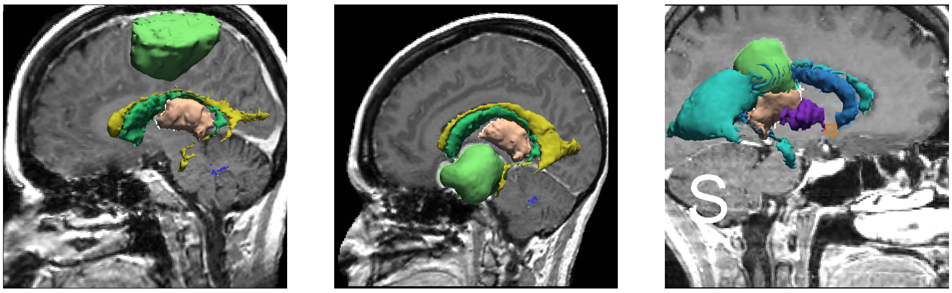


Figure 3.5: *Segmentation results obtained after alignment of the atlas on 3 patients with large tumors(each row show a different patient, last patient is shown in Figure 3.6). Contours of the tumor, ventricles, thalamus and central nuclei are overlayed on the patient image. A radial tumor growth deformation model is used inside the lesion. Outside the lesion, a mutual information variational flow is used to compensate inter-subject variations.*



(3)

Figure 3.6: *Segmentation results obtained after alignment of the atlas on 3 patients with large tumors, patient 3.*



Case 1 Case 2 Case 3

Figure 3.7: *3D rendering of tumor, ventricles, thalamus and central nuclei after atlas alignment on 3 patients with large tumors*

Maximizing Mutual Information for Prior Transformation Models

4

This chapter introduces first the classical registration pipe-line when using prior transformation models. The rigid and non-rigid transformation models we used in our work are then presented. A short description of the optimization methods we have used in our experiments is described, with a particular stress on the SPSA optimization method. In the second part of this chapter, we present different clinical applications using the concepts described in the first part.

4.1 Registration Pipeline

A common representation of automatic registration algorithms for prior transformation models distinguish three main components in the design : a transformation model representing the prior knowledge on the transformation to be found, a metric measuring the degree of alignment between the fixed and the moving image for a given transformation and an optimizer which seeks for the best transformations parameters (i.e. the parameters yielding the best metric value). These concepts are illustrated in Figure 4.1 [Ibanez and Schroeder 2005]. The applications presented in this chapter (Sections 4.4 and further) all use mutual information as similarity metric (Chapter 2). The next two sections describe the transformation models and optimizers used in the applicative part of this chapter.

4.2 Transformation Models

4.2.1 Rigid and Affine Transformation Models

A classical family of geometrical transformations is the affine transformation model which represents translations, rotations, scaling and skew.

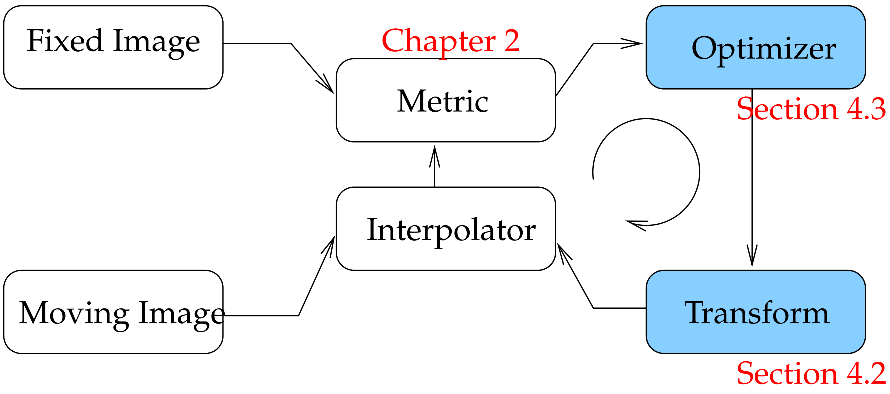


Figure 4.1: *Registration pipeline.The registration pipeline is made of three main components : the metric, the transformation model and the optimizer. The transformation takes coordinates of points in the fixed image domain and map them in the moving image domain. To measure the grey level at these mapped coordinates, an interpolator is required. The optimizer iteratively solves for the best transformation parameters optimizing the metric.*

It encompasses rigid transformations which are made up from chaining translations and rotations transformations.

Using the affine transformation model a point of coordinates x_0, x_1, x_2 is mapped to a point y_0, y_1, y_2 following

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} t_0 \\ t_1 \\ t_2 \end{pmatrix} \tag{4.1}$$

This transformation can be broken down to

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = S \cdot R_X \cdot R_Y \cdot R_Z \cdot C \cdot \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} t_0 \\ t_1 \\ t_2 \end{pmatrix} \tag{4.2}$$

In Equation (4.2), S stands for a scaling matrix, R_X, R_Y and R_Z are rotations matrices around the X, Y and Z axes and C is a skew matrix.

$$S = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{pmatrix} \quad ,$$

$$R_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(r_x) & -sin(r_x) \\ 0 & \sin(r_x) & \cos(r_x) \end{pmatrix} \quad ,$$

$$R_Y = \begin{pmatrix} \cos(r_y) & 0 & \sin(r_y) \\ 0 & 1 & 0 \\ -\sin(r_y) & 0 & \cos(r_y) \end{pmatrix} \quad ,$$

$$R_Z = \begin{pmatrix} \cos(r_z) & -\sin(r_z) & 0 \\ \sin(r_z) & \cos(r_z) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad ,$$

$$R_Z = \begin{pmatrix} 1 & \tan(c_x) & \tan(c_y) \\ 0 & 1 & \tan(c_z) \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

The affine transformations is therefore characterized by the following vector of parameters

$$\vec{\alpha} = (s_x, s_y, s_z, r_x, r_y, r_z, c_x, c_y, c_z, t_x, t_x, t_z)^t \tag{4.3}$$

Usually, t_x, t_y et t_z are expressed in millimeters and c_x, c_y et c_z, r_x, r_y and r_z in degrees. The scaling parameters s_x, s_y et s_z are unitless.

For the optimization of the transformations parameters, the s_i parameters are often replaced by

$$s_i' = 100 \cdot \ln(s_i)$$

This way, the search domain of s_i cover positive and negative numbers in the same manner as rotation and translation parameters. More complex representations of affine transformations using quaternions are described in the Insight software guide [Ibanez and Schroeder 2005].

4.2.2 Finite Elements as a Non-Rigid Transformation Model

To represent the space of allowed deformations, a set of basis functions need to be defined. The amplitude assigned to each basis function is found by the iterative optimization method described in Section 4.3.2. For local deformations, each basis function is non-null on a limited fraction of the image domain. A common way of modeling the deformations of an elastic body is to split the domain of interest in small elements assuming the deformation inside each element interpolates the node displacements of this element. This interpolation is defined by *shape functions* associated to each node. Using this quantization scheme, different mechanical properties can be assigned to different elements.

Numerous techniques have been proposed for designing elements which follow accurately the borders of the different structures. In our laboratory (Ferrant et al. [2002; 2001], du Bois d'Aische et al. [2004]), some research has been undertaken to design meshing strategies providing faithful surfaces representation and well shaped elements. Mutli-resolution meshing strategies generally allow to find a compromise between these two conflicting requirements.

Shape Functions

The decomposition of the displacement field as a weighted sum of basis functions is a flexible manner of representing arbitrary deformation fields with easily controlled complexity and local fidelity. A broad class of basis functions have been used (Radial Basis Function [Ruekert et al. 1999], B-splines [Kybic and Unser 2000],...). In the finite element method, the domain of interest is divided into a mesh of elements, and the displacement field u is estimated inside each element by

$$u_l(x) = \sum_{n \in \text{Nodes}} u_l^n N_{el}^n(x) \tag{4.4}$$

where u_l^n is the l^{th} component of the displacement for the n^{th} node. N_{el}^n represents the shape function associated with the n^{th} node. Shape functions are non-null in the element volume only. The shape functions of a node tends to zero around the other node of the same element. In the case of tetrahedral meshes, the shape functions are often chosen as linear functions. Ranking the node displacements in a vector $\boldsymbol{U}$ and the shape functions in a vector $\boldsymbol{N}_{el}(x)$, Equation (4.4) can be compacted as [Zienkewickz and Taylor 1987]

$$u(x) = \boldsymbol{N}_{el}(x) \cdot \boldsymbol{U} \tag{4.5}$$

Stress and Strain

Stress can be defined as the intensity of internal force on an infinitesimal surface element at each point of the volume of interest. Two types of stress can be distinguished :

- Normal (or extensional): act normal to the plane of an infinitesimal surface element

- Shear: act in-plane of this element.

Considering surface orientations with normals along x, y and z lead to the representation of this physical property by a second order symmetric tensor. For the convenience of notations, the non-redundant values of the stress tensor will be ranked in a vector denoted by $\boldsymbol{\sigma}$.

Strain can be defined as the percentage deformation of an infinitesimal surface element. Two types of strain can be distinguished :

- Normal (or extensional) : local elongation of the material at this point
- Shear: angular change between two orthogonal basis vectors of the infinitesimal surface element.

Considering x, y and z orientations of the surface normal lead to the representation of strain by a second order symmetric tensor. As for stress, the non-redundant values are ranked in a vector denoted by $\boldsymbol{\epsilon}$.

$$\boldsymbol{\epsilon} = \left(\frac{\partial u_0}{\partial x_0}, \frac{\partial u_1}{\partial x_1}, \frac{\partial u_2}{\partial x_2}, \frac{\partial u_0}{\partial x_0} + \frac{\partial u_1}{\partial x_1}, \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2}, \frac{\partial u_0}{\partial x_0} + \frac{\partial u_2}{\partial x_2}\right)^t = \boldsymbol{L} \cdot u(x) \tag{4.6}$$

where $\boldsymbol{L}$ is a differential operator defined by

$$\boldsymbol{L} = \begin{pmatrix} \partial/\partial x_0 & 0 & 0 \\ 0 & \partial/\partial x_1 & 0 \\ 0 & 0 & \partial/\partial x_2 \\ \partial/\partial x_0 & \partial/\partial x_1 & 0 \\ 0 & \partial/\partial x_1 & \partial/\partial x_2 \\ \partial/\partial x_0 & 0 & \partial/\partial x_2 \end{pmatrix} \tag{4.7}$$

Since $u(x)$ can be expressed as $\boldsymbol{N}_{el}(x) \cdot \boldsymbol{U}$ (Equation (4.5)), the differential $\boldsymbol{L}$ operator is actually applied to the shape functions in Equation (4.6) :

$$\boldsymbol{\epsilon} = \boldsymbol{L}\boldsymbol{N}_{el}\boldsymbol{U} \triangleq \boldsymbol{B}_{el}\boldsymbol{U} \tag{4.8}$$

The link between $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ depends on the material considered. In the specific case of linear elastic body, the relation between $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ becomes linear :

$$\boldsymbol{\sigma} = (\sigma_x \, \sigma_y \, \sigma_z \, \tau_{xy} \, \tau_{xz} \, \tau_{yz})^t \tag{4.9}$$

$$= \boldsymbol{D} \cdot \boldsymbol{\epsilon} \tag{4.10}$$

In the case of an isotropic material, $\boldsymbol{D}$ can be written as

$$\boldsymbol{D} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1 & \frac{\nu}{1-\nu} & \frac{\nu}{1-\nu} & 0 & 0 & 0 \\ \cdot & 1 & \frac{\nu}{1-\nu} & 0 & 0 & 0 \\ \cdot & \cdot & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \frac{1-2\nu}{2(1-\nu)} & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \frac{1-2\nu}{2(1-\nu)} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1-2\nu}{2(1-\nu)} \end{pmatrix} \tag{4.11}$$

where E is the Young's modulus

$$E_x = \frac{\partial \sigma_x}{\partial \boldsymbol{\epsilon}_x} \tag{4.12}$$

and ν the Poisson's ratio

$$\nu_{xy} = -\frac{\boldsymbol{\epsilon}_y}{\boldsymbol{\epsilon}_x} \tag{4.13}$$

In the isotropic case, $\nu_{xy} = \nu_{yz} = \nu_{xz}$ and $E_x = E_y = E_z$.

Regularization Metric : Linear Elastic Energy

The density of elastic energy stored at a given of a body under stress can be expressed as $\boldsymbol{\sigma}^t \cdot \boldsymbol{\epsilon}$. This density has to be integrated in the whole volume of interest

$$E = \int_X \boldsymbol{\sigma}^t \boldsymbol{\epsilon} \, \mathrm{d}\Omega \tag{4.14}$$

Decomposing the integral in (4.14) as a sum over all elements enables to rewrite the linear elastic energy as

$$E = \sum_{el} E_{el} = \sum_{el} \boldsymbol{U}^t \cdot \boldsymbol{K}^{el} \boldsymbol{U} \tag{4.15}$$

where the $(i,j)^{th}$ element of $\boldsymbol{K}^{el}$ is [Zienkewickz and Taylor 1987]

$$\boldsymbol{K}^{el}_{ij} = \int_{\Omega_{el}} \boldsymbol{B}^{el^t}_i \boldsymbol{D} \boldsymbol{B}^{el}_j \, \mathrm{d}\Omega_{el} \tag{4.16}$$

$\boldsymbol{K}^{el}$ is the *local stiffness matrix* associated to the element el. A global stiffness matrix $\boldsymbol{K}$ can be defined by summing for each pair of nodes (i,j) the $\boldsymbol{K}^{el}_{ij}$ entries for all elements adjacent to this node.

Using this definition, the elastic energy can be rewritten as

$$E = \frac{1}{2} \boldsymbol{U}^t \cdot \boldsymbol{K} \cdot \boldsymbol{U} \tag{4.17}$$

As the elastic energy is conservative, the force derivating of this energy is equal to $\boldsymbol{KU}$. This gives an intuitive explanation of the elastic energy : if a node is moved in a given direction, the neighboring nodes will be moved according to the column coefficients of the stiffness matrix. The elastic energy therefore enforce each node to move consistently with its neighbors. This "smooth" behavior is a crucial requirement for estimating intra-subject deformations (e.g. induced by differences in the patient position between two acquisitions).

Regularization Strategy

To find the non-rigid spatial correspondence between a fixed and a moving image, a classical strategy (see for example [Bajcsy and Kovačič 1989, Dengler and Schmidt 1988, Gee 1999]) in image processing is the minimization of a functional containing a similarity term S and a regularization term R (a weight α is defined between S and R):

$$\boldsymbol{U}_{opt} = \arg\max_{\boldsymbol{U}} S(F, M_{\boldsymbol{U}}) + \alpha \cdot R(\boldsymbol{U}) \tag{4.18}$$

In Equation 4.18, $\boldsymbol{U}$ is the vector of parameters modeling the displacement field from the fixed to the moving image, F and $M_{\boldsymbol{U}}$ are the features extracted from the fixed and moving images (grey levels in our case) for the current value of $\boldsymbol{U}$ parameters.

In this chapter, mutual information is used as similarity metric and the linear elastic energy acts as regularization term. Equation (4.18) can thus be rewritten as

$$\boldsymbol{U}_{opt} = \arg\max_{\boldsymbol{U}} MI(\boldsymbol{U}) + \alpha/2 \cdot \boldsymbol{U}^t \boldsymbol{KU} \tag{4.19}$$

Mesh Generation

The splitting of the volume of interest in elements following the geometry of the objects has been a broad field of research for years. In this work, the image domain was split into cubic elements. Each cell of this lattice is then divided into tetrahedral elements.

The first scheme envisaged is described in Equation (4.2). Each cube of a regular grid is split in a five tetrahedra pattern (two pattern are alternated).

Interleaving of two cubic lattices (Body Centered Cubic lattice - BCC) has been described by Molino [Molino et al. 2003]. Molino describes this

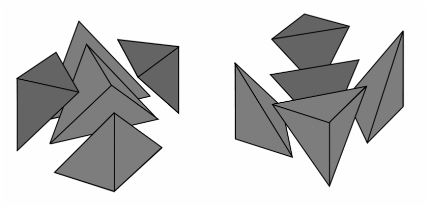
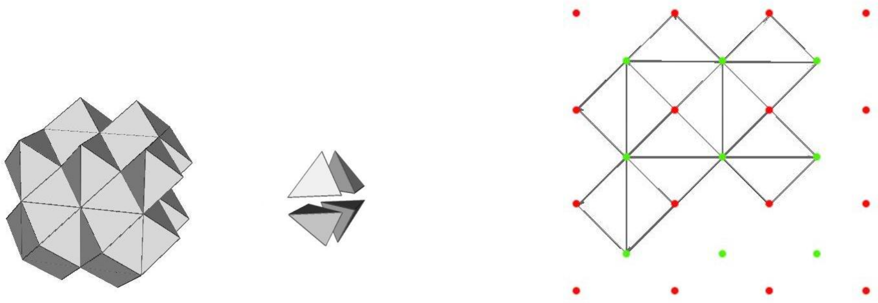


Figure 4.2: *Tetrahedral mesh generation from a regular cube lattice*

lattice as initialization of a multi-resolution mesh generation strategy [du Bois d'Aische et al. 2004, Molino et al. 2003]. BCC gives a more dense lattice of nodes than a regular grid. The tetrahedrons generated from such lattice tend to be more regular than the tetrahedrons obtained from Figure 4.2 splitting. Figure 4.3 shows the generation of the tetrahedral mesh from the BCC lattice.



(a)

3D rendering of two interlaced cubic grids splitting into tetrahedra

(b)

2D view of dividing a BCC lattice in tetrahedra

Figure 4.3: *Starting from a lattice of two interlaced cubic grids, a regular mesh of tetrahedrons can be generated.*

4.3 Optimization Strategies

4.3.1 Gradient Descent

When the metric implements a first order derivative as seen in Chapter 2 for the implementations of Viola and Wells [1996; 1995] and Mattes et al. [2003], a first optimization strategy is simply to follow the gradient direction for climbing the nearest local optimum of the cost function.

Similarity Metric Gradient As previously seen in Section 2.4.3, the transformation Jacobian has to be passed to the metric object for being inserted in the general metric derivatives (Eqs. (2.11) and (2.14) for Viola and Eqs. (2.20) and (2.19) for Mattes).

In the case of a finite element transformation model as presented in Section 4.2.2, the parameters of the transformations are simply the node displacements. From Equation (4.4), it can be seen that the transformation Jacobian as defined in Section 2.4.3 is simply

$$\frac{du_i}{du_l^n} = N^n(\boldsymbol{x_i} + \boldsymbol{u_i}) \tag{4.20}$$

where by definition u_l^n is the l^{th} component of the displacement for the n^{th} node.

Regularization Metric Gradient When a second term is added to the cost function as seen in Section 4.2.2 for ensuring a smooth matching transformation, the gradient of the regularization metric has to be computed at each iteration. A nice feature of linear elastic energy (Section 4.2.2) is the straitforward expression of its gradient

$$\nabla E_{el} = \alpha \cdot \boldsymbol{KU} \tag{4.21}$$

4.3.2 Simultaneous Perturbation Stochastic Approximation for Gradient Descent

Simultaneous Perturbation Stochastic Approximation (SPSA) has been firstly introduced by Spall[Spall 1998]. SPSA has attracted considerable attention for solving optimization problems for which it is impossible or time consuming to directly obtain a gradient of the objective function with respect to the parameters being optimized. SPSA is based on a highly efficient gradient approximation that relies only on measurements of the objective function to be optimized. It does not require an explicit knowledge of the gradient of the objective function.

High Dimensional Optimization Problems

SPSA is especially efficient in high-dimensional problems in terms of providing a good solution for a relatively small number of measurements of the objective function. A well known asymptotical result about SPSA efficiency is mentioned on the `http://www.jhuapl.edu/SPSA` web site (p stands here for the dimension of the optimization space).

> Under reasonably general conditions, SPSA and the standard finite-difference SA method achieve the same level of statistical accuracy for a given number of iterations even though SPSA uses **p times fewer measurements** of the objective function at each iteration (since each gradient approximation uses only 1/p the number of function measurements). This indicates that SPSA will converge to the optimal solution within a given level of accuracy with p times fewer measurements of the objective function than the standard method.

A formal demonstration and associated conditions of this statement are given in Spall [1992].

Noisy Measurements of the Cost Function

SPSA, like other stochastic search methods, accommodates noisy measurements of the cost function. This is an important concern in many practical problems like problems involving Monte Carlo simulations or incomplete knowledge. In mutual information registration, a practical way of reducing the computing time is the selection at each iteration of random samples to measure the joint probability density between fixed and moving image features. This implementation introduces a noise on the measurements of the cost function at each iteration. Moreover, some challenging registration problems involve noisy images requiring to opt for optimization schemes designed for dealing with noisy measurements.

Comparison with Other Stochastic Optimization Schemes

Formal theoretical and numerical algorithmic comparisons of SPSA with other state-of-the-art optimization methods (simulated annealing, evolutionary computation, etc.) have generally shown SPSA to have equal or greater efficiency in terms of the overall cost of the optimization process (Chin [1994], Maryak and Chin [2001], Spall [2000], Spall et al. [1999]). This is especially the case when only noisy values of the objective function are available. Maryak and Chin [2001] compare the performances of SPSA, simulated annealing and genetic algorithm for a cost function of 10 variables presenting many local minima. The number of cost function evaluations to get an acceptable estimate of the global optimum is about $2,500$ for SPSA and $50,000$ for the genetic algorithm. The simulated annealing algorithm seems to perform poorly on the cost function used in this experiment. Even if these figures are simply an indication, they give an idea of the relative efficiency of SPSA.

Stochastic Gradient Estimate using Finite Differences

Finite differences are a practical solution when it is not possible to get an analytical estimate of the cost function gradient. A classical two sided finite difference approximation estimates the ith component of the cost function y gradient by ($i = 1 \ldots p$ where p is the number of parameters) :

$$g_{ki}(\theta_k) = \frac{y(\theta_k + c_k e_i) - y(\theta_k - c_k e_i)}{2c_k} \tag{4.22}$$

where e_i is the ith unit vector given by $e_i = (0 \ldots \underbrace{1}_{i} \ldots 0)$ and k refers to the current iteration in the optimization process.

Two sided finite differences therefore require $2p$ computations of the cost function. This can become quite consuming for high dimensional optimization problems. An alternative is to generate a stochastic perturbation to all the parameters at the same time. As this perturbation is regenerated at each iteration, an iterative optimization routine will converge to an optimum for certain distributions of the stochastic perturbation. Spall [1998] has shown than a simple Bernouilli with equiprobable ± 1 values (with a decreasing gain c_k over the iterations) was suitable to lead to convergence. The resulting SPSA approximation is thus

$$g_{ki}(\theta_k) = \frac{y(\theta_k + c_k \delta_k) - y(\theta_k - c_k \delta_k)}{2c_k \delta_{ki}} \tag{4.23}$$

In Equation (4.23), only two measurements of the cost function y are needed since the numerator is the same for each component of the gradient.

Iterative Gradient Descent

To maximize the cost function, an easy evolution rule is to update the set of parameters at each iteration by

$$u_k = u_{k-1} + a_k \hat{g}_k(u_{k-1}) \tag{4.24}$$

where a_k is defined as the *learning rate*. The gain sequences a_k and c_k are decreasing during the iterations following

$$\begin{aligned} c_k &= c/k^\gamma & (4.25)\\ a_k &= a/(k+A)^\gamma & (4.26) \end{aligned}$$

To stabilize the process, several gradient estimations can be generated with different perturbations and then averaged. This scheme makes stochastic search algorithms easy to implement on shared memory parallel architectures since several estimates of the gradient can be generated independently.

4.3.3 Genetic Optimization

It exists a wide range of genetic optimization schemes. Basically though, genetic optimization is a biologically inspired optimization algorithm where a population of individuals evolves over several generations. The various schemes of genetic optimization schemes differentiate themselves by the strategy used for evolving from the population at the current time step of the algorithm to the population at the next time step.

Each individual is characterized by a set of values for the optimization parameters, called the genes. Random operators, modeling crossovers and mutations, generate new individuals from generation to generation. The crossover operator generates a new individual from two "parents" by intertwining their genes. The mutation operator is obtained by randomly substitution of certain array elements (in the bounds of the optimization domain). A mutation probability has to be defined.

Furthermore a selection scheme retains good individuals (with respect to the optimization objective) and rejects bad ones, so that the population size remains constant. The individuals with a poor value of the objective function are replaced using the crossover and mutation operators. The number of individuals to be replaced at each generation is determined by a fixed replacement probability. This is another parameter to be fixed. The algorithm stops when a fixed convergence to the best element of the current population is reached. A maximal number of generations is also fixed to avoid excessive computation time.

The reader can refer to Goldberg [Goldberg 1989] as an introductory reference. In Section 4.7, we use the steady-state implementation of the open-source genetic optimization library by Matthew Wall [Wall 1996].

4.4 Application : Mutual Information based Rigid Registration using SPSA

4.4.1 Implementation

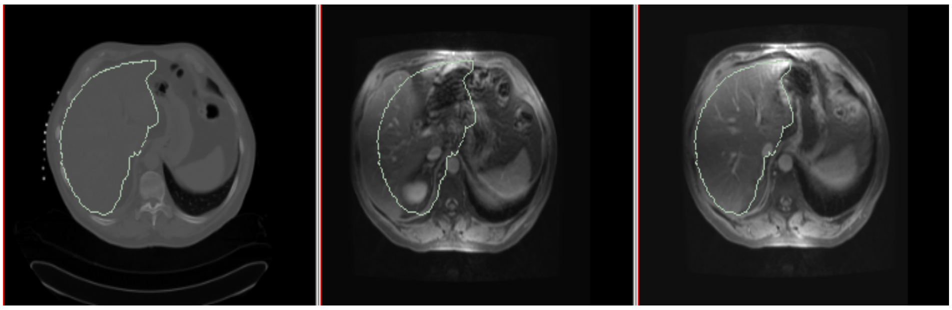
The SPSA optimization routine was implemented in the Insight segmentation and registration toolkit[Ibanez and Schroeder 2005] as an object derivating from `itkSingleValuedNonLinearOptimizer`. This way, this optimizer is integrated in the registration framework of the ITK library. ITK offers the possibility to compare several metrics, optimizers and transformation models by using a C++ object oriented interfacing approach. In this section, we present some results obtained by interfacing the SPSA optimizer with the Mattes[Mattes et al. 2003] implementation of the mutual information metric and a rigid deformation model allowing three rotations and translations.

4.4.2 CT-MR Liver Registration

The rigid alignment of a 3D MR liver image on a CT of the same subject was performed within less than 5 minutes on a standard computer. The Mattes Metric used 50 bins in the fixed and moving image and 10 000 samples randomly selected in the fixed image domain. The results are shown in Figure 4.4.

4.4.3 Brain MR T1/T2 Registration and DTI/MR Registration

Figures 4.5 and 4.6 illustrate two rigid registration results. Figure 4.5 shows the alignment of an intra-operative brain T2 image on a pre-operative T1 image of the same subject. Figure 4.6 shows the matching of a T1 intra-operative image on the baseline component of a pre-operative Diffusion Tensor Image (DTI). The DTI modality is over-imposed on the other image and covers a more limited region than the T1 image. The DTI image was therefore chosen as fixed image and the transformation was inverted for visualization purpose.



(a) (b) (c)

Figure 4.4: *SPSA optimization of mutual information for rigid registration of CT-MR liver images.*
Sub-figure (a) shows a slice of the - fixed image - CT. The contours of the target liver have been highlighted for visualization only (the algorithm is blind to these contours). Sub-figure (b) shows the initial alignment given to the algorithm. After optimization using the SPSA routine, the resulting alignment is plotted in Sub-figure (c).

4.5 Non-rigid FEM Registration : Preliminary Experiments

In this section, we present the result of preliminary experiments obtained using the hybrid cost function of Equation (4.19) and a simple gradient descent optimization scheme. The implementation of the mutual information metric is these experiment is the Viola and Wells [1996; 1995] implementation provided by the ITK [Ibanez and Schroeder 2005] toolkit.

4.5.1 Synthetic Luminance Distortion

The purpose of this experiment was to evaluate the robustness of the MI criterion in the presence of significant nonlinearities between the intensities of the scans to be aligned. We evaluated both MI and L_2 metrics.

For this experiment, we have first rigidly registered a pre-operative MR image on an intra-operative image of the same patient. For rigid registration Powell's optimization method was used to identify a rigid transformation maximizing the mutual information metric. The implementation is similar to [Maes and Collignon 1997].

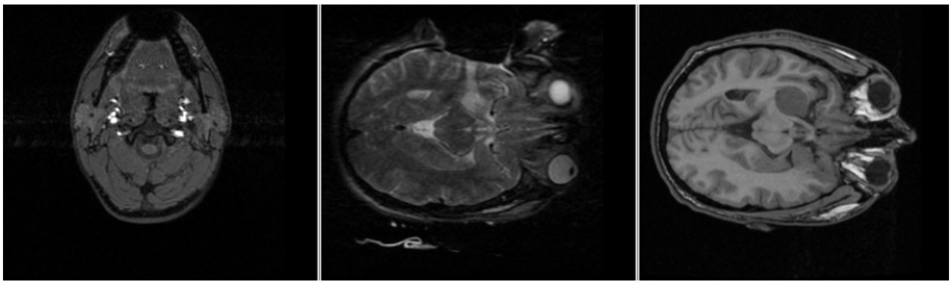


Figure 4.5: *Rigid registration of pre-operative MR T1 image on a intra-operative T2 image of the same subject.*
The first image shows the pre-operative -target- image. The intra-operative image is shown with a center to center alignment in the middle. After optimization of the mutual information metric, the intra-operative registered image is plotted in the right part of the figure.

Following the 3D rigid registration, a particular slice was extracted in the registered volumes and the brain was manually segmented in this specific slice. Then, to evaluate the robustness of the behavior of the mutual information criterion, we applied a luminance distortion to the intra-operative image. The distorted luminance $L'(x)$ is obtained from the original $L(x)$ by

$$L'(x) = \exp\left(\frac{L(x)}{C}\right) \tag{4.27}$$

The source (preoperative image) is shown in Figure 4.7(a) and the corrupted target image is shown in (b). Figures 4.7(c) and 4.7(d) show the result of the registration process using a mutual information criterion(warped image and resulting joint histogram are presented). Figures 4.7 (e) and (f) show the same results with a classical Root Mean Square (RMS) error. It is clear that the RMS criterion does not converge to a satisfying matching. The resulting joint histogram is not as compact as the result of an MI metric optimization.

4.5.2 Synthetic Transformation and Landmark Validation

In this experiment, a synthetic brain shift transformation was applied to a segmented slice of a pre-operative image. The synthetic displacement field is an arctangent in the vertical y direction. A set of landmarks (shown on Figure 4.8) was selected on the main contours of the image (brain, tumor

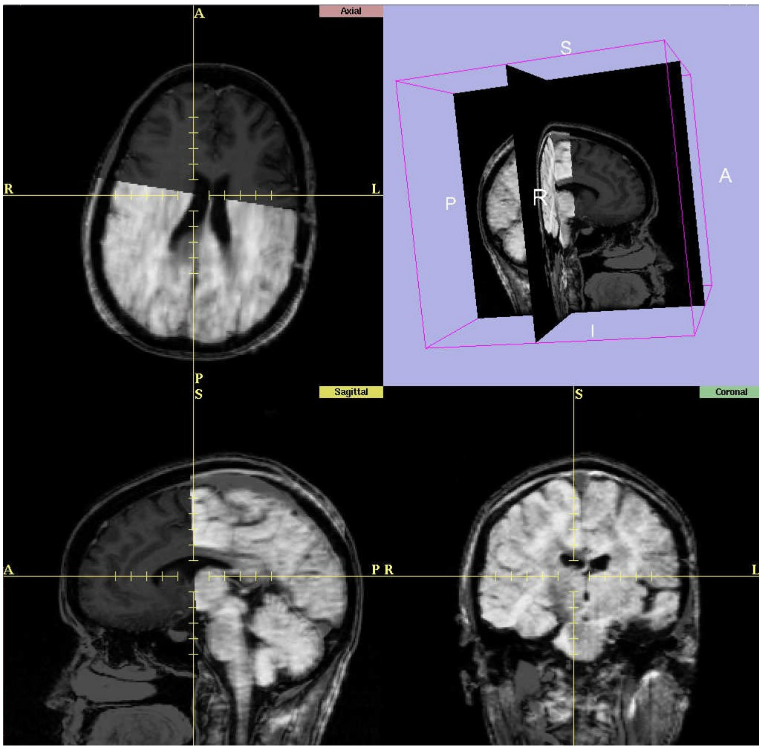


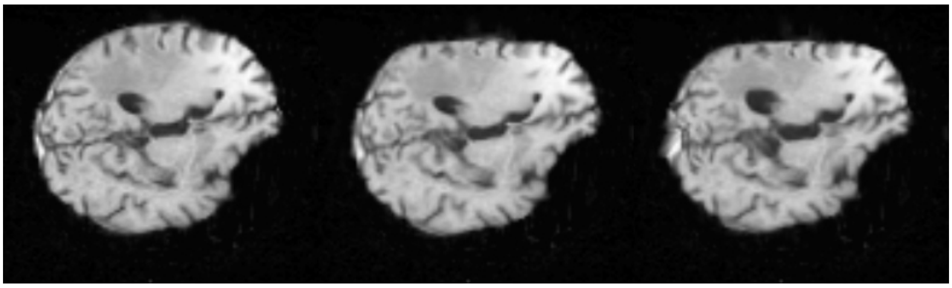
Figure 4.6: *Registration of the baseline component of a pre-operative DTI image on a T1 MRI acquired in the operating room.*
The DTI image (with a clearer contrast) was mapped on the T1 image. Manually registration would be made difficult in that case due to the slight slope in the acquisition plane of the DTI image.

and ventricles). In this case, the ideal transformation is known, and the registration error can be precisely evaluated. The first part of Figure 4.8 shows the initial set of landmarks (dots) and the target set of landmarks (crosses). The second part of Figure 4.8 match the target landmarks with the landmarks positions transformed by our algorithm. A significant decreasing of the registration error was observed.

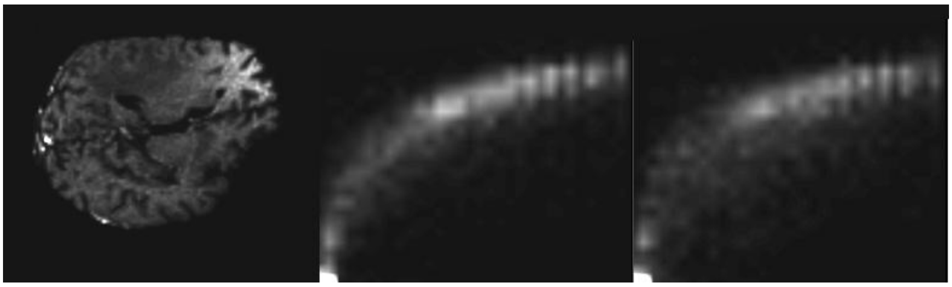
Registration error (voxels)	mean	min.	max.
initial	2.040	0.029	3.120
final	0.697	0.006	2.126

4.5.3 Intra-Operative Case

This experiment illustrates the registration of a pre-operative SPGR image to a post-operative MR T2 image of the same patient. These images were first rigidly registered in 3D as in the first experiment. In this case, a multi-level pyramid registration was used to catch the important rotation around the z axis. Two slices have been segmented manually and non rigidly registered. A multi-resolution strategy has been used. In this scheme, the transformation found in the former level is applied to each node of a finer mesh as initialization for the next level. The Figure 4.9 shows the result of the non rigid registration with and without regularization term (R in Eq. 4.18). It is clear that the ventricles are better matched if a regularization term is included.



(a) Source (c) MI reg. (e) RMS reg.



(b) Target (d) MI reg. scatterplot (f) RMS reg. scatterplot

Figure 4.7: *Synthetic luminance distortion for evaluating MI robustness: 2D experiment.*
A T1 weighted preoperative scan (a) was registered on a T1 weighted postoperative scan with an exponential luminance distortion (b) using a finite element elastic model. Two metrics are compared. Sub-figure (c) shows the deformed image with a mutual information criterion. Its joint histogram is plotted in the sub-figure (d). Sub-figure (e) shows the deformed image with a root mean square criterion. Its joint histogram is plotted in the sub-figure (f). The use of a MI metric leads to a better match and the resulting joint histogram is more compact.

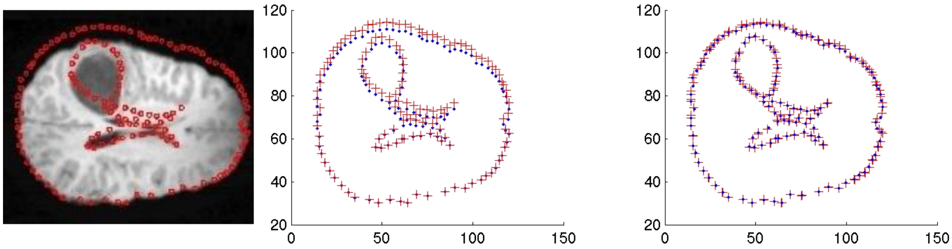
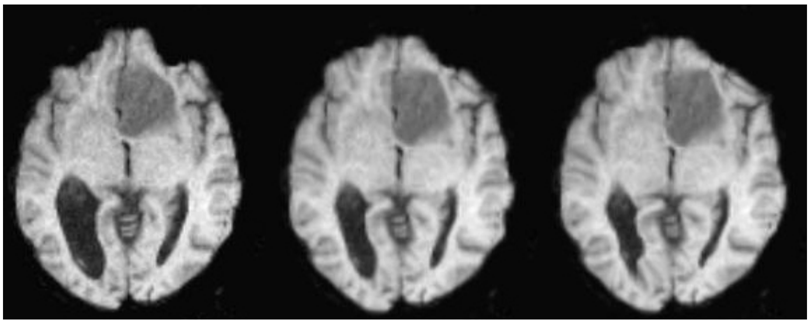
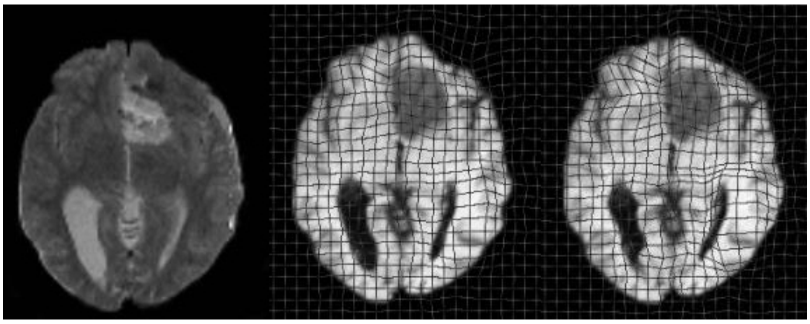


Figure 4.8: *Synthetic transformation and landmark validation: 2D experiment. Set of landmarks for registration accuracy measurements (left). Synthetic transformation on landmarks (middle). Comparison between target and landmarks after registration (right).*



(a) Source (c) MI result with regularization (e) MI result without regularization



(b) Target (d) With regul. (f) Without regul.

Figure 4.9: *Intra-operative brain shift case : 2D experiment.*
A T1 weighted pre-operative scan (a) was registered on a post-operative T2 weighted scan (b) using a finite element elastic model and a mutual information metric. Sub-figure (c) shows the deformed image with elastic regularization and the corresponding deformed mesh is plotted in sub-figure (d). If no regularization term is included in the optimized cost function, some inconsistencies appear in the resulting deformation field (especially in the ventricles area) as shown is (e) and (f).

4.6 Application : Non-rigid Registration of Pre-Operative and Intra-Operative Images

4.6.1 Background and Related Work

The development of fast non-rigid automatic registration algorithms is an important issue to fuse during surgery pre-operative and intra-operative data. T1-weighted SPGR MRI is used during brain surgery at Brigham and Women's Hospital (Boston, MA) for intra-operative surgical navigation. However, after the injection of a contrast agent (Gd-DTPA) to emphasize the tumor, the intensity properties of the tumor and surrounding tissues may be significantly altered. This can disrupt registrations which assume a linear dependence between the voxel intensities of the scans [Hata et al. 1999; 2000]. The surgery can also induce locally important modifications between the pre-operative and intra-operative modalities (tumor resection, hematomas, …). Local similarity measures such as block matching could diverge in these areas. Mutual information(MI) has been extensively used in medical images registration. It has be proven to be robust to contrast changes. MI is based on the joint histogram between the images to be aligned and is therefore a global metric which can be robust to local dissimilarities between the images.

In MRI-guided prostate biopsy, tumor foci can be better localized and identified on the pre-operative 1.5-T MRI than in the 0.5-T used for surgical navigation. For this reason, deformable registration of the pre-operative MRI to the intra-operative MRI has to be investigated. MR and CT images of the liver are relevant to assess tumor coverage, which is the most decisive parameter used for a successful procedure after image guided therapy such as RF-Ablation or cryotherapy. They also play an important role in providing information about the ablation of healthy tissue. A contrast enhanced pre-procedural MR (usually within seven days before), giving information about the tumor size and location, has to be matched with the contrast enhanced post-procedural MR (24 hours post) which provides relevant information about the area of necrosis. Rigid and non-rigid deformations caused by breath-hold pattern and the physiological changes in the micro-structure of the liver require the development of non-rigid registration tools. Also the intervention itself leads to a change in the liver shape. As well as in MR guided neuro-surgery, contrast changes are important between the two images to be registered in these two applications. In the liver application, some organs (e.g. stomach, colon, small bowel, …) can vary in an important way between the two acquisitions. For these reasons, a robust and global similarity measure like the mutual information metric sounds like a well adapted choice.

Ferrant[Ferrant et al. 2002; 2001; 2000] described the use of a finite element mesh as an efficient set of shape functions to model non-rigid deformations of the brain over the course of a surgery. Intra-operative segmentation of the brain was used to estimate displacements of the ventricles and cortex contours. These surface deformations were then propagated through the whole volume using a linear elastic interpolation. In this approach, the quality of the mapping depends on the segmentations in the source and target images. The quality of the match can be increased by adding more surfaces (i.e. more boundary conditions) but no guaranty is given about the quality of the registration in regions far from the boundary surfaces. The use of a volumetric similarity measure does not favor specific regions than others and does not require the use of intra-operative segmentations. Mattes, Ruekert, Bookstein, Davis and al [Bookstein 1989, Davis et al. 1997, Mattes et al. 2003, Ruekert et al. 1999] have presented a mutual information based non-rigid registration using B-Spline Radial Basis kernel Functions(RBF). RBF kernels can lead to a non-bijective function if two neighboring splines have an opposite contribution to the displacement field (which is not the case with volumetric meshes if to neighboring nodes are moving in opposite directions).

4.6.2 Registration Strategy

Objective Function

The cost function used in this work is the sum of mutual information and elastic energy (Equation (4.19)). The weight α remains a critical parameter to be chosen by the user. For coarse meshes, a smaller α can be used since the deformation model does not allow high frequencies in the deformation field. α can then be increased when finer meshes are considered (see Section 4.6.2).

Similarity Measure

The similarity measure we used is the Mattes [Mattes et al. 2003] implementation of mutual information. The main advantage of this implementation is to perform a quick estimation of the joint and marginal probability densities by using histograms. To have a cost function with sufficient continuity properties is a necessary condition for optimization convergence.

Therefore, the histograms are smoothed by a continuous kernel.

Regularization Term

The regularization term R in (4.18) provides a mechanism for including prior information regarding the nature of the deformation field. A usual *a priori* assumption is a smooth deformation field (see [Bauchemin and Barron 1995] for different regularization terms applied to optical flow techniques).

In this work, the regularization term is the linear elastic energy defined in Section 4.2.2. This regularization term acts to propagate the effect of active regions (regions where changing the amplitude of local basis functions significantly changes the value of the similarity metric) to the whole image domain.Segmentation of the fixed image is not required but allows to employ different material characteristics for different anatomical structures.

Implementation

The cost function given in Equation (4.19) is optimized using a gradient ascent strategy as described in Section 4.3.2. The Mutual Information gradient regarding u is estimated using the SPSA approximation described in Section 4.3.2. The gradient of the elastic energy is simply equals to $\alpha \cdot \boldsymbol{K}\boldsymbol{u}$ (Equation (4.21)) where $\boldsymbol{K}$ is the stiffness matrix described in Section 4.2.2 and is computed once for all before starting the SPSA iterations.

The flowchart in Figure 4.10 summarizes the following steps of our implementation :

- **Step 1 -** The derivative of the mutual information metric is estimated for each vertex of the mesh using a finite difference estimation. The vertices where the norm of the derivative is above a fixed threshold are selected as 'active vertices' for the MI gradient SPSA approximation. Only the contribution of the points belonging to the adjacent elements of a disturbed vertex (in the x,y or z directions) needs to be recomputed in the joint histogram. This 'active' part of the joint histogram is smoothed by the use of a B-Spline kernel to enforce smooth variations of the objective function[Mattes et al. 2003]. The use of a Symmetric Multi-Processor (SMP) architecture enables the delegation of the computation of mutual information for different perturbations to several threads.

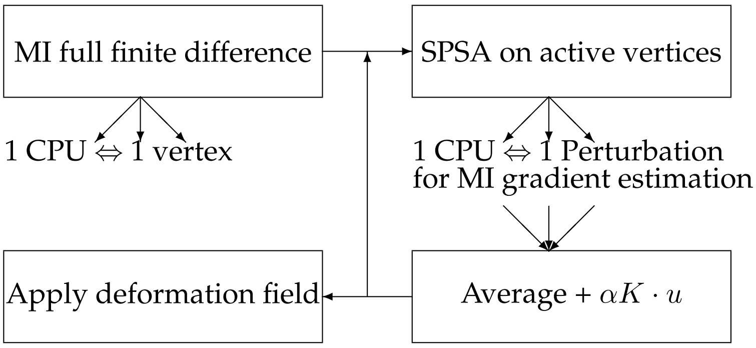


Figure 4.10: *Non-rigid registration flowchart for a finite elements deformation model with SPSA optimization.*

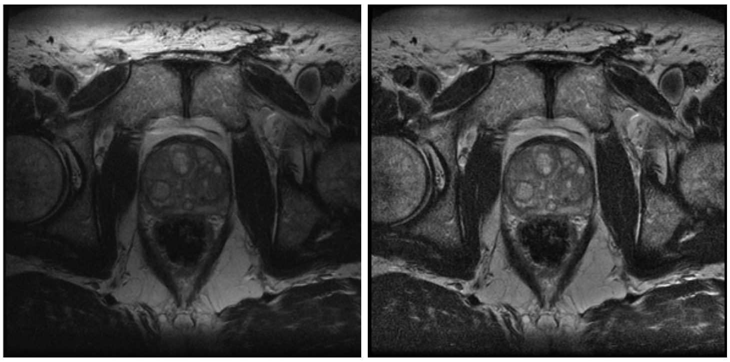
- **Step 2 -** At each iteration, each CPU performs a different perturbation of the subset of active vertices to stochasticly estimate the mutual information gradient. As in the first step, only the contributions of the points belonging to elements adjacent to active vertices are recomputed.
- **Step 3 -** These estimations are then averaged and added to the gradient of the elastic metric ($\alpha \cdot Ku$) before updating the current vector u of parameters.
- **Step 4 -** The algorithm loops to the second step until a convergence threshold is reached. After optimization, the resulting transformation is applied to the image and the displacement field is stored to be reloaded at the next resolution. In some applications, it is useful to update the set of active vertices by recomputing a full derivative with finite differences

This approach can easily be extended to a "coarse to fine" estimation of the displacement field : the algorithm is first run with a low resolution mesh. The resulting displacement field after optimization is then applied to the nodes of a finer mesh as initial condition for the next level. A similar approach allows to work with images pyramids by aligning in a first time sub-sampled versions of the fixed and moving images. The resulting displacement field can then be interpolated and applied as initialization to the next level.

4.6.3 Prostate Biopsy

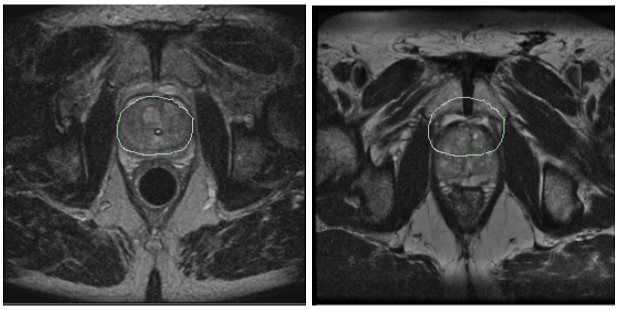
In this application, an intra-operative image of the prostate is acquired during prostate biopsy. The tumor and internal structures of the prostate are better localized on the pre-operative image of the prostate. The dimensions of the intra-operative image are 256x256x20 with a voxel size of 0.78125x0.78125x5 mm^3. The pre-operative image size is 512x512x30 with a 0.390625x0.390625x5 mm^3 voxel size. The pre-operative image suffers from important artifacts in the gray levels, due to the presense of an endo-rectal coil. Since mutual information is a similarity measure based on the image densities, such artifacts can lead to a bad convergence of the optimization. A bias correction had to be applied before the registration process. The approach used is described in Mangin[Mangin 2000, Weisenfeld and Warfield 2004]. Figure 4.11 shows the non-corrected (a) and the corrected (b) image. It can be seen that the bias is concentrated in this case in the top left corner of the image. The bias model used here allows more degree of freedoms than the model described in Section 4.7.3. Figure 4.12 shows how a rigid registration process can diverge in the absence of a pre-processing bias correction step.

The pre-operative and intra-operative images are then rigidly registered with the same method as described in the first experiment. The intra-operative (target) image is plotted in Figure 4.13(a). The pre-operative image before registration (center to center translation) is plotted in sub-figure (b). The result of the rigid registration is shown in Figure 4.13(c). The rigid transformation is taken as the initial condition for the non-rigid algorithm described in section 4.6.2. Several resolutions were used to approximate the deformation between the two images. The first mesh resolution for the regular grid before splitting in tetrahedra was 64x64x4 pixels, the second resolution uses a 32x32x4 resolution and the finest mesh 16x16x2. For validation purposes, the prostate volume was segmented in the two images before registration. The overlap between the source and target volumes was 72% after rigid registration and 85.4% after non-rigid registration (Figure 4.13(d)).

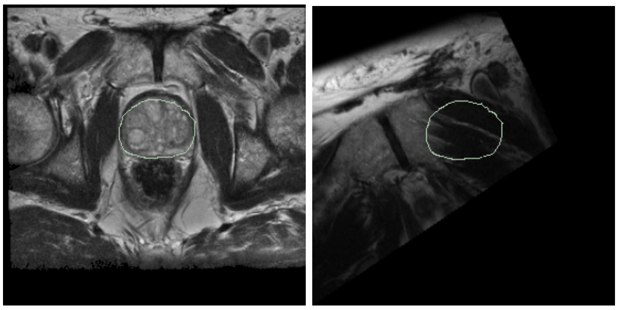


(a) Without correction. (b) With correction.

Figure 4.11: *Bias correction.*
This figure shows the effect of the bias correction using the implementation described in [Weisenfeld and Warfield 2004] and [Mangin 2000]. This correction is a necessary pre-processing to avoid a divergence of the registration process as shown in Figure 4.12.



(a) Intra-op (b) Pre-op : Initial



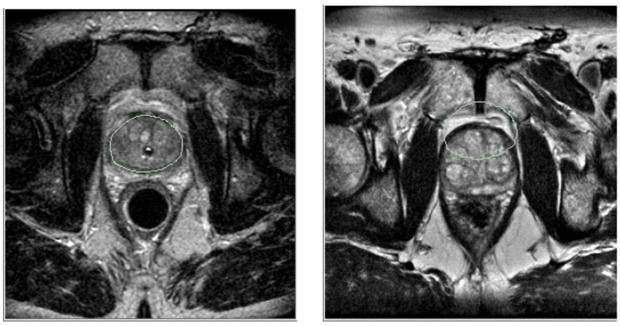
(c) Rigid reg.

with bias corr.

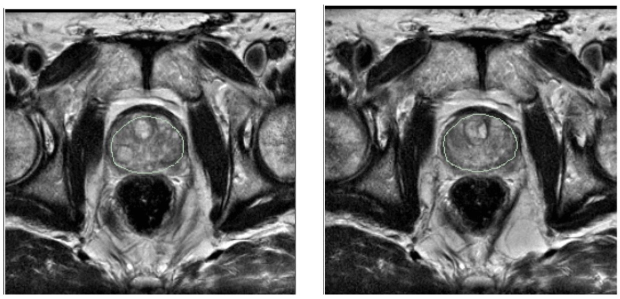
(d) Rigid reg.

without bias corr.

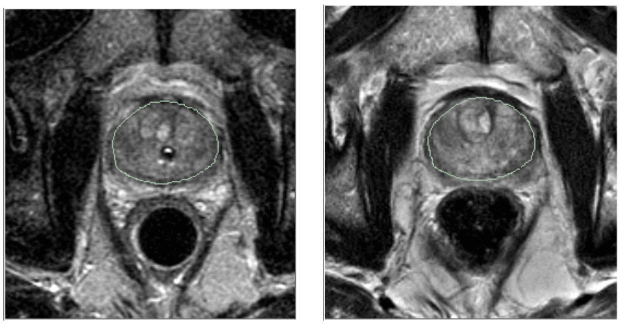
Figure 4.12: *Bias field inducing a divergence of the registration process*
This figure shows how the potentially large bias field due to the presence of an endo-rectal coil can lead to a divergence of the registration process. Subfigure (a) shows an axial slice of the intra-operative image (taken in this experiment as the fixed image) with the prosate boundary. Sub-figure (b) shows the initial alignment of a bias corrected version of the pre-operative image. The offset to recover can be seen by the distance to the intra-op. prostate contour (redrawn from Sub-figure (a)). Sub-figure (c) shows a satisfactory alignment obtained by maximizing mutual information between (a) and (b). Sub-figure (c) shows the mis-alignment got after maximizing mutual information if the bias in the pre-operative image is not corrected.



(a) Intra-operative image (b) Pre-operative image



(c) Rigid reg. of (b) to (a)(d) Non-rigid reg. of (b) to (a)



(e) detailed of (a) (f) detailed of (d)

Figure 4.13: *Rigid and non-rigid registration of a 1.5 pre-operative image of the prostate to 0.5-T intra-operative image(a).*
The initial condition for the rigid registration is shown in (b). The rigid registration was performed using a stochastic gradient descent (SPSA) optimizer and a mutual information metric, the result is shown in (c). The application of our non-rigid registration strategy is shown in (d), both the contours and the internal structures are closer to the target intra-operative image.

4.6.4 RF Ablation of Liver Tumors

In this experiment, our non-rigid registration algorithm was used to efficiently compare the pre-procedural tumor volume with the post-procedural necrosis after RF ablation.

First, the pre-procedural image is rigidly aligned to the post-procedural image. The dimensions of the pre-procedural image are 256x256x91 with a 1.5625x1.5625x2.5 mm^3 voxel size and 256x256x80 (with a 1.3672x-1.3672x2.5 mm^3 voxel size) for the post-procedural image. The result of the alignment is shown is Figure 4.15(a) and (b). This rigid registration uses the Mattes[Mattes et al. 2003] implementation of the mutual information metric. The rigid registration algorithm has been described in Section 4.4.

Next, the non-rigid registration described in Section 4.6.2 was applied to map the post-procedural image to the rigidly aligned pre-procedural image. For this case, an image pyramid was used for faster computation. At the first level of the pyramid, three resolutions were used to gradually refine the tetrahedral volumetric mesh. The contours of the rigidly aligned pre-procedural liver are mapped in Figure 4.15 on the post-procedural image before (a) and after (c) non-rigid registration. Figures 4.14 (a) and (b) show the difference between the source and target livers before and after non-rigid registration. In this experiment, the displacement is localized in the left liver lobe. After segmentation of this lobe in the fixed and moving image, an improving in overlapping from 62.79% to 94.93% was observed after non-rigid registration. No significant change is observed in the right lobe.

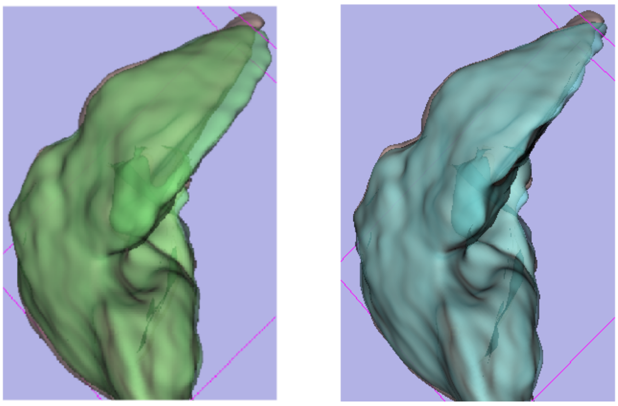
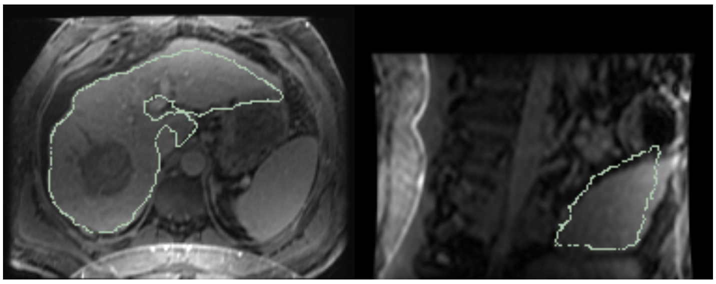
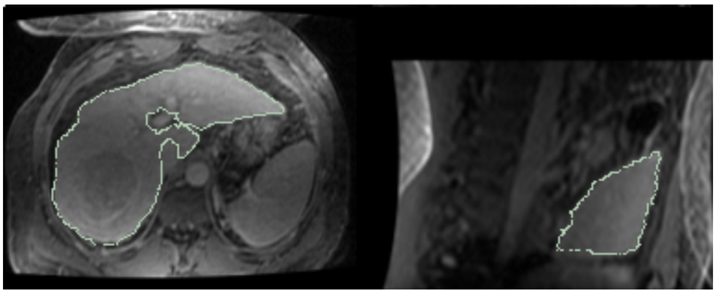

(a) Before registration (b) After registration

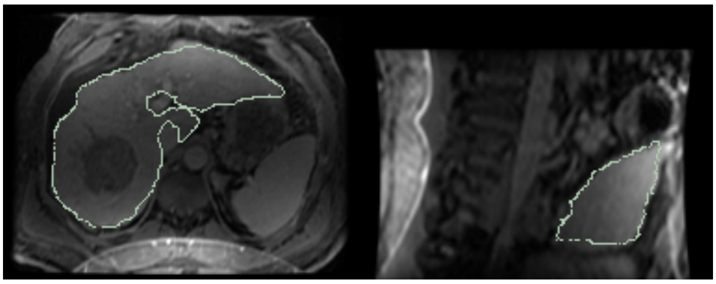
Figure 4.14: *Non-rigid registration of MR liver images.*
Sub-figures (a) and (b) show the liver surface (obtained from manual segmentations) in pre and post-procedural images before and after non-rigid registration. These contours are not used during the non-rigid registration.



(a) Post-procedural image



(b) Pre-procedural image



(c) Deformed post-procedural image

Figure 4.15: *Non-rigid registration of MR liver images.*
The pre-procedural (a) MR image of the liver was first aligned on a post-procedural (b) image of the same patient. The non-rigid registration of (a) on (b) is shown in (c).

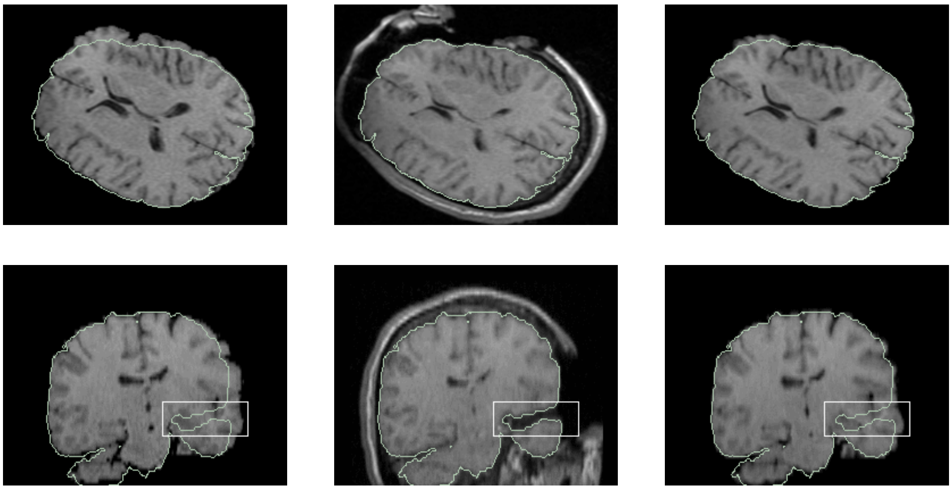
4.6.5 Brain Shift Estimation

In this section, the algorithm described in section 4.6.2 was applied for compensating the brain shift occurring during surgery after craniotomy. Such deformations cause misalignment of the surgical planning with respect to the preoperative imaging it is based on. The major factors of these non-rigid deformations include the gravity, the tissue retraction or resections caused by the tumor ablation, and the cerebrospinal fluid leakage or draining. Intra-operative MR images are acquired by an 0,5 T Open Magnet System (Signa SP, GE Medical Systems, Milwaukee,WI) (see Nabavi[Nabavi et al. 2001] for a description of the intra-operative procedure). Registration results for two patients are presented in this paper. A first image before skull opening is acquired (dimensions are 256x256x60, the voxel size is 0.94x0.94x2.5 mm^3). The brain is then manually segmented on this image.

In a pre-processing step, the pre-operative image was rigidly aligned with the intra-operative image. This registration process has been described in Section 4.4.

Estimating the non-rigid displacement from the pre-operative brain to the intra-operative is made difficult by the artifacts induced by the skull opening. Furthermore, where the skull is present in both images, it imposes a null displacement which brings a discontinuity with the displacement of the brain surface a few millimeters below. Such discontinuities are difficult to be estimated by a linear elastic model such as described in Section 4.2.2. Since we are mainly interested in modeling the deformations of the brain, we use a segmentation of the pre-operative brain as a mask to measure mutual information. The linear elastic model will then propagate the brain deformations outside the mask which means that the deformation will have no physical meaning outside this mask. No segmentation of the intra-operative image is required since mutual information is a statistical measure based on the joint intensity distribution between the source and target images.

Figure 4.16 shows in the first column the segmented pre-operative brain. The second column plots the target intra-operative brain and the third column shows the deformed pre-operative brain on the intra-operative image. The displacement field found by the registration algorithm is defined from the pre-operative to the intra-operative image. On the other hand, to warp the pre-operative on the intra-operative image, the displacement needs to be defined from the intra-operative image to the pre-operative image. Therefore, the resulting transformation has been

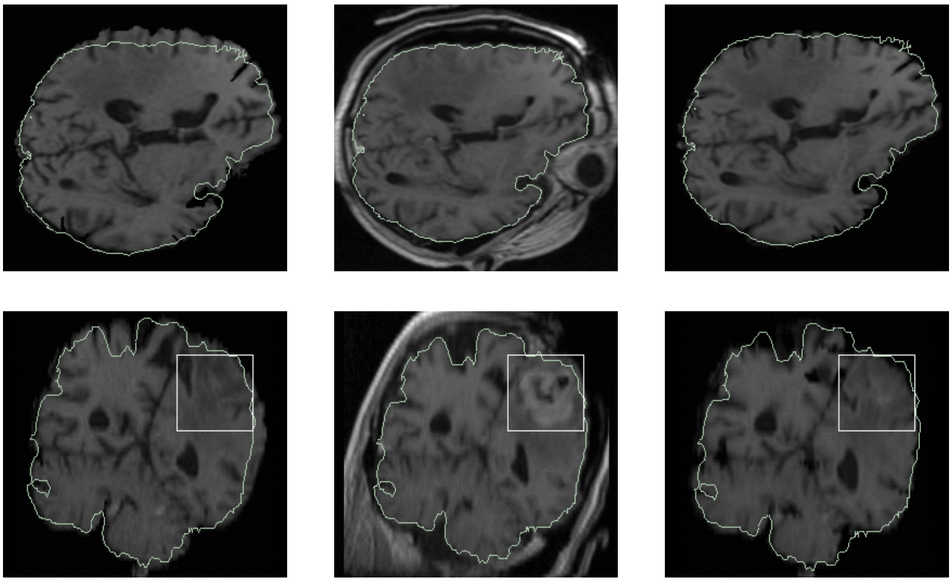


(a) Pre-operative (b)Intra-operative (c) Deformed pre-op.

Figure 4.16: *Non-rigid registration of brain shift MR images.*
Case 1 - *Non-rigid registration from a pre-operative image to an intra-operative image of the same patient. For visualization purpose, the intra-operative contours of the brain are displayed over the pre-operative brain (a), the intra-operative image (b) and the deformed pre-operative brain on the intra-operative image (c). The rectangle shows a region where a local similarity measure such as block matching[Rexilius 2001] could fail to find similarities between the pre-operative and the intra-operative brain.*

inverted. A simple inversion scheme is described at Figure 4.18. A new mesh is generated by applying the optimal displacements to the mesh vertices. The displacements associated to each vertex of the new mesh are simply the opposite of the initial displacements as shown in Figure 4.18. This solution is exact at the mesh vertices but inaccurate inside the elements. Therefore, this inversion can be inaccurate for coarse meshes.

For visualization purposes, the contours of the target intra-operative brain are mapped on the pre-operative image before (Figure 4.16(a))and after (Figure 4.16(b)) registration. This intra-operative segmentation is not used during the registration. The boxed area in Figure 4.16 shows a region of local dissimilarity between the pre-operative and the intra-operative image where a local matching strategy could fail to find correspondences. Figure 4.17 plots the results in the same way as in Figure 4.16 for the sec-



(a) Pre-operative (b) Intra-operative (c) Deformed pre-op.

Figure 4.17: *Non-rigid registration of brain shift MR images.*
Case 2 - *The results of the non-rigid algorithm are plotted in the same way as in Figure 4.16. Only the pre-operative segmentation of the brain is used by the registration algorithm as a mask to measure the matching. This is done to be robust to artifacts induced by the opening of the skull. The rectangle shows a region where deformation induced by surgery and contrast changes could disturb a mean square error criterion.*

ond subject of this work. For this case, a contrast agent was injected between the first and the second image acquisition. This makes the tumor light up in the second scan. Such contrast change between the modalities could disrupt registrations assuming a linear dependence between the voxel intensities of the scans[Hata et al. 1999; 2000].

Figure 4.19 plots the estimated displacement field by the registration algorithm. The gravity is acting from the top to the low of the axial views and the right to the left of the sagittal views. The first column plots the displacement field for the first case (same as in Figure 4.16) and the second column the second case (same as in Figure 4.17). The deformation estimated in the second case has an important component in the z direc-

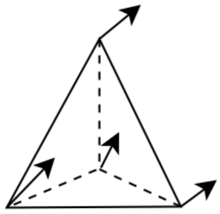
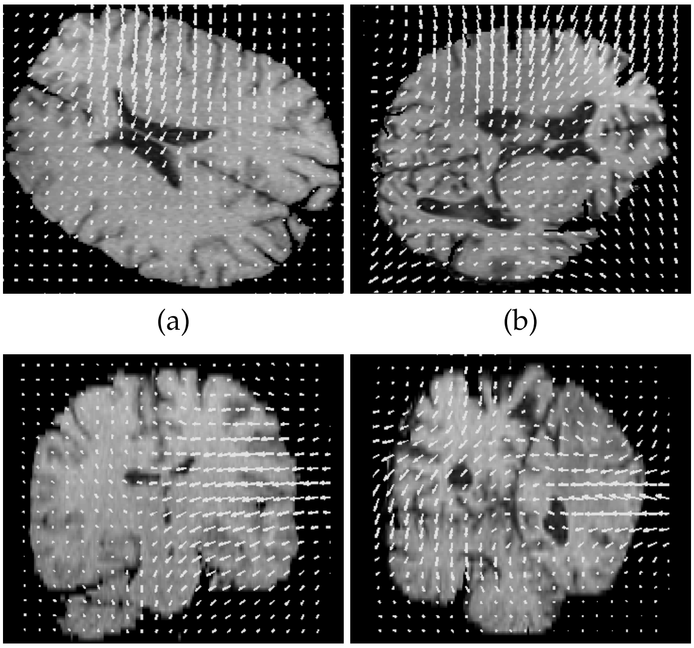


Figure 4.18: *Non-rigid transformation inversion.*
This figure shows the approximation used to inverse the transformation at the end of the optimization. This allows to compute the displacement field from the intra-operative to the pre-operative coordinate system even if the transformation was computed in the other direction during the optimization.

tion. Therefore, a simple slice to slice 2D registration would be inaccurate in this case.

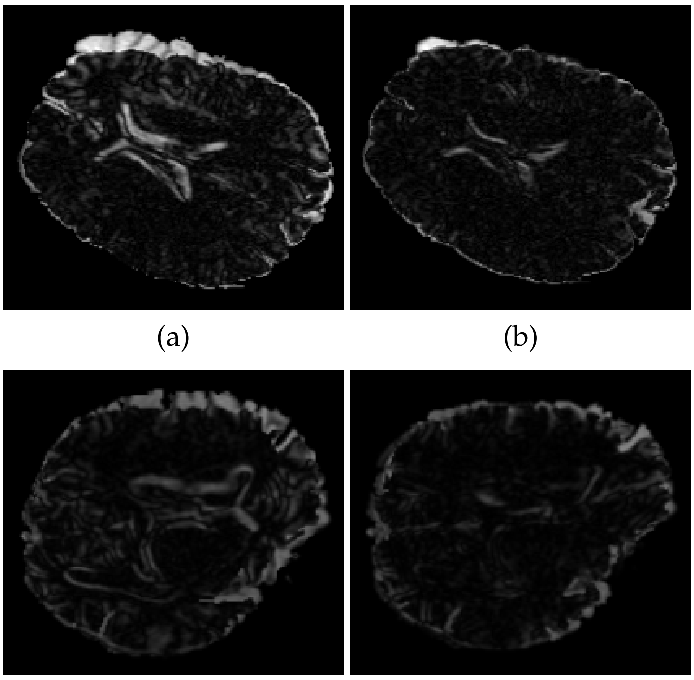
To estimate the registration accuracy, we plot the image difference before and after registration since in this case, the modality for acquiring the pre-operative image and the intra-operative image are the same, the difference should tend to zero (except in regions where the contrast agent is emphasizing the lesion as well as in resected or cut areas). The image difference before and after registration is plotted in Figure 4.20(a,b) for the first case and in Figure 4.20(c,d) for the second case.



(a) (b)

(c) (d)

Figure 4.19: *Non-rigid brain shift displacement field.*
The displacements fields are plotted in axial (first row) and sagittal (second row) for the first case (a,c)(see figure 4.16) and the second case (b,d)(see figure 4.17). Gravity is acting from the top to the bottom on the axial view, and from the right to the left on the sagittal view. In the second case, deformations induced by surgery cause a more complex displacement field than in the first case.



(a) (b)

(a) (b)

Figure 4.20: *Brain shift : difference images before and after registration.*
Case 1 (a,b) - *The difference before (a) and after (b) registration is plotted for the first case (see figure 4.16) to assess registration accuracy.* **Case 2 (c,d) -** *The difference before (c) and after (d) registration is plotted for the second case (see figure 4.17) to assess registration accuracy.*

4.7 Application : Simultaneous Registration and Bias Correction of Brain Intra-Operative MR Images

4.7.1 Introduction

Chapter 2 has given a short review of the wide range of registration applications based on the maximization of mutual information. The success of this similarity measure comes from its ability to catch complex dependencies between multi-modal signals such as a Magnetic Resonance (MR) and Positron Emission Tomography (PET).

Mutual information is in fact a non-linear measure of the divergence between the marginal and joint distributions of the gray levels of the source and target images. Therefore the registration can fail if one of these images suffer from *spatially* variant luminance distortions. Mutual Information based registration is obviously able to deal with uniform luminance bias. For this reason, it is important to examine robust ways to correct luminance deformations before or while performing the registration.

Different approaches can be followed to correct the bias field in the gray levels of the intra-operative image. First, a model of the luminance distortions must be chosen and inverted. Secondly, to optimize the parameters of this model, an optimization strategy must be defined.

If no other information is available, a first possibility consists of achieving the correction using the biased image itself because luminance distortions introduce disruptions in the statistical distribution of the image. Of course, the correction strategy is easier if an *a priori* model on the marginal distribution of the unbiased image is available. Van Leemput[Leemput et al. 1999] *et all* introduced Gaussian assumptions about the distribution of each *class* in the input image. This leads to simultaneous classification and bias correction.

In the context of intra-operative image acquisition, a first *diagnosis* MR image is acquired before surgery. To calibrate the intra-operative device, a second image acquisition is performed in the operating room before surgery. Because of the small size of the magnets and the short acquisition time, the images produced by such devices are often subject to luminance distortions, and are characterized by a limited field of view and low spatial resolution. Therefore, it would be interesting to perform a registration between the two modalities. Since registration can fail if one of the two images is biased[Butz et al. 2002], it is convenient to examine simultaneous registration and bias correction algorithms. The resulting bias correction will be based on the preoperative unbiased MR image.

4.7.2 Marginal Bias Correction

Bias correction Model

In our implementation, we used the model developed by Likar[Likar et al. 2000] to correct the distortions in the intra-operative image. It is assumed that it is possible to get an unbiased version $U(\vec{x})$ of the biased luminance $N(\vec{x})$ with multiplicative and additive corrections

$$U(\vec{x}) = M(\vec{x}) \cdot N(\vec{x}) + A(\vec{x}) \tag{4.28}$$

The additive and multiplicative bias corrections consist of a weighted polynomial basis functions sums.

$$A(\vec{x}) = \sum_i a_i A_i(\vec{x}) = \sum_i a_i \frac{P_i(\vec{x}) - a_{ni}}{a_{ei}} \tag{4.29}$$

$$M(\vec{x}) = 1 + \sum_i m_i M_i(\vec{x}) = 1 + \sum_i m_i \frac{P_i(\vec{x}) - m_{ni}}{m_{ei}} \tag{4.30}$$

where $P_i(\vec{x})$ is a polynom with degree lower than a fixed order.

The coefficients noted by a_{ei} and m_{ei} normalize a_i and m_i to get coefficients evolving in the same dynamic. This is particularly useful while optimizing. The role of a_{ni} and m_{ni} coefficients is to keep the mean of the luminance function. This is made to guarantee a certain *stability* to the transform. More concretely, it can be written by the following conditions (N_Ω refers to the number of voxels in the input image)(see Likar[Likar et al. 2000] for more details)

- the *mean preserving* condition :

$$1/N_\Omega \cdot \sum_{\vec{x}\in\Omega} U(\vec{x}) = 1/N_\Omega \cdot \sum_{\vec{x}\in\Omega} N(\vec{x}) \tag{4.31}$$

- and *parameters normalization* conditions :

$$\begin{aligned} 1/N_\Omega \cdot \sum_{\vec{x}\in\Omega} |N(\vec{x}) \cdot M_i(\vec{x})| &= 1 \\ 1/N_\Omega \cdot \sum_{\vec{x}\in\Omega} |A_i(\vec{x})| &= 1 \end{aligned} \tag{4.32}$$

The mean preserving condition (4.31) can be rewritten as

$$1/N_\Omega \sum_i \left[\sum_\Omega m_i N(\vec{x}) \frac{P_i(\vec{x}) - m_{ni}}{m_{ei}} + \sum_\Omega a_i \frac{P_i(\vec{x}) - a_{ni}}{a_{ei}} \right] = 0 \tag{4.33}$$

To get mean preserving and normalization coefficients independent from m_i and a_i, a possible choice is to take

$$\sum_\Omega N(\vec{x})(P_i(\vec{x}) - m_{ni}) = 0 \quad \forall i \tag{4.34}$$

$$\text{and}$$

$$\sum_\Omega (P_i(\vec{x}) - a_{ni}) = 0 \quad \forall i \tag{4.35}$$

This leads to a really simple expression

$$m_{ni} = \frac{\sum_\Omega N(\vec{x}) P_i(\vec{x})}{\sum_\Omega N(\vec{x})} \tag{4.36}$$

$$\text{and}$$

$$a_{ni} = \frac{1}{N_\Omega} \sum_\Omega P_i(\vec{x}) \tag{4.37}$$

It can be shown[Solanas and Thiran 2001] that the same conditions leads to

$$m_{ei} = \left\{ \begin{array}{ll} \frac{1}{N_\Omega} \sum_\Omega (N(\vec{x})(P_i(\vec{x}) - m_{ni})) & \text{if } M_i(\vec{x}) > 0 \\ \frac{1}{N_\Omega} \sum_\Omega (N(\vec{x})(m_{ni} - P_i(\vec{x}))) & \text{if } M_i(\vec{x}) > 0 \end{array} \right. \tag{4.38}$$

$$a_{ei} = \left\{ \begin{array}{ll} \frac{1}{N_\Omega} \sum_\Omega (P_i(\vec{x}) - m_{ni}) & \text{if } M_i(\vec{x}) > 0 \\ \frac{1}{N_\Omega} \sum_\Omega (m_{ni} - P_i(\vec{x})) & \text{if } M_i(\vec{x}) > 0 \end{array} \right. \tag{4.39}$$

Cost Function

In former section, a bias correction model was introduced. The best a_i and m_i coefficients have now to be chosen by optimizing an appropriate criterion. In the following, we show how a simple criterion derives from a Maximum Likelihood (ML) approach. If a bias correction vector C contains the a_i and m_i parameters, and $\vec{y}$ is defined as a vector containing the gray values y_i of all the pixels in the input image, the best bias correction will be found by a classical ML approach

$$C_{opt} = \arg\max_C \log(p(\vec{y}|C)) \tag{4.40}$$

where p denote the estimated probability density function using a classical estimation technique (histogram, Kernel Estimation,...).

It is now assumed that all the voxel intensities are statistically independent[Leemput et al. 1999]. The probability density distribution of the image $\vec{y}$ for a given model becomes

$$p(\vec{y}|C) = \prod_i p(y_i|C) \tag{4.41}$$

Injecting (4.41) in (4.40) leads to the maximization of

$$\sum_{i=1}^{N_\Omega} \log(p(y_i|C)) \tag{4.42}$$

The sum in (4.42) provides (with an N_Ω factor) an estimation of $\mathrm{E}\{\log(p(y_i|C))\}$. It is now obvious that maximizing (4.42) is equivalent to minimizing the marginal entropy of the $\vec{y}$ image given the model C defined by

$$H(\vec{y}) = -\mathrm{E}\{\log(p(\vec{y}|C))\} \tag{4.43}$$

This leads naturally to the conclusion that the bias field produces a dispersion in the probability distribution of the input image. The use of marginal entropy as criterion to correct this dispersion is an efficient (but radical) way to compensate the bias. Of course, the assumption that all voxel intensities were independent is not really satisfied... To take into account the dependence between intensities of adjacent pixels, other criterions can be introduced.

For example, it is possible to use conditional entropy between voxel values and their neighbor[Solanas and Thiran 2001]. Such strategy avoids divergence in the optimization process (minimizing entropy can lead to change the density probability function into a Dirac delta function !).

Another optimization strategy is the use of conditional entropy between the voxel intensities and the gradient norm of the input image. The gradient takes into account dependencies between a voxel and its neighbor. There is thus sense to use it as an *a priori* knowledge (see Solanas[Solanas and Thiran 2001] fore more details).

4.7.3 Joint Registration and Bias Correction

Cost Function

As described in section 4.7.2, a possible bias correction strategy is the minimization of the marginal entropy in the input image. In the intra-operative

(ODIN) case, another approach could be using the unbiased preoperative image as a reference to correct the bias.

Instead of marginal entropy, the criterion to optimize becomes a robust similarity measure between the preoperative and intra-operative images. It remains now to determine the most suitable criterion. The theory developed by Butz[Butz et al. 2002, Butz and Thiran 2002] about multi-modal signal processing introduces a general framework about robust similarity measures. If X and Y are the features (luminance, norm of luminance gradient, ...) extracted from the source and target image, a class of similarity measures can be derived from the general expression

$$\begin{aligned} e_n(X,Y) &= \frac{MI(X,Y)^n}{H(X,Y)^{1-n}} \quad \text{with}\, n \in [0,1] \qquad (4.44) \\ \text{where} \quad MI(X,Y) &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

In (4.44), $H(X,Y)$ represent the Joint Entropy between features X and Y. The Mutual Information between X and Y is noted by $MI(X,Y)$. Both Joint Entropy and Mutual Information are estimated using a joint histogram between random variables X and Y. It can be observed that equation (4.44) summarizes different well known similarity measures used in image registration.

1. If $n = 0$, maximizing (4.44) is equivalent to minimizing the Joint Entropy between X and Y. Because $H(X,Y) \geq H(X)$ and $H(X,Y) \geq H(Y)$, this approach consist of selecting *efficient* features (i.e. features with low entropy). This is similar to the bias correction approach we discussed in former section.

2. If $n = 1$, maximizing (4.44) is equivalent to maximizing Mutual Information. Butz[Butz et al. 2002] shows that this approach is equivalent to minimizing a lower error bound on the probability error. Nevertheless, it has been observed (see Gil[guez Gil 2002]) that Mutual Information will select features containing as much information as possible. This could increase needlessly the marginal entropy of selected features. If one of these features is the output of a polynomial bias correction filter, this could increase the bias instead of correcting it.

3. If $n = \frac{1}{2}$, a compromise is made between the efficiency of features ($H(X,Y)$) and the minimization of the lower error bound ($MI(X,Y)$). This criterion is equivalent to Normalized Entropy.

> In our case, it seems to be the best compromise to correct the bias and perform a registration simultaneously.

To evaluate the criterion (4.44), an estimation of the joint probability density between X and Y is needed. This is usually performed using joint histograms[Maes and Collignon 1997]. The quantization step in the histogram (i.e. the number of bins) in both axis is an important parameter to be fixed in the registration process. It can be represented by a first feature selection filter before evaluating the criterion. This filter simply consists in a first quantization of the input images in order to select the representative information in both images discarding the redundant noise. Butz[Butz et al. 2002] has shown that an optimal number of bins in both images can be selected with an adapted similarity measure. In as much that decreasing the number of bins obviously decreases the marginal entropies of the image representations, simply maximizing Mutual Information is dangerous and converges to the maximum number of bins (high entropy features). With a Normalized Entropy criterion, a compromise can be made between Mutual Information and low entropy characteristics. This *efficient* behavior explains why Normalized Entropy is preferred to perform a correct quantification in the modalities to register.

Bias and Affine Transformations

If we note by T the spatial transform applied from ODIN image to the diagnosis image, and by P the polynomial transform applied to the gray levels of the ODIN image, the optimization objectives can be formalized as follows

$$[C,\vec{t},n_x,n_y] = \arg \max_{C\in \mathrm{R}^q,\vec{t}\in \mathrm{R}^p,n_x\in \mathrm{Z}^+,n_y\in \mathrm{Z}^+} e_n\{P(O_{n_x}(X)),T^{-1}(D_{n_y}(X))\} \tag{4.45}$$

where X is an uniform Random Variable on the space domain of the ODIN image, O_{n_x} and D_{n_y} denote the luminance functions in the ODIN and diagnosis image quantized with n_x and n_y bins. During the optimization process, the transformation is applied to the diagnosis image and resized to the same domain as the ODIN image.

The spatial transform is implemented in our case by a rigid transform (3 translations and 3 rotations). The polynomial transform we used is multiplicative and linear, 3 m_i independent parameters have to be determined. The number of bins in both images was also optimized. The

S_{pop}	population size	500
p_{conv}	convergence percentage for termination condition	0.001
n_{gen}	maximum number of generations	50
p_{mut}	mutation probability	0.5
p_{cross}	crossover probability	0.7
$p_{replace}$	proportion of indiv. to be replaced at each generation	0.5

Table 4.1: *Genetic Optimization parameters.*

resulting number of parameters to optimize was thus set to 11. For visualization purpose, the spatial transform is inverted before being applied to the ODIN image. A global view of the algorithm is sketched in figure 4.21.

Optimization Method

We investigated the use of different optimization methods. Powell optimization seemed not robust enough to perform a correct registration. Moreover, these kind of optimization algorithm can not deal with a complex search domain. In our case, the number of bins (n_x and n_y) are discrete numbers whereas polynomial and rigid (translation and rotation) coefficients are real parameters. We have finally chosen a genetic algorithm (see Section 4.3.3) and used the GALIB [Wall 1996] library implementing different genetic algorithms. The particular values we used in our implementation are presented in table 4.1.

In our case, an acceptable initial condition is needed due to the limited field of view of the ODIN modality. It is also necessary to get reasonable computation time. The *search domain* for the genetic routine is a (10x10x10)mm box in translation centered on the initial condition. The bounds for polynomial bias correction were fixed to $[-10, 10]$ for each m_i coefficient. The values for p_{mut} and p_{cross} (c.f. table 4.1) are bigger than usual. An important value for p_{mut} avoids convergence to local minima because of exploring the whole *search domain*.

4.7.4 Results

The Genetic Algorithm (figure 4.21) simultaneously optimizes three blocks : the number of bins in the preoperative and intra-operative modalities, the polynomial transform parameters (bias correction) and the geomet-

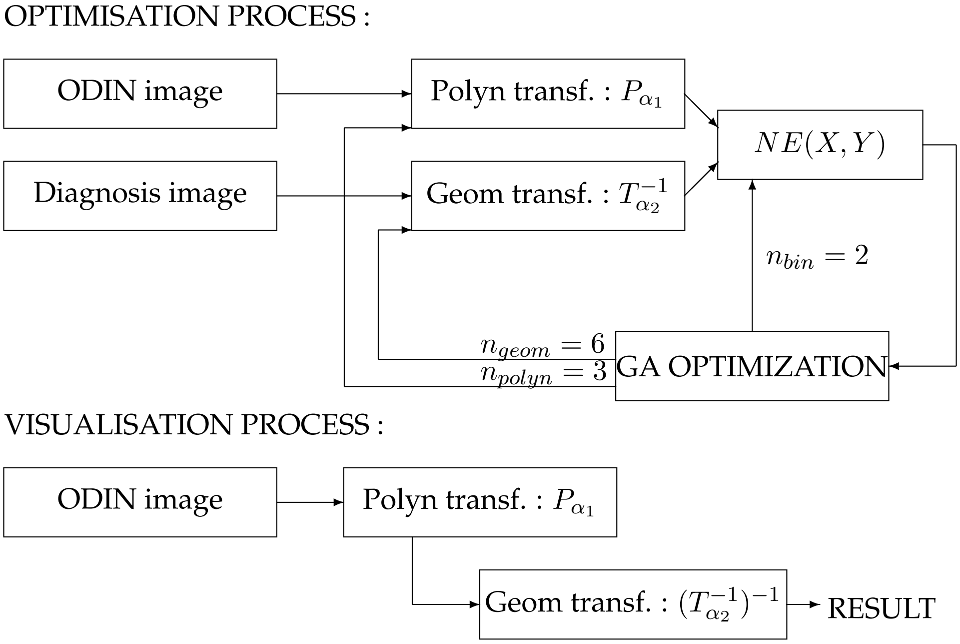


Figure 4.21: *Algorithmic flow chart*
Our algorithm uses a genetic optimization routine to find the best parameters in order to register preoperative and intra-operative modalities (6 parameters), correct the gray levels of the intra-operative image (3 parameters) and find the optimal number of bins to build the joint histogram (2 parameters). Once the algorithm has converged, we apply the luminance and the geometric transform to the intra-operative image. The similarity measure used is the Normalized Entropy between the source (diagnosis image) and the target (ODIN) image.

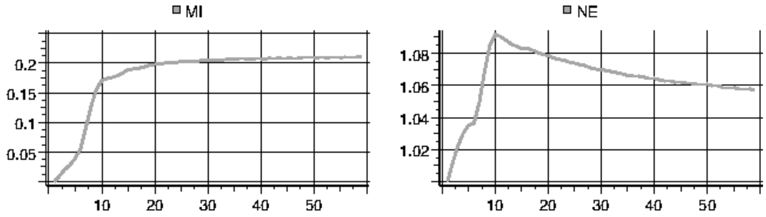


Figure 4.22: *MI/NE comparison for features selection.*
Selection of the numbers of bins in the ODIN image using Mutual Information (MI) or Normalized Entropy (NE). Both MI and NE presents a global optimum but MI converges to the maximum number of bins. The selection is therefore not relevant. On the contrary, NE converges to an intermediate number of bins.

ric parameters (translations and rotations). In this section, the results obtained for each block will be successively discussed.

Selecting the number of bins in both modalities allows to compare two similarity measures : Mutual Information (MI) and Normalized Entropy (NE). After optimization, the selected numbers of bins were fixed to 5 in the diagnosis image and to 10 in the ODIN image. The value of Mutual Information and Normalized Entropy is plotted in figure 4.22 as a function of the number of bins in the intra operative modality (keeping 5 bins in the diagnosis image). It clearly appears that NE presents a maximum for 10 bins in the ODIN image. On the contrary, MI selects always the maximum number of segments and emphasizes the marginal entropy in both images.

To examine if the selected number of bins is anatomically relevant, the corresponding threshold (figure 4.23) is applied to both images. The quantization in the diagnosis image allocates different bins to the ventricles, the brain or the skull. Due to the noise in the intra operative image, the quantization is more confused even if the polynomial correction try to correct the bias. However, this seems enough to converge to a satisfying alignment.

Our algorithm converges to a multiplicative bias correction visible in the middle part of figure 4.24. It can be clearly observed that the correction emphasizes the inferior left corner. Figure 4.25 compares the joint histograms after simple rigid registration and after simultaneous registration and bias correction. It appears that the resulting joint histogram after simultaneous optimization strategy is more compact.

Once the polynomial and the rigid registration applied to the ODIN

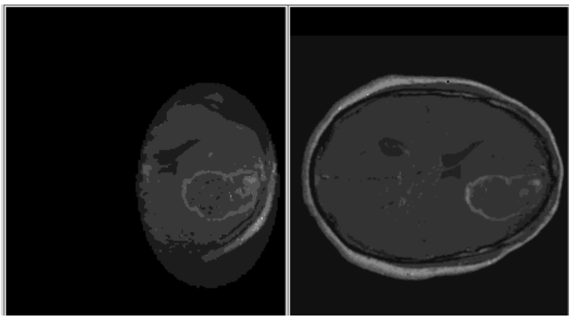


Figure 4.23: *Optimal number of bins joint selection using Normalized Entropy.*
Relevant objects are isolated (brain, ventricles, skull) emphasizing common information between the involved modalities.

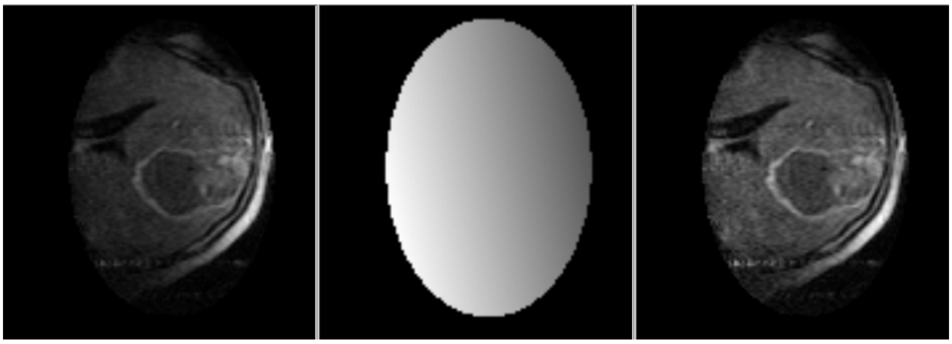


Figure 4.24: *Optimal bias correction using Normalized Entropy.*
A view of the bias correction obtained by our algorithm. We used a multiplicative and linear correction to correct the bias in the intra-operative modality. The coefficient of the polynomial correction in the gray levels and the geometric transform parameters were obtained by maximization of a similarity measure with a standard MR image of the same patient.

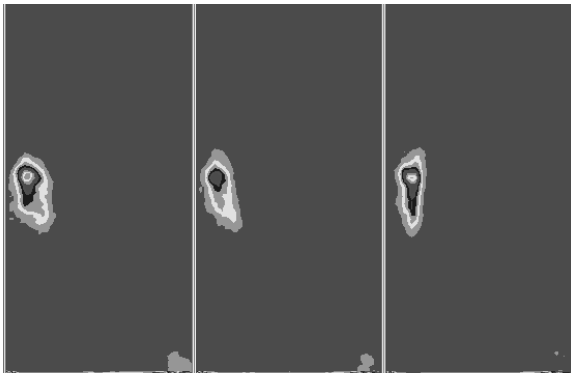


Figure 4.25: *Joint histograms before and after alignment.*
The first histogram we show is the initial joint histogram between the gray levels in the two images (ODIN and diagnosis MR image). The second histogram is obtained after rigid registration without bias correction. The third one is the joint histogram we get after simultaneous registration and bias correction of the ODIN and the diagnosis image. We see that in our case, the bias correction is necessary to avoid dispersion in the joint distribution.

image, we obtained the result showed in figure 4.26 (the contours of the preoperative image are over-imposed to the corrected ODIN image). It can be observed that the tumor and the ventricles are well registered. However, near the boundaries of the ODIN image, geometric distortions appear. More data would be needed (for example phantom images) to evaluate the importance of these distortions.

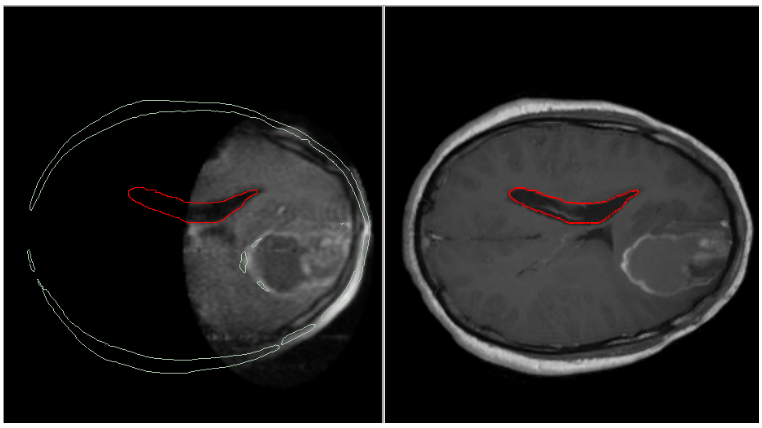


Figure 4.26: *Contours of the preoperative image reported on the intra-operative (ODIN) image.*

Multi-subject Registration for Unbiased Probabilistic Atlas Generation

5

This chapter describes an original multi-subject registration algorithm. We first review the existing probabilistic atlas construction methods. A short review of the EM technique (which has numerous applications in the signal processing field) is then given in Section 5.2. The STAPLE algorithm introduced by Warfield for the validation of image segmentation and expert performances evaluations is described in Section 5.3. The next section shows how the STAPLE framework can be used for multi-subjects alignment by addition to the model of a vector of transformations mapping each subject on the probabilistic atlas which is in this framework the hidden variable of the EM algorithm. We present results on a data base of 80 brains segmented into 4 labels (white matter, gray matter, ventricles and background).

5.1 Existing Probabilistic Atlas Construction Methods

Over the last century, the construction of brain atlases has been a central issue to the understanding of brain anatomy. The first atlases were obtained from post-mortem acquisitions. With the emergence of three dimensional imaging modalities such as Magnetic Resonance images, research has been directed towards the development of 3D digital atlases. So far, most of the available atlases are based on single subjects. The work of Talairach and Tournoux [1998] defines a 3-dimensional proportional system for comparing brains of different sizes. They organize the brain into areas delimited by anatomical plans and define proportional rules to map coordinates acquired on the subject to atlas coordinates. The Surgical Planning Laboratory atlas [Kikinis, R. 1996] is obtained from a Magnetic Resonance image manually segmented by medical experts.The atlas labels are organized into hierarchical categories.

Although these atlases provide a standard system of coordinates, they lack to illustrate the variability of anatomical structures since a single subject can not represent structural variabilities between individuals.

For characterizing this variability, an intense research effort has been dedicated to the development of volumetric registration algorithms mapping a single reference subject on a population. [Dinggang and Davatzikos 2002, Miller et al. 99, Toga 1999]. A statistical analysis of the transformations can also be used to characterize different populations[Csernansky et al. 1998, Joshi et al. 2003, Thompson et al. 1998]. This approach brings the problem of multi-subjects registration back to a pairwise registration problem.The drawback of this technique is to introduce a bias in the choice of the reference subject.

The problem of building unbiased atlases has therefore generated a wide interest in the medical imaging research community. Guimond et al. [1999] developed an iterative algorithm for reducing the bias : the average of displacement fields is computed and applied to the reference to generate an unbiased reference. Marsland et al. [2003] has proposed to change the reference subject over the alignment process. A distance is defined between each pair of subjects by composing transformations through the current reference. At each step, the reference is updated as the subject minimizing the sum of distances to the other subjects. Bhatia et al. [2004] proposed to pick one subject as a reference subject for intensities and computes a joint histogram by incrementing all bins corresponding to the following realizations

$$\text{(intensity in the reference subject, intensity in subject } i\text{)}$$

for $i = 1 \ldots N_{subjects} \backslash \{i_{ref}\}$. In this scheme, the transformation parameters for all subjects are optimized simultaneously. Zollei et al. [2005] also performs a joint optimization of transformation parameters across all subjects and uses the sum of voxel-wise entropies as a joint alignment criterion.

Joshi et al. [2004] computes a reference subject at each step by averaging all subjects for the current alignment. The velocity field for each subject is computed at each iteration by applying the inverse of a Navier-Stokes differential operator to the force function. The force function is simply the optical flow force between the reference and the subject.

5.2 The EM algorithm

EM is an iterative method to estimate some unknown parameters $\boldsymbol{\theta}$ given some measurement data $\boldsymbol{D}$. The model also includes hidden (also called nuisance) variables $\boldsymbol{W}$ which need to be integrated out. A first choice for the $\boldsymbol{\theta}$ parameters is to choose for the one optimizing a posterior probability of the parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{D}$ marginalized over the hidden variable $\boldsymbol{W}$

$$\boldsymbol{\theta}_{opt} = \arg\max_{\theta} \ln \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}) \right) \tag{5.1}$$

The idea behind EM is to iteratively estimate a distribution of the hidden variable conditioned over the observations and to use this posterior distribution to integrate the complete likelihood function.

There are many ways of deriving the EM approach, but one of the most intuitive being in term of lower bound maximization [Dellaert 2002, Minka 1998]. The idea behind EM is to start with a guess $\boldsymbol{\theta}^t$ for the parameters $\boldsymbol{\theta}$, compute a lower bound $B(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$ to the cost function in Equation (5.1) and maximize this lower bound instead.

The key problem in Equation (5.1) is the presence of of the logarithm of a big sum, which is difficult to deal with, especially for computing analytically the zero of the first derivative. Fortunately, we can construct a lower bound $B(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$ that contains a sum of logarithms. For deriving this bound, an arbitrary probability distribution $f^t(\boldsymbol{W})$ is introduced over the space of the hidden variable

$$\ln \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}) \right) = \ln \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} \frac{f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta})}{f^t(\boldsymbol{w})} f^t(\boldsymbol{w}) \right) \tag{5.2}$$

By using Jensen's inequality [Dellaert 2002],

Jensen's inequality.
If f is a convex function and X a random variable, then

$$Ef(X) \geq f(EX)$$

we can define a lower bound by

$$B(\boldsymbol{\theta}, \boldsymbol{\theta}^t) \overset{\text{def}}{=} \sum_{\boldsymbol{w} \in \boldsymbol{W}} f^t(\boldsymbol{w}) \ln \frac{f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta})}{f^t(\boldsymbol{w})} \leq \ln \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} \frac{f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta})}{f^t(\boldsymbol{w})} f^t(\boldsymbol{w}) \right) \tag{5.3}$$

5.2.1 Optimal Bound

An optimal bound can be derived by computing the optimal distribution $f(\boldsymbol{w})$ *for the current guess of* $\boldsymbol{\theta}^t$. For computing this bound [Dellaert 2002, Minka 1998], the constrain that $\sum_{\boldsymbol{w}} f^t(\boldsymbol{w}) = 1$ is added to the cost function using a Lagrange multiplier

$$\begin{aligned} C(f^t(\boldsymbol{w})) &= B(\boldsymbol{\theta}^{\boldsymbol{t}}, \boldsymbol{\theta}^t) - \lambda \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} f^t(\boldsymbol{w}) - 1 \right) \qquad (5.4) \\ &= \sum_{\boldsymbol{w} \in \boldsymbol{W}} f^t(\boldsymbol{w}) \ln f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}^t) - \sum_{\boldsymbol{w} \in \boldsymbol{W}} f^t(\boldsymbol{w}) \ln f^t(\boldsymbol{w}) \\ &\quad -\lambda \left(\sum_{\boldsymbol{w} \in \boldsymbol{W}} f^t(\boldsymbol{w}) - 1 \right) \qquad (5.5) \end{aligned}$$

Maximizing Equation (5.5) following $f(\boldsymbol{w})$ and λ yields the following solution

$$f^t(\boldsymbol{w}) = \frac{f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}^t)}{\sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}^t)} = f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^t) \qquad (5.6)$$

The optimal bound used by the EM algorithm is then

$$B(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}}) \ln \frac{f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta})}{f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}})} \qquad (5.7)$$

It is easy to see that$B(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$ touches the objective function in $\theta = \theta^t$:

$$B(\boldsymbol{\theta}^t, \boldsymbol{\theta}^t) = \sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}}) \ln f(\boldsymbol{D} | \boldsymbol{\theta}^{\boldsymbol{t}}) = \ln f(\boldsymbol{D} | \boldsymbol{\theta}^{\boldsymbol{t}}) \qquad (5.8)$$

which is identical to Equation (5.2).

Since $B(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$ can be decomposed in two terms

$$B(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = Q^t(\boldsymbol{\theta}^{\boldsymbol{t}}, \boldsymbol{\theta}^t) - \mathcal{H} \qquad (5.9)$$

where

$$\begin{aligned} Q^t(\boldsymbol{\theta}, \boldsymbol{\theta}^{\boldsymbol{t}}) &\overset{\text{def}}{=} \sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}}) \ln f(\boldsymbol{w}, \boldsymbol{D} | \boldsymbol{\theta}) \qquad (5.10) \\ &= E_W \left\{ \ln f(\boldsymbol{W}, \boldsymbol{D} | \boldsymbol{\theta}) | \, \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}} \right\} \qquad (5.11) \end{aligned}$$

and

$$\mathcal{H} \overset{\text{def}}{=} \sum_{\boldsymbol{w} \in \boldsymbol{W}} f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}}) f(\boldsymbol{w} | \boldsymbol{D}, \boldsymbol{\theta}^{\boldsymbol{t}}) \qquad (5.12)$$

which is not depending in $\boldsymbol{\theta}$, maximizing $B(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$ is equivalent to maximizing $Q^t(\boldsymbol{\theta}, \boldsymbol{\theta^t})$. Therefore the EM algorithm can be summarized as

```
while convergence is not reached, perform
    (E Step) calculate $f^t(\boldsymbol{w}) = f(\boldsymbol{w}|\boldsymbol{D}, \boldsymbol{\theta^t})$
    (M Step) calculate $\boldsymbol{\theta}^{t+1} = \arg\max_{\boldsymbol{\theta}} Q^t(\boldsymbol{\theta}, \boldsymbol{\theta^t})$
```

5.3 The STAPLE Algorithm

STAPLE [Warfield et al. 2004] is an algorithm for characterizing the performances of a set of experts or algorithms segmenting the same anatomical structures. It can be used for evaluating intra-observers and extra-observers variabilities. The expert performances are characterized by transitional probabilities for an expert to classify a pixel as belonging to the class s' when the true label is s.

The evaluation of this conditional probabilities require to estimate a ground truth which is by definition not an observable variable. For this reason, the EM algorithm shortly introduced in Section 5.2, is an ideal framework for iteratively estimating the performance parameters characterizing each expert and simultaneously estimating the hidden distribution of classes at each voxel.

As seen is Section 5.2, the EM algorithm is an iterative strategy for optimizing a complete likelihood cost function

$$\ln f(\boldsymbol{W}, \boldsymbol{D}|\boldsymbol{\theta}) \tag{5.13}$$

where in our case, $\boldsymbol{D}$ is the observed variable of expert decisions (set of segmentations), $\boldsymbol{W}$ is the hidden variable represented by a probability density of each tissue class at each voxel, $\boldsymbol{\theta}$ contains the set of conditional probabilities describing the expert performances.

5.3.1 E Step

At each iteration of the algorithm, two steps are performed. The first called *expectation step* (E Step) estimates the *a posteriori* density of the hidden variable conditioned over the observations of the expert decisions and the current estimate of the expert performance parameters. Typically the E step uses the Bayes law to write the posterior probabilities

$$p(\text{hidden variable} \mid \text{observations})$$

as a function of

$$p(\text{observations} \mid \text{hidden variable})$$

which are related to the model used to compute the likelihood cost function. In our context, the hidden variable density function W_{si} giving the probability to find the tissue class s at voxel i is obtained from the conditional probabilities θ

$$W_{si} = \frac{1}{C_i} f(A_i = s) \prod_{j}^{N_{\text{subjects}}} \hat{\theta}^{j}_{D_{ij}|s} \tag{5.14}$$

where $\theta^{j}_{s'|s}$ is defined as

$$\theta^{j}_{s'|s} = \text{prob. for expert } j \text{ to say } s' \text{ when the true label is } s \tag{5.15}$$

The constant C_i is fixed so that the W_{si} at each voxel are summed to 1

$$C_i = \sum_{s'} f(A_i = s') \prod_{j}^{N_{\text{subjects}}} \hat{\theta}^{j}_{D_{ij}|s'} \tag{5.16}$$

5.3.2 M Step

The second step called *M Step* solves for the θ parameters optimizing a modified likelihood score (5.11)

$$\begin{aligned} \boldsymbol{\theta}_{opt} &= \arg\max_{\boldsymbol{\theta}} E_W \left\{ \ln f(\boldsymbol{W}, \boldsymbol{D}|\boldsymbol{\theta}) \middle| \boldsymbol{D}, \boldsymbol{\theta^t} \right\} & (5.17) \\ &= \arg\max_{\boldsymbol{\theta}} E_W \left\{ \ln f(\boldsymbol{D}|\boldsymbol{W}, \boldsymbol{\theta}) f(\boldsymbol{W}) \middle| \boldsymbol{D}, \boldsymbol{\theta^t} \right\} & (5.18) \end{aligned}$$

where it we have assumed in Equation (5.18) that the hidden variable was independent on the $\boldsymbol{\theta}$ parameters i.e. $f(\boldsymbol{W}|\boldsymbol{\theta}) = f(\boldsymbol{W})$. Since $f(\boldsymbol{W})$ is independent on $\boldsymbol{\theta}$, optimizing Equation (5.18) is equivalent to optimize

$$\begin{aligned} \boldsymbol{\theta}_{opt} &= \arg\max_{\boldsymbol{\theta}} E_W \{ \ln f(\boldsymbol{D}|\boldsymbol{W}, \boldsymbol{\theta}) \} & (5.19) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{i}^{N_{\text{voxels}}} \sum_{j}^{N_{\text{subjects}}} E_{\boldsymbol{W}_i} \ln f(D_{ij}|\boldsymbol{W}_i, \boldsymbol{\theta}) & (5.20) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{i}^{N_{\text{voxels}}} \sum_{j}^{N_{\text{subjects}}} \sum_{s}^{N_{\text{labels}}} W_{si} \cdot \ln f(D_{ij}|\boldsymbol{W_i} = s, \boldsymbol{\theta}) & (5.21) \end{aligned}$$

Since all subjects are considered as independent, optimizing the cost function in Equation (5.21) can be done separately for each subject and the sum over j can therefore be considered term by term. The cost function to optimize for each subject is then

$$\sum_{s}\sum_{s'}\sum_{i:D_{ij}=s'} W_{si}\ln(\theta_{s's}) \tag{5.22}$$

We must associate to this cost function the constrain that θ has to be summed to 1 over s'. This can be done by adding a Lagrange multiplier to Equation (5.22).

$$\sum_{s}\sum_{s'}\sum_{i:D_{ij}=s'} W_{si}\ln(\theta_{s's}) + \lambda\cdot\left(1-\sum_{s'}\theta_{s's}\right) \tag{5.23}$$

Taking the derivative of Equation (5.23).

$$\sum_{i:D_{ij}=s'} W_{si}\frac{1}{\theta_{s's}} - \lambda = 0 \tag{5.24}$$

$$\Rightarrow \theta_{s's} = \sum_{i:D_{ij}=s'} W_{si}/\lambda \tag{5.25}$$

From the constrain $\sum_{s'}\theta_{s's} = 1$, it comes that $\lambda = \sum_i W_{si}$ and then

$$\theta_{s's} = \frac{\sum_{i:D_{ij}=s'} W_{si}}{\sum_i W_{si}} \tag{5.26}$$

The EM algorithms iterates until convergence between Equation (5.14) and Equation (5.26) until convergence. The convergence can be measured by comparing at two successive iterations the values of the θ^j matrix for each expert. In practice, averaging over all experts the trace of θ^j is a good indicator of convergence [Warfield et al. 2004].

5.4 Multi-subjects Registration Algorithm

In this section, we examine how the STAPLE algorithm can be adapted for performing the joint alignment of a collection of segmentations of the same tissue classes acquired from different subjects. The optimization process described in Section 5.3 must therefore take into consideration an additional set of parameters modeling the transformations aligning each subject on the current version of the hidden variable estimate. In this case, the hidden variable W represents a probabilistic atlas giving at each voxel (and for the current alignment), the probability of having a tissue class in the population.

Two families of transformations have been considered in our study : affine transformations and free-form displacement fields. The optimization procedures for these different transformation models are related to Chapters 3 and 4.

5.4.1 E Step

If the parameters vector θ is augmented by a vector of parameters modeling the transformation of each subject subject on the atlas, the tissue class at each voxel depends on the current transformation for this subject. Each D_{ij} value has then to be rewritten like

$$D_{ij} = D_j(T^j(i)) = D_j(T_i^j) \tag{5.27}$$

In the loop of the STAPLE algorithm, the *E step* performed at each iteration remains unchanged since all the parameters to optimize are kept constant during this phase. Equation (5.14) is still valid and computes at each iteration of the EM algorithm the atlas weights W_{si}.

5.4.2 M Step

If we get back to Equation (5.26), we can rewrite it using the notation of Equation (5.27) and a Kronecker delta for having the sum over pixels covering all the pixel indexes.

$$\hat{\theta}_{s'|s}^{j,T^j} = \frac{\sum_i W_{si} \cdot \delta(D_j(T^j(i)) - s')}{\sum_i W_{si}} \tag{5.28}$$

When it comes to include the transformation as parameters in the optimization procedure, a critical point is to have continuous changes in the cost function for continuous changes in the set of parameters. Since the cost function (Equation (5.23)) includes the set of θ parameters, a natural choice for ensuring continuity of the cost function is to substitute the Kronecker function in Equation (5.28) by a continuous kernel β (typically a Gaussian or BSpline) :

$$\hat{\theta}_{s'|s}^{j,T^j} = \frac{\sum_i W_{si} \cdot \beta(D_j(T^j(i)) - s')}{\sum_i W_{si}} \tag{5.29}$$

We can also remark that if the solution of $\theta^j_{s'|s}$ is injected into the cost function to optimize for each subject, Equation (5.23) can be rewritten as

$$\begin{aligned} H(D_j|A) &= \sum_{ss'} \theta^j_{s'|s} \cdot \ln(\theta^j_{s'|s}) \cdot \sum_i W_{si} \qquad (5.30) \\ &= \sum_{ss'} \theta^j_{s'|s} \cdot \theta_s \cdot \ln(\theta^j_{s'|s}) \qquad (5.31) \end{aligned}$$

where we have defined

$$\theta_s \overset{\text{def}}{=} \sum_i W_{si} \qquad (5.32)$$

The criterion in Equation (5.30) is by definition the conditional entropy of the subject over the atlas for the current alignment[1]. For a given transformation, the θ parameters optimizing the cost function given by Equation (5.21) can be computed from Equation (5.28). Since no close form of the cost function Equation (5.30) is available for the transformation parameters, an iterative optimization method must be used.

However, it can be seen that minimizing Equation (5.30) as it is can lead the optimization process to diverge. The following inequality

$$0 \leq H(D_j|A) \leq H(D_j) \qquad (5.33)$$

shows that $H(D_j|A)$ can be minimized simply by minimizing its upper bound $H(D_j)$. This means that a set of transformations moving all subjects outside the atlas domain would make tend the entropy of each subject to 0 and therefore would minimize the criterion in Equation (5.30) without yielding a satisfactory solution. This undesired solution can be discarded by adding the marginal entropy of the subject for the current transformation to the cost function. The criterion to optimize is thus the mutual information between the subject and the probabilistic atlas

$$MI_j = H(D_j) - H(D_j|A) \qquad (5.34)$$

where $H(D_j)$ is measured using the marginal histogram of the subject labelled image. The general optimization scheme is then

---

[1] An additional normalization of θ_s is required for having θ_s representing a probability but this only affects the cost function by a multiplicative constant.

(A) Compute the W_{si} for the current alignment

(B) Align each subject on the current probabilistic atlas

- ▷ for each transformation, compute θ^j using Equation (5.28)
- ▷ compute MI_j (Equation (5.34))
- ▷ iterate until convergence

(C) If W_{si} has not converged yet, get back to **(A)**

5.4.3 Multi-Subjects Rigid Alignment

In a first time, a rigid and scale deformation model is used to increase the similarity between the atlas and each subject. This transformation model allows 9 degrees of freedom for rotations, translations and scaling. To speed up the convergence, the set of transformations is initialized by translations mapping the center of gravity of each subject segmentation to the center of the common space of coordinates. For this transformation model, we use the SPSA optimization method introduced by Spall[Spall 1998].

5.4.4 Multi-Subjects Non-Rigid Alignment

For maximizing Equation (5.34), an iterative variational displacement field is derived in a similar way as described in Chapter 3.

At each iteration,the current displacement field $v(i)$ is updated by computing a $\Delta v(i)$ proportional to

$$\begin{aligned} \nabla D_j(T(i)) \cdot \sum_{ss'} -W_{si} \quad & \cdot \quad \ln \frac{p(A_i = s, D_j(T(i)) = s')}{p(D_j(T(i)) = s')} \\ & \cdot \quad \beta'(D_j(T(i)) - s') \end{aligned} \tag{5.35}$$

This equation is similar to Eqs. (3.17) and (3.18) but the β^1 box function over one bin of the fixed images is replaced by the weighting function W_{si}.

The presence of $\nabla T(i)$ in Eq.5.35 reminds a classical optical flow incremental field estimation. Instead of being multiplied by the difference of intensity functions in fixed and moving images, the gradient in the subject (moving image) is multiplied by a weighted function of $W_{si} \cdot \beta'(D_j(T(i)) - s')$. The weights are proportional to joint and marginal probability densities measured over the entire atlas domain. This proves that mutual information (in opposition to mean squared differences) acts a global metric, even for estimating local displacements, taking into account global statistical properties of both images to be aligned.

5.5 Experiments

The multi-subject registration algorithm has been run on 80 segmentations ($256 \times 256 \times 124$ voxels) of adult brains containing 4 labels : background, ventricles, white and gray matter). The set of transformations has been initialized by translations bringing the center of gravity of each subject to the center of the atlas domain.

For aligning such a large collection of subjects, a slight variant of the multi-subject alignment algorithm described in Section 5.4 has been implemented. Since the $\theta^j_{s'|s}$ have a close form solution in the M step (Equation (5.28)), the parameters of the transformation are first held constant and the usual STAPLE algorithm iterates between the E steps and M steps described in Section 5.3. In a second time, each subject is aligned on the W_{si} probabilistic atlas using first a rigid and scale transformation. Once the rigid alignment procedure has reached convergence, a non-parametric displacement field is estimated following Equation (5.35).

5.5.1 Rigid Alignment

Two iterations (each iteration includes (a) atlas and expert parameters estimation and (b) sequential subject to atlas alignment) were used to find the optimal parameters of a rigid and scale transformation model parameterized by 9 parameters. In our transformation model, the scaling parameters are mapped to the $(-\inf, +\inf)$ domain by optimizing their logarithm (this avoids to have different search domains for each parameter). The alignment of a subject on the current atlas estimate takes less than 5 minutes. The SPSA optimization routine and the mutual information metric for probabilistic atlas to image registration have been integrated in the Insight Toolkit[Ibanez and Schroeder 2005].

5.5.2 Non-Rigid Alignment

For non-rigid registration, the flow in Eq.5.35 has been integrated in the ITK[Ibanez and Schroeder 2005] dense multi-threaded non-rigid registration framework. The displacement field is smoothed at each iteration by a Gaussian kernel. Figure 5.2 plots the sum of segmentations weighted by the probabilistic atlas after rigid (left) and non-rigid (right) registration. Each label is plotted in a different channel (R,G,B). Bright color on this figure show areas where the subject labels and the probabilistic atlas where in accordance. It can be seen that the volume of 'bright' white matter has increased after non-rigid registration. Figure 5.1 plots a simple unweighted sum of the labels across all subjects after rigid (left) and non-rigid (right) alignment. The contours of the ventricles in the sum after non-rigid alignment appear more precisely after non-rigid alignment.

To figure out how smooth the transformation is, Figures 5.3 and 5.4 plot the same 9 subjects after affine and non-rigid alignment for a sagittal slice. Figures 5.5 and 5.6 show the same subjects as in Figures 5.3 and 5.4 but in a coronal view. It appears clearly on these figures, that the non-rigid transformations do not not affect the topology of the white matter matter ramifications but tend to better align the external contour of the brains and ventricles. Figure 5.1 show also that some small clusters of white matter appear more intense after non-rigid alignment. This show that the number of coherent white matter pixels across all subjects increase with the non-rigid alignment of these subjects on the probabilistic atlas.

5.5.3 Deformations Principal Components Analysis

To characterize the principal deformations modes mapping the probabilistic atlas on each subject in the data base, we run an Principal Component Analysis (PCA) on the collection of non-rigid displacement fields resulting from the MI flow computation (Equation (5.35).

Since the dimension of the data is enormous ($3 \times 256 \times 256 \times 124$), some precautions must be taken in the implementation.

If we first define the U matrix as the juxtaposition of the 80 displacement fields

$$U = \begin{pmatrix} v^1_{1x} & \cdots & v^{80}_{1x} \\ v^1_{1y} & \cdots & v^{80}_{1y} \\ v^1_{1z} & \cdots & v^{80}_{1z} \\ \vdots & \vdots & \vdots \\ v^1_{N_V x} & \cdots & v^{80}_{N_V x} \\ v^1_{N_V y} & \cdots & v^{80}_{N_V y} \\ v^1_{N_V z} & \cdots & v^{80}_{N_V z} \end{pmatrix} \tag{5.36}$$

where the superscript number refers to the subject $(1 \cdots 80)$, the first index covers the range of voxels $(1 \cdots N_V)$ and the second index refers to the x, y and z dimensions of the displacement vector.

A direct computation of the deformation modes would require to compute the eigen values and eigen vectors of the UU^t matrix. Such computation is not affordable for stability and memory reasons. To bypass this issue, we can look at the eigen vectors and eigen values of the U^tU matrix. If λ and ν are linked by

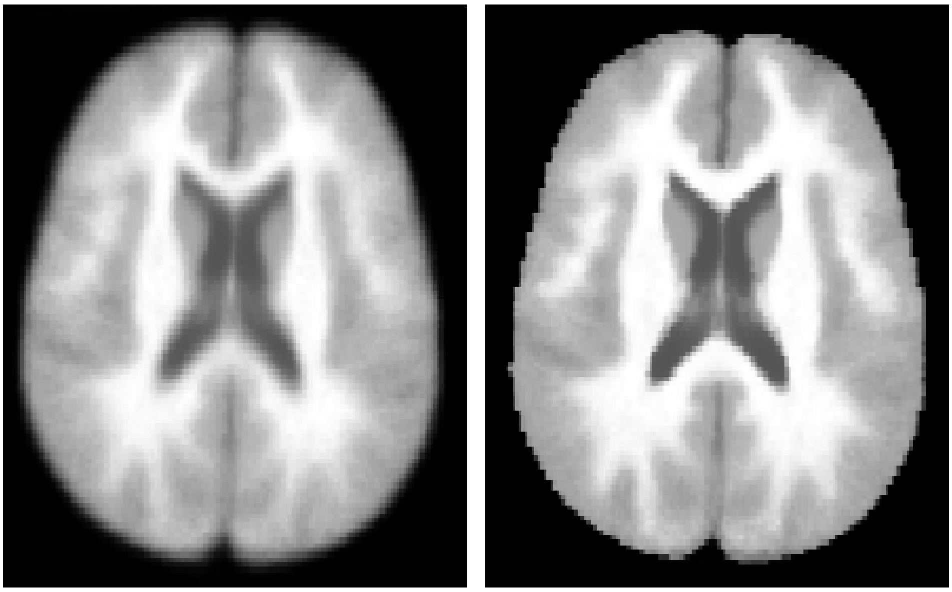
$$U^tU \cdot \nu = \lambda\nu \tag{5.37}$$

we can pre-multiply both members of this equality by U to get

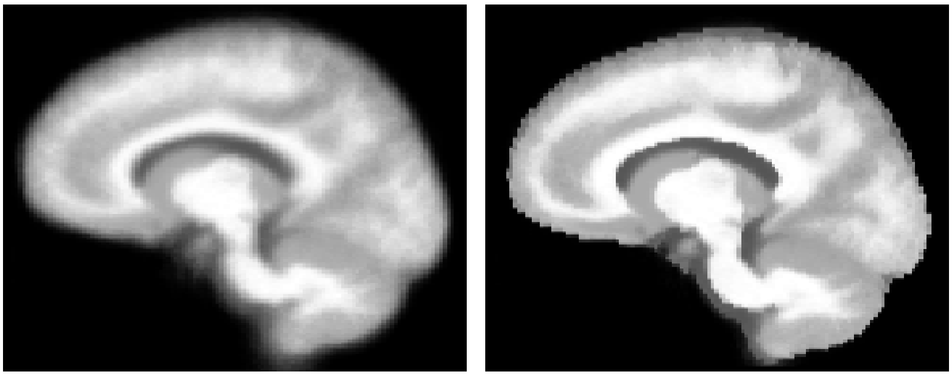
$$(UU^t) \cdot U\nu = \lambda U\nu \tag{5.38}$$

This equation shows that, if λ and ν are eigen value and vector of U^tU, then λ and $U\nu$ are eigen value and vector of UU^t. Since the dimensions of U^tU is 80 × 80 (the number of subjects in the data base), the computation has a reasonably low complexity. Moreover, the $U\nu$ product can be implemented to spare memory.

Figure 5.7 shows the first four modes of variations of the 80 displacements fields. The z component of the arrows are coded accordingly to the color bar. These deformations modes could be used to accelerate the alignment of a new subject on the data base or in an atlas based segmentation algorithm.



(rigid & scale sum axial) (non-rigid sum axial)



(rigid & scale sum sagit.) (non-rigid sum sagit.)

Figure 5.1: *Sum of the 80 segmentations (four labels) after rigid and scale alignment and after 2 iterations of non-rigid warping.*
The value plotted at each pixel is simply $\sum_j D_j(T_j(i))$. *The regions of white matter high variability tend to be more peripheral after non-rigid alignment.*

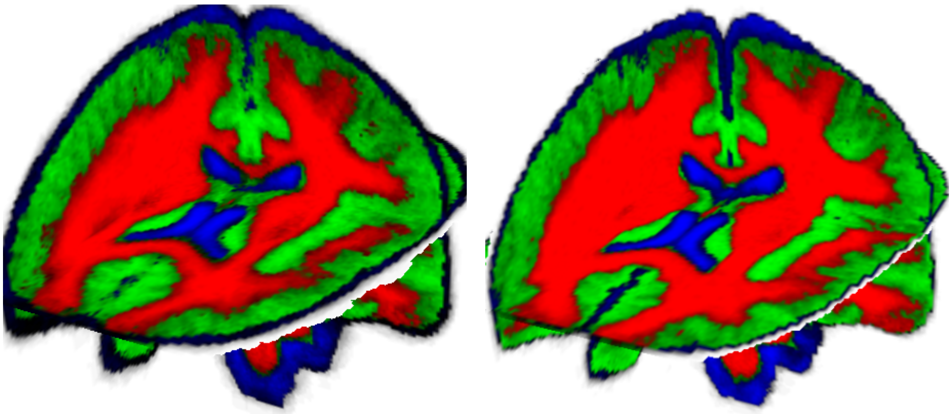
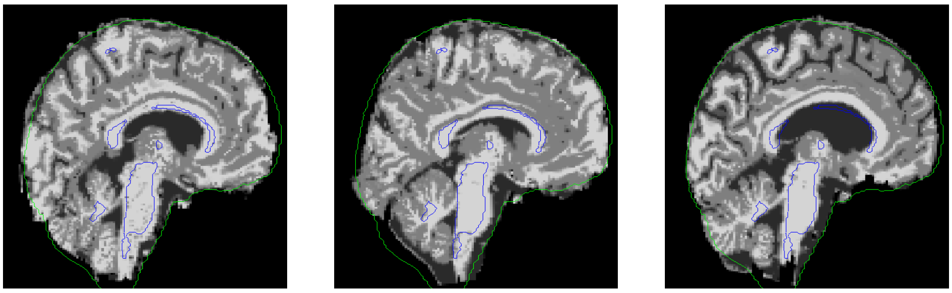
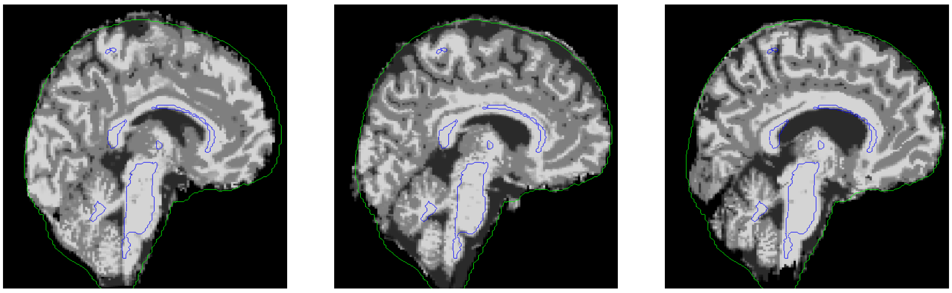

(weighted sum rigid & scale) (weighted sum non-rigid)

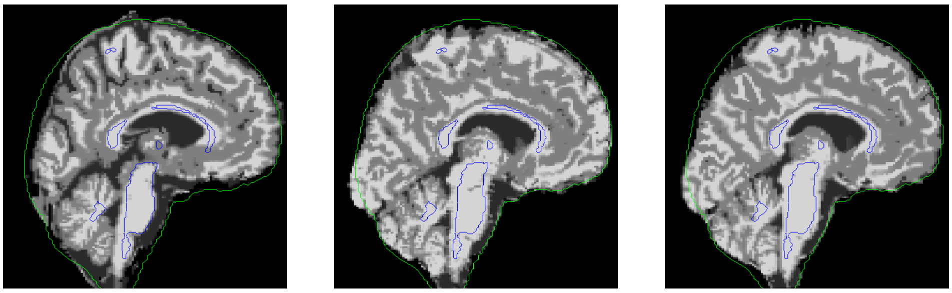
Figure 5.2: *Weighted sum.*
In this image, each subject voxel segmentation is summed in separate channels. Each subject label is weighted by the probability value of the atlas. Thus the value plotted at each pixel is $\sum_j W_{si}\mathrm{Channel}\{D_j(T_j(i))\}$*. Bright colors mean that the subject labels were coherent with the atlas at this point.*



(case 10 rigid & scale) (case 14 rigid & scale) (case 29 rigid & scale)

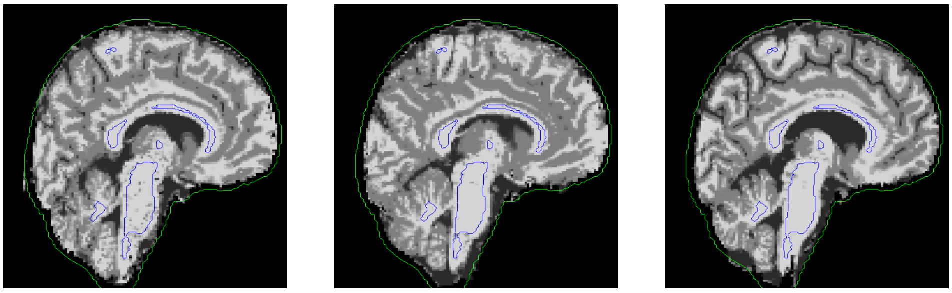


(case 3 rigid & scale) (case 33 rigid & scale) (case 53 rigid & scale)

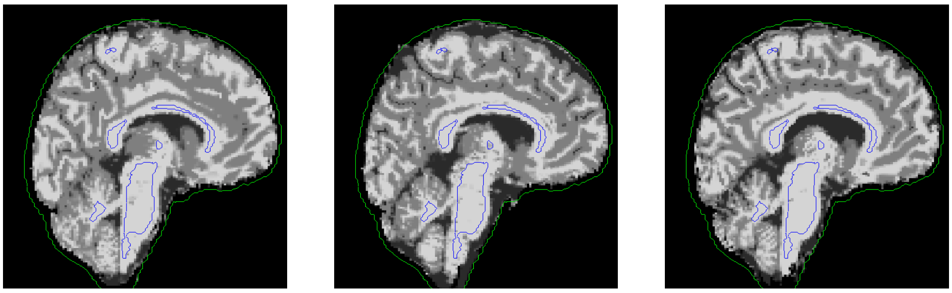


(case G54 rigid & scale) (case 57 rigid & scale) (case G24 rigid & scale)

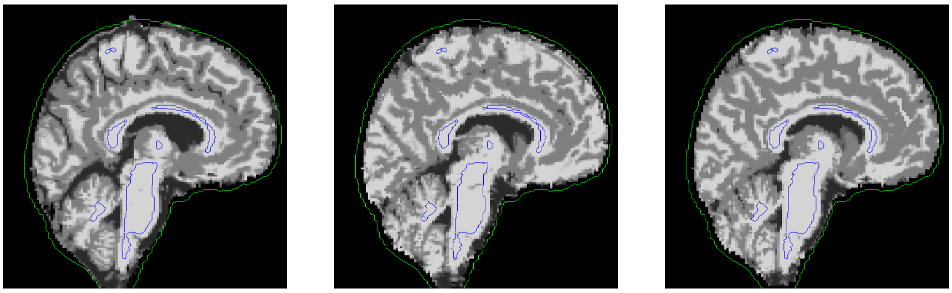
Figure 5.3: *Rigid and scale alignment of typical subjects (sagittal slice).. This figure shows the alignment of 9 subjects (chosen randomly in the 80 subjects) after rigid and scale alignment. The contours of the resulting probabilistic atlas are show in blue (white matter border) and red (brain-background border).*



(case 10 non-rigid) (case 14 non-rigid) (case 29 non-rigid)

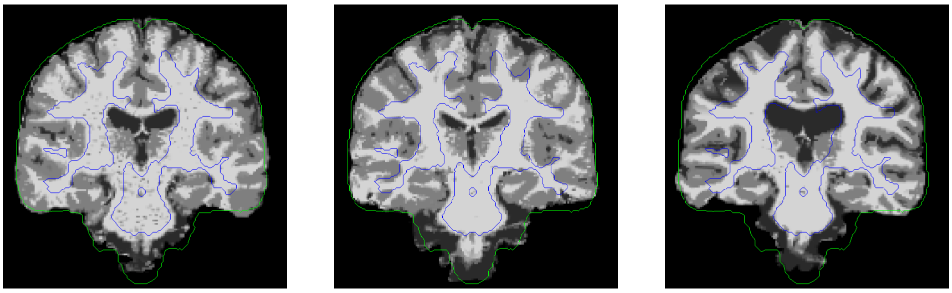


(case 3 non-rigid) (case 33 non-rigid) (case 53 non-rigid)

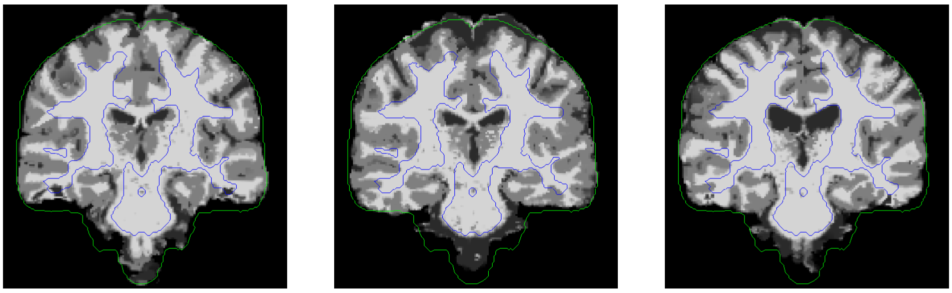


(case G54 non-rigid) (case 57 non-rigid) (case G24 non-rigid)

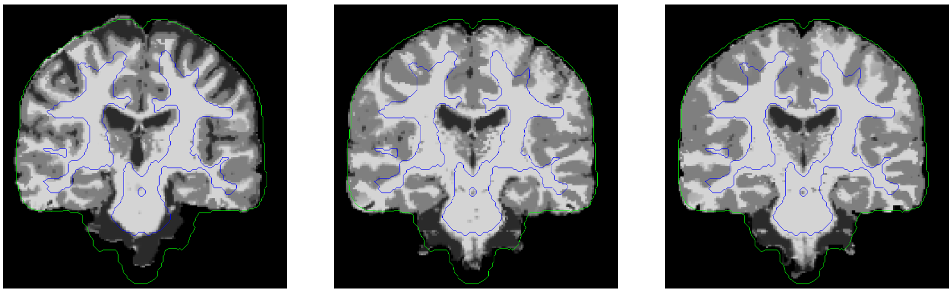
Figure 5.4: *Non-rigid alignment of typical subjects (sagittal slice).This figure shows the alignment of the same 9 subjects as in Figure 5.3 after non-rigid alignment. The contours drawn in blue and red are identical to the contours shown in Figure 5.3.*



(case 10 rigid & scale) (case 14 rigid & scale) (case 29 rigid & scale)

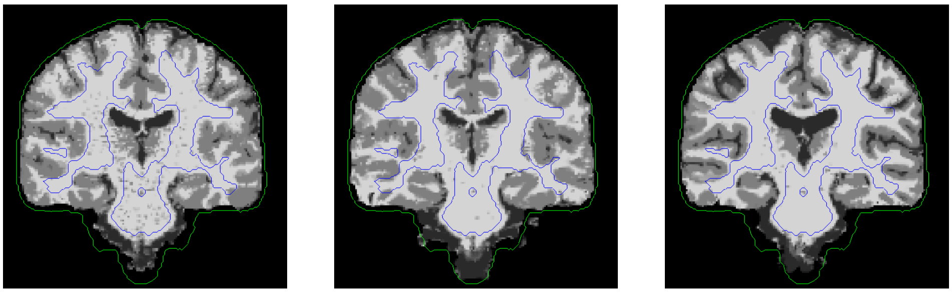


(case 3 rigid & scale) (case 33 rigid & scale) (case 53 rigid & scale)

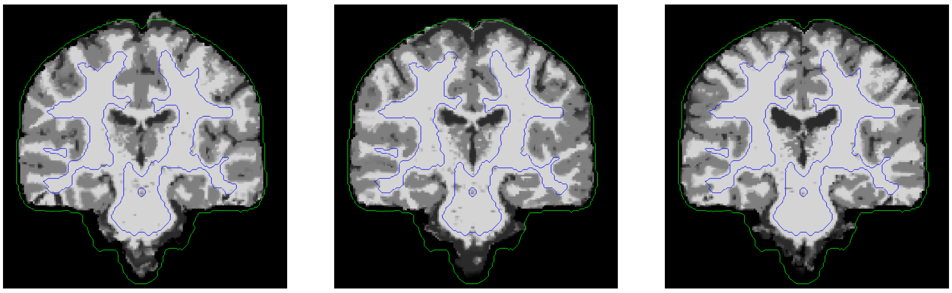


(case G54 rigid & scale) (case 57 rigid & scale) (case G24 rigid & scale)

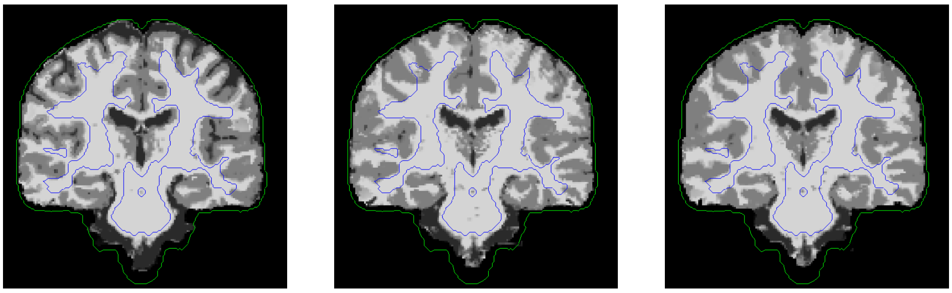
Figure 5.5: *Rigid and scale alignment of typical subjects (coronal slice). This figure shows the rigid and scale alignment of the same subjects as in Figures 5.3 and 5.4 on a coronal slice*



(case 10 non-rigid) (case 14 non-rigid) (case 29 non-rigid)

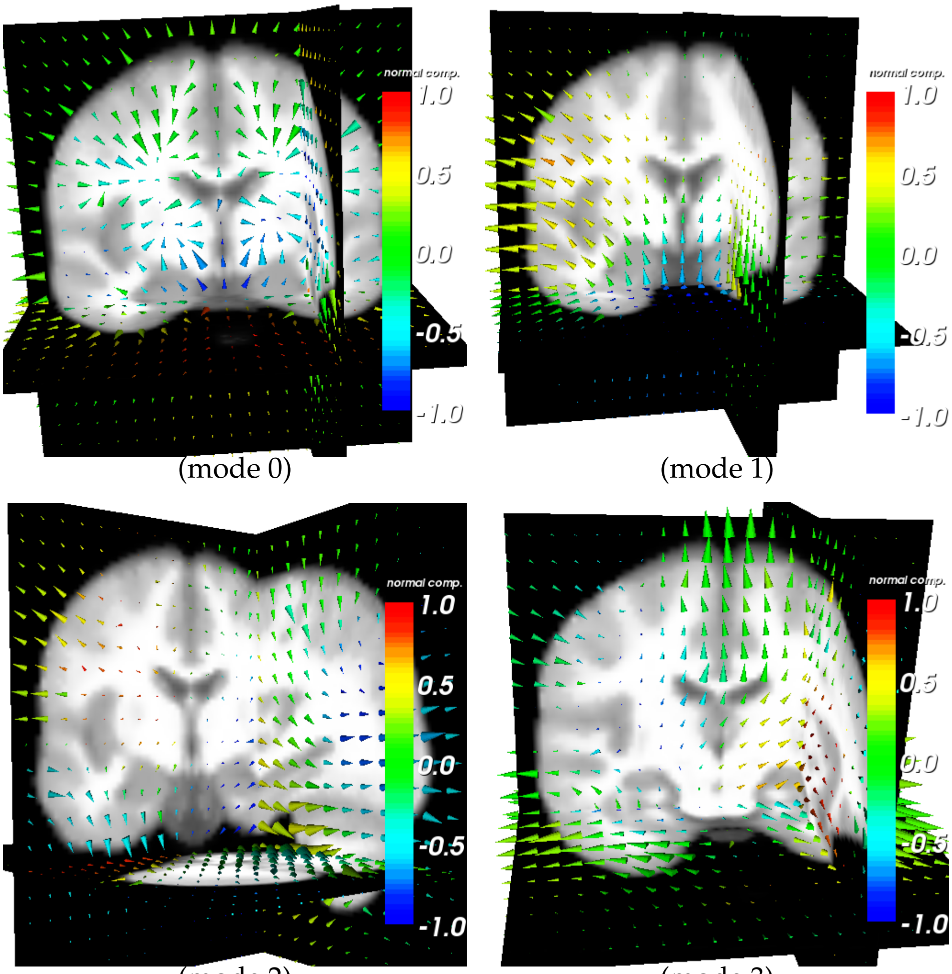


(case 3 non-rigid) (case 33 non-rigid) (case 53 non-rigid)



(case G54 non-rigid) (case 57 non-rigid) (case G24 non-rigid)

Figure 5.6: *Non-rigid alignment of typical subjects (coronal slice).This figure shows the non-rigid alignment of the same subjects as in Figures 5.3 and 5.4 on a coronal slice*



(mode 0) (mode 1)

(mode 2) (mode 3)

Figure 5.7: *Principal deformation modes.*
This figure shows the 4 first variation modes extracted from the 80 displacement fields obtained after non-rigid registration. The 0 mode shows the mean value of the displacement fields.

Conclusion and Perspectives

6

This chapter summarizes the contributions of this thesis, lists related publications and suggests improvements both on applicative and algorithmic point of views.

6.1 Achievements and Contributions

The goal of this thesis has been the development of non-rigid registration algorithms addressing well known challenges in the medical images processing community.

A first challenge is the design of registration algorithms including prior knowledge about volumetric deformations. Following the approach of Ferrant et al. [2002; 2001], we first have investigated the use of biomechanical deformation models in image registration. To address multimodal applications, we designed a gradient descent optimization of a two terms cost function combining mutual information and linear elastic energy. This algorithm is described in Chapter 4 (Section 4.5) and has been recently published in Media (Elsevier) [du Bois d'Aische et al. 2004].

A second challenge is to address high dimensional optimization problems in non-rigid registration. No optimizer performs better than all other in all applications but some classes of optimizers better fits specific classes of problems. In the case of mutual information based non-rigid registration, the optimizer has to keep reasonable performances even for a large number of parameters and must behave well in the presence of a noisy cost function. The SPSA algorithm, as proposed by Spall [1998] and described in Chapter 4, implements a fast estimate of the gradient (independently of the number of parameters). SPSA is well known to have good convergence properties in the presence of a noisy cost function (see the convergence proof in Maryak and Chin [2001]). We chose to investigate the use of SPSA in our work because of the noisy nature of mutual information. SPSA has been applied for rigid registration and this solution has been implemented as a module in the Medical Studio software, currently evaluated at the St Luc hospital. The SPSA has been committed to the ITK toolkit. We also investigated the used of SPSA in non-rigid registration combined with a Finite Elements deformation model. The algorithm and the results we got on brain shift, liver and prostate images are described in Chapter 4 (Section 4.6).

An alternative to stochastic optimization is the use of a variational method for optimizing mutual information. This approach has first been introduced by Hermosillo et al. [2001] but is not often used in practice and is currently absent from ITK. Since it combines the flexibility of a dense deformation field and the robustness of mutual information, we thought interesting to work on an implementation of this algorithm in ITK. It has been submitted for inclusion and we hope it will be integrated soon in the toolkit. First results are described in the last section of Chapter 3.

A last challenge in registration problems, is to address the matching of a large collection of subjects instead of searching correspondences between a pair of images. Our approach proposed in Chapter 5 brings this problem into the EM formalism. This work has been inspired by the paper of Warfield et al. [2004] for the validation of image segmentation and expert performance evaluation and the paper of Joshi et al. [2004]. As in [Joshi et al. 2004], a reference is first estimated for the current alignment in a first step and each subject is then aligned on the current reference in a second step. The algorithm iterates between these two steps until convergence. Our approach differs by using the STAPLE probabilistic atlas as a reference instead of a simple average of all subjects as in Joshi et al. [2004]. Our approach searches both for rigid and non-rigid transformations. Another difference is the use of a mutual information driving force for non-rigid registration instead of an optical flow force.

6.2 Perspectives and Future Works

Different perspectives could be investigated on the algorithmic and applicative levels.

From an algorithmic point of view, it would be interesting to investigate the following paths:

- To compare the approach of a two terms cost function like presented in Chapter 4 with the propagation of the elastic regularization by solving a linear system at each iteration. The deformation could be constrained by other physics-based models, such as proposed by Christensen ([Christensen and He 2001])

- To compare the performances of the first order SPSA that we have implemented with a second order scheme like presented in Spall [2000]. The algorithm could be stabilized using adaptive gains and a gradient smoothing procedure like suggested in [Vande Wouwer and Renotte 2003]
- To use the atlas constructed using the algorithm described in Chapter 5 in a joint segmentation-atlas alignment procedure (see for instance Warfield et al. [2000]
- To adapt the algorithm described in Chapter 5 for the alignment of gray level images. This would require to include some assumptions of gaussianity for each class of the probabilistic atlas and to estimate the parameters of these gaussians within the EM procedure.

From an applicative point of view, our next work will be to apply the techniques developed for brain probabilistic atlas construction to the neck area.

6.3 Related Publications

This thesis has led to several publications with Aloys du Bois d'Aische in international conferences. One journal paper has been accepted and two more journal papers will soon be submitted.

The joint registration and bias correction scheme described in Chapter 4 (Section 4.7) has been presented at the *SPIE Medical Images* in San Diego in january 2003. This work has mostly been done during my master thesis at EPFL - ITS (Lausanne)

The brain shift results of the non-rigid registration method described in Chapter 4 (Section 4.6.5) have been presented during the conference *SPIE Medical Images* in San Diego in January 2004. The results obtained on liver and prostate images (Chapter 4, Section 4.6.4) have been presented at the ISBI conference (Washington, April 2004) [De Craene et al. 2004a]. The inclusion in this implementation of the BCC mesher code developed by Aloys du Bois d'Aische has lead to a publication at the Miccai conference in Saint-Malo [du Bois d'Aische et al. 2004] (September 2004).

A journal paper has been published about mutual information combined with a bio-mechanical model for the alignment of histological and section images of the neck, in Elsevier's *Medical Image Analysis* in 2005 (see [du Bois d'Aische et al. 2004]).

I have been granted coauthor for the work of Aloys about articulated registration. This work will be presented at the next ICIP conference (Genoa 2005) [du Bois d'Aische et al. 2005]. A journal paper on this topic will soon be submitted.

Our atlas construction algorithm has been presented at the Miccai 2004 conference [De Craene et al. 2004b] and will be presented at ICIP 2005 [De Craene et al. 2005]. This work will soon be submitted for a journal publication.

Bibliography

M. Bach Cuadra, C. Pollo, Bardera A., Cuisenaire O., J.-G. Villemure, and J.-P. Thiran. Atlas-based segmentation of pathological mr brain images using a model of lesion growth. *IEEE Trans Med Imag*, 2004.

Ruzena Bajcsy and Stane Kovačič. Multiresolution Elastic Matching. *Computer Vision, Graphics, and Image Processing Journal.*, 46:1–21, 1989.

S. Bauchemin and J.L. Barron. The Computation of Optical Flow. *ACM computing surveys*, 27(3), 1995.

K.K Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent Groupwise Non-Rigid Registration for Atlas Construction. In *ISBI*, 2004.

F.L. Bookstein. Principal warps: Thin-plates splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Matching Intelligence*, 11(6):567–585, June 1989.

R.P. Brent. *Algorithms for Minimization Without Derivatives*, chapter 3-4. Prentice-Hall, 1973.

T. Butz, O. Cuisenaire, and J. Thiran. Multi-modal medical image registration: from information theory to optimization objective. In *14th International Conference on Digital Signal Processing*, 2002.

T. Butz and J. Thiran. Feature-space mutual information for multi-modal signal processing, with application to medical image registration. In *EUSIPCO 2002, Toulouse, France*, 2002.

D.C. Chin. A more efficient global optimization algorithm based on styblinski and tang. *Neural Networks*, 7:573–574, 1994.

G. Christensen and J. He. Consistent nonlinear elastic image registration. In *Proc. of MMBIA*, 2001.

G.E. Christensen, M.I. Miller, J.L. Marsh, and M.W. Vannier. Automatic analysis of medical images using a deformable textbook. *Computer assisted radiology*, pages 146–151, 1998.

A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marshal. Automated multimodality image registration using information theory. *Information Processing in Medical Imaging*, 1995a.

A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3d multimodality medical image registration using feature space clustering. In N. Ayache, editor, *Computer Vision, Virtual Reality, and Robotics in Medicine*, volume 905 of *Lecture Notes in Computer Science*, pages 195–204. Springer-Verlag, Berlin, 1995b.

J.C. Csernansky, S. Joshi, L. Wang, M. Gado, P. Miller, U. Grenader, and I. Miller. Hippocampal morphometry in schizophrenia by high dimensional brain mapping. In *Proceedings of the National Academy of Science*, pages 11406–11411, 1998.

O. Cuisenaire, J.-P. Thiran, B. Macq, C. Michel, A. De Volder, and F. Marques. Automatic Registration of 3D MR images with a Computerized Brain Atlas. In *SPIE Medical Imaging*, volume 1719, pages 438–449, 1996.

M.H. Davis, A. Khotanzad, D.P. Flamig, and S.E. Harms. A physics-based coordinate transformation for 3-d image matching. *IEEE Transactions on Medical Imaging*, 16(3):317–328, June 1997.

B.M Dawant, S.L. Hartmann, and Gadamsetty S. Brain atlas deformation in the presence of large space-occupying tumors. In *MICCAI 1999*, pages 589–596. Springer-Verlag, 1999.

M. De Craene, A. du Bois d'Aische, B. Macq, F. Kipfmueller, N. Weisenfeld, S. Haker, and S. K. Warfield. Multi-modal non-rigid registration using a stochastic gradient approximation. In *International Symposium on Biomedical Imaging*, pages 1459–1462. IEEE, 2004a. URL `http://www.tele.ucl.ac.be/~decraene/isbi2004/FINAL_ISBI.pdf`. Arlington, VA, USA.

M. De Craene, A. du Bois d'Aische, B. Macq, and S. K. Warfield. Multi-subject variational registration for probabilistic unbiased atlas generation. In *International conference on image processing*. IEEE, 2005. URL `http://www.tele.ucl.ac.be/~decraene/icip2005/icip.pdf`. Genova, It.

M. De Craene, A. du Bois d'Aische, and S. K. Warfield. Multi-subject registration for unbiased statistical atlas construction. In *7th Medical Image Computing and Computer-Assisted Intervention*, volume 3216 of *Lecture Notes in Computer Science*, pages 655–662. Springer, 2004b. URL `http://www.tele.ucl.ac.be/~decraene/miccai2004/atlas.pdf`. Saint-Malo, France.

F. Dellaert. The expectation maximization algorithm. Technical Report GIT-GVU-02-20, Georgia Insititue of Technology, GVU, 2002.

Joachim Dengler and Markus Schmidt. The Dynamic Pyramid – A Model for Motion Analysis with Controlled Continuity. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(2):275–286, 1988.

S. Dinggang and C. Davatzikos. Hammer, hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

A. du Bois d'Aische, M. De Craene, B. Macq, and S.K. Warfield. An improved articulated registration method for neck images. In *Proc. of Engineering in Medicine and Biology Conference EMBC*, 2005.

A. du Bois d'Aische, B. De Craene, M. Macq, and S. K. Warfield. An articulated registration method. In *International conference on image processing*. IEEE, 2005. URL `http://www.tele.ucl.ac.be/~decraene/icip2005/final_icip05_duboisda.pdf`. Genova, It.

A. du Bois d'Aische, M. De Craene, X. Geets, V. Gregoire, B. Macq, and S.K. Warfield. Efficient multi-modal dense field non-rigid registration : alignment of histological and section images. *Medical Image Analysis, ITK special issue*, 2004. URL `http://www.tele.ucl.ac.be/~decraene/media/media_duboisda.pdf`.

A. du Bois d'Aische, M. De Craene, S. Haker, N. Weisenfeld, C. Tempany, B. Macq, and S. K. Warfield. Improved non-rigid registration of prostate mri. In *7th Medical Image Computing and Computer-Assisted Intervention*, volume 3216 of *Lecture Notes in Computer Science*, pages 845–852. Springer, 2004. URL `http://www.tele.ucl.ac.be/~decraene/miccai2004/miccai-2004-prostate-reg%.pdf`. Saint-Malo, France.

V. Duay, P.F D'Haese, R. Li, and B. Dawant. Non-rigid registration algorithm with spatially varying stiffness properties. In *ISBI04*, pages 408–411, 2004.

Matthieu Ferrant, Arya Nabavi, Benoit Macq, Peter McL. Black, Ferenc A. Jolesz, Ron Kikinis, and Simon K. Warfield. Serial Registration of Intraoperative MR Images of the Brain. *Med Image Anal*, 6(4):337–359, 2002.

Matthieu Ferrant, Arya Nabavi, Benoit Macq, Ferenc A. Jolesz, Ron Kikinis, and Simon K. Warfield. Registration of 3D Intraoperative MR Images of the Brain Using a Finite Element Biomechanical Model. *IEEE Trans Med Imag*, 20:1384–1397, Dec 2001.

Matthieu Ferrant, Simon K. Warfield, Arya Nabavi, Ferenc A. Jolesz, and Ron Kikinis. Registration of 3d intraoperative MR images of the brain using a finite element biomechanical model. In *MICCAI*, pages 19–28, 2000. URL `citeseer.nj.nec.com/376273.html`.

R Fransens, C. Strecha, and Van Gool L. Multimodal and multiband image registration using mutual information. In *ESA-EUSC, Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation*, 2004.

J. C. Gee. On matching brain volumes. *Pattern Recognition*, 32:99–111, 1999.

D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Reading, MA, 1989.

Victor Rodríguez Gil. Feature selection for multi-modal image registration using mutual information. Technical report, Swiss Federal Institute of Technology (EPFL),Signal Processing Institute (ITS), March 2002.

A. Guimond, J. Meunier, and J.-P. Thirion. Average brain models: A convergence study. *Computer Vision and Image Understanding*, 77(2):192–210, 1999.

Nobuhiko Hata, Arya Nabavi, Simon Warfield, William M. Wells, Ron Kikinis, and Ferenc A. Jolesz. A volumetric optical flow method for measurement of brain deformation from intraoperative magnetic resonance images. In Chris Taylor and Alan Colchester, editors, *Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 928–935, Heidelberg, Germany, 1999. Springer-Verlag.

Nobuhiko Hata, Arya Nabavi, William M. Wells, Simon K. Warfield, Ron Kikinis, Peter McL. Black, and Ferenc A. Jolesz. Three-Dimensional Optical Flow Method for Measurement of Volumetric Brain Deformation from Intraoperative MR Images. *J Comput Assist Tomogr*, 24(4):531–538, Jul 2000.

Gerardo Hermosillo, Christophe Chefd'Hotel, and Olivier Faugeras. A variational approach to multi-modal image matching. Technical report, INRIA, Feb 2001.

D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff. A strategy for automated multimodality image registration incorporating anatomical khill94nowledge and imager characteristics. In H. H. Barrett and A. F. Gmitro, editors, *Information Processing in Medical Imaging*, volume 687, pages 182–196. Berlin, Germany:Springer-Verlag, 1993a.

D. L. G. Hill, C. Studholme, and D. J. Hawkes. Voxel similarity measures for automated image registration. In WA: SPIE Bellingham, editor, *Visualization in Biomedical Computing, R. A. Robb*, volume 2359, pages 205–216, 1994.

D. L. G. Hill, C. Studholme, and D. J. Hawkes. Automated 3-d registration of mr and ct images of the head. *Medical Physics*, 1:163–175, 1996.

D.L.G. Hill, D.J. Hawkes, N. A. Harrisson, and C.F. Ruff. A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics. *Information processing in medical imaging*, 687:182–196, 1993b.

Luis Ibanez and Will Schroeder. *The ITK Software Guide*. The Insight Consortium, 2005.

S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, August 2004. To appear.

S. Joshi, P. Lorenzen, G. Gerig, and E. Bullitt. Structural and radiometric assymetry in brain images. *Medical Image Analysis*, 7(2):155–170, 2003.

Kikinis, R. A digital brain atlas for surgical planning, model driven segmentation and teaching. *IEEE Trans. on Visual. and Comput. Graph.*, 2(3), 1996.

Jan Kybic and Michael Unser. Multidimensional elastic registration of images using splines. In *Proceedings of ICIP*, volume II, pages 455–458, Vancouver, Canada, 2000.

S.K. Kyriacou and C. Davatzikos. Nonlinear registration of brain images with tumor pathology using a biomechanical model. *IEEE Transaction on Medical Imaging*, 18:580–592, 1999.

Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated model-based bias field correction of mr images of the brain. *IEEE Transactions on Medical Imaging*, 18, October 1999.

Bostjan Likar, Max A. Viergever, and Franjo Pernua. Retrospective correction of mr intensity inhomogeneity by information minimization. In *Medical Image Computing and Computer-Assisted Intervention*, 2000.

F. Maes and A. Collignon. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16, 1997.

J. B. A. Maintz, P.A. Van den Elsen, and M. A. Viergever. Comparison of edge-based and ridge-based registration of ct and mr brain images. *Medical Image Analysis*, 1(2), 1996.

J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

J.-F. Mangin. Entropy minimization for automatic correction of intensity non uniformity. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysi s*, pages 162 –169, June 2000.

S. Marsland, C. Twining, and C. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *Medical Image Computing and Computer-Aided Intervention (MICCAI*, volume 2879, pages 771 – 779. Lecture Notes in Computer Science, 2003.

J.L. Maryak and D.C. Chin. Global random optimization by simultaneous perturbation stochastic approximation. In *Proceedings of the American Control Conference*, pages 756–762, Arlington, VA, June 2001.

D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank. Pet-ct image registration in the chest using free-form deformations. *IEEE Transaction on Medical Imaging*, 22(1):120–128, January 2003.

C.R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195206, 1997.

M.I. Miller, S.C. Joshi, and G.E. Christensen. *Brain Warping*, chapter 7, Large deformation fluid diffemorphisms for landmark and image matching. Academic Press, 99.

T. Minka. Expectation-maximization as lower bound maximization. `http://research.microsoft.com/~minka/papers/em.html`, 1998.

N. Molino, R. Bridson, J. Teran, and R. Fedkiw. A crystalline, red green strategy for meshing highly deformable objects with tetrahedra. In *12th International Meshing Roundtable*, pages 103–114, Sandia National Laboratories, september 2003.

A. Nabavi, Peter McL. Black, David T. Gering, C. F. Westin, Vivek Mehta, Richard S. Pergolizzi, Matthieu Ferrant, Simon K. Warfield, Nobuhiko Hata, Richard B. Schwartz, William M. Wells III, R. Kikinis, and Ferenc A. Jolesz. Serial Intraoperative MR Imaging of Brain Shift. *Neurosurgery*, 48:787–798, Apr 2001.

J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer-Verlag, 2000.

X. Papademetris, A. Sinusas, D. Dione, and R.T. Constableand J. Duncan. Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *IEEE Transactions on Medical Imaging*, 21(7): 786–799, July 2002.

G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2d-3d medical image registration. *IEEE Transactions on Medical Imaging*, 1999.

J.P.W. Pluim, J.B.A. Maintz, and M.A Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE transactions on medical imaging*, 19(8), 2000.

J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

J Rexilius. Physics-Based Nonorigid Registration for Medical Image Analysis. Technical Report 268, Medical University of Luebeck, Germany, SPL, 2001.

A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proc. of MICCAI*, volume 1496, pages 1115–1124, 1998.

G.K. Rohde, A. Aldroubi, and B.M. Dawant. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging*, 22(11):1470–1479, 2003.

D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. In *Medical Imaging: Image Processing*. SPIE, 2000.

D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Trans Med Imaging*, 22(8):1014–25, 2003.

L. Ruekert, I. Sonoda, C. Hayes, DLG Hill, MO Leach, and DJ Hawkes. Non-rigid registration using free-form deformations: Applications to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

Eduard Solanas and Jean-Philippe Thiran. Exploiting voxel correlation for automated mri bias field correction by conditional entropy minimization. In *Medical Image Computing and Computer-Assisted Intervention*, pages 1220–1221, 2001.

J. C. Spall. Overview of the simultaneous perturbation method for efficient optimization. *`http://techdigest.jhuapl.edu/td/td1904/spall.pdf`, Hopkins APL Technical Digest*, 19:482–492, 1998. URL `citeseer.nj.nec.com/536028.html`.

J.C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37:332–341, 1992.

J.C. Spall. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Transactions on Automatic Control*, 45:1839–1853, 2000.

J.C. Spall, S.D. Hill, and D.R. Stark. Theoretical comparisons of evolutionary computation and other optimization approaches. In *Proceedings of the Congress on Evolutionary Computation*, pages 1398– 1405, Washington, DC, July 1999.

R. Stefanescu, O Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, and X Pennec. Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy. In *MICCAI*, pages 704–711, 2004.

C. Studholme, D. L. G. Hill, and D. J. Hawkes. Multiresolution voxel similarity measures for mr-pet registration. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, pages 287–298. Kluwer Academic Publishers, Dordrecht., 1995a.

C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 1999.

C. Studholme, D.L.G. Hill, and D.J. Hawkes. Comparison of edge-based and ridge-based registration of ct and mr brain images. *Information Processing in Medical Imaging*, pages 287–298, 1995b.

J. Talairach and P. Tournoux. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system - an approach to cerebral imaging. *New York, NY: Thieme Medical Publishers*, 1998.

P. Thevenaz and M. Unser. Spline pyramids for inter-modal image registration using mutual information. In A. Aldroubi, A. F. Laine, and M. A. Unser, editors, *Wavelet Applications in Signal and Image Processing*, volume 3169, page pp. 236247. SPIE Press, 1997.

J.-P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demons. *Med. Image Anal.*, 2(3):243–260, 1998.

P.M. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, Khan A. A., M.S. Mega, G.W. Small, J.L. Cummings, and Toga A.W. Cortical variability and asymmetry in normal aging and alzheimer's disease. *Cerebral Cortex*, 8(6):492–509, 1998.

A.W. Toga. *Brain Warping*. Academic Press, 1999.

M. Unser, P. Thevenaz, C. Lee, and U.E. Ruttimann. Registration and statistical analysis of pet images using the wavelet transform. *IEEE Engineering in medicine and biology*, 14(5):603–611, 1995.

A. Vande Wouwer and C. Renotte. Stochastic approximation techniques applied to parameter estimation in a biological model. In *Proceedings of the IEEE Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Lviv, Ukraine, 9-11 September 2003 (nonlinear system identification for biological models)Proceedings of the National Academy of Science*, 2003. URL `http://www.jhuapl.edu/SPSA/PDF-SPSA/vandewouwer_IDAACS03.pdf`.

P. Viola and W. M. Wells. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

P. Viola and W.M. Wells. Alignment by maximization of mutual information. In *Fifth Int. Conf. on Computer Vision*, pages 16–23, 1995.

Matthew Wall. Galib : A c++ library of genetic algorithm components. Technical report, Mechanical Engineering Department, Massachusetts Institute of Technology, August 1996.

Simon K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis. Adaptive, Template Moderated, Spatially Varying Statistical Classification. *Medical Image Analysis*, 4(1):43–55, March 2000.

Simon K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation. *IEEE Trans Med Imag*, 2004.

Neil Weisenfeld and Simon K. Warfield. Normalization of joint image-intensity statistics in mri using the kullback-leibler divergence. In *ISBI*, pages 101–104. IEEE, 2004.

W.M. Wells, P. Viola, H. Atsumi, S. Nakajima, , and R. Kikinis. Multimodal volume registration by maximization of mutual information. *Med. Image Anal.*, 1(1):35–51, 1996.

J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, R. Maciunas, and al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Comput Assist Tomogr*, 21(4):554–566, Jul-Aug 1997.

R. P. Woods, S. R. Cherry, and J. C. Mazziotta. Rapid automated algorithm for aligning and reslicing pet images. *J. Comput. Assist. Tomogr.*, 16(4): 620–633, 1992.

R. P. Woods, J. C. Mazziotta, and S. R. Cherry. Mri-pet registration with automated algorithm. *J. Comput. Assist. Tomogr.*, 17(4):536–546, 1993.

O.C. Zienkewickz and R.L. Taylor. *The Finite Element Method*. McGraw Hill Book Co., 1987.

L. Zollei, E. Learned-Miller, E. Grimson, and W.M. III Wells. Efficient population registration of 3d data. In *Computer Vision for Biomedical Image Applications*. ICCV, 2005.

C.S. Zuo, A. Jiang, B.L. Buff, T.G. Mahon, and T.Z. Wong. Automatic motion correction for breast mr imaging. *Radiology*, 198(3):903–906, 1996.