Graph-based deformable registration : slice-to-volume mapping and context specific methods

Enzo Ferrante

NNT : 2016SACLC039

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
CENTRALESUPELEC

ÉCOLE DOCTORALE N°580
Sciences et Technologies de l'Information et de la Communication
Spécialité de doctorat : Mathématiques & Informatique

Par

M. Enzo Ferrante

Recalage déformable à base de graphes: mise en correspondance coupe-vers-volume et méthodes contextuelles

Thèse présentée et soutenue à Châtenay-Malabry, le 03 Mai 2016 :

Composition du Jury :

M Nicholas Ayache, Directeur de Recherche, *INRIA (Asclepios)*, Président du Jury
Mme Julia Schnabel, Professeure, *King's College London*, Rapporteur
M William M. Wells, Professeur, *Harvard / MIT*, Rapporteur
M Pierre Jannin, Directeur de Recherche, *Inserm (LTSI) / Univ. de Rennes*, Examinateur
M Christos Davatzikos, Professeur, *Univ. of Pennsylvania*, Examinateur
M Nikos Paragios, Professeur, *CentraleSupelec (CVN) / INRIA*, Directeur de thèse

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Recalage déformable à base de graphes: mise en correspondance coupe-vers-volume et méthodes contextuelles

Mots clés : recalage déformable, champs aléatoires de Markov, recalage slice-to-volume, modèles graphiques, optimisation discrète

Résumé : Les méthodes de recalage d'images, qui ont pour but l'alignement de deux ou plusieurs images dans un même système de coordonnées, sont parmi les algorithmes les plus anciens et les plus utilisés en vision par ordinateur. Les méthodes de recalage servent à établir des correspondances entre des images (prises à des moments différents, par différents senseurs ou avec différentes perspectives), lesquelles ne sont pas évidentes pour l'œil humain. Un type particulier d'algorithme de recalage, connu comme « les méthodes de recalage déformables à l'aide de modèles graphiques » est devenu de plus en plus populaire ces dernières années, grâce à sa robustesse, sa scalabilité, son efficacité et sa simplicité théorique. La gamme des problèmes auxquels ce type d'algorithme peut être adapté est particulièrement vaste. Dans ce travail de thèse, nous proposons plusieurs extensions à la théorie de recalage déformable à l'aide de modèles graphiques, en explorant de nouvelles applications et en développant des contributions méthodologiques originales.

Notre première contribution est une extension du cadre du recalage à l'aide de graphes, en abordant le problème très complexe du recalage d'une tranche avec un volume. Le recalage d'une tranche avec un volume est le recalage 2D dans un volume 3D, comme par exemple le mapping d'une tranche tomographique dans un système de coordonnées 3D d'un volume en particulier. Nos avons proposé une formulation scalable, modulaire et flexible pour accommoder des termes d'ordre élevé et de rang bas, qui peut sélectionner le plan et estimer la déformation dans le plan de manière simultanée par une seule approche d'optimisation. Le cadre proposé est instancié en différentes variantes, basés sur différentes topologies du graph, définitions de l'espace des étiquettes et constructions de l'énergie. Le potentiel de notre méthode a été démontré sur des données réelles ainsi que des données simulées dans le cadre d'une résonance magnétique d'ultrason (où le cadre d'installation et les stratégies d'optimisation ont été considérés).

Les deux autres contributions inclues dans ce travail de thèse, sont liées au problème de l'intégration de l'information sémantique dans la procédure de recalage (indépendamment de la dimensionnalité des images). Actuellement, la plupart des méthodes comprennent une seule fonction métrique pour expliquer la similarité entre l'image source et l'image cible. Nous soutenons que l'intégration des informations sémantiques pour guider la procédure de recalage pourra encore améliorer la précision des résultats, en particulier en présence d'étiquettes sémantiques faisant du recalage un problème spécifique adapté à chaque domaine.

Nous considérons un premier scénario en proposant un classificateur pour inférer des cartes de probabilité pour les différentes structures anatomiques dans les images d'entrée. Notre méthode vise à recaler et segmenter un ensemble d'images d'entrée simultanément, en intégrant cette information dans la formulation de l'énergie. L'idée principale est d'utiliser ces cartes estimées des étiquettes sémantiques (fournie par un classificateur arbitraire) comme un substitut pour les données non-étiquettées, et les combiner avec le recalage déformable pour améliorer l'alignement ainsi que la segmentation. Notre dernière contribution vise également à intégrer l'information sémantique pour la procédure de recalage, mais dans un scénario différent. Dans ce cas, au lieu de supposer que nous avons des classificateurs arbitraires pré-entraînés à notre disposition, nous considérons un ensemble d'annotations précis (vérité terrain) pour une variété de structures anatomiques. Nous présentons une contribution méthodologique qui vise à l'apprentissage des critères correspondants au contexte spécifique comme une agrégation des mesures de similarité standard à partir des données annotées, en utilisant une adaptation de l'algorithme « Latent Structured Support Vector Machine ».

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Title : Graph-based deformable registration: slice-to-volume mapping and context specific methods

Keywords : deformable image registration, markov random field, slice-to-volume registration, graphical models, discrete optimization

Image registration methods, which aim at aligning two or more images into one coordinate system, are among the oldest and most widely used algorithms in computer vision. Registration methods serve to establish correspondence relationships among images (captured at different times, from different sensors or from different viewpoints) which are not obvious for the human eye. A particular type of registration algorithm, known as graph-based deformable registration methods, has become popular during the last decade given its robustness, scalability, efficiency and theoretical simplicity. The range of problems to which it can be adapted is particularly broad. In this thesis, we propose several extensions to the graph-based deformable registration theory, by exploring new application scenarios and developing novel methodological contributions.

Our first contribution is an extension of the graph-based deformable registration framework, dealing with the challenging slice-to-volume registration problem. Slice-to-volume registration aims at registering a 2D image within a 3D volume, i.e. we seek a mapping function which optimally maps a tomographic slice to the 3D coordinate space of a given volume. We introduce a scalable, modular and flexible formulation accommodating low-rank and high order terms, which simultaneously selects the plane and estimates the in-plane deformation through a single shot optimization approach. The proposed framework is instantiated into different variants based on different graph topology, label space definition and energy construction. Simulated and real-data in the context of ultrasound and magnetic resonance registration (where both framework instantiations as well as different optimization strategies are considered) demonstrate the potentials of our method.

The other two contributions included in this thesis are related to how semantic information can be encompassed within the registration process (independently of the dimensionality of the images). Currently, most of the methods rely on a single metric function explaining the similarity between the source and target images. We argue that incorporating semantic information to guide the registration process will further improve the accuracy of the results, particularly in the presence of semantic labels making the registration a domain specific problem.

We consider a first scenario where we are given a classifier inferring probability maps for different anatomical structures in the input images. Our method seeks to simultaneously register and segment a set of input images, incorporating this information within the energy formulation. The main idea is to use these estimated maps of semantic labels (provided by an arbitrary classifier) as a surrogate for unlabeled data, and combine them with population deformable registration to improve both alignment and segmentation.

Our last contribution also aims at incorporating semantic information to the registration process, but in a different scenario. In this case, instead of supposing that we have pre-trained arbitrary classifiers at our disposal, we are given a set of accurate ground truth annotations for a variety of anatomical structures. We present a methodological contribution that aims at learning context specific matching criteria as an aggregation of standard similarity measures from the aforementioned annotated data, using an adapted version of the latent structured support vector machine (LSSVM) framework.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Para Pepa, quien se fue justo antes de llegar al final de este viaje.

To Pepa, who departed this life just before reaching the end of this journey.

Acknowledgments

First and foremost, I would like to thank Prof. Nikos Paragios for trusting and giving me the opportunity to do my PhD in the Center for Visual Computing of CentraleSupelec. His ideas are the basis of this thesis. His insights and advice together with our discussions guided my journey through the long path towards my PhD. But more importantly, his support after occasional failures, his jokes and constant stimulation allowed me to keep on going and do not give up in the difficult moments.

I am extremely grateful to the members of the jury. To Julia Schabel and Sandy Wells for reviewing this manuscript, and Pierre Jannin, Nicholas Ayache and Christos Davatzikos for reading this thesis and accepting to be part of the jury. It is a great honor to have them in my jury.

Many thanks to Prof. Silvio Savaresse for receiving me in Stanford during three amazing months, to Roland Angst for his valuable advises and discussions during this period, and to all the guys from CVGL lab for the nice moments we shared in the Bay Area.

I would like to thank the French State for the financial support for this thesis, and the Argentinian State for allowing me to study during 6 years in our public universities before moving to France, and for keeping the Argentinian House in Paris (where I lived during three amazing years) alive during more than eight decades.

The fantastic four years I spent in France would have not been the same without all my friends, which are too many to be named but too important to be omitted. The new ones that I met in the lab and in the Argentinian House at Cite Universitaire, and those that were always there supporting me from Argentina. Without them, my days in Paris would not be as happy. Special thanks to Puneet, Evgenios, Ishan and Aline, for being the best office mates one could ever ask for, to Natalia and Carine for always helping me with the paper work, to Vivien for the translation of the thesis abstract to French and to Fede for his help with the thesis writing until the last minute.

Finally, I would like to thank my family for their moral support; specially to my parents for their daily call, which made me feel in my hometown when I was actually thousands of kilometers away from home. To my Dad, for his never-ending optimism towards me and his wise words, always pronounced in the right moment. And to my Mum, because perhaps unconsciously, she was the one who transferred me her passion for knowledge, teaching and science. The same passion that brought me into the amazing world of research.

Contents

List of Figures

Chapter 1

Introduction

Image registration algorithms aim to align two or more images into the same coordinate system and they are widely used in computer vision [Szeliski 2006]. Inspired by Horn and Shunk [Horn 1980] and the work of Lucas and Kanade[1] in the eighties [Lucas 1981], the research community of computer vision has made major efforts towards developing more accurate and efficient strategies. Registration methods seek to establish correspondence relationships among images (captured at different times, from different sensors or from different viewpoints [Brown 1992]) which are not obvious for the human eye. The important image variations in conjunction with their diverse origin, make image registration one of the most challenging tasks in the field. Consequently, we observe the absence of a general algorithm suited to solve the general registration problem [Zitová 2003], and research has still to be done in order to address it. The literature in the area is vast and, every year, a significant number of new papers relevant to this problem are published in the most prestigious conferences and journals.

Graph-based deformable registration methods [Glocker 2011] have become popular during the last decade due to the fact that they exhibit robustness, scalability, efficiency and theoretical simplicity. The range of problems to which it can be adapted is broad. In this thesis, we propose several extensions to the graph-based deformable registration theory, by exploring new application scenarios and introducing novel methodological contributions. Before presenting our work, let us briefly motivate the research on image registration.

1.1 Motivation

Image registration is a central component for a large number of visual perception problems. The classical optical flow estimation problem [Fortun 2015], for example, can be formulated as a registration problem where we aim at recovering the displacement of intensity patterns in two different images. Motion analysis, video restoration, vehicle navigation systems and video surveillance, just to name a few, are cases in which optical flow estimation and image registration are critical. Another application involves remote sensing images. Accurate registration algorithms are essential in supporting Earth scientists as they mosaic satellite images and track changes of the planet's surface over time for environmental, political and fundamental science studies [Le Moigne 2011]. In the context of medical image analysis (main

[1] In this seminal work, they presented the well known patch-based translational alignment technique to estimate the optical flow between two images.

interest of this thesis), image registration methods are ubiquitous: diagnostic, planning, surgical and radiotherapy procedures are a few examples where these methods are essential.

The information provided by different anatomical imaging modalities (such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), X-ray, video sequences obtained by laparoscopy or endoscopy, etc.) as well as functional modalities (single-photon emission computed tomography (SPECT), positron emission tomography (PET), functional MRI (fMRI), etc.) is of complementary nature. Therefore, proper integration or fusion of data obtained from the independent images is desirable [Maintz 1998]. Image registration plays a key role in this task, since it allows bringing different image modalities into spatial correspondence. Longitudinal studies [Csapo 2012] -where images are gathered for the same subjects repeatedly over time- require image registration to monitor tumor growth, post surgery control, etc. Motion correction when acquiring 3D images of moving objects (such as fetal brain imaging [Rousseau 2006, Jiang 2007a] or fMRI reconstruction [Kim 1999]) is another problem tackled by image registration, where acquired images can once registered compensate the motion that occurred during the their acquisition. Another example is image guided surgery, where pre-operative 3D volumes and intra-operative 2D images need to be fused to guide surgeons during medical interventions. Image registration is fundamental in this process: it allows doctors to navigate 3D pre-operative high-resolution annotated data using low-resolution 2D images acquired in real-time during surgery. Several examples can be cited in this case such as image guided surgeries and therapies [Fei 2002], biopsies [Xu 2014a], radio frequency ablation [Xu 2013], tracking of particular organs [Gill 2008] and minimally-invasive procedures [Liao 2013, Huang 2009].

1.2 Graph-based Deformable Registration

Image registration methods can be classified according to different criteria, such as the transformation model (used to establish spatial correspondences), the matching criterion (means of defining similarities between images) and the optimization method (used to minimize the aforementioned matching criterion and thus infer the optimal transformation model). Given the importance of this problem, a large number of techniques has been proposed. We refer to the work by [Maintz 1998, Sotiras 2013] presenting a comprehensive study about medical image registration.

An active research approach to address deformable registration during the last years, has been the use of discrete methods and graphical models [Glocker 2011]. The term deformable is used to denote the fact that the observed images are associated through a nonlinear dense transformation, or a spatially varying deformation model [Sotiras 2013]. Discrete methods and graphical models [Wang 2013, Paragios 2014] are powerful formalisms that have been successfully applied to many vision tasks including image registration. Casting image registration as a labeling problem offers a simple yet robust and efficient way to model this problem. In such

a formulation, the graphical model consists of nodes corresponding to a deformation grid and the connectivity is used to impose regularization constraints. These nodes are labeled with displacement vectors corresponding to a discrete search space of allowable transformations.

Such a formulation, introduced in [Glocker 2008] is modular with respect to the matching criterion and can be efficiently optimized through different discrete optimization methods (see Section 2.4). Glocker proposed a formulation to deal with registration of images of the same dimension (i.e. 3D to 3D or 2D to 2D registration). Numerous extensions to this work were then introduced, proving the robustness of the framework. Just to name a few, let us consider: (i) the work of [Sotiras 2009] extending this framework to perform population registration (i.e. to identify an homology between more than two images at the same time), (ii) the work from the same authors coupling geometric and iconic registration method [Sotiras 2010a], where the matching criterion is modified so that it accounts for landmark correspondences and intensity based similarity measures, (iii) multi-atlas segmentation using graph-based deformable registration was proposed by [Alchatzidis 2014a, Alchatzidis 2015], while in (iv) [Fecamp 2015a, Fecamp 2015b] authors focused on how linear and deformable parameters can be decoupled when performing graph based registration. In another line of research, regularization constraints based on learned deformation priors were incorporated to the framework [Glocker 2009b], while an adaptive co-registration and segmentation of brain tumors where the sampling resolution of the label space is adapted according to the image content is proposed in [Parisot 2013]. Last but not least, a modified version of the initial work was also applied to detect temporal changes in remote sensing images [Vakalopoulou 2015].

The number of extensions proposed to the graph-based deformable registration framework, demonstrates how flexible and powerful it can be. Let us now present the main ideas developed in this thesis, which contemplate new application scenarios and introduce a number of methodological contributions.

1.3 Contributions

We begin this thesis by including, in chapter 2, an initial overview concerning discrete models in computer vision. We derive the standard semantic segmentation and image registration problems within this framework, and we present the inference algorithms used in this dissertation to optimize such models.

The first class of contributions is presented in chapter 3. We introduce a generic graph-based deformable registration framework, to deal with the challenging slice-to-volume registration problem. As we mentioned, the initial work by [Glocker 2008] proposes a formulation for deformable registration of images in dimensional correspondence. However, in slice-to-volume registration, we aim at registering a 2D image within a 3D volume, i.e. we seek a mapping function which optimally maps a tomographic slice I to the 3D coordinate space of a volume J. This problem is

substantially different from the 3D-3D or 2D-2D case; therefore, the standard registration approach cannot be directly extrapolated to the slice-to-volume scenario. We instantiate this generic framework into three different models to solve this challenging problem. Our aim is to demonstrate how flexible and powerful the graph theory is in terms of expressive potential of the modeling process, while solving a new problem using graphical models. This general formulation can be expressed through different discrete labeling problems on a graph by changing its topology, label space and energy terms.

First, we propose the so-called *overparameterized* method, which combines linear and deformable parameters within a coupled formulation on a 5-dimensional label space [Ferrante 2013]. We adopt a metric free locally over-parametrized graphical model seeking simultaneously a linear plane transformation and an in-plane dense deformation. Image similarity is encoded in singleton terms, while geometric linear consistency of the solution (common/single plane) and in-plane deformations smoothness are modeled in a pair-wise term. The main advantage of such a model is the simplicity provided by its pairwise structure, while the main disadvantage is the dimensionality of the label space which makes inference computationally inefficient and approximate (limited sampling of search space).

Motivated by the work of [Shekhovtsov 2008], we present an alternative model referred as the *decoupled* approach. In this case, linear and deformable parameters are separated into two interconnected subgraphs which refer to lower dimensional label spaces [Ferrante 2015b]. The main contribution is the ability to decouple the plane selection and the in-plane deformation parts of the transformation towards reducing the complexity of the model, while being able to obtain simultaneously the solution for both of them. To this end, the plane selection process is expressed as a local graph-labeling problem endowed with planarity satisfaction constraints, which is then directly linked with the deformable part through the data registration likelihoods. The resulting model is modular with respect to the image metric, can cope with arbitrary in-plane regularization terms and inherits excellent properties in terms of computational efficiency.

The last slice-to-volume registration model presented in this work is the *high-order* approach [Ferrante 2015a]. Differently from the strategy adopted for the decoupled model, here we aim at reducing the dimensionality of the label space by augmenting the order of the graphical model. Using third-order cliques which exploit the expression power of this type of variable interactions, we simplify the label space to standard 3D displacement vectors. Geometrical consistency (unique plane selection) and deformation smoothness (in-plane deformations) as well as image similarity are encoded in different third order cliques. Such a model provides better satisfaction of the global deformation constraints at the expense of quite challenging inference.

The second class of contributions is presented in chapter 4. The objective is to introduce semantic information within the registration process (independently of the dimensionality of the images). The notion of semantic segmentation refers to either partial or full labeling of the source/target volumes. Currently, most of

the methods rely on a single metric function that explains the similarity between the source and target images. We argue that incorporating semantic information to guide the registration process, will further improve the accuracy of the results, in particular, in the presence of semantic labels which then make the registration a domain specific problem.

In section 4.2, we consider the scenario where we are given a classifier inferring probability maps for the anatomical structures being present in the images. The method seeks to simultaneously register and segment a set of input images, incorporating this information within the energy formulation. The central idea is to use these estimated maps of semantic labels (provided by an arbitrary classifier) as a surrogate for unlabeled data, and combine them with deformable registration to improve both alignment and segmentation. To this end, images are deformed towards a common space where consensus with respect to image similarities (according to a given arbitrary metric) and semantic labels is achieved. We employ an approach sharing conceptual similarities with α-expansion [Boykov 2001], combined with efficient linear programming optimization methods. Promising results demonstrate the potential of our method on two different datasets, containing annotations of challenging brain structures.

Our last contribution -presented in section 4.3- also exploits semantic information in the registration process with a rather different objective. In this case, instead of supposing that we have more or less accurate classifiers at our disposal, we are given a set of ground truth annotations for different anatomical structures. We present a method that aims at learning context-specific matching criteria as a weighted aggregation of standard similarity measures from the aforementioned annotated data. As we will show in section 2.3, graph-based registration is modular with respect to the matching criterion. In other words, alternative similarity measures can be used depending on the type of images we are registering. Even so, the definition of this metric function is among the most critical and complicated aspects of the registration process. To alleviate this task, we show how different content specific metrics can be combined towards defining more efficient matching criteria. To this end, we adapted the standard latent structured support vector machine (LSSVM) framework [Yu 2009] so that optimal combinations of standard similarity measures can be learned from examples.

We conclude this work in chapter 5, presenting some general conclusions and giving further ideas and future research directions.

Chapter 2

Discrete Models in Computer Vision

In this chapter we provide an overview as it concerns discrete models in computer vision. First, we describe how classical discrete optimization and graphical models theory can be used to cast vision tasks as a graph labeling problem, giving a simple example where the semantic segmentation problem is solved within this framework. Second, following [Glocker 2008, Glocker 2011], we focus on how deformable image registration can be formulated using discrete models. In an effort to make this thesis self-contained, we also describe some standard discrete inference algorithms, applied in various parts of our work to perform the optimization task. We conclude with a small discussion about the relevance of discrete methods when modeling vision tasks.

Contents

2.1 Graphical Models in Computer Vision

Graphical models have been part of the computer vision field for more than three decades. In late nineties, they started to gain more attention in the community, thanks to the advent of efficient inference methods. They are also known as *probabilistic* graphical models (PGMs) [Koller 2009], since they encode probability distributions over high-dimensional spaces, and can be efficiently manipulated using ideas from discrete optimization and computer sciences. They enjoy a combination of simplicity -in mathematical terms- and efficiency that makes them the choice of preference for a wide spectrum of vision problems. Low, mid and high-level vision tasks can be modeled within this framework. Visual perception is addressed by defining a specific parametric model, associating it to the available observations (images) through an objective function and optimizing the model parameters given both, the objective and the observations [Paragios 2014].

Markov Random Fields (MRF) are a particular class of graphical models, ubiquitous in computer vision research [Wang 2013]. This powerful modeling tool corresponds to an undirected graph $\mathcal{G}$ encoding a probability distribution governed by the local independence assumption (also known as local Markov or markovian property). Each node of this graph represents a variable from the probability distribution. Dependence among the variables is modeled using the graph neighbor system, meaning that two variables are dependent of each other if and only if there is an edge between them. This dependency can be extended from order two to any order, through the concept of clique. A clique is defined as set of fully connected nodes; therefore, all the variables represented by nodes in a given clique depend on each other. The Markov property imposes that a node is independent of any other node given all its neighbors. This holds for any distribution that can be represented by a MRF.

Formally, a MRF is an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{C} \rangle$, where $\mathcal{V}$ is a set of nodes (every node $i \in V$ is associated to a variable $x_i \in \mathcal{X}$) and $\mathcal{C}$ is the set of cliques defining the graph connectivity. The Hammersley-Clifford [Hammersley 1971] theorem states that the probability distribution of any MRF can be factorized over its cliques as follows:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{x_c}), \quad \mathbf{x} = \{\mathbf{x_c}\}, \tag{2.1}$$

where Z is the so-called partition function (used to normalize the probability distribution so that $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$), ψ_c is the positive potential function associated to clique c and $\mathbf{x_c}$ is a possible configuration (assignment) for the variables x_i being part of the clique. Note that this definition is general enough to accommodate discrete and continuous variables. In this work, we focus on the discrete case, where variables x_i associated to nodes $i \in \mathcal{V}$ take values l from a discrete *label space* $\mathcal{L}$. This is known as a discrete MRF where $\mathbf{x}$ is a labeling (or assignment or realization) that imputes a label $l_i \in \mathcal{L}$ to every variable $x_i \in \mathcal{X}$, $\mathbf{x_c} \subset \mathbf{x}$ is a subset of $\mathbf{x}$ that assigns a label to every variable in the clique c and ψ_c is a clique potential mapping a positive real value to every possible configuration $\mathbf{x_c}$.

In this framework, nodes represent the variables of interest while domain specific/prior information is expressed through the clique potentials. Once the objective function is defined, the optimal solution can be obtained by performing inference via Maximum a posteriori (MAP) estimation (i.e. the solution $\hat{\mathbf{x}}$ will be the one that maximizes the probability function p):

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\in X} p(\mathbf{x}) = \arg\max_{\mathbf{x}\in X} \frac{1}{Z} \prod_{c\in\mathcal{C}} \psi_c(\mathbf{x_c}). \tag{2.2}$$

A more convenient representation of the joint distribution $p(\mathbf{x})$ can be obtained, so that the objective function is defined as a sum of clique potentials instead of a product. To this end, let us define a new type of potentials ϕ such that $\phi_c(\mathbf{x_c}) = -\log \psi_c(\mathbf{x_c})$. Therefore, the probability distribution $p(\mathbf{x})$ is redefined in terms of ϕ_c as:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\{-\sum_{c\in\mathcal{C}} \phi_c(\mathbf{x_c})\}, \quad \mathbf{x} = \{\mathbf{x_c}\}. \tag{2.3}$$

Let us denote $E(\mathbf{x}; \Psi)$, the *energy* of a MRF, as the sum of the clique potentials $\Phi = \{\phi_c\}, \forall c \in \mathcal{C}$ given by:

$$E(\mathbf{x}; \Phi) = \sum_{c\in\mathcal{C}} \phi_c(\mathbf{x_c}). \tag{2.4}$$

We reformulate the aforementioned MAP inference problem in a discrete MRF from equation 2.2 as an equivalent minimization problem where the objective function is the MRF energy $E(\mathbf{x}; \Phi)$ as follows:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in X} E(\mathbf{x}; \Phi) = \arg\min_{\mathbf{x}\in X} \sum_{c\in\mathcal{C}} \phi_c(\mathbf{x_c}) \tag{2.5}$$

Note that this equivalent energy minimization problem, does not compute the actual probability function $p(\mathbf{x})$. Therefore, the partition function Z (frequently defined as the sum of all the potentials evaluated in all possible labelings) is not required. This is desirable since computing Z as a normalizing factor for $p(\mathbf{x})$ requires the evaluation of the potentials for all possible labelings, usually intractable considering its definition.

The underlying graph can involve cliques of different sizes. The size of the maximum clique indicates the *order* of the MRF, which is one of its most distinctive properties. MRFs of order two are known as pairwise, whereas those whose order is higher than two are referred as high-order. In the same way, the energy terms of equation 2.5 involving two variables are called pairwise terms, while those with more than two variables are the high-order terms. The terms associated to only one variable are the unary terms. It is worth mentioning that the vast majority of the graph-based vision methods exploit pairwise models. The reason is that exact or approximate efficient inference solutions can be obtained in these cases. The quality of the inference depends on several facts such as the cardinality of the label space or different mathematical properties of the energy function [Kolmogorov 2004]. Nonetheless, during the last few years, novel high-order

models and inference algorithms have been developed leading to more accurate solutions [Kohli 2012, Komodakis 2011, Ishikawa 2009]. In chapter 3, we explore the expressive power of such models in the context slice-to-volume deformable registration.

The cardinality of the label space defines either a binary problem (two labels) or a multi-label problem (more than two labels). Kolmogorov and coworkers [Kolmogorov 2004] showed that, for a particular type of binary problems where the energy fulfills a condition known as submodularity, the global minimum can be found in polynomial time using graph-cuts [Boykov 1998].

An alternative structure frequently used to represent graphical models is factor graph. Some inference libraries, such as OpenGM2 [Kappes 2013] (used in this thesis), model the MAP problem using factor graphs since they explicitly describe the factorization of the joint distribution in the graph. It leads to a clear conceptualization of the underlying factorization. For the sake of completeness, we include in the appendix (section A.1) a brief description of this representation.

Several approaches were proposed to perform MAP inference on graphs, depending on the structure, order and energy of the discrete MRF. We refer the reader to the comprehensive survey [Kappes 2013] where a comparative study of modern inference techniques is presented. For demonstration purposes, in the remaining of this chapter we present two generic formulations of graphical models towards addressing two fundamental problems in computer vision and medical imaging. The semantic segmentation problem is used as an introductory and simple example. Then, image registration (the problem that we tackle in this thesis) is formulated within the same framework.

2.2 Semantic Segmentation as a Discrete Labeling Problem

Semantic segmentation is a typical example illustrating how a vision task can be solved through MAP inference on a discrete MRF. Given an image $I(i) : \Omega \in \Re^2 \rightarrow \Re$, we would like to assign a semantic class to every pixel $i \in \Omega$. Let us consider the case where we have a map indicating the probability of every pixel to be part of a given semantic class $l \in \mathcal{L}$ (label). This map could be obtained using any classifier. This probability map is represented by function $P_i(l) : \mathcal{L} \rightarrow [0, 1]$, which is a prior of the semantic classes, where i indicates a pixel from the original image.

Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a graph representing a MRF, where nodes in $\mathcal{V}$ are associated to variables (pixels) and $\mathcal{E}$ is a standard 4-connected neighborhood system defining a two-dimensional lattice (grid). In such a simple 4-connected neighborhood system each node is connected to its adjacent neighbors in left, right, top, and bottom directions. Variables x_i associated to nodes $i \in \mathcal{V}$ can take labels l_i from a label space $\mathcal{L}$. A labeling $\mathcal{S} = \{l_i \mid i \in \mathcal{V}\}$ assigns one label to every variable. We define the energy $E(\mathcal{S}; \Phi)$ consisting of unary potentials $V_i \in \Phi$ and pairwise potentials $V_{ij} \in \Phi$ such that it is minimum when $\mathcal{S}$ corresponds to the best labeling.

Unary terms $V_i(l_i) = -\log(P_i(l_i))$ assign low energy to high probability values. Pairwise terms encode an spatial homogeneity constraint by simply encouraging neighbor variables to take the same semantic label $V_{i,j}(l_i, l_j) = [l_i \neq l_j]$. This type of pairwise term is known as Potts model, and it is defined as:

$$[l_i \neq l_j] = \begin{cases} 1, & \text{if } l_i \neq l_j \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

Finally, the energy minimization problem is:

$$\mathcal{S}^* = \arg\min E(\mathcal{S}; \Phi) = \arg\min \sum_{i \in \mathcal{V}} V_i(l_i) + \lambda \sum_{(i,j) \in \mathcal{E}} V_{i,j}(l_i, l_j). \tag{2.7}$$

$\mathcal{S}^*$ represents the optimal label assignment. Note that this energy is a metric in the space of labels $\mathcal{L}$. According to [Boykov 2001], V is called a metric in the space of labels $\mathcal{L}$ if, given any three labels $\alpha, \beta, \gamma \in \mathcal{L}$ it satisfies:

$$V(\alpha, \beta) = 0 \quad \Leftrightarrow \quad \alpha = \beta, \tag{2.8}$$

$$V(\alpha, \beta) \quad = \quad V(\beta, \alpha) \geq 0, \tag{2.9}$$

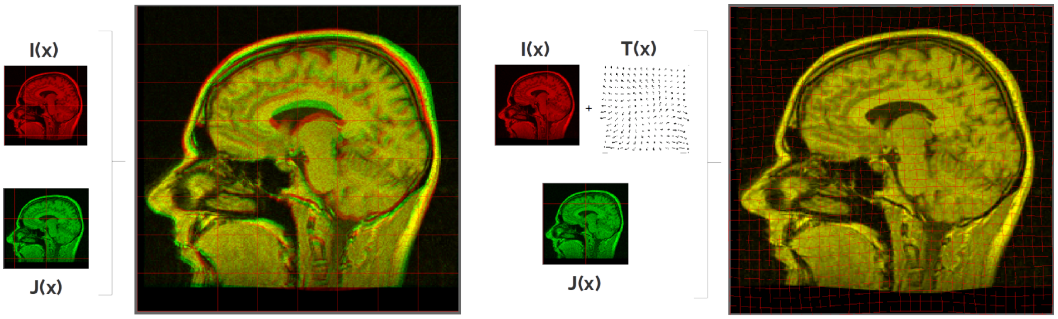
$$V(\alpha, \beta) \quad \leq \quad V(\alpha, \gamma) + V(\gamma, \beta). \tag{2.10}$$

Thus, if the energy fulfills these conditions, it is guaranteed that by using α-expansion technique we can find a solution $\hat{\mathcal{S}}$ whose energy lies within a factor of 2 with respect to the optimal energy (i.e. $E(\hat{\mathcal{S}}) \leq 2.E(\mathcal{S}^*)$). α-expansion is a well known move-making technique to perform approximate inference using graph-cuts [Boykov 1998]. It proved to be accurate in a broad range of vision problems. We refer the reader to [Boykov 2001] for a complete discussion on energy minimization using α-expansion.

2.3 Deformable Image Registration as a Discrete Labeling Problem

In the previous section, we described how semantic segmentation can be modeled through MAP inference on a discrete MRF. This problem is inherently discrete since the solution is given by a finite set of discrete labels indicating the class that corresponds to every pixel of the input image. However, in other cases like deformable image registration, the nature of the problem is not discrete. Thus, modeling it through discrete methods is not an obvious task. In this section, following [Glocker 2008, Glocker 2009c, Glocker 2011], we describe how deformable image registration can be intelligently formulated as a discrete labeling problem.

Deformable image registration is an NP-complete problem [Keysers 2003] and consists in recovering a transformation $T : \Re^d \rightarrow \Re^d$ mapping a source image I



(a) (b)

Figure 2.1: (a) Overlapping between the source (I) and target (J) images, before registration. As it can be observed, the aligment bewteen the images is not perfect. (b) Overlapping between the deformed source image and the target, after registration. The deformable registration algorithm recovered a deformation field T(x) that was applied to the source image, in order to align them.

to a target image J ($I, J : \Omega \subset \Re^d \rightarrow \Re$), so that optimal alignment is achieved[1] (the warped image is denoted by $I \circ T$). Figure 2.1 illustrates this idea. This can be mathematically formulated as follows:

$$\hat{T} = \arg\min_{T} \mathcal{M}(I, J; T) + \mathcal{R}(T), \tag{2.11}$$

where $\mathcal{M}$ represents the data term (similarity between the warped source image $I \circ T$ and the target image J) and $\mathcal{R}$ the regularization term (geometric consistency/realistic deformations). Here $\hat{T}$ indicates the optimal transformation.

T is usually defined using a dense deformation field $D : \Re^d \rightarrow \Re^d$ (relative with respect to the current position), mapping every position $x \in \Omega$ from image I to its corresponding position $T(x)$ in image J:

$$T(x) = x + D(x). \tag{2.12}$$

In such a setting, the dense deformation field D has as many parameters as pixels in the image (one d-dimensional displacement vector for every $x \in \Omega$). Such dimensionality of the graph results on high complexity which should be avoided (e.g. for a standard volumetric image of $256 \times 256 \times 256$ voxels, the degrees of freedom of our parameteric model would be $256^3 = 16777216$ 3-dimensional vectors). Therefore, a sparse representation of the deformation field is adopted. It is a reparametrization which uses k control points distributed uniformly within

[1] Note that we are assuming that both images have the same dimension d (i.e. we are performing monodimensional deformable registration, e.g. using 2D or 3D images). When we mix different dimensionalities (for example, we register 2D to 3D images), the formulation of the problem must be different. In the chapter 3 of this thesis, we present an extension to this standard formulation, adapted to the case where we aim at registering 2D slices within a 3D volume.

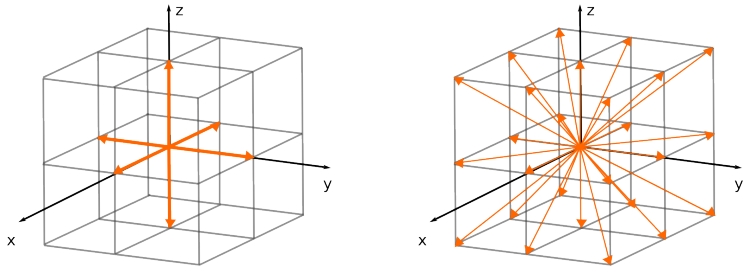
the image domain to represent the complete deformation field. An interpolation strategy acting on the displacement of these control points is used to reconstruct the final solution. The simplification introduced by such a model is significant: we can interpolate the displacement vectors that correspond to every image position x by simply manipulating a few control points. Given a set of control point displacements $\mathcal{D} = \{\mathbf{d_1}..\mathbf{d_k}\}$, the dense deformation field $D(x)$ used in equation 2.12 is reparameterized as:

$$D(x) = \sum_{i=1}^{k} \omega_i(x)\mathbf{d_i}, \tag{2.13}$$

where ω_i are weighting (or interpolation) factors determining the influence of the control point in the interpolated final position (the closer the control point is to the current position x, the higher the influence of the displacement $\mathbf{d_i}$). A common choice to define these weighting factors is the use of B-splines, resulting in the well-known Free Form Deformation (FFD) model. FFDs were popularized in the medical image registration community by [Rueckert 1999]. They adopt a regular grid as parametric model. Every control point contributes locally to the interpolation function. Other interpolation models like Thin-Plate Splines (see section 3.1.3.2) could be adopted. However, the local support for deformations and inherent computational efficiency make FFDs a good choice to deal with deformable image registration in a discrete scenario.

Once the parametric model is chosen, let us proceed with a formal definition of the corresponding graphical model and the energy function associated with it. Let us consider an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where nodes are associated to variables representing control points, and $\mathcal{E}$ is a regular grid (2D or 3D depending on the dimensionality of the images) with a 4-connected neighborhood system in case of 2D registration, and a 6-connected neighborhood system in case of 3D registration (where edges connecting nodes along the z axis are added to the standard 4-connected system). The edges $(i, j) \in \mathcal{E}$ are used to impose smoothness on the deformation field and render the problem well posed.

The label space $\mathcal{L}$ consists of labels $l_i \in \mathcal{L}$ associated to d-dimensional displacement vector $\mathbf{d_{l_i}} \in \Re^d$ (i.e. the integer label l_i could be seen as an index value for the displacement vector $\mathbf{d_{l_i}}$). Several strategies can be used to discretize the space. Two common approaches are the so-called dense and sparse sampling (illustrated in figure 2.2). When we assign a label l_i to a node i, the control point $\boldsymbol{p_i}$ is moved according to the displacement vector $\mathbf{d_{l_i}}$. A labeling Γ assigns one label l_i to every node $i \in \mathcal{V}$. Recall that every discrete label l_i is associated in our framework to a d-dimensional displacement vectors $\mathbf{d_{l_i}}$. Thus, we define $\mathcal{D}^\Gamma$ as the set formed by the displacement vectors associated to the labeling Γ. Hence, by considering different labelings Γ, we explore a discretized version of the naturally continuous space of deformation fields (see example in Figure 2.3. For example, let us consider a simple 2D label space formed by 5 labels $\mathcal{L} = \{0, 1, 2, 3, 4\}$. We assign to every label a vector of length 1 in each direction (according to the sparse sampling shown



(a) (b)

Figure 2.2: Different type of sampling strategies for the discrete label space used for image registration: sparse (a) and dense (b). Note that every label in this formulation corresponds to a displacement vector (in orange).

in figure 2.2), together with the null vector. Therefore, the vectors associated to every label would be: $\mathbf{d_0} = (0,0), \mathbf{d_1} = (1,0), \mathbf{d_2} = (0,1), \mathbf{d_3} = (-1,0), \mathbf{d_4} = (0,-1)$. If we are dealing with a graph of 6 nodes, then the labeling $\Gamma = \{0,3,1,1,3,2\}$ is interpreted as the following control point based representation of the displacement field $\mathcal{D}^{\Gamma} = \{\mathbf{d_0}, \mathbf{d_3}, \mathbf{d_1}, \mathbf{d_1}, \mathbf{d_3}, \mathbf{d_2}\}$.

The search across this space is guided by the minimization of the MRF energy defined as:

$$E(\Gamma; G, F) = \sum_{i \in \mathcal{V}} g_i(l_i) + \lambda \sum_{(i,j) \in \mathcal{E}} f_{i,j}(l_i, l_j), \tag{2.14}$$

where Γ is a labeling assigning one label l_i to every node $i \in \mathcal{V}$, G are the unary potentials playing the role of data (or matching) terms and F are the pairwise terms acting as regularizers of the estimated deformation field. The MAP inference problem (a discretized version of the continuous formulation from equation 2.11, where $G = \{g_i(\cdot)\}_{i \in \mathcal{V}}$ is associated to data terms $\mathcal{M}$ and $F = \{f_{ij}(\cdot,\cdot)\}_{(i,j) \in \mathcal{E}}$ to regularization terms $\mathcal{R}$) is therefore defined as:

$$\hat{\Gamma} = \arg\min_{\Gamma} E(\Gamma; G, F) = \arg\min_{\Gamma} \sum_{i \in \mathcal{V}} g_i(l_i) + \lambda \sum_{(i,j) \in \mathcal{E}} f_{i,j}(l_i, l_j). \tag{2.15}$$

The final dense deformation field $\hat{D}$ is interpolated from the set of displacement vectors $\mathcal{D}^{\hat{\Gamma}}$ associated to the parameters $\hat{\Gamma}$, as previously explained.

The data term measures the matching quality (according to the similarity measure) between the deformed source image I and the target image J, given the displacement vectors. In this thesis, we only consider iconic matching criteria (i.e. image intensities based functions). However, other matching criteria could be considered (we refer the reader to the review by [Sotiras 2013] for complete discussion about this topic in the context of deformable medical image registration). Every pixel (or voxel) will contribute to the data term of a given control point, according to its distance. The control points have only a limited local image support. We

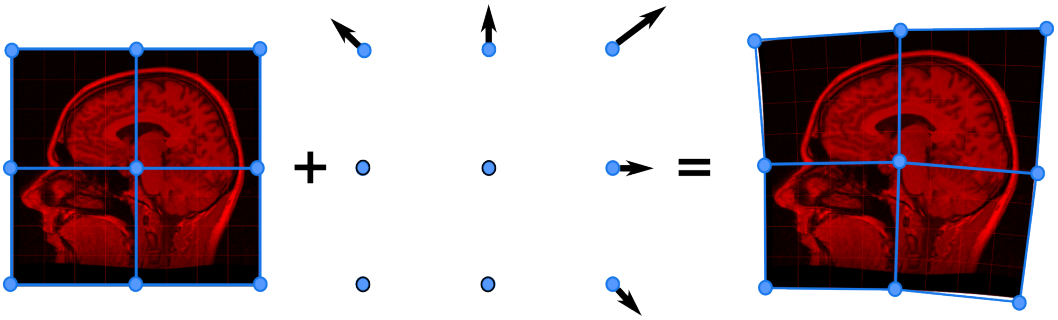


Figure 2.3: Deformable registration is formulated as a discrete labeling problem on a grid-graph, superimposed to the source 2D image. The nodes of the graph correspond to the variables of a Markov Random Field, and model the control points of a FFD. The space of plausible deformations is explored by choosing different labelings.

consider a patch-based similarity measure where the region of support for control point $\boldsymbol{p_i}$ is given by the image patch Ω_i (note that, most frequently, the complete image domain Ω is covered by the union of patches Ω_i, i.e. $\Omega = \bigcup_i \Omega_i$). Therefore, the unary potentials are defined as:

$$g_i(l_i) = \int_{\Omega_i} \omega_i(x)\delta(I \circ T_{\mathbf{d_{l_i}}}(x), J(x))dx, \tag{2.16}$$

where δ is an iconic similarity measure. Examples of standard similarity measures are the sum of absolute differences (SAD) and the sum of squared differences. In the context of different image modalities, the intensity correlation is not obvious and usually more complex functions based on statistical measures such as mutual information (MI) or cross-correlation (CC) may be adopted (for further information about iconic similarity measures, we refer the reader to section 3.1.2.1).

The data term is the driving force of the optimization process. However, even if the matching criterion is optimally minimized, the resulting deformation field could be erroneous from geometrically view point. The regularization term plays a key role in this process, since it imposes geometrical and smoothness constraints towards anatomically realistic solutions. We define it as a pairwise term acting on the two labels (displacement vectors) assigned to neighbor nodes. Using pairwise terms, we can encode different smoothness constraints, such as the standard Pott's model (penalizes deviations in the displacement assignment), truncated absolute or quadratic difference (penalize the magnitude of the vectors difference up to a bound), approximated curvature penalty (approximates a second order derivative of the displacement field) or the distance preserving approach (penalizes changes in the distances between neighboring control points with respect to their initial position) [Glocker 2009a]. Let us illustrate, for example, the distance preserving approach, defined as the ratio between the current position of the control points $\boldsymbol{p_i}, \boldsymbol{p_j}$ and their original position $\boldsymbol{p_{o,i}}, \boldsymbol{p_{o,j}}$:

$$f_{i,j}(l_i, l_j) = \frac{|| (\boldsymbol{p_i} + \mathbf{d_{l_i}}) - (\boldsymbol{p_j} + \mathbf{d_{l_j}}) ||}{|| (\boldsymbol{p_{o,i}}) - (\boldsymbol{p_{o,j}}) ||}. \tag{2.17}$$

As mentioned, alternative definitions could be considered for this term, depending on different constraints and domain specific restrictions.

A common strategy to improve the accuracy of the method and increase its capture range consists in using a pyramidal approach. It allows a progressive search space exploration, while keeping low complexity. The idea is to start with displacement vectors with important distance and progressively reduce their spread, centering them around the current solution. It is clear that increasing the number of labels to infinity will converge to a continuous formulation. Since this is intractable from computational perspective [Glocker 2008], the idea behind this approach is to iteratively compose several deformation fields, while refining the grid of nodes and the label space, so that big as well as small local deformations are captured. At every iteration we consider a fixed label space and constant number of grid nodes, we perform MAP inference for equation 2.15 and we compose the resulting deformation field with the previous one.

2.4 MAP Estimation on Graphical Models

The computer vision tasks that were presented in the previous sections (i.e. semantic segmentation and deformable image registration) were formulated as MAP inference problems. In this section, we focus on how these MAP problems can be solved through discrete optimization.

Discrete optimization of MRFs is, in general, an NP-hard problem [Shimony 1994]. However, in special cases, it can benefit from very efficient solutions. The trivial brute force algorithm (i.e. trying all possible combination of labels for each and every variable) has an exponential complexity that makes such an approach unsuitable. More efficient algorithms have been developed during the last two decades which boosted the use of graphical models in the field of computer vision. They can be classified in three main categories according to [Kappes 2013]:

i. Polyhedral and combinatorial methods, solving a continuous linear programming (LP) relaxation of the discrete energy minimization problem. The central idea is to relax the integrality condition of the variables in order to simplify the problem. Once the integrality constraint is relaxed, standard linear programming methods can be applied to solve the optimization problem, and rounding strategies are used to recover the integral solution. Examples of such approaches are Linear Programming Relaxations over the Local Polytope, Quadratic Pseudo Boolean Optimization (QPBO) [Rother 2007] and Dual Decomposition [Komodakis 2011].

ii. Message passing methods, in which messages are calculated and propagated between nodes in a graph. This propagation can be seen as a re-parametrization

of the original problem aiming to establish special properties in the re-weighted function that makes inference easier. Examples are the standard Loopy Belief Propagation (LBP) [Murphy 1999] and Three Re-weighted Belief Propagation (TRBP) [Wainwright 2005].

iii. Max-flow and move-making algorithms make use of the well know max-flow min-cut [Boykov 2004] algorithm from graph theory, which can optimally solve some instances of discrete energies. These methods are usually combined with greedy strategies that iteratively minimize over the label space by solving a sequence of max-flow min-cut sub problems. Examples are α-expansion, $\alpha\beta$ swap [Boykov 2001] and FastPD [Komodakis 2007a] algorithms. Simpler move-making algorithms not using max-flow, but still applying the strategy of starting with an initial labeling and iteratively moving to a better one until a convergence criterion is met, are the classical Iterated conditional modes (ICM) [Besag 1986] and its generalization Lazy Flipper [Andres 2012].

In the spirit of making this thesis self contained, in this coming section we describe three MAP inference algorithms used in our work. Loopy Belief Propagation and Lazy Flipper are used in chapter 3, since they can be applied to arbitrary graph topologies, label spaces and energy terms. In that chapter, we consider three different approaches to solve slice-to-volume deformable registration, varying in terms of graph topology, label space definition and energy construction. Therefore, we adopted these two algorithms which can be used to perform MAP inference in all these models. FastPD is used in chapter 4 where we propose context-specific pairwise deformable registration models which can be efficiently optimized through this method. In what follows, we briefly describe these three approaches.

2.4.1 Loopy Belief Propagation

Message passing algorithms estimate a solution by iteratively passing messages around the variables of the random field. These messages m_{ij} (sent from a node i to a node j) are actually vectors of size $\mid \mathcal{L} \mid$ (cardinality of the label space), where every scalar entry represents what node i *thinks* about assigning label l to the node j (every entry is therefore $m_{ij}(l_j)$). Given, for example, a pairwise random field with unary potentials ϕ_i and pairwise potentials ϕ_{ij}, the message $m_{ij}(l_j)$ is computed as:

$$m_{ij}(l_j) = \min_{l_i \in \mathcal{L}} (\phi_i(l_i) + \phi_{ij}(l_i, l_j) + \sum_{r \in \mathcal{N}(i) \setminus \{j\}} m_{rj}(l_i)), \tag{2.18}$$

where $\mathcal{N}(i)$ is the set of neighbors of i. Once a node (variable) i receives all the messages from its neighbors, it computes its beliefs (also vectors of size $\mid \mathcal{L} \mid$) in a label l_i as follows:

$$bel_i(l_i) = \phi_i(l_i) + \sum_{j \in \mathcal{N}(i)} m_i j(li). \tag{2.19}$$

The messages are iteratively passed from one node to its neighbors until no change occurs from one iteration to the next one. When convergence is achieved, the MAP labeling is obtained for every node i as the label l_i that minimizes the belief $bel_i(l_i)$.

Note that both, messages and beliefs computed for a given node i, depend on the messages received from its neighbors $j \in \mathcal{N}(i)$. Therefore, we need a way to initialize the message calculation process. If the graph that underlies the MRF is a tree, this process is initialized in the roots since messages for these nodes can be calculated considering just the potentials (the set $\mathcal{N}(i) \setminus \{j\}$ from equation 2.18 is empty). In this case, at convergence, the solution is guaranteed to be optimal for arbitrary energies. If the structure is not a tree, messages are passed in arbitrary order, and the algorithm is not guaranteed to converge in a finite number of iterations. LBP has shown good experimental performance in empirical studies [Murphy 1999] with arbitrary graphs and it was successfully applied in this thesis to optimize different type of graphical models in 3.

2.4.2 Lazy Flipper

Lazy Flipper is a movie-making algorithm proposed by [Andres 2012]. It is a generalization of the well-known ICM [Besag 1986] which explores (exhaustively or not) the search space.

The idea is to start from an arbitrary initial assignment and perform successive flips of variables that reduce the local (and consequently the global) energy to be minimized. A greedy strategy is adopted to explore the space of solutions: as soon as a flip reducing the energy is found, the current configuration is updated accordingly. In a first stage, only one variable is flipped at a time (in the same way that ICM). However, once a configuration is found whose energy can no longer be reduced by flips of one variable, a new stage starts where all subsets of two connected variables (i.e. variables that are linked by an edge in the graph) are considered. This strategy is applied, considering sets of maximum size k. This parameter controls the search depth. For $k = 1$, it specializes to ICM. For bigger values of k a tradeoff between approximation quality and runtime is established, which in the limit converges to an exhaustive search over only the connected subgraphs (intractable in most of the cases). The algorithm is implemented using two efficient data structures: the connected subgraph tree (CS-tree, used to efficiently and uniquely enumerate the subgraphs so that only connected subsets of variables are considered) and a tag list that prevents the repeated assessment of unsuccessful flips.

The solutions obtained using Lazy Flipper are guaranteed to be optimal within a given Hamming distance in the label space. This move-making algorithm can be applied to problems modeled with arbitrary graphs, energies and label spaces. That is the reason why we choose such a basic but robust approach to compare different types of models in chapter 3. We refer the reader to [B Andres, JH Kappes 2010, Andres 2012] for a complete description and discussion about this method.

2.4.3 FastPD

FastPD is a discrete optimization algorithm based on principles from linear programming and primal dual strategies, which at the same time generalizes α-expansion [Komodakis 2008]. Therefore, it may be included in categories (i) and (iii) from the classification scheme proposed at the beginning of this section. One of the main advantages of FastPD is its modularity/scalability, since it deals with a much wider class of problems than α-expansion, being an order of magnitude faster while providing the same optimality guarantees when performing metric labeling [Komodakis 2007b].

FastPD solves a series of max-flow min-cut problems on a graph. In that sense, it is similar to α-expansion which also performs MAP inference on multi-label problems by solving successive binary max-flow min-cut problems. The main difference between these approaches is the construction of the graph where max-flow min-cut algorithm is applied. α-expansion constructs the binary problem by restricting the label space, so that the only options for a given variable are to remain in its current assignment, or to take a label α (which varies in every iteration). Instead, FastPD constructs these binary problems by performing a Linear Programming Relaxation (LPR) of the integer program that represents the discrete MRF formulation. It builds upon principles drawn from the duality theory of linear programming, applying the well known primal-dual schema to the relaxed version of the MRF integer programming formulation.

For the sake of clarity, let us explicitly show how we can cast a standard MAP inference problem on a pairwise MRF as an integer program [Komodakis 2005]. Given, for example, the MAP formulation for the image registration problem presented in equation 2.15, we define the following integer program:

$$\begin{aligned}
\min \quad & \textstyle\sum_{i\in\mathcal{V}}\sum_{\alpha\in\mathcal{L}} g_i(\alpha)x_i(\alpha) \quad + \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}(i)}\sum_{\alpha\beta\in\mathcal{L}} f_{ij}(\alpha,\beta)x_{ij}(\alpha,\beta), && (2.20)\\
\text{s.t.} \quad & \textstyle\sum_{\alpha\in\mathcal{L}} x_i(\alpha) = 1 && \forall i \in \mathcal{V}, && (2.21)\\
& \textstyle\sum_{\alpha\in\mathcal{L}} x_{ij}(\alpha,\beta) = x_j(\beta) && \forall \beta \in \mathcal{L}, \forall i \in \mathcal{V} \wedge j \in \mathcal{N}(i), && (2.22)\\
& \textstyle\sum_{\beta\in\mathcal{L}} x_{ij}(\alpha,\beta) = x_i(\alpha) && \forall \alpha \in \mathcal{L}, \forall i \in \mathcal{V} \wedge j \in \mathcal{N}(i), && (2.23)\\
& x_i(\cdot), x_{ij}(\cdot,\cdot) \in \{0,1\} && && (2.24)
\end{aligned}$$

In this formulation $x_i(\cdot)$ and $x_{ij}(\cdot,\cdot)$ are indicators for the discrete labels l_i assigned to each node i, used to linearize the MRF energy. A binary variable $x_i(\alpha)$ is equal to 1 when node i chooses label α. The same holds for the pairwise case $x_{ij}(\cdot,\cdot)$. Equation 2.21 guarantees that every variable is labeled with only one label at a time, while equations 2.22 and 2.23 keep consistency between variables $x_i(\cdot)$ and $x_j(\cdot)$. This integer formulation is still an NP-hard problem, given the integrality constraint expressed in equation 2.24. Therefore, this integer program becomes a standard linear program when this constraint is relaxed, by allowing variables $x_i(\cdot)$ and $x_{ij}(\cdot,\cdot)$ to take continuous positive values.

The aforementioned linear program task is used by FastPD as the primal problem, while the dual of the resulting LP is used as the FastPD dual problem. Then, the primal-dual schema is iteratively applied. At every iteration, a max-flow min-cut problem is solved, improving both primal optimality and dual feasibility.

Given that FastPD can be applied to a wide range of pairwise energies (it only requires the pairwise potentials to be positive) providing good balance between efficiency and accuracy, it is used in chapter 4 where both models introducing context specific information to the registration process are conceived as pairwise MRFs.

2.5 Why Discrete Methods?

Discrete methods have several advantages when compared with continuous approaches for image registration. First, discrete algorithms are inherently gradient-free, while most part of continuous methods require the objective function to be differentiable. Gradient-free methods do not require computation of the energy derivative. Therefore, it may be applied to any complex energy function (allowing the user to define its own similarity measures in case of registration problems). The only requirement is that this function must be evaluable in a variety of possible discrete labelings. Second, continuous methods are quite often prone to be stuck in local minima when the functions are not convex. In case of discrete methods, even complicated functions could potentially be optimized using large neighbor search methods. The main limitation is the discretization of the continuous space; however, as suggested by [Glocker 2010], 'the optimality is bounded by the discretization, but with intelligent refinement strategy the accuracy of continuous methods can be achieved'. Third, parallel architectures can be used to perform non-sequential tasks required by several discrete algorithms (such as message calculation in LBP) leading to more efficient implementations. Fourth, by using a discrete label space we can explicitly control its range and resolution (it can be useful to introduce prior information, as it will be shown in chapter 3), while in continuous models it is not clear how this type of information can be used to constraint the solution. Last but not least, discrete frameworks such as discrete MRF provide a modular and principled way to combine prior knowledge with data likelihood, what makes it applicable to a wide range of vision tasks [Wang 2013].

Chapter 3

Slice-to-Volume Registration

In this chapter we investigate the use of graphical models in the context of slice-to-volume registration. We start with a comprehensive literature review, that presents most of the methods that have been proposed to deal with this challenging problem. Then, we introduce a scalable, modular and flexible formulation which can accommodate low-rank and high-order terms, that simultaneously selects the plane and estimates the in-plane deformation through a single shot optimization approach. The proposed framework is instantiated into different variants seeking either a compromise between computational efficiency (soft plane selection constraints and approximate definition of the data similarity terms through pairwise components) or exact definition of the data terms and the constraints on the plane selection. Simulated and real data in the context of US and MRI registration (where both framework instantiations as well as different optimization strategies are considered) demonstrate the potentials of our method.

Contents

3.1 Literature Review

Image registration is the process of aligning and combining data coming from more than one image source into a unique coordinate system. During the last decades, image registration has become one of the pillars of computer vision and medical imaging. In medical imaging, the problem of slice-to-volume registration has received further attention during the last decade. In this case, instead of registering images which same dimension, we seek to determine the slice (corresponding to an arbitrary plane) from a given 3D volume that corresponds to an input 2D image.

Several applications requiring slice-to-volume mapping have emerged and pushed the community towards developing more accurate and efficient strategies. Such medical imaging tasks can be classified in two main categories: those related to image fusion for image guided interventions; and those related to motion correction and volume reconstruction. In the first category, pre-operative 3D images and intra-operative 2D images need to be fused to guide surgeons during medical interventions. Slice-to-volume registration plays a key role in this process, allowing the physicians to navigate 3D pre-operative high-resolution annotated data using low-resolution 2D images acquired in real-time during surgery. In the second category, the goal is to correct for misaligned slices when reconstructing a volume of a certain modality. A typical approach to solve this task consists in mapping individual slices onto a reference volume in order to correct for inter-slice misalignment. Again, the development of accurate slice-to-volume registration algorithms is crucial to successfully tackle this problem.

Slice-to-volume registration is also known as 2D/3D registration, due to the dimension of the images involved in the registration process. However, this term is ambiguous since it describes two different problems depending on the technology used to capture the 2D image: it may be a projective (e.g. x-ray) or sliced/tomographic (e.g. US) image. Even if both problems share similarities in terms of image dimensionality, every formulation requires a different strategy to estimate the solution. The lack of perspective and different image geometry [Birkfellner 2007] inherent to both modalities, make it necessary to come up with distinct strategies to solve these registration problems. Moreover, a pixel in any 2D projective image does not correspond to only one voxel from the volume (this is the case for slice-to-volume), but to a projection of a set of them in certain perspective. Therefore, the type of functions used to measure similarities between the images is necessarily different in every case. While most of the projective 2D/3D image registration methods require to bring the images into dimensional correspondence (by different strategies like projection, back-projection or reconstruction [Markelj 2010]), in case of slice-to-volume registration, pixels from the 2D image can be directly compared with voxels from the volume. In this introduction we focus on the latter case, while a more comprehensive overview about projective 2D to 3D image registration is presented in [Markelj 2010]

Slice-to-volume registration could be considered as an extreme case of 3D-3D registration, where one of the 3D images contains only one slice. Even if theoreti-

cally true, 3D-3D registration methods can not be extrapolated in a straightforward way to the slice-to-volume scenario. This holds, particularly, for registration methods based on image information, since the descriptors used to quantify similarities between images, normally assume that the amount of information available from both images is balanced. The fact that a single slice (or even a few sparse slices) provides less information than an entire volume, should be explicitly considered in the problem formulation. Moreover, specific geometrical constraints like planarity satisfaction and in-plane deformation restrictions, arise in the case of slice-to-volume registration, which are not applicable in the setting of dimensional correspondence.

In this introduction, we discuss the literature of this emerging field, offering a comprehensive survey of the articles based on slice-to-volume registration, proposing a categorical study of the related algorithms according to a taxonomy and analyzing the advantages and disadvantages of each category. We propose a taxonomy based on previous surveys on image registration [Markelj 2010, Sotiras 2013], and adapted to the particular slice-to-volume case. We classified the related work according to several principles: (i) matching criterion (section 3.1.2), which specifies the strategies to define similarities between the images; (ii) transformation model (section 3.1.3), indicating the nature of the models used to bring images into spatial correspondence; (iii) optimization (section 3.1.4), differentiating the approaches according to the strategy used to infer the optimal transformation model; (iv) number of slices (section 3.1.5), which splits the methods in two groups, according to whether they require a single or multiple source image slices; and finally the (v) applications (section 3.1.6), where we identify the main clinical scenarios that have motivated research in the field.

3.1.1 Definition of Slice-to-Volume Registration

Let us start by giving a formal definition of slice-to-volume registration. Given a 2D image I and a 3D volume J, we seek a mapping function $\hat{\Theta}$ which optimally aligns the tomographic slice I to the 3D coordinate space of J, through the minimization of the following objective function:

$$\hat{\Theta} = \arg\min_{\Theta} \mathcal{M}(I, J; \Theta) + \mathcal{R}(\Theta), \tag{3.1}$$

where $\mathcal{M}$ represents the image similarity term (the so called matching criterion) and $\mathcal{R}$ the regularization term. Note that this mapping may be rigid or non-rigid, depending on whether we allow image I (or its corresponding reformatted slice from J) to be deformed or not. If we estimate only a rigid mapping (i.e. we calculate a 6 degrees of freedom rigid transformation or even a more restrictive one), we name the problem rigid slice-to-volume registration. In case we also infer some sort of deformation model or we consider more expressive linear transformations (such as affine transformations), we call it non-rigid registration. We refer the reader to section 3.1.3 for more information about different transformation models.

Matching criterion $\mathcal{M}$ measures the similarity between the 2D image and its corresponding mapping (slice) to the 3D volume. Usually, it is defined using inten-

sity information or salient structures from I and J. A complete discussion about matching criteria in the context of slice-to-volume registration (where we also consider methods that do not use any image information but rely on external sensors) is presented in section 3.1.2.

The regularization term $\mathcal{R}$ imposes constraints on the solution that can be used to render the problem well posed. It also may encode geometric properties on the extended (plane selection and plane deformation in case of non-rigid registration) transformation model. The choice of regularizer depends on the transformation model. While simple models like rigid body transformations can be explicitly estimated even without regularizer, the term $\mathcal{R}$ becomes crucial in more complex non-rigid scenarios to ensure realistic results.

We aim at optimizing the energy defined in equation 3.1, by choosing the best $\hat{\Theta}$ that aligns the 2D and 3D images. Depending on the variables we are trying to infer, and their optimality guarantees, they can be classified in different categories. A full study of this topic is included in section 3.1.4.

The general definition given in equation 3.1 considers a single slice as the input to the registration process. However, for the sake of completeness, in this survey we also consider the so called multi slice-to-volume registration approaches, for which several (but sparse) slices are registered to a full 3D volume. It is worth noting that we do not include methods that perform a prior 3D reconstruction from the input 2D slices before registration (see, for example, the work by [Arbel 2001, Arbel 2004]), since these methods reduce the problem to the classic 3D-3D scenario, which is not within the scope of our work. On the contrary, we consider those methods that directly register the subset of 2D slices (which can be orthogonal, parallel, arbitrary or even without an a-priori known spatial relation), rather than reconstructing a volume. Discussion and classification of the approaches according to this criterion is presented in section 3.1.5.

3.1.2 Matching Criterion

The matching criterion (also known as (dis)similarity measure, merit function or distance function) quantifies the level of alignment between the images, and it is typically used to guide the optimization process of the transformation model. Depending on the nature of information exploited in the matching process, registration methods can be classified as iconic -we use voxel intensities to quantify similarity-, geometric -we use a sparse set of salient image locations to guide the registration- or hybrid methods -we combine both strategies-. In the particular case of slice-to-volume registration, there are also some approaches which instead of using image information, they rely on other non-image technologies; we will refer to them as sensor based methods.

Slice-to-volume registration excluding medical imaging-based methods is mainly performed using two different technologies: optical (OTS) and electromagnetic (EMTS) tracking systems [Birkfellner 2008]. Optical systems employ different types of cameras (such as infrared (IR), standard RGB or laser cameras) to track markers

which help to identify the current position of the objects. Electromagnetic positioning systems perform tracking based on a system of transmitter, sensors and processing unit that localizes the position and orientation of a target object by measuring electromagnetic field properties. Both methods inherit constraints on the nature of transformation that they can estimate, since only rigid transformations can be obtained through these technologies. Moreover, accurate initial calibration is usually required at the beginning, due to the fact that positioning is done through the estimation of the relative displacement from the previous one and therefore errors can be propagated and accumulated through all the process. Instead, registration algorithms exploiting image based information -iconic or geometric- can deal with elastic anatomical changes and they are less sensitive to initial errors (since it is simple to correct them iteratively during the registration process). On the negative side, such methods can be more sensitive to image noise, which is frequently present in intra-operative, real-time and low-quality modalities, normally corresponding to the input 2D image. In addition, the amount of information of an image slice is sparse when compared to a volumetric image, resulting on ambiguities in terms of image matching that render the registration problem ill-posed. Different strategies have been developed to deal with these problems depending on the choice of the matching criterion.

Image registration can be monomodal -when the slice and the volume are captured with the same type of image technology- or multimodal -when slice and volume refer to different modalities, e.g. US slice and MRI or CT volume-. In the former case, the task of measuring the similarity between the images is simpler, since pixel/voxel intensity values corresponding to the same anatomical structure are highly correlated, or even identical, in both images. Therefore, iconic methods may perform better since the main issue associated to them -the difficulty to explain image similarities using pixel/voxel correspondences- is already solved. In case of multimodality, where the relation between pixel intensities is not obvious, there are two major alternatives: to continue using the iconic matching criterion but defining more complex similarity measures, or to adopt a geometric or sensor based strategy which appears to be more robust when dealing with different image modalities.

3.1.2.1 Iconic

Iconic matching criteria are defined using image intensity information. Similarity between images is measured using functions that act on the pixel/voxel intensity level. Standard signal processing tools, information theoretical approaches or even similarity measures defined for particular image modalities can be considered. The challenge lies in describing both images on a common space where they can be compared, in particular in the context of multimodal setting, where voxel intensities corresponding to the same anatomical or functional structures are dissimilar. There are two desirable properties sought at the definition of any iconic similarity measure: (i) to be convex, since it simplifies the optimization process and (ii) to be discriminative, in the sense that it assigns distinct values to different tissues or

anatomical structures.

Typically, in the context of slice-to-volume registration, an iconic matching criterion $\delta(I_1, I_2) : \Omega_1 \times \Omega_2 \rightarrow \Re$ is defined to quantify the similarity between two slices (or patches, according to whether we specify a global or local function). Such a similarity measure varies depending on the modalities we are trying to register. In a monomodal scenario, simple similarity measures such as the sum of absolute differences (SAD) [Ferrante 2013, Ferrante 2015a, Ferrante 2015b, Leung 2010], sum of square sifferences (SSD) [Fogtmann 2014, Heldmann 2009, Leung 2010, Marami 2011, Miao 2014a, Osechinskiy 2011b, Po Su 2013, Seshamani 2013, Xu 2008, Yu 2008, Yu 2011] or even mean of square differences [Fogtmann 2014, Gholipour 2010, Gholipour 2009, Honal 2010, Kim 2008b, Kim 2010a, Tadayyon 2010a, Tadayyon 2010b] of the intensity values can be used. In vector notation, SAD can be also seen as the L^1 norm of vectorized image, whereas SSD would correspond to the L^2 norm. These metrics assume that there is a straightforward correspondence between the intensity values in both images, which is not necessarily the case.

More complex metrics, exploiting statistical properties of the observed intensity values in both images, have also been proposed. These methods, on top of handling identity transformations, can cope with piece-wise linear relationships between the intensities in the images to be registered. In these cases, image pixels in both images are seen as entries of two random vectors X and Y. Cross-correlation (CC) is a well known function widely used in the fields of signal processing and statistics, also applied in several slice-to-volume registration studies [Birkfellner 2007, Elen 2010, Fei 2002, Fei 2003a, Fei 2003b, Fei 2004b, Frühwald 2009, Hummel 2008, Jiang 2007a, Jiang 2007b, Jiang 2009, Kainz 2015, Kim 2005, Miao 2014a, Noble 2005, Osechinskiy 2009, Osechinskiy 2011b, Xu 2008, Yan 2012, Zarow 2004]. CC measures the correlation between the entries of X and Y. It is simple to compute and, more importantly, invariant to shifts and scaling in the intensity domain. Another metric is the correlation ratio (CR), which has shown promising results even in multimodal image registration [Roche 1998]. It measures functional dependencies between X and Y, taking values between 0 (no functional dependence) and 1 (purely deterministic dependence). CR is intrinsically asymmetrical, since the two variables (images) do not play the same role in the functional relationship. In other words, unlike CC, CR offers different values depending on the order that images were considered. CR has been used as an iconic criterion for slice-to-volume registration in [Marami 2011, Osechinskiy 2011b, Smolíková-Wachowiak 2005].

Information theoretic similarity measures are usually the choice of preference for multimodal registration, and slice-to-volume registration is not an exception. The most popular is mutual information (MI), which measures the statistical dependence or information redundancy between the image intensities of corresponding distributions in both images, that is assumed to be maximal if the images are geometrically aligned [Maes 1997]. It requires an estimation of joint and marginal probability density functions (PDFs) of the intensities in every image. Given the information sparse nature of slice-to-volume registration when compared to the volume-to-volume scenario, the estimation of these joint PDFs for every slice -especially in slices of low im-

age resolution/number of samples- is a hard task and may redound to poor MI-based registration results. An alternative approach to improve MI-based slice-to-volume registration was developed by incorporating informative PDF priors in the context of fMRI time-series registration [Bhagalia 2009]. First, it was shown that slices located near the middle of the head scans give more reliable PDFs and MI estimations because they refer to a richer information space than the end slices (top or bottom). End-slices registration is then guided by a joint PDF prior based on intensity counts from registered center-slices. Alternatively, a better MI calculation in slice-to-volume registration can be achieved by using a PDF estimate that retains as much information about voxel intensities as possible from the higher resolution anatomical data set, when registering 2D MR scout scan to a complete 3D MR brain volume [Chandler 2004]. MI is a widely used similarity measure for slice-to-volume registration, adopted in an important number of methods in the last decades (e.g. [Birkfellner 2007, Brooks 2008, Eresen 2014, Fei 2002, Ferrante 2015b, Fogtmann 2014, Gill 2008, Huang 2009, Kim 1999, Museyko 2014, Park 2004, Rousseau 2006, Seshamani 2013, Smolíková-Wachowiak 2005, Tadayyon 2010a, Xiao 2011, Xu 2014a, Yeo 2004, Zakkaroff 2012, Zarow 2004]). One of the main drawbacks of MI, is that it varies when the overlapping area between the images changes, i.e. it is not invariant to changes in the overlap region throughout registration. It could happen that while estimating the transformation model, some potential solutions lie out of the volume. In such cases, an overlap invariant function would be of choice. To this end, a modified version of MI, the normalized mutual information (NMI), can be applied, which is simply the ratio of the sum of the marginal entropies and the joint entropy [Studholme 1999]. Another advantage of NMI with respect to MI is its range: it conveniently takes values between 0 and 1. NMI has been used as well for slice-to-volume registration [Chandler 2006, Elen 2010, Gefen 2008, Hummel 2008, Jiang 2007a, Kainz 2015, Kuklisova-Murgasova 2012, Leung 2010, Marami 2011, Miao 2014a, Rousseau 2005].

Using prior knowledge like segmentation masks during the iconic registration process can be useful. These approaches, also known as region-based methods, employ intensities information or statistics to describe a pre-segmented region. Chan-Vese metric [Chan 2001] for instance, aims to minimize the intensity variances on the regions inside and outside a given segmentation contour. Nir and coworkers [Nir 2011, Nir 2014] applied this matching criterion to the problem of aligning multiple slices of histological images to in vivo MR images of the prostate.

Border information is another low level visual cue that was exploited by iconic methods. It is usually determined from the intensity gradient of the images, which gives an idea of the image structure defined by intensity changes, independently of their actual value. MI as well as SAD or SSD can be applied on top of the gradient magnitudes of both images. In [Brooks 2008], for example, MI between the gradient magnitudes of an US slice and MRI volume is used. [Po Su 2013] applies SSD on both image intensities and gradients, combining them in a unique similarity measure. In [Xu 2014b], the matching criterion is defined using the normalized gradient field of the images, while CC of both intensity and gradient

magnitude was adopted in [Xu 2008].

More robust similarity measures have been introduced more recently to deal with problems in medical image registration, and in the slice-to-volume case in particular. Remarkable contributions have been made by [Wein 2008a, Fuerst 2014] to this field. They propose different similarity measures based on the simulation of US images from MRI and from CT, which can deal with these challenging multimodal registration problems. In [Wein 2007, Wein 2008a], novel methods for simulation of ultrasonic effects from CT data are presented, together with a new similarity measure entitled Linear Correlation of Linear Combination (LC^2), which is invariant to missing simulation details, yielding smooth properties and a global optimal corresponding to the correct alignment. Since they simulate US imaging effects with respect to the probe geometry, the original B-mode scan planes of the sweep are used instead of 3D reconstruction, making it suitable for multi slice-to-volume registration. LC^2 was used by [Fuerst 2014], although without the simulation process: the similarity measure is defined by locally matching US intensities to both MRI intensity and gradient magnitude.

Another robust similarity measure, the modality independent neighborhood descriptor (MIND) was proposed by [Heinrich 2012] for multimodal rigid and deformable registration. It is based on the concept of local self-similarity at the patch level in order to create a multi-dimensional descriptor through ranking of the local intensity distribution of the two images, therefore providing a very good representation of the local shape of an image feature. It can be computed in a dense fashion across all the pixels (or voxels) of the images; once it is computed, the SSD of the MIND representations can be used as a similarity measure. Thanks to its point-wise (pixel or voxel-wise) calculation nature, it can be adapted to almost any registration algorithm. An extension to MIND, named Self Similarity Context (SSC) is also estimated using patch-based self-similarities [Heinrich 2013]. In [Cifor 2013b], MIND is successfully used in a multi slice-to-volume registration framework to align untracked freehand 2D US sweeps to CT volumes.

A last family of robust similarity measures was introduced by [Rivaz 2014c, Rivaz 2014b, Rivaz 2014a] in the context of mutlimodal US/MRI image registration. In [Rivaz 2014c], a metric called Contextual Conditioned Mutual Information (CoCoMI) was proposed. The metric aims at tackling one of the main drawbacks of classic MI based methods, that is taking into account the intensity values of corresponding pixels and not of neighbor. Images are treated as "bag of words" and consequently contextual information is ignored. CoCoMI overcomes this limitation by conditioning the MI estimation on contextual information. In [Rivaz 2014b], Self Similarity α-MI (SeSaMI) -another MI based matching criterion- is proposed. α-MI is usually calculated on multiple features like intensities and their gradients, as opposed to standard MI which is usually calculated on intensities only. SeSaMI combines this multi-feature α-MI formulation with self-similarities in a kNN α-MI registration framework by penalizing clusters (i.e. the nearest neighbors) that are not self-similar. Finally, in [Rivaz 2014a], a CR similarity measure is introduced. Robust Patch Based Correlation Radio (RaPTOR) computes local CR values on

small patches and adds them to form a global cost function. Authors claim a property that makes suitable such methods: their metric is invariant to important spatial intensity inhomogenity, what is especially useful when dealing with US images due to wave attenuation, shadowing and enhancement artifacts. One of the main advantages of these Rivaz's metrics (CoCoMI, SeSaMI and RaPTOR) is that their gradient can be derived analytically, and therefore the cost function can be efficiently optimized using stochastic gradient descent methods.

3.1.2.2 Geometric

Geometric registration finds correspondences between meaningful anatomical locations or salient landmarks [Joshi 2000, Glocker 2011]. These methods aim at minimizing an energy function that, for a given transformation, measures the discrepancy between the key-points detected in both, the 2D slice and the volumetric image. Simplicity of the registration process once the landmarks are appropriately determined, no sensitivity to initializations and a wider capture range in terms of deformation are the main strengths of such approaches. On the other hand, the landmark detection and matching processes are not that trivial, and errors on their position compromise the accuracy of the registration process. Moreover, due to the sparsity of the key-points, the quality of the deformations may become insufficient (due to the limited support on the interpolation).

The early work by [Gourdon 1994] presents a geometric method to perform slice-to-volume registration between a curve and a surface. In this work, Gourdon and Ayache exploited the knowledge about the differential properties computed on both, the curve and the surface, to constrain the rigid matching problem. The most relevant contribution of this work is their discussion about how differential constraints can be used to register rigidly a curve on a surface. However, they used the basic Marching Cubes algorithm [Lorensen 1987] to extract them from simulated and medical data. The extraction of these structures remains an arduous task for images with low resolution, a scenario often valid in slice-to-volume registration.

Extracting distinctive features becomes even more complicated for medical images (opposed to natural ones), since the first ones are usually not as discriminative (lack of texture) as the last ones. Invariance to scaling, rotation and changes in illumination or brightness constitute useful properties for methods seeking to extract salient points. Highly distinctive features (in both spatial and frequency domains) simplify the matching task since it is likely that they are correctly matched. Classical examples of such descriptors successfully applied in different computer vision tasks are the Scale-invariant feature transform (SIFT) [Lowe 1999], the Harris detector [Harris 1988], the Histogram of Oriented Gradients (HOG) [Dalal 2005], Speeded Up Robust Features (SURF) [Bay 2008], etc. More recently, features learned using deep learning have been successfully applied for scenarios involving massive amount of annotated data for training [Long 2014]. These features are used to extract the salient landmarks that will guide the geometric registration process. Once the landmarks are established in both slice and volumetric images, there are

two choices. First, we can perform a matching process which results in establishing correspondences between pairs of points from the two sets. Once the matching process is finished, correspondences can be used to estimate the desired -rigid or non-rigid- transformation models. Alternatively, methods directly estimating the transformation model without inferring any correspondence could also be applied. Sometimes, both correspondences and transformation model can be inferred at the same time, like in case of Iterative Closest Point (ICP) [Besl 1992] algorithm. ICP is an algorithm subsequently improving the matching of point pairs. It minimizes the sum of geometric distances between the transformed set of source image landmarks and the closest detected landmarks in the target one. It is a simple and fast method that follows the closest neighbor principle to perform the matching task, which often converges to local minimum though.

[Dalvi 2008] proposed a slice-to-volume registration approach which uses ICP. First, it extracts phase congruency information from the slices/volume using oriented 2D Gabor wavelets. Then, using local non maximum suppression, a robust and accurate set of feature points is automatically obtained, which are subsequently matched by ICP inferring a rigid body transformation. [Nir 2014] presented a particle filtering framework using a geometric matching criteria where the closest point principle is applied (they use the SSD between pairs of closest points in the two sets). In their proposal, the inferred transformation model involves 9-DOF (affine) matrices instead of 6-DOF transformations (rigid body). Nir's method was applied by [Moradi 2013] as a pre-processing step in a tumor detection learning based framework, to register US to whole-mount histopathology references of prostate images, mapping the location of tumors to the US image domain. Another method that uses ICP was proposed by [Yavariabdi 2015] for deformable registration. They register 2D transvaginal US (TVUS) images with 3D MRI volumes to localize endometrial implants. They use contour to surface correspondences through a novel variational one-step deformable ICP method, finding a smooth deformation field while establishing point correspondences automatically. The main drawback of this approach is that it relies on the user to segment the organs. Manual pelvic organ segmentation is a laborious and time consuming task, and consequently the applicability of the approach in real scenarios is limited. Note that in the three papers discussed in this paragraph, different types of transformation models where estimated based on the closest point principle, showing the flexibility of such approach.

Signed distance functions (or signed distance maps) can be used to avoid the landmark matching step when performing non-rigid registration of shapes and points. In this case, the landmarks or shapes are assigned to zero distance, while the rest of the pixels of the image are labeled with the distance to the nearest geometrical primitive (landmark or boundary). Once the distance map is created, the optimal transformation model can be estimated by means of standard iconic registration (e.g. using SSD on the distance maps) [Paragios 2003, Xiaolei Huang 2006, Taron 2009]. In [Boer 2007], signed distance functions are used to register slices of histological images with a pre-reconstructed 3D model. Authors proposed to use a naive brute force approach for optimization, which incurs in

extremely high computational cost.

3.1.2.3 Sensor Based Methods

Sensor based systems are an alternative to image-based methods (even if images are often used for their initialization). Slice position and orientation are continuously updated from the current one. Such information can be used to extract the corresponding slice from a given volume assuming an initial correspondence. Optical (OTS) and electromagnetic (ETMS) tracking systems are the most widely used technologies to perform this task. OTS systems determine in real-time the position of an object by tracking the positions of either active or passive markers attached to the object. It requires a line-of-sight to be maintained between the tracking device and the instrument to be tracked. This might cause an inconvenience to the physician -especially in image guided interventions- during his work, resulting highly inconvenient. EMTS systems usually consist of three basic components: the electromagnetic field generator, a system control unit that interfaces with a PC, and tracked sensor coils together with their interfaces to the system control unit. By measuring the behavior of each coil, the position and orientation of the object can be determined. EMTS do not inherit the same constraints as the line-of-sight requirements but it is sensitive to distortion from nearby metal sources and exhibits limited accuracy compared to optical tracking [Birkfellner 2008]. Moreover, neither OTS nor EMTS can deal with elastic deformations between volume and slice. In general, one can claim that sensor based methods are usually more reliable (when they are calibrated correctly) than image based methods since they are not affected by ambiguities that arise during the image interpretation process.

In [Gholipour 2011], a three-dimensional magnetic field sensor is used to track the motion of a subject during MRI scanning. It allows estimating the location of the slices for volume reconstruction. The rigid 6-DOF transformation of the sensor in three dimensions (6-DOF) is calculated in real-time using the native gradient fields of the MRI scanner. The relative 3D location of each slice is computed through the sensor motion parameters at the time of slice acquisition. Other works exploiting EMTS systems can be found here [Hummel 2008, Olesch 2011a, Olesch 2011b, Xu 2008].

Optical tracking systems can also replace classical image based slice-to-volume registration algorithms. In [Schulz 2014] a slice-by-slice motion correction for fMRI image reconstruction (see section 3.1.6.2 for a complete discussion about this problem) is achieved thanks to an optical tracking system. Authors proposed to use three tracking cameras with embedded image processing, to track the position of the three optical markers attached to the skull using goggles. Tracking information is then used to replace the classical slice-to-volume registration step [Kim 1999] necessary to account for motion correction during fMRI image reconstruction. In [Bao 2005], optical tracking is used to register laparoscopic US to CT images of a phantom liver, where an an infrared camera was used to localize the US probe. The advantage of infrared cameras is that the position sensor does not pick up interference

from reflections and ambient light. However, the line-of-sight requirement still holds and therefore sensor must not be blocked. Other papers using OTS to track the position of image slices when registering them with volumetric images can be found here [Huang 2009, Penney 2006, Penney 2004, Schulz 2014].

More recently [Eresen 2014], smartphone tracking technology was considered as a navigation tool to initialize a slice-to-volume registration process between a histological 2D slice and a MR volume. The Inertial Measurement Unit (IMU) of the smartphone (also available as standalone components) is used to define the orientation of the slice. Given the orientation and position data of an IMU sensor, one can then interpolate the corresponding slice from the MR volume. This interactive alignment is applied to determine initial orientation for the 2D slice, and is refined using an iconic MI based registration process optimized via brute force. IMU tracking systems use a combination of accelerometers and gyroscopes to measure acceleration and angular velocity, respectively. Since acceleration is the second derivative of position with respect to time, and angular velocity is the first derivative, angular changes integrated over time from a known starting position yield translation and rotation (i.e. a 6-DOF transformation) [Birkfellner 2008]. These sensors are cheap and widely available. However, a major issue of using IMUs for tracking refers to accumulation of errors (either systematic or statistical), leading to degradation estimation over time. Kalman filters [Kalman 1960] can be used to deal with this type of issues and have been often adopted from the scientific community.

Note that, as described in the next section, several papers that propose sensor-based methods, are actually a combination of this technology with some image-based registration technique, resulting in what we call a hybrid method.

3.1.2.4 Hybrid

[Sotiras 2010b] states that the hybrid matching criteria take advantage of both, iconic and geometric approaches, in an effort to get the best of both worlds. This category can be extended by including sensor based technologies as it can also be combined with iconic or geometric methods.

[Mitra 2012b, Mitra 2012a] proposed a hybrid slice-to-volume registration approach, combining geometric and iconic matching criteria in a probabilistic framework to register transrectal US (TRUS) with MR images of the prostate. The geometric component is based on shape-context [Belongie 2002] representations of the segmented prostate contours; Bhattacharyya distance [Bhattachayya 1943] between the shape-context histograms of the two shapes is used to find point to point correspondence in every axial MR image. The Chi-square distances between the TRUS slice and each of the MR slices are calculated and used to determine the matching slice. Once the TRUS-MR slice pair with the minimum Chi-square distance is determined, it is used to retrieve a 2D rigid transformation (in-plane rotation and translation) between them. This transformation is applied to the remaining MR slices to ensure similar 2D in-plane rigid alignment with the 2D TRUS slice.

The iconic step is performed by measuring the similarity (NMI and CC are used as metric) between the TRUS image and every rotated axial slice of the MRI. Finally, shape and image similarity measures are transformed into probability density functions and mapped into a statistical similartity framework towards retrieving the MR slice that resembles the TRUS image. Two issues limit the applicability of such an approach in clinical practice. First, the registration is determined by the manual segmentations of the prostate on both images. Second, the method assumes that the TRUS slices are parallel to the MR axial planes; therefore, the result of the registration does not consider any out-of-plane rotation, which could occur in a realistic scenario.

Tracking information (coming from a EMTS or OTS system) can be combined in a hybrid approach with iconic or geometric criteria to perform slice-to-volume registration. In this case, the tracking signal is used to calculate the additive position of a new 2D slice with respect to the previous one, bringing them to the same coordinate system. Iconic or geometric registration is used to calculate the initial transformation that relates this unique coordinate system with the volumetric image. This approach was used by [Penney 2004] to register intra-operative US images to a pre-operative MR volume. The algorithm was extended in [Penney 2006] to deal with 2D US to volumetric CT registration. [San José Estépar 2009] refines the rigid tracking information provided by an EM system attached to a laparoscopic or endoscopic probe. The refinement is made in terms of translation parameters, using an edge-based iconic registration method combined with a phase correlation technique. In [Xu 2008], EM tracking and intra-operative iconic image registration are used to superimpose MRI data on TRUS images of the prostate. To this end, a three-step algorithm is applied, with the intermediate one corresponding to slice-to-volume registration. At the beginning of the TRUS procedure, the operator performs a 2D tracked axial sweep from the prostate's base to its apex, which is used to produce a volumetric US reconstruction of the prostate volume. The reconstructed US volume and the MRI are manually aligned. Thus, 2D TRUS slices are registered to the US reconstructed volume (reducing the MRI/US multimodality problem to US/US monomodality) while the estimated transformation can still be used to recover the corresponding MRI slice. During the intervention, out-of-plane motion compensation is achieved using intermittent multi slice-to-volume registrations between nearly real-time 2D US images and the 3D US image. This rigid body registration is based on minimizing SSD with a Simplex algorithm. The final step refers to a 2D-2D registration of US (using CC and gradient based similarity measures) and it seeks to account for in-plane misalignment. In this workflow, EM tracking information is used to simplify the problem from multi to monomodality registration, and combined with iconic slice-to-volume registration to account for motion compensation. Other hybrid methods that combine tracking information with iconic or geometric registration approaches are [Sun 2007, Yan 2012]

In [Ghanavati 2010] a method that combines the information from a tracked freehand 2D US transducer with an iconic matching criterion based on US image simulation was proposed, in the context of Total Hip Replacement (THR) surgeries.

They use a statistical shape model (SSM) of the pelvis, constructed from several CT images. They developed a multi-slice to volume registration method, to register multiple 2D US slices to a statistical atlas of the pelvis. The mean shape of the atlas is first registered rigidly to the freehand 2D US images, based on an US simulation method [Wein 2008a]. Then, the atlas is instantiated by measuring the iconic similarity between the actual US and the US images simulated from the instance of the atlas. This two-step algorithm estimating the rigid transformation and followed by the optimization of the deformable parameters, was turned into a one-shot optimized approach in [Ghanavati 2011].

In [Wein 2008b], a position sensor attached to the patient's skin was used to extract a scalar surrogate measurement, which represents an anterior-posterior translation used to detect and compensate for respiratory movements of the liver. This information is combined with an iconic criterion to estimate an affine 12-DOF transformation model which maps pre-operative plans and imaging into the interventional scenario.

[Huang 2009] proposed a hybrid slice-to-volume registration strategy exploiting multiple temporal frames. A multimodality image navigation system was introduced that integrates 2D US images with pre-operative cardiac CT volumes, using electrocardigram (ECG) information, optical tracking and iconic MI based registration. The ECG is used to synchronize the US images to the corresponding pre-operative dynamic 3D CT, depending on the cardiac phase indicated by the ECG. Spatial information given by the optical tracker is used to produce a near-optimal starting point for every slice, that is refined by maximizing the MI similarity measure. Such an approach works in real-time, making it applicable to real interventional scenarios. A different combination of sensors to solve the registration problem was proposed in [Hummel 2008], where EM and OTS sensors are combined with iconic registration to register endoscopic US to CT data.

3.1.3 Transformation Model

Transformation models explain the relation between the slice and the volume being registered, and are the outcome of the registration process. They are often classified according to their degrees of freedom. Rigid transformations deal with global rotations and translations, while deformable models -the complex case- can produce local in-plane and out-of-plane deformations. The richness of the deformation model is proportional to the number of parameters we need to specify it, and therefore the trade-off is to be found between the model complexity and power of expression.

3.1.3.1 Rigid

The simplest transformation model accounts for rotation and translation parameters. It is usually expressed as a 6 degrees of freedom (6-DOF) transformation Θ^R composed by 3 rotation and 3 translation parameters. Such a basic model is

the most common choice in the literature for slice-to-volume registration. Rigid transformations are expressive enough to explain simple slice-to-volume relations. They can deal with in-plane and out-of-plane translations and rotations. Clinical scenarios that do not inherit image distortion -like simple inter-slice motion correction [Jiang 2007b, Kim 2008b, Rousseau 2005, Smolíková-Wachowiak 2005] or basic nature image guided surgeries [San José Estépar 2009, Birkfellner 2007, Gill 2008]- can be modeled with rigid transformations.

When out-of-plane motion is avoided, even simpler models can be used. [Zakkaroff 2012] proposed to recover in-plane slice rotations in cardiac MR series, using the stack alignment transform. In-plane translation along X and Y and rotation around a user-supplied center of rotation for the individual slices were parameterized independently. It also includes a parameter for global translation along the Z direction. Hence, it combines the individual slice transforms into a unique space of parameters -that contains 3-DOF per slice instead of 6- so that all of them can be optimized simultaneously. This type of models has been largely criticized suggesting that (especially in the context of volume reconstruction) the exact corresponding slice can only be found through a slice-to-volume registration which considers out-of-plane rotation and translations as well. In [Xiao 2011], authors considered a slice-to-slice matching initial step, where every 2D image from a histological volume is matched to one axial MRI slice, via a group-wise MI based comparison. It implies a restricted transformation model involving simple slice correspondences (1-DOF). Subsequently, they correct the out-of-plane misalignment applying 2D-2D affine registration for each pair of matched slices, and a posterior 3D-3D affine registration between the histology pseudo-volume (reconstructed after transforming every 2D slice) and the MRI volume. Restricted rigid body transformations can be a convenient initialization component of a complete slice-to-volume registration pipeline.

6-DOF rigid body transformations are part of nearly all slice-to-volume registration algorithms. Literature seeking deformable registration [Xu 2014a], often initially employs rigid alignment to account for big range displacements. The standard way to estimate 6-DOF rigid transformations, consists in minimizing an energy functional (based on an iconic or geometric matching criterion) often with a continuous optimization algorithm (see section 3.1.4.1) where the search space is part of the Euclidean group $SE(3)$ of rigid transformations. Sensors as described in section 3.1.2.3 can be used as well for rigid alignment as proposed in a number of papers [Chandler 2008b, Elen 2010, Fei 2002, Fei 2003a, Frühwald 2009, Fuerst 2014, Gholipour 2010, Huang 2009, Jiang 2007a, Kainz 2015, Kim 1999, Kim 2010a, Nir 2011, Penney 2004, Rousseau 2006, Smolíková-Wachowiak 2005, Xu 2014b, Yeo 2004, Yu 2011].

3.1.3.2 Non-Rigid

In this review, we call non-rigid transformations to those models which perform -at some extent- changes in the structure of the images. These changes vary from simple operations -which can still be modeled using linear transformations, such as

scale, flip or shear- to more complex models that produce local deformations.

Linear Models

Linear transformations are the first step towards non-rigid registration. In [Wein 2008a], authors suggested that affine models can eliminate most of the large-scale deformations when registering intra-operative 2D US slices to CT images. These deformations are particularly observed between CT and US exams in different respiratory configurations, and could be handled with affine transformations. However, in a different scenario where histological slices of the prostate are registered to a MRI volume, Nir and coworkers [Nir 2014] concluded that the affine transformation model was unable to capture deformations, especially in the rectum area. This is due to the fact that local deformations cannot be expressed from global affine models. In such cases, deformable registration is required (further described in the next section).

Recently, affine models were used for registering 2D slices and 3D images of the bones [Hoerth 2015]. Their proposal includes a novel way to initiate the registration between 2D slices and 3D images of the bones. The method uses the Generalized Hough Transform (GHT) [Ballard 1981] to identify suitable starting positions. They create a template version of the 2D slice by thresholding its gradient vector field, used to explore the 3D volume trough a GHT shape-detection process, providing a set of initial positions. Standard affine registration process is then used to deform images according to the initial configurations of the GHT space and updated using a MI based criterion. Such a principle could be considered to handle alternative transformation models. Other papers that estimate affine models to solve slice-to-volume registration are [Gefen 2008, Micu 2006, Museyko 2014, Wein 2008b, Xiao 2011].

Deformable Models

Elastic deformations are powerful transformations with strong adoption in slice-to-volume mapping. The expressive power of these models depends on the number of parameters used to define them. An extensive description of different types and categories of deformable models used for non-rigid registration is presented in [Sotiras 2013] and [Holden 2008]. Here, we focus on models applied for slice-to-volume mapping.

Thin-Plate Splines (TPS) are frequently used to generate a dense deformation field from a sparse set of control points. These methods involve a set of control points that can be located in arbitrary positions, which are usually obtained by detecting salient structures. Radial basis function (RBF) -where the value at any interpolation point is given as a function of its distance from the control points- are centered at the control points and combined with affine terms to define an interpolation function. TPS minimize a bending energy based on this interpolation function, which gives a closed-form solution whose uniqueness is guaranteed in most cases. TPS can be decomposed into an affine and a local component. In

[Osechinskiy 2011a, Osechinskiy 2011b], TPS are used to parametrize a smooth 3D deformation of a 2D surface (slice). Control points are placed in a regular grid on the 2D image domain, and a 3D warp is defined using three independent TPS functions. Similarly, in [Miao 2014a], several regular 2D grids of control points define one TPS deformation model per slice in a multi slice-to-volume registration scenario. [Kim 2000] also applied TPS to a single slice-to-volume registration problem, and compared two variants of the elastic mapping (involving different number of control points to support the TPS model) with a rigid body registration algorithm. Authors extracted two main conclusions: (i) TPS based registration outperformed rigid body registration, at least in their multi slice fMRI scenario where local deformations are encountered due to the local field induced deformations or localized out-of-plane motion artifacts; and (ii) the number and location of the control points have a significant impact on the final results.

Free-form deformation (FFD) models are also widely applied for medical image registration. Originated in computer graphics [Sederberg 1986], they became popular thanks to the seminal work by [Rueckert 1999]. In this model, the weighting functions are cubic B-splines. The control points, have limited local support (in contrast to TPS where the control points influence the complete domain during interpolation) and are uniformly distributed over the image domain in a grid-like manner [Glocker 2011]. FFDs were presented in section 2.3 of this thesis since they have been used when casting deformable image registration as a discrete problem [Glocker 2008]. In [Ferrante 2013, Ferrante 2015a, Ferrante 2015b], FFDs were used in a graph-based discrete optimization framework to perform slice-to-volume deformable mapping. In this model, a 2D grid-like graph encodes at the same time the plane position (rigid body transformation) and the in-plane deformation of a slice with respect to its corresponding position at the 3D. The in-plane deformation is obtained through a FFD interpolation, where the nodes of the graph play the role of control points. Similarly, but in a continuous setting, [Osechinskiy 2011b] optimizes the position of the control points defined on a 2D grid, with unknown variables corresponding to 3D displacements. In a different scenario, [Fuerst 2014] proposed the use of a 3D grid based FFD within the bounding box of a tracked US sweep. Therefore, they estimated a single 3D deformation field out of a 3D grid, which can be used to deform all the slices contained in the US sweep. The main advantage of the FFD over TPS, is that it does not require solving a linear system for computing weights from displacements. However, FFD imposes constraints in terms of modeling, since the control points must be placed in a regular grid and must enclose the domain boundaries, whereas for TPS, control points can -in principle- be arbitrarily placed in the domain [Osechinskiy 2011b]. Other interesting works combining B-splines to model a deformable transformation in the context of slice-to-volume registration can be found here [Brooks 2008, Honal 2010, Museyko 2014, Po Su 2013, Rivaz 2014c, Rivaz 2014b, Rivaz 2014a, Tadayyon 2010a, Tadayyon 2011, Xu 2014a].

An alternative model based on finite-element (FE) meshes has been used in [Marami 2011] to model a dynamic linear elastic deformation field. It imposes a regularization constraint on permissible volume transformations based on an iconic

image similarity criterion. The advantage of FE models is that they allow incorporating the dynamic behavior of tissue deformation into the registration process, through physically meaningful constraints.

3.1.4 Optimization Method

Optimization methods aim to determine the instance of the transformation model that minimizes a function based on the aforementioned matching criterion (see section 3.1.2). Depending on the nature of the variables being involved, those methods can be classified to continuous or discrete. The continuous approaches exploit the entire space of parameters, while the discrete ones a discretized/quantized version of the admissible solutions.

Numerous problems in computer vision and medical imaging are inherently discrete (like semantic segmentation); however, this is not the case of slice-to-volume image registration, where the search space is continuous. Most of the published methods about slice-to-volume registration adopt a continuous formulation. Nevertheless, recent works on image registration have focused on discrete formulations [Glocker 2011, Glocker 2008], both for projective and sliced 2D/3D registration [Zikic 2010b, Zikic 2010a, Ferrante 2013, Ferrante 2015a, Ferrante 2015b]. The next subsections present existing work and discuss the limitations of continuous and discrete inference methods in the context of slice-to-volume registration. We also consider a third category of heuristic methods, which are independent of the nature of the variables and can be applied to a wider range of problems at the expense of not providing optimality warranty. These methods are usually applied when finding an optimal solution is impossible or impractical given the nature of the problem objective function.

3.1.4.1 Continuous

Continuous optimization algorithms are usually iterative methods. They infer the best value for a set of parameters by iteratively updating them. A common mathematical formulation for this strategy is given by:

$$\Theta_{t+1} = \Theta_t + \omega \mathbf{d_t} \qquad t = 0, 1, 2, 3.... \tag{3.2}$$

where Θ is the vector of parameters, $\mathbf{d_t}$ is the search direction at iteration t and ω is the step size or gain factor. The search direction can be calculated using different strategies.

Continuous optimization methods can be classified according to different criteria (e.g. convex or concave minimization, solvers for linear or non-linear functions, smooth or non smooth problems, etc.). Here we simply classify them depending on whether they perform gradient- or non-gradient-based optimization, i.e., whether they exploit first (or higher order) derivatives of the energy function to compute the search direction, or they rely on other strategies. An interesting comparative anal-

ysis where methods coming from both classes are used to solve slice-to-volume registration of histological slices with MR images is presented by [Osechinskiy 2011b].

Gradient-based Methods

Gradient-driven methods use the derivative of the objective function to guide the optimization process. In case of convex and differentiable functions, these methods are endowed with optimality guarantees. Otherwise, convergence to local minima is possible, and they could be sensible to initialization. Their main drawback is the requirement of analytical derivation or numerical estimation of the energy function derivatives, reducing their applicability since it is usually complicated.

Gradient descent is the simplest strategy in this category, where the search direction $\mathbf{d_t}$ is given by the negative gradient of the energy function. It refers to the standard continuous optimization method, and it has been widely applied to the problem of slice-to-volume registration [Rousseau 2005, Rousseau 2006, Bhagalia 2009, Huang 2009, Tadayyon 2010a, Tadayyon 2011, Kim 2010b, Marami 2011, Yu 2011, Zakkaroff 2012, Miao 2014a, Xu 2014a].

Conjugate gradient methods use conjugate directions instead of the local gradient to estimate $\mathbf{d_t}$. Energy function with the shape of a long and narrow valley, can be optimized using fewer steps than standard gradient descent approach, resulting in faster convergence. [Elen 2010, Osechinskiy 2011b] have applied this strategy to estimate rigid and non-rigid slice-to-volume mapping functions, respectively.

Quasi-Newton optimization strategies are used when the Jacobian or Hessian of the energy function cannot be calculated or it is too expensive to be computed. In this case, the search direction at time t is computed based on an estimation of the Hessian, calculated using information provided by the previous iteration $t-1$. Quasi-Newton optimization was applied to perform deformable registration of brain slices to MR images in [Kim 2005], using several similarity measures. It was also used to register 2D intra-operative US images with pre-operative volumes in [Brooks 2008]. One of the most popular quasi-Newton methods is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm used in [Honal 2010] to correct for breathing motion artifacts during MRI acquisitions. A limited-memory version of BFGS (L-BFGS) is particularly suited for problems involving large numbers of variables and was applied to solve multi-slice to volume registration in [Fogtmann 2014, Xu 2014a].

Gauss-Newton methods can deal with non-linear least squares functions. An approximation of the Hessian matrix is used which, once combined with the gradient, provides a good estimate of the search direction for such functions. Thus, it only requires computing first order derivatives (in contrast to the standard Newton method where the actual Hessian matrix must be computed). In [Heldmann 2010, Olesch 2011a, Olesch 2011b], Gauss-Newton method is used for multimodal registration, where 2D US slices of the liver were registered to a pre-operative CT volume. Their matching criterion computes the Sum of Squared Differences (SSD) on the segmentation of the liver vessel structure in both, 3D and 2D slices.

SSD is a least square function and therefore, Gauss-Newton method can be efficiently applied. This method was also used in [Lasowski 2008], solving an Iteratively Reweighter Least Squares (IRLS) process in order to estimate a rigid transformation between fluoroCT and CT images. The Levenberg-Marquardt algorithm is an alternative approach to solve non-linear least squares problems, that requires also only first order derivatives. For well-behaved smooth functions, this algorithm can take more time than standard Gauss-Newton method. However, the method is more robust than the standard Gauss-Newton, meaning that it could find a solution even with bad initializations. To the best of our knowledge, this optimization algorithm was only applied to estimate rigid body transformations; however, it has been used to optimize several iconic matching criteria in a variety of domains such as reconstruction of 3D cell images [Yu 2011], endoscopic interventions [Hummel 2008] and fetal brain imaging [Kim 2008b, Kim 2010a].

Non-Gradient based Methods

Derivative free methods eliminate the differentiability condition of the objective function. They are applicable to a wider range of functions, including noisy, non-differentiable or even unknown functions, where we only have a black-box that returns an output value given a set of input parameters. These are cases that frequently arise when dealing with medical image registration.

The simplest non-gradient based method that has been used to solve slice-to-volume registration is the local search or best neighbor method. In this case, each parameter to be estimated is perturbed in turn using the stepsize ω, and the value of the objective function is calculated. The search direction d_t is then estimated as the one that produced the greatest improvement in the objective function. The main drawback of this approach is its dependency from the choice of the initial stepsize ω. A big stepsize can result in an algorithm moving outside its capture range. On the other hand, for stepsizes that are too small, the optimization may become trapped in a local optimum [Penney 2004]. Such a greedy algorithm has been applied to estimate rigid body and affine transformations mapping slice-to-volume, which have less DOF than deformable models. Different image modalities like US to MRI [Penney 2004] or CT [Penney 2006, Sun 2007] images, as well as fluoroCT images [Micu 2006] were considered in the clinical setting.

Nelder-Mead [Nelder 1965] (also known as downhill simplex method or amoeba method) is the most popular derivative-free method. It relies on the notion of simplex (a $n+1$ polytope living in a n-dimensional space) to explore the space of solutions in a systematic way. At every iteration, the method constructs a simplex over the search surface, and the objective function is evaluated on its vertices. The algorithm moves across the surface by replacing, at every iteration, the worst vertex of the current set by a point reflected through the centroid of the remaining n points. The method can converge to a local optimum for objective functions that are smooth and unimodal. However, it exhibits more robust behavior for complex parameter space compared to standard gradient-based methods, providing a good compromise

between robustness and convergence time [Leung 2010]. It has been widely used for slice-to-volume registration applications [Fei 2002, Fei 2003a, Fei 2003b, Fei 2004b, Frühwald 2009, Gill 2008, Hummel 2008, Kim 1999, Kim 2000, Leung 2010, Noble 2005, Osechinskiy 2011b, Park 2004, Wein 2008a, Xiao 2011, Xu 2014b, Yeo 2006, Yeo 2004].

Another popular non-gradient based method is the well-known Powell's method. It explores the search space by performing bi-directional searches along N different vectors. Usually, these vectors initially refer to the canoninc directions. Then, the search directions are updated using linear combinations of the earlier ones. The algorithm iterates until no significant improvement is made. [Smolíková-Wachowiak 2005] applied Powell's method to register two dimensional cardiac images to preoperative 3D images. Authors claim that similarly to all local techniques, Powell's method converge to local minima, but it is generally robust and accurate, and exhibits fast convergence. [Wein 2008b] takes advantage of the fact that Powell's method performs line search in specific directions. They apply Principal Component Analysis (PCA) on the 12 parameters of an affine transformation, reducing the search space to the 3 most significant PCA modes. Therefore, Powell direction search is initilized with the most significant PCA modes, which assures good performance and robustness. Other papers applying Powell's method to our problem can be found here [Fei 2003a, Gholipour 2010, Gholipour 2009, Jiang 2007a, Jiang 2007b, Jiang 2009, Micu 2006, Osechinskiy 2011b].

3.1.4.2 Discrete

In a discrete scenario, slice-to-volume registration is expressed as a discrete labeling problem over a graphical model. This is, indeed, the approach considered in this thesis. For the sake of completeness, in this section we are briefly reviewing our own publications, which are (to the best of our knowledge) the only discrete based slice-to-volume registration algorithms proposed to date.

Graphical models and discrete optimization are powerful formalisms that have been successfully used during the past years in the field of computer vision [Wang 2013]. Image registration can be formulated as a minimal cost graph problem where the nodes of the graph correspond to the control points of a deformation grid and the graph connectivity imposes regularization constraints. Even if this technique has been applied mainly to mono-dimensional cases (2D-2D or 3D-3D) [Glocker 2011], our works were focused on extending this theory to the case of slice-to-volume registration [Ferrante 2013, Ferrante 2015a, Ferrante 2015b]. In section 2.3, we showed how monodimensional deformable image registration can be formulated within the discrete MRFs framework, as proposed by [Glocker 2008, Glocker 2009c, Glocker 2011]. We refer to [Wang 2013] for a complete discussion about discrete methods in computer vision.

[Ferrante 2013] addressed deformable slice-to-volume registration using graphical models and discrete optimization in the so-called overparameterized approach (see section 3.2.1). Grid nodes encode at the same time the plane location and the in-plane deformations. A pairwise model which combines linear and deformable

parameters within a coupled formulation on a 5-dimensional label space was used. FastPD was adopted as the optimization algorithm given the properties of the energy function (pairwise terms and non-submodularity). Motivated by the work of [Shekhovtsov 2008], in an effort to reduce the dimensionality of the label space, [Ferrante 2015b] presented a different model (the so called decoupled model, see section 3.2.2 for a complete description) where linear and deformable parameters are now separated into two interconnected subgraphs which refer to lower dimensional label spaces. It reduces the dimensionality of the label space by increasing the number of edges and vertices, while keeping a pairwise graph that was optimized using loopy belief propagation. Finally, in [Ferrante 2015a], a high-order approach is presented (see section 3.2.3), where the label space dimensionality reduction is achieved by augmenting the order of the graphical model, using third-order cliques which exploit the expression power of this type of variable interactions. Such a model provides better satisfaction of the global deformation constraints at the expense of complex inference. Loopy belief propagation was also used as the optimization algorithm in this case.

3.1.4.3 Miscellaneous

Independently of whether the variables are continuous or discrete, different strategies can be used to explore the space of solutions. Heuristic or metaheuristic can be used to obtain acceptable solutions especially in cases where we deal with non-linear, non-convex or black-box optimization problems, even if we know that no optimality guarantee is provided. In some cases (e.g. initializing a more complex registration procedure) having an approximately good solution is enough; in these scenarios, different strategies can be envisioned.

Evolutionary algorithms are among the most popular bio-inspired metaheuristics. In evolutionary computation, the concept of biological evolution is used to explore the search space. The individuals of a given population are the candidate solutions to our problem and they evolve according to different laws (such as mutation, recombination and selection). Evaluation of the quality for a given solution is performed through a fitness function, which is in fact our objective function. The idea is that, over the generation sequence, individuals with better and better fitness are generated, leading to a good solution. These algorithms are usually stochastic. If this is the case, new candidate solutions might be sampled according to a multivariate normal distribution in the space of parameters. One evolutionary algorithm applied in slice-to-volume registration problems is the covariance matrix adaptation evolutionary strategy (CMA-ES) [Ghanavati 2010, Tadayyon 2010a, Tadayyon 2011]. This method updates the covariance matrix of the aforementioned distribution, by estimating a second order model of the underlying objective function similar to the approximation of the inverse Hessian matrix. Another case based on genetic algorithms to solve our problem of interest was proposed by [Gefen 2005, Gefen 2008]. Different optimization strategies may be combined as well depending on the parameter space. In [Tadayyon 2010a, Tadayyon 2011], for example, the CMA-ES was

applied to optimize the translation parameters of the rigid transformation model, while rotation and deformable ones were estimated using a standard gradient descent approach. In their application to prostate MR images, the CMA-ES was not able to optimize a 6-DOF search space as it diverged on rotations regardless of scaling.

Another popular metaheuristic is the well known simulated annealing (SA) method. It mimics the physical process that metal atoms suffer when the material is heated and then slowly cooled. In order to avoid local minima, the algorithm explores new directions which lower the objective function, but also, with a certain probability, those that raise the objective. This probability decreases with the number of iterations, so that it stays away from local minima in early iterations and is able to explore globally for better solutions. [Birkfellner 2007] used this method to register fluroCT slice to CT volumes optimizing several standard iconic matching criteria, while [Cifor 2013b] applied it to a multi-slice to volume registration case where a robust modality-independent similarity measure was optimized.

3.1.5 Number of Slices

In this review we include single and multi slice-to-volume registration methods. While in the first case the estimated slice-to-volume mapping function considers just a single 2D slice, in the second case several slices are mapped to the 3D volume.

A trivial extension of any single slice-to-volume registration method to the multi-slice case could be achieved by simply applying the algorithm to every input slice independently. It would allow parallelizing the process -since every registration could be performed in parallel- but, at the same time, contextual information would be lost since no relation among the input slices would be considered. Several methods of this type have been proposed. We can cite [Fei 2003a, Ferrante 2015a, Osechinskiy 2011b, Yavariabdi 2015] among others.

Different strategies can be adopted to register multiple slices to a given volume. Here we only review methods that perform a consistent transformation to the set of input slices, but still conceive the process as slice-based without seeing them as a complete and unique volume. Methods that reconstruct a new volume using the input slices, and then perform a standard 3D-3D registration process, are not taken into account. Note that, in some works like [Brooks 2008, Yan 2012, Xu 2014a], experiments have been performed to compare the performance of multi slice-to-volume versus volume-to-volume registration. In [Xu 2014a], the comparison was done in the context of MR prostate images, in order to show that multi slice-to-volume registration is sufficient in capturing prostate motion intra-operatively. The conclusion was that multi slice-to-volume registration was able to produce results that were close enough to volume-to-volume registration. This is an encouraging result, considering the data computation reduction archived when using only a few slices instead of a complete volume. [Brooks 2008] and [Yan 2012] arrived at similar conclusions in different medical scenarios. This consensus indicates the reliability of using sparse slices instead of the full volume when possible. Several advantages are

associated to this strategy: (i) the computational requirements are lower thanks to the reduction in the amount of data to be processed (only a few slices vs a complete volume), (ii) the omission of the potentially complex reconstruction step, (iii) greater adaptability of the similarity measures and (iv) easier parallelization [Brooks 2008].

In the presence of multiple slices, multi slice-to-volume procedure -as opposed to single slice procedure- greatly increases the quality of the matching leading to more robust registration methods [Yu 2011]. This is even more interesting when the relative position between the slices is known. In [Chandler 2006, Chandler 2008b], for example, multi slice-to-volume registration was used to correct for misaligned cardiac anatomy in Short Axis (SA) images by registering stacks of two parallel slices (which are supposed to be aligned between them) to a high-resolution 3D MR axial cardiac volume. In other cases, different configurations like orthogonal slices [Gill 2008,Leung 2010,Tadayyon 2011,Miao 2014b,Xu 2014a], slices aligned in arbitrary positions (obtained using a tracking system) [Wein 2008a,Heldmann 2009, Heldmann 2010,Olesch 2011a,Olesch 2011b,Cifor 2013b] or even temporal sequences of slices [Miao 2014a], are registered to a volume. Another case is [Yu 2008,Yu 2011], where a bi-protocol was proposed to reconstruct a microscopic volume of a cell. The bi-protocol is composed by two sets of multiple slices imaging the same cell, which are captured with two different geometries. Both sets of slices are then registered to each other and a final volume is reconstructed using a simple interpolation strategy.

Let us recall that multi slice-to-volume registration involves lower computational complexity compared to 3D-3D registration methods. Furthermore, when compared to single slice methods, multi slice-to-volume registration certainly improves the robustness of the registration process by augmenting the image support. Dealing with several slices requires more computing power than the single slice case; however, on occasions like freehand US sweeps, slices contain redundant information that could be avoided. In that sense, as the complexity of the scheme is proportional to the number of input slices, [Wein 2008a] proposed a strategy which starts out by selecting only a few key-slices. These are used to estimate a rigid or affine transformation model, mapping the US slices to a CT volume. Since neighboring frames of the freehand US sweep contain overlapping information, only a few key slices are selected. They chose the most informative slices by selecting those with the highest image entropy. [Olesch 2011a] applied the same key-slice selection technique to the variational deformable registration framework proposed by [Heldmann 2009]. To this end, well distributed slices throughout the volume were considered, containing meaningful information in terms of entropy. In another work, [Olesch 2011b] extended this key-slice selection technique by introducing a focused registration strategy that only considers slices which are in a given region of interest (ROI).

In [Jiang 2007b], a multi slice-to-volume registration method -entitled Snapshot magnetic resonance imaging (MRI) with Volume Reconstruction (SVR)- is proposed to deal with reconstruction of MRI of moving subjectss. After imaging the target object of interest repeatedly (producing multiple overlapped stacks of slices of the same moving object) to guarantee sufficient sampling, one of the stacks is chosen.

Then, an iterative mutli slice-to-volume registration strategy is applied where the stacks are subsequently registered (in a rigid way) to the reference. Once all the stacks are registered, they are averaged to produce a new reference. Then, the number of temporally contiguous slices in every stack is reduced so that the data is divided into subpackages that are temporally contiguous although the slices in each subpackage may not be spatially adjacent. The process is then repeated: the subpackages are registered to the average image, which is again updated, and the number of slice per stack is once more decreased; this process is repeated until each slice is treated in isolation (reducing to the single slice-to-volume case) and the final average volume is reconstructed. In [Jiang 2007a, Jiang 2009] the idea was extended to allow the reconstruction of diffusion tensor images in the same scenario. In the same work, [Jiang 2007b] tested the benefit of using both, parallel and orthogonally acquired slices with prospectively acquired data and simulated cases as well. More recently, a fast multi-GPU accelerated framework to perform SVR was presented [Kainz 2015]. They proposed a fully parallel SVR approach for the reconstruction of high-resolution volumetric data from motion corrupted stacks of images. Parallelization is performed at two levels: (i) at the slice level, multiple slices are treated separately for large parts of the reconstruction process and (ii) since pixel/voxel based operations are independent of each other, they can be also executed in parallel. Authors claimed that their approach is five to ten times faster than the fastest currently available multi-CPU frameworks.

[Rousseau 2005, Rousseau 2006] proposed the registration of multiple sets of orthogonal 2D MRI slices into a high resolution MRI volume. The work of [Jiang 2007b] shares similarities some with them where, instead of reducing the number of slices per subpackage until each slice is treated in isolation, a two-step approach is performed. In this setting, a global alignment of the low resolution images was performed first and then each slice of the low resolution images was registered to the reconstructed high resolution volume. Only orthogonal sets of slices were considered for reconstruction.

The work by [Kim 2008b,Kim 2010a], built on top of [Rousseau 2005,Rousseau 2006], proposed a new approach entitled Slice Intersection Motion Correction (SIMC). It considers the registration process directly in terms of the intersections of each pair of slices in the stacks, avoiding the intermediate volume estimation process. Independent per-slice rigid transformations are estimated by minimizing a global energy function defined by the sum of dissimilarity measures of all intersection profiles between any two orthogonal slices. SIMC can be used in this scenario, considering intersecting lines of voxels instead of standard patch or global based calculation. [Kim 2008b, Kim 2010a] applied SIMC to the problem of reconstructing MR images of a moving human fetal brain. Since then, several extensions to this work have been published, like for example the work presented in [Kim 2010b] expanding the method to the case of Diffusion Tensor Images and [Kio Kim 2010, Kim 2011] which modified it in order to account for bias field inconsistency correction on fetal brain MR images.

An approach based on particle filtering (PF) was considered in [Nir 2011,Nir 2013,

Nir 2014] to deal with multi slice-to-volume registration of histological images to MR or US volumes. PF framework derives an optimal estimation of the parameters in a Bayesian fashion, tackling two of the main issues that arise from multimodal registration, such as susceptibility to initialization and optimal solutions.

3.1.6 Applications

A broad number of clinical settings and applications can benefit from slice-to-volume registration. In this review we classify them in two main categories.

3.1.6.1 Image Fusion and Image Guided Interventions (IGI)

Several medical procedures such as image guided surgeries and therapies [Fei 2002], biopsies [Xu 2014a], radio frequency ablation [Xu 2013], tracking of particular organs [Gill 2008] and minimally-invasive procedures [Liao 2013, Huang 2009] belong to this category. In this context, slice-to-volume registration brings high resolution annotated data into the operating room. Generally, pre-operative 3D images such as computed tomography (CT) or magnetic resonance images (MRI) are acquired for diagnosis and manually annotated by expert physicians. During the surgical procedure, 2D real time images are generated using different technologies (e.g. fluoroCT, US or interventional MRI slices). The alignment of intra-operative images with pre-operative volumes augments the information that physicians have access to, and allows them to navigate the volumetric annotation while performing the operation. These intra-operative images inherit lower resolution and quality than the pre-operative ones. Moreover, tissue shift collapse as well as breathing and heart motion during the procedure, cause elastic deformation in the images, what makes slice-to-volume registration an extremely challenging task. A statistically significant improvement in alignment has been demonstrated when comparing automatic methods to manual (human) results, showing the importance of automatic slice-to-volume registration algorithms in the context of image fusion and IGI [Fühwald 2009].

The pioneering work of [Fei 2001, Fei 2002, Fei 2003a, Fei 2004a, Fei 2004b] introduced iconic slice-to-volume registration to the problem of image fusion in the context of image guided surgeries. The motivation was that low-resolution Single Photon Emission Computed Tomography (SPECT) can be brought to operating by pre-registering it with a high-resolution MRI volume, which could be subsequently fused with live-time iMRI. That is how, by registering the high-resolution MR image with live-time iMRI acquisitions, Fei and coworkers could map the functional data and high-resolution anatomic information to live-time iMRI images for improved tumor targeting during thermal ablation.

[Birkfellner 2007] used slice-to-volume registration to fuse 2D fluoroCT with volumetric CT, which is a well know tool for image-guided biopsies in interventional radiology. In this case, the pre-interventional diagnostic high resolution CT with contrast agent is used to localize a lesion in the liver. However, during the inter-

vention, the lesion is no longer visible. Thus, localizing the slice of the CT that corresponds to the intra-operative fluoroCT allows doctors to find the lesions during the biopsy. This approach only considers rigid transformations. However, interventional procedures like radio frequency ablation (RFA) or image-guided biopsies, which use fluoroCT as image guiding technology, are performed while the patient is breathing continuously. Therefore, deformations should also be taken into account when registering with the pre-operative static CT image. The influence of such deformations and the reliability of performing non-rigid registration in such scenario was discussed in [Micu 2006, Lasowski 2008]. It was claimed that a 2D-3D non-rigid registration solution -based on the single low quality fluoroCT- cannot be precise as required to performed medical procedures. This is mainly due to the poor support in terms of liver anatomical features (mainly vessels) provided by the fluoroCT slices. They proposed to overcome this limitation by providing an adaptive visualization [Lasowski 2008] of the volume area surrounding the minimum estimated pose. This approach copes with the uncertainty in estimating the deformation and brings more information than a single registered slice. Their method performs rigid slice-to-volume registration, and includes views of the CT-Volume determined along flat directions of the out-of-plane motion parameters next to the minimum pose. Laparoscopic and endoscopic interventional procedures also exploit slice-to-volume registration. [San José Estépar 2009], proposed a method to register endoscopic and laparoscopic US images with pre-operative computed tomography volumes in real time. It is based on a new phase correlation technique called LEPART accounting for rigid registration. Other methods applying slice-to-volume registration on minimally invasive procedures can be found here [Heldmann 2010, Bao 2005].

3.1.6.2 Motion Correction / Volume Reconstruction

The second category is motion correction and volume reconstruction. Here, the goal is to correct misaligned slices when reconstructing a volume of a certain modality. A typical approach to solve this task consists in mapping individual slices within a volume onto another reference volume in order to correct the inter-slice misalignment. The popular map-slice-to-volume (MSV) method introduced this idea [Kim 1999]. More recently, applications of slice-to-volume registration to the same problem in different contexts like cardiac magnetic resonance (CMR) [Chandler 2008a, Elen 2010], fetal images [Seshamani 2013] and diffusion tensor imaging (DTI) [Jiang 2009] have shown promising results. In these problems, it is usually assumed that a single slice is coherent, in the sense that spatial inconsistency only happens at the inter-slice level.

Slice motion correction in the context of volume reconstruction, typically involves iterative registration of slices to a target volume. The target volume may be an anatomical reference or it could be reconstructed at each iteration using current estimates of slice motion, considering all possible views of the subject.

[Kim 1999] introduced the map-slice-to-volume (MSV) approach for inter-slice motion correction in time series of fMRI image. In such cases, head motion rep-

resents the major source of error in measuring intensity changes related to given stimuli in fMRI time series [Yeo 2004]. The aim of MSV is to retrospectively remap slices that are shifted by head motion to their spatially correct location using an anatomical MRI volume as reference. The MSV approach estimates 6-DOF rigid transformations independently for every fMRI slice, by minimizing the energy based on mutual information, using a Nelder-Mead downhill simplex method for optimization. This method was presented as an alternative to the slice-stack approach. Instead of considering slice-wise registration, it assumes stacks of slices being already registered among them, and ignores the inter-slice motion inherent to multi slice echo planar imaging (EPI) acquisition sequence (since each slice is excited at a sequential time interval). The MSV showed better performance than previous volume-to-volume registration methods, but the reliability of the estimated position parameters for the end cap slices was low due to the limited information support of smaller regions (less textured area and more background). [Park 2004] proposed to overcome this limitation using Joint Mapping of Slices into Volume (JMSV). JMSV is a multi-slice registration method that jointly estimates a rigid body transformations per slice, while penalizing the implied acceleration in the motion trajectory of the subject -i.e. abrupt changes in the motion parameters of sequentially acquired slices. Other extensions to the standard MSV include (i) accounting for deformable registration [Kim 2000] through TPS transformation models; (ii) improving motion correction capability of MSV with concurrent iterative field-corrected reconstruction [Yeo 2004, Yeo 2008]; (iii) extensive evaluation of the activation detection performance of MSV and effects of temporal filtering of motion parameter estimates [Yeo 2006] and spin saturation effect [Kim 2008a]; (iv) improving the MI matching criterion estimation in the low-complexity end-slices (slices near top or bottom of the head scans, presenting poorer information) by incorporating joint probability density functions of image intensities estimated from successfully registered center-slices in the same time-series.

After the fundamental work of [Kim 1999] and its extensions (most of them related to fMRI image reconstruction), another problem requiring to correct for inter-slice motion started to attract attention from the community: fetal brain MR imaging. In this case, there is no anatomical reference available to be used as a target volume; therefore, a reference volume is calculated during the registration process using the current estimate of the slices. [Jiang 2007a] and [Rousseau 2006] established the basis of a new family of methods which rely on iterations of slice-to-volume registration and scattered data interpolation (SDI) to perform super-resolution reconstruction of moving objects -in particular, fetal brain MRI-. Both approaches share a similar and iterative slice-to-volume registration scheme, but differ on the SDI method: while [Rousseau 2006] use Gaussian kernel-based SDI, [Jiang 2007a] relied on a regular grid of control-points that acts as a cubic B-spline to perform SDI. In a more recent work, [Gholipour 2009, Gholipour 2010] criticized SDI-based approaches stating that none of these techniques can guarantee the convergence of the reconstruction to, at least, a local optimal solution. Thus, they proposed the use of a maximum likelihood (ML) error minimization approach,

which guarantees the convergence of the reconstructed image to match the motion-corrected slices through steepest descent error minimization. An extension to this method was presented in [Gholipour 2011], using a real-time sensor based tracking (i.e. a non-image based approach) to estimate an initial position of every slice, which is then refined through a retrospective slice-to-volume registration approach based on ML error minimization.

3.2 Graph-based Slice-to-Volume Deformable Registration

In the previous section, we presented a literature review about slice-to-volume registration of biomedical images. In the current section, we introduce a novel slice-to-volume registration framework, which models this challenging task as a discrete labeling problem. As mentioned, our work is pioneer in using discrete methods to solve this problem. We already discussed how image registration can (in general) be formulated using MRFs and discrete inference in section 2.3. Here we extend this approach to the particular case of slice-to-volume registration, and we propose three different derivations that can solve this problem.

In section 3.1.1 we give a general definition of slice-to-volume registration, which is reflected in equation 3.1. This definition contains a generic transformation model Θ that maps the 2D input image I to the corresponding slice from volume J. In our non-rigid discrete setting, that function includes a 2D-2D in-plane local deformation field $\hat{T}_D$ and a plane $\hat{\pi}[J]$ (i.e. a rigid transformation that specifies a bi-dimensional slice from the volume J). The general optimization problem defined in equation 3.1 is therefore reformulated as:

$$\hat{T}_D, \hat{\pi} = \arg\min_{T_D, \pi} \mathcal{M}(I \circ T_D(\boldsymbol{x}), \pi[J](\boldsymbol{x})) + \mathcal{R}(T_D, \pi), \tag{3.3}$$

where $\mathcal{M}$ represents the data similarity term and $\mathcal{R}$ the regularization term. The data term $\mathcal{M}$ measures the matching quality between the deformed 2D source image and the corresponding 3D slice. The regularization term $\mathcal{R}$ imposes certain constraints on the solution that can be used to render the problem well posed. It also imposes certain expected geometric properties on the extended (plane selection and plane deformation) deformation field. The plane $\hat{\pi}$, that minimizes the equation, indicates the location of the 3D volume slice that best matches the deformed source image. The deformation field $\hat{T}_D$ represents the in-plane deformations that must be applied to the source image in order to minimize the energy function.

The fundamental idea behind our approaches is quite intuitive: we aim at deforming a planar 2D grid in the 3D space, which encodes both the deformation field $\hat{T}_D$ and the plane $\hat{\pi}$ at the same time. This grid is super-imposed to the 2D source image and consists of control points that jointly represent the in-plane deformation and the current position of the 2D image into the 3D volume. The source image is positioned within the volume by applying different displacement vectors with

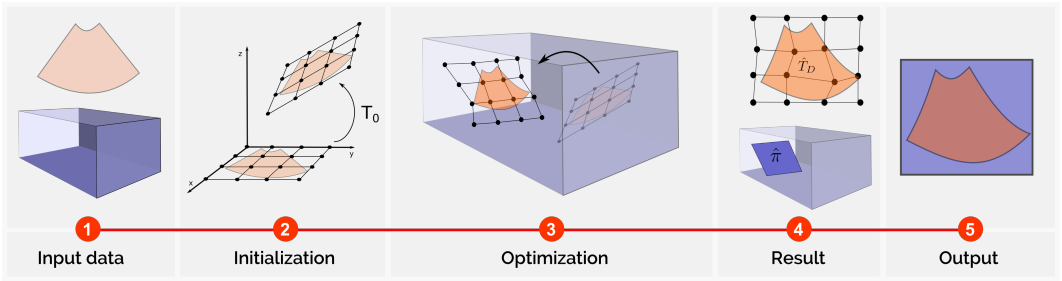


Figure 3.1: Basic workflow to perform slice-to-volume registration based on graphical models. (1) A 2D input image I and a 3D target volume J are given as input data. (2) A grid is superimposed to image I. The process is initialized using a 6-DOF rigid transformation T_0 that specifies the initial position of the grid within the volume J. (3) The grid is deformed by optimizing an energy function. (4) The plane $\hat{\pi}$ and the deformation field $\hat{T}_D$ are reconstructed from the final state of the optimized grid. (5) $\hat{T}_D$ is used to deform image I, and it is provided as output together with the corresponding slice $\hat{\pi}[J]$.

respect to the control points of the superimposed grid. These displacements are chosen such that a given energy (see Eq. 3.3) is minimized to best fit the matching criterion $\mathcal{M}$. Since they can be moved without any restriction, geometric constraints are imposed through the regularization term $\mathcal{R}$ in order to keep a smooth deformation field and a planar grid. Given that we impose a soft planar constraint, the resulting grid is approximately planar. Therefore, we reconstruct the final solution by projecting all the points into a regression plane which is estimated out of the current position of the points. The rigid transformation that indicates the position of the regression plane is considered as $\hat{\pi}$. Finally, the projected grid is interpreted as a 2D Free Form Deformation model (FFD) [Rueckert 1999] where each control point has local influence on the deformation and is used to approximate the dense deformation field $\hat{T}_D$ (see Figure 3.2). Having said that, alternative control point interpolation models can be used as well. Figure 3.1 illustrates this workflow.

This general formulation can be expressed through different discrete labeling problems on a graph by changing its topology, the label space definition and the energy terms. As we mentioned, in this work we propose three different approaches to derive slice-to-volume registration as a discrete graph labeling problem. First, we propose the so-called *overparameterized* method, which combines linear and deformable parameters within a coupled formulation on a 5-dimensional label space [Ferrante 2013]. The main advantage of such a model is the simplicity provided by its pairwise structure, while the main disadvantage is the dimensionality of the label space which makes inference computationally inefficient and approximate (limited sampling of search space). Motivated by the work of [Shekhovtsov 2008], we present a *decoupled* model where linear and deformable parameters are separated into two interconnected subgraphs which refer to lower dimensional label

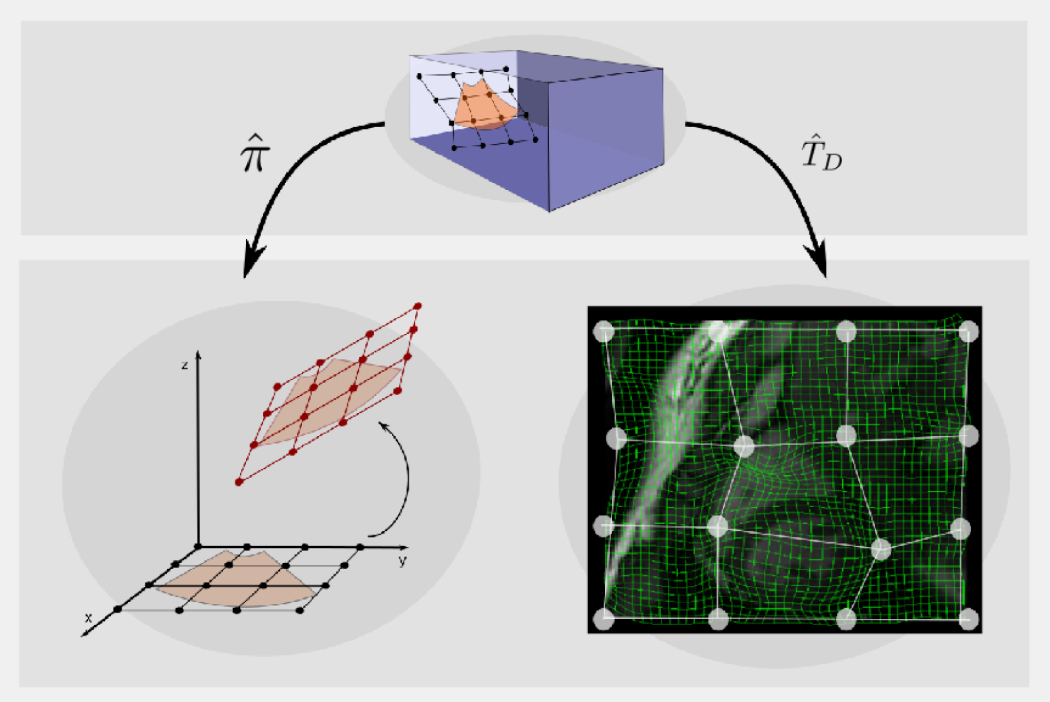
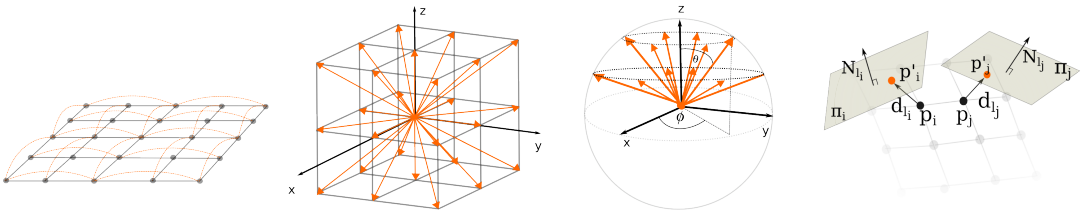


Figure 3.2: Recovering the rigid and deformable parameters. Since the resulting grid after inference is approximately planar, we recover the final solution by projecting all the points into a regression plane which is estimated out of the current position of the points. The rigid transformation $\hat{\pi}$ indicates the position of the regression plane. The projected grid is interpreted as a 2D FFD, used to approximate the dense deformation field $\hat{T}_D$.

spaces [Ferrante 2015b]. It allows us to reduce the dimensionality of the label space by increasing the number of edges and vertices, while keeping a pairwise graph. Finally, in the *high-order* approach [Ferrante 2015a], we achieve this dimensionality reduction by augmenting the order of the graphical model, using third-order cliques which exploits the expression power of this type of variable interactions. Such a model provides better satisfaction of the global deformation constraints at the expense of quite challenging inference.

3.2.1 Overparameterized Approach

Let us consider an undirected pairwise graph $G_O = \langle V, E \rangle$ super-imposed to the 2D image domain with a set of nodes V and a set of cliques E. The nodes V (a regular lattice) are interpreted as control points of the bi-dimensional quasi-planar grid that we defined in the previous section. The set of edges E is formed by regular 4-neighbors grid connections and some extra edges introduced to improve the propagation of the geometrical constraints (see figure 3.3.a). The vertices $v_i \in V$



(a) (b) (c) (d)

Figure 3.3: (a) Connectivity structure of the graph for a grid of size 5x5. The gray edges are standard 4-neighbor connections while the orange ones correspond to the extra cliques introduced to improve the geometrical constraints propagation. (b) Displacement vectors corresponding to the first three elements of a label from the overparameterized approach $\boldsymbol{d_{l_i}} = (d_x, d_y, d_z)$. (c) Unit vectors in spherical coordinates corresponding to the last two coordinates of a label from the overparameterized approach $\boldsymbol{N_{l_i}} = (\phi, \theta)$. (d) Displacement of the control points $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$ when the corresponding labels l_i associated to $(\boldsymbol{d_{l_i}}, \boldsymbol{N_{l_i}})$ and l_j associated to $(\boldsymbol{d_{l_j}}, \boldsymbol{N_{l_j}})$ are applied. The planes π_i and π_j are those that contain the control points $\boldsymbol{p_i} + \boldsymbol{d_{l_i}}, \boldsymbol{p_j} + \boldsymbol{d_{l_j}}$ and whose normals are $\boldsymbol{N_{l_i}}, \boldsymbol{N_{l_j}}$ respectively.

are moved by assigning them different labels $\boldsymbol{l_i} \in L$ (where L corresponds to the label space) until an optimal position is found. Γ_O represents a complete labeling that assigns one label to each variable (control point).

In order to deform the graph, we define a label space describing the inplane deformations and the plane selection variables. To this end, we consider a label space L where every label l_i is associated to 5-tuples $(d_x, d_y, d_z, \phi, \theta)$, where the first three parameters (d_x, d_y, d_z) define a displacement vector $\boldsymbol{d_{l_i}}$ in the cartesian coordinate system (see figure 3.3.b), and the angles (ϕ, θ) define a vector $\boldsymbol{N_{l_i}}$ on a unit sphere, expressed using spherical coordinates (see figure 3.3.c). Let us say we have a control point $\boldsymbol{p_i} = (p_{xi}, p_{yi}, p_{zi})$ and we assign to this point the label l_i, associated to the 5-tuple $(d_{xi}, d_{yi}, d_{zi}, \phi_i, \theta_i)$. So, the new point position $\boldsymbol{p'_i}$ after assigning the label is calculated using the displacement vector as given by the following equation:

$$\boldsymbol{p'_i} = (\boldsymbol{p_{xi}} + \boldsymbol{d_{xi}}, \boldsymbol{p_{yi}} + \boldsymbol{d_{yi}}, \boldsymbol{p_{zi}} + \boldsymbol{d_{zi}}). \tag{3.4}$$

Additionally, we define a plane π_i containing the displaced control point $\boldsymbol{p'_i}$ and whose unit normal vector (expressed in spherical coordinates and with constant radius $r = 1$) is $\boldsymbol{N_{l_i}} = (\phi_i, \theta_i)$. One of the most important constraints to be considered is that our transformed graph should have a quasi-planar structure, i.e. it should be similar to a plane; the plane π_i associated with every control point $\boldsymbol{p_i}$ is used by the energy term to take into account this constraint. Figure 3.3.d shows how to interpret the labels for two given points $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$.

The energy to be optimized is formed by data terms $G = \{g_i(\cdot)\}$ (or unary potentials) associated with each graph vertex and regularization terms $F = \{f_{ij}(\cdot,\cdot)\}$ (or pairwise potentials) associated with the edges. As it was mentioned, the first ones are typically used for encoding some sort of data likelihood, whereas the later ones act as regularizers and thus play an important role in obtaining high-quality results [Glocker 2011]. The energy associated to the overparameterized formulation is defined as:

$$\mathcal{P}_{\mathcal{O}}(\Gamma_O; G, F) = \sum_{i \in V} g_i(l_i) + \gamma \sum_{(i,j) \in E} f_{ij}(l_i, l_j), \tag{3.5}$$

where $l_i, l_j \in L$ are the labels assigned to the vertices $v_i, v_j \in V$ respectively. Then, we can define slice-to-volume registration as a MAP inference problem using the overparameterized formulation as:

$$\hat{\Gamma}_O = \arg\min_{\Gamma_O} \mathcal{P}_{\mathcal{O}}(\Gamma_O; G, F), \tag{3.6}$$

where $\hat{\Gamma}_O$ is the optimal labeling used to recover the rigid transformation and dense deformation field as explained in the previous section.

The formulation of the unary potentials that we propose is independent of the similarity measure. It is calculated for each control point given any intensity based metric δ capable of measuring the similarity between two 2D images (e.g sum of absolute differences, mutual information, normalized cross correlation). This calculation is done for each control point $\boldsymbol{p_i}$, using its associated plane π_i in the target image J and the source 2D image I. An oriented patch Ω_{l_i} over the plane π_i (centered at $\boldsymbol{p_i}$) is extracted from the volume J, so that the metric δ can be calculated between that patch and the corresponding area from the source 2D image (see Figure 3.4):

$$g_i(l_i) = \int_{\Omega_{l_i}} \delta(I(\boldsymbol{x}), \pi_i[J](\boldsymbol{x})) d\boldsymbol{x}. \tag{3.7}$$

One of the simplest and commonly used similarity measures is the Sum of Absolute Differences (SAD) of the pixel intensity values. It is useful in the monomodal scenario, where two images of the same modality are compared and, therefore, the grey intensity level itself is discriminant enough to determine how related are the two images. Its formulation in our framework is:

$$g_{SAD,i}(l_i) = \int_{\Omega_{l_i}} \mid I(\boldsymbol{x}) - \pi_i[J](\boldsymbol{x}) \mid d\boldsymbol{x}. \tag{3.8}$$

In multimodal scenarios, where different modalities are compared (e.g. CT with US images), statistical similarity measures such as Mutual Information (MI) are generally used since we can not assume that corresponding objects have the same intensities in the two images. MI is defined using the joint intensity distribution $p(i,j)$ and the marginal intensity distribution $p(i)$ and $p(j)$ of the images as:

$$g_{MI,i}(l_i) = -\int_{\Omega_{l_i}} \log \frac{p(I(\boldsymbol{x}), \pi_i[J](\boldsymbol{x}))}{p(I(\boldsymbol{x}))p(\pi_i[J](\boldsymbol{x})))} d\boldsymbol{x}. \tag{3.9}$$

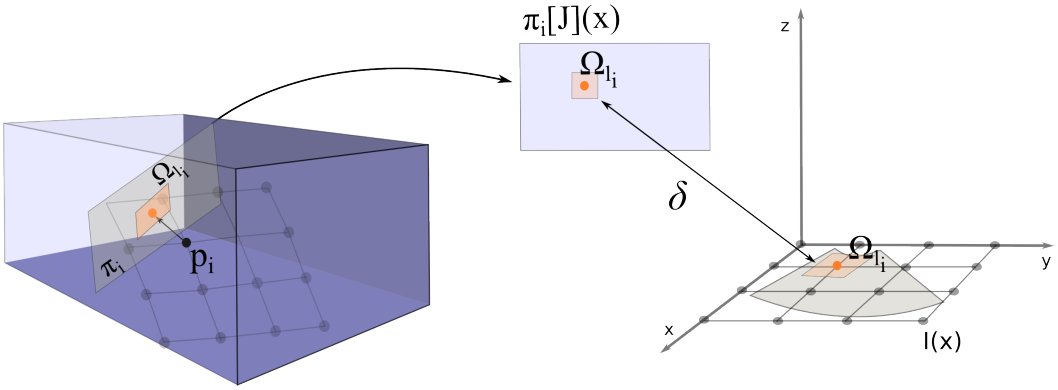


Figure 3.4: Data term formulation for the overparameterized approach. The points $\boldsymbol{x} \in \Omega_{l_i}$ are used to calculate the unary potential. $\pi[J](\boldsymbol{x})$ returns the intensity of the point in the 2D slice corresponding to the plane π_i in the 3D image, whereas $I(\boldsymbol{x})$ returns the 2D image intensity. δ represents the similarity measure.

As we can see in the previous examples, our framework can encode any local similarity measure defined over two two-dimensional images.

Let us now proceed with the definition of the regularization term. Generally, these terms are used to impose smoothness on the displacement field. In our formulation, the pairwise potentials are defined using a linear combination of two terms: the first (F_1) controls the grid deformation assuming that it is a plane, whereas the second (F_2) maintains the plane structure of the mesh. They are weighted by a coefficient α as indicates the following equation:

$$f_{ij}(l_i, l_j) = \alpha F_{1i,j}(l_i, l_j) + (1 - \alpha) F_{2i,j}(l_i, l_j). \tag{3.10}$$

The in-plane deformation is controlled using a distance preserving approach: it tries to preserve the original distance between the control points of the grid. Since this metric is based on the euclidean distance between the points, it assumes that they are coplanar. We use a distance that is symmetric, based on the ratio between the current position of the control points $\boldsymbol{p_i}, \boldsymbol{p_j}$ and their original position $\boldsymbol{p_{o,i}}, \boldsymbol{p_{o,j}}$:

$$\psi_{i,j}(\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}}) = \frac{|| \, (\boldsymbol{p_i} + \boldsymbol{d_{l_i}}) - (\boldsymbol{p_j} + \boldsymbol{d_{l_j}}) \, ||}{|| \, (\boldsymbol{p_{o,i}}) - (\boldsymbol{p_{o,j}}) \, ||}. \tag{3.11}$$

Once we have defined ψ_{ij}, the regularizer should fulfill two conditions: (i) it has to be symmetric with respect to the displacement of the points, i.e. it must penalize equally whenever the control points are closer or more distant; (ii) the energy has to be zero when the points are preserving distances and monotonically increasing with respect to the violation of the constraint. The following regularization term fulfills both conditions for a couple of nodes $i, j \in V$ labeled with labels l_i, l_j:

$$F_{1i,j}(l_i, l_j) = (1 - \psi_{i,j}(\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}}))^2 + (1 - \psi_{i,j}(\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}})^{-1})^2, \tag{3.12}$$

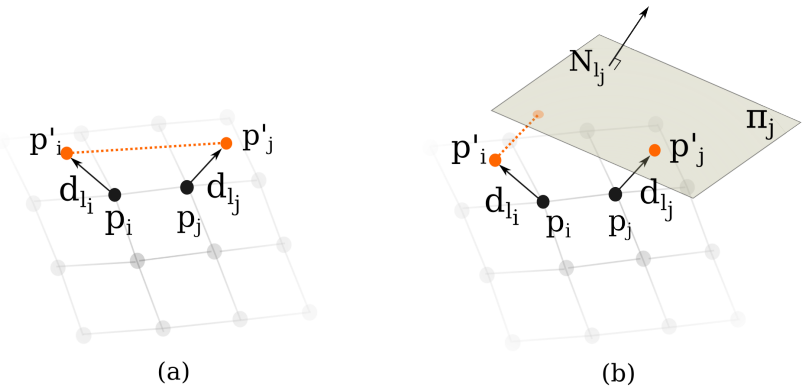


Figure 3.5: (a) In plane regularization term: the dotted line represents the distance used in F_1, i.e. the distance between the points assuming they are coplanar. (b) Plane structure regularization term: the dotted line represents the distance between one of the control points and the plane corresponding to the other one. This information is used to compute the term F_2.

The plane preservation term is based on the average distance between a given control point and the plane defined from the neighboring ones (see Figure 3.5.b). The aim is to maintain the quasi-planar structure of the grid. Given that the distance between a point and a plane is zero when the point lies on the plane, this term will be minimum when both of the control points are on the same plane.

The distance between a point $\boldsymbol{p} = (p_x, p_y, p_z)$ and a plane π defined by the normal vector $\boldsymbol{N} = (n_x, n_y, n_z)$ and the point $\boldsymbol{q} = (q_x, q_y, q_z)$ is calculated as:

$$D_\pi(\boldsymbol{p}) = \frac{\mid n_x(p_x - q_x) + n_y(p_y - q_y) + n_z(p_z - q_z) \mid}{\sqrt{n_x^2 + n_y^2 + n_z^2}}. \tag{3.13}$$

F_2 is defined using this distance (equation 3.13) and corresponds to the average of $D_{\pi_j}(\boldsymbol{p_i} + \boldsymbol{d_{l_i}})$ and $D_{\pi_i}(\boldsymbol{p_j} + \boldsymbol{d_{l_j}})$:

$$F_{2i,j}(l_i, l_j) = \frac{1}{2}(D_{\pi_j}(\boldsymbol{p_i} + \boldsymbol{d_{l_i}}) + D_{\pi_i}(\boldsymbol{p_j} + \boldsymbol{d_{l_j}})). \tag{3.14}$$

Recall that normal vectors in our label space are expressed using spherical coordinates with a fixed radius $r = 1$ (unit sphere). However, the formulation that we presented uses cartesian coordinates. Therefore, the mapping from one space to another is done as follows:

$$x = r\sin(\theta)\cos(\phi), y = s\sin(\theta)\sin(\phi), z = r\cos(\theta). \tag{3.15}$$

Note that such pairwise terms are non submodular since we include the current position of the points (which can be arbitrary) in their formulation and therefore the submodularity constraint is not fulfilled. In this context, even if there is no energy bounding that guarantees certain quality for the solution of the optimization

problem, good empirical solutions are feasible since we are in a pairwise scenario. Still, two issues do arise: (i) high dimensionality of the label space and consequently high computational cost, (ii) insufficient sampling of the search space and therefore suboptimal solutions. In order to address these issues while maintaining the pairwise nature of the methods, we propose the *decoupled* method inspired by [Shekhovtsov 2008]. We consider decoupling the label space into two different ones and redefining the topology of the graph, so that we can still capture rigid plane displacements and in-plane deformation.

3.2.2 Decoupled Approach

We propose to overcome the limitations of the overparameterized method by decoupling every node of the previous approach in two different ones: one modeling the in-plane deformation and another the position of the plane. This is somewhat analogous to creating two separated graphs of the same size and topology corresponding to different random variables and label spaces. Once spaces have been decoupled, different sampling strategies can be used for them. Another advantage of this approach is that we can define distinct regularization terms for edges connecting deformation nodes or plane position nodes. It allows to regularize in a different way the deformation and the plane position, imposing alternative geometrical constraints for every case.

Since data term computation requires the exact location of the node, both position and deformation labels are necessary. Both graphs can thus be connected through a pairwise edge between every pair of corresponding nodes. Therefore, new pairwise potentials are associated with these edges in order to encode the matching measure.

Formally, the decoupled formulation consists of an undirected pairwise graph $G_D = \langle V, E \rangle$ with a set of nodes $V = V_I \cup V_P$ and a set of cliques $E = E_I \cup E_P \cup E_D$. V_I and V_P have the same cardinality and 4-neighbor grid structure. Nodes in V_I are labeled with labels that model in-plane deformation, while labels used in V_P model the plane position. Edges from E_I and E_P correspond to classical grid connections for nodes in V_I and V_P respectively; they are associated with regularization terms. Edges in E_D link every node from V_I with its corresponding node from V_P, creating a graph with a three dimensional structure; those terms encode the matching similarity measure. Note that E_I and E_P can be extended with the same type of extra edges defined in section 3.2.1 (see figure 3.3.a) to improve the satisfaction of the desired geometrical constraints. The proposed graphical model is depicted in figure 3.6 and corresponds to a pairwise MRF, without unary potentials.

We define two different label spaces, one associated with V_I and one associated with V_P. The first label space, L_I, is a bidimensional space that models in-plane deformation using displacement vectors $\boldsymbol{d}_{l^I} = (d_x, d_y)$ associated to labels l^I. The second label space, L_P, indicates the plane in which the corresponding control point is located and consists of labels l^P representing different planes. In order to specify the plane and the orientation of the grid on it, we consider an orthonormal basis

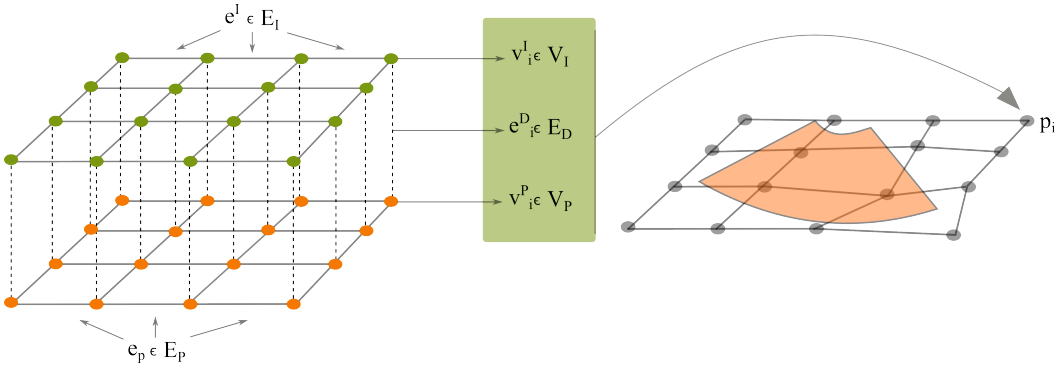


Figure 3.6: Structure of the decoupled graph. The green nodes are included in V_I while the orange ones in V_P, modeling in-plane deformations and plane position respectively. Edges connecting V_I nodes are part of E_I and those connecting V_P nodes are part of E_P; they are associated with regularization terms. Dotted lines represent cliques in E_D that encode the matching similarity measure. Using this information we can reconstruct an imaginary grid similar to the one presented in section 3.2.1. In the image we can appreciate how we associate two nodes of the graph with one control point of the grid.

acting on a reference point in this plane. Using this information, we can reconstruct the position of the control points of the grid. The planes parametrization is given by vectors (ϕ, θ, λ) associated to labels l^P, where angles ϕ and θ define a vector $\boldsymbol{N}$ over a unit sphere, expressed through its spherical coordinates (see figure 3.3.c). This value, together with parameter λ, defines the position of the plane associated with the given control point. This is an important advantage of our method: we could use prior knowledge to improve the way we explore the plane space, just by changing the plane space sampling method.

As it concerns the considered plane sampling method, the final position of every control point $\boldsymbol{p_k}$ of the grid is determined using the pairwise term between two graph nodes ($v_k^I \in V_I$ and $v_k^P \in V_P$) and their respective labels ($l_k^I \in L_I$ and $l_k^P \in L_P$). Imagine we have a plane π_k with normal vector $\boldsymbol{N}$ that contains the displaced control point $\boldsymbol{p_k} + \boldsymbol{d_{l_k^I}}$. Parameter λ indicates the magnitude of the translation we apply to π_k in the direction given by $\boldsymbol{N}$ in order to determine the plane's final position (see figure 3.7 for a complete explanation). Given that we can associate different planes to different control points (by assigning them different labels l^P), we need to impose constraints that will force the final solution to refer to a unique plane.

The energy that guides the optimization process involves three different pairwise terms, which encode data consistency between the source and the target, smooth-

ness of the deformation and unique plane selection:

$$\mathcal{P}_{\mathcal{D}}(\Gamma_D; I, P, D) = \alpha \sum_{(i,j)\in E_I} e^I_{i,j}(l^I_i, l^I_j) + \beta \sum_{(i,j)\in E_P} e^P_{i,j}(l^P_i, l^P_j) + \sum_{(i,j)\in E_D} e^D_{i,j}(l^I_i, l^P_j), \tag{3.16}$$

where α, β are scaling factors, $e^I_{i,j} \in I$ are *in-plane* deformation regularizers (associated to edges in E^I), $e^P_{i,j} \in P$ are *plane* consistency constraints (associated with edges in E^P) and $e^D_{i,j} \in D$ are *data* terms (associated with edges in E^D). l^I_i, l^P_i are labels from label spaces L_I and L_P respectively, while Γ_D is a complete labeling that assigns one label to every variable. Thus, we aim at solving the following MAP inference problem:

$$\hat{\Gamma}_D = \arg\min_{\Gamma_D} \mathcal{P}_{\mathcal{D}}(\Gamma_D; I, P, D). \tag{3.17}$$

Note that, in this case, we reconstruct the final position of the control points $\boldsymbol{p_k}$ as a combination of the labels l^I_k and l^P_k assigned by the optimal labeling $\hat{\Gamma}_D$. Once the grid is reconstructed, we can apply the same strategy (projecting the control points to the regression plane) to extract both, the ridig transformation and the dense deformation field, as explained at the beggining of section 3.2.

The data term is defined for every control point of the imaginary grid $\boldsymbol{p_k}$ using the information provided by two associated graph nodes. It is encoded in the pairwise term $e^D \in E^D$. To this end, we extract an oriented patch $\Omega_{l^I_k, l^P_k}$ over the plane π_k (centered at $\boldsymbol{p_k}$) from the volume J, so that the similarity measure δ can be calculated between that patch and the corresponding area over the source 2D image (see figure 3.7):

$$e^D_{i,j}(l^I_i, l^P_j) = \int_{\Omega_{l^I_i, l^P_j}} \delta(I(\boldsymbol{x}), \pi_k[J](\boldsymbol{x}))d\boldsymbol{x}. \tag{3.18}$$

We define two different regularization terms. The first controls the in-plane deformation; it is defined on V_I and corresponds to a symmetric distance preserving penalty:

$$e^I_{i,j}(l^I_i, l^I_j) = (1 - \psi_{i,j}(\boldsymbol{d_{l^I_i}}, \boldsymbol{d_{l^I_j}}))^2 + (1 - \psi_{i,j}(\boldsymbol{d_{l^I_i}}, \boldsymbol{d_{l^I_j}})^{-1})^2, \tag{3.19}$$

where $\psi_{i,j}$ is the distance defined in equation 3.11.

The second term penalizes inconsistencies in terms of plane selection, and is defined on V_P. We use the earlier defined (at is concerns the overparameterized model, in equation 3.14) point-to-plane distance:

$$e^P_{i,j}(l^P_i, l^P_j) = \frac{1}{2}(D_{\pi_j}(\boldsymbol{p_i}') + D_{\pi_i}(\boldsymbol{p_j}')). \tag{3.20}$$

where $\boldsymbol{p_i}'$ and $\boldsymbol{p_j}'$ are the positions after applying label l^P_i, l^P_j to $\boldsymbol{p_i}$, $\boldsymbol{p_j}$ respectively.

Note that these terms are similar to the ones of the former approach. However, there is an important difference regarding the parameters they use. In case

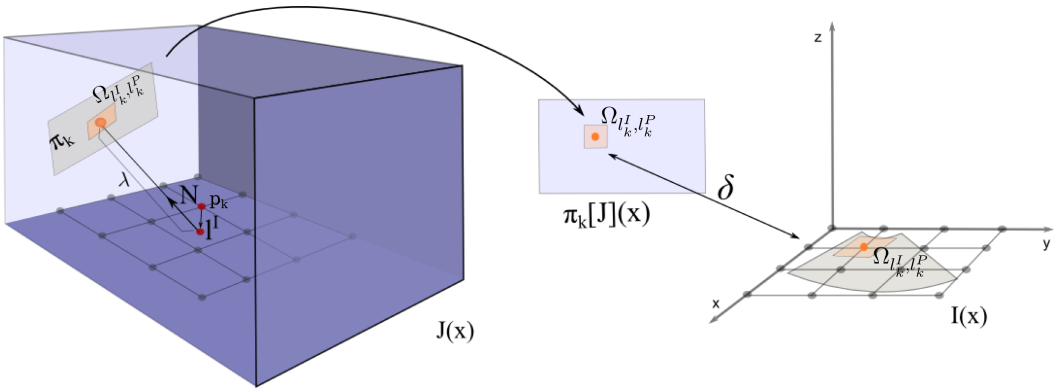


Figure 3.7: Data term formulation for the decoupled approach. It is similar to the formulation shown in figure 3.4, but it combines labels from different label spaces. The points $\boldsymbol{x} \in \Omega_{l_k^I,l_k^P}$ are used to calculate the unary potential. $\pi_k[J](\boldsymbol{x})$ returns the intensity of the point in the 2D slice corresponding to the plane π_k in the 3D image, whereas $I(\boldsymbol{x})$ returns the 2D image intensity. δ represents the similarity measure. In order to compute the final position of the sampled patch in the volume, the displacement $\boldsymbol{d}_{\boldsymbol{l_k^I}} = (d_x, d_y)$ associated to the in-plane deformation label l_k^I is applied to the corresponding imaginary grid point $\boldsymbol{p_k}$. Then, label tuple (N, λ) associated to the plane selection label l_k^P is used: the point is translated in the direction given by vector N as indicates scalar λ. Finally, the patch $\Omega_{l_k^I,l_k^P}$ is sampled from plane π_k with normal N, centered at the displaced point $\boldsymbol{p_k}$ (in orange).

of the overparameterized approach, parameters are always 5-dimensional labels. In the current approach, parameters are at most 3-dimensional, thus reducing the complexity of the optimization process while also allowing a denser sampling of the solution space. Conventional pairwise inference algorithms could be used to optimize the objective function corresponding to the previously defined decoupled model. Such a model offers a good compromise between expression power and computational efficiency. However, the pairwise nature of such an approach introduces limited expression power in terms of energy potentials. The smoothness (regularization) terms with second order cliques are not invariant to linear transformations such as rotation and scaling [Glocker 2009c], while being approximate in the sense that plane consistency is imposed in a rather soft manner. These concerns could be partially addressed through a higher order formulation acting directly on the displacements of the 2D grid with 3D deformation labels. Furthermore, the data term is just a local approximation of the real matching score between the deformed source 2D image and the corresponding target plane; by introducing high-order terms we could define it more accurately.

3.2.3 High-Order Approach

The new formulation consists of an undirected graph $G_H = \langle V, E \rangle$ with a set of nodes V and a set of third-order potentials $E = E_D \cup E_R$. The nodes are control points of our two-dimensional quasi-planar grid and they are displaced using 3D vectors $\boldsymbol{d_{l_i}} = (d_x, d_y, d_z)$ associated to labels $l_i \in L_H$. We define two types of cliques in E. Cliques in E_D are triplets of vertices with a triangular shape and they are associated with data terms. Those in E_R, are collinear horizontal and vertical third-order cliques associated with regularization terms.

Unlike the previous methods, which require extra labels to explicitly model the plane selection, high-order potentials explicitly encode them. Furthermore, third-order triangular cliques can also explicitly encode data terms, since the corresponding plane can be precisely determined using the position of these 3 vertices. We use triplets of collinear points for regularization terms. According to [Kwon 2008], this allows us to encode a smoothness prior based on the discrete approximation of the second-order derivatives using only the vertices' position. Therefore, we define a simple three dimensional label space of displacement vectors which is sampled as shown in figure 3.3.b.

The energy to be minimized consists of data terms D_{ijk} associated with triangular triplets of graph vertices $(i, j, k) \in E_D$ and regularization terms R_{ijk} associated with collinear horizontal and vertical triplets $(i, j, k) \in E_R$. It is defined in the following equation:

$$\mathcal{P}_{\mathcal{H}}(\Gamma_H; D, R) = \sum_{(i,j,k)\in E_D} D_{ijk}(l_i, l_j, l_k) + \gamma \sum_{(i,j,k)\in E_R} R_{ijk}(l_i, l_j, l_k), \tag{3.21}$$

where γ is a scaling factor and l_i is a label associated to a displacement vector $\boldsymbol{d_{l_i}} = (d_x, d_y, d_z)$ assigned to the node i. Γ_H is a labeling that assigns one label to every variable used to reconstruct the final rigid transformation and dense deformation field. Slice-to-volume registration is thus posed as a MAP inference problem on a high-order graph following:

$$\hat{\Gamma}_H = \arg\min_{\Gamma_H} \mathcal{P}_{\mathcal{H}}(\Gamma_O; D, R). \tag{3.22}$$

The data term is defined over a disjoint set of triangular cliques, covering the entire 2D domain, as shown in figure 3.8.a. Its formulation is independent of the similarity measure δ and it is calculated for each clique $\boldsymbol{c} = (i, j, k) \in E_D$ using the source 2D image I and the corresponding plane $\pi_d[J]$ extracted from the target volume J, defined by the three control points of the clique. For a given similarity measure δ, the data term associated with the clique $\boldsymbol{c}$ is thus defined as:

$$D_{ijk}(l_i, l_j, l_k) = \int_{\Omega_{l_i,l_j,l_k}} \delta(I(\boldsymbol{x}), \pi_d[J](\boldsymbol{x}))d\boldsymbol{x}, \tag{3.23}$$

where $\boldsymbol{x} \in \Omega_{l_i,l_j,l_k}$, and Ω_{l_i,l_j,l_k} corresponds to the triangular area defined by the control points of clique $\boldsymbol{c} = (i, j, k)$ over the plane $\pi_d[J]$, after applying the corre-

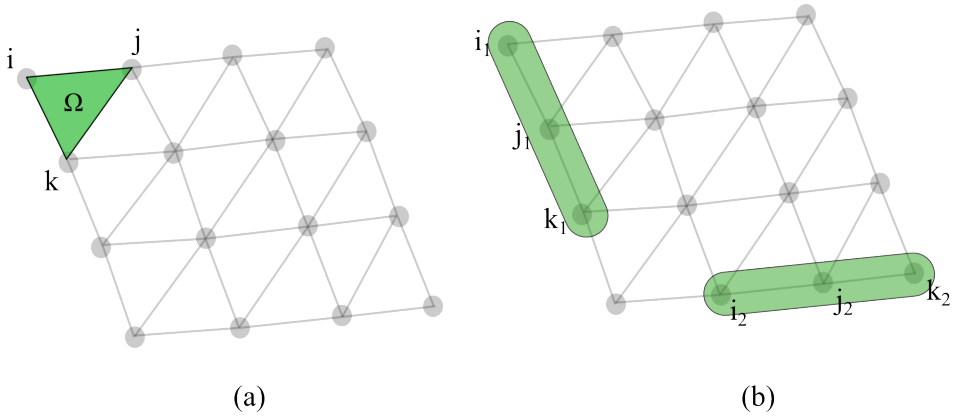


Figure 3.8: Different types of cliques used in the formulation. (a) Example of a triangular clique used for data term computation. The green patch Ω corresponds to the clique (i, j, k) and it is used to calculate the data term. (b) Examples of vertical (i_1, j_1, k_1) and horizontal (i_2, j_2, k_2) collinear third-order cliques used to regularize the grid structure.

sponding labels l_i, l_j, l_k to the vertices.

Smoothness and plane consistency are also imposed using higher order cliques. We define a clique for every set of three collinear and contiguous grid nodes (in horizontal and vertical directions as depicts figure 3.8.b). We also introduce extra cliques formed by nodes that are collinear but not contiguous. The aim is to propagate the regularization so that the planar structure is conserved. The regularization term, as previously, seeks to satisfy the plane structure of the grid and the smoothness nature of the in-plane deformations.

Planar consistency can be easily enforced by propagating a null second-derivative constraint among collinear triplets of points. In fact, a null second-derivative for these cliques does not impose just a planarity constraint but it also aims at regularizing the grid structure. Thanks to the third-order cliques, we can accurately approximate a discrete version of the second-order derivative [Kwon 2008]. Given three contiguous control points $\boldsymbol{p_i}, \boldsymbol{p_j}, \boldsymbol{p_k}$ and their corresponding displacement vectors $\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}}, \boldsymbol{d_{l_k}}$ associated to the labels l_i, l_j, l_k, it can be approximated as follows: $|| \, (\boldsymbol{p_i} + \boldsymbol{d_{l_i}}) + (\boldsymbol{p_k} + \boldsymbol{d_{l_k}}) - 2 \cdot (\boldsymbol{p_j} + \boldsymbol{d_{l_j}}) \, ||$.

Based on this idea, we define the following energy term that is proportional to the second derivative, and normalized with the original distance between the control points, d:

$$R^A_{ijk}(l_i, l_j, l_k) = \frac{|| \, (\boldsymbol{p_i} + \boldsymbol{d_{l_i}}) + (\boldsymbol{p_k} + \boldsymbol{d_{l_k}}) - 2 \cdot (\boldsymbol{p_j} + \boldsymbol{d_{l_j}}) \, ||^2}{d^2}, \tag{3.24}$$

In-plane deformation smoothness is reinforced in the same manner as the previous models, through a symmetric distance preserving approach. For the sake of clarity, we redefine equation 3.12 as $\Psi_{ij}(l_i, l_j) = (1 - \psi_{i,j}(\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}}))^2 + (1 -$

$\psi_{i,j}(\boldsymbol{d}_{\boldsymbol{l}_i}, \boldsymbol{d}_{\boldsymbol{l}_j})^{-1})^2$, and we apply it to both pairs of contiguous points that form the clique (i, j, k):

$$R^B_{ijk}(l_i, l_j, l_k) = \frac{\Psi_{ij}(\boldsymbol{d}_{\boldsymbol{l}_i}, \boldsymbol{d}_{\boldsymbol{l}_j}) + \Psi_{jk}(\boldsymbol{d}_{\boldsymbol{l}_j}, \boldsymbol{d}_{\boldsymbol{l}_k})}{2}. \tag{3.25}$$

The equation that regularizes the grid is a weighted combination of both terms R^A_{ijk} and R^B_{ijk}:

$$R_{ijk}(l_i, l_j, l_k) = (1 - \alpha)R^A_{ijk}(l_i, l_j, l_k) + \alpha R^B_{ijk}(l_i, l_j, l_k), \tag{3.26}$$

where α represents a weighting factor used to calibrate the regularization term.

3.3 Results and Discussion

Let us now proceed with a systematic evaluation of the proposed methods. One of the main aspects shared across methods is the inference algorithms used to produce the desired solution.

3.3.1 Inference Methods

Depending on their cardinality and regularity, objective functions can be optimized using a variety of discrete optimization algorithms which offer different guaranties. It must be noted that the regularization terms presented in our three models are non submodular, since we include the current position of the points (which can be arbitrary) in their formulation. Therefore, submodularity constraint is fulfilled neither in the pairwise nor in the high-order terms (for a clear definition of submodularity in pairwise and high-order energies, we refer the reader to the work of [Ramalingam 2008]).

In [Ferrante 2013], the overparameterized approach was optimized using the FastPD algorithm [Komodakis 2007b] while for the decoupled [Ferrante 2015b] and the higher order models [Ferrante 2015a], we consider loopy belief propagation networks. For the sake of fairness, in order to improve the confidence of the comparison among the three methods, in this work we reimplement it to be optimized with the same algorithms. Therefore, results in this work can not be directly compared with our previous works.

Given the variety of models presented in this work, we choose two different inference methods that can deal with arbitrary graph topologies and clique orders, coming from two standard inference algorithm classes: (i) Loopy Belief Propagation (LBP), a well know message passing algorithm that has been extensively used in the literature; and (ii) the Lazy Flipper (LF) by [Andres 2012], a move-making algorithm which is a generalization of the classical Iterated Conditional Modes (ICM) [Besag 1986] and has provided good approximations for several non-submodular models in different benchmarks . Both are approximate inference methods that can accommodate arbitrary energy functions, graph topologies and label

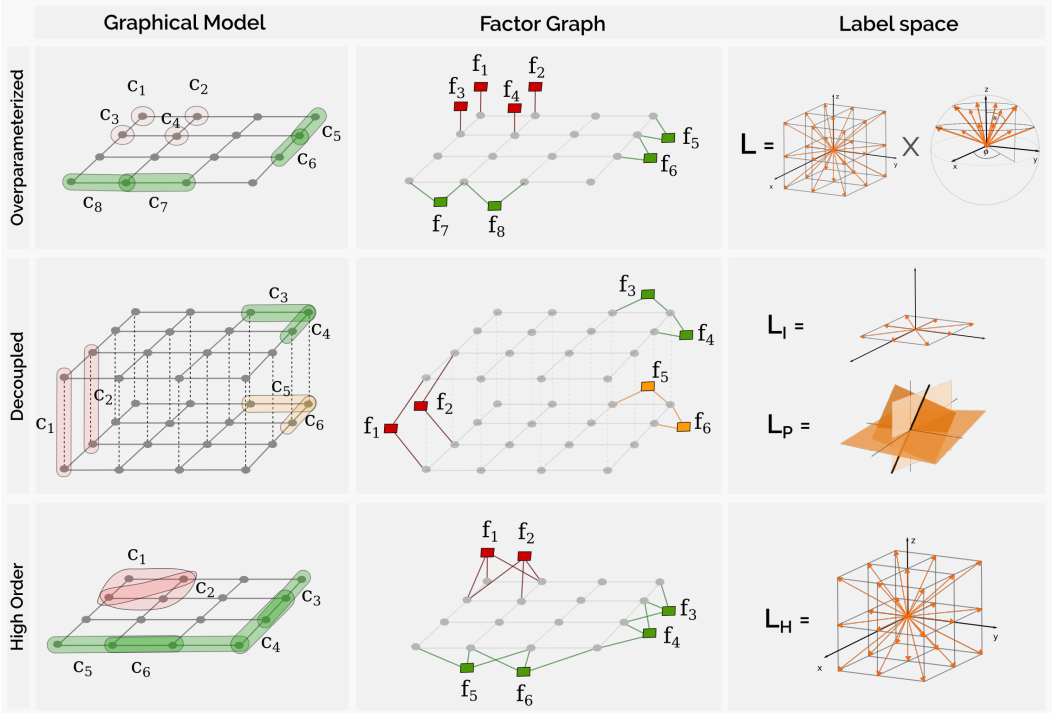


Figure 3.9: Factor graph derivation and labels spaces corresponding to the over-parameterized, decoupled and high-order approaches. It shows the equivalence between cliques c_i in the first column (unary, pairwise and high-order, depending on the model) and the corresponding factors f_i in the second column. In red, we observe the cliques and factors associated with data terms, while in green and orange we represent those associated with regularization terms. In the third column, we include a figure representing the label space associated to every model (orange vectors and planes are associated to different labels). Note that the overparameterized approach is defined as the Cartesian product between the displacements and plane selection labels, while in the decoupled approach these label spaces are independent.

spaces, and allow us to show how the three proposed approaches perform under different optimization strategies. We refer the reader to sections 2.4.1 and 2.4.2 for a brief description about LBP and LF, respectively.

We adopt the OpenGM2 library [Kappes 2013] which implements both inference methods, and makes it possible to perform fair comparisons. It requires construction of a factor graph for every scheme (we refer the reader to the appendix in section A.1 for a complete definition of factor graphs). Figure 3.9 shows a comparison between the three models and the derivation of the corresponding factor graph in each case.

Algorithm 1 Pseudo-code corresponding to the pyramidal approach based slice-to-volume registration

procedure 2D3DREGISTRATION(I: Source,J:Target,T_0:Initial guess)
  $G \leftarrow initializeGraph(T_0)$ ▷ Initialize the graph in the position indicated by T_0
  bestEnergy $\leftarrow \infty$
  for i=1 to gridLevels **do**
    $L \leftarrow updateLabelSpace(L, i)$ ▷ Update the label space for the given level
    for j = 1 to iterationSteps **do**
      newEnergy, newLabeling $\leftarrow$ optimizeGraphicalModel(G, L)
      if newEnergy $<$ bestEnergy **then**
        $applyLabeling(G$, newLabeling)
        bestEnergy = newEnergy
      end if
      $refineLabelSpace()$
    end for
  end for
  return bestEnergy, G
end procedure

3.3.2 Experimental Validation

We compute results on two different datasets for the three methods, using the two inference algorithms (LBP and LF) in order to validate both the resulting 2D-2D deformation field and the final plane estimation. The first dataset is a monomodal MRI heart dataset while the second one corresponds to multimodal US-MRI brain images.

We also run the same experiments using a continuous approach to estimate a rigid transformation model that serves as baseline for comparison. The continuous optimization is performed using the Simplex algorithm [Nelder 1965], which is a gradient-free method that has been successfully used to deal with continuous formulations of slice-to-volume registration, as discussed in section 3.1.4.1. We run the continuous optimization until convergence, which was always achieved in a few seconds. We optimized a global energy where the similarity measure was computed for the complete image, since no local deformation model is considered.

In the three methods, we adopt a pyramidal approach, using different grid resolution levels, from coarse to fine spacing between the control points. For each grid resolution, some iterations of the registration algorithm are performed, choosing the optimal set for each one and updating the control point positions with this information. During the inner iterations of one grid level, the size of the displacement vectors that form the deformation label space as well as the parameter variation of the plane label space are reduced in order to improve the search space sampling. A

pseudocode of the algorithm is shown in Algorithm 1. For every registration case, we run the inference algorithm several times (more precisely, the inference method is executed a number of times equal to the product between grid refinement levels and label refinement levels). For a single execution of both discrete inference methods, we use the same compound stopping criterion based on the energy gap between iterations and maximum running time. The algorithms run until the energy improvement between two iterations is smaller than a fraction of the current energy (we use $\varepsilon = 0.01\%$) or until a timeout is reached (we use 60 seconds as timeout). For LF we use a maximum depth of 2 (for details about LF, we refer the reader to section 2.4.2 or to [Kappes 2013]).

In the following subsections we describe the datasets and present quantitative results.

Monomodal Dataset Experiment

The monomodal dataset is derived from a temporal series of 3D heart MRI volumes. It consists of 10 sequences of 19 MRI slices which have to be registered with an initial volume. The slices are extracted from random positions in the volumes while satisfying spatio-temporal consistency. The ground truth associated with this dataset refers to the rigid transformation used to extract every 2D slice of every sequence (it is used to validate the plane estimation or rigid registration) and a segmentation mask of the left endocardium, that can be used to validate the quality of the estimated deformation field.

The dataset is generated from a temporal series of 3D heart MRI volumes M_i as shown in figure 3.10. For a given sequence in the heart dataset, every 2D slice I_i is extracted from the corresponding volume M_i at a position which is calculated as follows. Starting from a random initial translation $T_0 = (T_{x_0}, T_{y_0}, T_{z_0})$ and rotation $R_0 = (R_{x_0}, R_{y_0}, R_{z_0})$, we extract the first 2D slice I_0 from the initial volume M_0. Then, gaussian noise is added to every parameter of the transformation in order to generate the position of the next slice at the next volume. We use $\sigma_r = 3°$ as rotation and $\sigma_t = 5$mm as translation parameters. Those parameters generate maximum distances of about 25mm between the current and the succeeding plane. In this way, we generate 2D sequences that correspond to trajectories inside the volumes. Since the initial 3D series consists of temporally spaced volumes of the heart, there are local deformations between them due to the heartbeat; therefore, extracted slices are also deformed.

The resolution of the MRI volume is $192 \times 192 \times 11$ voxels and the voxel size is 1.25mm $\times$ 1.25mm $\times$ 8mm. The slices of the 2D sequences are 120×120 pixels with a voxel size of 1.25mm $\times$ 1.25mm.

Experiments for the 3 methods were performed using equivalent configurations. In all of them we used 3 grid refinement levels, 4 steps of label refinement per grid level, initial grid size of 40mm and minimum patch size (for similarity measure calculation) of 20mm. In case of the overparameterized approach we used $\alpha = 0.8$, $\gamma = 1$ and 342 labels; for the decoupled approach we used $\alpha = 0.8$, $\beta = 0.2$, 25

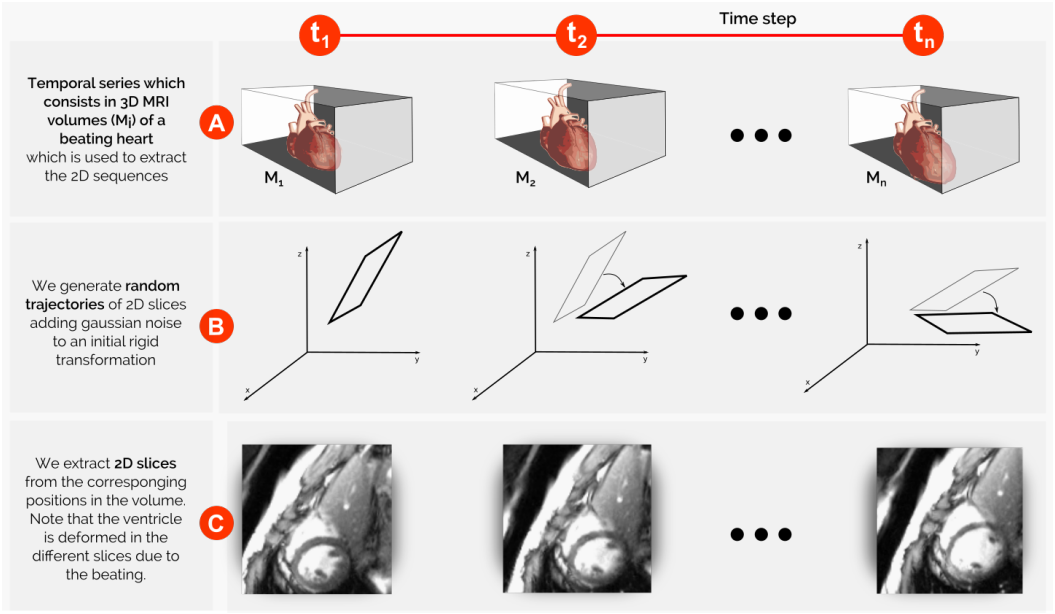


Figure 3.10: Heart dataset construction. Given a series of 3D MRI volumes of a beating heart (A), we extract ten different random trajectories (B). Every trajectory is composed of twenty different positions from which we extract the 2D slices (C).

labels in the 2D deformation space and 91 in the plane selection space; and finally, for the high-order approach we used $\alpha = 0.5$, $\gamma = 1.10$ and 19 labels. Parameters α, β, γ were chosen using cross-validation. The number of labels in every label space was chosen to make the search spaces as similar as possible.

Results are reported (for every approach and every inference method) for 10 sequences of 19 images, giving a total of 190 registration cases. We use SAD as similarity measure since we are dealing with monomodal registration. The idea is to register every 2D slice I_i (which plays the role of an intra-operative image) to the same initial volume M_0 (which acts as the pre-operative image). The resulting position of the slice I_i is used to initialize the registration of slice I_{i+1}.

Figure 3.13 shows results in terms of rigid transformation estimation. We measured the distance between the transformation parameters, and reported the average of the 190 registration cases. It resulted in less than 0.02rad (1.14°) for rotation and less than 1.5mm for translation parameters in all the discrete approaches and optimization methods. The decoupled method outperforms the others by orders of magnitude in terms of reduction of the standard deviation and the mean error.

To measure the influence of the deformation in the final results, we use the dataset annotations. We compute statistics for the segmentation overlapping at three different stages: before registration (i.e. between the source image and the target volume slice corresponding to the initial transformation), after rigid registration (i.e. between the source image and the target volume slice corresponding to

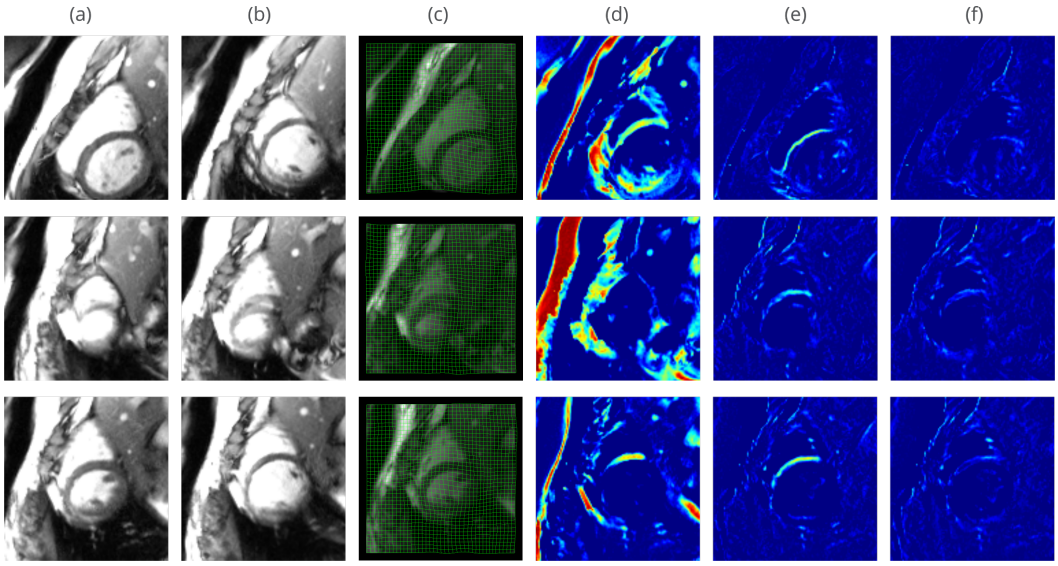


Figure 3.11: Slices extracted from three different sequences of the heart dataset before and after registration. Input slice (a) is initialized in a position specified by a rigid transformation within the volume, whose slice corresponds to (b). After deformable registration, a deformation field (c) is estimated. (d) shows the difference between initial images (a) and (b), while (e) shows the difference between (a) and the corresponding slice extracted after rigid registration. Finally, (f) corresponds to the results after deformable registration (i.e., the difference between the deformed version of slice (a) and the slice corresponding to the estimated transformation). Red indicates bigger differences between the images. Note how these values are changing before (d), after rigid (e) and after deformable (f) registration.

the estimated transformation) and after deformable registration (i.e. between the deformed source image and the target volume slice corresponding to the estimated transformation). We evaluate accuracy computing DICE coefficient, Hausdorff distance and Contour Mean Distance (CMD). We also provide sensitivity (which measures how many pixels from the reference image are correctly segmented in test image) and specificity (which measures how many pixels outside the reference image are correctly excluded from the test image) coefficients to complete the analysis. Results presented in figure 3.14 show the mean and standard deviation of the indicators at the three stages, for the three approaches and the two inference methods. Observe that results improve at each stage, achieving DICE coefficient of around 0.9 after deformation. Hausdorff distance and CMD decrease at each stage until a total reduction of around 66%. Decoupled method still outperforms the others after deformation in all the indicators, and presents a substantial improvement in terms of standard deviation reduction with respect to them (it is consistent with the results we showed in figure 3.13 for the rigid parameters). All the three discrete methods outperform the continuous rigid baseline. Figure 3.15 complements these

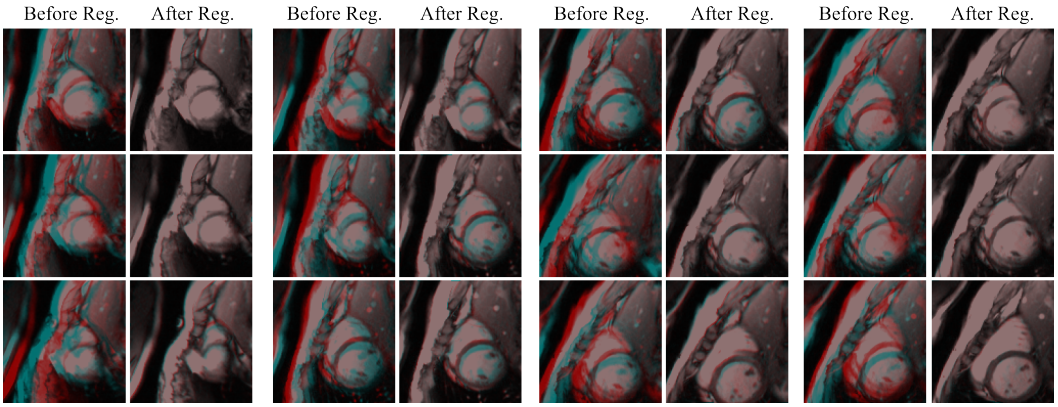


Figure 3.12: 12 registration cases of the same sequence, before and after deformable registration. The overlapping images (in light blue we show the source image and in red the target) showed before registration corresponds to the source image and a slice taken from the volume at the initial position. The overlapping after registration corresponds to the deformed source image and the slice taken from the volume at the estimated plane position.

results by showing DICE values disaggregated per sequence, while figures 3.11 and 3.12 shows some qualitative results before and after registration.

Finally, in terms of running time, figure 3.16 presents the average value for the three approaches and the two inference methods, together with the distribution with respect to data cost computation and optimization time. As we can see, the decoupled method again outperforms the other two when inference is performed using LBP. We run all the experiments (brain and heart datasets) on an Intel Xeon W3670 with 6 Cores, 64bits and 16GB of RAM.

Multimodal Experiment

Another dataset was used to test our approaches on multimodal image registration. The dataset consists of a preoperative brain MRI volume (voxel size of $0.5\text{mm} \times 0.5\text{mm} \times 0.5\text{mm}$ and resolution of $394 \times 466 \times 378$ voxels) and 6 series of 9 US images extracted from the patient 01 of the database MNI BITE presented in [Mercier 2012]. The intra-operative US images were acquired using the prototype neuronavigation system IBIS NeuroNav. We generated 6 different sequences of 9 2D US images of the brain ventricles, with resolution around 161×126 pixels and pixel size of $0.3\text{mm} \times 0.3\text{mm}$. The brain ventricles were manually segmented in both modalities. The estimated position of the slice n was used to initialize the registration process of slice $n+1$. Slice 0 was initialized in a position near the ground truth using the rigid transformation provided together with the dataset. We computed statistics as we did in the previous experiment, but in this case based on the overlap between ventricle segmentations. Since we registered different modalities, we used

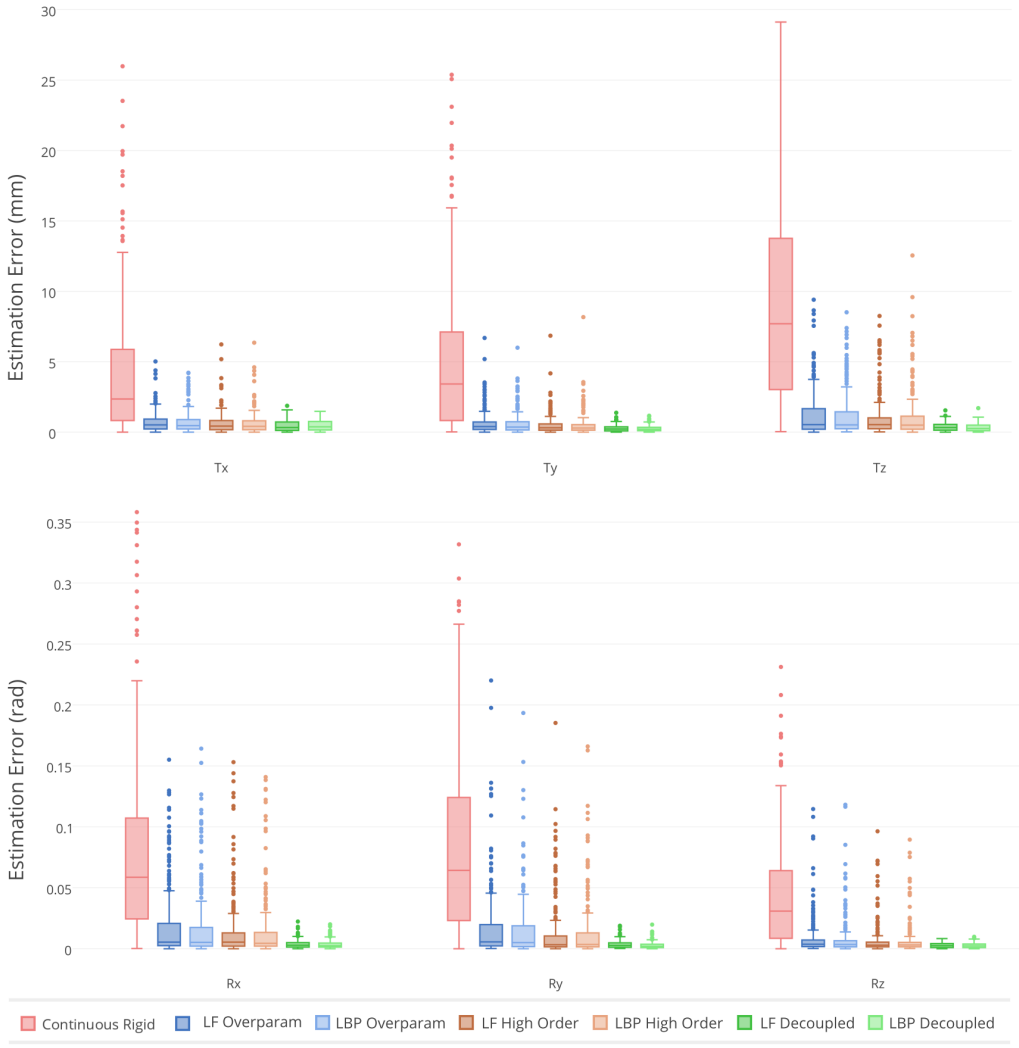


Figure 3.13: Rigid transformation estimation error for the heart dataset. We measured the distance for every one of the 6 rigid parameters, for the three approaches using LF and LBP as inference methods and for the continuous rigid approach. The discrete methods outperform the results obtained using the baseline. Independently of the discrete inference method, the decoupled approach outperforms the other two in terms of average and standard deviation of the estimated error, for all the 6 parameters.

Mutual Information as similarity measure instead of SAD.

Figure 3.14 summarizes the average DICE, specificity, sensibility, Hausdorff distance and Contour Mean Distance coefficients for all the series, while figure 3.16 reports the running times. Figure 3.15 complements these results by showing DICE values disaggregated per sequence. Figure 3.19 show some qualitative results. While

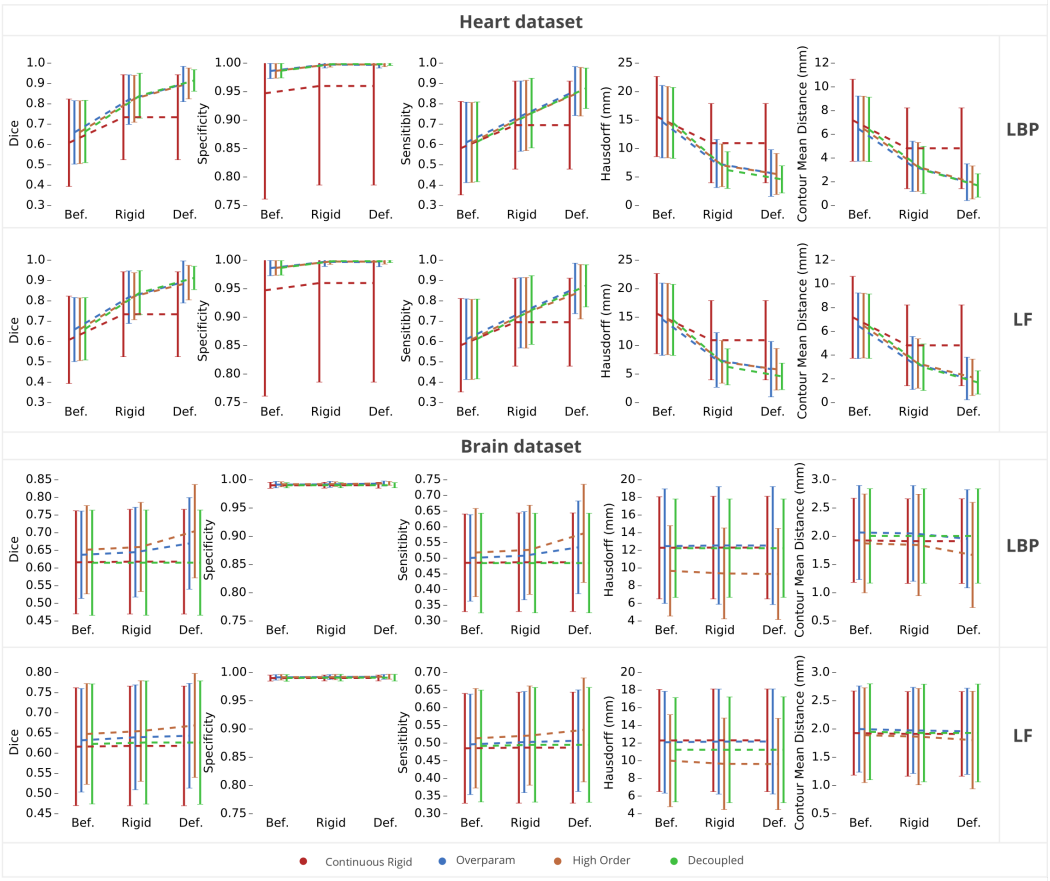


Figure 3.14: Segmentation overlapping statistics computed before, after rigid and after deformable registration for both datasets (190 registration cases for the heart dataset and 54 for the brain dataset). In the case of deformable registration, the source segmentation mask was deformed using the estimated deformation field. Results are reported for the continuous approach that only estimates rigid parameters using the Simplex method, and for three discrete methods (overparameterized, high-order and decoupled) using both inference strategies (LBP and LF).

the decoupled method does better in terms of computational time (independently of the inference method), the high-order method achieves better results in terms of segmentation statistics. It must be noted that, in this case, we are dealing with a more complex problem than in the case of monomodal registration; consequently, the increment obtained in terms of accuracy for both, rigid and deformable registration, is smaller. Given that we are dealing with highly challenging images of low resolution being heavily corrupted from speckle, those results are extremely promising. It is known to the medical imaging community that explaining correspondences between different modalities is a difficult task (see section 3.1.2).

In all brain experiments we use initial grid size of 8mm, minimum patch size of

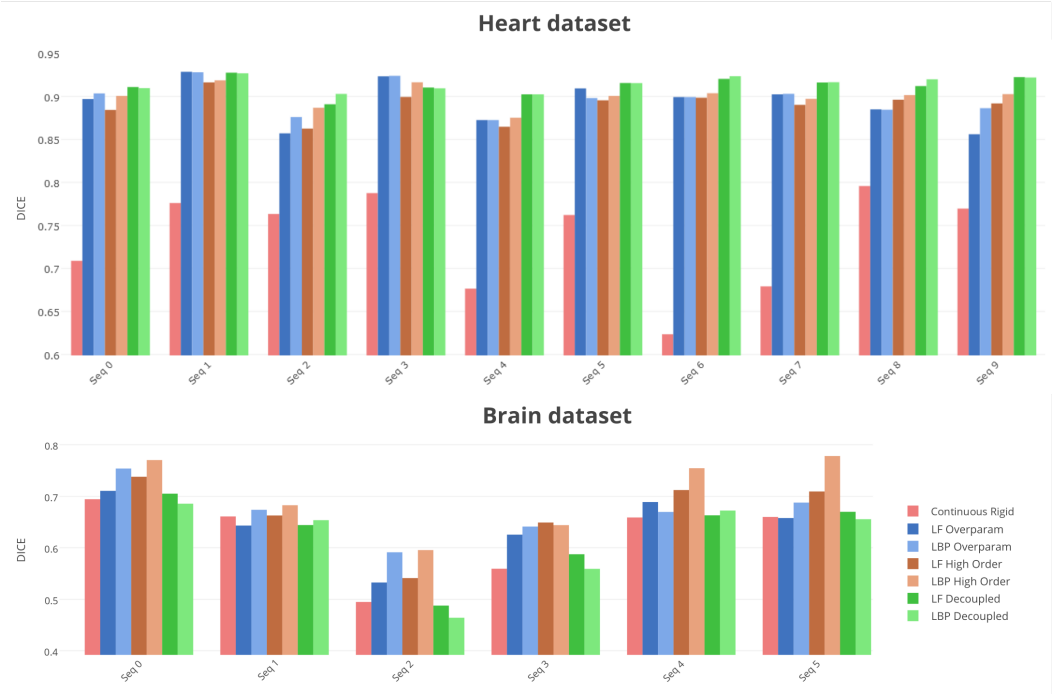


Figure 3.15: Final DICE (after deformation) comparison for every sequence (10 sequences in the heart dataset and 6 sequences in the brain dataset). Results are reported for the continuous approach that only estimates rigid parameters using the Simplex method, and for three discrete methods (overparameterized, high-order and decoupled) using both inference strategies (LBP and LF). In case of the heart dataset, decoupled method outperforms the other two in most part of the sequences. In the brain dataset, the high-order approach shows better performance in most cases. This is coherent with the aggregated results shown in Figure 3.14.

$13mm$, 16 bins to construct the mutual information histograms, 3 grid levels and 4 steps of label refinement per grid level. In case of the overparameterized approach we used $\alpha = 0.9$, $\gamma = 0.1$ and 342 labels; for the decoupled approach $\alpha = 0.015$, $\beta = 0.135$, 25 labels in the 2D deformation space and 91 in the plane selection space; finally, for the high-order approach $\alpha = 0.7$, $\gamma = 0.05$ and 19 labels. Parameters are chosen similarly as in the heart experiments.

3.3.3 Comparative Analysis

In this section, we aim at comparing different aspects of the three approaches we present in this paper, namely label spaces, graph topology and computational time. Without loss of generality, some assumptions are made regarding the models. First, we consider only square grids where N is the number of control points and consequently $\sqrt{N}$ is the number of nodes per side. Second, for the sake of simplicity we

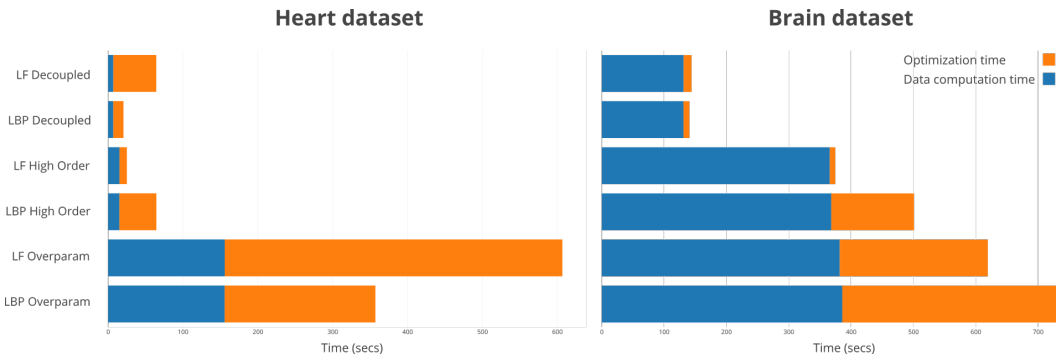


Figure 3.16: Average running time expressed in seconds, for one registration case, for the three approaches running on the heart dataset (a) and the brain dataset (b), using LF and LBP. Blue part corresponds to data cost computation while orange part corresponds to the optimization time. As we can observe, data cost computation represents a bigger portion of the total time in the brain dataset than in the heart dataset. This is due to the similarity measure: while in the monomodal case (heart) we use a simple SAD, in the multimodal case (brain) we need a more complex measure like mutual information. Note that data cost computation time remains constant when we vary the inference method (with small fluctuations due to operating system routines which ran during the experiment) but not across different models. Note that we do not compare with the rigid estimation using Simplex methods, since it estimates only 6 parameters (in a few seconds) while the discrete methods deal with hundreds.

do not consider the extra cliques introduced to improve the geometrical constraints propagation, since they are contemplated as an alternative strategy which may or may not be adopted.

Figure 3.17 shows a comparative analysis between the three approaches, using the two proposed inference methods, in terms of optimization time and final energy. Table 3.1 presents a compendium of the most critical parameters related to the proposed methods. Let us start with the label spaces. We divide them into two types: displacement space (L_D) and plane selection space (L_P). The first one contains the displacement vectors (2D or 3D, depending on the model) applied to the control points, while the second one contains the set of planes that can be chosen. In terms of cardinality of the label spaces, the overparameterized approach has the highest complexity, given by the cartesian product between the displacements and all the possible planes, $|L_D \times L_P|$. The decoupled model is dominated by the maximum of the cardinality of both label spaces, $\max(|L_D|, |L_P|)$. Finally, for the high-order model it depends only on $|L_D|$ since it is not necessary anymore to explicitly model which planes can be chosen - the triangles defined by the triplets of points describe a plane (and even more, a patch on this plane) by themselves. It clearly illustrates how we can reduce the complexity of a given label space by

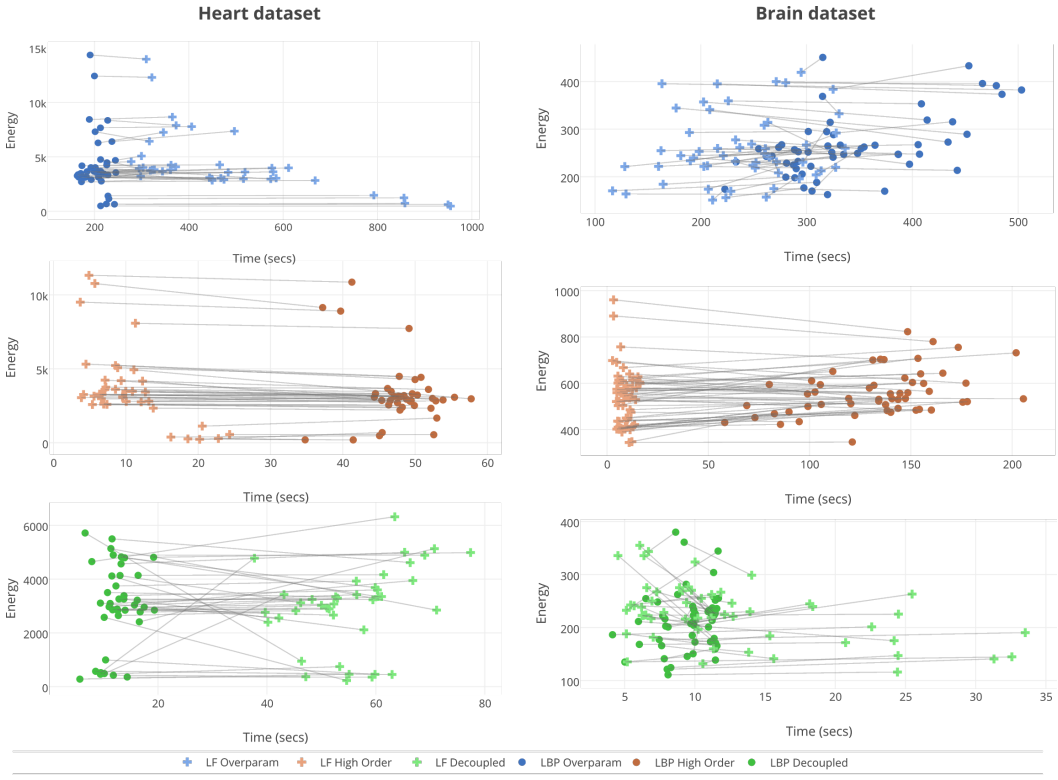


Figure 3.17: Comparison between total optimization time and final energy using two different optimizers (LBP corresponds to circles and LF to crosses). Results are shown for the overparameterized approach (in blue), the high-order approach (in orange) and the decoupled approach (in green). The gray lines connect data points corresponding to the same registration case. As it can be observed with respect to the final energy, both methods are equivalent in general (without considering the outliers). However, there are more important differences in terms of computational time. In the high-order approach, where the label space is small, LF outperforms LBP since convergence is achieved in a few seconds, independently of the dataset. For bigger label spaces (like decoupled and overparameterized approaches), LBP converges faster in case of the heart dataset, where SAD is used as similarity measure and therefore the energy is smooth. The last case is when we use MI as similarity measure (brain dataset) and we have big label spaces: there is no clear pattern in this case. Note that we do not compare with the rigid estimation using Simplex methods, since it estimates only 6 parameters (in a few seconds) while the discrete methods deal with hundreds.

making smart decisions in terms of energy definition and graph topology.

However, there is always a trade-off. This strong reduction in the size of the label space, has an effect on other parameters like number of cliques and number of variables. In case of the decoupled model, the main advantage is related to the fact

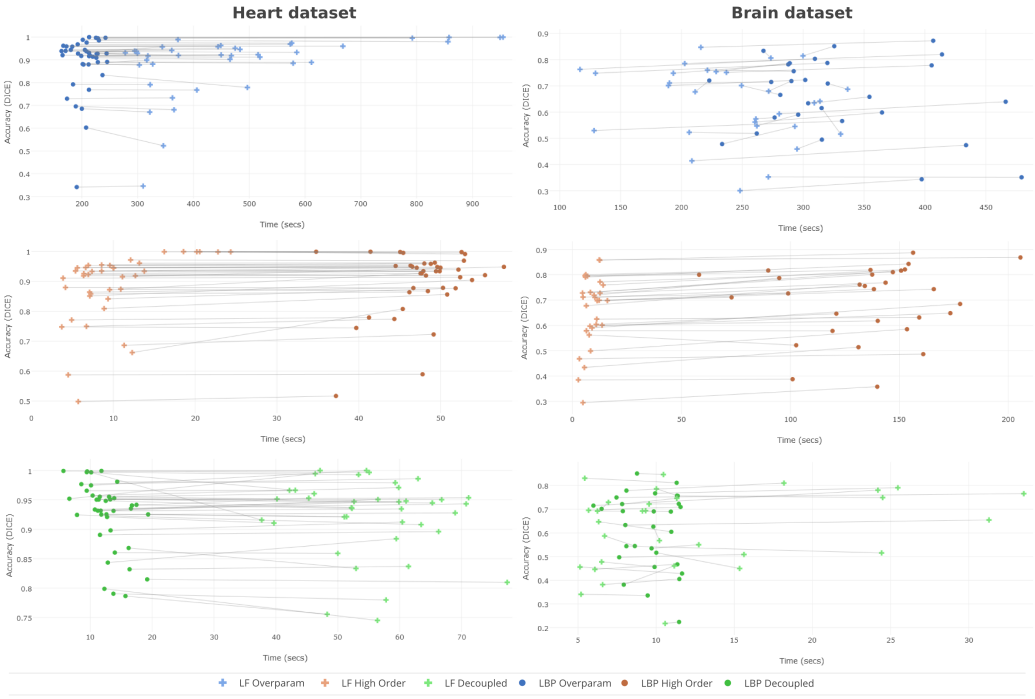


Figure 3.18: Comparison between total optimization time and results accuracy (measured using DICE coefficient) using two different optimizers (LBP corresponds to circles and LF to crosses). Results are shown for the overparameterized approach (in blue), the high-order approach (in orange) and the decoupled approach (in green). The gray lines connect data points corresponding to the same registration case. Note that these results are consistent with those shown in figure 3.17. Indeed, one can observe that graphs in figure 3.17 are essentially a flipped version (over the X axis) of graphs included in this figure. This evidences a high correlation between low energy values and high accuracy of the results, proving that the energy is appropriately modeled.

that while the number of variables and edges augment linearly (it goes from N to $2N$ in case of variables, and from $2N-2\sqrt{N}$ to $5N-4\sqrt{N}$ in case of pairwise edges), the number of labels decreases quadratically (from $|L_D \times L_P|$ to $\max(|L_D|, |L_P|)$). It results in better performance for the decoupled method as can be observed in figure 3.16. A consequence of the third-order cliques in the high-order method is higher computation costs. Even then, judging from the running times reported in figure 3.16, we achieve good experimental computation time because of the smaller label space.

Finally, we include a comparison in terms of memory footprints (see table 3.2) among the three methods, using two different optimizers. We reported the maximum amount of memory that a process consumed while running one registration case for the heart dataset. As expected, the overparemterized model requires more

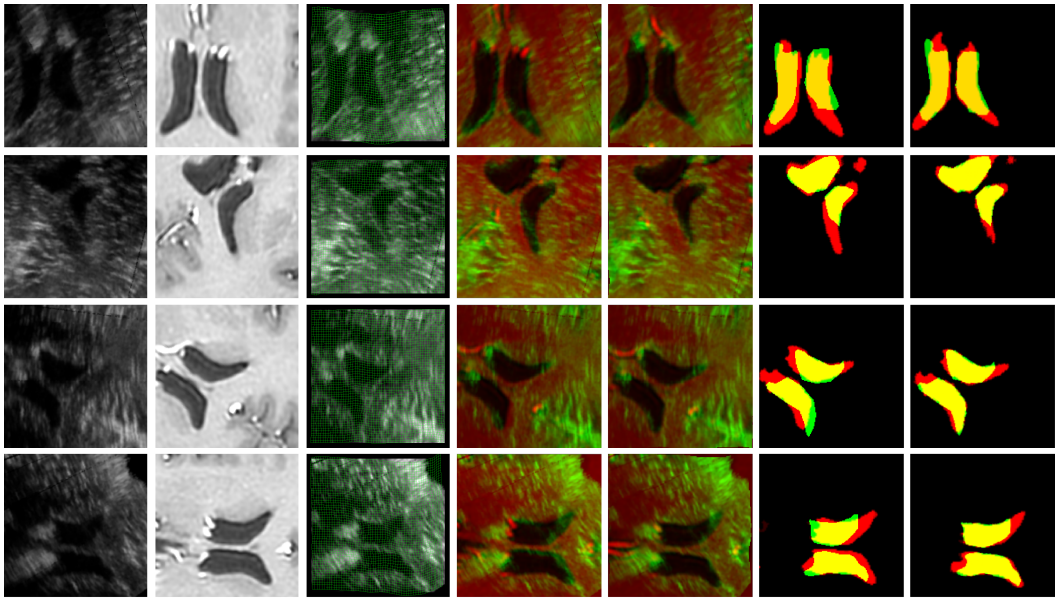


Figure 3.19: Results for one slice from four of the six brain sequences (each row correspond to a different sequence). (a) Source 2D US image. (b) Slice extracted from the MRI corresponding to the initial position of the plane. (c) Deformed source image overlapped with the estimated deformation field. (d) Blending between initial images (US and corresponding MRI slice). (e) Blending between final images (Deformed US image and estimated MRI slice). (f) Overlapping between initial segmentations. (g) Overlapping between segmentations after registration.

Parameter	**Overparameterized**	**Decoupled**	**High-order**
Label space	$\|L_D \times L_P\|$	$\max(\|L_D\|, \|L_P\|)$	$\|L_D\|$
# variables	N	$2N$	N
# 1st order cliques	N	-	-
# 2nd order cliques	$2N - 2\sqrt{N}$	$5N - 4\sqrt{N}$	-
# 3rd order cliques	-	-	$4N - 6\sqrt{N} + 2$

Table 3.1: Comparison among the three methods in terms of label space and graph topology.

memory than the other two approaches. Results also suggest that LF is more efficient in terms of memory consumption than LBP, since given the same graphical model, LF always outperforms LBP in terms of memory consumption.

	Overparameterized	Decoupled	High-order
LBP	362 MB	37.9 MB	65.4 MB
LF	342 MB	34.9 MB	44.7 MB

Table 3.2: Memory footprint comparison among the three methods, using two different optimizers. The reported value corresponds to the maximum amount of memory that the process used while running in every case.

3.4 Conclusions and Further Ideas

We derive three new models from the standard graph-based deformable registration theory for slice-to-volume registration. We show promising results in a monomodal and a multimodal case, using different inference methods, and we compare them with a baseline rigid approach were inference is performed using continuous optimization. The proposed framework inherits the advantages of graph-based registration theory: modularity with respect to the similarity measure, flexibility to incorporate new types of prior knowledge within the registration process (through new energy terms) and scalability given by its parallelization potential.

The three methods we present aim at optimizing different types of energy functions in order to get both, rigid and deformable transformations that can be applied independently, according to the problem we are trying to solve. An extensive evaluation in terms of different statistical indicators is presented, together with a comparative analysis of the algorithmic and computational complexity of each model. This work constitutes a clear example of the modeling power of graphical models, and it pushes the limits of the state-of-the-art by showing how a new problem can be solved not just in one, but in three different ways.

Numerous future developments built upon the proposed framework can be imagined. Alternative optimization methods and in particular second order methods in the context of higher order inference could improve the quality of the obtained solution while decreasing the computational complexity. The integration of geometric information (landmark correspondences) combined with iconic similarity measures [Sotiras 2010b] could also be an interesting additional component of the registration criterion. Moreover, sensor based technologies (like the ones described in section 3.1.2.3) could be combined with our iconic method to improve its stability when image occlusions or abrupt changes occur. Domain/problem specific parameter learning [Baudin 2013,Komodakis 2015] towards improving the proposed models could have a positive influence on the obtained results. The learning framework that will be presented in section 4.3 follows this direction.

The applications explored in this chapter are mainly related to pre and intra-operative image fusion. However, as it was mentioned in section 3.1.6, other scenarios related to motion correction and volume reconstruction require slice-to-volume registration methods. It would be interesting to explore how our approaches perform in these cases.

The three models we proposed aim at finding the optimal deformation field

that relates a 2D slice to a 3D volume. From this deformation field, we extract rigid and deformable parameters as explained in section 3.2. However, it was mentioned that in some specific problems, rigid models are enough since deformations are not considered (see section 3.1.3.1). If we would like to apply our method to these cases, unnecessary efforts would be done to infer the deformation field and then extract the rigid parameters. Therefore, rigid (only) slice-to-volume registration using discrete models constitutes an open research direction. In this regard, the work of [Zikic 2010b] about linear image registration using MRF and discrete methods could be extended to the case of rigid slice-to-volume registration.

Chapter 4

Context Specific Registration Methods

In this chapter we explore how semantic information can be used to improve the registration process. Existing methods for medical image registration rely on a single metric function explaining the similarity between the source and target images. We argue that incorporating semantic information to guide the registration process, will further improve the accuracy of the results. To this end, we consider two different scenarios. In the first scenario, we are given a set of unlabeled input images and pre-trained classifiers. These classifiers can estimate semantic probability maps indicating how likely every voxel corresponds to a given anatomical structure. This information is incorporated within a co-registration and segmentation algorithm that jointly maps all the input volumes to a common reference space while generating a semantic segmentation mask per image. In the second scenario, we consider the case where hard semantic segmentation masks are available for a training dataset. We use them towards learning context-specific aggregations of standard similarity measures. The learned matching criterion is integrated to the standard graph based image registration formulation, resulting in a multi-metric algorithm endowed with a spatially varying similarity metric function conditioned on the anatomical structures.

Contents

4.1 Literature Review

Medical images are n-dimensional vectors of intensity values. Context specific/semantic information can augment the expressive power of such images. This information may be encoded in different ways such as data tags (e.g age, gender, weight, disease, etc), anatomical or functional annotations, landmarks, etc. In this chapter, we discuss how such context specific or semantic information can be exploited to improve registration.

Various approaches were envisioned to incorporate priors into the registration process. Without being exhaustive, let us present the most commonly adopted approaches towards incorporating context information to the registration process: (i) knowledge-based transformations, where the information is encoded within the deformation model (ii) modality-specific similarity measures, adopted when we know a priori the type of images we want to register (iii) metric learning, where we aim at learning more accurate matching criteria from examples and (iv) segmentation-aware strategies, which directly incorporate segmentation priors to the registration process.

Knowledge-based Transformations

Prior knowledge as it concerns the elastic mapping within deformable registration can be achieved either by learning a high dimensional statistical model of deformations or by exploiting biomechanical/biophysical properties of tissues [Sotiras 2013]. In the first case, statistical deformation models are used to capture information about deformation across a population of subjects. Principal Component Analysis (PCA) was used in several studies (e.g [Wang 2000, Tang 2009]) to represent such statistical models. Then, registration can be performed estimating the linear combination of the PCA modes of variation that optimize a given criterion. In [Wouters 2006] the PCA parameters were estimated by minimizing the mutual information similarity criterion. The advantage of such models is that they constrain the set of admissible deformations to be a linear combination of a basis learned from examples. Such methods are manifold constrained since only "observable" mappings can be reproduced. Manifold enhanced methods is an alternative that seek to minimize the distance of the mapping from the space of admissible solutions. One can refer to the work presented in [Glocker 2009b] where the deformation was expressed through a graph-based control-driven method and relative constraints on the deformation of control points were learned. Such constraints were used in the form of costs during the inference of the deformations in new data.

The second alternative, exploiting the same principle, uses biomechanical or biophysical inspired models. Building such models is a challenging task since not only the material properties but also the geometry and boundary conditions of the organs must be considered. Tumor growth models, for example, have been successfully used to guide the estimation of the deformation field [Clatz 2005] through such biphysical constraints. Breast simulation based on Finite Elements Model

(FEM) was used by [del Palomar 2008] to accurately predict real deformations of the breast. Other organs like the prostate have also been considered. In [Mohamed 2002] a biomechanical model was used to produce admissible simulations of the deformations generated by the TRUS probe insertion. This information is then considered when performing registration between pre-operative and intra-operative prostate images.

The advantage of such approaches relates to the DOF used to represent the transformation. The aforementioned models decrease the number of transformation parameters when compared to the standard dense deformation field. Thus, exploring the search space becomes simpler, leading to improved efficiency and lower computational complexity at the expense of decreasing the expressive power of the mapping through constraints on the admissible transformations.

Modality-Specific Similarity Measure

The use of context-specific metric functions is an alternative constraint to the deformation priors. Clinical settings/applications usually involve known image modalities to be registered. Therefore, an effort was invested to design modality specific similarity measures adapted to them. The simplest approach consists in simulating the target modality from the source (through either a physical simulation system or a machine learning approach), reducing the problem to a monomodal scenario. Simulating one modality from another can be achieved considering the physical properties of the imaging device [Sotiras 2013]. In [Wein 2007, Wein 2008a] a novel method to generate simulated US images from CT data was proposed. They introduce a new similarity measure entitled Linear Correlation of Linear Combination (LC^2), which is invariant to missing simulation details, yielding smooth properties and a global maximum at the correct alignment. The same LC^2 was used in [Fuerst 2014] for registering US to MRI images. The similarity measure was defined by locally matching US intensities to both MRI intensity and gradient magnitude. These techniques exploit knowledge relative to specific types of image modalities. Therefore, they achieve accurate results when the underlying physical principles are known and allow the simulation of new modalities. However, such methods fail when generalizing to arbitrary image technologies.

Distance Metric Learning

Metric learning is an alternative way to incorporate context specific and prior information to the similarity measure. It seeks to determine a mapping that will bring samples from similar classes close (corresponding anatomical structures in the different modalities) and far apart the data from different classes [Boukouvalas 2011]. This concept can be applied to different fields using distance metrics, such as classification, clustering and retrieval tasks. A complete survey about distance metric learning was presented in [Yang 2006]. From this general definition, the meaning of metric learning in the context of image registration can be simply interpreted as

learning a domain specific matching criterion that allows the comparison of any two given image modalities. Hence, instead of using single universal similarity measures such as MI, SAD or NCC, we learn specific matching criteria optimally adapted for a given task.

In [Yang 2006] methods were classified as supervised or unsupervised. Supervised methods require a set of annotated data (i.e. some sort of similarity information, either in the form of a full scalar distance between samples, a ranking or some proximity of some samples to others or just a separation between samples that are deemed similar and dissimilar [Michel 2014]). Metric learning consists in determining a mapping that will bring all data points within the same classes close, while separating all the data points from different classes. In unsupervised methods (referred as manifold learning approaches in [Yang 2006]) the main idea is to learn an underlying low-dimensional manifold where geometric relationships (e.g. distance) between most of the observed data are preserved. Then, we can map the original data into this simpler representation endowed with a metric that correctly represents the data similarity [Bronstein 2010].

Several image registration methods using supervised metric learning can be found in the recent literature. [Lee 2009] derived a rigid multi-modality registration algorithm where the similarity measure is learned in a discriminative manner, such that the target and correctly deformed source image receive high similarity scores. The learning was performed using a max-margin approach known as structured prediction. The structured prediction approach of [Lee 2009] uses pairs of pre-aligned images as training data. They assume that the similarity of two images decomposes into the similarities of local regions (patches). Therefore, the learning is performed at the patch level. In a different but still supervised setting, [Bronstein 2010] proposed the use of sensitive hashing to learn a multi-modality distance metric, that can be applied to data coming from two different spaces. The idea is to embed the input data from two arbitrary spaces into the binary space where Hamming distance is minimum for corresponding patches and high for the non corresponding ones. This mapping is expressed as a binary classification problem and can be efficiently learned using boosting algorithms. The authors applied this method to perform multimodal non-rigid registration. The training set involved pairs of perfectly aligned images and consisted on a collection of positive and negative pairs of patches. Such a metric learning approach was plugged into the standard graph-based deformable registration framework in [Michel 2011]. In [Toga 2008], instead of requiring pre-aligned images, the training set consisted of non-aligned images with manually annotated patch pairs (landmarks). Coarse-to-fine strategies were adopted where a global similarity measure was learned first, followed by a fine similarity measure that can capture correspondences at the higher resolution image level. Learning was performed by selecting and fusing a large number of image features. More recently, approaches based on deep learning have started to gain popularity. An interesting discussion about this topic was presented in [Long 2014], suggesting that features learned using convnets are at least as useful (and sometimes considerably more useful) than conventional ones when performing alignment

of elements of the same class.

Unsupervised learning methods for image registration have also been studied. [Wachinger 2010] proposed to apply manifold learning (through Laplacian eigenmaps) to learn structural representations of multimodal images. The idea was to calculate dense descriptors that represent the structural information of image patches which do not depend on the intensity values of the images but on the structures in the patch. Once the original images were projected to the learned manifold, standard L1 or L2 norm were used for the registration of the structural images. In [Ou 2011] the notion of mutual saliency was considered to locally adapt the metric function involving a weighted sum over a large space of features. A different approach based on unsupervised deep learning was proposed by [Wu 2013]. The composition of a basis of filters was learned to effectively represent the observed image patches. These features were then incorporated in state-of-the-art registration methods and tested on MR images of the brain.

Segmentation-Aware Strategies

Segmentation labels can be an interesting alternative to encode prior information. Given an image I (2D or 3D) and a finite set of semantic classes $\mathcal{C}$, a hard segmentation mask assigns a label $c \in \mathcal{C}$ to every pixel (voxel) of I. Soft segmentation masks (or probability maps) indicate how likely a given pixel (voxel) is to be associated with a class $c \in \mathcal{C}$. We will call segmentation-aware strategies to the methods that incorporate -at some extent- these segmentation maps (hard or soft) directly to the registration process.

A popular way to incorporate segmentation information into the registration process is by performing both tasks at the same time. The idea is to exploit the inter-dependencies between the two problems to improve both, registration and segmentation results. This is known as co-registration and segmentation. A unified segmentation framework integrating tissue classification and image registration within the same generative model was proposed in [Ashburner 2005]. Their model is based on a mixture of Gaussians (MOG) incorporating smooth intensity variation, non-linear registration and soft segmentations. Another example is given in [Mahapatra 2012], where segmentation information was integrated to improve MRF-based elastic image registration. They formulated the data term of the MRF energy as a combination of image intensity and mutual dependence of registration and segmentation information. Hence, the labels were defined as the joint occurrence of displacement fields and segmentation class probability. The main disadvantage of this approach is the manual intervention required to define a mask around the object of interest. The area within this mask was considered to get the intensity distributions of the object, which was then used to determine the mutual dependence of the segmentation mask and the displacement vectors. [Wyatt 2003] also showed that one can obtain more accurate and robust results by seeking a joint solution to registration and segmentation. They used a MRF-based approach as well, inducing spatial regularization into the registration process through a class-based entropy

measure. Gaussian Mixture Models were coupled with an MRF prior to perform the segmentation task, while the registration relied on the segmentation by minimizing the joint class histogram between both images. In [Parisot 2014], an elegant formulation seeking to take advantage of concurrent segmentation and registration of brain tumor images was proposed. Such a framework used a data-driven adaptive graphical model endowed with uncertainties and was used to adapt the registration task to the presence of the tumor increasing also the segmentation quality. One can also refer to early works like [Yezzi 2003], where a geometric, variational framework that uses active contours to simultaneously segment and register features from two or more images was proposed. They minimize a set of energy functionals, one for each image. Similar work in the context of computer vision and motion analysis and tracking was presented in [Paragios 2005].

Another alternative considering segmentation masks within the registration process is the atlas-based method. In the context of medical image analysis, an atlas is usually defined as the combination of an intensity image (template) and its corresponding semantic annotations (the atlas labels). These annotations can be expressed in the form of discrete segmentation labels or probability maps (probabilistic atlas). The image segmentation problem can be solved using either a single [Cabezas 2011] or multiple atlases [Iglesias 2015]. The idea is to register the atlas(es) to the target image and, then, transfer the labels using the obtained deformation field. In the multi-atlas setting, a label fusion strategy must be considered to generate the final (consensus) segmentation mask. Several works have been presented where simultaneous registration and segmentation are performed, leading to improvements in both tasks by exploiting the inter-dependencies between the two problems. [Gass 2014] introduced a novel MRF-based approach for segmenting medical images while simultaneously registering an atlas non-rigidly. Even if they modeled both type of variables (deformation vectors and segmentation classes) they did it in a sum-label way, i.e. the solution label space is the union of segmentation and registration label spaces and therefore the total number of labels is the sum of the number of registration and segmentation labels (instead of the product as in [Mahapatra 2012]). This strategy reduces the complexity of the label space by using a two-layer model where variables in every layer take values from different spaces. The solution of this pairwise model was efficiently found using α-expansion [Boykov 2001]. A variational formulation that simultaneously register an atlas shape to a novel brain image and segments it at the same time was presented in [Wang 2005]. They solved a coupled set of nonlinear PDEs using efficient numerical schemes. The main disadvantage of this approach is that it is limited to binary segmentation problems.

Methods mentioned in the previous paragraph make use of a single atlas. However, multi-atlas approaches where prior information is incorporated to the registration process have also been considered. In [Heckemann 2010], a hierarchical multi-atlas approach was proposed. It introduced tissue-class information to improve image registration and, consequently, segmentation results. The modification used only tissue classification information (learned offline) to guide the three first

hierarchy registration levels (using cross-correlation matching criterion to align the tissue probability maps). The resulting transformation was then used as a starting point for detailed registration using the gray intensity values and NMI as the similarity measure. Another multi-atlas framework that incorporates soft segmentation priors was proposed in [Alchatzidis 2014a]. The motivation was based on the idea that prior segmentation probabilities can help registration to avoid local minima, specially in areas of high anatomical variability like brain cortex. They used a MRF-based formulation where each atlas is associated with a deformation field (registration variables), while the target image is associated with a segmentation map (segmentation variables). Prior per voxel probabilities, produced through learning of local features, are taken into account in the energy formulation. Intensity values, smoothness constraints and a coupling term imposing agreement between the segmentation label and the segmentation hypotheses produced from the deformed atlases are considered as well in the energy formulation.

4.2 Prior-based Co-Registration and Segmentation

In this section we propose a modular and scalable framework for dense co-registration and co-segmentation that incorporates (soft) prior information. The central idea of this segmentation-aware method is to use estimated maps of semantic labels (provided by an arbitrary classifier) as a surrogate for unlabelled data, and combine them with (population) deformable registration to improve both alignment and segmentation. The method seeks to deform all volumes towards an arbitrary reference space where the transformed images exhibit consensus with respect to image similarities (according to some arbitrary metric) and semantic labels. We employ an approach sharing conceptual similarities with α-expansion, combined with efficient linear programming.

4.2.1 Introduction

In recent years, multi-atlas segmentation methods have become one of the most widely used segmentation techniques for biomedical applications. Multi-atlas based segmentation principle employs an annotated dataset of atlases (images with their corresponding ground truth segmentation), in order to segment an unseen image (target). The atlases are registered to the target image and the resulting deformed candidate segmentations are fused, generating the final segmentation mask. Normally, these approaches allow to segment one target image at a time. In this work, we propose to boost the spatial transformation estimation process by integrating learned soft priors on the structures of interest. To this end, we propose a population deformable registration framework [Sotiras 2009] that, given a set of input volumes, creates an optimal (unknown) reference space through their mutual deformation. We achieve this through a discrete groupwise co-registration process endowed with segmentation consistency constraints which makes possible to segment not one, but multiple images at the same time. This gives rise to a ground truth-free, data-driven registration process that leads to more accurate results than standard multi-volume segmentation strategies.

Multi-atlas segmentation approaches involve two stages: registration and label fusion. Different non-linear approaches have been proposed in the literature with regards to registration [Sotiras 2013]. Methods differ in terms of the deformation model (physical models, interpolation theory or based on specific knowledge of the domain), matching criterion (geometric, iconic or hybrid) and optimization technique (continuous or discrete).

Label fusion is a post processing step, that merges the candidate segmentations obtained after wrapping the atlases to the target image. Such strategies may be simple, like for example majority voting, or more complex involving probabilistic frameworks [Rohlfing 2004, Sabuncu 2010]. An extensive review of multi-atlas seg-

mentation methods can be found in [Iglesias 2015]. A comprehensive evaluation of the accuracy of the atlas-based brain structure segmentation with several different publicly available registration methods was provided in [Klein 2009].

Our method bears some similaries with a standard multi-atlas segmentation algorithm, but targets a different objective. The segmentation results are a by-product of the population registration where context labels are used to improve the registration performance. Moreover, we aim at segmenting all input volumes simultaneously, while standard approaches do it independently. In the literature, concept similarities with the proposed framework can be found in [Heckemann 2010] and [Alchatzidis 2014a]. These methods use probabilistic priors obtained through a pre-trained classifier to improve segmentation and registration results. Here, opposite to [Heckemann 2010], we do not perform independent prior and intensity based registration steps, but we do it using a compound matching criterion that considers both at the same time. Furthermore, in contrast to the work of [Alchatzidis 2014a] where segmentation variables are explicitly modeled, here we only model registration variables, reducing the number of parameters to be estimated and making the approach tractable and scalable. Moreover, both methods aim at segmenting a single target image, whereas ours considers a population of images simultaneously.

Given a set of unseen images, our method infers the atlases on-the-fly by using previously learned classifiers, and incorporates this information to the energy formulation. We evaluate the effectiveness of our approach on the task of segmenting challenging sub-cortical structures in two brain image datasets. We compare the performance of the proposed method with the standard pairwise atlas-based segmentation, and we evalute it using accurate as well as weak priors. Comparison with ground truth driven co-registration and segmentation is also included for the sake of completeness.

4.2.2 Prior-based Co-Registration and Segmentation

The co-registration and segmentation algorithm is formulated as an energy minimization problem. The input is the set of 3D images $D_I = \{I_1, I_2, ..., I_N\}$ (with $I_i : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$) and their corresponding segmentation priors (soft or hard) $D_S = \{S_1, S_2, ..., S_N\}$ which are associated to the possible segmentation classes $c \in \mathcal{C} = \{0, ..., C-1\}$ as $S_i : \Omega \times \mathcal{C} \rightarrow \{0, 1\}$. The desired output is a set of final multi-label segmentation masks $\hat{D}_S = \left\{\hat{S}_1, \hat{S}_2, ..., \hat{S}_N\right\}$ together with the deformation fields $\hat{\mathcal{T}} = \left\{\hat{T}_1, \hat{T}_2, ..., \hat{T}_N\right\}$ which warp every image to a common coordinate space. We optimize an objective function E based on a compound matching criterion where both image similarities and learned priors are considered as well as deformation smoothness for the deformation fields. Let us now, without loss of generality, proceed with a concept definition of our approach. The objective function that could be used to solve such a problem involves three terms:

$$E(\mathcal{T}; D_I, D_S, \delta, \eta, \mathcal{R}) = E_I\left(\mathcal{T}; D_I, \delta\right) + E_S\left(\mathcal{T}; D_S, \eta\right) + E_R\left(\mathcal{T}; \mathcal{R}\right). \tag{4.1}$$

The first term, $E_I(\mathcal{T}; D_I, \delta)$, imposes compactness on appearance of the registered population at the pixel level. It depends on a similarity measure δ and the image intensity of every volume in the input set D_I. It is defined as:

$$E_I(\mathcal{T}; D_I, \delta) = \int_\Omega \delta(I_1 \circ T_1(x), I_2 \circ T_2(x), \ldots, I_N \circ T_N(x))dx, \tag{4.2}$$

where δ is a function measuring the coherence between equivalent pixels (or patches centered at given positions, depending on the formulation) in every deformed image. It is nothing more than the notion of similarity measure, which is extended so that it accounts for more than one image at the same time. Such a term can be either higher order (explicit modeling of global similarity between the deformed volumes), or lower order (decomposition of the global term into a sum of local terms).

The second term, $E_S(\mathcal{T}; D_S, \eta)$, measures the degree of agreement among the deformed priors. It depends on the segmentation masks D_S and it uses a function η that quantifies this agreement. Hamming distance of the thresholded probability maps or more complex functions could be used in this case. The term is defined as follows:

$$E_S(\mathcal{T}; D_S, \eta) = \sum_{c \in \mathcal{C}} \int_\Omega \eta(S_1 \circ T_1(x, c), S_2 \circ T_2(x, c), \ldots, S_N \circ T_N(x, c))dx. \tag{4.3}$$

Note that we sum over all possible semantic labels $c \in \mathcal{C}$. If we have priors which are more reliable than others, we could weight them according to the confidence level.

Last but not least, we define a regularization term E_R that introduces geometric or anatomical constraints to the deformation fields. In medical image registration, we usually assume that the deformation field is smooth. Different type of regularizers $\mathcal{R}$ could be used, which are commonly defined as a convex function acting on the gradient of the deformation field. We apply this constraint independently to every deformation field T_i as follows:

$$E_R(\mathcal{T}; \mathcal{R}) = \sum_{i=1}^{N} \int_\Omega \mathcal{R}(T_i(x))dx. \tag{4.4}$$

The estimation of the desired transformations can be formulated as a MAP inference problem where we aim at minimizing the following:

$$\hat{\mathcal{T}} = \arg\min_{\mathcal{T}} E(\mathcal{T}; D_I, D_S, \delta, \eta, \mathcal{R}). \tag{4.5}$$

Once the optimal deformation fields $\hat{\mathcal{T}}$ are estimated, we can reconstruct the final segmentation masks $\hat{D}_S$. To this end, we first propagate every original segmentation prior from D_S to the common space, generating warped segmentation masks $S_i \circ \hat{T}_i$. Then, say we want to estimate the final segmentation mask for volume k, then we back propagate all the warped segmentation masks $S_i \circ \hat{T}_i$ from the common space, to the space of image k by using the inverse of the deformation

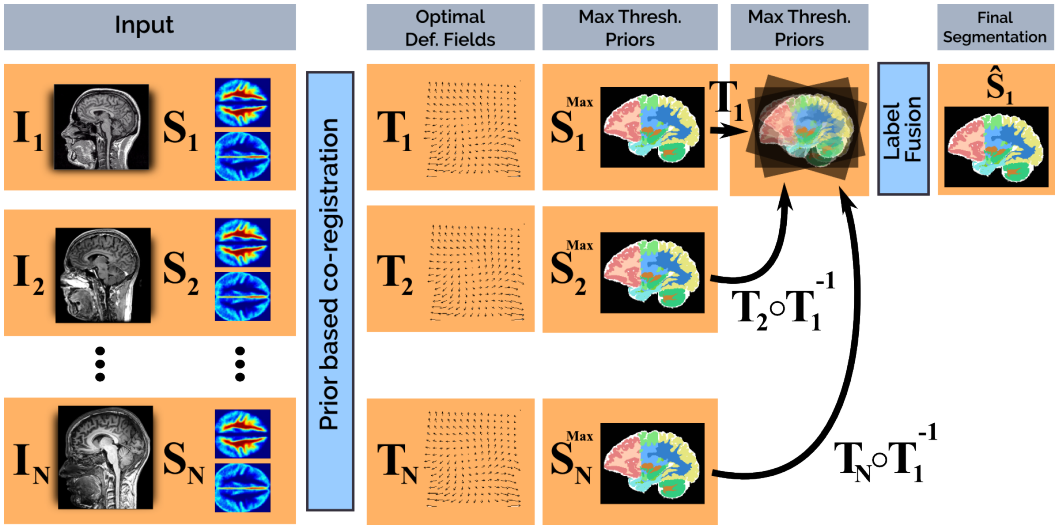


Figure 4.1: Co-registration and segmentation workflow. The input consists of images (I_i) and their corresponding segmentation likelihoods S_i. After running the prior co-registration algorithm (which considers the priors in the energy formulation through the semantic consistency term), the optimal deformation fields are used to deform the *maximum-a-posteriori* of the segmentation likelihoods, which are considered as segmentation hypothesis. A label fusion strategy is then applied to generate the final segmentation mask. The example in this figure illustrates the process to generate the final segmentation for image I_1. The same strategy is followed to generate the rest of the segmentation.

field. This results in the following equation with respect to the segmentation of the k example:

$$\hat{S}_k = fuse(S_1 \circ \hat{T}_1 \circ \hat{T}_k^{-1}, S_2 \circ \hat{T}_2 \circ \hat{T}_k^{-1}, \ldots, S_k, \ldots, S_N \circ \hat{T}_N \circ \hat{T}_k^{-1}). \tag{4.6}$$

In an abuse of notation, $S \circ T$ indicates that the deformation field T is applied to the initial domain with images intensities being replaced with semantic classes $c \in \mathcal{C}$.

The method is modular with respect to the fusion strategy applied to the back-propagated priors to compute the final segmentation mask. We used a simple majority voting, assigning to every voxel the class $c \in \{0, ..., C-1\}$ with the highest number of votes after back-propagation; although simple, this scheme proved to be fairly effective. Figure 4.1 illustrates the complete process. Let us now proceed with a formal definition of the approach in the context of graphical models.

Graphical models and discrete optimization techniques can address quite effectively this type of vision problems. However, note that the energy to be minimized in the MAP problem (defined in equation 4.1) involves high-order dependencies in the data terms E_I and E_S since all the deformation fields must be considered at the same time. A proper derivation of a discrete MRF using this formulation will

produce high-order terms which are difficult to optimize and significantly increase the computational complexity of the method. Alternative models can be derived for this MAP problem, by doing a smarter choice in terms of graph topology, discrete energy formulation and label space. We propose to follow the strategy adopted by [Sotiras 2009], that involved an approximation of the global terms. It consists of assuming that, when estimating a given deformation field T_k, the rest of the images do not move within the current iteration. This is somehow analogous to the expansion moves of the well known α-expansion [Boykov 2001], in the sense that we follow the philosophy of the 'very large neighborhood search' techniques [Ahuja 2002]: we start with a feasible solution (in our case, the identity deformation fields) and iteratively try to obtain a better solution that improves the solution of a volumes towards all "static" volumes.

4.2.2.1 Iterative Prior-based Co-Registration and Segmentation

The co-registration and segmentation algorithm presented in the previous section is reformulated as an iterative process that solves several energy minimization problems defined on independent Markov Random Fields. The step-by-step procedure is described in Algorithm 2. The energy terms E_I, E_S and E_R (intensity matching, semantic prior consistency and regularization terms, respectively) are replaced by $\mathcal{E}_I$, $\mathcal{E}_S$ and $\mathcal{E}_R$ which consider a single transformation T_k at a time (at a given iteration), instead of considering the complete set $\mathcal{T}$. The co-registration process is performed iteratively: only one source image is chosen at each step and registered considering all other images in the set (simultaneously) as targets. To this end, we optimize an objective function $\mathcal{E}$ based on a compound criterion where image similarities ($\mathcal{E}_I$), learned priors ($\mathcal{E}_S$) and deformation field regularization ($\mathcal{E}_R$) are considered.

The matching component $\mathcal{E}_I$ uses again an iconic similarity measure δ. It could be Sum of Absolute Differences (SAD), Mutual Information (MI), Normalized Cross Correlation (NCC) or any other metric that captures similarity between the two images. It is defined as follows:

$$\mathcal{E}_I\left(T_k; \bar{I}, I_k\right) = \sum_{i=0, i\neq k}^{N} \int_{\Omega} \delta(\widetilde{I}_i, I_k \circ T_k(x))dx. \tag{4.7}$$

Note that we refer to $\widetilde{I}_i$ as the latest deformed version of every original input image I_i. Given C semantic classes $\mathcal{C} = \{0, ..., C-1\}$, the segmentation prior component $\mathcal{E}_S$ takes the following form:

$$\mathcal{E}_{\mathcal{S}}\left(T_k; \bar{S}, S_k\right) = \sum_{c\in\mathcal{C}} \sum_{i=0, i\neq k}^{N} \int_{\Omega} \eta(\widetilde{S}_i, S_k \circ T_k(x, c))dx, \tag{4.8}$$

where η is the same function measuring the degree of agreement among the deformed segmentation masks (used to define E_S in equation 4.3) and $\widetilde{S}_i$ is the latest

Algorithm 2 Iterative Co-registration Segmentation algorithm

procedure ICS($D_I : \{I_1, I_2, ..., I_N\}$, $D_S : \{S_1, S_2, ..., S_N\}$)
  Initialize the deformation fields $\left\{\hat{T}_1, \hat{T}_2, ..., \hat{T}_N\right\}$ as null (identity) deformation fields
  repeat
    repeat
      Randomly select one image from D_I as temporal target image I_k
      Register I_k to the other images in the set $\bar{I} = \{I_i\}_{i\in\{1,...,N\}, i\neq k}$, while optimizing $\mathcal{E}$:

$$\ddot{T}_k = \arg\min_{T_k} \mathcal{E}_I\left(T_k; \bar{I}, I_k\right) + \mathcal{E}_S\left(T_k; \bar{S}, S_k\right) + \mathcal{E}_R\left(T_k\right) \tag{4.10}$$

      Update D_I and D_S as: $I_k \leftarrow I_k \circ T_k$ and $S_k \leftarrow S_k \circ T_k$
      Update deformation field $\hat{T}_k \leftarrow \hat{T}_k \circ \ddot{T}_k$
    until All members of D_I have been selected as target once
  until No more changes in the transformation T_i or the maximum of iterations is reached
  Obtain $\hat{\mathcal{T}} = \left\{\hat{T}_1, \hat{T}_2, ..., \hat{T}_N\right\}$
  for each image $I_k \in D_I$ **do**
    for each segmentation prior $S_i \in D_S$ **do**
      Deform S_i to the native space of I_k:

$$S'_i = S_i \circ \hat{T}_k^{-1} \tag{4.11}$$

    end for
    Apply label fusion (e.g., *Majority Voting*) on $\{S'_i\}_{i\in\{1,...,N\}}$ to obtain $\hat{S}_k$
  end for
  Output: $\hat{\mathcal{T}} = \left\{\hat{T}_1, \hat{T}_2, ..., \hat{T}_N\right\}$ and $\hat{D}_S = \left\{\hat{S}_1, \hat{S}_2, ..., \hat{S}_N\right\}$
end procedure

deformed version of the original input mask S_i. $\bar{S}$ is defined as $\{S_i\}_{i\neq k}$ and S_k is the segmentation prior for the target image.

Finally, the regularization term E_R from equation 4.4 is redefined as $\mathcal{E}_R$ following:

$$\mathcal{E}_R(T_k) = (N-1)\int_\Omega \mathcal{R}\left(T_k(x)\right) dx, \tag{4.9}$$

where we scale the regularization term by $(N-1)$ for normalization purposes.

This iterative process is repeated until convergence, meaning that all images are aligned in a common pose space. The same back-propagation and label fusion strategy defined in equation 4.6 is used to produce the final segmentation masks.

4.2.2.2 Discrete Formulation

The non-rigid registration between each target image I_k and the rest of the images $\bar{I} = \{I_i\}_{i\in\{1,...,N\}, i\neq k}$ is formulated as a discrete energy minimization problem. We consider Free Form Deformation (FFD) [Rueckert 1999] as our discrete transformation model, which parametrizes the deformation fields T_k as a linear combination of K control points. These are regularly placed forming a uniformly distributed grid.

We cast the energy minimization task as a discrete optimization problem of a pairwise discrete MRF. As usual, it is modeled using an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{X} \rangle$

superimposed to the 3D image I_k, where $\mathcal{V}$ and $\mathcal{X}$ denote the nodes and the edges of the graph, respectively. The nodes are interpreted as random variables, while the edges encode the interaction between the variables. The random variables of the MRF model correspond to the displacements of the control points and edges encode dependencies between them. The solution space that can be explored around a control point is given by labels l living in a label space $\mathcal{L}$, where every label $l \in \mathcal{L}$ is associated to a displacement vector $\boldsymbol{d_l}$. Thus, in the context of discrete optimization, the energy of equation 4.10 becomes:

$$\mathcal{E}_{\text{MRF}}\left(\Gamma; G, F\right) = \sum_{p \in \mathcal{V}} g_p\left(l_p\right) + \lambda \sum_{(p,q) \in \mathcal{X}} f_{pq}\left(l_p, l_q\right), \tag{4.12}$$

where l_p, l_q are the labels assigned to the nodes $p, q \in \mathcal{V}$, λ is a scaling factor and Γ is a labeling that assigns a label l_i to every variable in the MRF. The data term $g_p\left(l_p\right) \in G$ is formulated as the combination of the matching and the segmentation prior term:

$$g_p(l_p) = \int_{\Omega_p} \delta(\widetilde{I}_i, I_k \circ T_k^{\mathbf{d_{l_p}}}(x))dx + \beta \sum_{c \in \mathcal{C}} \int_{\Omega_p} \eta(\widetilde{S}_i, S_k \circ T_k^{\mathbf{d_{l_p}}}(x, c))dx, \tag{4.13}$$

where $T_k^{\mathbf{d_{l_p}}}$ is the transformation induced by the movement of the control point p when displacement vector $\mathbf{d_{l_p}}$ (associated to label l_p) is applied. β is a coefficient, which determines the influence of segmentation priors on the optimization problem. The data term is defined for an image patch Ω_p centered at control point p, as it was discussed in section 2.3. Note that for $\beta = 0$, the problem reduces to the standard approach; the only difference in this case is that it considers all images in the input set, instead of considering just one target image like in the pairwise case.

The pairwise potentials are defined as the regularization of the deformation field. A discrete approximation of the gradient of the spatial transformation can be computed by taking the difference between the displacements of the neighbouring nodes:

$$f_{pq}\left(l_p, l_q\right) = \left\| \mathbf{d_{l_p}} - \mathbf{d_{l_q}} \right\|. \tag{4.14}$$

The method is modular with respect to the regularization term. Therefore, different type of regularizers could be considered depending on the problem.

In order to infer the best possible labeling, we employ the Fast-PD [Komodakis 2008] algorithm described in section 2.4.3. It is an efficient discrete optimization method based on linear programming relaxation, that has shown promising results when applied to multi-label problems with similar types of energies.

4.2.3 Experiments and Results

We evaluate the performance of our approach on two medical image datasets by segmenting subcortical structures of the brain. These structures are located below the cerebral cortex and refer to complex shapes difficult to segment. We tested

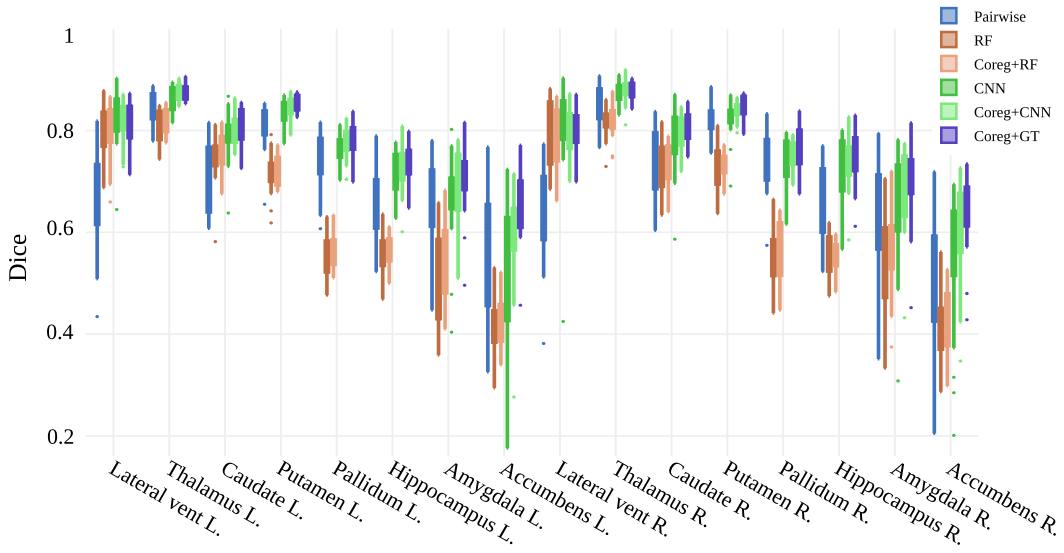


Figure 4.2: Average Dice coefficient of subcortical structures in IBSR dataset.

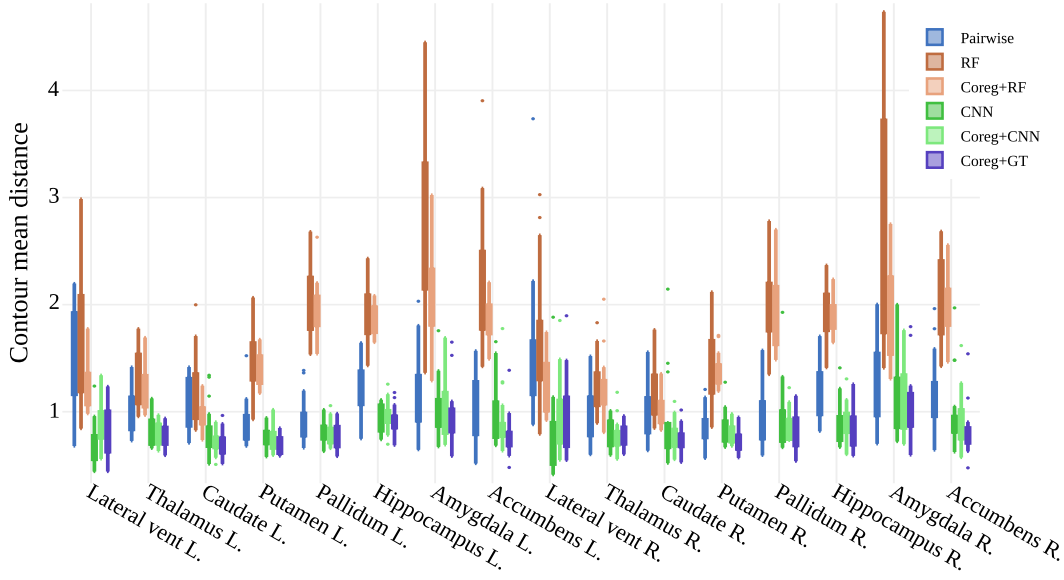


Figure 4.3: Average Hausdorff distance of subcortical structures in IBSR dataset.

using two different types of priors: reliable priors generated using a convolutional neural network (CNN) and weaker segmentations obtained through random forest (RF) classification.

The first dataset is the publicly available Internet Brain Segmentation Repository (IBSR) [Rohlfing 2012]. It includes 18 annotated 3D T1-weighted MR scans with slice thickness of around 1.3 mm. We use the set of 16 subcortical annota-

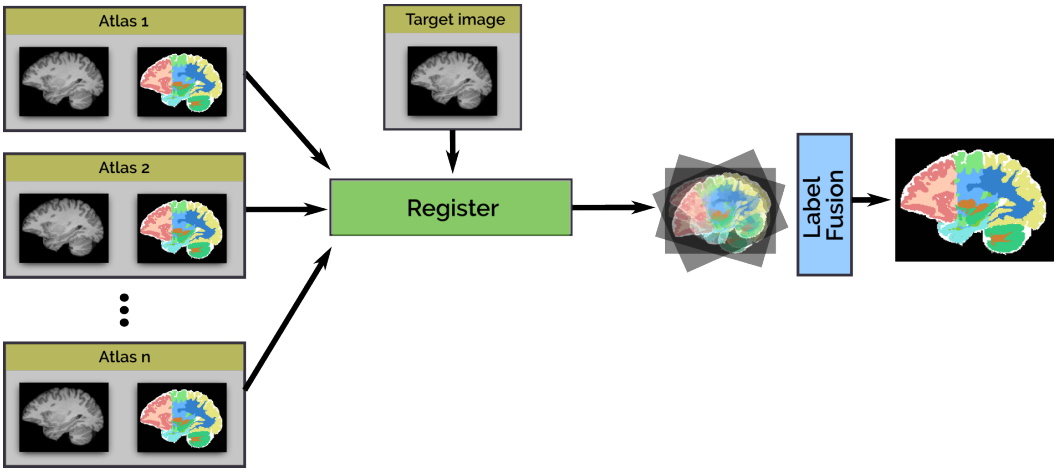


Figure 4.4: Standard pairwise multi-atlas registration approach. Given a set of atlases (image + ground truth segmentation mask) and a target image to segment, the pairwise multi-atlas registration strategy consists in registering (individually) every atlas image to the target. Once all of them are registered, the resulting deformation fields are used to warp the ground truth segmentation masks to the target space. These deformed masks are considered as segmentation hypotheses. A label fusion step is then applied to generate the final segmentation mask considering the hypotheses.

tions, including left and right lateral ventricle, thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens. The second dataset involves patients with Rolandic Epilepsy (RE), and contains 17 children with epilepsy and 18 matched healthy individuals. T1-weighted magnetic resonance images (MRI) scans were performed by a 3 T scanner (Philips Acheiva) with an in-plane resolution of 256×256 and slice thickness of 1 mm. The left and right putamen structures were manually annotated by an expert.

Pairwise Multi-Atlas Baseline. As a baseline, in the first experiment, we implement the standard multi-atlas segmentation based on pairwise registration. In the pairwise approach, all atlases are independently registered to the target image, then the ground truth annotations are deformed to generate the segmentation on the target space (see Figure 4.4). Pairwise registration has been conducted in a discrete fashion using the standard formulation presented in section 2.3, without any prior guidance. Figures 4.2- 4.5 show the values for Dice overlap coefficient, Haussdorff distance (HD) and contour mean distance (CMD) between the output segmentation of the pairwise approach and the ground truth on IBSR dataset. Figure 4.6 illustrates the same measurements for the RE dataset. In both cases our approach using CNN priors (Coreg+CNN) performs better than the pairwise segmentation baseline.

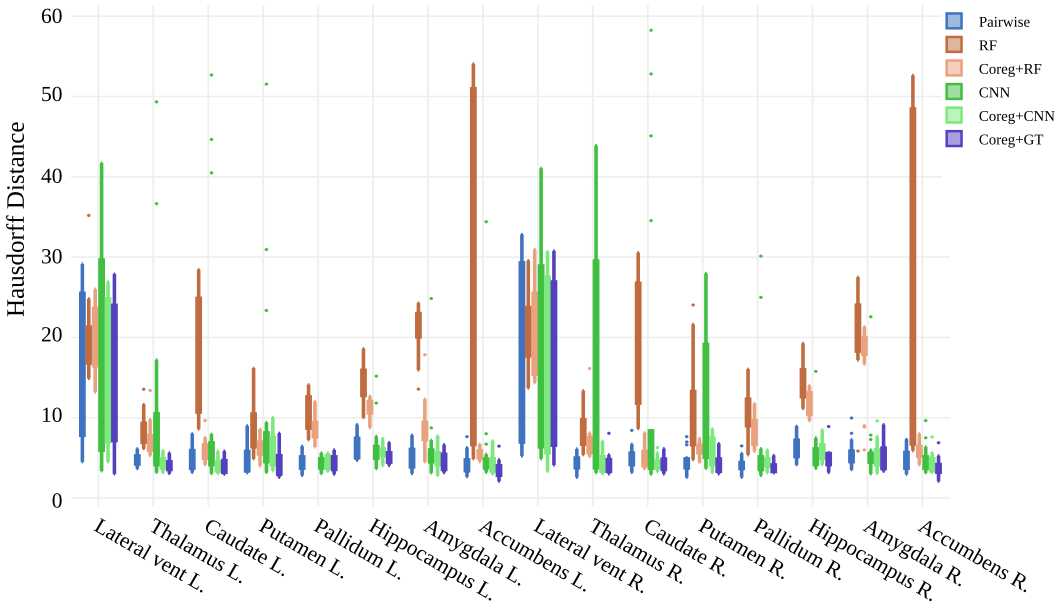


Figure 4.5: Average contour mean distance on sub-cortical structures of IBSR dataset.

It should be noted though that the pairwise baseline has a clear advantage with respect to the proposed approach: the segmentation masks that are deformed and used to generate the final segmentation result are the actual ground truth. Instead, the prior based methods (Coreg+CNN and Coreg+RF) do not use at any point the ground truth masks; they only use the estimated probability maps.

CNN Priors (Coreg+CNN) We train a CNN classifier using the formulation presented in Appendix A.2. To this end, we split the IBSR dataset into three sets. Each time, we use two of the sets as training data and the third set as test data. Similarly, we split RE into two subsets of equal size, using one for training and one for testing, each time. We train on both datasets for 35 epochs starting with a learning rate of 0.01 and dropping it at a logarithmic rate until 0.0001. We train using standard SGD with a momentum of 0.9 and a softmax loss. For all our experiments we use MATLAB and the deep learning library MatConvNet [Vedaldi 2014][1].

RF Priors (Coreg+RF) In a second experiment, we compare the performance of the proposed segmentation method using CNN priors, with an approach based on less accurate priors obtained using Random Forest (RF) classification. The RF-based probability maps are computed according to [Alchatzidis 2014b], and the same co-registration process, as explained in 4.2.2, is applied to produce the final

[1]The source code of the classifier can be downloaded from `https://github.com/tsogkas/brainseg`

segmentation. Note that the quality of these priors is not as good as the one of the CNN priors.

Thresholded Priors (CNN and RF) Figures 4.2- 4.6 show the average Dice coefficient, HD and CMD for the co-registration-segmentation method based on RF priors. To further demonstrate the interest of using the iterative co-registration process on top of CNN/RF priors, we also report the results of the simple segmentation methods based on taking the maximum probability given by CNN and RF without co-registration. In this setting, given a CNN/RF prior, the segmentation class of every voxel is simply chosen as the class with the highest probability. These results show that the CNN priors (without any registration process) are more accurate than the weak RF priors.

Ground Truth Based Oracle (Coreg+GT) So far we have highlighted the positive effect of using priors in our proposed co-registration-segmentation method. As mentioned before, the main idea is to use the probability maps as a surrogate for the ground truth annotations to guide the co-registration process. In order to assess the maximum potential of our approach, we implement an *oracle* variant that can provide us with an upper-bound for the performance of our method. Hence, we employ the ground truth segmentation masks for all 3D volumes, except for the target image, for which we keep the priors computed by the CNN (this could be considered as an almost-ideal case in which our classifier can generate extremely accurate probability maps). We refer to this method as Coreg+GT in the figures.

Unsurprisingly, the results of Coreg+GT after running Algorithm 2 outperform all other combinations (See Figures 4.2- 4.6). Nonetheless, performance of the Coreg+CNN variant is close to Coreg+GT in most cases. This evidence solidifies our original claim, that reliable priors can act as a practical substitute for gold standard annotations for multi-atlas segmentation.

In Figure 4.7 we show some qualitative results, while in 4.2.4 we report the average Dice coefficient, HD and CMD for all structures in the IBSR and RE datasets. The oracle combination Coreg+GT performs best, as expected, with Coreg+CNN a close second, outperforming the other approaches. However, Coreg+CNN has the additional advantage of using learned priors instead of ground truth annotations, which are not always available.

4.2.4 Discussion

In this section, we propose a generic co-registration-segmentation method that can benefit from (soft) segmentation priors and produce results which are close to those generated using ground-truth. Our method is not a multi-atlas segmentation algorithm, since the segmentation results are a by-product of the co-registration where context labels are used to improve the registration performance. We have shown that, by using accurate segmentation priors (like CNN based segmentation maps) we can obtain such results close to the ground truth. More importantly, we also

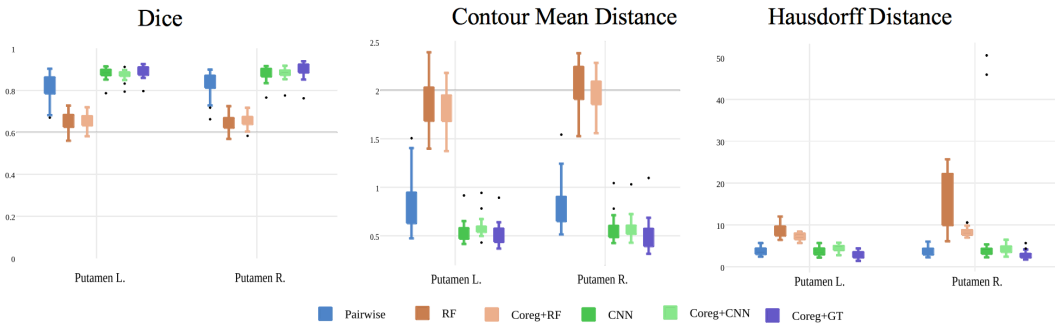


Figure 4.6: Average Dice coefficient, Hausdorff distance, and contour mean distance for the left and right putamen structures in RE dataset.

			Pairwise	RF	Coreg+RF	CNN	Coreg+CNN	Coreg+GT
IBSR	Dice	mean	0.70	0.64	0.65	0.75	0.76	0.78
		std	0.12	0.15	0.14	0.13	0.10	0.09
	CMD	mean	1.10	1.79	1.58	0.87	0.87	0.81
		std	0.37	0.68	0.46	0.26	0.22	0.21
	HD	mean	6.80	16.40	9.80	9.70	6.48	6.09
		std	6.07	10.50	5.73	10.97	5.66	5.88
RE	Dice	mean	0.82	0.65	0.65	0.88	0.88	0.89
		std	0.05	0.04	0.03	0.03	0.02	0.03
	CMD	mean	0.81	1.95	1.88	0.54	0.58	0.49
		std	0.23	0.24	0.19	0.11	0.09	0.12
	HD	mean	3.69	11.71	7.70	4.93	4.18	2.80
		std	0.86	5.58	0.98	7.48	0.83	0.76

Table 4.1: Comparison of the average and standard deviation of the Dice coefficient, HD and CMD on the ISBR and RE datasets. As we can observe, when we use accurate priors (Coreg+CNN), we achieve results close to the ground truth after co-registration and segmentation in the majority of the cases. Moreover, when we use weak priors like (Coreg+RF), we improve the initial results given by the thresholded priors, specially in terms of CMD and HD.

showed that by using less accurate priors (like those learned using RF), we also obtain good results after co-registration-segmentation as shown in Figures 4.3, 4.2, 4.5 and 4.6. There, we observe significant improvements between RF and Coreg+RF, especially in terms of Hausdorff and contour mean distances. This is due to our formulation, which includes a novel energy term that enables any prior to help during registration. Our method may be used in combination with weak priors learned from weak annotations, such as bounding boxes, requiring less time to be produced than accurate segmentation masks, yet delivering good results.

Our method has considerable advantages over standard multi-atlas segmentation. Given a set of target volumes, multi-atlas segmentation would repeatedly register a set of ground-truth masks and perform label fusion individually for each

test volume. Contrary to that, we compute the segmentation probabilities once, and then segment all the volumes "simultaneously". If numerous ground-truth masks were to be used for the registration step, our method benefits from substantial computational gains, as complexity depends only on the number of input volumes we want to segment. In case of large datasets, it removes the burden of selecting an appropriate ground-truth subset to perform multi-atlas segmentation more efficiently; one simply has to compute the probability masks on the input volumes. Another difference is that, in classical multi-atlas, appearance features are used to compute the deformation fields between source and target during the registration process, while no prior on the target image is available. We exploit more sophisticated, learned features to drive the co-registration process. These features are computed for all volumes involved, and are directly related to the desired final output.

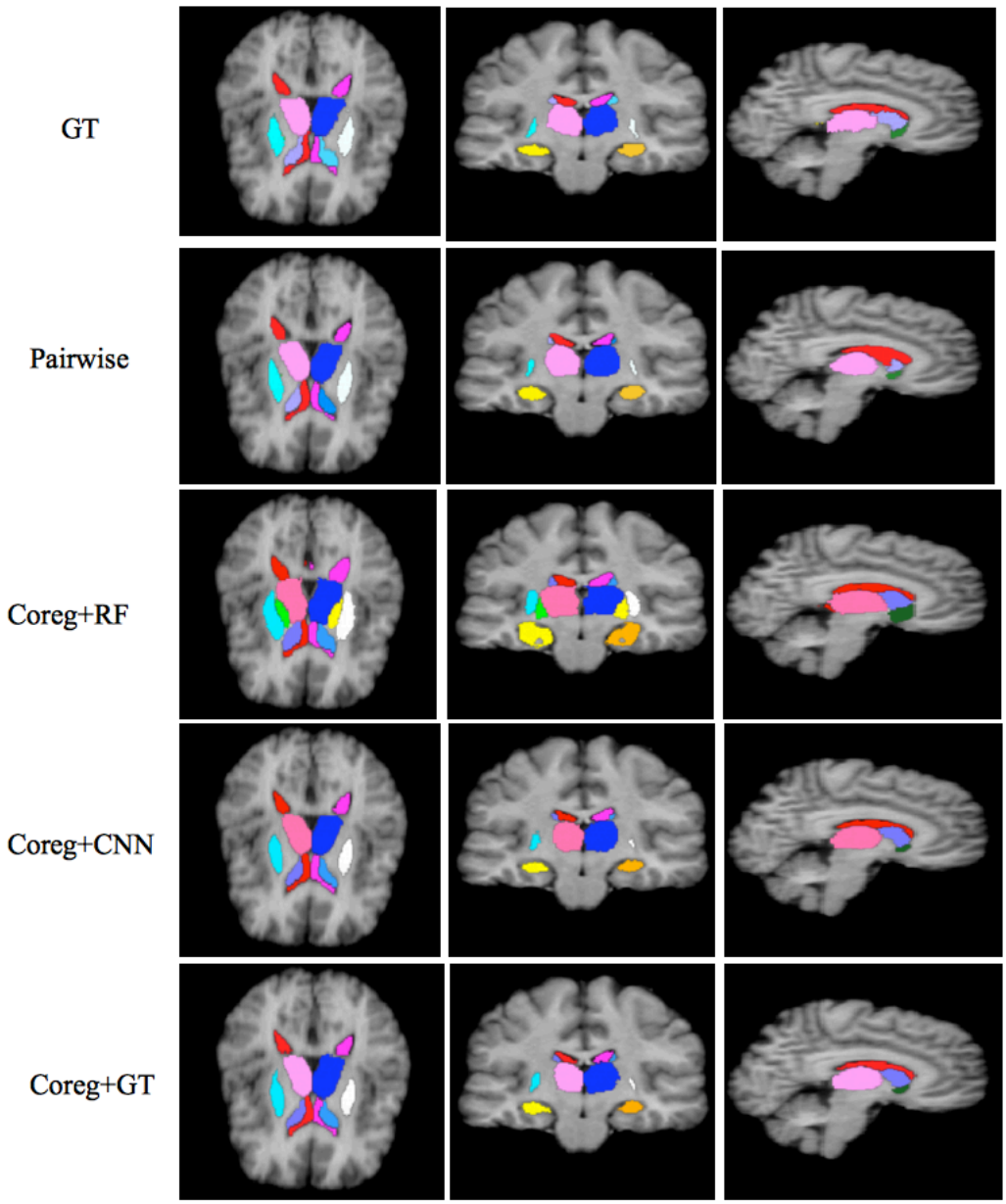


Figure 4.7: Segmentation results for three different views (axial, coronal, and sagittal). Coreg-CNN and Coreg-GT (oracle) generate more accurate segmentations compared to pairwise and RF-based approaches. Coreg-CNN can be used as a reliable substitute for ground truth annotations in multi-atlas co-registration-segmentation (better viewed in color and magnified).

4.3 Deformable Registration through Learning of Context-Specific Metric Aggregation

In this section we describe the second contribution of this chapter: a novel weakly supervised discriminative algorithm for learning context specific registration metrics as a linear combination of conventional similarity measures. The advantage of combining conventional metrics rather than performing actual metric learning is computational efficiency. Such metrics have been well optimized over the past two decades and therefore both their strengths and limitations are known while efficient implementation strategies have been developed for them as well. The challenge remains to find the optimal relative weighting (or parameters) of different metrics forming the data term. Hand-tuning these parameters would result in sub optimal solutions and quickly become infeasible as the number of metrics increases. Furthermore such hand-crafted combination can only happen at global scale (entire volume) and therefore will not be able to account for the different tissue properties of the observed anatomy. In order to alleviate this problem, we propose a weakly supervised learning algorithm for estimating these parameters locally, conditioned to the data semantic classes. The objective function of our formulation turns out to be a special case of non-convex function, known as the difference of convex function. We use the well known concave convex procedure to obtain the minima or the saddle point of the optimization problem. We show the efficacy of our approach on three highly challenging datasets in the field of medical imaging, for different anatomical structures and image modalities.

4.3.1 Introduction

Deformable registration involves the definition of a similarity criterion that, once endowed with a deformation model and a smoothness constraint, determines (through an optimization strategy) the optimal transformation to align two given images. The definition of this metric function is among the most critical components of the registration process. In that spirit, recent efforts have been made towards learning them from data. A discussion about metric learning for image registration was presented at the beginning of this chapter. As mentioned, machine learning methods, like support vector machines [Lee 2009] or similarity sensitive hashing [Bronstein 2010] have been considered to find mappings between different image modalities. However, these methods rely on the existence of exact massive correspondences at learning which are usually difficult to obtain. Instead, we focus on cases where the ground-truth is given in the form of segmentation masks, which is more common in real scenarios than availability of dense deformation fields or pre-aligned images.

In a number of applications, segmentation maps might be available in one of the volumes to be registered (e.g. adaptive radiotherapy, patient follow up exams or multi-atlas segmentation). Human anatomy is domain specific (deformations as well) and depending on the organ to be imaged one can expect different ade-

quacy between the local structure and different mono or multimodal metrics (which somehow reflect the observed tissue properties). Hence, the available segmentation maps could be used to locally suggest efficient similarity measures depending on the organ. The aim of this section is to propose a novel theoretical framework where domain specific optimal combination of metrics is derived according to the clinical task / observed images. The degrees of freedom of the learning procedure are the class weight combination of the different metrics.

It has been empirically shown that the data terms in the energy function have great influence on the accuracy of the solution to the registration problem. Normally, a data term is the function of a metric such as mutual information, sum of absolute differences, normalized cross correlation, etc. A particular data term is thus chosen based on the application. We argue that instead of using one metric, the locally adaptive and content-specific combination of different metrics can further improve the accuracy of the registration task, in particular, in the presence of semantic labels which then make the registration a domain specific problem.

The idea of combining different similarity measures (not necessarily considering them as domain or content-specific) has already been explored. In [Cifor 2012] it was shown that multichannel registration produces more robust registration results when compared to using the features individually. In this case, they used gray intensity value, phase congruency and local phase as features. In a posterior work, [Cifor 2013a] proposed a methodology that does not require to explicitly weight the features, by estimating different deformation fields from each feature independently, and then composing them into a final diffeomorphic transformation. Such a strategy produces multiple deformation models (as many as the number of metrics) that in general are locally inconsistent. Therefore, their combination will not be trivial and, in the general case, not anatomical meaningful.

In deformable registration, it is crucial to choose the right relative weighting between the different metrics and the pairwise smoothness term. One naive way would be to choose these relative weights by cross validation (or hand tuning) over the parameters. Clearly, this approach quickly becomes infeasible as the number of metrics increases while it can be applied only at global scale and cannot accommodate local anatomical differences. In order to circumvent the above mentioned problem, we would like to learn the relative weights from a given training dataset using a learning framework. We propose a novel discriminative learning framework, based on the well known structured support vector machines (SSVM) [Taskar 2003, Tsochantaridis 2004], to learn the relative weights (or the parameters). The SSVM and its extension to latent models LSSVM [Yu 2009] have received considerable attention in the recent years for parameter learning. Few recent examples of such frameworks are [Baudin 2013, Kumar 2015].

Our focus is mainly on the 3D to 3D deformable registration problem where the input and the output images are 3D volumes. However, the same framework can be trivially used for other registration problems as well, as 2D to 2D, or even for slice-to-volume registration. One of the key issues we face in all these problems is that the ground truth deformations are not known. This leads us to adopt a weakly

supervised learning framework where we treat the ground truth deformations as the latent variables. We model the latent variable imputation problem as the deformable registration problem with additional constraints. These constraints ensure that when the latent deformations are applied to the source image, the deformed source image is maximally aligned with the target image. The alignment accuracy is measured based on a loss function.

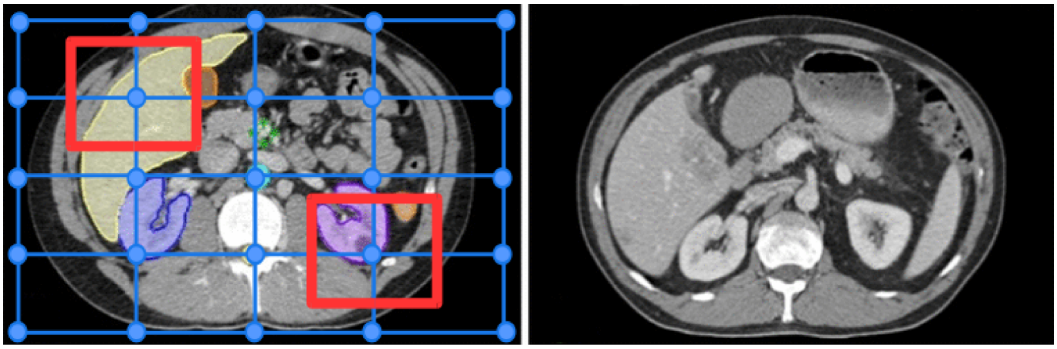
Our learning framework, similar to the LSSVM, is a special case of non-convex optimization problems, known as the difference of convex functions. The local optimum or the saddle point of such non-convex function can be obtained efficiently using the well known CCCP algorithm [Yuille 2003]. We demonstrate the efficacy of our framework using three challenging medical imaging datasets.

4.3.2 Multi-Metric Deformable Registration

Let us assume that we are given a source $3D$ volume (or image) I, source $3D$ segmentation mask S^I, and the target $3D$ volume (or image) J. The size of the segmentation mask is the same as that of the corresponding image. The segmentation mask is formed by the elements (or voxels) $s_k \in \mathcal{C}$, where $\mathcal{C}$ is the set of classes. Without loss of generality, we assume that the elements in the class set $\mathcal{C}$ are the discrete variables starting from one.

The problem is formulated using a MRF that consists of a regular grid graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges. Each node $i \in V$ corresponds to a control point $\boldsymbol{p_i}$. Each control point $\boldsymbol{p_i}$ is allowed to move in the $3D$ space, therefore, can be assigned a label $l_i \in \mathcal{L}$ associated to a displacement vector $\boldsymbol{d_{l_i}} \in \Re^3$ (in figure 2.2 we illustrate how this label space is built as a discretization of the 3D space). Notice that each $3D$ displacement vector is a tuple defined as $\boldsymbol{d_{l_i}} = (d_x, d_y, d_z)$, where d_x, d_y, and d_z are the displacements in the x, y, and z directions, respectively. A quantized version of the deformation field $\mathcal{D}^\Gamma$ is associated to a labeling Γ of the graph $\mathcal{G}$. In another words, the labeling $\Gamma \in \mathcal{L}^{|V|}$ assigns to a each node $i \in V$ a label l_i associated to a displacement vector $\boldsymbol{d_{l_i}}$. Note that $\mathcal{D}^\Gamma$ is a control point based representation of a dense deformation field, which is interpreted as a FFD. We will call $\mathring{\mathcal{D}}^\Gamma$ to the dense deformation field obtained as an FFD interpolation of the control point based representation $\mathcal{D}^\Gamma$. We denote the control point $\boldsymbol{p'_i}$ as the new control point when the displacement $\boldsymbol{d_{l_i}}$ is applied to the original control point $\boldsymbol{p_i}$. Let us define a patch $\bar{\Omega}^I_{l_i}$ on the source image I centered at the displaced control point $\boldsymbol{p'_i}$ (after applying label l_i). Similarly, we define Ω^J_i as the patch on the target image J centered at the original control point $\boldsymbol{p_i}$, and $\bar{\Omega}^{S^I}_{l_i}$ as the patch on the input segmentation mask centered at the displaced control point $\boldsymbol{p'_i}$.

Given the definition of these variables, let us define the unary feature vector corresponding to the i^{th} node for a given label l_i as $\mathcal{U}_i(l_i; I, J) = (u_1(\bar{\Omega}^I_{l_i}, \Omega^J_i), \cdots, u_n(\bar{\Omega}^I_{l_i}, \Omega^J_i)) \in \mathbb{R}^n$, where n is the number of metrics and $u_j(\bar{\Omega}^I_{l_i}, \Omega^J_i)$ is the unary feature corresponding to the j^{th} metric evaluated using the patches $\bar{\Omega}^I_{l_i}$ and Ω^J_i. For example, in case the j^{th} metric is the mutual information (MI) then the unary feature



(a) (b)

Figure 4.8: Context-specific metric aggregations. The multi-metric deformable registration algorithm uses a different aggregation of metrics depending on the context. It is determined by the dominant class in the corresponding image support area. In the example, we can observe that the liver (in yellow) and the kidney (in violet) are the dominant classes for the two highlighted control points, respectively. Note that source image (a) is annotated with semantic labels, while the target (b) is a simple intensity image.

$u_{MI}(\bar{\Omega}^I_{l_i}, \Omega^J_i)$ is the mutual information between the patches $\bar{\Omega}^I_{l_i}$ and Ω^J_i. In case of single metric, then n is equal to 1. Recall that we have $|\mathcal{C}|$ number of classes. Therefore, given a weight matrix $W \in \mathbb{R}^{n\times|\mathcal{C}|}$, where $W(i,j)$ denotes the weight for the i^{th} metric corresponding to the class j, the unary potential for the i^{th} node for a given label l_i is computed as follows:

$$\bar{\mathcal{U}}_i(l_i; I, J, S^I, W) = \mathbf{w}(\bar{c})^\top \mathcal{U}_i(l_i; I, J) \in \mathbb{R}, \tag{4.15}$$

where, $\mathbf{w}(\bar{c}) \in \mathbb{R}^n$ is the $\bar{c}^{th}$ column of the weight matrix W and $\bar{c}$ is the most dominant class in the patch on the source segmentation mask $\bar{\Omega}^{S^I}_{l_i}$. The dominant class $\bar{c}$ is obtained as follows:

$$\bar{c} = \arg\max_{c\in\mathcal{C}} f(\bar{\Omega}^{S^I}_{l_i}, c), \tag{4.16}$$

where, $f(\bar{\Omega}^{S^I}_{l_i}, c)$ is the number of voxels of class c in the patch $\bar{\Omega}^{S^I}_{l_i}$. Notice that one can use other criteria to find the dominant class.

The pairwise clique potential between the control points $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$ is defined as $f(l_i, l_j)$, where $f(\cdot,\cdot)$ is the L_1 norm between the two displacement vectors $\boldsymbol{d_{l_i}}, \boldsymbol{d_{l_j}}$ corresponding to the input labels (other regularizers could be used as well, as it was mentioned in section 2.3). Under this setting, the multi-class energy function corresponding to the deformable registration task is defined as:

$$\mathcal{E}(\Gamma; I, J, S^I, W) = \sum_{i\in V} \bar{\mathcal{U}}_i(l_i; I, J, S^I, W) + \sum_{(i,j)\in E} f(l_i, l_j). \tag{4.17}$$

Therefore, we aim at finding the optimal labeling $\hat{\Gamma}$ (associated to the quantized version of the deformation field $\mathcal{D}^{\hat{\Gamma}}$) by solving the following problem:

$$\hat{\Gamma} = \arg\min_{\Gamma \in \mathcal{L}^{|V|}} \mathcal{E}(\Gamma; I, J, S^I, W). \tag{4.18}$$

Approximate Inference Algorithm Similar to [Glocker 2008], we adopt a pyramidal approach that allows us to refine the search space at every level and, at the same time, capture a big range of deformations. We use the well known FastPD [Komodakis 2007a] (see section 2.4.3) as the inference algorithm at every level of the pyramid. Notice that the energy function from equation 4.17 is defined over the nodes and the edges of the sparse graph G which represents the deformation field, and not over the dense voxels and the neighbourhood system defined over the input image I. The reason being that the input images are too big and thus can not be optimized efficiently. Once we obtain the optimal labeling $\hat{\Gamma}$, we estimate the dense deformation field $\mathring{\mathcal{D}}^{\hat{\Gamma}}$ from its quantized representation $\mathcal{D}^{\hat{\Gamma}}$ using the FFD interpolation model [Rueckert 1999] in order to warp the input image, as it was already mentioned.

4.3.3 Learning the parameters

In the previous section we assumed that the weight matrix W is given to us. However, this assumption becomes unrealistic quickly as the number of metrics and the classes increases. In order to circumvent this problem, we propose an algorithm to learn the weights using a given dataset. Our algorithm is based on the well known latent structured SVM framework [Taskar 2003, Tsochantaridis 2004, Yu 2009] which optimizes an upperbound on the empirical risk. Instead of learning the complete weight matrix at once, we learn the weights for each class $c \in \mathcal{C}$ individually. From now onwards, the weight vector $\mathbf{w}_c$ denotes a particular column of the weight matrix W, which represents the weights corresponding to a particular class. We use the words 'parameters' and 'weights' interchangeably. In what follows we talk about the learning algorithm in details.

4.3.3.1 Preliminaries

Dataset We consider a dataset $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1,\cdots,N}$. Each $\mathbf{x}_i$ is a pair represented as $\mathbf{x}_i = (I_i, J_i)$, where I_i is the source volume (or the source image) and J_i is the target volume (or the target image). Similarly, each $\mathbf{y}_i$ is a pair represented as $\mathbf{y}_i = (S_i^I, S_i^J)$, where S_i^I and S_i^J are the segmentation masks for the source and target images, respectively. The size of each segmentation mask is the same as that of the corresponding images. As stated earlier, the segmentation mask is formed by the elements (or voxels) $s_k \in \mathcal{C}$, where $\mathcal{C}$ is the set of classes.

The Loss Function The loss function $\Delta(S^I, S^J) \in \mathbb{R}_{\geq 0}$ evaluates the similarity between the two segmentation masks S^I and S^J. Higher value of $\Delta(.,.)$ implies

stronger dissimilarity between the segmentations. Since our final evaluation is based on the dice coeffcient, therefore, we use a dice based loss function. Note that we follow the empirical risk minimization strategy, in which we should optimize the risk over which the evaluation is performed. The loss is therefore defined as:

$$\Delta(S^I, S^J) = 1 - DICE(S^I, S^J). \tag{4.19}$$

We approximate the dice between the segmentation masks as defined below:

$$DICE(S^I, S^J) = 2 \sum_{i \in V} \frac{|\phi(\boldsymbol{p_i^I}) \cap \phi(\boldsymbol{p_i^J})|}{|\phi(\boldsymbol{p_i^I})| + |\phi(\boldsymbol{p_i^J})|}, \tag{4.20}$$

where, $\phi(\boldsymbol{p_i^I})$ and $\phi(\boldsymbol{p_i^J})$ are the patches at the control point $\boldsymbol{p_i}$ on the segmentation masks S^I and S^J, respectively. The function $|.|$ represents the cardinality of a given set. The above approximation of the dice makes it decomposable over the nodes of the graph G. As will be discussed shortly, this decomposition allows us to train our algorithm very efficiently.

Joint Feature Map Given the parameters $\mathbf{w}_c$ for a particular class, the labeling Γ, and the input tuple $\mathbf{x}$, the mutli-class energy function (4.17) can be trivially converted into class based energy function as follows:

$$\mathcal{E}_c(\Gamma; \mathbf{x}, \mathbf{w}) = \mathbf{w}_c^\top \sum_{i \in V} \mathcal{U}_i(l_i; \mathbf{x}) + w_p \sum_{(i,j) \in E} f(l_i, l_j). \tag{4.21}$$

where, $w_p \in \mathbb{R}_{\geq 0}$ is the parameter associated with the pairwise term. Let us denote the parameter vector $\mathbf{w} \in \mathbb{R}^{n+1}$ as the concatenation of $\mathbf{w}_c$ and w_p. The energy function (4.21) is linear in $\mathbf{w}$ and can be written as:

$$\mathcal{E}_c(\Gamma; \mathbf{x}, \mathbf{w}) = \mathbf{w}^\top \Psi(\Gamma; \mathbf{x}). \tag{4.22}$$

where, $\Psi(\Gamma; \mathbf{x}) \in \mathbb{R}^{n+1}$ is the joint feature map defined as:

$$\Psi(\Gamma; \mathbf{x}) = \begin{pmatrix} \sum_{i \in V} \mathcal{U}_i^1(l_i; \mathbf{x}) \\ \sum_{i \in V} \mathcal{U}_i^2(l_i; \mathbf{x}) \\ \vdots \\ \sum_{i \in V} \mathcal{U}_i^n(l_i; \mathbf{x}) \\ \sum_{(i,j) \in E} f(l_i, l_j) \end{pmatrix} \tag{4.23}$$

Notice that the energy function (4.21) does not depend on the source segmentation mask S^I. Source segmentation masks in the energy function (4.17) are used to obtain the dominant class using the equation (4.16), which in this case is not required. However, we will shortly see that the source segmentation mask S^I plays a crucial role in the learning algorithm to compute the loss function.

Latent Variables Ideally, the dataset D must contain the ground truth deformation labeling Γ corresponding to the source image I in order to compute the energy term defined in the equation (4.21). Since annotating the dataset with the ground truth deformation is non-trivial, we use them as the latent variables in our algorithm. As will be seen shortly, we impute these deformations using the given dataset ensuring that the loss (as defined in the equation (4.19)) between the source segmentation mask when deformed using the imputed deformation field, and the target segmentation mask is minimized. In what follows, we will refer to this step indistinctly as 'latent assignment step' or 'segmentation consistent registration'.

4.3.3.2 The Objective Function

Given the dataset D, we would like to learn the parameter vector $\mathbf{w}$ such that minimizing the energy function (4.21) leads to a deformation field which when applied to the source segmentation mask gives minimum loss with respect to the target segmentation mask. Let us denote $S \circ \mathring{\mathcal{D}}^{\Gamma}$ the deformed segmentation when the dense deformation $\mathring{\mathcal{D}}^{\Gamma}$ is applied to the segmentation mask S. Therefore, ideally, we would like to learn $\mathbf{w}$ such that:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \frac{1}{N} \sum_i \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}}, S_i^J). \tag{4.24}$$

where, $\bar{\Gamma} = \arg\min_{\Gamma} \mathcal{E}(\Gamma; \mathbf{x}_i, \mathbf{w})$. The above objective function is the empirical risk minimization based formulation. However, the objective function is highly non-convex in $\mathbf{w}$, therefore, minimizing it directly makes the algorithm sensitive in terms of convergence to bad local minima. In order to circumvent this problem, we optimize a regularized upper bound on the loss as follows:

$$\min_{\mathbf{w},\{\xi_i\}} \quad \frac{1}{2}||\mathbf{w}||^2 + \alpha||\mathbf{w} - \mathbf{w}_0||^2 + \frac{C}{N}\sum_i \xi_i, \tag{4.25}$$

$$s.t. \quad \min_{\Gamma, \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\Gamma}, S_i^J)=0} \mathbf{w}^\top \Psi(\mathbf{x}_i, \Gamma) \leq \mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{\Gamma}) - \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}}, S_i^J) + \xi_i, \forall \bar{\Gamma}, \forall i \tag{4.26}$$

$$w_p \geq 0, \xi_i \geq 0, \forall i. \tag{4.27}$$

The above objective function minimizes an upper bound on the dice based loss function denoted as the variable ξ_i, known as the slack. The first term in the objective function $||\mathbf{w}||^2$, is the regularization term used to avoid overfitting. The effect of the regularization term is controlled by the hyper-parameter C. The second term is the proximity term. This ensures that the learned $\mathbf{w}$ is close to the initialization $\mathbf{w}_0$. The effect of the proximity term can be controlled by the hyperparameter α.

Intuitively, for a given input sample, the constraints of the above objective function tries to enforce the condition that the energy corresponding to the best possible deformation field (with minimum loss, additional constraint to enforce coherence) must always be less than the energy corresponding to any other deformation field with a margin proportional to the loss. Notice that, the term

Algorithm 3 The CCCP Algorithm.

1: D, $\mathbf{w}_0$, C, α, η, the tolerance ε.
2: $t = 0$, $\mathbf{w}_t = \mathbf{w}_0$.
3: **repeat**
4: Segmentation consistent registration. For a given $\mathbf{w}_t$, impute the latent variables $\hat{\Gamma}_i$ for each sample by solving the problem (4.29).
5: Obtain the updated parameters $\mathbf{w}_{t+1}$ by solving the convex optimization problem (4.30). The most violated constraint can be found by solving the problem (4.31).
6: $t = t + 1$
7: **until** The objective function of the problem (4.25) does not decrease more than ε.

$\min_{\Gamma, \Delta(S_i^I \circ \hat{\mathcal{D}}^\Gamma, S_i^J)=0} \mathbf{w}^\top \Psi(\mathbf{x}_i, \Gamma)$ in equation 4.26, makes the problem non convex. Shortly we will see how to upperbound this term, also known as the latent variable imputation step, in order to make the problem convex.

4.3.3.3 The Learning Algorithm

The objective function (4.25) that optimizes an upper bound on the empirical risk, is non-convex. Hence, it can not be optimized efficiently to obtain the optimal set of parameters. However, it can be shown that the objective function is a non-convex function that can be re-written as the difference of convex functions, which can be seen as the sum of the convex and the concave functions. For such family of non-convex functions, the well known CCCP algorithm [Yuille 2003]) can be used to obtain efficiently a local minima or a saddle point (the cutting plane algorithm can produce the optimal solution, but in practice this is not possible due to the computational constraints associated with the dimension of the problem). Broadly speaking, the CCCP algorithm consist of three steps — (1) upperbounding the concave part at a given $\mathbf{w}$, which leads to an affine function in $\mathbf{w}$; (2) optimizing the resultant convex function (sum of convex and affine functions is convex); (3) repeating the above steps until the objective can not be decreased beyond a given tolerance of ε.

The CCCP algorithm for the optimization of the objective function (4.25) is shown in the Algorithm 3 (a detailed version of the algorithm is presented in the appendix A.3). The first step of upperbounding the concave functions (Algorithm 3, Line 4) is the same as the latent imputation step, which we call the *segmentation consistent registration* problem. The second step is the optimization of the resultant convex problem (Algorithm 3, Line 5), which, in this case, is the optimization of the SSVM. The optimization leads to updating the parameters. We use the well known cutting plane algorithm [Joachims 2009] for this purpose. In what follows we discuss these steps in detail.

Segmentation Consistent Registration As already discussed, the ground truth deformation is not known a priori. Thus, in this step, we generate the best possible ground truth deformation field at a given $\mathbf{w}$. This is same as the latent imputation step of the CCCP algorithm. Recall that we are interested in learning the parameters $\mathbf{w}$ such that the upper bound on the loss function, defined in equation (4.19), is minimized. This leads us to formulate the latent imputation step as an inference problem with additional constraints. These additional constraints ensure that the imputed deformation field deforms the input image such that the loss between the deformed input image and the target image is minimized. Mathematically, for a given parameter vector $\mathbf{w}$, the latent deformation is imputed by solving the following problem:

$$\hat{\Gamma}_i = \underset{\Gamma \in \mathcal{L}^{|V|}, \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\Gamma}, S_i^J)=0}{\arg\min} \mathbf{w}^\top \Psi(\mathbf{x}_i, \Gamma). \tag{4.28}$$

The above problem is hard to solve and may not have unique solution. Thus, we solve the relaxed version of the above problem as defined below:

$$\hat{\Gamma}_i = \underset{\Gamma \in \mathcal{L}^{|V|}}{\arg\min}\, \mathbf{w}^\top \Psi(\mathbf{x}_i, \Gamma) + \eta \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\Gamma}, S_i^J). \tag{4.29}$$

where, η controls the relaxation trade-off parameter. Since the loss function used is decomposable, the above problem is equivalent to the inference of the deformable registration with simple modifications on the unary potentials. Thus, it can be solved efficiently using the FastPD inference algorithm discussed in the section 2.4.3.

Updating the Parameters Once the latent variables have been imputed or the concave functions have been upperbounded, the resultant objective function can be written as:

$$\begin{aligned} \min_{\mathbf{w}, \{\xi_i\}} \quad & \frac{1}{2}||\mathbf{w}||^2 + \alpha||\mathbf{w} - \mathbf{w}_0||^2 + \frac{C}{N}\sum_i \xi_i, \\ s.t. \quad & \mathbf{w}^\top \Psi(\mathbf{x}_i, \hat{\Gamma}_i) \leq \mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{\Gamma}) - \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}}, S_i^J) + \xi_i, \forall \bar{\Gamma}, \forall i \\ & w_p \geq 0, \xi_i \geq 0, \forall i. \end{aligned} \tag{4.30}$$

where, $\hat{\Gamma}_i$ is the labeling associated to the quantized latent deformation field $\mathcal{D}^{\hat{\Gamma}_i}$ imputed by solving the problem (4.29). Intuitively, the above objective function tries to learn the parameters $\mathbf{w}$ such that the energy corresponding to the imputed deformation field is always less than the energy of any other deformation field with a margin proportional to the loss function with some positive slack. Notice that the above objective function has exponential number of constraints, one for each possible labeling $\bar{\Gamma} \in \mathcal{L}^{|V|}$. In order to alleviate this problem we use the cutting plane algorithm [Joachims 2009]. Let us briefly talk about the idea behind the cutting plane algorithm. For a given $\mathbf{w}$, each labeling $\bar{\Gamma}$ gives a slack. Therefore, instead of minimizing all the slacks for a particular sample at once, we rather find the labeling that leads to the maximum value of the slack and store this in a set known

as the working set. This is known as finding the most violated constraint. Therefore, instead of using exponentially many constraints, we now optimize our algorithm over the constraints stored in the working set. This process is repeated till no constraints can be added to the working set. The main ingredient of the above discussed cutting plane algorithm is *finding the most violated constraint*. As discussed earlier, the most violated constraint for the i^{th} sample is the deformation field associated to the labeling that maximizes the slack corresponding to this sample. Rearranging the terms in the constraints of the objective function (4.30) to obtain the slack, ignoring the constant term $\mathbf{w}^\top \Psi(\mathbf{x}_i, \hat{\Gamma}_i)$, and maximizing it with respect to the possible deformations (which is equivalent to minimizing the negative of it), leads to the following problem solving which gives the most violated constraint:

$$\bar{\Gamma}_i = \underset{\Gamma \in \mathcal{L}^{|V|}}{\arg\min} \Big(\mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{\Gamma}) - \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}}, S_i^J) \Big). \tag{4.31}$$

Since the loss function is decomposable, again, the above problem is equivalent to the deformable registration with modifications on the unary potentials. Thus, it can be solved efficiently using the FastPD based approximate inference algorithm discussed in the section 2.4.3.

4.3.3.4 Prediction

Once we obtain the learned parameters $\mathbf{w}_c$ for each class $c \in \mathcal{C}$ using the Algorithm 3, we form the matrix W where each column of the matrix represents the learned parameter for a specific class. This learned matrix is then used to solve the registration problem defined in the equation 4.17 using the approximate inference algorithm in a new setting where semantic labels are available in the source image but not in the target (like for example, the scenario of atlas-based segmentation).

4.3.4 Experiments and Results

We evaluated our registration algorithm and the learning framework on three different medical datasets – (1) RT Parotids, (2) RT Abdominal, and (3) IBSR. These datasets involve several anatomical structures, different image modalities, and inter/intra patient images, which makes the deformable registration task on these dataset highly challenging. In all the experiments, we cross validate the hyper parameters C and α, and use $\eta = 50$. We use four different metrics in all the experiments: (1) sum of absolute differences (SAD), (2) mutual information (MI), (3) normalized cross correlation (NCC), and (4) discrete wavelet coefficients (DWT). In all the experiments (single and multi-metric) we used the same set of parameters for the pyramidal approach based inference algorithm (discussed in the section 4.3.2). These parameters are as follows: 2 pyramid levels, 5 refinement steps per pyramid level, 125 labels, and distance between control points of 25mm in the finer level. The running time for each registration case is around 12 seconds. For the training, we initialized $\mathbf{w}_0$ with the hand tuned values for each metric: $\mathbf{w}_0 = (0.1, 10, 10, 10)$,

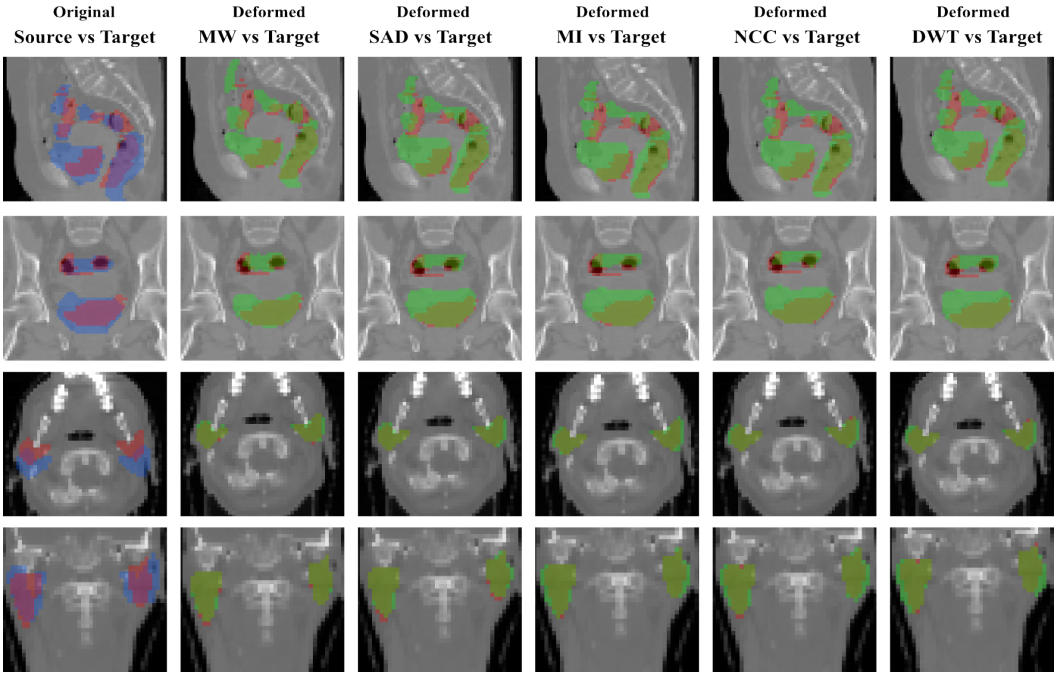


Figure 4.9: Overlapping of the segmentation masks in different views for one registration case from **RT Abdominal** (first and second rows) and **RT Parotids** (third and fourth rows) datasets. The first column corresponds to the overlapping before registration between the source (in blue) and target (in red) segmentation masks of the different anatomical structures of both datasets. From second to sixth column, we observe the overlapping between the warped source (in green) and the target (in red) segmentation masks, for the multiweight algorithm (MW) and for the single metric algorithm using sum of absolute differences (SAD), mutual information (MI), normalized cross correlation (NCC) and discrete wavelet transform (DWT) as similarity measure. This is coherent with the numerical results reported in Figures 2 and 3 from the main paper.

for SAD, MI, NCC, and DWT, respectively. Below we give details about the different datasets and discuss the results obtained in each of them.

RT Parotids The first dataset (RT Parotids) contains 8 CT volumes of head, obtained from 4 different patients, 2 volumes per patient. The volumes are captured in two different stages of a radiotherapy treatment in order to estimate the radiation dose. Right and left parotid glands were segmented by the specialists in every volume. The dimensions of the volumes are $56 \times 62 \times 53$ voxels with a physical spacing of 3.45mm, 3.45mm, and 4mm, in x, y, and z axes, respectively. We generated 8 pairs of source and target volumes using the given dataset. Notice that, while generating the source and target pairs, we did not mix the volumes coming from different patients. We splitted the dataset into train and test, and

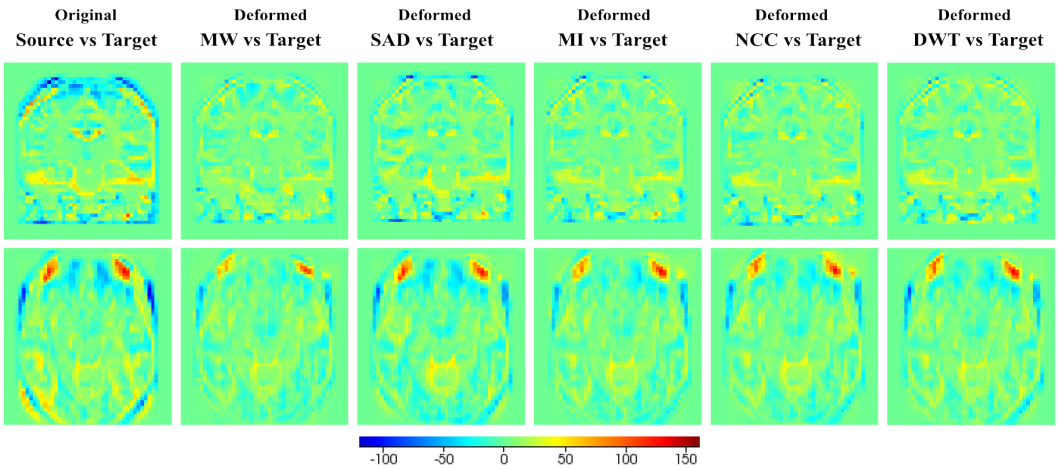


Figure 4.10: Qualitative results for one slice of one registration case from **IBSR dataset**. In this case, since showing overlapped structures in the same image is too ambiguous given that the segmentation masks almost cover the complete image, we are showing the intensity difference between the two volumes. This is possible since images are coming from the same modality and they are normalized. The first column shows the difference of the original volumes before registration. From second to sixth column we observe the difference between the warped source and the target images, for the multiweight algorithm (MW) and the single metric algorithm using sum of absolute differences (SAD), mutual information (MI), normalized cross crorrelation (NCC) and discrete wavelet transform (DWT) as similarity measure.

cross validated the hyperparameters C and α on the train dataset. The average result on the test dataset are shown in the Figure 4.11, while qualitative results can be found in image 4.9.

RT Abdominal The second dataset (RT Abdominal) contains 5 CT volumes of abdomen for a particular patient captured with a time window of about 7 days during a radiotherapy treatment. Three organs have been manually segmented by the specialists: (1) sigmoid, (2) rectum, and (3) bladder. The dimensions of the volumes are $90 \times 60 \times 80$ voxels with a physical spacing of 3.67mm, 3.67mm, and 4mm, in x, y, and z axes, respectively (there are small variations depending on the volume). We generated a train dataset of 6 pairs and test dataset of 4 pairs. The results on the test dataset are shown in the Figure 4.12, while qualitative results can be found in image 4.9.

IBSR The third dataset (IBSR) is the well known Internet Brain Segmentation Repository dataset, which consists of 18 brain MRI volumes, coming from different patients. Segmentations of three different brain structures are provided: white mater (WM), gray mater (GM), and cerebrospinal fluid (CSF). We used a down-

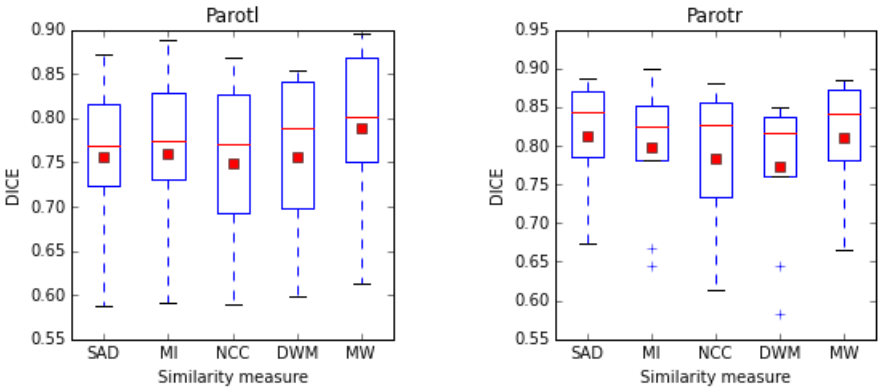


Figure 4.11: Results for the RT parotids dataset for the single-metric registration (SAD, MI, NCC, DWT) and the multi-metric registration (MW). The weights for the multi-metric registration are learned using the framework proposed in this work. 'Parotl' and 'Parotr' are the left and the right organs. The red square is the mean and the red bar is the median. It is evident from the results that using the learned linear combination of the metrics outperforms the single-metric based registration.

sampled version of the dataset to reduce the computation. The dimension of the volumes are $64 \times 64 \times 64$ voxels with a physical spacing of 3.75mm, 3.75mm, and 3mm in x, y, and z axes, respectively. To perform the experiments, we divided the 18 volumes in 2 folds of 9 volumes on each fold. This gave a total of 72 pairs per fold. We used an stochastic approach for the learning process, where we sample 10 different pairs from the training set, and we tested on the 72 pairs of the other fold. We run this experiment 3 times per fold, giving a total of 6 different experiments, with 72 testing samples and 10 training samples randomly chosen. The results on the test dataset are shown in the Figure 4.13.

As evident from the Figures 4.11, 4.12 and 4.13, the linear combination of similarity measures weighted using the learned coefficients outperforms the single metric based registration. In all the cases the dice for the multi-metric is higher than the dice for the single metric based registration, or it is as good as the best dice obtained using the single metric registration (please refer to the Figure 4.12, 'Parotr' to see the case in which the multi-metric is at least as good as the best obtained using the single metric). The results for the 'Sigmoid' organ in the Figure 4.12 show that in some cases the multi-metric based registration can significantly outperform the single metric based registration. Table 4.3.4 shows the average dice value per organ for the three datasets, considering the single and multi-metric approaches. The last column indicates the average dice improvement that our proposed method provides when compared with the single metric version. We achieved maximum average improvement of 0.033, 0.082 and 0.037 in terms of dice coefficient for RT Parotids, RT Abdominal and IBSR respectively.

Figure 4.14 shows the examples of the slices from the $3D$ volumes corresponding

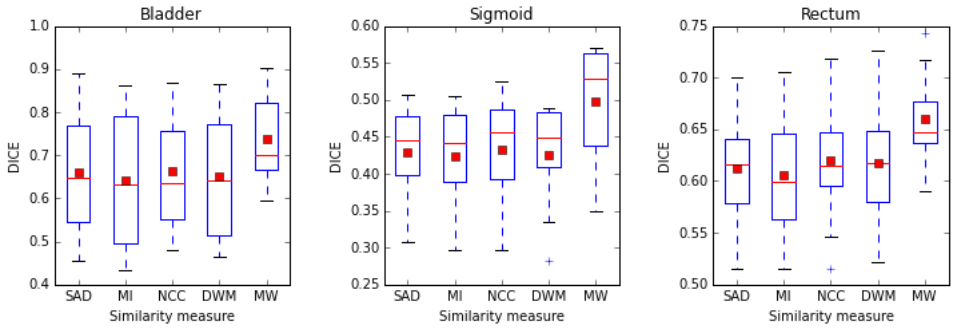


Figure 4.12: Results for the RT abdominal dataset for the single-metric registration (SAD, MI, NCC, DWT) and the multi-metric registration (MW). The weights for the multi-metric registration are learned using the framework proposed in this work. 'Bladder', 'Sigmoid', and 'Rectum' are the three organs on the dataset. The red square is the mean and the red bar is the median. It is evident from the results that using the learned linear combination of the metrics outperforms the single-metric based registration.

to each dataset. In figures 4.9 and 4.10 we include some qualitative results on the three challenging datasets in order to highlight the effects of learning the weights of different metrics. In the first one (figure 4.9), we present the overlapping of the segmentation masks in different views for one registration case from **RT Abdominal** and **RT Parotids** datasets, using single and multi-metric approaches. The observed results are coherent with the numerical results reported in figures 4.11,4.12. We observe that multiweight algorithm gives a better fit between the deformed and ground truth structures than the rest of the single similarity measures, which are over segmenting most of the structures showing a poorer registration performance. In the second graph (figure 4.10), we include results for the IBSR dataset. According to the scale in the bottom part of the image, extreme values (which mean high differences between the images) correspond to blue and red colors, while green indicates no difference in terms of intensity. Note how most of the big differences observed in the first column (before registration) are reduced in the multiweight algorithm, while some of them (specially in the peripheral area of the head) remain when using single metrics.

4.3.5 Discussion

The main contribution presented in this section is a novel and sophisticated framework for learning aggregations of image similarities in the context of deformable image registration. We also proposed a multi-metric MRF based image registration algorithm that incorporates such metric aggregations by weighting different similarity measures depending on the anatomical regions. We showed that associating different similarity criteria to anatomical regions yields results superior to the classic single metric approach. In order to learn this mapping in real scenarios where

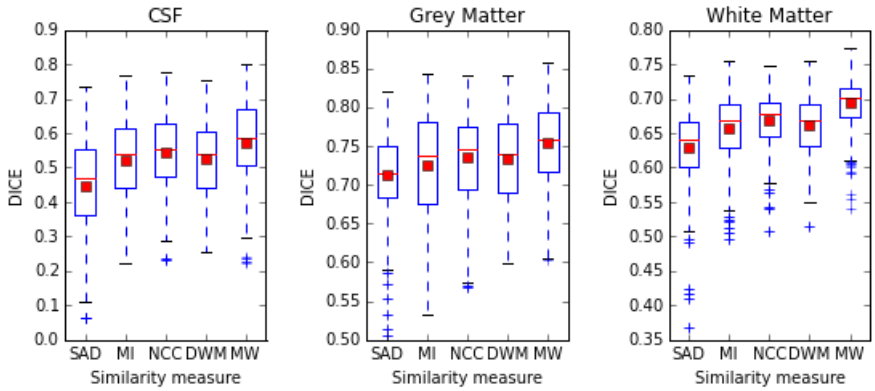


Figure 4.13: Results for the IBSR dataset for the single-metric registration (SAD, MI, NCC, DWT) and the multi-metric registration (MW). The weights for the multi-metric registration are learned using the framework proposed in this work. 'CSF', 'Grey Mater', and 'White Mater' are different structures in the brain. The red square is the mean and the red bar is the median. It is evident from the results that using the learned linear combination of the metrics outperforms the single-metric based deformable registration.

ground truth is generally given in the form of segmentation masks, we proposed to conceive deformation fields as latent variables and implement our problem using the LSSVM framework. The main limitation of our method is the need of segmentation masks for the source images in testing time. However, different real scenarios like radiation therapy or atlas-based segmentation methods fulfill this condition and can be improved through this technique.

Note that, at prediction (testing) time, the segmentation mask is simply considered to determine the metrics weights combination per control node (as indicated in equation 4.16). The segmentation labels are not used explicitly at testing time to guide the registration process that is purely image based. This is different from the approach presented in section 4.2, where we discussed how segmentation priors can be directly incorporated to registration process in a population-registration scenario. However, as it was mentioned, the co-registration and segmentation method requires a classifier able generate priors for both, source and target images. In the multi-metric approach proposed in this section, segmentation masks are only required (at testing time) for the source image and are only used to choose the best learned metric aggregation. The idea could be further extended to unlabeled data (as it concerns the source image) where the dominant label class per control node is the output of such a classification/learning method.

Dataset	Organ	SAD	MI	NCC	DWT	MW	Average dice increment for MW
RT Parotids	Parotl	0,756	0,760	0,750	0,757	0,788	0,033
	Parotr	0,813	0,798	0,783	0,774	0,811	0,019
RT Abdominal	Bladder	0,661	0,643	0,662	0,652	0,736	0,082
	Sigmoid	0,429	0,423	0,432	0,426	0,497	0,070
	Rectum	0,613	0,606	0,620	0,617	0,660	0,046
IBSR	CSF	0,447	0,520	0,543	0,527	0,546	0,037
	GM	0,712	0,725	0,735	0,734	0,761	0,035
	WM	0,629	0,658	0,669	0,661	0,682	0,028

Table 4.2: Average dice value per organ, for the single and multi-metric approaches, are reported for the three datasets. The last column indicates the average dice improvement that our proposed method provides when compared with the single metric approaches. We can observe improvements of a maximum of 8% points in terms of dice coefficient.

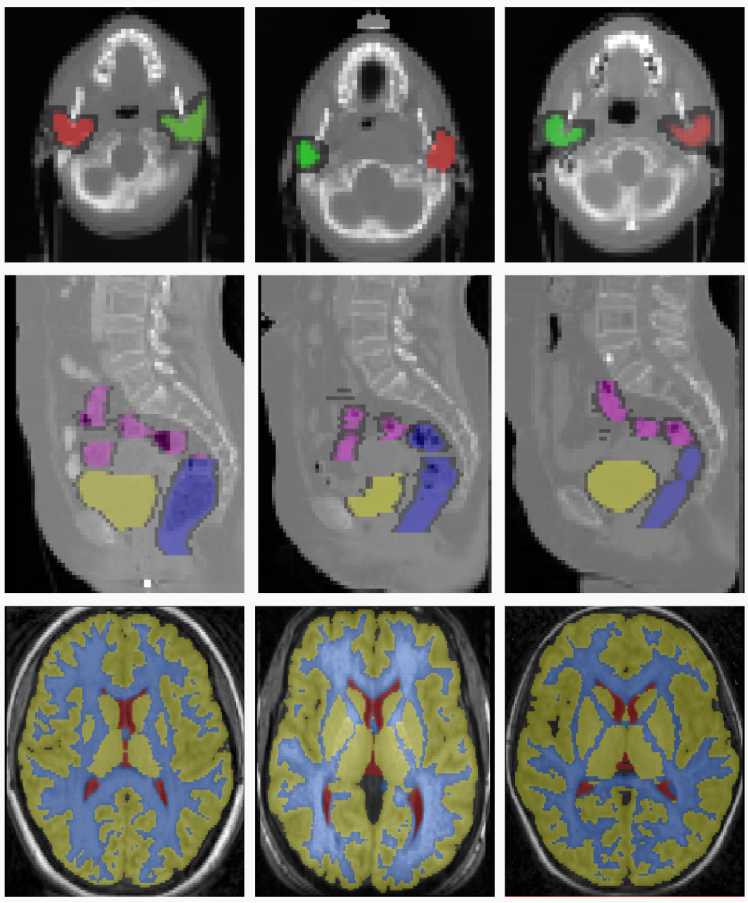


Figure 4.14: The top row represents the sample slices from three different volumes of the RT Parotids dataset. The middle row represents the sample slices of the RT Abdominal dataset, and the last row represents the sample slices from the IBSR dataset.

4.4 Conclusions and further ideas

In this chapter we explored two different strategies to incorporate prior knowledge to the registration process. In section 4.2 we proposed a novel method for co-registration and segmentation of a population of images, guided by semantic label likelihoods. We evaluated two different types of priors: probability maps obtained from a deep CNN and weaker annotations obtained with standard RF. Our approach aimed to: i) infer deformations that are anatomically plausible (smoothness constraints); ii) establish visual consistencies between all volumes according to an arbitrary metric; iii) enforce segmentation consistencies among all volumes, according to the predicted likelihoods. Experimental evaluation on conventionally adopted benchmarks, as well as on an additional clinical problem, demonstrates the effectiveness of our approach. Our experiments also show the importance of reliable segmentation priors. Label likelihoods extracted with a deep CNN outperform alternative methods and can replace ground truth annotations in co-registration with minimum loss in performance. In this context, future research directions include formulating the problem using high-order models to allow simultaneous estimation of all the deformation fields and to avoid the iterative approach. The use of decomposed optimization methods allowing efficient, one-shot optimization [Komodakis 2011] is under investigation.

Our second contribution in this chapter was a novel, generic, context-specific registration algorithm that is based on the principle of learning the combination of different metrics. Rather than employing highly customized solutions that suffer from scalability, portability and modularity, our approach relies on producing class-dependent metrics as linear combination of widely known and conventional monomodal and multimodal metrics. Consequently, the proposed registration method is scalable and modular and can be adjusted in any setting by simply changing the linear weights. These weights could be learned trivially offline using partially annotated data or could easily integrate new metrics. Extensive experimental validation on various challenging datasets demonstrated the potentials of the proposed method. The registration weights were learned by minimizing an upperbound on the dice based loss function. Dice is conventionally used but does not offer a very convincing picture as it concerns registration performance. The integration of alternative accuracy measures such as the Hausdorff distance between surfaces or even real geometric distances for anatomical landmarks could further enhance the performance and the robustness of the method. The use of alternative parameter learning mechanisms [Komodakis 2015] is also another interesting approach to explore.

Last but not least, an important research direction that involves the two contributions is testing the proposed methods on other real clinical problems, where co-registration and segmentation as well as context specific metrics are of great importance, such as adaptive radiotherapy.

Chapter 5

Conclusions and future directions

5.1 Contributions

In this thesis we propose several contributions to the theory of graph-based deformable registration. The first one (presented in chapter 3), focuses on extending it to a new application scenario not yet explored. We present three different models to deal with the challenging problem of slice-to-volume registration. We define low-rank and high-order graphical models, that simultaneously select the plane and estimates the in-plane deformation through a single shot optimization approach. We discuss the peculiarities of each model in a comparative study, emphasizing the robustness of this theory and illustrating the always underlying compromise between model complexity and power of expression. While exploring different approaches, we transfer the complexity from the label space to the topology of the graph, and we also experiment with high-order terms. Different types of discrete optimizers are used to perform parameter inference on two different datasets of medical images. We also compare the quality of our results with a standard rigid approach based on Simplex optimization of a global similarity measure. The results show that, even in a problem which is continuous by nature like slice-to-volume deformable registration, we can achieve equivalent performance by defining the label space in a smart way and exploring it using iterative large neighborhood search strategies.

Two different strategies incorporating prior information to the standard graph-based registration process are proposed in chapter 4. Both approaches demonstrate that semantic and context specific information can boost the registration algorithms by further improving the accuracy of the results. The first one (described in section 4.2) shows how prior information about the anatomical structures of a given set of images, can be encompassed within the energy function of a discrete population registration algorithm. We seek at performing joint registration/segmentation of the input images while exploiting context information provided by pre-trained classifiers. We experiment with both reliable as well as weaker priors. The first were obtained using a CNN classifier, while for the second we used random forests. The experimental validation shows that by using accurate segmentation priors we can obtain results close to the ground truth. Moreover, we also show that less accurate priors will also lead to acceptable results after performing co-registration-segmentation. The second contribution (discussed in section 4.3), adapts the well known LSSVM framework to learn content specific aggregations of metrics from anno-

tated data. We implemented this idea within a multi-metric version of the standard graph-based deformable registration algorithm, and we evaluate its performance on three datasets involving several organs, different image modalities, and inter/intra patient images, which make the deformable registration task highly challenging. In all these cases, the multi-metric approach produced more robust registration results than using the features individually. One of the key issues we faced when designing this approach, is that the ground truth deformations are not known a priori. Instead, annotations usually come in the form of hard segmentation or probability masks. This led us to adopt a weakly supervised learning framework where we treat the ground truth deformations as the latent variables. In such a way, our proposal is applicable to a wider range of problems than those methods requiring ground truth deformations.

5.2 Future directions

The contributions presented in this dissertation push the limits of the state-of-the-art methods concerning medical image registration. However, they are only small steps towards improving this challenging task in different directions where research has still to be done. To conclude this work, let us discuss some envisioned future research directions.

Boosting context specific methods

The advent of accurate classifiers efficiently augmenting the information provided by standard image modalities, rises a number of interesting questions regarding image registration, which are only partially answered in this thesis. Can this information be used to further improve registration results? Would it be possible to directly learn the deformation fields from segmentation masks, instead of conceiving them as latent variables in this process? Could we benefit from the results of the co-registration and segmentation process to boost the classifiers used to create the initial priors? All these open questions define new directions in which research can be conducted.

Combining our two context specific strategies might be possible. If we think about the co-registration and segmentation strategy, considering that we have the inferred semantic labels at our disposal, the proposed approach can support class-specific combinations of metrics that could be learned offline using the learning approach from section 4.3. Once the metric aggregations are learned, incorporating a multi-metric data term within the co-registration-segmentation approach would be simple since the inferred priors could be used as part of equation 4.16 to find the dominant classes. In such a way, having a classifier alleviates the main limitations of our multi-metric registration algorithm, i.e. the requirement of segmentation masks for the source image at testing time. The long-term goal would be to study the gains of combining different metrics per class and using them as content-adaptive potentials in the registration/segmentation energy.

Developing novel high-order models

Other interesting open problem is related to the use of high-order graphs and inference methods especially adapted to these cases. As it was shown, the power of expression of graphical models grows together with their complexity. However, different types of high-order models can be efficiently optimized [Komodakis 2009, Dokania 2015]. In section 3.2.3 we present a high-order model to solve slice-to-volume registration which is optimized using standard inference methods, whereas in section 4.2 we formulate prior based co-registration and segmentation as a high-order problem, which is then approximated through a pairwise formulation. In both cases more accurate results could be obtained if proper high-order inference adapted to the particular type of energies would be considered. Accordingly, dual decomposition methods [Komodakis 2009] could be contemplated as an alternative solution.

Extending and improving graph-based slice-to-volume registration

Numerous future developments built upon the proposed slice-to-volume registration framework can be envisioned. New hybrid data terms integrating geometric information, iconic similarity measures [Sotiras 2010b] and sensor based technologies (like the ones described in section 3.1.2.3) may lead to improvements in terms of accuracy and stability when image occlusions or abrupt changes occur. Domain/problem specific parameter learning [Baudin 2013, Komodakis 2015] towards improving the proposed models could have a positive influence on the obtained results. The learning framework presented in section 4.3 could be simply adapted to perform this task.

Slice-to-volume applications explored in this thesis are mainly related to pre and intra-operative image fusion. However, as it was mentioned in section 3.1.6, other scenarios like motion correction and volume reconstruction require slice-to-volume registration methods. It would be interesting to explore how our approaches performs in these cases. Moreover, a single slice at a time is considered in our formulation. However, as discussed in section 3.1.5, employing a multi slice-to-volume procedure -as opposed to single slice procedure- certainly improves the robustness of the registration process by augmenting the image support. This is even more interesting when the relative position between the slices is known a priori (e.g. the works of [Gill 2008, Heldmann 2009, Cifor 2013b, Yu 2011]). Our single slice-to-volume registration framework could be directly extended to the multi slice scenario by performing independent registrations for each slice. However, the prior knowledge about the relative positions or interaction dynamics among the slices as well as the wider image support would be ignored. A possible extension considering one grid-graph per slice (modeling in-plane deformations and plane selection parameters) and inter-graph connections (encoding inter-slice dependencies) could do the job.

Considering discrete rigid slice-to-volume registration where only 6-DOF are estimated, could also be of interest. The three models we proposed aim at perform-

ing deformable slice-to-volume registration finding the optimal dense deformation field. From this deformation field, we extract rigid and deformable parameters as explained in section 3.2. However, we mentioned that in some specific problems, rigid models are enough since deformations are not considered (see section 3.1.3.1). Unnecessary efforts are invested in exploring an immense parameter space when applying our method to these cases, where only 6-DOF should be considered. Therefore, rigid (only) slice-to-volume registration using discrete models constitutes an open research direction. In this regard, the work of [Zikic 2010b] about linear image registration using MRF and discrete methods could be extended to the case of rigid slice-to-volume registration.

Appendix A

Appendix

A.1 Factor Graphs

In this appendix, we give a brief description of the factor graph data structure, which was used in this thesis to implement the three graphical models proposed in Chapter 2 that aim at solving the challenging slice-to-volume registration problem.

A factor graph G' is a bipartite graph that factorizes a given global energy function, expressing which variables are arguments of which local functions [Kschischang 2001]. Given a graphical model of any order $G = \langle V, E \rangle$, we can derive a factor graph $G' = \langle V', F', E' \rangle$. Here, V' is the set of variable nodes formed by the nodes of G, F' is a the set of all the factors $f \in F'$ (where every f is associated to one clique G), and the set $E' \subset V' \times F'$ defines the relation between the nodes and the factors. Every factor f has a function $\varphi_f : V'^n \rightarrow \mathbb{R}$ associated with it, that might correspond to one of the data or regularization terms defined in previous sections. The energy function of a discrete labeling problem like the one presented in 2.4 in the context of factor graphs is then given by:

$$\mathcal{E}(x) = \sum_{f \in F'} \varphi_f(l_1^f, ..., l_n^f), \tag{A.1}$$

where x corresponds to a given labeling for the complete graph and $l_1^f ... l_n^f$ are labels given to the variables in the neighborhood (or scope) of the factor f.

A.2 Convolutional Neural Networks for Brain Segmentation

In this appendix, we give a short description of the Convolutional Neural Network (CNN) based classifier used to learn the segmentation priors applied in section 4.2. This classifier is part of our recent segmentation method that was presented in [Shakeri 2016].

In the past few years, deep learning techniques, and particularly CNNs, have rapidly become the tool of choice for tackling challenging computer vision tasks. CNNs were popularized by Lecun, after delivering state of the art results on handwritten digit recognition [LeCun 1998]. However, they fell out of favor in the following years, mostly due to hardware and training data limitations. Nowadays, the availability of large-scale datasets (e.g. ImageNet), powerful GPUs and appropriate

software libraries, have rekindled the interest in deep learning and have made it possible to harness their power. Krizhevsky et al. published [Krizhevsky 2012] results demonstrating clear superiority of deep architectures over hand-crafted features or shallow networks, for the task of image classification. Since then, CNNs have helped set new performance records for many other tasks; object detection [Girshick 2014], texture recognition [Cimpoi 2015] and object semantic segmentation [Chen 2014] just to name a few.

The CNN used in this work is similar in spirit to [Prasoon 2013], but with some notable differences. In [Prasoon 2013] the authors train one CNN for each of the three orthogonal views of MRI scans, for knee cartilage segmentation, with the loss being computed on the concatenated outputs of the three networks. The inputs to each CNN are 28×28 image patches and the output is a softmax probability of the central pixel belonging to the tibial articular cartilage. In contrast, our method operates on full 2D image slices, exploiting context information to accurately segment regions of interest in the brain. In addition, we use *fully convolutional* CNNs [Long 2015] to construct dense segmentation maps for the whole image, instead of classifying individual patches. Furthermore, our method handles multiple class labels instead of delivering a foreground-background segmentation, and it does that efficiently, performing a single forward pass in $5ms$.

We build a CNN for semantic segmentation, using ideas that were recently introduced for the Deeplab model by Chen et al [Chen 2014]. Deeplab is a fully convolutional neural network that computes a dense labelling of the input image. Compared to other systems, such as the one proposed by Long et al [Long 2015], Deeplab avoids excessive downsampling of the feature maps by skipping the subsampling steps in the last convolutional layers. In order to keep the same effective receptive field size, the authors introduce *holes* in a spirit similar to the "atrous" (with holes) algorithm [Mallat 1999]. For more details, we refer to [Chen 2014].

In our case we are interested in computing accurate priors for the location of sub-cortical regions in the human brain. Employing the hole trick, allows us to retain a reasonable output resolution, increasing the probability of capturing these fine structures. For training, the straightforward way would be to simply fine-tune the last three convolutional layers in the original Deeplab network, starting from the VGG-16 initialization for the remaining weights. However, given the very different nature of natural RGB images and MR image data (RGB vs grayscale, varying vs black background), we decided to train a fully convolutional network from scratch.

Our choice is not free of technical hurdles. Medical image datasets are usually smaller in size compared to natural image datasets, and high-quality ground truth annotations for segmentation are hard and costly to obtain. In our case, we only have a few 3D scans at our disposal, which increases the risk of overfitting. As we mentioned, the repeated pooling and sub-sampling steps that are applied to the input image, decrease the output resolution, making it difficult to detect and segment finer structures in the human brain. We attempt to overcome these hurdles through a series of design choices for the network. First, we use a shallow network, compared to VGG-16/Deeplab, composed of five pairs of convolutional/max pooling

Table A.1: Layers used in our architecture. One-pixel stride is used in all convolutional layers. A hole stride of "1" means that we introduce no holes.

Block	conv	# filters	hole stride	pool(stride)	dropout
1	7×7	64	1	3×3 (2)	no
2	5×5	128	1	3×3 (2)	no
3	3×3	256	2	3×3 (1)	yes
4	3×3	512	2	3×3 (1)	yes
5	3×3	512	2	3×3 (1)	yes
6	4×4	1024	4	no pooling	yes
7	1×1	39	1	no pooling	no

layers. Sub-sampling takes place only in the first two max-pooling layers; in the remaining layers we use a stride of 1 and introduce holes as in [Chen 2014]. A 256×256 input image gives as output a $64 \times 64 \times C$ array (sub-sampled by a factor of 4), where C is the number of class labels. We use $1-$pixel stride for all convolutional layers and 0.5 activation probability for all dropout layers. In A.1 we report the layer types and parameter values in detail.

At test time, our network processes 2D slices from a 3D volume. The output is a three-dimensional array of probability maps (one for each class/brain structure), obtained via a softmax operation. We can readily obtain a brain segmentation at this stage, by simply resizing the output probability array to the input image dimensions using bilinear interpolation and assigning at each pixel the label with the highest probability. As we discuss in section 4.2, these segmentations can be used by the co-registration-segmentation algorithm to guide the deformation process.

A.3 CCCP algorithm (detailed version)

Here we include a detailed version of the CCCP algorithm used to solve the learning problem of section 4.3 (see algorithm 3).

Algorithm 4 The CCCP algorithm (detailed version).

1: $\mathcal{D}$, $\mathbf{w}_0$, C, α, η, the tolerance ε.
2: $t \leftarrow 0$.
3: $\mathbf{w}_t \leftarrow \mathbf{w}_0$
4: **repeat**
5: For a given $\mathbf{w}_t$, impute the latent variables $\hat{\Gamma}_i$ for each sample by solving the problem:

$$\hat{\Gamma}_i = \underset{\Gamma \in \mathcal{L}^{|V|}}{\arg\min} \Big(\mathbf{w}_t^\top \Psi(\mathbf{x}_i, \Gamma) + \eta \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\Gamma}, S_i^J) \Big).$$

6: Initialize the constraint set for each sample: $\mathcal{W}_i \leftarrow \emptyset, \forall i$.
7: **repeat**
8: Obtain the most violated constraint (compute $\bar{\Gamma}_i$ for each sample):

$$\bar{\Gamma}_i = \underset{\bar{\Gamma} \in \mathcal{L}^{|V|}}{\arg\min} \Big(\mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{\Gamma}) - \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}_i}, S_i^J) \Big).$$

9: Update constraint set if $\bar{\Gamma}_i$ is sufficiently violated.

$$\mathcal{W}_i \leftarrow \mathcal{W}_i \cup \bar{\Gamma}_i, \forall i.$$

10: Solve the following optimization problem to obtain $\mathbf{w}$:

$$\begin{aligned} \min_{\mathbf{w}, \{\xi_i\}} \quad & \frac{1}{2}||\mathbf{w}||^2 + \alpha||\mathbf{w} - \mathbf{w}_0||^2 + \frac{C}{N}\sum_i \xi_i, \\ s.t. \quad & \mathbf{w}^\top \Psi(\mathbf{x}_i, \hat{\Gamma}_i) \leq \mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{\Gamma}_i) - \Delta(S_i^I \circ \mathring{\mathcal{D}}^{\bar{\Gamma}}, S_i^J) + \xi_i, \forall \bar{\Gamma}_i \in \mathcal{W}_i, \forall i \\ & w_p \geq 0, \xi_i \geq 0, \forall i. \end{aligned}$$

11: **until** No working set $\mathcal{W}_i$ can be further updated.
12: $t \leftarrow t + 1$
13: Update the parameters: $\mathbf{w}_t \leftarrow \mathbf{w}$
14: **until** Objective of the problem does not decrease more than ε.

Bibliography

[Ahuja 2002] Ravindra K. Ahuja, Özlem Ergun, James B. Orlin and Abraham P. Punnen. *A survey of very large-scale neighborhood search techniques.* Discrete Applied Mathematics, vol. 123, no. 1-3, pages 75–102, 2002. (Cited on page 92.)

[Alchatzidis 2014a] S Alchatzidis, A Sotiras and N Paragios. *Discrete Multi Atlas Segmentation using Agreement Constraints.* In BMVC, 2014. (Cited on pages 5, 87 and 89.)

[Alchatzidis 2014b] S Alchatzidis, A Sotiras and N Paragios. *Discrete Multi Atlas Segmentation using Agreement Constraints.* In BMVC, 2014. (Cited on page 97.)

[Alchatzidis 2015] Stavros Alchatzidis, Aristeidis Sotiras and Nikos Paragios. *Local atlas selection for discrete multi-atlas segmentation.* In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pages 363–367. IEEE, apr 2015. (Cited on page 5.)

[Andres 2012] Bjoern Andres, Jörg H. Kappes, Thorsten Beier, Ullrich Köthe and Fred A. Hamprecht. *The Lazy Flipper: Efficient Depth-Limited Exhaustive Search in Discrete Graphical Models.* pages 154–166. 2012. (Cited on pages 19, 20 and 64.)

[Arbel 2001] Tal Arbel, Xavier Morandi, Roch M Comeau and D Louis Collins. *Automatic Non-linear MRI-Ultrasound Registration for the Correction of Intra-operative Brain Deformations.* no. i, pages 913–922, 2001. (Cited on page 26.)

[Arbel 2004] Tal Arbel, Xavier Morandi, Roch M Comeau and D Louis Collins. *Automatic non-linear MRI-ultrasound registration for the correction of intra-operative brain deformations.* Computer aided surgery : official journal of the International Society for Computer Aided Surgery, vol. 9, no. December, pages 123–136, 2004. (Cited on page 26.)

[Ashburner 2005] John Ashburner and Karl J. Friston. *Unified segmentation.* NeuroImage, vol. 26, no. 3, pages 839–851, 2005. (Cited on page 85.)

[B Andres, JH Kappes 2010] U Köthe B Andres, JH Kappes. *The Lazy Flipper: MAP inference in higher-order graphical models by depth-limited exhaustive search.* Rapport technique, 2010. (Cited on page 20.)

[Ballard 1981] D.H. Ballard. *Generalizing the Hough transform to detect arbitrary shapes.* Pattern Recognition, vol. 13, no. 2, pages 111–122, jan 1981. (Cited on page 38.)

[Bao 2005] P Bao, J Warmath, R Galloway and A Herline. *Ultrasound-to-computer-tomography registration for image-guided laparoscopic liver surgery.* Surgical Endoscopy, vol. 19, no. 3, pages 424–429, mar 2005. (Cited on pages 33 and 49.)

[Baudin 2013] Pierre-Yves Baudin, Danny Goodman, Puneet Kumar, Noura Azzabou, Pierre G. Carlier, Nikos Paragios and M. Pawan Kumar. *Discriminative Parameter Estimation for Random Walks Segmentation.* pages 219–226. 2013. (Cited on pages 78, 103 and 121.)

[Bay 2008] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool. *Speeded-Up Robust Features (SURF).* Computer Vision and Image Understanding, vol. 110, no. 3, pages 346–359, jun 2008. (Cited on page 31.)

[Belongie 2002] S. Belongie, J. Malik and J. Puzicha. *Shape matching and object recognition using shape contexts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pages 509–522, apr 2002. (Cited on page 34.)

[Besag 1986] Julian Besag. *Statistical analysis of dirty pictures.* Journal of the Royal Statistical Society. Series B (Methodological), pages 259—-302, jan 1986. (Cited on pages 19, 20 and 64.)

[Besl 1992] P.J. Besl and H.D. McKay. *A method for registration of 3-D shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 239–256, 1992. (Cited on page 32.)

[Bhagalia 2009] Roshni Bhagalia, Jeffrey a. Fessler, Boklye Kim and Charles R. Meyer. *Improved fMRI time-series registration using joint probability density priors.* Proceedings of SPIE, vol. 7259, pages 72590J–72590J–9, 2009. (Cited on pages 29 and 41.)

[Bhattachayya 1943] A. Bhattachayya. *On a measure of divergence between two statistical population defined by their population distributions.* Bulletin Calcutta Mathematical Society, 1943. (Cited on page 34.)

[Birkfellner 2007] Wolfgang Birkfellner, Michael Figl, Joachim Kettenbach, Johann Hummel, Peter Homolka, RuÌdiger Schernthaner, Thomas Nau and Helmar Bergmann. *Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images.* Medical Physics, vol. 34, no. 1, page 246, 2007. (Cited on pages 24, 28, 29, 37, 45 and 48.)

[Birkfellner 2008] Wolfgang Birkfellner, Johann Hummel and Emmanuel Wilson. *Image-Guided Interventions Technology and Applications (Chapter 2, Tracking Devices).* In Terry Peters and Kevin Cleary, editeurs, Image-Guided Interventions Technology and Applications, chapitre Chapter 2, pages 23–45. Springer, 2008. (Cited on pages 26, 33 and 34.)

[Boer 2007] B.A. Boer, J.M Ruijter and F. P Voorbaak. *Towards the automatic registration of histological sections into a 3D reference mode.* In Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence., 2007. (Cited on page 32.)

[Boukouvalas 2011] Zois Boukouvalas. *Distance Metric Learning for Medical Image Registration.* no. August, 2011. (Cited on page 83.)

[Boykov 1998] Y. Boykov, O. Veksler and R. Zabih. *Markov random fields with efficient approximations.* In IEEE Conference on Computer Vision and Pattern Recognition, pages 648–655. IEEE Comput. Soc, 1998. (Cited on pages 12 and 13.)

[Boykov 2001] Y. Boykov, O. Veksler and R. Zabih. *Fast approximate energy minimization via graph cuts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pages 1222–1239, 2001. (Cited on pages 7, 13, 19, 86 and 92.)

[Boykov 2004] Y. Boykov and V. Kolmogorov. *An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pages 1124–1137, sep 2004. (Cited on page 19.)

[Bronstein 2010] Michael M Bronstein, Alexander M Bronstein, Fabrice Michel and Nikos Paragios. *Data fusion through cross-modality metric learning using similarity-sensitive hashing.* In CVPR 2010, pages 3594–3601. IEEE, 2010. (Cited on pages 84 and 102.)

[Brooks 2008] Rupert Brooks, D Louis Collins, Xavier Morandi and Tal Arbel. *Deformable Ultrasound Registration without Reconstruction.* vol. 5242, no. 2008, 2008. (Cited on pages 29, 39, 41, 45 and 46.)

[Brown 1992] Lisa Gottesfeld Brown. *A survey of image registration techniques.* ACM Computing Surveys, vol. 24, no. 4, pages 325–376, dec 1992. (Cited on page 3.)

[Cabezas 2011] Mariano Cabezas, Arnau Oliver, Xavier Llado, Jordi Freixenet and Meritxell Bach Cuadra. *A review of atlas-based segmentation for magnetic resonance brain images.* Computer Methods and Programs in Biomedicine, vol. 104, no. 3, pages e158–e177, 2011. (Cited on page 86.)

[Chan 2001] T.F. Chan and L.A. Vese. *Active contours without edges.* IEEE Transactions on Image Processing, vol. 10, no. 2, pages 266–277, 2001. (Cited on page 29.)

[Chandler 2004] Adam G. Chandler, Thomas Netsch, Chris A. Cocosco, Julia A. Schnabel and David J. Hawkes. *Slice-to-volume registration using mutual*

information between probabilistic image classifications. pages 1120–1129, may 2004. (Cited on page 29.)

[Chandler 2006] A.G. Chandler, R.J. Pinder, T Netsch, J.A. Schnabel, D.J. Hawkes, D.L.G. Hill and R Razavi. *Correction of Misaligned Slices in Multi-Slice MR Cardiac Examinations by Using Slice-To-Volume Registration.* In 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2006., volume 10, pages 474–477. IEEE, 2006. (Cited on pages 29 and 46.)

[Chandler 2008a] a G Chandler, R J Pinder, T Netsch, J a Schnabel, D J Hawkes, D L Hill and R Razavi. *Correction of misaligned slices in multi-slice cardiovascular magnetic resonance using slice-to-volume registration.* J Cardiovasc Magn Reson, vol. 10, page 13, 2008. (Cited on page 49.)

[Chandler 2008b] Adam G Chandler, Richard J Pinder, Thomas Netsch, Julia A Schnabel, David J Hawkes, Derek L G Hill and Reza Razavi. *Correction of misaligned slices in multi-slice cardiovascular magnetic resonance using slice-to-volume registration.* Journal of Cardiovascular Magnetic Resonance, vol. 10, no. 1, page 13, 2008. (Cited on pages 37 and 46.)

[Chen 2014] L Chen, G Papandreou, I Kokkinos, K Murphy and A L Yuille. *Semantic image segmentation with deep convolutional nets and fully connected crfs.* arXiv preprint arXiv:1412.7062, 2014. (Cited on pages 124 and 125.)

[Cifor 2012] Amalia Cifor, Laurent Risser, Daniel Chung, Ewan M. Anderson and Julia A. Schnabel. *Hybrid feature-based Log-Demons registration for tumour tracking in 2-D liver ultrasound images.* Proceedings - International Symposium on Biomedical Imaging, pages 724–727, 2012. (Cited on page 103.)

[Cifor 2013a] Amalia Cifor, Laurent Risser, Daniel Chung, Ewan M. Anderson and Julia A. Schnabel. *Hybrid feature-based diffeomorphic registration for tumor tracking in 2-D liver ultrasound images.* IEEE Transactions on Medical Imaging, vol. 32, no. 9, pages 1647–1656, 2013. (Cited on page 103.)

[Cifor 2013b] Amalia Cifor, Laurent Risser, Mattias P. Heinrich, Daniel Chung and Julia A. Schnabel. *Rigid Registration of Untracked Freehand 2D Ultrasound Sweeps to 3D CT of Liver Tumours.* Abdominal Imaging. Computation and Clinical Applications, MICCAI 2013, 2013. (Cited on pages 30, 45, 46 and 121.)

[Cimpoi 2015] M Cimpoi, S Maji and A Vedaldi. *Deep filter banks for texture recognition and segmentation.* In CVPR, 2015. (Cited on page 124.)

[Clatz 2005] O. Clatz, M. Sermesant, P.-Y. Bondiau, H. Delingette, S.K. Warfield, G. Malandain and N. Ayache. *Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation.* IEEE Transactions on Medical Imaging, vol. 24, no. 10, pages 1334–1346, oct 2005. (Cited on page 82.)

[Csapo 2012] Istvan Csapo, Brad Davis, Yundi Shi, Mar Sanchez, Martin Styner and Marc Niethammer. *Longitudinal Image Registration with Non-uniform Appearance Change*. pages 280–288. 2012. (Cited on page 4.)

[Dalal 2005] N. Dalal and B. Triggs. *Histograms of Oriented Gradients for Human Detection*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 886–893. IEEE, 2005. (Cited on page 31.)

[Dalvi 2008] Rupin Dalvi and Rafeef Abugharbieh. *Fast feature based multi slice to volume registration using phase congruency.* Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, vol. 2008, pages 5390–3, 2008. (Cited on page 32.)

[del Palomar 2008] A. Pérez del Palomar, B. Calvo, J. Herrero, J. López and M. Doblaré. *A finite element model to accurately predict real deformations of the breast*. Medical Engineering & Physics, vol. 30, no. 9, pages 1089–1097, nov 2008. (Cited on page 83.)

[Dokania 2015] Puneet K. Dokania and M. Pawan Kumar. *Parsimonious Labeling*. ICCV 2015, jul 2015. (Cited on page 121.)

[Elen 2010] An Elen, Jeroen Hermans, Javier Ganame, Dirk Loeckx, Jan Bogaert, Frederik Maes and Paul Suetens. *Automatic 3-D Breath-Hold Related Motion Correction of Dynamic Multislice MRI*. IEEE Transactions on Medical Imaging, vol. 29, no. 3, pages 868–878, mar 2010. (Cited on pages 28, 29, 37, 41 and 49.)

[Eresen 2014] Aydin Eresen, Peng Li and Jim Xiuquan Ji. *Correlating 2D histological slice with 3D MRI image volume using smart phone as an interactive tool for muscle study*. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 6418–6421. IEEE, aug 2014. (Cited on pages 29 and 34.)

[Fecamp 2015a] Vivien Fecamp, Aristeidis Sotiras and Nikos Paragios. *Modular linear iconic matching using higher order graphs*. In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pages 1097–1101. IEEE, apr 2015. (Cited on page 5.)

[Fecamp 2015b] Vivien Fecamp, Aristeidis Sotiras and Nikos Paragios. *Simultaneous Linear and Deformable Registration Through a Higher Order MRF Model*. In BAMBI, MICCAI 2015, 2015. (Cited on page 5.)

[Fei 2001] Baowei Fei, Andrew Wheaton, Zhenghong Lee, Kenichi Nagano, Jeffrey L Duerk and David L Wilson. *<title>Robust registration method for interventional MRI-guided thermal ablation of prostate cancer</title>*. In

Seong K. Mun, editeur, Medical Imaging 2001. International Society for Optics and Photonics (SPIE), volume 4319, pages 53–60, may 2001. (Cited on page 48.)

[Fei 2002] Baowei Fei, Jeffrey L. Duerk and David L. Wilson. *Automatic 3D registration for interventional MRI-guided treatment of prostate cancer*. Computer Aided Surgery, vol. 7, no. 5, pages 257–267, 2002. (Cited on pages 4, 28, 29, 37, 43 and 48.)

[Fei 2003a] Baowei Fei, Jeffrey L Duerk, Daniel T Boll, Jonathan S Lewin and David L Wilson. *Slice-to-volume registration and its potential application to interventional MRI-guided radio-frequency thermal ablation of prostate cancer*. IEEE Transactions on Medical Imaging, vol. 22, no. 4, pages 515–525, 2003. (Cited on pages 28, 37, 43, 45 and 48.)

[Fei 2003b] Baowei Fei, Kristin Frinkley and David L Wilson. *Registration Algorithms for Interventional MRI-guided Treatment of the Prostate*. Library, vol. 5029, pages 192–201, 2003. (Cited on pages 28 and 43.)

[Fei 2004a] Baowei Fei, Zhenghong Lee, Daniel T. Boll, Jeffery L. Duerk, D. Brace Sodee, Jonathan S. Lewin and David L. Wilson. *Image Registration and Fusion for Interventional MRI Guided Thermal Ablation of the Prostate Cancer*. MICCAI 2003, vol. 51, no. 1 I, pages 177–183, 2004. (Cited on page 48.)

[Fei 2004b] Baowei Fei, Zhenghong Lee, Daniel T. Boll, Jeffery L. Duerk, D. Brace Sodee, Jonathan S. Lewin and David L. Wilson. *Registration and fusion of SPECT, high-resolution MRI, and interventional MRI for thermal ablation of prostate cancer*. IEEE Transactions on Nuclear Science, vol. 51, no. 1 I, pages 177–183, 2004. (Cited on pages 28, 43 and 48.)

[Ferrante 2013] Enzo Ferrante and Nikos Paragios. *Non-rigid 2D-3D Medical Image Registration Using Markov Random Fields*. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, volume 8151, pages 163–170. 2013. (Cited on pages 6, 28, 39, 40, 43, 52 and 64.)

[Ferrante 2015a] Enzo Ferrante, Vivien Fecamp and Nikos Paragios. *Implicit planar and in-plane deformable mapping in medical images through high order graphs*. In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pages 721–724. IEEE, apr 2015. (Cited on pages 6, 28, 39, 40, 43, 44, 45, 53 and 64.)

[Ferrante 2015b] Enzo Ferrante, Vivien Fecamp and Nikos Paragios. *Slice-to-volume deformable registration: efficient one-shot consensus between plane selection and in-plane deformation*. International Journal of Computer Assisted Radiology and Surgery, vol. 10, no. 6, pages 791–800, jun 2015. (Cited on pages 6, 28, 29, 39, 40, 43, 44, 53 and 64.)

[Fogtmann 2014] Mads Fogtmann, Sharmishtaa Seshamani, Christopher Kroenke, Xi Cheng, Teresa Chapman, Jakob Wilm, Francois Rousseau and Colin Studholme. *A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy.* IEEE Transactions on Medical Imaging, vol. 33, no. 2, pages 272–289, 2014. (Cited on pages 28, 29 and 41.)

[Fortun 2015] Denis Fortun, Patrick Bouthemy and Charles Kervrann. *Optical flow modeling and computation: A survey.* Computer Vision and Image Understanding, vol. 134, pages 1–21, 2015. (Cited on page 3.)

[Frühwald 2009] Laura Frühwald, Joachim Kettenbach, Michael Figl, Johann Hummel, Helmar Bergmann and Wolfgang Birkfellner. *A comparative study on manual and automatic slice-to-volume registration of CT images.* European radiology, vol. 19, no. 11, pages 2647–53, nov 2009. (Cited on pages 28, 37, 43 and 48.)

[Fuerst 2014] Bernhard Fuerst, Wolfgang Wein, Markus Müller and Nassir Navab. *Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D LC(2) Metric.* Medical image analysis, vol. 18, no. 8, pages 1312–9, dec 2014. (Cited on pages 30, 37, 39 and 83.)

[Gass 2014] Tobias Gass, Gabor Szekely and Orcun Goksel. *Simultaneous Segmentation and Multiresolution Nonrigid Atlas Registration.* IEEE Transactions on Image Processing, vol. 23, no. 7, pages 2931–2943, jul 2014. (Cited on page 86.)

[Gefen 2005] Smadar Gefen, Louise Bertrand, Nahum Kiryati and Jonathan Nissanov. *Localization of Sections Within the Brain Via 2D to 3D Image Registration.* In Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., volume 2, pages 733–736. IEEE, 2005. (Cited on page 44.)

[Gefen 2008] Smadar Gefen, Nahum Kiryati and Jonathan Nissanov. *Atlas-based indexing of brain sections via 2-D to 3-D image registration.* IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pages 147–156, 2008. (Cited on pages 29, 38 and 44.)

[Ghanavati 2010] Sahar Ghanavati, Parvin Mousavi, Gabor Fichtinger, Pezhman Foroughi and Purang Abolmaesumi. *Multi-Slice to Volume Registration of Ultrasound Data to a Statistical Atlas of Human Pelvis.* SPIE proceedings: Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, vol. 7625, pages 76250O–76250O–10, 2010. (Cited on pages 35 and 44.)

[Ghanavati 2011] Sahar Ghanavati, Parvin Mousavi, Gabor Fichtinger and Purang Abolmaesumi. *Phantom validation for ultrasound to statistical shape model*

registration of human pelvis. In Kenneth H. Wong and David R. Holmes III, editeurs, Proc. of SPIE, volume 7964, page 79642U, mar 2011. (Cited on page 36.)

[Gholipour 2009] Ali Gholipour and Simon K Warfield. *Super-resolution reconstruction of fetal brain MRI.* MICCAI Workshop on Image Analysis for the Developing Brain (IADBâ2009). London, UK, pages 45–52, 2009. (Cited on pages 28, 43 and 50.)

[Gholipour 2010] Ali Gholipour, J A Estroff and S K Warfield. *Robust Super-Resolution Volume Reconstruction From Slice Acquisitions: Application to Fetal Brain MRI.* IEEE Transactions on Medical Imaging, vol. 29, no. 10, pages 1739–1758, oct 2010. (Cited on pages 28, 37, 43 and 50.)

[Gholipour 2011] Ali Gholipour, M. Polak, A. van der Kouwe, E. Nevo and S. K. Warfield. *Motion-robust MRI through real-time motion tracking and retrospective super-resolution volume reconstruction.* In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 5722–5725. IEEE, aug 2011. (Cited on pages 33 and 51.)

[Gill 2008] Sean Gill, Purang Abolmaesumi, Siddharth Vikal, Parvin Mousavi and Gabor Fichtinger. *Intraoperative Prostate Tracking with Slice-to-Volume Registration in MR.* In Proceedings of the 20th International Conference of the Society for Medical Innovation and Technology, 2008. (Cited on pages 4, 29, 37, 43, 46, 48 and 121.)

[Girshick 2014] R Girshick, J Donahue, T Darrell and J Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation.* In CVPR, 2014. (Cited on page 124.)

[Glocker 2008] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab and Nikos Paragios. *Dense image registration through MRFs and efficient linear programming.* Medical image analysis, vol. 12, no. 6, pages 731–41, 2008. (Cited on pages 5, 9, 13, 18, 39, 40, 43 and 106.)

[Glocker 2009a] Ben Glocker. *drop-User Guide Version 1.05.* Rapport technique, 2009. (Cited on page 17.)

[Glocker 2009b] Ben Glocker, Nikos Komodakis, Nassir Navab, Georgios Tziritas and Nikos Paragios. *Dense Registration with Deformation Priors.* In IPMI, volume 5636 LNCS, pages 540–551, 2009. (Cited on pages 5 and 82.)

[Glocker 2009c] Ben Glocker, Nikos Komodakis, Nikos Paragios and Nassir Navab. *Approximated Curvature Penalty in Non-rigid Registration Using Pairwise MRFs.* pages 1101–1109. 2009. (Cited on pages 13, 43 and 61.)

[Glocker 2010] Ben Glocker. *Random Fields for Image Registration.* PhD Thesis, 2010. (Cited on page 22.)

[Glocker 2011] Ben Glocker and Aristeidis Sotiras. *Deformable Medical Image Registration: Setting the State of the Art with Discrete Methods*. Annual review of biomedical engineering, vol. 13, no. April, pages 219–244, aug 2011. (Cited on pages 3, 4, 9, 13, 31, 39, 40, 43 and 55.)

[Gourdon 1994] A Gourdon and N Ayache. *Registration of a curve on a surface using differential properties*. 1994. (Cited on page 31.)

[Hammersley 1971] J. Hammersley and P. Clifford. *Markov Fields on Finite Graphs and Lattices*. Unpublished Manuscript, 1971. (Cited on page 10.)

[Harris 1988] Chris Harris and Mike Stephens. *A combined corner and edge detector*. In Proc. of Fourth Alvey Vision Conference, 1988. (Cited on page 31.)

[Heckemann 2010] Rolf A. Heckemann, Shiva Keihaninejad, Paul Aljabar, Daniel Rueckert, Joseph V. Hajnal and Alexander Hammers. *Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation*. NeuroImage, vol. 51, no. 1, pages 221–227, 2010. (Cited on pages 86 and 89.)

[Heinrich 2012] Mattias P. Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V. Gleeson, Sir Michael Brady and Julia a. Schnabel. *MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration*. Medical Image Analysis, vol. 16, no. 7, pages 1423–1435, 2012. (Cited on page 30.)

[Heinrich 2013] Mattias Paul Heinrich. *Deformable lung registration for pulmonary image analysis of MRI and CT scans*. PQDT - UK & Ireland, vol. U608438, 2013. (Cited on page 30.)

[Heldmann 2009] Stefan Heldmann and Nils Papenberg. *A Variational Approach for Volume-to-Slice Registration*. In Scale Space and Variational Methods in Computer Vision, volume 1, pages 624–635, 2009. (Cited on pages 28, 46 and 121.)

[Heldmann 2010] S Heldmann, B Beuthien and J Olesch. *Improved Minimal-Invasive Laparoscopic Liver Surgery by Registra- tion of 3D CT and 2D Ultrasound Slices*. Proc BMT, no. November, pages 2–4, 2010. (Cited on pages 41, 46 and 49.)

[Hoerth 2015] R.M. Hoerth, D. Baum, D. Knötel, S. Prohaska, B.M. Willie, G.N. Duda, H.-C. Hege, P. Fratzl and W. Wagermaier. *Registering 2D and 3D imaging data of bone during healing*. Connective Tissue Research, vol. 56, no. 2, pages 133–143, 2015. (Cited on page 38.)

[Holden 2008] M. Holden. *A Review of Geometric Transformations for Nonrigid Body Registration*. IEEE Transactions on Medical Imaging, vol. 27, no. 1, pages 111–128, jan 2008. (Cited on page 38.)

[Honal 2010] Matthias Honal, Jochen Leupold, Sandra Huff, Tobias Baumann and Ute Ludwig. *Compensation of breathing motion artifacts for MRI with continuously moving table.* Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 63, no. 3, pages 701–12, 2010. (Cited on pages 28, 39 and 41.)

[Horn 1980] Berthold K.P. Horn and Brian G. Schunck. *Determining Optical Flow.* Artificial Intelligence, vol. 17, pages 185–203, 1980. (Cited on page 3.)

[Huang 2009] Xishi Huang, John Moore, Gerard Guiraudon, Douglas L Jones, Daniel Bainbridge, Jing Ren and Terry M Peters. *Dynamic 2D ultrasound and 3D CT image registration of the beating heart.* IEEE Transactions on Medical Imaging, vol. 28, no. 8, pages 1179–1189, 2009. (Cited on pages 4, 29, 34, 36, 37, 41 and 48.)

[Hummel 2008] Johann Hummel, Michael Figl, Michael Bax, Helmar Bergmann and Wolfgang Birkfellner. *2D/3D registration of endoscopic ultrasound to CT volume data.* Physics in Medicine and Biology, vol. 53, no. 16, pages 4303–4316, aug 2008. (Cited on pages 28, 29, 33, 36, 42 and 43.)

[Iglesias 2015] Juan Eugenio Iglesias and Mert R. Sabuncu. *Multi-Atlas Segmentation of Biomedical Images: A Survey.* Medical Image Analysis, vol. 24, no. 1, pages 205–219, 2015. (Cited on pages 86 and 89.)

[Ishikawa 2009] Hiroshi Ishikawa. *Higher-order clique reduction in binary graph cut.* 2009 IEEE CVPR, pages 2993–3000, 2009. (Cited on page 12.)

[Jiang 2007a] Shuzhou Jiang, Hui Xue, Serena J Counsell, Mustafa Anjari, Joanna Allsop, Mary A Rutherford, Daniel Rueckert and Joseph V Hajnal. *In-utero 3D high resolution fetal brain diffusion tensor imaging.* In Nicholas Ayache, Sébastien Ourselin and Anthony Maeder, editeurs, Medical Image Computing and Computer-Assisted Intervention â MICCAI 2007, volume 4791 of *Lecture Notes in Computer Science*, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. (Cited on pages 4, 28, 29, 37, 43, 47 and 50.)

[Jiang 2007b] Shuzhou Jiang, Hui Xue, Alan Glover, Mary Rutherford, Daniel Rueckert and Joseph V. Hajnal. *MRI of moving subjects using multislice Snapshot images with Volume Reconstruction (SVR): Application to fetal, neonatal, and adult brain studies.* IEEE Transactions on Medical Imaging, vol. 26, no. 7, pages 967–980, 2007. (Cited on pages 28, 37, 43, 46 and 47.)

[Jiang 2009] Shuzhou Jiang, Hui Xue, Serena Counsell, Mustafa Anjari, Joanna Allsop, Mary Rutherford, Daniel Rueckert and Joseph V. Hajnal. *Diffusion tensor imaging (DTI) of the brain in moving subjects: Application to in-utero fetal and ex-utero studies.* Magnetic Resonance in Medicine, vol. 62, no. 3, pages 645–655, sep 2009. (Cited on pages 28, 43, 47 and 49.)

[Joachims 2009] T Joachims, T Finley and C.-N. Yu. *Cutting-plane Training of Structural {SVM}s.* Machine Learning, 2009. (Cited on pages 109 and 110.)

[Joshi 2000] S C Joshi and M I Miller. *Landmark matching via large deformation diffeomorphisms.* IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 9, no. 8, pages 1357–70, jan 2000. (Cited on page 31.)

[Kainz 2015] Bernhard Kainz, Markus Steinberger, Wolfgang Wein, Maria Kuklisova-Murgasova, Christina Malamateniou, Kevin Keraudren, Thomas Torsney-Weir, Mary Rutherford, Paul Aljabar, Joseph V Hajnal and Daniel Rueckert. *Fast Volume Reconstruction From Motion Corrupted Stacks of 2D Slices.* IEEE Transactions on Medical Imaging, vol. 34, no. 9, pages 1901–1913, sep 2015. (Cited on pages 28, 29, 37 and 47.)

[Kalman 1960] RE Kalman. *A new approach to linear filtering and prediction problems.* Trans ASME â J Basic Eng, vol. 82, pages 35–45, 1960. (Cited on page 34.)

[Kappes 2013] Jorg H Kappes, Bjoern Andres, Fred A Hamprecht, Christoph Schnorr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X Kausler, Jan Lellmann, Nikos Komodakis and Carsten Rother. *A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems.* In 2013 IEEE Conference on Computer Vision and Pattern Recognition, pages 1328–1335. IEEE, jun 2013. (Cited on pages 12, 18, 65 and 67.)

[Keysers 2003] Daniel Keysers and Walter Unger. *Elastic image matching is NP-complete.* Pattern Recognition Letters, vol. 24, no. 1-3, pages 445–453, 2003. (Cited on page 13.)

[Kim 1999] Boklye Kim, Jennifer L. Boes, Peyton H. Bland, Thomas L. Chenevert and Charles R. Meyer. *Motion correction in fMRI via registration of individual slices into an anatomical volume.* Magnetic Resonance in Medicine, vol. 41, no. 5, pages 964–972, 1999. (Cited on pages 4, 29, 33, 37, 43, 49 and 50.)

[Kim 2000] Boklye Kim, Charles R Meyer and Peyton H Bland. *Correction of local deformations in fMRI by 3D non-linear warping in map-slice-to-volume approach.* vol. 8, page 59412, 2000. (Cited on pages 39, 43 and 50.)

[Kim 2005] Tae-Seong Kim. *Co-registration of Multiple Postmortem Brain Slices to Corresponding MRIs Using Voxel Similarity Measures and Slice-to-Volume Transformation.* Journal of Biomedical Engineering Research, pages 231–241, 2005. (Cited on pages 28 and 41.)

[Kim 2008a] Boklye Kim, Desmond T B Yeo and Roshni Bhagalia. *Comprehensive mathematical simulation of functional magnetic resonance imaging time series including motion-related image distortion and spin saturation effect.* Magnetic resonance imaging, vol. 26, no. 2, pages 147–59, 2008. (Cited on page 50.)

[Kim 2008b] K. Kim, M. Hansen, P. Habas, F. Rousseau, O. Glenn, a. J. Barkovich and C. Studholme. *Intersection based registration of slice stacks to form 3D images of the human fetal brain.* 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI, no. November 2015, pages 1167–1170, 2008. (Cited on pages 28, 37, 42 and 47.)

[Kim 2010a] Kio Kim, P.A. Habas, Francois Rousseau, O.A. Glenn, A.J. Barkovich and Colin Studholme. *Intersection Based Motion Correction of Multislice MRI for 3D Fetal Brain Image Formation.* IEEE Transactions on Medical Imaging, vol. 29, no. 1, pages 146–158, jan 2010. (Cited on pages 28, 37, 42 and 47.)

[Kim 2010b] Kio Kim, Piotr a. Habas, Francois Rousseau, Orit a. Glenn, a. J. Barkovich, Meriam Koob, Jean-Louis Dietemann, Ashley J. Robinson, Kenneth J. Poskitt, Steven P. Miller and Colin Studholme. *Reconstruction of a geometrically correct diffusion tensor image of a moving human fetal brain.* Methods, no. November 2015, pages 76231I–76231I–9, 2010. (Cited on pages 41 and 47.)

[Kim 2011] Kio Kim, Piotr A Habas, Vidya Rajagopalan, Julia A Scott, James M Corbett-Detig, Francois Rousseau, A James Barkovich, Orit A Glenn and Colin Studholme. *Bias field inconsistency correction of motion-scattered multislice MRI for improved 3D image reconstruction.* IEEE transactions on medical imaging, vol. 30, no. 9, pages 1704–12, 2011. (Cited on page 47.)

[Kio Kim 2010] Kio Kim, Piotr Habas, Vidya Rajagopalan, Julia Scott, James Corbett-Detig, Francois Rousseau, Orit Glenn, James Barkovich and Colin Studholme. *Non-iterative relative bias correction for 3D reconstruction of in utero fetal brain MR imaging.* In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pages 879–882. IEEE, aug 2010. (Cited on page 47.)

[Klein 2009] Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, D Louis Collins, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P Woods, John J Mann and Ramin V Parsey. *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.* NeuroImage, vol. 46, no. 3, pages 786–802, 2009. (Cited on page 89.)

[Kohli 2012] P. Kohli and C. Rother. *Higher-Order Models in Computer Vision*. In Chapter 1, Image Processing and Analysis with Graphs (CRC Press). 2012. (Cited on page 12.)

[Koller 2009] Daphne Koller and Nir Friedman. Probabilistic Graphical Models, Principles and Techniques. MIT Press, 2009. (Cited on page 10.)

[Kolmogorov 2004] Vladimir Kolmogorov and Ramin Zabih. *What energy functions can be minimized via graph cuts?* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pages 147–159, feb 2004. (Cited on pages 11 and 12.)

[Komodakis 2005] N. Komodakis and G. Tziritas. *A new framework for approximate labeling via graph cuts*. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, pages 1018–1025 Vol. 2. IEEE, 2005. (Cited on page 21.)

[Komodakis 2007a] N. Komodakis and G. Tziritas. *Approximate Labeling via Graph Cuts Based on Linear Programming*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pages 1436–1453, aug 2007. (Cited on pages 19 and 106.)

[Komodakis 2007b] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Fast, approximately optimal solutions for single and dynamic MRFs*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007. (Cited on pages 21 and 64.)

[Komodakis 2008] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies*. Computer Vision and Image Understanding, vol. 112, no. 1, pages 14–29, oct 2008. (Cited on pages 21 and 94.)

[Komodakis 2009] N Komodakis and N Paragios. *Beyond pairwise energies: Efficient optimization for higher-order MRFs*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 2985–2992, 2009. (Cited on page 121.)

[Komodakis 2011] Nikos Komodakis, Nikos Paragios and Georgios Tziritas. *MRF energy minimization and beyond via dual decomposition*. IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 3, pages 531–52, mar 2011. (Cited on pages 12, 18 and 118.)

[Komodakis 2015] Nikos Komodakis, Bo Xiang and Nikos Paragios. *A Framework for Efficient Structured Max-Margin Learning of High-Order MRF Models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 7, pages 1425–1441, jul 2015. (Cited on pages 78, 118 and 121.)

[Krizhevsky 2012] A Krizhevsky, I Sutskever and G E Hinton. *Imagenet classification with deep convolutional neural networks.* In NIPS, 2012. (Cited on page 124.)

[Kschischang 2001] F.R. Kschischang, B.J. Frey and H.-A. Loeliger. *Factor graphs and the sum-product algorithm.* IEEE Transactions on Information Theory, vol. 47, no. 2, pages 498–519, 2001. (Cited on page 123.)

[Kuklisova-Murgasova 2012] Maria Kuklisova-Murgasova, Gerardine Quaghebeur, Mary A. Rutherford, Joseph V. Hajnal and Julia A. Schnabel. *Reconstruction of fetal brain MRI with intensity matching and complete outlier removal.* Medical Image Analysis, vol. 16, no. 8, pages 1550–1564, 2012. (Cited on page 29.)

[Kumar 2015] M. Pawan Kumar, Haithem Turki, Dan Preston and Daphne Koller. *Parameter Estimation and Energy Minimization for Region-Based Semantic Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 7, pages 1373–1386, jul 2015. (Cited on page 103.)

[Kwon 2008] Dongjin Kwon, Kyong Joon Lee, Il Dong Yun and Sang Uk Lee. *Nonrigid Image Registration Using Dynamic Higher-Order MRF Model.* In Computer Vision â ECCV 2008, pages 373–386. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. (Cited on pages 62 and 63.)

[Lasowski 2008] Ruxandra Lasowski, Selim Benhimane, Jakob Vogel, Tobias F. Jakobs, Christoph J. Zech, Christoph Trumm, Christian Clason, Nassir Navab, Medical Solutions and Computed Tomography. *Adaptive visualization for needle guidance in RF liver ablation: taking organ deformation into account.* Proceedings of SPIE, pages 69180A–69180A–12, 2008. (Cited on pages 42 and 49.)

[Le Moigne 2011] Jacqueline Le Moigne, Nathan S. Netanyahu and Roger D. Eastman. Image Registration for Remote Sensing. 2011. (Cited on page 3.)

[LeCun 1998] Y LeCun, L Bottou, Y Bengio and P Haffner. *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. (Cited on page 123.)

[Lee 2009] Daewon Lee, Matthias Hofmann, Florian Steinke, Yasemin Altun, Nathan D Cahill and Bernhard Scholkopf. *Learning similarity measure for multi-modal 3D image registration.* In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 186–193. IEEE, 2009. (Cited on pages 84 and 102.)

[Leung 2010] K. Y. Esther Leung, Marijn van Stralen, Attila Nemes, Marco M. Voormolen, Gerard van Burken, Marcel L. Geleijnse, Folkert J. Ten, Johan Cate and H. C. Reiber. *Sparse Registration for Three-Dimensional Stress*

Echocardiography. no. November 2015, 2010. (Cited on pages 28, 29, 43 and 46.)

[Liao 2013] Rui Liao, Li Zhang, Ying Sun, Shun Miao and Christophe Chefd'Hotel. *A Review of Recent Advances in Registration Techniques Applied to Minimally Invasive Therapy.* IEEE Transactions on Multimedia, vol. 15, no. 5, pages 983–1000, 2013. (Cited on pages 4 and 48.)

[Long 2014] Jonathan L. Long, Ning Zhang and Trevor Darrell. *Do Convnets Learn Correspondence?* In Advances in Neural Information Processing Systems. NIPS., pages 1601–1609, 2014. (Cited on pages 31 and 84.)

[Long 2015] Jonathan Long, Evan Shelhamer and Trevor Darrell. *Fully convolutional networks for semantic segmentation.* CVPR, 2015. (Cited on page 124.)

[Lorensen 1987] William E. Lorensen and Harvey E. Cline. *Marching cubes: A high resolution 3D surface construction algorithm.* ACM SIGGRAPH Computer Graphics, vol. 21, no. 4, pages 163–169, aug 1987. (Cited on page 31.)

[Lowe 1999] D.G. Lowe. *Object recognition from local scale-invariant features.* In Proceedings of the Seventh IEEE International Conference on Computer Vision, pages 1150–1157 vol.2. IEEE, 1999. (Cited on page 31.)

[Lucas 1981] Bruce D Lucas and Takeo Kanade. *An Iterative Image Registration Technique with an Application to Stereo Vision.* Imaging, vol. 130, no. x, pages 674–679, 1981. (Cited on page 3.)

[Maes 1997] Frederik Maes, Dirk Vandermeulen, Guy Marchal and Paul Suetens. *Multimodality Image Registration by Maximization of Mutual Information.* vol. 16, no. 2, pages 187–198, 1997. (Cited on page 28.)

[Mahapatra 2012] Dwarikanath Mahapatra and Ying Sun. *Integrating segmentation information for improved mrf-based elastic image registration.* IEEE Transactions on Image Processing, vol. 21, no. 1, pages 170–183, 2012. (Cited on pages 85 and 86.)

[Maintz 1998] J B Maintz and M a Viergever. *A survey of medical image registration.* Medical image analysis, vol. 2, no. 1, pages 1–36, 1998. (Cited on page 4.)

[Mallat 1999] Stéphane Mallat. A wavelet tour of signal processing. Academic press, 1999. (Cited on page 124.)

[Marami 2011] Bahram Marami, Shahin Sirouspour and David W Capson. *Model-based 3D/2D deformable registration of MR images.* Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology

Society. Conference, vol. 2011, no. November, pages 4880–3, 2011. (Cited on pages 28, 29, 39 and 41.)

[Markelj 2010] P Markelj, D Tomaževič, B Likar and F Pernuš. *A review of 3D/2D registration methods for image-guided interventions.* Medical Image Analysis, vol. 16, no. 3, pages 642–661, apr 2010. (Cited on pages 24 and 25.)

[Mercier 2012] Laurence Mercier, Rolando F. Del Maestro, Kevin Petrecca, David Araujo, Claire Haegelen and D. Louis Collins. *Online database of clinical MR and ultrasound images of brain tumors.* Medical Physics, vol. 39, no. 6, page 3253, 2012. (Cited on page 70.)

[Miao 2014a] Shun Miao, Rui Liao, Gerald Moran, John Butler, Li Pan and Z. Jane Wang. *Dynamic MR-based respiratory motion compensation for hybrid PET/MR system.* 2014 9th IEEE Conference on Industrial Electronics and Applications, pages 1915–1920, 2014. (Cited on pages 28, 29, 39, 41 and 46.)

[Miao 2014b] Shun Miao, Z Jane Wang, Rui Liao and B C Vt. *NON-PARAMETRIC ORTHOGONAL SLICE TO VOLUME DEFORMABLE REGISTRATION : APPLICATION TO PET / MR RESPIRATORY MOTION COMPENSATION Department of Electric and Computer Engineering , University of British Columbia , Imaging and Computer Vision , Siemens Corporatio.* vol. D, no. Mc, pages 530–534, 2014. (Cited on page 46.)

[Michel 2011] Fabrice Michel, Michael Bronstein, Alex Bronstein and Nikos Paragios. *Boosted metric learning for 3D multi-modal deformable registration.* In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, numéro 1, pages 1209–1214. IEEE, mar 2011. (Cited on page 84.)

[Michel 2014] Fabrice Michel. *Multi-Modal Similarity Learning for 3D Deformable Registration of Medical Images.* PhD thesis, Ecole Centrale Paris, 2014. (Cited on page 84.)

[Micu 2006] Ruxandra Micu, TF Jakobs, Martin Urschler and Nassir Navab. *A new registration/visualization paradigm for ct-fluoroscopy guided rf liver ablation.* Medical Image Computing and . . . , pages 882–890, 2006. (Cited on pages 38, 42, 43 and 49.)

[Mitra 2012a] J Mitra, S Ghose, D Sidibe, R Marti, A Oliver, X Llado, J C Vilanova, J Comet and F Meriaudeau. *Joint probability of shape and image similarities to retrieve 2D TRUS-MR slice correspondence for prostate biopsy.* In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 2012, pages 5416–5419. IEEE, aug 2012. (Cited on page 34.)

[Mitra 2012b] J Mitra, S Ghose, D Sidibe, A Oliver, R Marti, X Llado, J C Vilanova, J Comet and F Meriaudeau. *Weighted likelihood function of multiple statistical parameters to retrieve 2D TRUS-MR slice correspondence*

for prostate biopsy. In 2012 19th IEEE International Conference on Image Processing, pages 2949–2952. IEEE, sep 2012. (Cited on page 34.)

[Mohamed 2002] Ashraf Mohamed, Christos Davatzikos and Russell Taylor. *A Combined Statistical and Biomechanical Model for Estimation of Intra-operative Prostate Deformation.* volume 2489 of *Lecture Notes in Computer Science*, pages 452–460. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002. (Cited on page 83.)

[Moradi 2013] Mehdi Moradi and SS Mahdavi. *Ultrasound RF time series for tissue typing: first in vivo clinical results.* SPIE Medical . . . , vol. 8670, pages 1–8, 2013. (Cited on page 32.)

[Murphy 1999] Kevin P. Murphy, Yair Weiss and Michael I. Jordan. *Loopy Belief Propagation for Approximate Inference: An Empirical Study.* In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (UAI'99)„ 1999. (Cited on pages 19 and 20.)

[Museyko 2014] Oleg Museyko, Robert Percy Marshall, Jing Lu, Andreas Hess, Georg Schett, Michael Amling, Willi A Kalender, Klaus Engelke and Willi A. Kalendera & Klaus Engelkea Oleg Museykoa*, Robert Percy Marshallb, Jing Lua, Andreas Hessc, Georg Schettd, Michael Amlingb. *Registration of 2D histological sections with 3D micro-CT datasets from small animal vertebrae and tibiae.* Computer Methods in Biomechanics and Biomedical Engineering, vol. 5842, no. November, pages 1–16, 2014. (Cited on pages 29, 38 and 39.)

[Nelder 1965] J. A. Nelder and R. Mead. *A Simplex Method for Function Minimization.* The Computer Journal, vol. 7, no. 4, pages 308–313, jan 1965. (Cited on pages 42 and 66.)

[Nir 2011] Guy Nir and Allen Tannenbaum. *Temporal registration of partial data using particle filtering.* In IEEE International Conference on Image Processing, volume 17, pages 2177–2180. IEEE, sep 2011. (Cited on pages 29, 37 and 47.)

[Nir 2013] Guy Nir and Septimiu E. Salcudean. *Registration of whole-mount histology and tomography of the prostate using particle filtering.* page 86760E, mar 2013. (Cited on page 47.)

[Nir 2014] Guy Nir, Ramin S. Sahebjavaher, Piotr Kozlowski, Silvia D. Chang, Edward C. Jones, S. Larry Goldenberg and Septimiu E. Salcudean. *Registration of whole-mount histology and volumetric imaging of the prostate using particle filtering.* IEEE Transactions on Medical Imaging, vol. 33, no. 8, pages 1601–1613, 2014. (Cited on pages 29, 32, 38 and 47.)

[Noble 2005] Nicholas M I Noble, Redha Boubertakh, Reza S Razavi and Derek L G Hill. *Inter-breath-hold Registration for the Production of High Resolution*

Cardiac MR Volumes. In Medical image computing and computer-assisted intervention (MICCAI 2005), volume 8, pages 894–901, 2005. (Cited on pages 28 and 43.)

[Olesch 2011a] Janine Olesch. *Fast intra-operative non-linear registration of 3D-CT to tracked, selected 2D-ultrasound slices*. In SPIE Medical . . . , volume 7964, pages 3–8, 2011. (Cited on pages 33, 41 and 46.)

[Olesch 2011b] Janine Olesch and Bernd Fischer. *Focused Registration of Tracked 2D US to 3D CT Data of the Liver*. In Bildverarbeitung für die Medizin 2011, pages 79–83, 2011. (Cited on pages 33, 41 and 46.)

[Osechinskiy 2009] Sergey Osechinskiy and Frithjof Kruggel. *Quantitative comparison of high-resolution MRI and myelin-stained histology of the human cerebral cortex*. Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, pages 85–89, 2009. (Cited on page 28.)

[Osechinskiy 2011a] Sergey Osechinskiy and Frithjof Kruggel. *Deformable registration of histological sections to brain MR images using a hybrid boundary-based slice-to-volume approach*. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, vol. 2011, pages 4876–9, 2011. (Cited on page 39.)

[Osechinskiy 2011b] Sergey Osechinskiy and Frithjof Kruggel. *Slice-to-Volume Nonrigid Registration of Histological Sections to MR Images of the Human Brain*. Anatomy Research International, vol. 2011, pages 1–17, jan 2011. (Cited on pages 28, 39, 41, 43 and 45.)

[Ou 2011] Yangming Ou, Aristeidis Sotiras, Nikos Paragios and Christos Davatzikos. *DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting*. Medical Image Analysis, vol. 15, no. 4, pages 622–639, aug 2011. (Cited on page 85.)

[Paragios 2003] Nikos Paragios, Mikael Rousson and Visvanathan Ramesh. *Non-rigid registration using distance functions*. Computer Vision and Image Understanding, vol. 89, no. 2-3, pages 142–165, feb 2003. (Cited on page 32.)

[Paragios 2005] Nikos Paragios and Rachid Deriche. *Geodesic active regions and level set methods for motion estimation and tracking*. Computer Vision and Image Understanding, vol. 97, no. 3, pages 259–282, mar 2005. (Cited on page 86.)

[Paragios 2014] Nikos Paragios and Nikos Komodakis. *Discrete Visual Perception*. In 2014 22nd International Conference on Pattern Recognition, volume 2, pages 18–25. IEEE, aug 2014. (Cited on pages 4 and 10.)

[Parisot 2013] Sarah Parisot, William Wells, Stephane Chemouny, Hugues Duffau and Nikos Paragios. *Uncertainty-driven efficiently-sampled sparse graphical models for concurrent tumor segmentation and atlas registration.* Proceedings of the IEEE International Conference on Computer Vision, no. i, pages 641–648, 2013. (Cited on page 5.)

[Parisot 2014] Sarah Parisot, William Wells, Stephane Chemouny, Hugues Duffau and Nikos Paragios. *Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs.* Medical Image Analysis, vol. 18, no. 4, pages 647–659, 2014. (Cited on page 86.)

[Park 2004] Hyunjin Park, Charles R Meyer and Boklye Kim. *Improved Motion Correction in fMRI by Joint Mapping of Slices into an Anatomical Volume.* Proceedings of the 7th International Conference on Medical Image Computing and Computer Aided Intervention (MICCAI 2004), no. 3217, pages 745–751, 2004. (Cited on pages 29, 43 and 50.)

[Penney 2004] G.P. Penney, J.M. Blackall, M.S. Hamady, T. Sabharwal, A. Adam and D.J. Hawkes. *Registration of freehand 3D ultrasound and magnetic resonance liver images.* Medical Image Analysis, vol. 8, no. 1, pages 81–91, 2004. (Cited on pages 34, 35, 37 and 42.)

[Penney 2006] G P Penney, D C Barratt, C S K Chan, M Slomczykowski, T J Carter, P J Edwards and D J Hawkes. *Cadaver validation of intensity-based ultrasound to CT registration.* Medical image analysis, vol. 10, no. 3, pages 385–95, jun 2006. (Cited on pages 34, 35 and 42.)

[Po Su 2013] Po Su, Jianhua Yang, Kongkuo Lu, Nam Yu, Stephen T Wong and Zhong Xue. *A Fast CT and CT-Fluoroscopy Registration Algorithm With Respiratory Motion Compensation for Image-Guided Lung Intervention.* IEEE Transactions on Biomedical Engineering, vol. 60, no. 7, pages 2034–2041, jul 2013. (Cited on pages 28, 29 and 39.)

[Prasoon 2013] A Prasoon, K Petersen, C Igel, F Lauze, E Dam and M Nielsen. *Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network.* In MICCAI. 2013. (Cited on page 124.)

[Ramalingam 2008] Srikumar Ramalingam, Pushmeet Kohli, Karteek Alahari and Philip H. S. Torr. *Exact inference in multi-label CRFs with higher order cliques.* In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, jun 2008. (Cited on page 64.)

[Rivaz 2014a] Hassan Rivaz, Sean Chen and D Louis Collins. *Automatic Deformable MR-Ultrasound Registration for Image-Guided Neurosurgery.* IEEE transactions on medical imaging, vol. 0062, no. c, pages 1–15, sep 2014. (Cited on pages 30 and 39.)

[Rivaz 2014b] Hassan Rivaz, Zahra Karimaghaloo and D. Louis Collins. *Self-similarity weighted mutual information: A new nonrigid image registration metric.* Medical Image Analysis, vol. 18, no. 2, pages 343–358, 2014. (Cited on pages 30 and 39.)

[Rivaz 2014c] Hassan Rivaz, Zahra Karimaghaloo, Vladimir S Fonov and D Louis Collins. *Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information.* IEEE Transactions on Medical Imaging, vol. 33, no. 3, pages 708–725, mar 2014. (Cited on pages 30 and 39.)

[Roche 1998] Alexis Roche, Grégoire Malandain, Xavier Pennec and Nicholas Ayache. *The correlation ratio as a new similarity measure for multimodal image registration.* pages 1115–1124. 1998. (Cited on page 28.)

[Rohlfing 2004] Torsten Rohlfing, Robert Brandt, Randolf Menzel and Calvin R Maurer Jr. *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains.* NeuroImage, vol. 21, no. 4, pages 1428–1442, 2004. (Cited on page 88.)

[Rohlfing 2012] T Rohlfing. *Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable.* IEEE Transactions on Medical Imaging, 2012. (Cited on page 95.)

[Rother 2007] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky and Martin Szummer. *Optimizing Binary MRFs via Extended Roof Duality.* In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, jun 2007. (Cited on page 18.)

[Rousseau 2005] F Rousseau and O Glenn. *A novel approach to high resolution fetal brain MR imaging.* Medical Image Computing and Computer-Assisted Intervention â MICCAI 2005, vol. 3749, pages 548–555, 2005. (Cited on pages 29, 37, 41 and 47.)

[Rousseau 2006] Francois Rousseau, Orit A Glenn, Bistra Iordanova, Claudia Rodriguez-Carranza, Daniel B Vigneron, James A Barkovich and Colin Studholme. *Registration-Based Approach for Reconstruction of High-Resolution In Utero Fetal MR Brain Images.* Academic Radiology, vol. 13, no. 9, pages 1072–1081, sep 2006. (Cited on pages 4, 29, 37, 41, 47 and 50.)

[Rueckert 1999] D Rueckert, L I Sonoda, C Hayes, D L Hill, M O Leach and D J Hawkes. *Nonrigid registration using free-form deformations: application to breast MR images.* IEEE transactions on medical imaging, vol. 18, no. 8, pages 712–21, aug 1999. (Cited on pages 15, 39, 52, 93 and 106.)

[Sabuncu 2010] M R Sabuncu, B T T Yeo, K Van Leemput, B Fischl and P Golland. *A Generative Model for Image Segmentation Based on Label Fusion.* Medical Imaging, IEEE Transactions on, vol. 29, no. 10, pages 1714–1729, 2010. (Cited on page 88.)

[San José Estépar 2009] Raúl San José Estépar, Carl-Fredrik Westin and Kirby G Vosburgh. *Towards real time 2D to 3D registration for ultrasound-guided endoscopic and laparoscopic procedures.* International Journal of Computer Assisted Radiology and Surgery, vol. 4, no. 6, pages 549–560, nov 2009. (Cited on pages 35, 37 and 49.)

[Schulz 2014] J. Schulz, T. Siegert, P.-L. Bazin, J. Maclaren, M. Herbst, M. Zaitsev and R. Turner. *Prospective slice-by-slice motion correction reduces false positive activations in fMRI with task-correlated motion.* NeuroImage, vol. 84, pages 124–132, 2014. (Cited on pages 33 and 34.)

[Sederberg 1986] Thomas W. Sederberg and Scott R. Parry. *Free-form deformation of solid geometric models.* ACM SIGGRAPH Computer Graphics, vol. 20, no. 4, pages 151–160, aug 1986. (Cited on page 39.)

[Seshamani 2013] S. Seshamani, M. Fogtmann, X. Cheng, M. Thomason, C. Gatenby and C. Studholme. *Cascaded slice to volume registration for moving fetal FMRI.* In 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI), pages 796–799. IEEE, apr 2013. (Cited on pages 28, 29 and 49.)

[Shakeri 2016] Mahsa Shakeri, Stavros Tsogkas, Enzo Ferrante, Sarah Lippe, Samuel Kadoury, Nikos Paragios, Iasonas Kokkinos, Universite Paris-saclay and Polytechnique Montreal. *Sub-cortical brain structure segmentation using F-CN.* In ISBI 2016, 2016. (Cited on page 123.)

[Shekhovtsov 2008] Alexander Shekhovtsov, Ivan Kovtun and Václav Hlaváč. *Efficient MRF deformation model for non-rigid image matching.* Computer Vision and Image Understanding, vol. 112, no. 1, pages 91–99, oct 2008. (Cited on pages 6, 44, 52 and 58.)

[Shimony 1994] Solomon Eyal Shimony. *Finding MAPs for belief networks is NP-hard.* Artificial Intelligence, vol. 2, 1994. (Cited on page 18.)

[Smolíková-Wachowiak 2005] Renata Smolíková-Wachowiak, Mark P Wachowiak, Aaron Fenster and Maria Drangova. *Registration of two-dimensional cardiac images to preprocedural three-dimensional images for interventional applications.* Journal of Magnetic Resonance Imaging, vol. 22, no. 2, pages 219–228, aug 2005. (Cited on pages 28, 29, 37 and 43.)

[Sotiras 2009] Aristeidis Sotiras, Nikos Komodakis, Ben Glocker, Jean-François Deux and Nikos Paragios. *Graphical Models and Deformable Diffeomorphic Population Registration Using Global and Local Metrics.* Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009, pages 672–679, 2009. (Cited on pages 5, 88 and 92.)

[Sotiras 2010a] Aristeidis Sotiras, Yangming Ou, Ben Glocker, Christos Davatzikos and Nikos Paragios. *Simultaneous geometric–iconic registration.* Medical

image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13, no. Pt 2, pages 676–683, 2010. (Cited on page 5.)

[Sotiras 2010b] Aristeidis Sotiras, Yangming Ou, Ben Glocker, Christos Davatzikos and Nikos Paragios. *Simultaneous geometric–iconic registration.* Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13, no. Pt 2, pages 676–683, 2010. (Cited on pages 34, 78 and 121.)

[Sotiras 2013] Aristeidis Sotiras, Christos Davatazikos and Nikos Paragios. *Deformable Medical Image Registration: A Survey.* IEEE Transactions on Medical Imaging, vol. 32, pages 1153–1190, 2013. (Cited on pages 4, 16, 25, 38, 82, 83 and 88.)

[Studholme 1999] C. Studholme, D.L.G. Hill and D.J. Hawkes. *An overlap invariant entropy measure of 3D medical image alignment.* Pattern Recognition, vol. 32, no. 1, pages 71–86, jan 1999. (Cited on page 29.)

[Sun 2007] Yiyong Sun, Samuel Kadoury, Yong Li, Matthias John, Jeff Resnick, Gerry Plambeck, Rui Liao, Frank Sauer and Chenyang Xu. *Image guidance of intracardiac ultrasound with fusion of pre-operative images.* International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, vol. 10, no. Pt 1, pages 60–67, 2007. (Cited on pages 35 and 42.)

[Szeliski 2006] Richard Szeliski. *Image Alignment and Stitching: A Tutorial.* Foundations and TrendsÂ® in Computer Graphics and Vision, vol. 2, no. 1, pages 1–104, 2006. (Cited on page 3.)

[Tadayyon 2010a] Hadi Tadayyon. *Target Motion Compensation in MRI-guided Prostate Biopsy with Static Images.* In 2010 Annual International Conference of the IEEE EMBC, volume 18, pages 1199–1216, 2010. (Cited on pages 28, 29, 39, 41 and 44.)

[Tadayyon 2010b] Hadi Tadayyon, Siddharth Vikal, Sean Gill, Andras Lasso and Gabor Fichtinger. *MRI-guided prostate motion tracking by means of multislice-to- volume registration.* vol. 18, no. 9, pages 1199–1216, 2010. (Cited on page 28.)

[Tadayyon 2011] Hadi Tadayyon, Andras Lasso, Aradhana Kaushal, Peter Guion and Gabor Fichtinger. *Target motion tracking in MRI-guided transrectal robotic prostate biopsy.* IEEE transactions on bio-medical engineering, vol. 58, no. 11, pages 3135–42, 2011. (Cited on pages 39, 41, 44 and 46.)

[Tang 2009] Songyuan Tang, Yong Fan, Guorong Wu, Minjeong Kim and Dinggang Shen. *RABBIT: Rapid alignment of brains by building intermediate*

templates. NeuroImage, vol. 47, no. 4, pages 1277–1287, oct 2009. (Cited on page 82.)

[Taron 2009] M. Taron, N. Paragios and M.-P. Jolly. *Registration with Uncertainties and Statistical Modeling of Shapes with Variable Metric Kernels*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pages 99–113, jan 2009. (Cited on page 32.)

[Taskar 2003] B Taskar, C Guestrin and D Koller. *Max-margin {M}arkov networks*. In NIPS, 2003. (Cited on pages 103 and 106.)

[Toga 2008] Arthur W. Toga. *Learning based coarse-to-fine image registration*. 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–7, 2008. (Cited on page 84.)

[Tsochantaridis 2004] I Tsochantaridis, T Hofmann, T Joachims and Y Altun. *Support vector machine learning for interdependent and structured output spaces*. In ICML, 2004. (Cited on pages 103 and 106.)

[Vakalopoulou 2015] Maria Vakalopoulou, Konstantinos Karatzalos, Nikos Komodakis and Nikos Paragios. *Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data*. In 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 61–69. IEEE, jun 2015. (Cited on page 5.)

[Vedaldi 2014] A Vedaldi and K Lenc. *MatConvNet-convolutional neural networks for MATLAB*. arXiv preprint arXiv:1412.4564, 2014. (Cited on page 97.)

[Wachinger 2010] C Wachinger and N Navab. *Manifold Learning for Multi-Modal Image Registration*. Proc. BMVC, pages 82.1—-82.12, 2010. (Cited on page 85.)

[Wainwright 2005] M.J. Wainwright, T.S. Jaakkola and A.S. Willsky. *MAP Estimation Via Agreement on Trees: Message-Passing and Linear Programming*. IEEE Transactions on Information Theory, vol. 51, no. 11, pages 3697–3717, nov 2005. (Cited on page 19.)

[Wang 2000] Yongmei Wang and Lawrence H. Staib. *Physical model-based non-rigid registration incorporating statistical shape information*. Medical Image Analysis, vol. 4, no. 1, pages 7–20, mar 2000. (Cited on page 82.)

[Wang 2005] Fei Wang and Baba C Vemuri. *Simultaneous registration and segmentation of anatomical structures from brain MRI*. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 8, no. Pt 1, pages 17–25, 2005. (Cited on page 86.)

[Wang 2013] Chaohui Wang, Nikos Komodakis and Nikos Paragios. *Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey.* Computer Vision and Image Understanding, vol. 117, no. 11, pages 1610–1627, nov 2013. (Cited on pages 4, 10, 22 and 43.)

[Wein 2007] Wolfgang Wein, Ali Khamene, Dirk-André Clevert, Oliver Kutter and Nassir Navab. *Simulation and Fully Automatic Multimodal Registration of Medical Ultrasound.* In Medical Image Computing and Computer-Assisted Intervention â MICCAI 2007, pages 136–143. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. (Cited on pages 30 and 83.)

[Wein 2008a] Wolfgang Wein, Shelby Brunke, Ali Khamene, Matthew R Callstrom and Nassir Navab. *Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention.* Medical Image Analysis, vol. 12, no. 5, pages 577–585, oct 2008. (Cited on pages 30, 36, 38, 43, 46 and 83.)

[Wein 2008b] Wolfgang Wein, Jz Cheng and Ali Khamene. *Ultrasound based respiratory motion compensation in the abdomen.* MICCAI 2008 Worshop on Image Guidance and Computer Assistance for Softissue Intervetions, vol. 32, no. 6, page 294, 2008. (Cited on pages 36, 38 and 43.)

[Wouters 2006] Jeroen Wouters, Emiliano D'Agostino, Frederik Maes, Dirk Vandermeulen and Paul Suetens. *Non-rigid brain image registration using a statistical deformation model.* pages 614411–614411–8, mar 2006. (Cited on page 82.)

[Wu 2013] Guorong Wu, Minjeong Kim, Qian Wang, Yaozong Gao, Shu Liao and Dinggang Shen. *Unsupervised Deep Feature Learning for Deformable Registration of MR Brain Images.* In MICCAI 2013, volume 72, pages 649–656, 2013. (Cited on page 85.)

[Wyatt 2003] Paul P. Wyatt and J. Alison Noble. *MAP MRF joint segmentation and registration of medical images.* Medical Image Analysis, vol. 7, no. 4, pages 539–552, 2003. (Cited on page 85.)

[Xiao 2011] Gaoyu Xiao, B. Nicolas Bloch, Jonathan Chappelow, Elizabeth M. Genega, Neil M. Rofsky, Robert E. Lenkinski, John Tomaszewski, Michael D. Feldman, Mark Rosen and Anant Madabhushi. *Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer.* Computerized Medical Imaging and Graphics, vol. 35, no. 7-8, pages 568–578, 2011. (Cited on pages 29, 37, 38 and 43.)

[Xiaolei Huang 2006] Xiaolei Huang, N. Paragios and D.N. Metaxas. *Shape registration in implicit spaces using information theory and free form deformations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 8, pages 1303–1318, aug 2006. (Cited on page 32.)

[Xu 2008] Sheng Xu, Jochen Kruecker, Baris Turkbey, Neil Glossop, Anurag K Singh, Peter Choyke, Peter Pinto and Bradford J Wood. *Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies.* Computer Aided Surgery, vol. 13, no. 5, pages 255–264, jan 2008. (Cited on pages 28, 30, 33 and 35.)

[Xu 2013] Lijian Xu, Jun Liu, Weiwei Zhan and Lixu Gu. *A novel algorithm for CT-ultrasound registration.* 2013 IEEE Point-of-Care Healthcare Technologies (PHT), pages 101–104, jan 2013. (Cited on pages 4 and 48.)

[Xu 2014a] Helen Xu, Andras Lasso, Andriy Fedorov, Kemal Tuncali, Clare Tempany and Gabor Fichtinger. *Multi-slice-to-volume registration for MRI-guided transperineal prostate biopsy.* International Journal of Computer Assisted Radiology and Surgery, vol. 10, no. 5, pages 563–572, 2014. (Cited on pages 4, 29, 37, 39, 41, 45, 46 and 48.)

[Xu 2014b] Robert Xu, Prashant Athavale, Adrian Nachman and Graham A Wright. *Multiscale Registration of Real-Time and Prior MRI Data for Image-Guided Cardiac Interventions.* IEEE Transactions on Biomedical Engineering, vol. 61, no. 10, pages 2621–2632, 2014. (Cited on pages 29, 37 and 43.)

[Yan 2012] Charles X B Yan, BenoÇt Goulet, Donatella Tampieri and D. Louis Collins. *Ultrasound-CT registration of vertebrae without reconstruction.* International Journal of Computer Assisted Radiology and Surgery, vol. 7, no. 6, pages 901–909, 2012. (Cited on pages 28, 35 and 45.)

[Yang 2006] Liu Yang and R Jin. *Distance metric learning: A comprehensive survey.* Michigan State Universiy, pages 1–51, 2006. (Cited on pages 83 and 84.)

[Yavariabdi 2015] Amir Yavariabdi, Adrien Bartoli, Chafik Samir, Maxime Artigues and Michel Canis. *Mapping and Characterizing Endometrial Implants by Registering 2D Transvaginal Ultrasound to 3D Pelvic Magnetic Resonance Images.* Computerized Medical Imaging and Graphics, vol. 45, pages 11–25, oct 2015. (Cited on pages 32 and 45.)

[Yeo 2004] Desmond T.B. Yeo, Jeffery A. Fessler and Boklye Kim. *Motion Correction in fMRI by Mapping Slice-to-Volume with Concurrent Field-Inhomogeneity Correction.* In Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004), pages 752–760, 2004. (Cited on pages 29, 37, 43 and 50.)

[Yeo 2006] Desmond T B Yeo, Roshni R Bhagalia and Boklye Kim. *Improved map-slice-to-volume motion correction with B0 inhomogeneity correction: validation of activation detection algorithms using ROC curve analyses.* Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 9, no. Pt 2, pages 276–83, 2006. (Cited on pages 43 and 50.)

[Yeo 2008] Desmond T B Yeo, Jeffrey a Fessler and Boklye Kim. *Concurrent correction of geometric distortion and motion using the map-slice-to-volume method in echo-planar imaging.* Magnetic resonance imaging, vol. 26, no. 5, pages 703–14, 2008. (Cited on page 50.)

[Yezzi 2003] A. Yezzi, L. Zöllei and T. Kapur. *A variational framework for integrating segmentation and registration through active contours.* Medical Image Analysis, vol. 7, no. 2, pages 171–185, jun 2003. (Cited on page 86.)

[Yu 2008] Yong Yu, Alain Trouv and Bernard Chalmond. *A Fast 3D Volume Reconstruction for Confocal Micro-rotation Cell Imaging.* 2008 International Conference on BioMedical Engineering and Informatics, pages 775–778, 2008. (Cited on pages 28 and 46.)

[Yu 2009] Y. Yu, A. Trouvé, B. Chalmond, O. Renaud and S. L. Shorte. *Confocal bi-protocol: A new strategy for isotropic 3D live cell imaging.* Journal of Microscopy, vol. 242, no. 1, pages 70–85, 2009. (Cited on pages 7, 103 and 106.)

[Yu 2011] Y. Yu, a. Trouvé, B. Chalmond, O. Renaud and S. L. Shorte. *Confocal bi-protocol: A new strategy for isotropic 3D live cell imaging.* Journal of Microscopy, vol. 242, no. 1, pages 70–85, 2011. (Cited on pages 28, 37, 41, 42, 46 and 121.)

[Yuille 2003] A Yuille and A Rangarajan. *The concave-convex procedure.* Neural Computation, 2003. (Cited on pages 104 and 109.)

[Zakkaroff 2012] Constantine Zakkaroff, Aleksandra Radjenovic, John Greenwood and Derek Magee. *Recovery of Slice Rotations with the Stack Alignment Transform in Cardiac MR Series.* In Procedings of the British Machine Vision Conference 2012, pages 35.1–35.11. British Machine Vision Association, 2012. (Cited on pages 29, 37 and 41.)

[Zarow 2004] C. Zarow, T. S. Kim, M. Singh and H. C. Chui. *A standardized method for brain-cutting suitable for both stereology and MRI-brain co-registration.* Journal of Neuroscience Methods, vol. 139, pages 209–215, 2004. (Cited on pages 28 and 29.)

[Zikic 2010a] Darko Zikic, Ben Glocker, Oliver Kutter, Martin Groher, Nikos Komodakis, Ali Kamen, Nikos Paragios and Nassir Navab. *Linear intensity-based image registration by Markov random fields and discrete optimization.* Medical image analysis, vol. 14, no. 4, pages 550–62, aug 2010. (Cited on page 40.)

[Zikic 2010b] Darko Zikic, Ben Glocker, Oliver Kutter, Martin Groher, Nikos Komodakis, Ali Khamene, Nikos Paragios and Nassir Navab. *Markov random field optimization for intensity-based 2D-3D registration.* In SPIE Medical

Imaging. International Society for Optics and Photonics, 2010. (Cited on pages 40, 79 and 122.)

[Zitová 2003] Barbara Zitová and Jan Flusser. *Image registration methods: A survey.* Image and Vision Computing, vol. 21, no. 11, pages 977–1000, 2003. (Cited on page 3.)