Registration/Fusion of multimodal images using higher order graphs

Vivien Fécamp

NNT : 2016SACLC005

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À
L'ÉCOLE CENTRALESUPÉLEC

ÉCOLE DOCTORALE N° 580
Sciences et Technologies de l'Information et de la Communication

Spécialité de doctorat Mathématiques et Informatiques

Par

M. Vivien Fécamp

Fusion/Recalage d'images multimodales à l'aide de graphes d'ordres supérieurs

Thèse présentée et soutenue à l'École CentraleSupélec, Chatenay-Malabry, le 12/01/2016 :

Composition du Jury :

M. Thirion Bertrand	Professeur INRIA (Neurospin, PARIETAL)	Président
M. Delingette Hervé	Professeur INRIA, Épidaure	Rapporteur
M. Najman Laurent	Professeur, Université Paris Est	Rapporteur
M. Sotiras Aristeidis	Chercheur, University of Pennsylvania	Examinateur
M. Glocker Benjamin	Conférencier, Imperial College, London	Examinateur
M. Nikos Paragios	Professeur, École CentraleSupélec	Directeur de thèse

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

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

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

Titre : Fusion/Recalage d'images multimodales à l'aide de graphes d'ordres supérieurs

Mots clés : Champs de Markov Aléatoires, Recalage, Décomposition Duale, Alternating Direction Method of Multipliers

Résumé : L'objectif principal de cette thèse est l'exploration du recalage d'images à l'aide de champs aléatoires de Markov d'ordres supérieurs, et plus spéciquement d'intégrer la connaissance de transformations globales comme une transformation rigide, dans la structure du graphe. Notre cadre principal s'applique au recalage 2D-2D ou 3D-3D et utilise une approche hiérarchique d'un modèle de champ de Markov dont le graphe est une grille régulière. Les variables cachées sont les vecteurs de déplacements des points de contrôle de la grille.

Tout d'abord nous expliquerons la construction du graphe qui permet de recaler des images en cherchant entre elles une transformation ane, rigide, ou une similarité, tout en ne changeant qu'un potentiel sur l'ensemble du graphe, ce qui assure une exibilité lors du recalage. Le choix de la métrique est également laissée á l'utilisateur et ne modie pas le fonctionnement de notre algorithme. La diculté fut d'encoder la connaissance globale de la transformation linéaire dans des cliques ou hyperarêtes, qui sont par nature locales. En travaillant sur les propriétés de ces transformations et en utilisant la structure régulière de la grille de points de contrôle, nous avons réussi à dénir des contraintes simples et locales, qui si elles sont vériées sur l'ensemble des cliques, assurent que la transformation trouvée appartient au champ de recherche souhaité. Ces cliques sont composés de trois ou quatre points.

L'optimisation est un problème dans notre cas et les algorithmes classiques ne sont pas adaptés. Nous avons eu recours à une méthode de décomposition qui coupe le problème principal en sous-problèmes gérables et essaie d'accorder les sous-solutions trouvées. La décomposition duale permet donc de gérer les hyper-arêtes du graphe et garantit l'obtention du minimum exact de la fonction pourvu que l'on ait un accord entre les esclaves. Un graphe similaire est utilisé pour réaliser du recalage 2D-3D.

Ensuite, nous fusionnons le graphe précédent avec un autre graphe construit pour réaliser le recalage déformable. Le graphe résultant de cette fusion est plus complexe et, an d'obtenir un résultat en un temps raisonnable, nous utilisons une méthode d'optimisation appelée ADMM (Alternating Direction Method of Multipliers) qui a pour but d'accélérer la convergence de la décomposition duale. Nous pouvons alors résoudre simultanément recalage ane et déformable, ce qui nous débarrasse du biais potentiel issu de l'approche classique qui consiste à recaler anement puis de manière déformable. Il existe d'autres manières de fusionner les cliques du précédent graphe avec un modèle classique de recalage déformable. Nous donnons un exemple permettant de recaler de manière déformable une image tout en déformant une sous-partie de l'image de manière rigide.

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

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

Title : Registration/Fusion of Multimodal Images using Higher Order Markov Random Fields

Keywords : Markov Random Fields, Registration, Dual Decomposition, Alternating Direction Method of Multipliers

Abstract : The main objective of this thesis is the exploration of higher order Markov Random Fields for image registration, specifically to encode the knowledge of global transformations, like rigid transformations, into the graph structure. Our main framework applies to 2D/2D or 3D/3D registration and use a hierarchical grid-based Markov Random Field model where the hidden variables are the displacements vectors of the control points of the grid.

We first present the construction of a graph that allows to perform linear registration, which means here that we can perform affine registration, rigid registration, or similarity registration with the same graph while changing only one potential. Our framework is thus modular regarding the sought transformation and the metric used. Inference is performed with Dual Decomposition, which allows to handle the higher order hyperedges and which ensures the global optimum of the function is reached if we have an agreement among the slaves. A similar structure is also used to perform 2D/3D registration.

Second, we fuse our former graph with another structure able to perform deformable registration. The resulting graph is more complex and another optimisation algorithm, called Alternating Direction Method of Multipliers is needed to obtain a better solution within reasonable time. It is an improvement of Dual Decomposition which speeds up the convergence. This framework is able to solve simultaneously both linear and deformable registration which allows to remove a potential bias created by the standard approach of consecutive registrations.

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

Remerciements

Mes remerciements se dirigent en premier lieu vers mon directeur de thèse Nikos Paragios et Aristeidis Sotiras, mon second encadrant caché. Je ne saurais exprimer pleinement ma gratitude pour leur encadrement de qualité. Leur pédagogie, leur acuité scientifique et leur passion communicative n'ont cessé de m'impressionner, de m'inspirer et de me donner goût à la recherche pendant ces trois années. Je me souviendrai de ces rendez-vous sur Skype à essayer laborieusement de s'échanger nos idées.

En second lieu je souhaite exprimer ma profonde gratitude aux membres de mon comité de thèse : MM. Bertrand Thirion et Benjamin Glocker et plus spécifiquement aux rapporteurs, MM. Hervé Delingette et Laurent Najman, pour le temps passé à lire et évaluer mon travail ainsi qu'à leurs pertinentes remarques et questions.

Je souhaite remercier Samuel Kadoury de m'avoir accueilli dans son laboratoire à l'École polytechnique de Montréal ainsi que les membres du laboratoire pour leur accueil chaleureux et les bons moments passés ensemble, particulièrement William Thong et Julien Couet.

Merci à Enzo Ferrante pour sa bonne humeur et son épaule dans les moments difficiles, merci à Stavros Alchatzidis pour ses discussions théoriques, à Rafael Marini Silva pour son aide technique, à Siddharta Chandra, mon meilleur co-bureau et à tous les membres du laboratoire avec qui j'ai passé de bons moments.

Enfin je dédie cette thèse à mon grand-père qui n'est plus là pour me féliciter.

Acknowledgements

My first acknowledgements head to my advisor Nikos Paragios and Aristeidis Sotiras, who acted as a second advisor for me. I cannot express my gratitude for the quality of their work with me. Their pedagogy, their scientific knowledge and their communicative passion never ceased to impress me, to inspire me and to teach me what research should be during those three years. I will remember those hangouts, trying laboriously to exchange our ideas.

Second, I would like to express my deepest gratitude towards the members of my thesis committee: the reviewers Dr. Hervé Delingette and Dr. Laurent Najman for the time spent to read, evaluate my work and their pertinent remarks and questions. I thank all the members of the jury, Dr. Bertrand Thirion the president of the jury, and Dr. Benjamin Glocker for their time and remarks.

I would like to thank Samuel Kadoury to have welcome me in his laboratory at the École polytechnique de Montréal as well as the members of his laboratory for their warm welcome and the good moments spent together, particularly William Thong and Julien Couet.

I thank Enzo Ferrante for his good mood and his shoulder in the difficult moments, Stavros Alchatzidis for our theoritical discussions, Rafael Marini Silva for his technical help, Siddartha Chandra, my best office mate and all the members of the Centre for Visual Computing with whom I spent enjoyable times.

Finally I dedicate this thesis to my grandfather who is not here to congratulate me.

Notations

Manuscript Conventions

G	Graph
V	Set of nodes
E	Set of edges
H	Set of hyperedges
L	Set of labels

Mathematical Acronyms

ADMM	Alternating Decomposition by Method of Multipliers
DD	Dual Decomposition
Fast-PD	Fast Primal Dual algorithm
FFD	Free Form Deformation
MRF	Markov Random Field

Medical Abbreviations

CT	Computed Tomography
LC2	Linear Correlation of Linear Combination
MI	Mutual Information
MRI	Magnetic Resonance Imaging
NMI	Normalized Mutual Information
PET	Positron Emission Tomography
SAD	Sum of Absolute Differences
SSD	Sum of Squared Differences
US	Ultrasound

Contents

Chapter 1

Introduction

This thesis lies at the intersection of Computer Vision and Medical Image Analysis. The work itself and the algorithms presented here are not specific to medical images but the tests and the examples come from the medical world. The main reason is that, despite important advances, many algorithms which work for natural images are not adapted to medical images, therefore registration is still a challenging problem. Registration is the problem of aligning two images in the same space. One can seek different kinds of registration between images by allowing different sets of transformations: affine transformations, rigid, similarities, projective, deformable. In this work, we will regroup affine, rigid, and similarities under the name linear registration. A similarity measure is used to quantify the registration quality. The choice of the similarity measure is important and depends on the modality of the images. An optimization procedure is performed to reach the minimum (or at least a local minimum) of the objective function, in either a discrete or continuous setting.

In this introductory chapter, in section 1.1 we give a brief description of the context of this work. Next, in section 1.2, we present the main ideas of this work. We conclude this chapter with an overview of the work done during this thesis in section 1.3.

1.1 Thesis Clinical Context

Medical Image Analysis is a booming research area because of the increasing number of patients and images. For example, the number of cancer patients and cancer mortality has increased by 10% between 2008 and 2012. Those patients need a regular follow-up of the evolution of their disease, which means more exams and more images to analyse. Despite this real need and the huge amount of algorithms and papers in this research area, most of the medical doctors do not use automatic softwares and rely on their own skills. Indeed, the results of the algorithms should reach an incredible precision and robustness to be considered and used on a daily basis by doctors, while at the same time the quality of the images may vary a lot due to technical or human mistakes. The three main areas of research in Medical Image Analysis are registration, segmentation, and physiological modeling.

Registration means aligning two images. It allows to study temporal changes like a disease progression. It can be either long-term (scoliosis, tumour) or short-term, (breathing, during surgery). It is also paramount when we need to fuse different informations from different modalities like anatomical and functional information. Registering many images from different patients is required to construct an atlas to understand the variability among a population.

Segmentation means finding the contour of an object, like an organ that we want to study. For example, the segmentation of the left ventricle of the heart after and just before a beat gives the stroke volume. Low contrast, noise, and artefacts due to the acquisition are common difficulties encountered. Image-based segmentation relies mainly on the value of the voxels of the image, shape-based is initialized with a template, and atlas-based uses a set of training segmented images for guidance. Atlas-based and shape-based methods need an initial good registration.

Physiological modeling means building a simplified version of an object in order to see how it behaves or interacts with other objects. The interaction we seek to model may be biomechanical or physiological. Instead of just following a disease progression, one could predict its evolution and thus adapt the therapy. The aim is to better understand human physiology and have a personalized medicine.

We understand now that registration is in the front end of many applications: fusion, intervention and treatment planning [49], computer-aided diagnosis [63], surgery simulation [74], atlas building [76], radiation therapy [151], anatomy segmentation [160] and so on and so forth. It is used for all the body parts: brain [71], eyes [121], lung [24], breast [58], abdomen [54], prostate, entire body [132], cervical, wrist, vascular structures [140], bones, knee [30] and spine [14].

Some reviews can be found on medical image registration [4,39,55,61,139], on softwares for multimodal images [136], or on some specific anatomic parts like the heart [101] and the brain [159], or on similarity measures [114].

Several free softwares are available on the Internet: ITK [66], Insight Segmentation and Registration Toolkit, a well-known C++ library, 3D Slicer for image analysis and visualisation [35, 50, 113], Elastix [81] both 3D Slicer and Elastix are based on ITK), FLIRT [73] has a source code in C++, is part of FSL and only available on Linux, ANTS [149, 150] the Advanced Normalization Tools for image registration with different transformation models and similarity measures. Comparison are available in [80, 136, 161].

To be used in so different applications, we need a highly versatile registration algorithm: it should be fast to be used during surgical operation and adapt to any type of modalities.

1.1.1 Image Modalities

Since the discovery of X-rays in 1895, medical doctors rely more and more on medical imaging to understand the structure and the activity of the organs they cannot directly see. These are the two main purposes of medical imaging. The structural modalities allow to visualize the positioning of the different organs or tissues while the functional modalities aim to understand the activity of an organ through physiological changes.

X-rays
X-rays are a kind high frequency electromagnetic radiation. Bones and teeth absorb X-ray efficiently and are then clearly visible on a radiography. A radiography is a very common exam, the image is obtained on an X-ray detector of a patient illuminated with a short X-ray pulse. The absorption of X-rays is different as they pass through the different tissues of the body. A contrast agent can be injected into the patient to highlight different structures like the cardiovascular system. X-rays are ionizing radiations so a long or repeated exposure can provoke burns or cancers.

Computed Tomography (CT)
Computed Tomography scans rely on X-ray imaging. A series of beams is taken at different angle around the patient. Each beam produces a 2D image in a specific direction and all of these are used to reconstruct a 3D volume. CT scans are well contrasted and high resolution images. As they are based on X-rays they present a risk. They expose the patient to between 150 and 1,100 times the amount of radiation of a conventional X-ray.

Ultrasound
Ultrasound is a non-invasive, cheap imaging procedure, and is totally harmless. Therefore it is commonly used to visualize babies in utero but it is also used to visualize internal body structures. A probe sends waves of ultrasound which are reflected differently depending on the tissue properties. Computing the time the waves need to return to the probe gives the depth of the different kind of tissues. The main noise in ultrasound images is called speckle and is caused by the interferences of the different waves.

Magnetic Resonance Imaging (MRI)
Magnetic Resonance Imaging is a non-invasive technique (at least, the contrary has not been proven) with a high resolution. A powerful magnetic field is produced by a superconducting electromagnet. The magnetic field magnetizes

the tissues by aligning their spin magnetic moments. Weaker oscillating magnetic fields are then applied at the appropriate resonance frequency and cause the targeted atoms to emit a radio frequency signal. The targeted atoms need a non-zero spin magnetic moment to resonate, which is the case for hydrogen atoms. As they are present in large quantity in the body tissues like in water, they are chosen as the targeted atoms most of the time. MRI is particularly used to visualize the nervous system, muscles, the heart and tumours, but not bones because they do not contain enough hydrogen atoms. MRI is usually preferred over CT since it does not use any ionizing radiation. However, it is more expensive, time-consuming and claustrophobia-exacerbating.

Functional Magnetic Resonance Imaging (fMRI)
Functional Magnetic Resonance Imaging is a functional imaging technique. It relies on the magnetization of red blood cells and was discovered by Seiji Ogawa in [107]. In active areas there is an increase in the blood oxygen consumption and an increase of the blood flow to compensate: it is called the BOLD signal (Blood Oxygen Level Dependent). Yet deoxygenated red blood cells contain a molecule sensitive to Magnetic Resonance. The changes induced to the magnetic field by this molecule are detected and give the precise position of where the blood is highly consumed. fMRI is paramount for tumour surgery planning to identify the areas which are affected (tumour areas require lots of blood because they reproduce constantly and then tend to dry out the areas around) and the areas which are the most important for the activity of the organs.

Positron Emission Tomography (PET)
PET imaging allows to measure in 3D the metabolic or molecular activity of an organ, hence it is a functional imaging technique. It relies on the positrons emitted by a nuclear product injected into the body. The positrons emit two photons when it is annihilated which allow to know precisely the 3D position of the annihilation. As positrons are anti-matter particles, they are annihilated as soon as they are emitted by the nuclear product. Thus we know exactly the position and the concentration of the nuclear product. The nuclear product has a short half-life and is injected directly in the bloodstream. Most of the time the nuclear product is fluor 18 incorporated into a molecule similar to glucose, so that it will attach to tissues like the brain, the cardiac muscle or cancerous tissues. Different tissues or organs can be targeted by different isotopes. Of course, the procedure is invasive and potentially dangerous due to the use of radioactive particles.

1.2 Thesis Scientifical Context

A graph is a representation of a set of objects called the nodes or vertices, which may interact by pairs. The interaction is represented by an edge on the graph. A graph is a powerful structure capable of modeling interactions between elements. For example social networks, phylogenetic trees, and e-sport matchmaking structures are represented by graphs and many tools seek to predict the unknown parts of those graphs. In Computer Vision they are also commonly used to model the interactions of pixels or voxels of an image. Each voxel is a node of the graph and each node is linked by an edge to its closest neighbour. The number of neighbours depends on the complexity of the problem.

But the interaction is limited to couple of nodes, therefore some constraints which involve a higher number of nodes cannot be modeled with a graph. To avoid this limitation, we need a hypergraph, which is an extension of a graph containing hyperedges, sometimes called cliques in computer vision. An hyperedge is simply a link between more than two nodes.

1.3 Thesis Overview

This thesis is divided into 3 main chapters. In chapter 3 we will detail our model to perform linear registration within a discrete Markov Random Field framework. We will prove our model with local constraints can guarantee that we stay in the set of the global transformations we seek. We will also explain how Dual Decomposition works, why we are using this optimization algorithm and this decomposition.

Our linear graph is extended and combined with another graph to handle simultaneously linear and deformable registration in chapter 4. The purpose of this model is to tackle simultaneously linear and deformable registration. Usually, linear registration is computed as a first step and then deformable registration is performed. This sequential procedure can create a bias. Our framework removes this bias by solving both problems at the same time. However, this new graph is more complex than the original one and the pairwise term is not submodular. Therefore we will explore some potential algorithms to solve the problem.

Chapter 5 will focus on 2D/3D registration and how our previous graph structure can adapt to the problem of plane selection. We will adapt our structure and our strategies considering the different setting and show how modular it is.

The thesis is concluded in chapter 6 with a discussion on the presented

work, its implications, and suggestions concerning research direction in the future.

Chapter 2

Markov Random Fields and Optimization

In this chapter we will remind some useful mathematical notions and algorithms.

2.1 Graphs and Hypergraphs

A graph $G = (V, E)$ is a set of nodes V and a set of edges E. If the elements of E are couples of terms of V, then G is a directed graph. If the elements of E are pairs, then G is an undirected graph. Here, we will consider undirected graphs, that each edge is a pair of distinct nodes $\{i, j\}$ (no self-loop) and that each edge is associated with a non-negative weight $w_{i,j}$. If there is no edge between two nodes, we consider its weight to be zero by extension. The neighbourhood $\mathcal{N}(i)$ of a node i is the set of all nodes j that are connected to i by an edge:

$$\mathcal{N}(i) = \{j \in N, \{i, j\} \in E\}. \tag{2.1}$$

A simple way to describe a graph is by giving its adjacency matrix W. W it the $n \times n$ matrix, where $n = |V|$, defined by:

$$W_{i,j} = w_{i,j}. \tag{2.2}$$

Example
Let G_1 be the undirected graph defined by $N = \{a, b, c, d\}$ and $E = \{\{a, b\}, \{a, d\}, \{c, b\}, \{c, d\}\}$. The edge weights are set to $w_{a,b} = 2$, $w_{a,d} = 1$, $w_{c,b} = 0.5$, and $w_{c,d} = 1$. We represent G_1 in Fig. 2.1 and its adjacency matrix is:

$$\begin{pmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

The adjacency is used to define the Laplacian matrix L which is paramount to compute many other properties of the graph (spectral gap, spectral clustering)

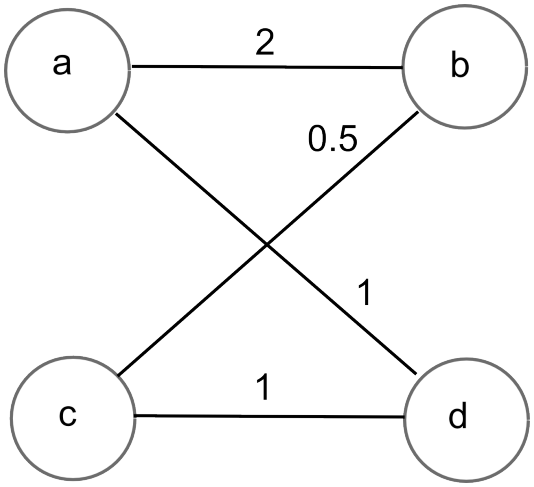


Figure 2.1: Visualization of the graph G_1.

as: $L = D - W$ where D is the degrees matrix. D is a diagonal matrix where $D_{i,i} = \sum_{j \in \mathcal{N}(i)} w_{i,j} \ \forall i$. Despite this simple definition, L is a true Laplacian is the sense of function analysis.

Usually, the nodes represent entities and the edges represent links between these entities. The graph represents the complexity of the interactions between the different entities. Graphs are widely used to solve a huge variety of problems. For example, social and street networks are naturally represented by a graph and used to optimise the traffic. Planar graph theory is used to minimize the number of integrated circuit required for different tasks. Graph colouring is used for task management or to optimize the allocation of radio frequencies.

A hypergraph is an extension of a graph. Hypergraphs were first introduced in [9]. Instead of having edges representing a link between two nodes, we have hyperedges representing a link between three or more nodes. Mathematically, a hypergraph $G = (V, E, H)$ is a set of nodes V, a set of edges E, and a set of hyperedges H. An element e of H is a set of nodes and is associated with a positive weight w_e. The key idea is to encode the fact that several nodes share a common property.

Example
Let G_2 be the undirected hypergraph defined by $N = \{a, b, c, d\}$ and $H = \{\{a, b, c\}, \{b, c, d\}\}$. The edge weights are set to $w_{a,b,c} = 1$, and $w_{b,c,d} = 2$. We represent G_2 in Fig. 2.2

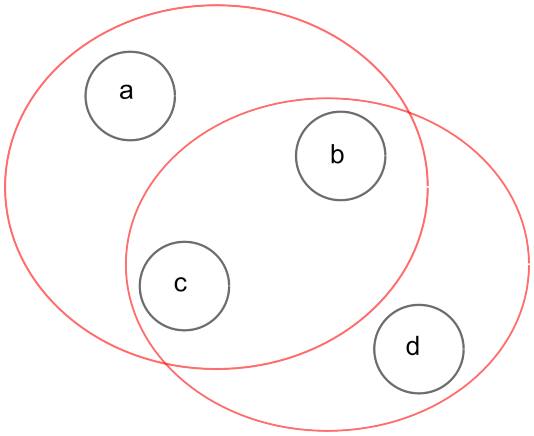


Figure 2.2: Hypergraph G_2 with two hyperedges $e_1 = \{a, b, c\}$ and $e_2 = \{b, c, d\}$.

2.2 Markov Random Fields

A growing number of graph-based approaches have been developed in Computer Vision. They can easily model different problems by representing different objects and their dependencies, and there exist many algorithms to perform inference. Moreover, in several cases, it can be proven that the solution is exact as long as the potential functions satisfy some properties.

A marginal distribution of a single variable X_i, where $X = \{X_i\}$ is a set of discrete variables with a joint mass function ρ, is:

$$P_{X_i}(x_i) = \sum_{x\prime \setminus x\prime_i = x_i} p(x\prime). \tag{2.3}$$

A Markov Random Field (MRF) is a probabilistic graphical model that represents a set of random variables and their dependencies through an undirected graph. Let $G = (V, E)$ be an undirected graph, V the set of nodes, E the set of edges, and $X = X_v, v \in V$ a set of random variables associated to the nodes of the graph. G and X form a Markov Random Field if the local Markov property is satisfied:

$$X_u \perp\!\!\!\perp X_{V \setminus \{\mathcal{N}(u) \cup u\}} | X_{\mathcal{N}(u)} \ \forall u \in V. \tag{2.4}$$

In practice it means a variable is independent of all others conditionally to its neighbourhood.

Let us define a clique c as a set of nodes where every pair is linked by an edge. A maximal clique is a clique such that, if we add another node, the resulting set is not a clique any more. When X follows a MRF, the joint density $p(x) = p(X = x)$ which denotes the probability of the random variables X

taking a specific configuration x, can be factorized over the maximal cliques of the graph:

$$p(x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c), \tag{2.5}$$

where C is the set of maximal cliques in G. x_c is the subset of variables which belongs to the clique c and a potential function $\phi_c(x)$ is defined over c. That is the mathematical definition of a clique. In computer vision, "clique" is often used as a synonym of hyperedge. In the rest of the thesis, we will use "clique" with this non-exact meaning.

Markov Random Fields were introduced in Computer Vision by [47]. They provide a probabilistic framework where knowledge can be integrated in the graph structure. For example for an image, the nodes will represent the pixels of the image and the edges will link a node to its closest neighbours. The hidden variable, or label we seek to recover may be discrete or continuous. The problem can be formulated as a maximum a posteriori (MAP) estimation:

$$x^\star = \arg\min_x p(x). \tag{2.6}$$

We can write the probability as a Gibbs distribution:

$$p(x) = \frac{1}{Z} exp(-E(x)), \tag{2.7}$$

with Z is a normalization constant and $E(x)$ is the energy defined on the graph as:

$$E(x) = \sum_{c \in C} \theta_c(x_c) \tag{2.8}$$

2.3 Stochastic Processes

Until 1920 statisticians used the method of moments for estimating distributions. This method is fast and easy to understand. The moments are the expected values of powers of the random variable we study, like the mean and the variance. The moments of a population are estimated by computing the moments of the data sample. This method works well for simple cases but not for more advanced problems. That is why Fisher introduced the maximum likelihood estimation.

When a given phenomenon has no accurate model, we can try to build a statistical one as precise as possible and based on the observations we can make. Let us note y the measured data, m the chosen model, and x the

parameters of the model, which are the labels in the MRF framework. y are the observable and m is the MRF structure. One solution is to find the most probable values of x once m is fixed and y is observed. So we want to maximise the probability of x knowing m and y:

$$\max_x P(x|m,y). \tag{2.9}$$

By application of Bayes' theorem, the problem is transformed into a Maximum a posteriori (MAP) estimation:

$$\max_x P(x|m,y) = \max_x \frac{P(y|m,x) \times P(x|m)}{P(y|m)}. \tag{2.10}$$

$P(y|m,x)$ is called the likelihood and measures how well the data and the predicted model fit. $P(x|m)$ is the a priori for given set of parameters. $P(y|m)$ the evidence is the sum of events representing y and is independent of the hidden variable x. We usually suppose that the parameters are uniformly distributed, which means $P(x|m)$ is constant, so the maximum a posteriori is equivalent to the maximum of likelihood:

$$\max_x P(x|y) = \max_x P(y|x). \tag{2.11}$$

2.4 A Review of MRF optimization algorithms

Here we present some well-known optimization algorithms used on graphs because we will use or mention them in other chapters. The problem has been modeled into a graph. We want to assign a label l_p from the label set L to each node p. For each node, there is a unary function, depending on the label, which gives a cost corresponding to the assignment of that label to that node. For each edge, there is a pairwise function, which gives a cost corresponding to the assignment of a couple of labels to the couple of nodes linked by the edge. A pairwise function is frequently modeled by a square matrix whose size is the number of labels. Here, we will focus on undirected graphs and not hypergraphs. Thus, the energy of the MRF is:

$$E(x) = \sum_{p \in V} U_p(l_p) + \sum_{(p,q) \in E} V_{p,q}(l_p, l_q), \tag{2.12}$$

where $x = \{l_p\}_{p \in V}$, l_p is the label corresponding to the node p, $U_p(l_p)$ denotes the *unary potentials* of node p, and $V_{p,q}(l_p, l_q)$ denotes the *pairwise potentials* between the nodes p and q. We search for the minimum of the energy and its argument:

$$x^\star = \arg\min_x E(x). \tag{2.13}$$

2.4.1 Viterbi Algorithm

The Viterbi algorithm is used to predict the most probable sequence of hidden variables x that produced the sequence of observations y. The observations and the model are supposed to be known. The Viterbi algorithm works for chains of nodes or trees.

In graph theory, a tree is an undirected, acyclic and connected graph. It looks like the branching of a tree. We choose one node to be the root of the tree; the nodes (except the root) which are connected to only one other node are called the leaves of the tree. For this sort of graphs, message passing gives an exact solution and is also called forward-backward algorithm. We first define a message from a leaf a to node b as:

$$\forall l_b \in L, m_{a\to b}(l_b) = \min_{l_a \in L}(U_a(l_a) + V_{a,b}(l_a, l_b)). \tag{2.14}$$

Then we define a message from a non-leaf node c to another node d:

$$\forall l_d \in L, m_{c\to d}(l_d) = \min_{l_c \in L}(U_c(l_c) + V_{c,d}(l_c, l_d) + \sum_{b \in \mathcal{N}(c)\setminus d} m_{b\to c}(l_c)). \tag{2.15}$$

A message may be viewed as the update of the unary potentials by taking into account the potentials of the sending node. The algorithm starts by sending messages from the leaves until they arrive in the root. The information contained in the root:

$$\forall l_r \in L, m_r(l_r) = U_r(l_r) + \sum_{b \in \mathcal{N}(r)} m_{b\to r}(l_r)) \tag{2.16}$$

is equal to the minimum of the energy of the whole tree for the different labels in the root node. We now know the most likely hidden variable of the root, which corresponds to the minimum of the energies. To determine the hidden variables of the other nodes, we simply have to look back at the minima we chose during the computation of messages.

2.4.2 Belief Propagation

Belief propagation or sum-product message passing is used to perform inference and to compute the marginals of the nodes of the graph. It is exact on trees, but gives good approximation on general graphs unlike the Viterbi algorithm. The idea is similar to the Viterbi algorithm as is relies on passing messages from nodes to nodes, updating the potentials and choosing the minimum of the marginals to select the label for each node. Very fast and efficient solutions have been developed as it is very general and gives good approximations even for general graphs for which it is called Loopy Belief Propagation and has no mathematical properties.

Belief propagation it a general algorithm, but may not converge when the graph contains cycles. Naively, the number of messages is exponential in the size of the largest clique, but optimization methods have been developed to be more efficient, like Potetz [115] which reaches linearity in the size of the largest clique for a class of potential functions.

2.4.3 Graph Cuts

A graph cut is a partition of the nodes of a graph into two sets. It is defined for graphs whose edges are associated to a weight. The cost of the cut is the sum of the weights of the edges which are destroyed by the cut. Formally, let us note S and T the two subsets forming the partition of V. The cut-set of the cut is the set of edges which have one node in S and one in T: $\{(u, v) \in E | u \in S, v \in T\}$. A common problem is to find the minimum cut or min-cut of a graph. A cut is minimum if there is no other cut whose cost is smaller. A s-t cut is a specific kind of cut where two predetermined nodes called the source, s, and the sink t, are not in the same subset. It is an important notion in flow network because the max-flow min-cut theorem shows that the cost of a s-t min cut is equal to the maximum flow in a network flow. The theorem was proven in 1956 independently by two teams, Ford and Fulkerson, and Elias, Feinstein, and Shannon. Intuitively, if the graph represents a net of pipes with only one entry point called the source, and one exit point called the sink, then the maximum flow we can reach is equal to the bottleneck pipe. An efficient algorithm in polynomial type was proposed in [43]. It is an iterative algorithm that, at each step tries to increase the current solution which is a possible flow by exploring one by one every node.

This algorithm can easily solve a binary MRF. Indeed, let us define two new nodes which will represent the two potential labels. These nodes will be the source and the sink. Let us add two edges per node, one linking the node to the source, the other one linking the node to the sink. After the min-cut, the regular nodes (i.e. neither the source nor the sink) in the same subset as the source are assigned to the first label, those in the same subset as the sink are assigned to the second label.

The values of the binary pairwise and the unary potentials have to be encoded into this new graph. Let us imagine that our original graph contains only two nodes a and b, that the two labels are 0 and 1, and the potentials are defined by the functions $U_a(x)$, $U_b(x)$, $V_{a,b}(x, y)$. The nodes in the same subset as the source will be assigned the label 0, the nodes with the sink will be assigned the label 1. The modified graph which will be used with the Ford-Fulkerson theorem will be the graph on Fig. 2.3. Let us put, at least for a first step, those potentials:

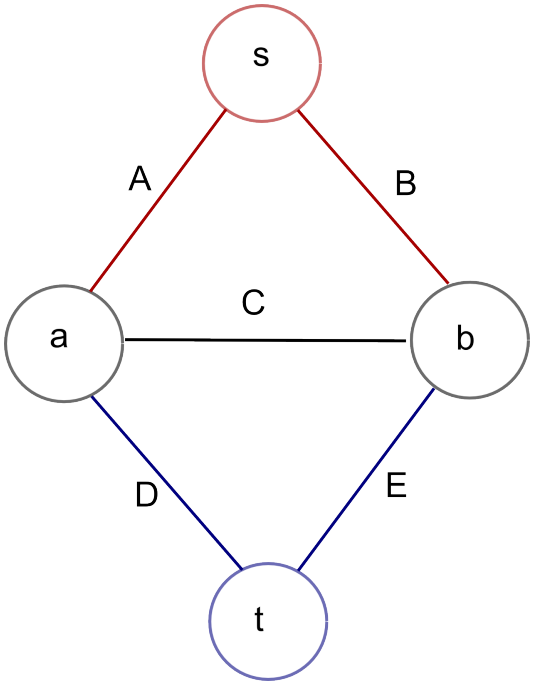


Figure 2.3: Graph cut graph settings.

$$A = U_a(1) + \frac{+V_{a,b}(0,0) + V_{a,b}(1,1) + V_{a,b}(1,0) - V_{a,b}(0,1)}{2} \tag{2.17}$$

$$B = U_b(1) + \frac{-V_{a,b}(0,0) + V_{a,b}(1,1) - V_{a,b}(1,0) + V_{a,b}(0,1)}{2} \tag{2.18}$$

$$C = \frac{-V_{a,b}(0,0) - V_{a,b}(1,1) + V_{a,b}(1,0) + V_{a,b}(0,1)}{2} \tag{2.19}$$

$$D = U_a(0) + V_{a,b}(0,0) \tag{2.20}$$

$$E = U_b(0). \tag{2.21}$$

Now a s-t cut will correspond to the corresponding energy of the assignment on the original graph. But the algorithm needs positive weights, otherwise it may fall into endless loops. If A, B, D or E is negative, there are simple manipulations to change the graph. For example, if A is negative and D positive, we can add the opposite of A to D and put A to 0; if they are both negative, the new A will be equal to the opposite of D and the new D will be

equal to the opposite of A. These operations do not change the corresponding energy as there are four possible s-t cuts, two of which cut the A edge, while the two others cut the D edge. After these operations, A, B, D and E are positive. So we only have one condition on our energy functions which comes directly from the positivity of C:

$$V_{a,b}(0,0) + V_{a,b}(1,1) <= V_{a,b}(1,0) + V_{a,b}(0,1). \tag{2.22}$$

This type of function is called submodular and the notion can easily be extended to non-binary functions [88]. Submodular functions have originally been defined on sets of objects for game theory and economics. As the two graphs are equivalent and as the Ford-Fulkerson algorithm is exact, the solution is the global optimum. [88] proved that, with graph cuts, the global optimum of the energy is reached if the MRF function is submodular. The procedure we showed for a very simple graph can obviously be extended to bigger graphs. Graph cuts is restricted to a class of energy : when E is submodular and x_i is a boolean, the minimum can be found in polynomial time thanks to the graph cuts algorithm (see [70] or [88]).

Graph cut based algorithms have been designed to handle multilabel MRFs in [13]: α-expansion and $\alpha - \beta$-swap. For α-expansion, the idea is to change progressively the labels by trying them one by one. At each step, the current assignment of every node is considered as the same label, the old label, while a new one called α is proposed. The graph cut is realised with only those two labels. For $\alpha - \beta$-swap, at each iteration a couple of labels is chosen and the nodes can change their label between those two. Torr et al. [144] showed those algorithms could solve even higher order energies in polynomial time, if the potential were a metric (for α-expansion) or a semi-metric (for $\alpha - \beta$-swap). The fusion move algorithm was introduced by Lempitsky et al. in [96]. It consists of finding the best combination of two potential labelling of the whole space. α-expansion is then a special fusion move, where one of the labelling is the constant α on the whole space.

Several methods have been developed to accelerate the optimization [152] or to handle dynamics MRFs [83], for example to handle interactive object segmentation [12]. Some authors extended the graph cuts to non-submodular functions, like Kolmogorev and Rother in [87]. They reparametrize the potential to reach a normal form. Then, they create a graph, which contains two nodes for each variable, x_i and $\neg x_i$. The algorithm, called Roof Duality or QPBO (Quadratic Pseudo-Boolean Optimisation) only gives a partial labelling, some nodes are still unknown, but it has some good properties known as Weak autarcy and Partial optimality. Those properties allowed Rother et al. [126] to extend Roof Duality to try to complete the partial solutions in some cases. Grabcut [125] is an algorithm based on graph cuts to perform

image segmentation while only drawing a simple box around the object, where they also optimize the convergence.

2.4.4 Spectral clustering

When there is no unary potential, one can use spectral clustering. A popular objective function used in spectral clustering is to minimize the normalized cut [135]. The cost of a cut is :

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$$

where A and B are the two parts of the graphs, and w is the weight function of the edges. This cost favours little cuts. Let us note $assoc(A, V) = \sum_{u \in A, v \in V} w(u, v)$ where V is the set of all nodes. This measures the total connection from nodes in A to nodes in the graph. The normalized cut is :

$$Ncut(A, B) = cut(A, B) \left(\frac{1}{assoc(A, V)} + \frac{1}{assoc(B, V)} \right)$$

The idea of spectral clustering is to use the similarity matrix, or the laplacian matrix, and to find interesting properties thanks to their eigenvectors. A known algorithm was developed by Ng et al. in [106]. Given a similarity matrix of all the couples, calculate the laplacian matrix and its eigenvectors. Those eigenvectors define points in $\mathbf{R}^n$ which can be more easily clustered and give the clustering of the initial objects.

2.4.5 Probabilistic clustering

In Bulò et al. [17], the authors try to cluster points without partitioning them. Let W be the empirical likelihood matrix, and Y be the matrix of the unknown cluster membership. They try to minimize :

$$\min \|W - \alpha Y^T Y\|^2$$
$$\text{s.t. } \alpha \in \mathbb{R}, \ \sum_i Y_{i,j} = 1 \ \forall j, \ Y_{i,j} \geq 0 \ \forall i, j$$

They solve it thanks to the Baum-Eagon inequality, which, for some polynomials, give a growth transformation.

2.4.6 Order Reduction Techniques

The order of the energy can be reduced by adding new variables. Ramalingam et al. proved we could transform submodular multi-labels energy functions

into submodular boolean functions in polynomial time [119], but that it was impossible for order higher than 4.

In [124], Rother et al. remarked that the potential function over a clique is often sparse, which means many configurations are considered bad and so, have the same value, and that the good configurations were often similar. So they try to parameterize compactly the energy, to minimize the number of variables. They transformed the minimizing sparse higher-order energy function into an equivalent quadratic minimization problem.

Ishikawa introduced in [67] a method to convert the minimization problem of any higher-order binary energy to that of a pairwise and unary energy. He then used fusion moves in [68] for higher order energies. The fusion move could be very efficient if we are able to propose a good solution at each step. He did so using the gradient of the energy. Recently [69], he showed that there were numerous methods to reduce a higher order energy, and that the method could influence the solving by the algorithm. So he pointed out some conditions such that a function is minimizable in polynomial time using graph cuts. Gallagher et al. developed that idea in [46]. Their algorithm called Order Reduction Inference, searches over a set of reduction methods to minimize the difficulty of the optimization problem.

Recently, Kahl and Strandmark [77] generalized Roof Duality to polynomials of any degree. Avoiding to reduce the function to a quadratic pseudo-boolean one gives them tight bounds.

2.4.7 Fast Primal-Dual

The discrete problem can be written as a linear integer program, and then be relaxed into a continuous problem. The continuous problem is generally easier to solve because the technique are very well-known, but the solution found has to be sent back in the discrete domain, and thus, may be far away from the true discrete optimum. That is the case of Fast Primal-Dual or Fast-PD. In [93], Komodakis et al. introduced the primal-dual schema to Computer Vision and Markov Random Fields optimization. Primal or Dual approaches take the problem and try to optimize a solution. Here, the idea is to find and update a pair (x, y) of primal and dual solutions of the problem, which gives us a framing of the value of the solution. This framing is called the primal-dual gap. When the ratio y/x is smaller than a value f, we say the solution is an f-approximation to the optimal solution. Practically, the algorithm iterates until a pair of solution satisfies a pre-defined constraint f. Every iteration consists in solving a max-flow problem for a certain graph which depends on the current primal-dual pair of solutions. As it uses the graph cuts algorithm explained previously 2.4.3, Fast-PD requires submodular

functions to give mathematical guarantees. And the speed of the algorithm depends on the complexity of the max-flow problem. Fast-PD uses the pair of primal-dual solutions from the previous iteration to reduce the number of augmenting paths required. Finally, if the pairwise of the original graph is a metric, Fast-PD is certified to be as precise as α-expansion and faster.

2.4.8 Graph Decomposition

According to Alahari [2], most MRF have a simple part and a difficult part. His idea is to solve and fix the easy part first, to have more time to detail the difficult part of the MRF. Dividing a complex graph into smaller and much easier to solve ones has been tried. Wainwright et al. [155] tried to decompose the graph in a convex combination of trees (Tree Reweighted). Optimizing the problems independently on the trees give an upper limit of the optimum, which is all the more tight as the trees have a compatible optimal structure. Kolmogorov developed a new version of the previous algorithm called Sequential Tree Reweighted in [85] which has better convergence guaranties.

The main problem is the synthesis of the different subgraphs, when they do not agree on the label which should be assigned. In order to force them into agreeing, Komodakis and Paragios [90] developed a cycle-repairing algorithm, to fix any inconsistent cycles that may appear during optimization. The repairing helps to tighten the relaxation and gives better results.

[91] presented a master-slave based framework. The graph (or hypergraph) and the energy function are divided in several parts which are solved individually by the slaves. The master see the results and change the division of the energy function until an agreement is found. This algorithm is called Dual Decomposition, and will be detailed in 3.3.

Batra et al. presents a new way to decompose a graph [6]. In order to minimize the coordination work necessary to harmonize the subgraphs (the master work in the previous paragraph), they give more complex work to the subproblems (the slaves). They construct their subgraphs in a wider set, called Outer-Planar Graphs. Finding a spanning family of such graphs seems quite easy in practice. The harmonization is realized by message passing between subgraphs.

2.4.9 Extensions to hypergraphs

Hypergraphs are a very difficult problem. Some of the previous algorithms had been extended to work for hypergraphs, but generally the computation time increases too.

In [145], Torr et al. use unsupervised segmentation to create textons (or image segments), using several Mean-Shift, thus limiting the number of points of the image (a texton contains a lot of pixels). Higher order terms of the energy try to have all the pixels of a texton with the same label, but with a decreasing energy (two pixels labelled differently is worse than only one). It is called the Robust Pn Model, compared to the Pn Model, where the cost is the same as soon as one pixel is labelled differently than the rest of the texton. HO terms are really restrictive, such that order can be easily reduced thanks to a few variable. Second order energy is then optimized using move making algorithms.

In [158], Werner proposed a max-sum diffusion algorithm as a generalization of n-airy LP relaxation, and to tighten relaxations. He uses the marginal polytope to define a way to add hyperedges without changing the problem while tighten, pace by pace, the relaxation.

Dual Decomposition and the related algorithms can handle hypergraphs but the hyperedges must be solved at some points. So the slaves solvers must deal with them, either with a specific algorithm like those we are reviewing, or with exhaustive search if the slave is small enough.

The use of factor graph [75, 99] allows to run message-passing algorithms on the factorized graph but the size of the graph and the number of iterations to reach convergence increase drastically.

2.4.9.1 Spectral Methods

Zhou et al. [165] extend the spectral clustering method to hypergraphs. The energy is inspired by the normalized cuts as :

$$\text{argmin}_{\delta S} vol(\delta S)(\frac{1}{\text{vol } S} + \frac{1}{\text{vol } S^c})$$

where S is a set of the nodes, S^c is the complement of S, $vol(\delta S)$ represent the weight of the cut hyperedges, and vol(S) is the weight of the hyperedges in S. This problem is then relaxed and written as a matrix multiplication of the form $f^T \Delta f$. Δ is then called the hypergraph Laplacian, and a method close to the 2D one is applied. Huang et al. [64] use that framework to realize video object segmentation. They use over-segmentation to create the nodes and motion cues to calculate the hyperedges. The results are encouraging.

In [31], Duchenne et al. proposed a tensor-based algorithm for graph matching. They formulate their problem with an assignment matrix X and a similarity tensor H, as searching $\max H \otimes_1 X \otimes_2 X \otimes_3 X$. The solution is found thanks to a higher-order power method developed in [120].
Zass and Shashua made a strong conditional independence assumption of the

matching [162], so the similarity matrix S is a power of the probability assignment matrix X : $S = \otimes^d X$, with d the size of the hyperedges.

2.4.9.2 Order reduction

The order of the energy can be reduced by adding new variables. Ramalingam et al. proved we could transform submodular multi-labels energy functions into submodular boolean functions in polynomial time [119], but that it was impossible for order higher than 4.

In [124], Rother et al. remarked that the potential function over a clique is often sparse, which means many configurations are considered bad and so, have the same value, and that the good configurations were often similar. So they try to parameterize compactly the energy, to minimize the number of variables. They transformed the minimizing sparse higher-order energy function into an equivalent quadratic minimization problem.

Ishikawa introduced in [67] a method to a convert the minimization problem of any higher-order binary energy to that of a pairwise and unary energy. He then used fusion moves in [68] for higher order energies. The fusion move could be very efficient if we are able to propose a good solution at each step. He did so using the gradient of the energy. Recently [69], he showed that there were numerous methods to reduce a higher order energy, and that the method could influence the solving by the algorithm. So he pointed out some conditions such that a function is minimizable in polynomial time using graph cuts. Gallagher et al. developed that idea in [46]. Their algorithm called Order Reduction Inference, searches over a set of reduction methods to minimize the difficulty of the optimization problem.

Recently, Kahl and Strandmark [77] generalized Roof Duality to polynomials of any degree. Avoiding to reduce the function to a quadratic pseudo-boolean one gives them tight bounds.

2.4.9.3 Hypergraphs decomposition

As graphs were divided in subgraphs to calculate the energy, hypergraphs can be decomposed too. In [145], Torr et al. use unsupervised segmentation to create textons (or image segments), using several Mean-Shift, thus limiting the number of points of the image (a texton contains a lot of pixels). Higher order terms of the energy try to have all the pixels of a texton with the same label, but with a decreasing energy (two pixels labelled differently is worse than only one). It is called the Robust P_n Model, compared to the Pn Model, where the cost is the same as soon as one pixel is labelled differently than the rest of the texton. HO terms are really restrictive, such that order can be

easily reduced thanks to a few variable. Second order energy is then optimized using move making algorithms.

In [158], Werner proposed a max-sum diffusion algorithm as a generalization of n-airy LP relaxation, and to tighten relaxations. He uses the marginal polytope to define a way to add hyperedges without changing the problem while tighten, pace by pace, the relaxation.

Hypergraph partitioning in VLSI design

The creation of integrated circuits containing thousands of components is called Very Large Scale Integration. The circuits is modelled by a hypergraph. We want to separate the circuits into different parts, while minimising the interactions between the parts, so engineers early used hypergraph partitioning.

A well-known, iterative improvement algorithm to partition is called Fiduccia-Mattheyses algorithm (see [18] for details and possible improvements). It is initialized with a random cut. At each pass, every node can change its place in the partition, and a gain is associated with every move. Each pass contains a loop of steps. At each step, gains are updated, the best one is chosen and the related move is done. The moved node is fixed in the rest of the pass. The inner loop stops when all the nodes are fixed. The best configuration seen during the inner loop is chosen of a departure point for the next pass. The algorithm stops when the energy has not been changed during a pass.

In [3], Kayaaslan et al. review some approaches to solve the partitioning problem. The dual of a hypergraph is a hypergraph where nodes and edges are reversed. The Clique-net graph transforms a hyperedge into a clique, but cannot assign perfectly the initial cost. The Net-intersection graph (NIG) represents a hypergraph $H = (U, N)$ as a graph $G = (V, E)$ such that the nodes of G are the hyperedges of H and the edges of G links two nodes if the corresponding hyperedges shared a hypernode in H. They use the last one and realize a vertex separation. To partition the hypergraph in several parts, they use Recursive Bipartitioning, which is commonly used in that domain.

A game theory perspective

In [111], Pellilo and Bulò question the fact that clustering is often thought of as partitioning. They say clustering is more about grouping points which are similar, so outliers may remain alone, and some points can belong to several clusters. From this point, they construct a k-players game [16], where k is the order of the interaction between the points. Each player chooses a point and they received a reward proportional to the similarity of the k points, and nothing if they have the same point. Thanks to game theory, he is able

to evaluate the best strategies which regroup the points having the higher similarities, i.e. a cluster.

2.5 Markov Random Fields for Image Registration

Image registration algorithms aims to find a transformation T that best aligns two images I, $J : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}$. d is the dimension of the image, usually two or three and Ω is the image domain. The image I will be deformed by the transformation T to fit the image J. I is called the source or moving image, J is called the target or fixed image. We call $I \circ T$ the deformed image. The set of transformations that is explored depends on the applications (deformable, rigid, affine). We seek to find the best transformation $\hat{T}$ considering an energy or cost function $\mathcal{E}$ created to evaluate how well the target image J and the deformed image $I \circ T$ fit.

$$\hat{T} = argmin_T \mathcal{E}(T). \tag{2.23}$$

As we saw before, the Bayes' theorem says that minimizing the energy is equivalent to maximizing the posterior distribution $P(T|I, J)$. The energy function contains two parts. The first part measures the fitting of the two images, while the second part is a regularization term that depends on the transformation set we are searching in. The second part may be always zero, is the algorithm explores exhaustively a small set of transformations, or may be essential to make the problem well-posed, like for deformable transformations.

2.5.1 Similarity Measures

In medical imaging, common similarity measures are pixel-based (or voxel-based), which means we superimpose the two images and compare the values of the intensity of the pixels or voxels. We describe here some of them. The first two, SAD and SSD are monomodal measures, which means they are used to compare images from the same modality. Multimodal measures are used for images of different modalities. Some are general and can adapt to different pairs of modalities, many of them are derived from the mutual information. Some are specific to a couple of modalities, like LC2, and should be performing better than mutual information on that specific couple.

In the following paragraphs, I and J are the two images we compare and Ω is the image domain, which means the coordinates of the points for which we have the values of both the images. In practice, the coordinates of the voxels do not coincide, so the voxel values of at least one image are

interpolated. The procedures to follow when we do not have the values of both the images (normalization, extrapolation) depends on the problem and will not be discussed here. Depending of the definition, the measure requires to be maximised or minimised. In practice, we will always minimise our energy, and just take the opposite if needed.

Sum of Absolute Differences

SAD is a basic similarity measure used for monomodal images. It compares the intensities of the pixels and penalizes when they are different.

$$SAD(I, J) = \sum_{x \in \Omega} |I(x) - J(x)|. \tag{2.24}$$

It is a fast and simple way to compare two images but may be unreliable because of changes of luminosity for example. It may be used with other methods to improve the precision of the results.

Sum of Squared Differences

SSD is similar to SAD but uses an L_2 norm.

$$SSD(I, J) = \sum_{x \in \Omega} (I(x) - J(x))^2. \tag{2.25}$$

Normalized Cross-Correlation

Cross-Correlation is a similarity measure between two signals. It is a basic statistical approach that assumes a linear relation between the two signals. It is similar to the convolution of two functions. In image registration, it is defined as:

$$CC(I, J)(u, v) = \frac{\sum_{(x,y) \in \Omega} (I(x, y) - J(x - u, y - v))}{(\sum I(x, y)^2 \sum J(x - u, y - v)^2)^{0.5}}, \tag{2.26}$$

where (x, y) are the coordinates of points in the image domain Ω, (u, v) are the translation parameters. The maximum of the energy is obtained for the best translation parameters. A fast way to compute it is developed in [97].

Mutual Information

Mutual Information was introduced in [153]. It measures the statistical dependency between two variables. Let us remind that statistical dependency and causality are two different concepts. If X and Y are two discrete random variables, $P(X = x)$ and $P(Y = y)$ the probabilities of events x and y,

$P(X = x, Y = y)$ the joint probability of events x and y, then their mutual information is defined as:

$$MI(X,Y) = \sum_{x,y} P(X = x, Y = y) \log \frac{P(X = x, Y = y)}{P(X = x)\, P(Y = y)}. \tag{2.27}$$

Mutual information is positive and symmetric. It is equal to zero if and only if the two variables are independent. It increases as the dependency increases.

To compare images, we consider the grey values of the two images as random variables. Because of the huge quantity of values, we split the grey values into a predefined number of bins (16, 32 or 64 usually). We wanted to take into account changes in the luminosity, as it was a limitation of SAD and SSD. Mutual information does not associate bins of the same value directly. On the contrary, the measure evaluates if there is an assignment of the bins of the two images. It means that if the second image is the same one as the first image in which colors have been swapped (for example blue is yellow, green is purple, and orange is white), then the mutual information will be equal to zero.

Normalized Mutual Information

We can use the definition of entropy $H(X) = -\sum_x P(X = x) \log P(X = x)$ to express the mutual information as:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \tag{2.28}$$

where $H(X,Y)$ is the joint entropy of X and Y. We see that $MI(X,X) = H(X)$ so in order to compare the mutual informations of different variables, we define the Normalized Mutual Information as:

$$NMI(X,Y) = \frac{MI(X,Y)}{\sqrt[2]{H(X)H(Y)}}. \tag{2.29}$$

Then the value is between 0 and 1.

Linear Correlation of Linear Combination

Linear Correlation of Linear Combination (LC2) is a similarity measure that was introduced in [156] for CT-ultrasound registration. It was latter shown in [157] to perform also well for MRI-ultrasound registration. This couple of modalities, CT and Ultrasound, or X-rays and Ultrasound similarly, is very difficult to registrate, because these two modalities show different structures and it is not handled very well by classical multimodal measures. Indeed, Computed Tomographies show the structure of the body but Ultrasound show also where the structure changes. It may be a change a texture in an organ,

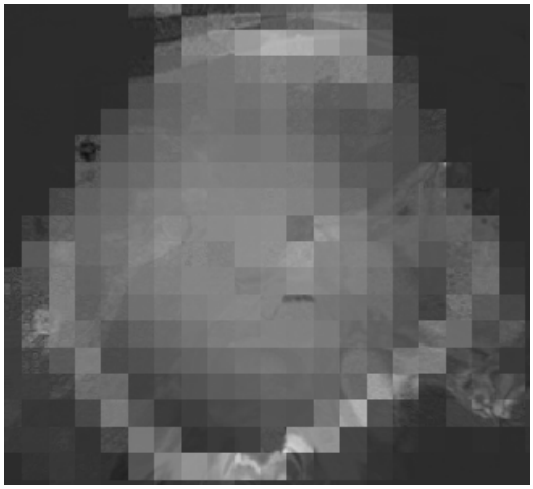


Figure 2.4: The LC2 metric tries to simulate a US image from the CT image.

or a change of organ, and there the intensity of the Ultrasound is high. So it makes sense to compare the intensity in the Ultrasound with the gradient of the CT to take that particularity into account. After explaining with more details the physics behind the two modalities, the authors of [156] determine the intensity of the ultrasound should be compared to a linear combination of the intensity and of the gradient of the image, plus a constant term. It is similar to simulating an ultrasound image from the CT image with the simulated intensity defined as:

$$x_i = \alpha p_i + \beta r_i + \gamma, \forall i \tag{2.30}$$

where x_i represent the simulated intensity of the point i, p_i is the intensity in the CT image, r_i is the gradient in the CT image and $\alpha, \beta, and \gamma$ are three unknown weighting parameters. An example of a such a simulation is shown in Fig. 2.4, where each square has its own parameters.

The unknown parameters are dependent of the organ, so they are not constant on the image. They are not even constant on an organ, because the gradient should be important on the edges but not in the middle where it mostly comes from noise during the image acquisition. Practically, the parameters are computed for different patches on the image. In their experiments [157], they search for the best size of the patch. We recommend to choose patches of about 15mm for linear registration. To compute the best set of parameters for a patch, we want to minimize:

$$\sum_{i=1}^{n} \|(x_i - u_i)\|^2, \tag{2.31}$$

where u_i is the intensity in the ultrasound and where the pixels of the patch are indexed by $1 \cdots n$. If we pose

$$M = \begin{pmatrix} p_1 & r_1 & 1 \\ \vdots & \vdots & \vdots \\ p_n & r_n & 1 \end{pmatrix}, \tag{2.32}$$

then we want to minimize:

$$\left\| M \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} - \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \right\|^2 \tag{2.33}$$

Therefore the solution is:

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = (M^T M)^{-1} M^T \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} \sum p_i^2 & \sum p_i r_i & \sum p_i \\ \sum p_i r_i & \sum r_i^2 \sum r_i & \\ \sum p_i \sum r_i n & & \end{pmatrix}^{-1} \begin{pmatrix} \sum p_i u_i \\ \sum r_i u_i \\ \sum u_i \end{pmatrix} \tag{2.34}$$

Finally, we can compute the LC2 measure with the formula:

$$1 - \frac{\sum_i^n (x_i - u_i)^2}{n \times Var(u)}, \tag{2.35}$$

where $Var(u)$ is the variance of the intensities $\{u_i, 1 \leq i \leq n\}$. This measure is to be maximised.

2.5.2 Our MRF Framework

In this thesis, we explain different MRF problems that we optimize to find the best displacement at each iteration. We detail this procedure and how the successive iterations are connected here.

2.5.2.1 The Data Term

At each iteration, a rectangular grid (in 2D or 3D) of control points is superimposed on the moving image. In Fig. 2.5 the control points are the red intersections. For each control point, we seek to determine the best displacement vector (the label of the corresponding node(s)) among a set of labels L. Let p be a control point and l_p a potential label. To evaluate the data term, we want to compute the similarity measure of two patches in the source and the target images. The patch is defined as a rectangle whose dimensions are the spacing of the grid. We use the patch centred at the control point in the source image as in Fig. 2.6, we note it B_p. We want to know where it fits in the target image, so the second patch is centred at the control point p plus the translation corresponding to the label l_p, we note it B_{l_p}. Several patches for different labels can be seen in Fig. 2.6.

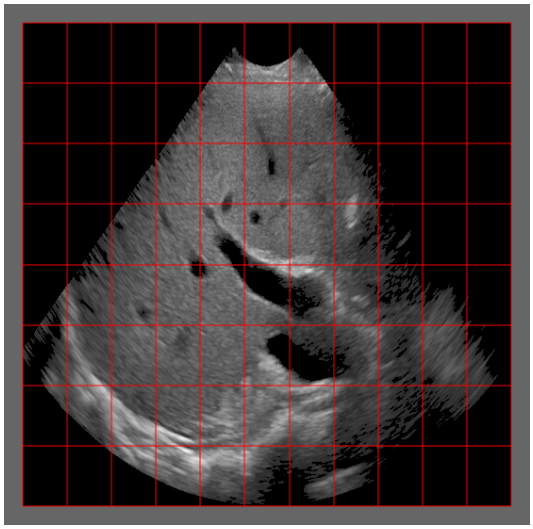


Figure 2.5: A grid of control points is superimposed to the image. The points are chosen at the red intersections.

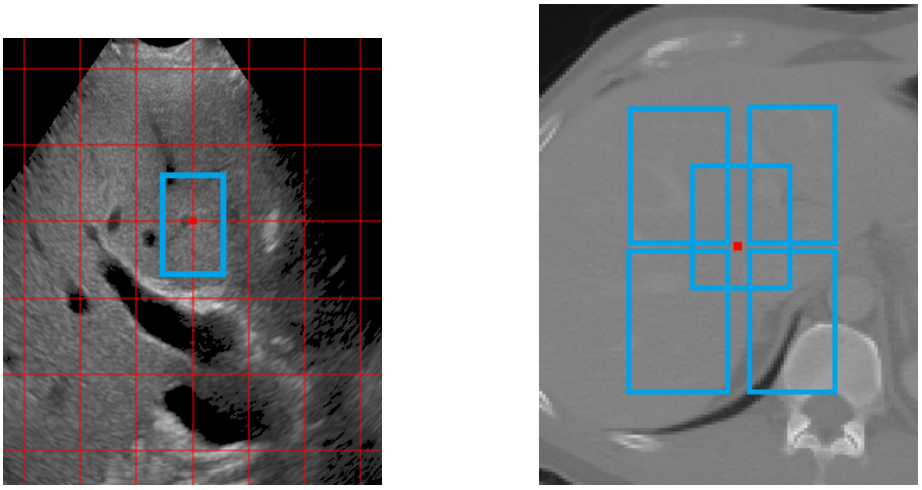
2.5.2.2 Free Form Deformations

As we saw, the similarity measures are voxel-based, and their computation time is an issue. Moreover, the data cost is not smooth so we do not want to compute a gradient. Thus, we use a discrete approach. When the best displacements of the control points is found, at the end of an iteration, the movement of all the points of the moving image is approximated using splines. This model is very common in medical imaging and is called Free Form Deformations (FFD). It was introduced in Computer Graphics in [130] and gained interest in Medical Imaging since its coupling with cubic-B Splines in [29, 95, 127].

The main advantages of the FFD is its simplicity and smoothness, as it can model, with few points, complex deformation fields. The resulting transformation of cubic-B Splines is generally smooth and preserves the topology. To make sure it is always the case, we impose the maximum displacement of every control point is no greater than 0.4 times the grid spacing. A proof regarding the preservation of the topology may be found in [21, 128].

2.5.2.3 Iterations in our MRF Framework

The discrete approach only allows a small number of displacement vectors for each control points, so many iterations are performed during which a larger search space of displacement is explored. In practice, large displacements are available at the first iteration to have a large search space, then the displacements are decreased, in size, during the other iterations until we reach



(a) Source image (b) Target image

Figure 2.6: The comparison of the data term: (a) A patch, in blue, is centered around the big control point in red. (b) Different patches are shown for different displacements vectors, around the control point in red. They are to be compared with the patch in the source image.

generally a sub-voxel precision. The deformation fields are composed through the different iterations.

Moreover, the quality of the images is often decreased at the beginning to the iterations to speed-up the computation, in a coarse-to-fine manner. The resolution increases and is the best possible for the finer iterations.

Chapter 3

Linear Registration

3.1 Introduction

Linear transformations include rotations, skewings, scalings, and translations. In this chapter, when we will say "linear registration", we will mean affine, rigid or similarity registration. These transformations are global and do not model local changes.

Linear registration is an important step in image processing pipelines. It is used in computer vision, video surveillance, medical imaging, and for military purposes. Difficulties come from the difference of the sensors which took the images, from the camera viewpoints and from the changes through time. The diversity of images makes it impossible to have a universal method that could handle all registration problem. When an algorithm or a procedure is chosen, it should take into account signal to noise ratio, the registration accuracy needed, the type of deformations, and every specific parameter which depends on the application and which is of importance. But there are two main families of registration methods: feature-based or geometric methods and image-based or iconic methods.

The importance of registration and the constant need for improvement regarding accuracy and speed caused it to be an old but still active research area. The first survey paper on registration [15, 79] focused on image correlation methods. Feature-based methods have been detailed in [168].

3.1.1 Feature-based Methods

A manual way to perform linear registration is to determine a set of control points in the two images. If the number of constraints imposed by the set of control points is bigger than the degree of freedom of the transformation we seek, then we use smoothing methods like least square regression or iterative algorithms like RANSAC [40]. This is also the main idea of feature-based or geometric methods. First, distinctive points like salient contours are detected in both images. These points are called features and can be significant regions, lines, or points. Second a correspondence between the two features sets of the two images is performed. The particularities of the images we mentioned

before is to be taken into account while choosing the feature descriptor and the similarity measure to create the assignment. Spatial relationship must also be part of the descriptors as there must be a coherence in the assignment. Third the best transformation is estimated based on this correspondence. Finally the transformation and the resampling is applied to the source image to fit the target image.

The choice of the feature descriptors is a key element of the algorithm. They should be robust regarding the quality or the degradation of the image and they have to be able to discriminate enough the control point to perform a suitable assignment. They have to be invariant to the different acquisition conditions (viewpoints, illuminations, cameras) so the assignment between the two images is possible. Moreover the control points should be spread all over the image to estimate accurately the transformation. The feature can either be based on a point, a line or a region.

The point features group uses very specific geographic points like road crossings [57] or centroids of water regions, line intersections [142], local extrema of wavelet transforms [42] (for satellite images), and corners [62]. [123] offers a review on corner detectors. One of the most famous is Harris corner detector [59]. The idea is to put a small window around an area. If and only if the point is close to a corner, then when we move in any direction, we should have a large change in intensity. In practice, we compute the Sum of Squared Differences between the small window (which is a copy of the area) and patches around the initial area. Harris detector is invariant to rotation, intensity shifts and scales but not to spatial scale. This issue has been addressed in [103] which searches for the best scale to compare the corners. Other ways to detect corners and which do not use gradient exist like the robust SUSAN method [137] which computes the area of the same color as that of the central pixel.

The line features group relies again on structures like roads [98] or anatomical structures [154], or mathematical lines like contours. Well-known methods like Canny detector [20] are efficient to detect line features. More detectors are described in [167].

The region features group includes structures which have been detected by segmentation algorithms [56]. It also includes the most used feature descriptors. Scale-invariant feature transform (SIFT) [100] uses descriptors containing local information from the image which describes the image as independently as possible regarding scale, viewpoint and luminosity. It is a robust method used in object detection, video tracking, 3D modelling, images stitching and so on and so forth. Speeded Up Robust Features (SURF) [7] is an extension of SIFT which uses Haar wavelets and which is faster to compute. Histogram of oriented gradients (HOG) [26] supposes the local form of an

object can be described by the distribution of the intensity of the gradient or by the direction of the contours. They are computed over a dense grid of cells all over the image, which is one major difference with SIFT. It is particularly efficient to detect people.

The correspondence of control points is done using the euclidean norm in the feature space. Different optimization methods are employed to search among the set of control points, like k-d trees or Best-bin-first [8] search method. Once the feature correspondence has been performed, we have to retrieve the transformation from the pairs of corresponding control points. For linear registration, the problem is well-posed and a least square fit gives the best solution.

The transformation that maps one point set to another given descriptors-driven correspondences is then estimated using either iterative methods [10], robust estimators [19, 141] or multiple hypothesis estimation/validation ones like Ransac [41] and its numerous variants [146]. These methods produce state of the art results in computer vision due to the fact that real-scenes often consist of numerous well identified and discriminative interest points. On the other hand these methods fail miserably in biomedical imaging simply because the identification of reliable interest points is not that evident while at the same time associating discriminative descriptors is problematic. The aforementioned concerns are amplified when considering multimodal imaging.

3.1.2 Image-based Methods

Feature-based methods require enough distinctive objects to work. That is often the case for natural images. For other images which do not contain so many details, like for medical images, image-based methods are necessary. These methods aims at comparing directly areas of the images. One of the limitation is the rectangular window which is most often used for the area. For linear registration which is often associated to a large displacement, it may be difficult to fit the window. Historically, these methods used cross-correlation (CC) [5, 118], Fourier methods to get rid of noise, or mutual information for multimodal images.

When a similarity measure has been defined globally on the whole image, it "only" remains to find the global maximum (or minimum for dissimilarity measure). For linear registration, the maximum number of parameters is 12. Even if it is small, an exhaustive search on a large range of values is computationally too costly. In special cases where the number of degrees of freedom can be drastically reduced, if it is translation only, or if we have specific information regarding the movement [65], exhaustive search is a good solution because it gives the global minimum. Gradient descent methods like

Gauss-Newton or Levenberg-Marquardt [129] are often used for more complex functions. In the paper introducing MI [153], the authors used the classical gradient descent optimization method for mutual information. Marginal Space Learning (MSL) was introduced to detect 3D anatomical structures [163]. The idea is to divide the search space and search for a subset of the transformation parameters one at a time. A comparison of this approach with a full exhaustive search has been made in [164].

Local methods establish a subset of image correspondences (often using local/block matching according to a given metric for a subset of the image domain) and then fit optimally the best linear model using inference techniques earlier presented for the case of geometric methods [110]. Global methods [72] use conventional similarity metrics like SAD, NCC, SSD, MI and seek the parameter set that corresponds to their lowest potential that is often determined through a gradient-driven optimization method. Their main strength is computational efficiency, while their main limitations refer to lack of robustness (sensitive to the initial conditions), lack of modularity (optimization depends on the linear model) and the fact that they can handle only differentiable objective functions. Simplex methods [131], cutting planes methods [84] or more recently discrete optimization ones [166] overcome to some extend the above mentioned limitations while suffering from computational complexity and precision.

Casting computer vision problems as labeling ones through the use of Markov Random Field (MRF) theory has gained great attention since [47] advocated the use of such frameworks for image restoration. The main advantages of this family of optimization methods stem from its efficiency and optimality guarantees. On the other hand, their main disadvantage is their inability to handle efficiently complex interactions between the latent variables.

Despite the fact that discrete optimization spans a great range of applications, it is rarely used to tackle global linear registration; to the best of our knowledge, only [166] has attempted to estimate global transformations with the use of MRFs. This was due to the fact that sampling efficiently the parameter space of linear transformations is not that trivial, while at the same time the connectivity of the graph becomes problematic. Indeed, in theory all image pixels should be connected with the graph representing the linear latent variables.

The registration problem is formulated with the use of the MRF theory. It casts a discrete label assignment problem where the labels correspond to a quantized set of possible displacements. In contrast to [166], the search space is smaller and tractable. Higher-order interactions between the variables are used to allow the encoding of the properties of global linear transformations. Different higher-order constraints are proposed for varying degrees of freedom

of linear transformation models. Our main contribution lies in showing how local second- and third-order constraints can be used to ensure global transformation properties. A dual-decomposition scheme is used to infer the optimal displacements.

3.2 Graphical Model

Given a source image and a target image, we seek to estimate the linear transformation that best aligns the source image to the target. The transformation is parametrized through the use of a deformation grid superimposed onto the source image. The deformation of the source image is governed by the movement of the control points. The goal is to find the displacements of the control points so that the global transformation of the whole image is a linear transformation. To find the displacement of a control point, we compare a patch of the target image centered at this control points with patches of the source image around the corresponding control point.

We use higher order MRFs to formulate our problem in a discrete context. The problem is represented by a hypergraph $G = (V, C)$, where V denotes the set of nodes that encode the latent variables (grid node displacement) and C the set of cliques (or hyperedges) that encode higher-order interactions between the variables. There are no edges in this hypergraph, so E is empty. V forms a grid of control points superimposed to the source image we want to register. Let $L = \{l_1, \dots l_n\}$ be the set of labels, where each element corresponds to a potential displacement vector. The algorithm associates a label l_p to each control point p, so that the final result best aligns the two images. The energy of the MRF can be written as:

$$E_{MRF} = \sum_{p \in V} U_p(l_p) + \sum_{c \in C} H_c(l_c), \tag{3.1}$$

where l_p is the label corresponding to the node p, $U_p(l_p)$ denotes the unary potentials, $l_c = \{l_p, p \in c\}$, is the set of labels assigned to the nodes in the clique or hyperedge c and $H_c(l_c)$ denotes the higher order potentials.

3.2.1 Unary potentials

Let p be a control point and l_p be the label associated to the control point p. Let B_p be the patch corresponding to the center point p in the source image, and B_{l_p} the patch corresponding to the image point p translated by the displacement vector l_p, in the target image. The unary potential quantifies how well the patch B_p matches the patch B_{l_p}:

$$U_p(l_p) = \rho(B_p, B_{l_p}), \tag{3.2}$$

where ρ, is an arbitrary intensity-based similarity measure. Being discrete, the proposed model is gradient-free and can encompass a wide choice of similarity measures, from SAD to statistical measures for multimodal registration like MI [153].

3.2.2 Higher order potentials

In order to constrain the global linear property of the transformation through local constraints, we exploit the property of linear transformations to preserve the barycentre. Let us denote by (p, q, r) three aligned control points in a clique c and (l_p, l_q, l_r) their respective labels. We denote (P) the following condition:

$$\vec{l_p} + \vec{l_r} - 2 * \vec{l_q} = \vec{0}. \tag{3.3}$$

In order to guarantee that the transformation is linear, the violation of the condition (P) must be penalized. Thus, the higher potential is defined as:

$$H_c(l_c) = \Phi(\vec{l_p} + \vec{l_r} - 2 * \vec{l_q}), \tag{3.4}$$

where $\Phi()$ is a cost function defined as:

$$\Phi(x) = \begin{cases} 0 & \text{if } x = 0 \\ \infty & \text{otherwise.} \end{cases} \tag{3.5}$$

The λ-clique potential encodes different kind of constraints depending on the number of degrees of freedom of the global linear transformation we seek to recover. Let us denote the 4 points of the λ-clique as s, t, u, and v (as shown in Fig. 3.1) and $s' = s + l_s$, t', u', and v' their respective images. In all three cases (affine, rigid, and similarity), the condition (P) should hold true for the three points of the diagonal, s, t, and u. Additionally, we have:

For similarity registration: The images of the points s, u, and v should form an isosceles right triangle, with the image of v being the vertex of the right angle,

$$\begin{aligned} H_c(l_c) &= \Phi((\vec{s'} - \vec{v'}).(\vec{u'} - \vec{v'})) \\ &+\Phi(\|(\vec{s'} - \vec{v'})\| - \|(\vec{u'} - \vec{v'})\|). \end{aligned} \tag{3.6}$$

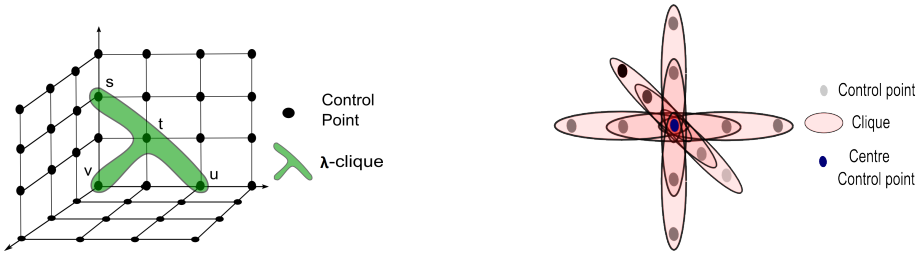
The first part of the equation checks the orthogonality and the second part checks the two sides have the same norm.

For rigid registration: The isosceles triangle formed by the points s, u, and v should have the same size as the one formed by the respective images

of the points. This leads to the definition of the following condition:

$$\begin{aligned} H_c(l_c) &= \Phi((\vec{s'} - \vec{v'}).(\vec{u'} - \vec{v'})) \\ &+\Phi(\|(\vec{s'} - \vec{v'})\| - \|(\vec{u'} - \vec{v'})\|) \\ &+\Phi(\|(\vec{s'} - \vec{v'})\| - \|(\vec{s} - \vec{v})\|). \end{aligned} \tag{3.7}$$

The additional third term penalizes differences in size between the triangle before and after the transformation.



(a) (b)

Figure 3.1: The different cliques in 3D: (a) A λ-clique contains 4 control points forming a λ-shape. There is one λ-clique on each face of the cube. (b) The 9 cliques containing the central control point in blue.

3.2.3 Graph construction

Let us now detail the construction of the graph G. The set of cliques C contains all the triplets consisting of collinear and neighbours points along each dimension of the grid. C also contains λ-cliques which are quadruplets, containing 3 points along a diagonal of the grid and the corner point, such that the 4 points form a T (see Fig. 3.1). The presence of λ-cliques is necessary for the global properties of the linear transformation to hold. The proof is given in 3.2.4.

In 2D, each point belongs to three horizontal cliques and three vertical cliques. The graph contains one λ-clique positioned in one corner.

In 3D, each point belongs to three cliques along each axis (see Fig. 3.1). The grid forms a cube and each face contains one λ-clique, therefore C contains 6 λ-cliques. Each λ-clique is similar to the one defined in the 2D case.

3.2.4 Mathematical Proof

Here, we prove for the 2D case, that if every clique satisfy the condition (P) (including the corresponding points in the λ-clique), then the transformation

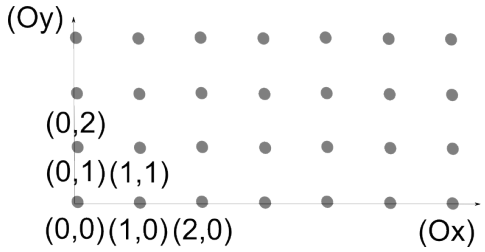


Figure 3.2: The grid in 2D.

is exactly an affine transformation. In the 2D affine case, the constraints we impose are: the condition (P) on every horizontal and vertical clique, and on the diagonal in the λ-clique.

Lemma 3.2.1. *If the images of two control points along a dimension of the grid (horizontally or vertically) are defined, then the images of all the points on the line they form are defined.*

Proof. The condition (P) guarantees that the points in a clique are aligned, and that the middle point is preserved. That means we have one equation for each clique. If the line contains n points, we have $(n-2)$ cliques. When 2 points are fixed, there remains as many equations as free points. The equations are clearly independent (the corresponding matrix is triangular with non-zero value on the diagonal), so the system contains exactly one solution: all the images are defined. □

Theorem 3.2.2. *If the images of three non-aligned points of the grid are defined, then there is one and only one image grid satisfying the condition (P) on all the cliques.*

Proof. Let us suppose those three non-aligned points are the points ((0,0), (0,1), (1,0)) (see fig. 3.2). Let us remind that the affine transformation satisfies the condition (P) on all the cliques, so there is at least one solution. We prove it is the only one by constructing it.
According to the lemma, the images of the points on the x-axis (similarly on the y-axis) are defined by the image points of (0,0) and (0,1) (similarly (0,0) and (1,0)). Let us now consider the three points in the λ-clique ((2,0), (1,1), (0,2)). Since the images of the points (2,0) and (0,2) are known the middle point is defined. This middle point (1,1) along with the point (0,1) ((1,0) similarly) define the images of all the points on the lines (y=1) ((x=1) similarly) thanks to the lemma.
We now have two image points for all the remaining horizontal and vertical lines, so we know the whole image grid. □

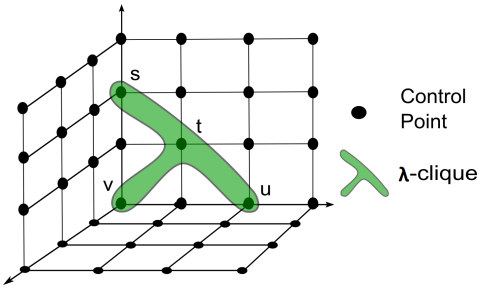


Figure 3.3: A λ-clique contains 4 control points forming a T-shape. There is one λ-clique on each face of the cube.

By definition our constraints are less strong than the ones of an affine transformation. The theorem demonstrates that our grid construction has as many degrees of freedom as an affine transformation, i.e. 3 points can be independently chosen in 2D, so our constraints define an affine transformation of the grid.

3.2.4.1 Extension to 3D

In 3D, the same idea holds, but we have an additional degree of freedom. To be precise, only four λ-cliques are enough if they are well chosen, even if we mentioned six λ-cliques in the method. The four λ-cliques are positioned on four different faces of the cube such that three of the λ-cliques share a common point (the point called v in the λ-cliques as shown in fig. 3.3), the origin (0,0,0) of the grid for example. If we have the image of the four points ((0,0,0), (0,0,1), (0,1,0), and (1,0,0), it defines a unique affine transformation in 3D. We construct the unique grid with the images of those points. As in 2D, the three faces sharing the origin are automatically defined. The fourth face, share at least two orthogonal edges with the three first faces. Again, we have all the conditions of the 2D case, so the fourth face is defined. Now, two opposed faces of the cube are defined, so each line orthogonal to those faces contains two imposed points. The lemma applies to all those lines and so the 3D grid is formed.

3.2.4.2 Extension to similarity and rigid case

Similarity is a subset of affine transformations. We have to eliminate shears and make sure the scaling factor is the same along the different axes. Those are the two condition defined by the λ-clique in the similarity case, by the right isosceles triangle. Rigid transformations is a subset of similarity without homothetic transforms, which corresponds to the last additional constraint.

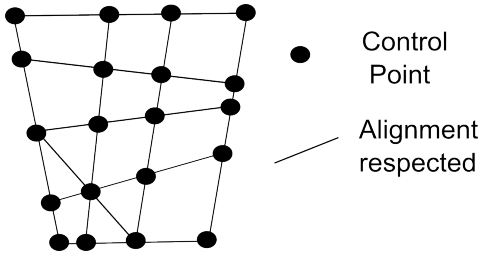


Figure 3.4: A potential grid transformation with only the alignment constraint on all the hyperedges defined.

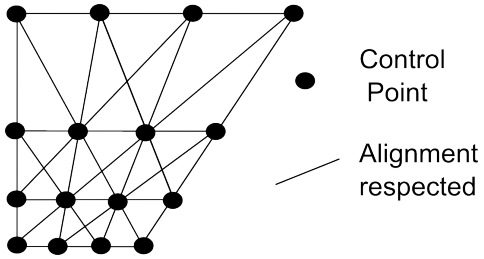


Figure 3.5: A potential grid transformation with only the alignment constraint on many triplets.

3.2.4.3 Some counterexamples

For a better understanding of the need of all the previous elements, we exhibit here some counterexamples of grid deformations when we remove one of those elements. If we remove the distance constraint on all the triplets, the problem gains many degrees of freedom and a 2D grid may be deformed like this 3.4. Increasing the number of hyperedges will not solve this problem as projective transformations follow this constraint too 3.5. Without the λ-clique, other cases may appear like 3.6, which is not a projective transformation.

3.2.5 Extension to the projective case

There are two different ways to extend our framework to projective transformations. Let us remind two properties of homographies or projective transformations. First the image of a line, a plane, or a space is a line, a plane or a space respectively. Second the cross-ratio or anharmonic ratio is preserved. The cross-ratio of 4 distinct and aligned points A, B, C, and D is defined as:

$$r = \frac{AC \times BD}{BC \times AD}. \tag{3.8}$$

If we only want to use the preservation of lines, then we must use a huge number of cliques. Indeed, as we saw previously 3.6, the current number of

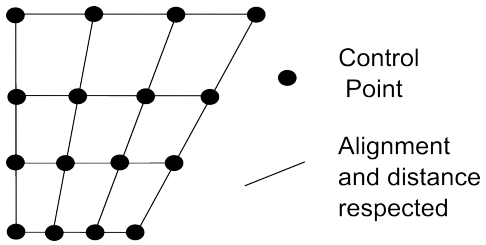


Figure 3.6: A potential grid transformation without the λ-clique constraint.

hyperedges was not sufficient to prevent other transformations. So we would need many cliques that will make the optimization a lot more difficult.

If we use the cross-ratio, we can design a very similar graph to the one before. The hyperedges now contain 4 aligned points and we need a diagonal hyperedge, similar to the λ-clique. By the same constructive proof, we can show we can only accept projective transformations. But the problem, in our framework, comes from the label set. Indeed, a discrete label set does not allow projective transforms which are not affine, as long as we keep a hard constraint on the hyperedge potential.

With a continuous label set, this new graph is able to handle projective transforms. Using a continuous label set would require different solvers but is possible.

3.3 Optimization algorithm

To solve the MRF, we use the Dual Decomposition [92]. The main idea consists in decomposing a global difficult problem into smaller solvable subproblems (referred to as slaves) and then extracting a solution by cleverly combining the solutions from these subproblems. In this case, the difficulty of the inference of the optimization displacements lies in the presence of the higher order cliques. Here, the graph is decomposed into trees that constitute the set of subproblems. The only requirement for the choice of the subproblems is that they cover (at least once) every node and hyperedge of the hypergraph G.

3.3.1 Dual Decomposition algorithm

Decomposition is an old [28] and general approach [11] to solve a problem by splitting it into smaller subproblems and solving them separately either consecutively or simultaneously. The original motivation was to solve very large problems which were too complex for other techniques. For example decomposition methods can also handle problems which could not be solv-

able for memory reasons, it can speed up via parallel computation or solve problems in a distributed way. Separable problem is the most trivial case of decomposition method because there is no need to recombine the solutions of the subproblems. The expectation-maximisation algorithm is a decomposition of the parameters and the log-likelihood. Schur complement method is also a decomposition method in the sense that the method can solve the initial equation by only computing inverses of smaller matrices.

We will consider the following problem with the coupling variable y:

$$\min_{x_i,y} \sum_i f_i(x_i, y). \tag{3.9}$$

It is equivalent to the master problem:

$$\min_{y} \sum_i \phi_i(y), \tag{3.10}$$

where the two subproblems are $\phi_i(y) = min_{x_i} f_i(x_i, y)$. This is a primal decomposition of the problem and can be solved by subgradient methods, cutting-plane, or Newton method if ϕ_i is differentiable for all i.

Dual decomposition is similar to primal decomposition but requires to introduce a Lagrangian. Let us now consider the problem:

$$\min_{x} \sum_i f_i(x). \tag{3.11}$$

We then introduce auxiliary variables x_i to decouple the problem into this equivalent version:

$$\min_{x,x_i} \sum_i f_i(x_i). \tag{3.12}$$

$$\text{such that } x_i = x \tag{3.13}$$

Now we form the dual problem. The Lagrangian is:

$$g(\lambda_i) = \min_{x_i,x} \sum_i f_i(x_i) + \sum_i \lambda_i(x_i - x), \tag{3.14}$$

where $\{\lambda_i\}$ is the set of lagrangian multipliers. As we search the maximum of this function over the set of lagrangian multipliers, we deduce we must have $\sum_i \lambda_i = 0$. Therefore we can rewrite the dual as:

$$g(\lambda_i) = \min_{x_i} \sum_i f_i(x_i) + \lambda_i(x_i). \tag{3.15}$$

Finally we call the following decoupled equation of the dual problem the master:

$$\max_{\sum_i \lambda_i = 0} g(\lambda_i) = \sum_i g_i(\lambda_i), \tag{3.16}$$

and those are the subproblems, or slave subproblems:

$$g_i(\lambda_i) = \min_{x_i} f_i(x_i) + \lambda_i \times x_i. \tag{3.17}$$

The master problem is convex (in the variable λ_i) and is solved using a projected subgradient method. The optimization procedure iterates between solving the slaves and sending the solutions to the master, which updates the slaves parameters according to $\lambda_i = Proj(\lambda_i + \alpha_t \nabla g_i(\lambda_i))$, where t is the iteration, α_t a predefined positive step parameter, ∇ is a projected subgradient, and $Proj()$ indicates a projection over the feasible set $\sum_i \lambda_i = 0$.

3.3.2 Choice of the Slave Decomposition

The strategy employed to combine the solutions of the different subproblems is of great importance. It should be able to handle conflicting suggestions from the slaves for the same node. As we showed before, a master problem acts as a coordinator between the slave subproblems. It gathers the solutions of all subproblems and adds penalties based on the levels of agreement for every node. For example, when two slaves A and B disagree about a node N, i.e. they proposed respectively the labels l_a and l_b for this node N, a penalty is added to the choice they made, and the choice the other slave made is awarded in the cost function: the cost function of the slave A is increased by a coefficient α if it chooses the label l_a for the node N again, and decreased by the same amount if it chooses label l_b, so that the slave A will explore new solutions and, most probably, include the label l_b the other slave B suggested.

In our case, a slave problem is defined for each line parallel to a coordinate axis, and a slave for each T-clique. An example of the different slaves in 2D is illustrated in fig. 3.7. Every clique and each node is then part of a slave so we recover the hypergraph. One slave is a tree, the others are chains, so it is very easy and fast to solve. One additional advantage of this optimization method is the independence of the slave problems, which allows a parallelization of the computation.

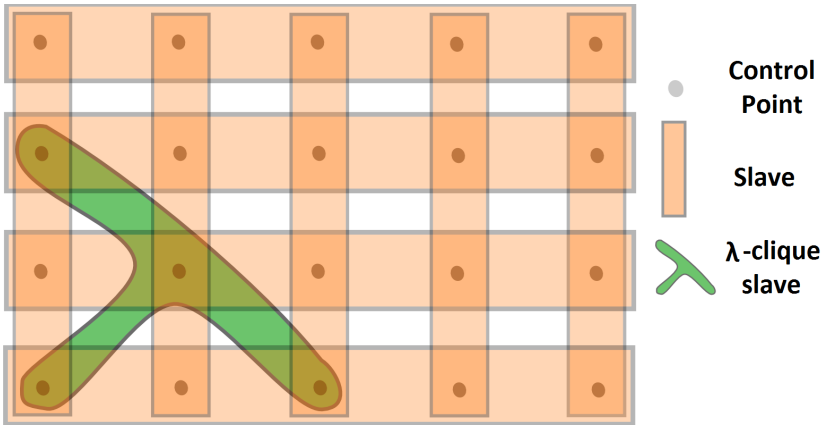


Figure 3.7: The decomposition of the original problem in slave subproblems in 2D: one for each line, one for each T-clique.

3.4 Experimental Validation

3.4.1 Implementation Details

We use a multiresolution registration in a coarse-to-fine manner. The different iterations are used to refine the label space. The deformation grid, which is the result of the algorithm, is reset after each iteration and the resulting displacement fields are incrementally composed on the image. We use a strict constraint on the higher order terms, so, at each iteration, the resulting displacement is exactly among our space search (affine, similarity or rigid registration in our different experiments). This ensures, by composition, that the final displacement is exactly affine, rigid or a similarity.

The grid size contains 3^n points, where n is the dimension of the space, and might be increased, up to 5^n in our experiments. We started with few control points because we expected large deformations. The label set is composed of 5^n displacement vectors. These displacement vectors are regularly spaced on a grid centered at the 0-displacement vector. The maximal length of the displacement vectors is 0.4 multiplied by the distance between two control points along each axis. This length is iteratively reduced, such that first iterations account for large displacements while latter iterations account for small displacements. We used up to 8 iterations in our experiments.

Our algorithm is implemented in C++. The results were obtained using a 64 bits machine with a Intel Xeon W3670 processor and 16 Go of RAM.

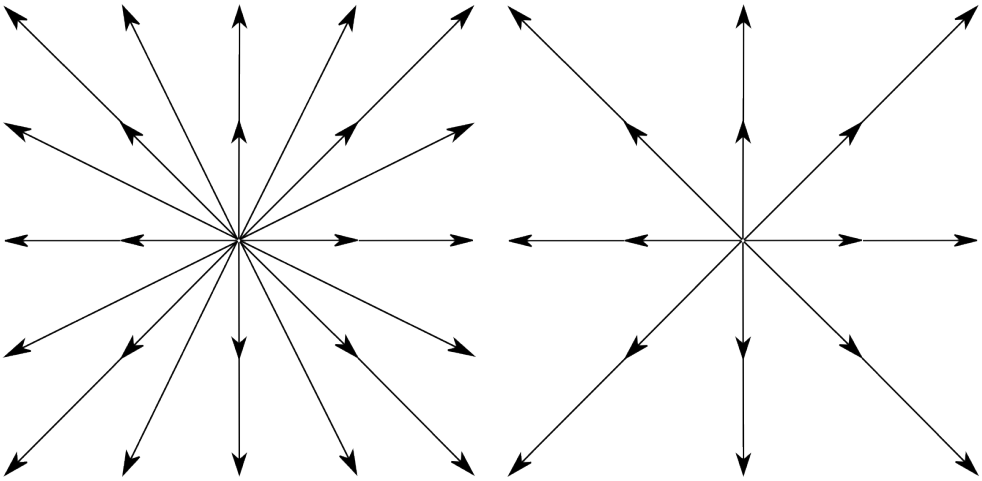


Figure 3.8: On the left, the dense label set in 2D contains all the discretized labels on a square. On the right, the sparse label set in 2D contains a subset of the dense label set: only vectors on main axes and main diagonals.

3.4.1.1 Choice of the Label Set

The choice of the label set is important. It must allow a large number of transformations in the space we require. There are two usual label sets used in deformable registration, one is called dense, the other is called sparse. They are both defined as the discretization of a small area. The dense set contains all the discrete vectors contained in a square in 2D or in a cube in 3D. The sparse set contains a subset of the precedent one, with only the vectors on the main axes and on the main diagonals. Both sets are shown in Fig. 3.4.1.1.

We want to allow as many transformations as possible. For translations, there is no constraint as all the control points would require the same displacement vector. For skews, we only need displacement vectors along a line so we have enough labels in both label sets. But for scales and rotations the sparse labels do not contain enough displacement vectors. To be precise, there is no scale or rotation allowed by the sparse set (except the identity). So we used, and recommend to use the dense label set if we want to seek an affine transformation or a similarity.

For rigid registration, we have another additional constraint which may be a problem if the images have different dimensions along the different axes. Let us remind the control grid is superimposed on the moving image. If the moving image is, for example, 10x20 centimetres, then the movement of the points along the second dimension are larger (there is a higher range) than those on the first dimension because the second dimension is bigger. If the dense label set contains enough displacement vectors, then some solutions exist. But there

is an easy way to increase the number of possibilities with a smaller label set: we keep the dense framework but the size of the rectangle in which we define the displacement vectors depends on the size of the image (see an example on Fig. 3.9). For our 20x10cm image, we take the vectors on a (b)x($2b$) rectangle, where b depends on the iteration of the registration as usual, an example is shown in Fig. 3.10. We have now an adapted label set for rigid registration. We will use this one in our experiments on rigid registration. Of course it can be used for affine registration, but is not necessary. It did not show improvements on the dense label set for the affine registration experiments, so we used the usual dense label set.

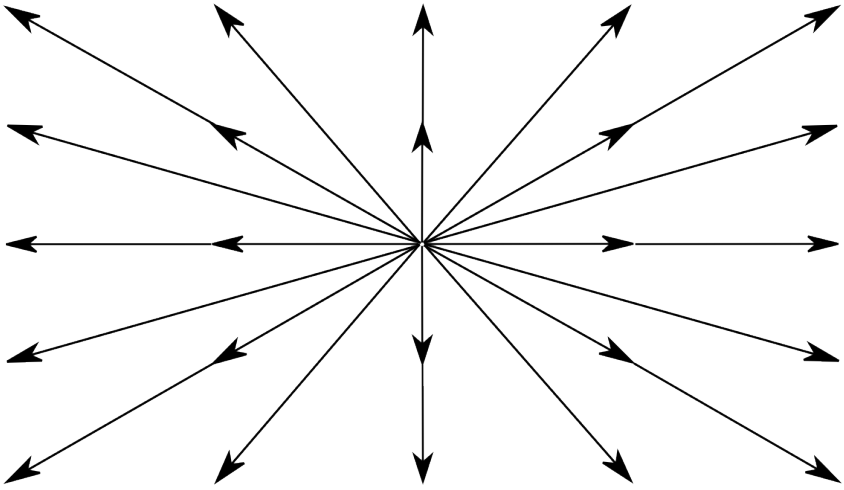


Figure 3.9: The rigid label set we used for rigid registration is very similar to the dense label set but the proportions of the square are similar to those of the image to allow more transformations.

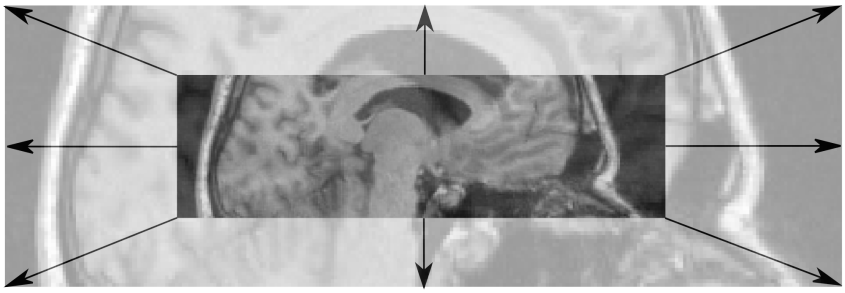


Figure 3.10: To rescale this rectangular image, the horizontal displacement vectors are longer than the vertical displacement vectors.

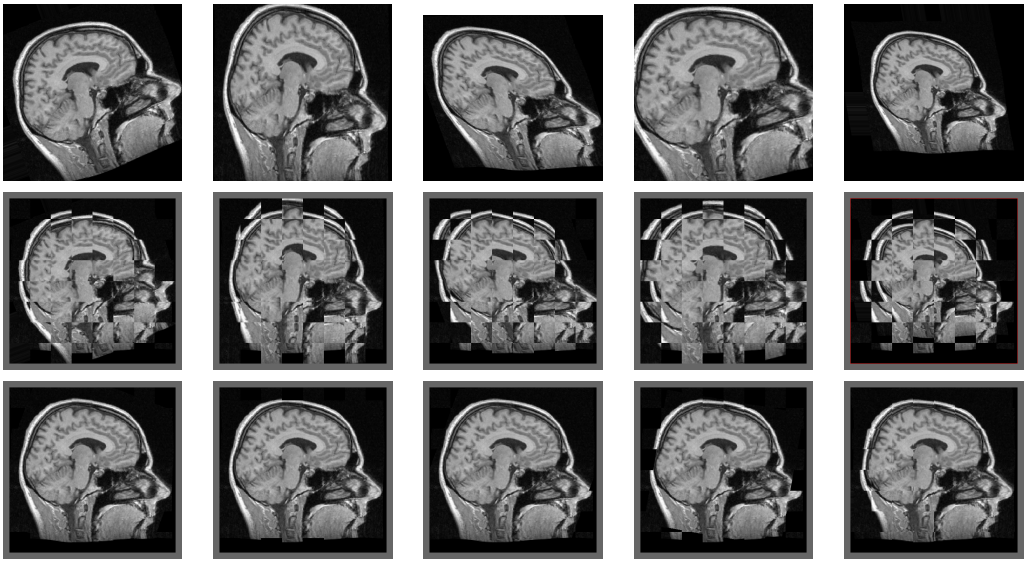


Figure 3.11: The different transformations are, from left to right: a 20 degrees rotation, a scale along the y-axis only, a shear, a combination of a 10 degrees rotation with a 120% scale and a combination of a shear with a 80% scale. First row shows the deformed image, second row the comparison with the target image, and last row the comparison between the result and the target images.

3.4.2 2D Study

We tested our method on a huge set of simulated data. We perform affine transformations of a head image. We then register this deformed image to the original image. The transformation parameters are uniformly drawn from certain ranges. Rotations lies between 0 and 20 $^\circ$; scales are chosen among 80, 90, 100, 110 and 120% ; translations reach 30mm and the shearing angle reach $\pm$ 15 $^\circ$. These transformation parameters will be denoted P_1. We will also denote P_2 the subset of parameters without shears, and P_3 the subset of P_2 without scales. P_1 is a set of 5625 affine transformations, P_2 a set of 1125 similarities, and P_3 a set of 225 rigid transformations. We used the Sum of Absolute Difference (SAD) similarity measure. The average runtime is about 10 seconds. On many images, a part was missing due to the transformation, mostly to the scale factor. Results are shown in table 3.1 and some examples are shown in Fig. 3.11.

Parameters set	SAD	Affine	Similarity	Rigid
P_1	44.58	13.04	-	-
P_2	43.39	8.23	9.97	-
P_3	37.23	6.57	8.57	8.77

Table 3.1: Results of the 2D study on a large sample of transformations of a head image. Second column shows the global SAD between the two images before registration, the others columns show the SAD after registration for the different frameworks we tested (affine, similarity and rigid cases).

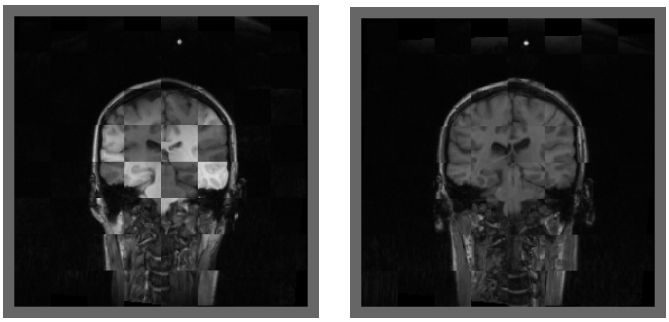


Figure 3.12: Two examples of registration, the source and the target are superimposed

3.4.3 IBSR dataset

We test our method on 3D images using manual segmentations. Seven MRI images of the brain are registered, and for all of them, manual expert segmentation are available. The data is part of the Internet Brain Segmentation Repository (IBSR), provided by the Center for Morphometric Analysis at Massachusetts General Hospital (available at http://www.cma.mgh.harvard.edu/ibsr). The T1-weighted images have been positionally normalized into the Talairach orientation (rotation only). We selected one data set as the template and register it to the remaining six data sets. Two examples of the registration are given in 3.4.3. The transformation found by our method is then used to warp the corresponding segmentation, which is finally compared to the segmentation of the template, using DICE score, the sensitivity and the specificity, given in Fig. 4.

3.4.4 Comparison with MedInria

First, we used an annotated 3D MR-T1 (Magnetic Resonance images produced by spin-lattice relaxation) calf muscle dataset in order to compare the proposed method against the publicly available image analysis MedInria soft-

Image	DICE	Sensitivity	Specificity
Brain 1	93,4477	92,0727	99,5325
Brain 2	88,8486	85,8105	99,3105
Brain 3	90,9922	79,5661	97,0576
Brain 4	85,9011	88,8448	98,3108
Brain 5	89,4162	86,6715	99,3257
Brain 6	80,4629	80,3937	98,1771
Average	88,1781	85,5599	98,6190

Table 3.2: Results of the brain registration evaluated on manual segmentations. Given is the DICE score, the sensitivity and the specificity.

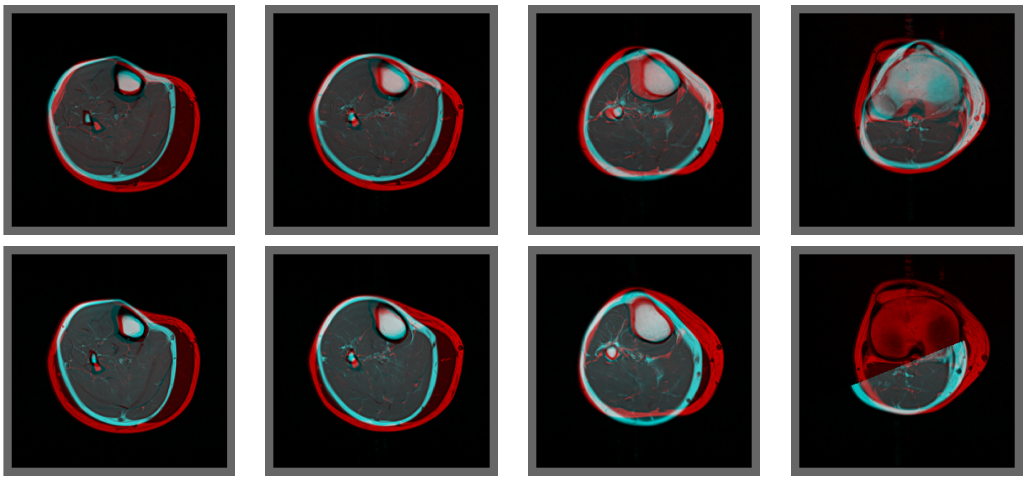


Figure 3.13: Comparison of a registration, first ligne is our algorithm, second line is MedInria algorithm. Different slices of the same image are shown on different columns. Our registration is more accurate.

Image	DICE	Sensitivity	Specificity
Before registration	70	71	95
Our registration	81	84	96
MedInria registration	71	73	95

Table 3.3: Results of the affine registration evaluated on manual segmentations of calf muscles MRI. Given is the DICE score, the sensitivity and the specificity.

ware. We register interpatient images. We report the DICE, the sensitivity and the specificity of the segmentations, before and after registration. We used the algorithm named Baloo [109], for affine registration, with the SSD similarity measure, with 5 iterations, 1 "coarsest pyramid level" and 1 "finest level".

The results are shown in table 3.3 and Fig. 3.14 compares the repartition of the DICE for the two methods. One visual comparison is showed in Fig. 3.4.4. Those results show that our algorithm performs better than MedInria registration in this dataset. The execution time was about 3 seconds for MedInria, and about 8 seconds for our algorithm.

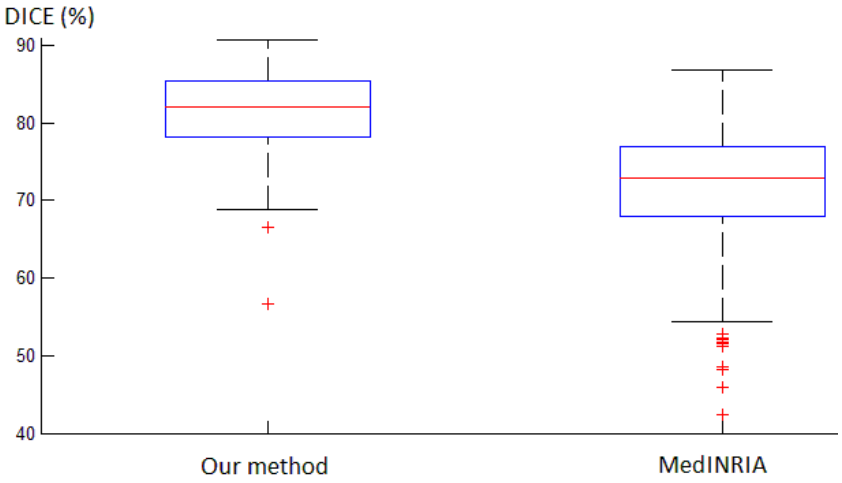


Figure 3.14: Comparison of the DICE between our method and MedInria software. The proposed algorithm is more accurate than MedInria.

3.4.5 RIRE database

We test our method performing a series of 3D multimodal registrations of brain images (CT (Computed Tomography), MR-PD (Proton Density), MR-

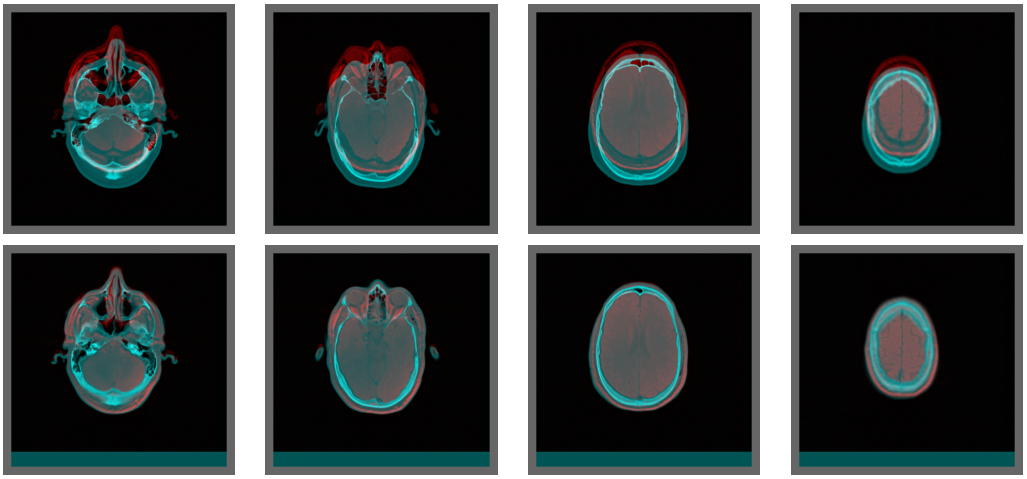


Figure 3.15: Qualitative evaluation of registration. Source and target images are shown using different color bands. Axial views before (top) and after (bottom) registration are shown. The images have been successfully aligned.

T1, and MR-T2 (spin-spin relaxation)) provided by *The Retrospective Image Registration Evaluation Project* (RIRE)[1]. The CT images have a resolution of $512 \times 512 \times 29$ and a physical voxel size of $0.65 \times 0.65 \times 4$ mm, while the MR images have a resolution of $256 \times 256 \times 26$ and a voxel size of about $1.25 \times 1.25 \times 4$ mm. We performed CT to MR registrations for 7 patients (Patient001 to Patient007, with a total of 21 registrations). We employed a histogram-based estimation of Normalized Mutual Information (NMI) using 32 bins as the similarity measure. We show in Fig. 3.4.5 a slice of the 3D volumes before and after registration. We also show in table 3.5 different results obtained with our algorithm for different multimodal similarity measures.

We compare our results with FLIRT (results provided by [104]), Elastix [82], Simplex [131], and another MRF method [166] in table 3.4. This last paper reports the results and implementation details for Simplex, Elastix and Zikic's method. The accuracy of the registration is evaluated by computing the Fiducial Registration Error (FRE) for 10 fiducial points. All results are available on the RIRE website. The running time for the proposed approach, when using NMI, was approximately 2 minutes. The other methods run on our machine in approximately 5 minutes. Let us note that we report here the previously published results for the other methods, and not the results that we obtained when running them. This is because we were not able to

[1] Available at: http://www.insight-journal.org/rire/

	Mean error				
Modality	Simplex	Elastix	Zikic	FLIRT	Ours
PD	3.889	4.174	3.901	2.267	2.180
T1	1.124	1.186	1.157	1.656	4.255
T2	1.931	1.903	1.724	2.227	1.923

Table 3.4: Results of the 3D multimodal CT-MR registrations on patient data of the RIRE database. We report the mean FRE (in mm) after registration using the Simplex, Elastix, Zikic MRF, FLIRT, and our method.

	Mean error		
Modality	NMI(32 bins)	NMI(16 bins)	MI(32 bins)
PD	2.180	2.258	2.260
T1	4.255	4.030	2.247
T2	1.806	4.419	2.250

Table 3.5: Results obtained with the proposed method for different metrics, and different number of histograms bins used for the metric estimation. The adopted similarity measure, NMI with 32 bins, is indeed the most suitable measure to our problem.

accurately reproduce them.

In their article, Zikic et al. [166] left out one image (P03 with MR-PD) because none of the methods could register it. We include it back for comparison. Our method is competitive with state of the art methods in terms of quality of results and speed. Our score is poor when registering the MR-T1 images because we significantly fail to align one image, which considerably increases the mean error.

3.4.6 Clinical Application

The registration of 3D ultrasound with MRI is a challenging task because of the differences of those two modalities. But it is also a standard medical procedure to have a pre-operative MRI, which gives a precise image of the body, and intra-operative US, because this is the only kind of imagery that can be done during a surgery. The better the registration is, the more accurate the surgeon will be able to operate the patient.

To perform a fast and accurate registration during the intervention, surgeons usually use magnetic sensors, but their use require time, space and precision. To remove these limitations, we propose here a fully automatic

framework to perform the registration.

For this specific couple of modalities, classic similarity measures like Sum of Squared Differences, Normalized Mutual Information or Correlation Ratio are not adapted. Some measures have been designed for specific organs, then they lack modularity. A promising and generic measure was developed in [122]. This measure compares US to both the MRI intensity and its gradient. In [156] a measure called Linear Correlation of Linear Combination (LC^2 or LC2) is designed using a fine understanding of the US and CT measures. We explained this method in 2.5.1.

We applied our method to a medical pipeline. First a global positioning of the ultrasound on the MRI space is performed. In the general case, an exhaustive search can be performed, but most of the time some a priori knowledge can, and should be used to accelerate the search. For example, in a database of liver ultrasounds, the positions of the patients are very similar. So the global positioning can focus on a subarea of the whole space. Secondly, a rigid registration using our framework developed in [34] is used. Before the whole procedure, a mask is generated on the fly on the ultrasound image to take only into account the voxels which have information and to remove the useless black background of the image. Those experiments were done on a Intel Core i7-4790 CPU with 32 Go of RAM. To evaluate our results, we manually created fiducial points on the two images and we compute the Target Registration Error (TRE). All distances are in mm.

3.4.6.1 MRI and CT dataset

We first tested our algorithm on a database of 6 pairs of images, MRI and CT of the same patient. We used the Normalized Mutual Information measure because it is more adapted to this couple of modalities, so we could test the algorithm independently of the LC2 measure. The images had approximately the same orientation but the dimensions were greatly different, so that the translations required to register the images reached 200mm in one dimension. So we design the global positioning to test a wide range of translations but no rotations. Our search area for each axis was [-200,+200] with a spacing of 20mm. The mean execution time was about 30 seconds. Then we did the rigid local registration using the MRF model. We did 13 iterations and began with a 30mm search range. The search range is decreased at each iteration by a factor of 0.67. The mean execution time was about 100 seconds. We show the results of those two steps in the table 3.6 and some visual results in Fig. 3.4.6.1. Four points were chosen to do the ground truth. We show the standard deviation σ in both cases.

Table 3.6: TRE in mm after global positioning and local rigid registration, and the standard deviation at the end of the algorithm.

Patient	Global	Local	σ
P1	12	9	4.9
P2	22	9	2.2
P3	15	11	4.2
P4	12	10	2.3
P5	13	9	2.4
P6	26	15	5.9

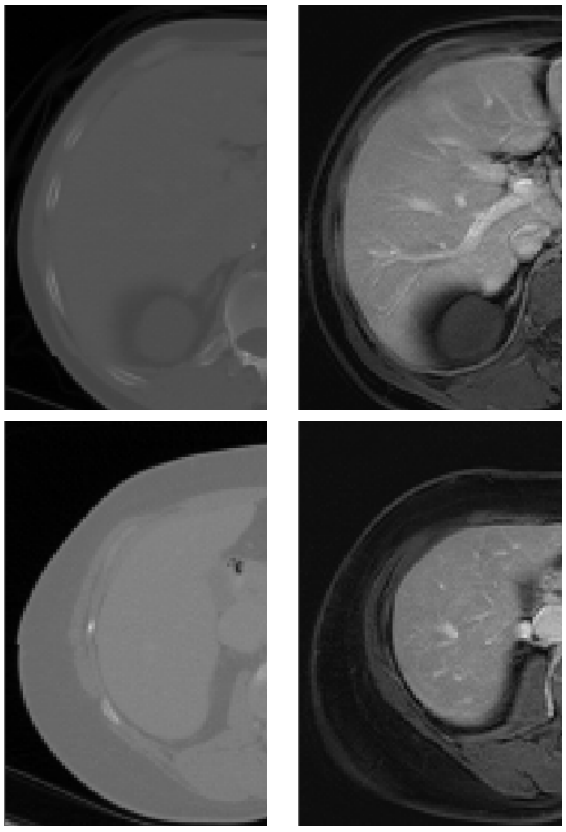


Figure 3.16: Comparison of CT, on the left, and MRI, on the right, of two different patients after the registration.

Table 3.7: TRE after global positioning.

Patient	Global
P1	14
P2	29
P3	54
P4	59
P5	27
P6	19
P7	44
P8	4
P9	44

3.4.6.2 US and CT images

We then tested our algorithm on a database of 9 pairs of US and CT images. We used the LC2 measure as described earlier. The images have totally different orientation. We used a search range around the ground truth of 20mm in every axis and we varied the orthonormal basis along two angles between -35 and +35 degrees. We removed the ground truth from the search area. We show the resulting TRE after the global positioning in table 3.7, which shows it was accurate enough to start the rigid algorithm, and an example of registration in Fig. 3.4.6.2. The execution time was about 400 seconds.

We then tested the rigid registration with the parameters like we did with the previous dataset. We also compared the two measure NMI and LC2. As expected, the NMI measure performed very bad. As for LC2, we found two high TRE for patients P1 and P3: visually, the results are good but are different from the ground truth. It happened because the two images are taken at two opposite moments of the breath cycle, and our result and the ground truth did not focused on the same area. The quality of the registrations is not very precise because of the quality of the US images, which presented many artefacts as on Fig. 3.18.

We tested our discrete MRF formulation to solve the problem of multimodal registration of US and MRI images. The global positioning does not take too much time as long as images are downsampled and as the search area is limited. The local registration shows good results that would be improved by deformable registration, but the quality of the images may prevent such a more precise registration.

Table 3.8: TRE in mm before and after rigid registration.

Patient	Initialization	LC2	NMI
P1	10	54	40
P2	17	9	17
P3	20	37	52
P4	33	11	41
P5	13	13	13
P6	35	5	39
P7	23	10	22
P8	26	9	42
P9	24	7	33

Figure 3.17: Example of registration, the source ultrasound on top, the deformed image in the middle, and the target at bottom.

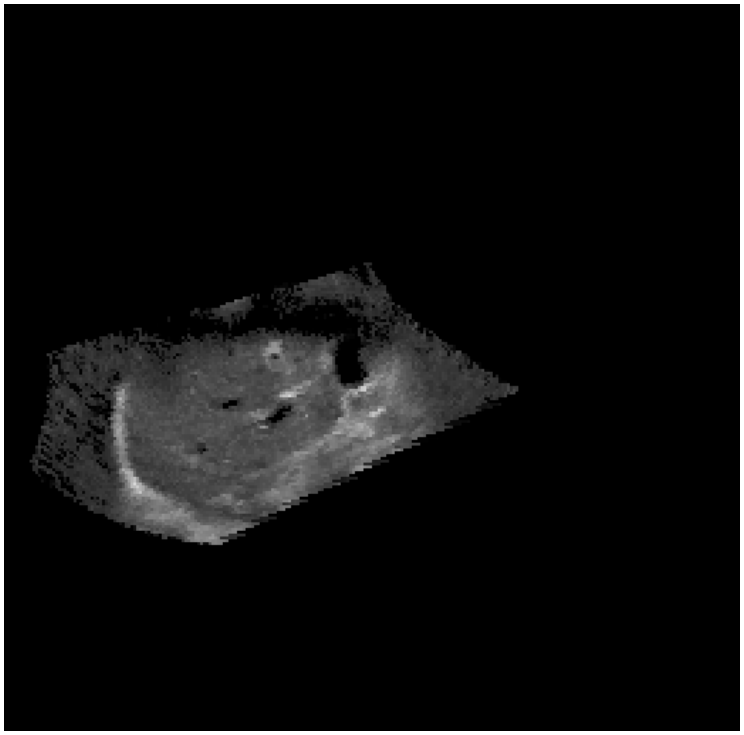


Figure 3.18: One of the ultrasound of the database. The quality limits the accuracy of the registration.

3.5 Discussion

In this chapter, we have presented a discrete MRF formulation to solve the problem of linear registration, using a local higher order graphical model with hidden variables being the displacement vectors of the control points. The proposed formulation is modular with respect to the similarity criterion and the nature of the linear transformation (rigid, similarity, affine and could be extended to projective). Moreover, it is computationally efficient by means of its relative local nature and the designed search space. The performance of the method on several data set along with comparisons with state of the art methods demonstrate the interest of our approach.

We tried to extend our concept to 3D point cloud. The first problem is that we lose the grid structure. Implicitly, our grid structure allows to reduce the dimensionality of the problem on the hyperedges. Indeed, as the points are aligned on the grid, we only have to verify that the barycentre is preserved in this one dimensional subspace. Hence we only require 3 points. For a point cloud, the points are randomly distributed and then we cannot use the same trick as before. Thus, the condition we have to check is the 3D placement of the points of the hyperedge. Mathematically, the whole transformation is

encoded in the displacement of the points of the hyperedge. We then need 5 points in each hyperedge. The purpose of our problem can only be an approximation now, so we use a soft constraint that checks if the movements of the 5 points is coherent, and we create enough hyperedges to transmit the information to the whole point cloud. In practice, the computation on the cliques takes a longer time, and there is many more cliques that makes the convergence nearly impossible. So we abandoned this potential extension.

With this framework, we are able to perform linear registration with a Markov Random Field. In the next chapter, we will find a way to mix a deformable model with our linear model to solve complex registration problems.

Chapter 4

Combining Linear and Deformable Registration

4.1 Introduction

We mentioned previously the importance of registration. The previous chapter was focusing on linear registration. This one will focus on deformable registration. It is a much more complex problem, and usually in medical imaging, there is no real truth to evaluate the deformable registration as even extremely accurate registrations by professionals do not include the inner transformations of tissues and organs.

In this section, we classify the different algorithms according to the transformation model chosen, then the type of similarity criterion (described in 2.5.1 and 3.1), and finally we briefly discuss about the different optimisation strategies. For a more complete review, see [138].

4.1.1 Transformation Model

There are different constraints that may be imposed on the displacement of the image. The two main categories are realistic models, which have or try to have a physical or an anatomical mean, and those which are inspired by interpolation theory. According to [105], the physical category can be subdivided into five: elastic model, viscous fluid flow model, diffusion model, curvature registration, and flows of diffeomorphisms.

Elastic model describes materials elastic objects. It uses the Navier-Cauchy equation. The main assumptions are small deformations and a linear dependency between stress and strain. This model is used for many engineering materials like structural analysis and is often solved with finite elements. The second subcategory models the image as a viscous fluid, and the associated equation is Navier-Stokes equation. This equation models the behaviour of most of the fluids: Newtonian fluids to be precise, those whose stress is linear in the strain rate. This model was introduced in medical imaging in [22, 23]. Diffusion model relies on the diffusion equation. It was introduced in [143] and inspired by Maxwell's demon, which was a thought experiment to break

the second law of thermodynamics. The demon apply forces on the membrane where the image is not constant. The membrane is one-way. It is an iterative process of small displacements. Regularization is based on Gaussian smoothing. The forces are usually computed using the SSD measure but NMI has been encoded in this framework in [147]. Curvature was introduced by Fischer and Modersitzki [38]. It does not penalize affine transformations. Finally, flows of diffeomorphisms is based on the Lagrange transport equation [23,148].

Interpolated deformation models interpolate the displacement of the whole image using only the movement of a small set of voxels of the image. This small set of points are called control points. They may be placed randomly on the image [134] or regularly on a grid like for FFD we developed in 2.5.2.2.

Of course, this problem is mathematically ill-posed so a regularization is paramount to have a proper framework.

4.1.2 Optimisation Strategy

The optimisation strategies are generally split into two groups: continuous and discrete. Continuous methods use the space as it was continuous while discrete methods relies on a subset of values to avoid gradient strong irregularities due to artefacts, local extrema or simply to reduce the computation. Sometimes the border is fuzzy because continuous methods are applied for problems with only a discrete set of measures, for example if we consider the pixellisation of images, then interpolation is used to consider the image space to be complete and regular.

Continuous methods usually relies on the gradient. That is the strong point of continuous function. Most methods rely on an update rules:

$$\theta_{i+1} = \theta_i + \alpha_i g_i(\theta_i), \tag{4.1}$$

where i is the iteration parameter, θ is the vector of parameters i.e. the unknown, and g_i is the search direction. The most common approaches are gradient-descent, conjugate gradient, Powell's method [116], Newton-type methods, Levenberg-Marquardt [33], Stochastic gradient descent [36].

Continuous methods are sensitive to local extrema and initial conditions. Moreover, they are not modular regarding the similarity criterion and the transformation model. Discrete methods overcome these limitations but they lack precision because the solution space is sampled. The more it is samples, the better accuracy is achieved, but also the most time it takes to compute. We already detailed several discrete algorithms in 2.4. Let us add that techniques like Fast-PD and Tree Reweighted Sequential [86] are based on Linear Programming. Instead of trying the original NP-hard problem, they use a LP relaxation.

4.2 Graphical Model

The main idea to solve this problem was to use the higher order cliques defined in the previous chapter and a classical MRF graph that performs deformable registration. So the basic formulation is the same as before, but the structure of the graph is not. We remind shortly the formulation.

We model the simultaneous linear and deformable registration problem through a hypergraph $G = (V, E, C)$, where V denotes the set of nodes, E the set of edges, and C the set of higher order cliques. Let $L = \{l_1, \dots l_n\}$ be the set of labels that corresponds to a quantized version of the solution space, and l_p denotes the label assigned to node p. The algorithm associates a label l_p to each control point p, in order to align the two images. The energy of the MRF can be written as:

$$E_{MRF} = \sum_{p \in V} U_p(l_p) + \sum_{(p,q) \in E} V_{p,q}(l_p, l_q) + \sum_{c \in C} H_c(l_c) \ , \tag{4.2}$$

where $U_p(l_p)$ denotes the unary potentials, $V_{p,q}(l_p, l_q)$ denotes the binary potentials, $l_c = \{l_p, p \in c\}$, is the set of labels assigned to the nodes in the clique c and $H_c(l_c)$ denotes the higher order potentials.

We want to separate and determine simultaneously the linear and deformable transformations. To do that, we decompose the displacement of each control point into a linear and a deformable part. A possible way is to put the two part of the displacement into the same node as before. This would keep the same number of nodes, but the label set would be the square of what it was before, and thus may be intractable. Instead, in order to infuse that knowledge into our MRF framework, we will duplicate (see Fig. 4.1) the original grid of control points: the first part will encode the linear displacements, the second part will encode the deformable displacements. Therefore, each control point is represented by two corresponding nodes of the graph, one representing its linear displacement, the other one representing its deformable displacement. That duplication allows to keep a reasonable number of nodes (twice as many as before) and the same number of labels, and therefore to greatly decrease the computational cost of performing inference in the graph. A similar decomposition was used in [133] for the x- and y- axes. We applied that idea for the linear and non-linear parts of the displacement here. Let us note $V1$ the nodes in the first part of the duplicated graph (linear part), and $V2$ those in the second part (deformable part):

$$V = V1 \cup V2. \tag{4.3}$$

For clarity, we will first explain the structure of the graph and then detail the values of the potentials. For the same reason, the diagrams in this part

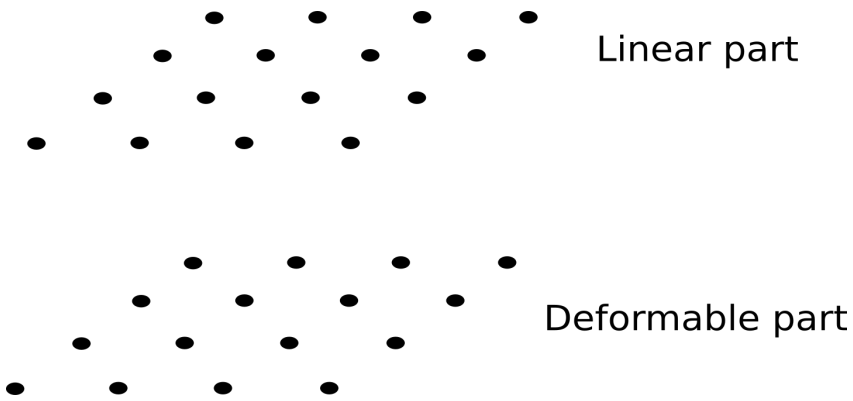


Figure 4.1: The graph contains two parts, each point of the grid is duplicated into two nodes, one in each part. We show here a 2D example for ease of visualization.

are for 2D-2D registration. The extension to 3D-3D is quite obvious and the written explanations apply to both cases.

4.2.1 Graph Construction

Each control point on the images correspond to two nodes, so the original grid is duplicated. The idea is to have the information about the linear transformation in $V1$, hence we will find there the hyperedges (triplets and λ-clique) defined in 3, and the deformable transformation encoded in $V2$, consequently we will find the usual grid with smoothing terms. A major difference is the data term. Indeed, we need to know the total displacement of a control point to evaluate the similarity measure, so we need to know both the linear displacement and the deformable displacement. The data term is now evaluated as a pairwise function linking a pair of duplicated nodes. Finally, we must remember that a deformation like a global translation is not penalized by smoothing terms, because this kind of displacement is allowed. However we want the total linear displacement to be in the linear part so we must create a penalty in the deformable part, such that the algorithm would always rather put a translation into the linear part than in the deformable part. We put a little penalty over the length of the deformable displacement in the deformable part $V2$. The complete graph is shown here 4.2.

Let us now define C, the set of cliques. In our framework, the cliques have very different goals. Those in $V2$ ensure the smoothness of the deformable displacements. So there is an edge between between each pair of neighbour nodes, which form a grid as used for computing deformable displacement alone as in [127]. Moreover, each node in $V2$ have an unary potential so deforming

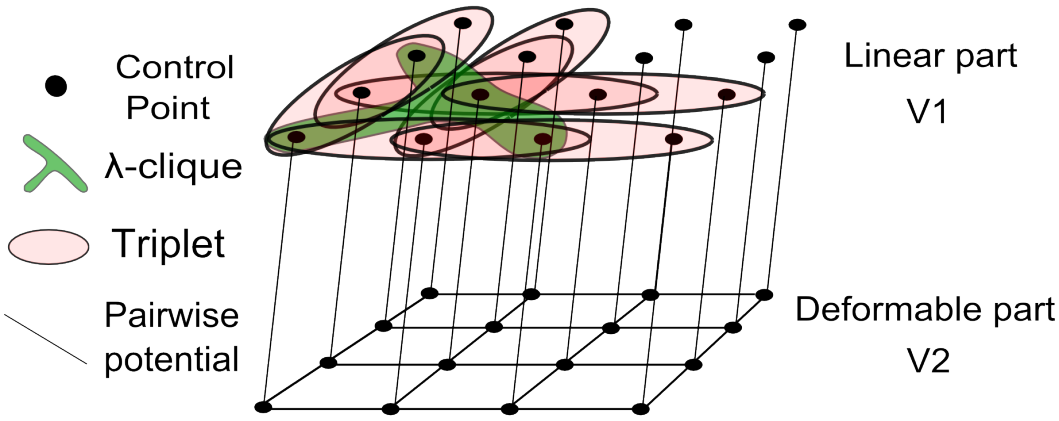


Figure 4.2: The complete graphical model for simultaneous linear and deformable registration in 2D (some triplets are omitted for clarity). Pairwise potentials are represented by black links connecting graph nodes; graph nodes that are related by triplets are grouped by ellipses, while the λ-clique is shown in green.

linearly the image in this part of the graph is penalized. The cliques in $V1$ ensures that the linear displacements of all the points form a coherent linear transformation of the image. What is left to be detailed is the data term. The data term should capture the interactions between pairs of linear and deformable displacements so each pair of duplicated nodes (one in $V1$, one in $V2$) will be linked by an edge.

4.2.1.1 Unary Potentials

To ensure the algorithm prefers large linear displacements instead of large deformable ones, we employ a unary potential penalizing the norm of the vector of the displacement vector.

$$U_p(l_p) = \|l_p\|. \tag{4.4}$$

This potential is defined for every node in $V2$, where l_p corresponds to a deformable displacement.

4.2.1.2 Pairwise Potentials

Regularization Term

A regularization term operating between nodes in $V2$ is necessary in order to ensure the deformable registration is smooth. This can be achieved by

penalizing the vector differences between neighboring nodes:

$$V_{p,q}(l_p, l_q) = \frac{\|q - p - l_p + l_q\|}{\|q - p\|} \ , \tag{4.5}$$

where p and q represent two neighbour control points, both in $V2$.

Data Term

In order to quantify the alignment of the two images, we employ a patch-based similarity criterion, or we compare a patch from the source image $B_{p,q}$ with a patch in the target domain $B_{l_p+l_q}$ that is chosen based on the composition of the affine and deformable part of the deformation. In order to model the composition of the two parts, it is necessary to use a pairwise term involving the corresponding nodes p and q in the two parts of the graph. The data term is defined as:

$$V_{p,q} = \rho(B_{p,q} B_{l_p+l_q}) \ . \tag{4.6}$$

4.2.1.3 Higher Order Terms

The higher order potentials are defined as in Chapter 3. Triplets and λ-clique ensure the linearity of the transformation. An example of those cliques for a 2D grid for a 2D registration is shown in Fig. 4.2.

4.2.2 Optimization algorithm

To solve the MRF, we use DD-ADMM [1]. Dual Decomposition [92] consists in decomposing a global difficult problem into smaller solvable subproblems (referred to as slaves) and then extracting a solution by cleverly combining the solutions from these subproblems. DD-ADMM is an improvement of Dual Decomposition which accelerates the convergence. In this case, the difficulty of the inference of the optimization displacements lies in the presence of the higher order cliques. Here, the graph is decomposed into trees that constitute the set of subproblems and into a grid. The only requirement for the choice of the subproblems is that they cover (at least once) every node and hyperedge of the hypergraph G.

Formerly, we had tried several algorithms to perform the optimisation like Loopy Belief Propagation and Tree Reweighted Belief Propagation. The hard constraint made it very difficult for those usual algorithms. And the factor graph was huge because of the higher order hyperedges. The label set was the square of what it was before and the execution time reached several days for Belief Propagation for example.

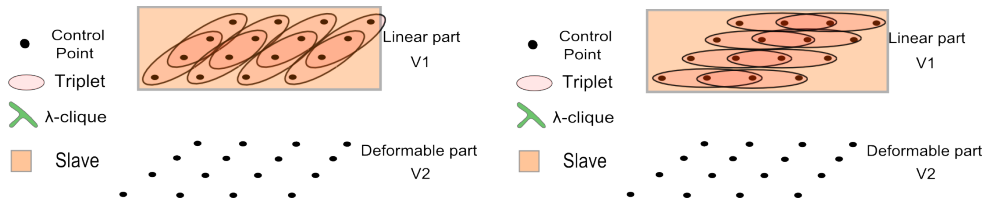


Figure 4.3: Two different slaves in 2D. One for horizontal and vertical lines in the affine part. In 3D another slave would be required for the z-axis, for a total of 5 slaves..

4.2.2.1 Alternating Descent by Method of Multipliers Algorithm

ADMM replaces the Maximum A Posteriori computation by a quadratic problem, at each iteration, so it penalises strongly any disagreement. Compared to DD, the Augmented Lagrangian with a quadratic problem is:

$$g(\lambda_i) = \min_{x_i,x} \sum_i f_i(x_i) + \sum_i \lambda_i(x_i - x) + \frac{\eta}{2} \sum_i \|x_i - x\|^2, \tag{4.7}$$

where η is a weighting factor. The traditional resolution method to compute the augmented lagrangian alternate between the joint minimisation of the function and an update of the multipliers [60, 117]. [1] proposes to use the Alternating direction method of multipliers [45, 53], which replaces the joint maximisation by a single Gauss-Seidel step.

The main difference is each slave must compute a quadratic problem, and not a maximum a posteriori. [1] gives some mathematical guarantees but we have lost many compared to DD. Principally, we have lost the certitude that, when an agreement is reached among the slaves, we are sure it is the global optimum for DD. It is not the case any more for ADMM. The article also give insight on how to compute the quadratic subproblems. In our case, it reduces to adding a quadratic penalty to the unary terms which is different at each iteration but which does not change fundamentally the optimizers.

4.2.2.2 Choice of the Slaves Decomposition

In our case, in $V1$, a slave problem is defined for each line parallel to a coordinate axis in the linear part of the grid. In 2D we have the two slaves shown in 4.2.2.2. Thoses slaves contain parralel chains of triplets which are consequently independent from each other. They may be easily solved by higher order message passing or exhaustive search. We used exhaustive search where the computation is greatly accelerated by precomputation by removal of impossible combinations.

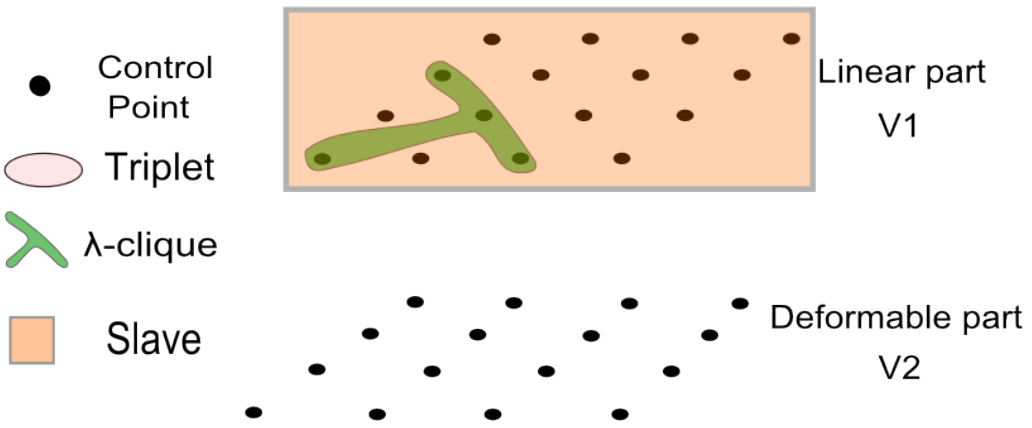


Figure 4.4: The slave containing the λ-clique in 2D.

There is also one slave for the λ-cliques 4.4. In 2D, there is only one clique so exhaustive search is fast. In 3D, there are four λ-cliques. Three of them share nodes but exhaustive search is still the fastest way to solve we found. By intelligently separating the constraints, we reduced the slave to explore the combination on six nodes forming a triangle.

The last slave contains all the nodes of the graph but only the edges of $V2$ and the edges encoding the data term. In this big slave, those edges contain one node (in $V1$) link to nothing else. Thus we can send a message from it to the other end of the edge, like in Message-Passing algorithm, to put all needed information into the unary of the node in $V2$. Then we have a simple slave and we optimize it using the Fast-PD algorithm [93]. With this decomposition (4 in 2D, 5 in 3D), DD-ADMM uses a master-slave framework similar to Dual Decomposition: the master send solvable subproblems to the slaves, which handle the subproblems and send back a solution 4.6.

4.3 Experimental Validation

4.3.1 Implementation Details

The algorithm uses an iterative coarse-to-fine refinement process. The quality of the image is reduced at the first steps to accelerate the computation. The label space is successively refined to explore a large number of displacements while keeping a reasonable execution time. The label space corresponds to a discretization of potential displacement vectors, regularly distributed on a grid around the 0-displacement vector. The maximal length of the displacement

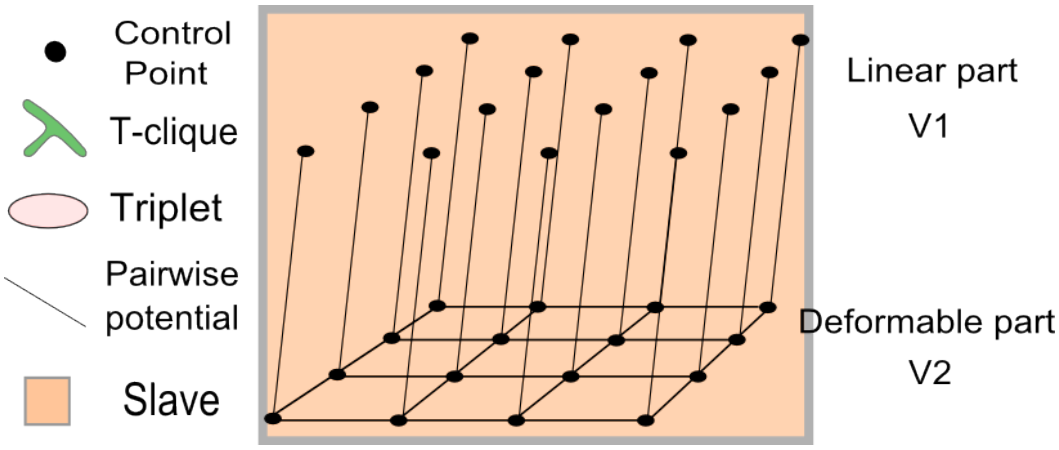


Figure 4.5: The deformable slave contains all the pairwise potentials of the graph. We send a message from the nodes only linked by an edge (containing the data term), then we can solve the problem using Fast-PD because the other pairwise, the smoothing term, is submodular.

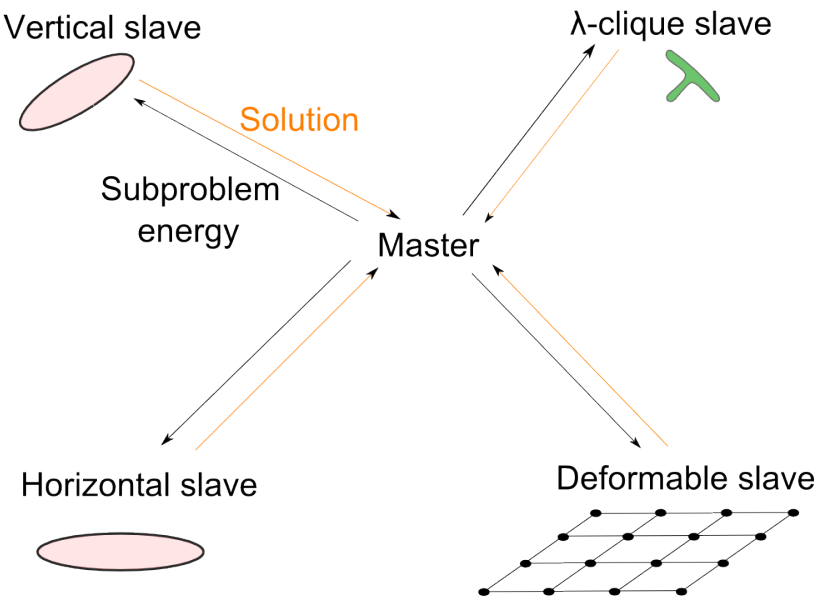


Figure 4.6: The master-slave framework for our simultaneous problem in 2D. There are four slaves (five in 3D). The master gives a subproblem and each slave answers with a solution.

vectors is 0.4 multiplied by the distance between two control points along each axis. The length is iteratively reduced along the iterations. We used up to 7 iterations in our experiments. The successive label space refinement allows to keep the number of labels quite small, 3^3 or 5^3, while reaching sub-millimetre registration accuracy. The grid contains 3^3 control points at the first iterations and is increased to 9^3.

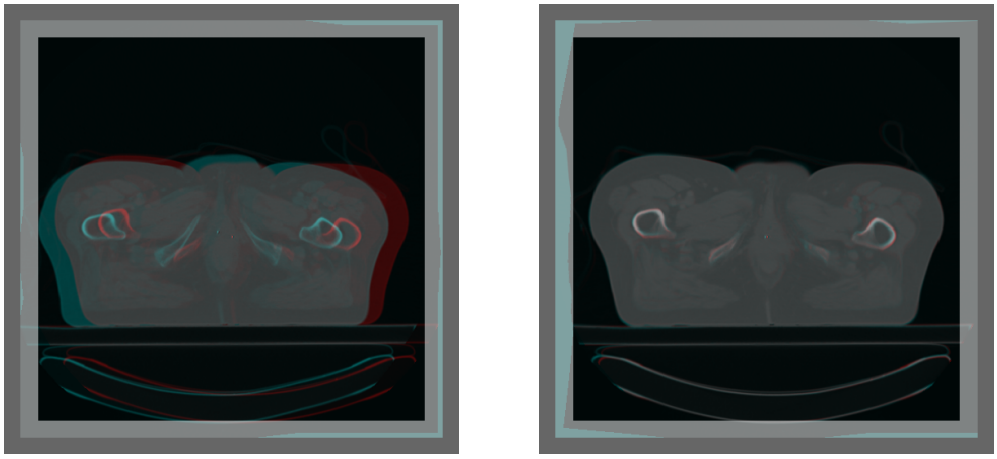
The algorithm is implemented in C++. The tests were performed on a 64 bits machine with a Intel Xeon W3670 processor and 16 Go of RAM. The mean running time for 3D volumes was about 160 seconds when using the similarity criterion SAD.

4.3.2 Affine Transformed Images

We use a database of abdomen 3D CT images, containing 6 images of the same patient at different moments. Two organs have been manually segmented by medical doctors, the sigmoid and the bladder. The image dimension is about 512*512*121 with a physical spacing of 0.92*0.92*4 mm, with small variations on the images. We perform several affine transformations of one image. We then applied a small deformation field to the transformed image. This deformation field is small in the sense it should not contain any global linear transformation. We then try to register these deformed images to the original one. Rotations lies between 0° and 5° and translations reach 20mm. We performed 22 different transformations, with a Sum of Absolute Differences (SAD) similarity measure. We want to compare the affine transformation we find with the one we initially applied. So we fixed 6 points in the images at some extremities of the bodies, and compute the mean distance between the two transformations. Our results show a mean distance of 2.61 mm. Most of the error come from rotations which are not captured by the data term. The results could be improved by using a rotation invariant measure. One example of registration is shown in Fig. 4.7.

4.3.3 Real Images

We then use intra-patient images from the same database to compare our method with a sequential linear and deformable registration. So images are initially aligned with a linear registration. Then we apply a deformable registration algorithm, DROP [52]. In parallel, we apply our algorithm. We compare the DICE we get from the two methods. Our results show a small improvement (cf. 4.1) of the DICE.



(a) (b)

Figure 4.7: A registration, the two images are superimposed in different colours: (a) Before registration. (b) After registration.

DICE	Bladder	Sigmoid
Before registration	45.61	39.383
Consecutive registration	78.15	68.55
Our registration	78.47	68.64

Table 4.1: Results of the DICE of two organs while comparing our simultaneous registration with a consecutive one.

4.4 Discussion

In this paper, we have a discrete MRF formulation to solve the problems of linear and deformable registrations simultaneously, using a local higher order graphical model acting with hidden variables being the displacement vectors as labels. The proposed formulation can deal with arbitrary similarity criterion, it is modular with respect to the nature of the linear transformation and exhibits computational efficiency due to its relative local nature and the designed search space. We use a multilevel strategy, in a coarse-to-fine manner. The performance of the method on 3D multi-modal medical data along with comparisons with state of the art methods demonstrate its potential for applications. Opposed to the usual sequential linear/deformable registration, our scheme is based on a sound mathematical framework, even if our experiments do not show the bias. This bias may be a problem for other kind of images or problem, for which our algorithm would reach its true potential. Moreover this approach is fast compared to state of the art methods.

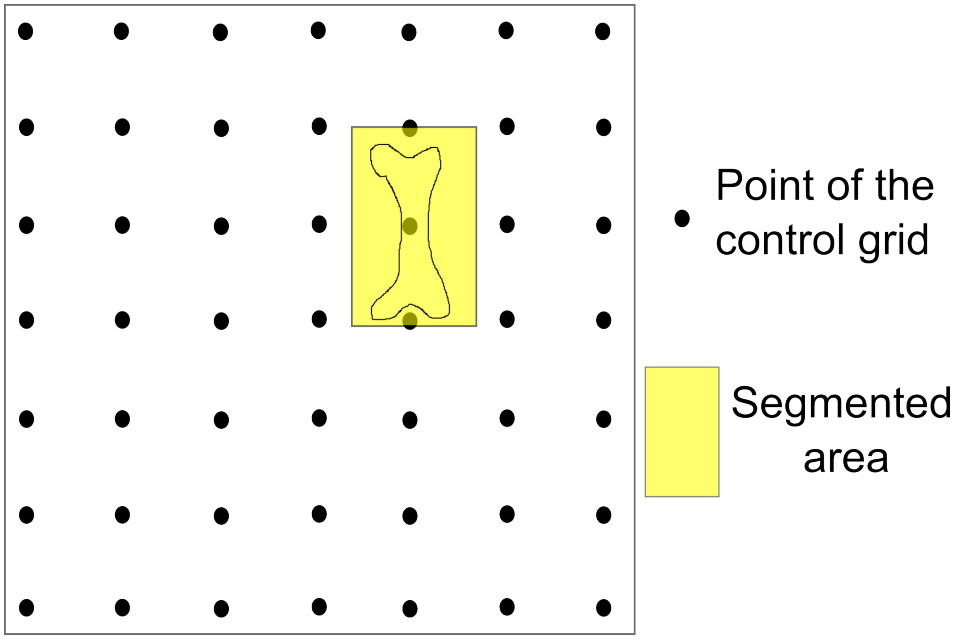


Figure 4.8: The object is roughly segmented, in the yellow area, and the grid of control points is superimposed on the image.

We also showed it was possible to combine linear and deformable registration and explored ways to do the optimisation. This opens a new range of possibilities of applications in 2D/3D registration in the next chapter or to separate linear and deformable registration in the same image as we briefly explain here.

4.4.1 Linear Within Deformable Registration

Now we present another way to combine linear and deformable registration. Another matter of interest for medical applications is to perform a global deformable registration with a subimage which should be rigidly or affinely deformed. Indeed, we know that the bones structures, for example, cannot be deformed (unless they are broken, so it is true most of the time). To improve the registration or to avoid artefacts, we could integrate this knowledge in the process. It can be easily performed with the linear structure we constructed and the classical deformable grid. For the sake of clarity, the images shown represent a 2D/2D registration. The algorithm extends very simply for 3D/3D registration.

We create a rough segmentation of the object whose transformation is linear. This area contains the higher order hyperedges necessary to the linear registration, we call it the linear area. This area do not need the smoothing

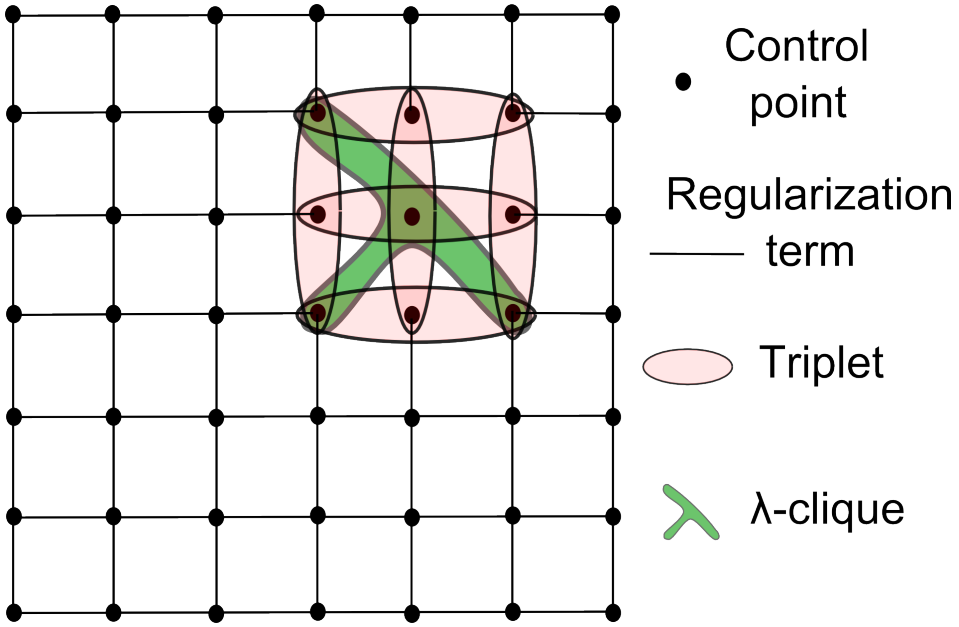


Figure 4.9: The graph contains a sub-part inspired by the linear registration.

pairwise between the inner nodes because the regularization is taken into account by the higher order terms. But we need a regularization between the outer nodes of the linear area and the rest of the graph: the pairwise regularization. The rest of the graph only contains smoothing pairwise terms between the nodes. The new graph is shown in Fig. 4.9

To compute the displacement of the whole image, we use interpolation. Those in the linear area are interpolated using linear transformation, because the displacement of this part of the image is linear. The rest of the points are interpolated using B-splines.

The optimization can be performed with the same strategy we used previously: the grid with Fast-PD or message passing algorithm, the hyperedges with exhaustive search, and a master-slave decomposition to assemble those parts.

Chapter 5

2D-3D registration

5.1 Introduction

2D/3D image registration is used in many clinical applications such as image-guided surgery, when a pre-operative 3D image and an intra-operative 2D view are to be merged to guide the surgeons. While there has been much work on 3D/3D and 2D/2D registrations, 2D/3D has received attention more recently [102]. There are different kinds of 2D/3D registrations, like the registration of a 2D image to a 3D shape [25], the registration of a projective 2D image (for example for X-Ray), or of a sliced image (for example when ultrasound are used). We will focus on the latter case in this paper. Searching the in-plane deformations simultaneously is necessary because the internal structures move due to breathing, beating heart and the medical intervention itself. These create large elastic deformations and make the registration extremely challenging.

5.2 Related work

A variety of methods has been proposed to deal with this problem. Many intensity-based and gradient-based methods use local iterative optimizations techniques such as gradient descent [44] or best neighbour search method [94] but they are very dependent on the initial position. [27] presents a feature based method that performs slice to volume registration, using several slices in order to improve the quality of the results. [51] tracks intra-operative MRI slices of prostate images with a pre-operative MRI volume. This monomodal registration (MRI intra-operative slices to MRI pre-operative volume) is designed to provide patient tracking information for prostate biopsy performed under MR guidance. [32] proposes a method to register endoscopic and laparoscopic ultrasound images with pre-operative computed tomography volumes that potentially could work in real time. It is based on a new phase correlation technique called LEPART and it manages only rigid registration. Phase correlation was used in [32] reporting quasi real time results on laparoscopic ultrasound to CT images registration. [108] presents a flexible framework for

intensity based slice-to-volume non-rigid registration algorithms that was used to register histological sections images to MRI of the human brain.

Casting computer vision problems as labelling ones through the use of Markov Random Field (MRF) theory has gained great attention since [48]. [37] presents a MRF framework in high dimension to which we compare our results.

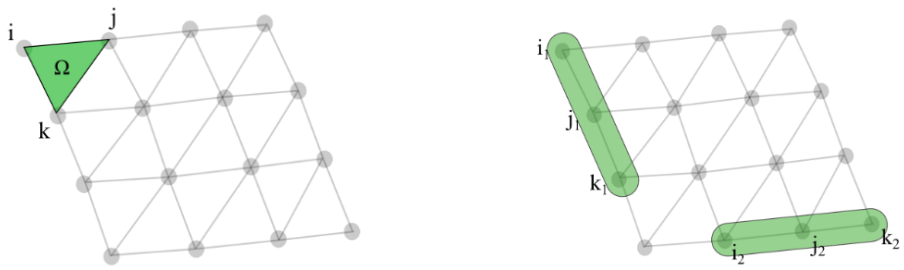
Our framework aims at reducing the dimensionality of the search space in a local pairwise discrete MRF. So we represent a physical control point into two nodes of the MRF graph, one representing the search of the 2D space, the other the in-plane deformations. So the number of labels is reduced and fastens the optimization algorithm. The framework is intensity based and independent of the metric that is being used, so it can be adapted to different image modalities or new measures [112].

5.3 Graphical Model

Given a 2D source image I and a 3D target volume J, we want to find the slice $\hat{\pi}[J]$ of the volume J that best matches the image I. In-plane deformations can affect the image I, those deformations are modeled by a 2D deformation field $\hat{T}_D$ that is inferred during the optimization process. It is possible to consider more general deformations, i.e. 3D deformations, but for our applications, the medical doctors were interested in showing the deformed 2D image and the chosen slice, hence our choice.

Our discrete formulation of the problem uses an undirected graph $G = (V, C)$. V is the set of nodes which corresponds to the control points superimposed on the 2D image. C is the set of cliques or hyperedges and is composed of two parts $C = C_D \cup C_R$. The nodes represent the control points of a quasi-planar grid that models the in-plane deformations and the current best selected plane simultaneously. We used a quasi-planar model because it allows more flexibility and produces better results than an exactly planar plane selection. The deformations are modeled in a Free Form Deformation (FFD) framework. The label set L corresponds to a set of 3D displacement vectors $L = \{l_i = (d_x, d_y, d_z)\}_i$, indicating the position of the control point in the 3D space (the plane selection is deduced from the position of the displaced control points).

Cliques in C_D are triplets of vertices with a triangular shape (see in Fig. 5.3) and they are associated with the data terms. The cliques in C_R are horizontal and vertical cliques of three aligned points (see in Fig. 5.3), similar to those defined in the previous section. The control points are displaced by assigning them different labels until an optimal position is found. Let us now define the corresponding high order functions.



(a) (b)

Figure 5.1: The different type of cliques: (a) The green area Ω is one of the triangular cliques $(i, j, k) \in C_D$ and is used to compute the data term. (b) Vertical clique $(i_1, j_1, k_1) \in C_R$ and horizontal clique $(i_2, j_2, k_2) \in C_R$ examples. They are used to regularize the grid and give a quasi-planar solution.

The energy to minimize regroups data terms D_{ijk} associated to triangular triplets in C_D and regularization terms R_{ijk} associated to triplets in C_R. The first ones encode the matching cost while the later ones act as regularizers: to ensure the control points keep a regular and quasi-planar grid structure. The energy of the MRF is thus defined as:

$$E_{MRF} = \sum_{(i,j,k)\in C_D} D_{ijk}(l_i, l_j, l_k) + \beta \sum_{(i,j,k)\in C_R} R_{ijk}(l_i, l_j, l_k), \tag{5.1}$$

where l_i is the label associated to the node i, and β is a weighting parameter balancing the contributions of the similarity and geometric terms.

5.3.1 Data Term

The data term is defined over a set of triangular cliques as in Fig. 5.3. Its formulation is independent of the similarity measure ρ and is computed for each clique $d = (i, j, k) \in C_D$ using the 2D source 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 ρ, the data term associated with the clique d is thus defined as:

$$D_{ijk}(l_i, l_j, l_k) = \sum_{\Omega_d} \rho(I(x), \pi_d[J](x))dx, \tag{5.2}$$

where $x \in \Omega_d$ and Ω_d corresponds to the triangular area defined by the control points of clique d over the plane $\pi_d[J]$, after applying the corresponding displacement vectors (identified by the labels $l_i, l_j, and l_k$) to the control points.

5.3.2 Regularization Term

5.3.2.1 Planar Consistency

Planar consistency can be easily enforced through a discrete approximation of the second-order derivatives of the grid. Here, the condition is similar as previously in Chapter 3, but because of the soft constraint, as soon as the second iteration, the grid may not be exactly aligned, so we have to conserve the current position of the control points in our equation. The condition is thus:

$$(p_i + l_i) + (p_k + l_k) - 2 \times (p_j + l_j) = \vec{0}. \tag{5.3}$$

Based on this idea, we define the following energy term using the euclidean norm of the resulting vector, normalized with the original distance between the control points d:

$$R^A_{ijk}(l_i, l_j, l_k) = \frac{\|(p_i + l_i) + (p_k + l_k) - 2 \times (p_j + l_j)\|^2}{d^2} \tag{5.4}$$

5.3.2.2 In-plane Deformation Smoothness

As usual with FFD, we require a distance preserving approach. We impose the original distances between the control points of the grid is preserved during the whole algorithm. Let us define ϕ_{ij} as the ratio of the euclidean distance between the displaced points p_i+l_i and p_j+l_j and the original distance between their initial positions $p_{o,i}$ and $p_{o,j}$, which gives the following equation:

$$\phi_{ij} = \frac{\|(p_i + l_i) - (p_j + l_j)\|}{\|p_{o,i} - p_{o,j}\|}. \tag{5.5}$$

We need our regularizer to fulfill two conditions. First, we need it to be symmetric with respect to the displacement of the points, i.e. to penalize with the same cost when the control points are closer or more distant. This is achieved by posing:

$$\Phi_{ij} = (1 - \phi_{ij})^2 + (1 - \frac{1}{\phi_{ij}})^2 \tag{5.6}$$

Second, we need the energy to be zero when the points are preserving distances and bigger than zero otherwise. The precedent expression fulfils both conditions for a couple of control points, so for a given clique (i, j, k), we define the second part of the regularizer as:

$$R^B_{ijk}(l_i, l_j, l_k) = \frac{\Phi_{ij} + \Phi_{jk}}{2} \tag{5.7}$$

The equation that regularizes our grid is a combination of both parts

$$R_{ijk}(l_i, l_j, l_k) = (1 - \alpha)R^A_{ijk}(l_i, l_j, l_k) + \alpha \times R^B_{ijk}(l_i, l_j, l_k) \tag{5.8}$$

where α is a weighting factor. The proposed objective function can be either optimized through its mapping to a factor graph or using higher order optimization methods like dual decomposition. We have adopted the factor graph approach and implemented it using OpenGM2 library [78].

5.4 Experiments

Despite the reasonable dimension of the label space, the use of factor graphs limits the number of labels that can be used during inference. In order to overcome this limitation we adopt a pyramidal approach where grid size is decremented in every pyramid level. For each grid resolution, some iterations of the registration algorithm are performed. We choose the optimal set for each one and update the control point positions while varying the size of the displacement vectors that form the label space to improve the search space sampling.

It should be noted that the planarity constraint is imposed in a soft manner. Furthermore, due to the incremental approach, errors can be accumulated and lead to inconsistent planar transformations. Therefore, regularization terms do not guarantee that the final solution is a plane. In order for the grid to be a plane, we project every control point to the regression plane estimated from the current position of these points. This projection corresponds to a 2D FFD that gives a good approximation of the deformation field.

Evaluation and comparison of the method was performed over the 2D/3D monomodal MRI heart dataset presented in [37]. It consists of 10 different temporal series of 20 bidimensional slices each one, that must be registered with a volumetric MRI, giving a total of 200 registration cases. For a complete description of the dataset refer to [37]. Plan estimation was evaluated measuring the error between the ground truth and the estimated planes (planes were represented by a 6-DOF rigid transformation, with 3 translation and 3 rotation parameters $(T_x, T_y, T_z, R_x, R_y, R_z)$).

The average error among all the registration cases is less than 0.0057rad $(0,3265^\circ)$ for rotation and less than 0.5386mm for translation parameters (see Figure 5.2). Given that image resolution in z axis is lower than in x and y, a bigger error is observed in the estimated translation for z coordinate. Experiments were performed using 3 grid levels, with 5 iterations each one, initial grid size of 40mm, maximum displacements of 25mm, 19 labels, $\alpha = 0.5$ and $\beta = 0.7$. The average running time was around 80 seconds (on an Intel

Xeon W3670 with 6 Cores and 16GB of RAM), where 57% of it corresponds to energy computation and 33% to optimization. Fig. 5.2 compares the results of our method with the results obtained using the algorithm proposed in [37]. Both mean and standard deviation errors of the estimated parameters are reduced by our method, meaning that the results are more accurate and less disperse than in the previous case. It shows that improving the quality of the energy terms by increasing the order of the cliques, results in an improvement of the final solution. It is important to remark that, for the same number of grid levels and iterations, our new approach reported slightly better computational time, running on the same computer (average of 80 sec per registration case for our method and 100 sec for the previous method). This difference is mainly due to the over-parametrized label space (5-dimensional) adopted by the previous method, that exponentially increments the number of labels needed to sample an equivalent solution space.

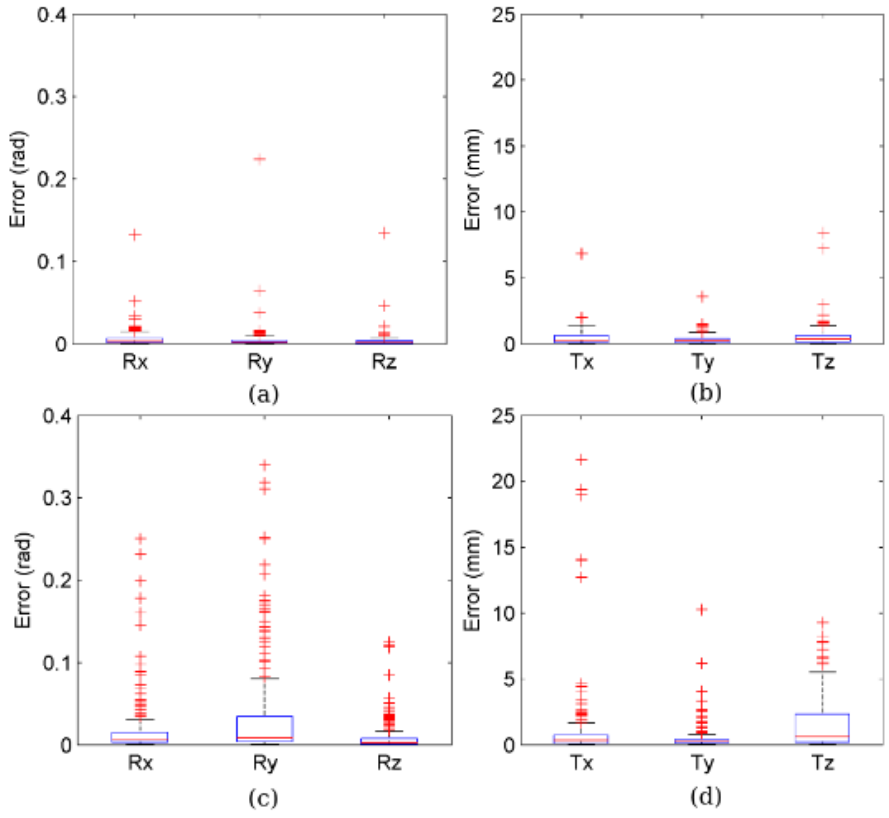


Figure 5.2: (a) and (b) Error estimation of the plane parameters for the method described here. (c) and (d) Error estimation for the method presented in [37] for the same dataset with equivalent setup.

Validation of in-plane deformation was performed over 20 registration cases. The dataset provides manual segmentations S of the left endocardium from a set of 20 slices. We register each slice with a starting from a random position around the ground truth (using Gaussian noise with $\sigma_r = 4.5\,\text{deg}$ and $\sigma_t = 5mm$ for rotation and translation parameters respectively. The estimated deformation field was applied to the corresponding initial segmentation and it was compared with the ground truth using DICE coefficient. The average DICE before deformation was 0.85 while after deformation we obtained 0.91, showing that in case of deformable organs like heart, deformation is important to guarantee reliable results.

5.5 Discussion

The method is independent of the similarity measure; consequently, it has the potential to be adapted to different scenarios with variety of image modalities. We compared our results with another method based on graph optimization, that uses lower dimensional cliques but higher dimensional label space. We showed that using higher order cliques we can model more powerful and accurate energies that lead to more reliable results.

Future work must be conducted mainly in three directions. First, the validation of the proposed formulation in real clinical scenarios (liver tumour resection guidance and MRI/US guidance for prostate biopsy) is under investigation. It should be noted that such a task is complex due to the complete absences of ground truth. Second, alternative optimization methods (in particular the ones acting directly on the objective function label space like dual decomposition) might lead to better results and are under investigation. Last, decoupling rigid and deformable parameters would allow us to have a better control over the algorithm behaviour, while also decreasing the computational complexity and reducing the hardness of the associated graphical model.

In this chapter, we showed another application of the triplets we defined previously. It proved they can be used in other contexts and with some differences we had mentioned in the method: hard or soft constraint and other algorithms.

Chapter 6

Conclusion

In this thesis, we construct a new way to solve linear registration within a higher order Markov Random Field framework, and we show different methods to couple this graph with a standard deformable model. The original motivation was to solve simultaneously linear and deformable registration to remove a potential bias coming from the fact the usual way to solve it uses a consecutive approach. We solve this problem and show our higher order framework can also be integrated into other problems. In this thesis, we address the registration task by solving two major concerns: how to perform linear registration within a Markov Random Field framework, and how to fuse this problem with deformable registration and how to solve it.

6.1 Contributions

Regarding the recent improvements of the optimization on Markov Random Fields, principally the Dual Decomposition and its developments, we searched for new applications that could be solved. Higher Order MRF already existed but all could not be solved before. With the decomposition methods, the number of solvable problems increase. With this view on optimization, we wanted to design and solve new registration problems.

We model the linear registration problem in a graphical model framework thanks to higher order terms. Those terms contain only local information and do not know the whole set of parameters of the transformation but they globally ensure the transformation lies in the set we seek to retrieve. The MRF registration tends to be as modular as possible regarding the pairwise potential, the similarity (or dissimilarity) measure, and the optimisation algorithm. From this standpoint, we designed a higher order model which, with slight changes, can handle affine, similarity or rigid transformations. We decompose the problem in different subproblems to use a decomposition technique called Dual Decomposition. We test our method on different problems and show state of the art results.

Solving linear registration with MRF was not the ultimate goal, because linear registration can already be solved with different methods with excellent results. Our aim was to perform simultaneously linear and deformable regis-

tration. They are usually performed successively but that creates a bias of the deformable registration towards the linear registration, and we want to get rid of it. We achieve this by fusing the previous model with a deformable model. The term "fuse" is very general as there are many ways to fuse two graphs, and our final model is quite different from the classic deformable model. For example the data term is a pairwise term, which is obviously non-submodular and has no property to be solved efficiently. That is the reason why we need again a decomposition method. As the graph is more complex than before, we use DD-ADMM, which is an extension of Dual Decomposition, and we design a clever decomposition to be able to avoid the problem of the pairwise term that contains the data term. The results show our model is working and apparently there was no bias to get rid of. We also show that the same ideas we developed here can be used to fuse linear and deformable registration in other ways and to solve the problem.

Furthermore, we propose a novel framework for 2D/3D registration. We use the triplets with soft constraints as regularizers of a 2D grid moving in a 3D space. With this soft constraint, we have to change the optimization. Indeed, the hard constraint of the higher order has the advantage to limit the computation of our slaves, but also has the disadvantage to disrupt many classical algorithms. At the opposite, the soft constraint would increase the computation of our slaves, but it does not disrupt any more classical algorithms, which is why we turned back to a factor graph formulation. The results on a challenging clinical setting demonstrate the efficiency of our method.

To sum up, the main contributions of this thesis are the following:

- We propose a higher order graphical model that captures linear transformations. The innovation comes from the design of the hyperedges and their combination which is able to solve a global problem locally. We show a decomposition able to optimize the problem. The model is flexible and compares to state of the art methods.

- We develop a global approach to jointly solve linear and deformable registration. We fuse our previous model with a deformable registration Markov Random Field. The complexity of the graph increases but we achieve good results thanks to DD-ADMM.

- We show an example on 2D/3D registration using the higher order terms we design. We use a different setup (soft constraint, factor graph) to prove the flexibility of our framework and how it can adapt for different purposes.

6.2 Future work

We mentioned that our linear framework could be extended to handle projective transformations. It would require many changes from the rest of our framework because we would need a continuous framework to handle the variety of transformations that are projections. Thus the data cost, which is already taking most of the time of the whole algorithms to compute, should be handled very efficiently, resulting in a happy medium between precision and computation time. That new model could handle projective 2D/3D registration for medical applications or on real images for drones location.

For linear registration alone, we used Dual Decomposition. The main advantage of this algorithm is that if an agreement is obtained, we know this is the optimal solution of the problem. Of course, it may not be reached. It is the case for our first model, which converges in more than 90% of the time. Usually, when it does not converge, we still know we are close (and how close) to the minimal energy and a happy medium is chosen between the solutions of the different slaves. However it does not work in every cases. Unfortunately, it does not work for our problem because it contains higher order constraints that makes voting or similar methods totally inefficient. Our linear and deformable model is more complex and Dual Decomposition does not achieve an agreement often enough. We used DD-ADMM but it does not show the same mathematical securities as DD, which means it converges more often but we have no guarantee that the agreement is the optimal solution. We tried many optimization algorithms on this problem but they failed due to the complexity of the graph. As new algorithms will probably emerge soon to solve higher order problems, attention should stay on them to find and compare them to DD-ADMM.

The second model fusing linear and deformable registration seems of great interest for medical doctors. Partnership with a medical lab interested in applications focusing on bones registration could lead to a software designed for medical doctors. Similar problems should be solved with this framework, as soon as medical applications will emerge.

Last but not at least, we designed our model to contain as few hyperedges as possible and to stay symmetric. But that may not be the best choice of hyperedges regarding results and optimization, especially when the size of the grid and the label set increase. A learning approach, which already exist with Dual Decomposition [89], could lead to a substantially different design and it could greatly enhance performance of the method.

Publications of the Author

International Conferences

- Fécamp Vivien, Sotiras Aris, and Paragios Nikos. Modular linear iconic matching using higher order graphs. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on* (pp. 1097-1101). IEEE.

- Fécamp Vivien, Sotiras Aris, and Paragios Nikos. Simultaneous Linear and Deformable Registration Through a Higher Order MRF Model. In *Bayesian and grAphical Models for Biomedical Imaging.* Springer International Publishing, 2015. (pp.2-12).

- Ferrante Enzo, Fécamp Vivien, and Paragios Nikos. Implicit Planar and In-Plane Deformable Mapping in Medical Images Through High Order Graphs. In *IEEE International Symposium on BIOMEDICAL IMAGING: From Nano to Macro (ISBI)*. 2015.

- Ferrante Enzo, Fécamp Vivien, and Paragios Nikos. Slice-to-volume deformable registration: efficient one-shot consensus between plane selection and in-plane deformation. *International Journal of computer assisted radiology and surgery*, 2015 (pp. 1-10).

Chapter 7

Appendix

We include here a paper for 2D/3D registration we did in collaboration with M. Enzo Ferrante, but which is not in the scope of this thesis because it does not use higher order MRF. Concerning this paper, my work consisted in the designing of the graph. My idea was to decouple the label set of one node into two nodes, similarly to what I did to combine linear and deformable registration, because the previous graph which had been designed could not be solved correctly. The results are slightly worse than those with the higher order model, probably due to the fact that those higher order model more precisely the data term. Their drawback is an increase of the complexity of the model and a need to use approximation to reach a solution.

Noname manuscript No.
(will be inserted by the editor)

Slice-to-Volume Deformable Registration

Efficient one shot consensus between plane selection and in-plane deformation

Enzo Ferrante · Vivien Fecamp · Nikos Paragios

Received: date / Accepted: date

Abstract
Purpose: This paper introduces a novel decomposed graphical model to deal with slice-to-volume registration in the context of medical images and image guided surgeries.
Methods: We present a new non-rigid slice-to-volume registration method whose main contribution is the ability to decouple the plane selection and the in-plane deformation parts of the transformation - through two distinct graphs - 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.
Results: The proof of concept for the proposed formulation is done using cardiac MR sequences of a beating heart (an artificially generated 2D temporal sequence is extracted using real data with known ground truth) as well as multimodal brain images involving ultrasound and computed tomography images. We achieve state of the art results while decreasing the computational time when we compare with another method based on similar techniques.
Conclusions: We confirm that graphical models and discrete optimization techniques are suitable to solve non-rigid slice-to-volume registration problems. Moreover, we show that decoupling the graphical model and labeling it using two lower dimensional label spaces, we can achieve state of the art results while substantially reducing the complexity of our method and moving the

Center for Visual Computing (CVN), CentraleSupelec — Galen Team, INRIA
92295 Chatenay-Malabry, France
Tel.: (+33) 01 41 13 16 30
E-mail: enzo.ferrante@ecp.fr

approach close to real clinical applications once considered in the context of modern parallel architectures.

Keywords Slice-to-Volume Registration · 2D-3D Registration · Discrete Optimization · Graphical Models · Markov Random Fields

1 Introduction

The problem of slice-to-volume deformable image registration consists in aligning a sliced 2D image (e.g. Ultrasound or US) to its corresponding plane from a 3D volume (e.g. Computer Tomography or CT). We call it deformable registration because the 2D image can be deformed during the registration process.

This problem finds applications in many medical image related contexts such as computer aided-biopsy [19], motion correction for image reconstruction [5], tumor ablation [22] and image-guided surgery (IGS) [23]. In the case of image guided procedures, a pre-operative 3D image and several intra-operative 2D acquisitions are to be fused towards providing position and navigation information to the surgeons. Nowadays, this fusion is mainly performed using two different tracking technologies: optical (OTS) and electromagnetic (EMTS) tracking systems. In the first case, OTS requires a line-of-sight to be maintained between the tracking device and the instrument to be tracked; this fact can disturb doctors during their work and is not always convenient. In the second case, EMTS does not have line-of-sight requirements but it is very susceptible to distortion from nearby metal sources and presents limited accuracy compared to optical tracking [4]. Moreover, nor OTS neither EMTS can deal with deformations between intra and pre-operative images. In this work, we propose to use 2D-3D slice-to-volume registration algorithms which are purely image based to solve this challenging problem and overcome the limitations presented by current technologies.

The problem of deformable image registration has been a pillar of computer vision (optical flow) and medical imaging (image fusion), and therefore one can cite numerous methods to perform 2D-2D and 3D-3D registration [11] [1]. However, the problem of 2D-3D registration, and particularly the problem of slice-to-volume registration, deserves separate investigation and specific methods development. While a single 2D slice contains less information than a 3D volume, the solution remains a 3D mapping function (a deformation field in case of non-rigid registration or a transformation matrix in case of rigid registration) as in the case of 3D-3D registration. This fact converts 2D to 3D slice-to-volume registration in a really challenging problem. The other case of 2D-3D registration problems, where projective 2D images such as X-Ray images are registered with volumetric images (CT for example) has received more attention in the last years [18] [15] and is not covered in this paper.

A variety of methods has been proposed to deal with slice-to-volume registration. In [3], standard optimization approaches and heuristics (as Simplex and Simulated Annealing algorithms) are applied on FluroCT to CT registration, testing with different intensity based similarity measures. [6] presents

a feature based method that performs slice-to-volume registration, using several slices in order to improve the quality of the results. [10] tracks intra-operative MRI slices of prostate images with a pre-operative MRI volume. This monomodal registration (MRI intra-operative slices to MRI pre-operative volume) is designed to provide patient tracking information for prostate biopsy performed under MR guidance. A similar problem is tackled by [25] where a two-step algorithm (rigid registration in the first step, and deformable registration in the second one) is applied to register three orthogonal intra-operative MR slices with a pre-operative volume. [23] proposes a method to register endoscopic and laparoscopic US images with pre-operative CT volumes. It is based on a new phase correlation technique called LEPART and it manages only rigid registration in quasi real time. [21] presents a flexible framework for intensity based slice-to-volume non-rigid registration algorithms that was used to register histological sections images to MRI of the human brain.

The main limitations of the aforementioned methods are their specificity to the clinical context (they are derived and can be used for specific clinical applications), the requirement of anatomical segmentations in some of them that increases their complexity and often their sequential nature where first plane is selected and then in-plane deformation is determined. Graphical models are powerful formalisms that could be amended to overcome these limitations. Casting computer vision problems as labeling ones through the use of Markov Random Field (MRF) theory has gained attention since [9]. It has been widely used to solve non-rigid image registration in the last years [11] [16] [17], mainly for 2D-2D or 3D-3D. In [26], a method based on MRFs to perform 2D-3D registration is presented, but it estimates just rigid transformations and works with projective images. Regarding slice-to-volume registration using MRF, our previous work [7] presents a MRF framework based on a high dimensional label space to solve this problem; we will refer to it as the overparameterized method.

In this work, our aim is to introduce a low rank graphical model that is able to simultaneously perform plane selection and estimate the in-plane deformation between the 2D source image and the corresponding slice from the 3D volume. We decouple a physical control point of a regular grid in two nodes of the MRF graph, one taking labels from the plane selection label space and the other one from the in-plane deformations label space. In that way, the complexity of the model reduces to the square of the cardinality of the biggest label space (instead of being quadratic in the product of the cardinalities of the two spaces), with a slight increase of the graphical model connectivity. This technique has been previously applied in 2D-2D registration [24]. The main advantage is related to the fact that, while the number of nodes augment linearly, the number of labels is decreased in a quadratic order.

The main contributions of our paper with respect to our previous work [7] are therefore two-fold. Firstly, we propose a new way of decoupling the plane selection and the in-plane deformation label spaces towards a novel low rank model of order 3 (instead of a model of order 5 as in [7]); it results into a more tractable problem in terms of getting the optimal solution. Secondly,

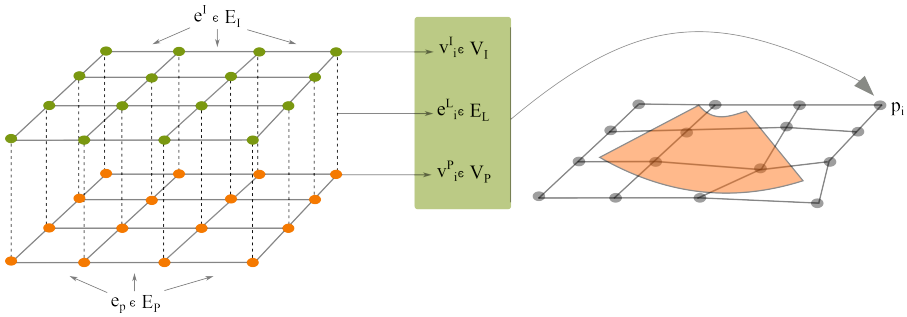


Fig. 1 Structure of the decoupled graph. The green nodes (top grid) are included in V_I and orange ones (bottom grid) in V_P modeling in-plane deformations and plane position respectively. Edges connecting V_I nodes are part of E_I and those connecting V_P nodes are part of E_P; they are associated with regularization terms. Dotted lines represent cliques in E_D that encode the matching similarity measure. Using this information we can reconstruct a deformed grid that is interpreted as a Free Form Deformation model. In the image we can appreciate how we associate two nodes of the graph with one control point of the grid.

we obtain substantial decrease of the search space size (order of 10), allowing much richer sampling of the label space, thus in theory more precise solutions. Moreover, by decoupling the label spaces it is possible to explore both of them with different sparseness levels.

The framework is intensity based and independent of the similarity measure, so it can be adapted to different image modalities or new measures. We tested our approach on two different datasets: a monomodal dataset where 2D MRI images of the heart are registered with MRI volumes, and another multimodal dataset where 2D US images are fused with CT volumes [20]. Both datasets were also used in [7].

The paper is organized as follows: in Section 2 we present the decoupled MRF formulation together with a complete explanation about the label spaces and the energy terms. In Section 3, the validation tests and results are presented and discussed. Finally, Section 4 concludes our paper and provides some ideas on relevant future directions.

2 Method Description

Non-rigid slice-to-volume registration can be seen as an optimization problem. We aim at optimizing an energy function by choosing the optimal plane (slice) $\hat{\pi}[J]$ from target volume J and the optimal deformation field $\hat{T}_D$ as indicates the following equation:

$$\hat{T}_D, \hat{\pi} = \underset{T_D, \pi}{\operatorname{argmin}} \mathcal{D}(I \circ T_D(\boldsymbol{x}), \pi[J](\boldsymbol{x})) + \mathcal{R}(T_D, \pi), \tag{1}$$

where I is the source 2D image, $\mathcal{D}$ represents the data term and $\mathcal{R}$ the regularization term. Given the 2D source image I and the 3D target volume J, we seek the slice $\hat{\pi}[J]$ from volume J that best matches the image I. We call

it non-rigid registration because image I can be deformed by the deformation field $\hat{T}_D$. The data term $\mathcal{D}$ measures the similarity between the source and the target, while the regularization term imposes smoothness constraints on the solution.

From this general optimization problem, we can derive different formulations. In [7], we proposed a high-dimensional label space based approach considering local labels of dimension five (plane + in-plane deformations). One of the main problems related to this high dimensionality is its consequently high computational cost. In this work, we try to avoid this problem by decoupling the label space in two different ones and reforming the structure of the graph to still capture rigid plane displacements and in-plane deformation.

Our formulation consists in an undirected pairwise graph $G_D =< V, E >$ with a set of nodes $V = V_I \cup V_P$ and a set of edges $E = E_I \cup E_P \cup E_D$. V_I and V_P have a 4-neighbor grid structure and the same cardinality. Nodes in V_I are labeled with in-plane deformation labels, while labels used in V_P represent the plane position. Edges from E_I and E_P correspond to a conventional pairwise neighborhood connection system for nodes in V_I and V_P respectively; they are associated with regularization terms (E_I corresponds to in-plane deformation regularizers and E_P to the plane selection regularizers). Edges in E_D link every node from V_I to its corresponding node from V_P, creating a graph with a sort of three dimensional structure (see Figure 1); those terms associated to E_D encode the data terms (i.e. the similarity measure).

In order to get a better understanding of the model, we can think of a single hypothetical grid similar to the one defined in [7], where every control point $\boldsymbol{p_k}$ from this grid is associated with two nodes from our approach, i.e. $v_k^I \in V_I$ and $v_k^P \in V_P$. This idea is depicted in Figure 1 and it will be useful to understand the energy terms.

Label Space

We define two different label spaces, one associated with nodes in V_I (called L_I) and the other one associated with nodes in V_P (called L_P).

The first label space, L_I, is a bidimensional space that models in-plane deformation using displacement vectors $\boldsymbol{l^I} \in E_I = (d_x, d_y)$.

The second label space, L_P, indicates the plane in which the corresponding control point is located. It consists of labels $\boldsymbol{l^P}$ associated to different planes. In order to specify the plane and the orientation of the grid on it, we store an orthonormal basis of this plane together with the position of a reference point in this plane. Using this information, we can reconstruct the position of the rest of the control points in the grid. This way of storing the planes, allow us to implement different plane space sampling methods. In this work, we chose a simple uniformly sampling around the current plane position, varying rotation and translation parameters in a given range. 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.

To compute the final position of a control point we use both labels. First, the corresponding label in L_P defines a 3D point belonging to a plane space with a given basis. Then, we use the corresponding label in L_I to move the point in the 2D plane thanks to its basis.

Objective Function

The energy that guides the optimization process is defined on the pairwise terms. Two types of edges represent regularization terms while the last one represents the data terms; the energy is thus defined as:

$$\mathcal{E}(I,P,D) = \min\{\gamma \sum_{(i,j)\in E_I} e^I_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^I_j}) + \alpha \sum_{(i,j)\in E_P} e^P_{i,j}(\boldsymbol{l^P_i}, \boldsymbol{l^P_j}) + \beta \sum_{(i,j)\in E_D} e^D_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^P_j})\}, \tag{2}$$

where γ, α and β are positive weighting factors, $e^I_{i,j} \in I$ are the in-plane regularizers (associated to edges in E^I), $e^P_{i,j} \in P$ are the plane regularizers (associated with edges in E^P) and $e^D_{i,j} \in D$ the data terms (associated with edges in E^D). $\boldsymbol{l^I_i}, \boldsymbol{l^P_i}$ are labels from both label spaces L_I and L_P respectively. Data and regularization terms are detailed in the following sections.

Data Likelihood

The data term is defined for interconnected pairs of nodes (i,j) between the two graphs (where $i \in V^I, j \in V^P$) and their corresponding labels $\boldsymbol{l^I} \in L_I, \boldsymbol{l^P} \in L_P$. It is encoded in the pairwise terms $e^D \in E_D$. As we described before, a plane and an in-plane deformation 2D-vector are associated with every control-point. Combining both labels, we calculate the final position of the control point $\boldsymbol{p_k}$ and extract an oriented patch Ω_k over the plane π_k (centered in $\boldsymbol{p_k}$) from the volume J, so that the similarity measure δ can be calculated between that patch and the corresponding area over the 2D source image:

$$e^D_{i,j}(\boldsymbol{l^I_i}, \boldsymbol{l^P_j}) = \int_{\Omega_k} \delta(I(\boldsymbol{x}), \pi_k[J](\boldsymbol{x}))d\boldsymbol{x}. \tag{3}$$

The patch-based similarity measure δ (defined on the sub-domain Ω_k) can encompass a wide choice of intensity-based measures. One of the simplest and most used similarity measures is the Sum of Absolute Differences (SAD). It is useful in the monomodal scenario, where two images of the same modality are compared. Its formulation is:

$$e^D_{SAD_{i,j}}(\boldsymbol{l^I_i}, \boldsymbol{l^P_j}) = \int_{\Omega_k} \mid (I(\boldsymbol{x}) - \pi_k[J](\boldsymbol{x}) \mid dx. \tag{4}$$

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 cannot assume that corresponding objects have the same intensities in the two images. MI is defined using the joint

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 ← initializeGraph(T_0)        ▷ Initialize the graph in the position indicated by T_0
    bestEnergy ← ∞
    for i=1 to gridLevels do
        L ← updateLabelSpace(L, i)        ▷ Update the label space for the given level
        for j = 1 to iterationSteps do
            newEnergy, newLabeling ← optimizeGraphicalModel(G, L)
            if newEnergy < bestEnergy then
                applyLabeling(G, newLabeling)
                bestEnergy = newEnergy
            end if
            refineLabelSpace()
        end for
    end for
    return bestEnergy, G
end procedure
```

intensity distribution $p(i,j)$ and the marginal intensity distribution $p(i)$ and $p(j)$ of the images as:

$$e^D_{MI_{i,j}}(\boldsymbol{l_i^I}, \boldsymbol{l_j^P}) = -\int_{\Omega_k} \log \frac{p(I(\boldsymbol{x}), \pi_k[J](\boldsymbol{x}))}{p(I(\boldsymbol{x}))p(\pi_k[J](\boldsymbol{x}))} dx. \tag{5}$$

As we could see in the previous examples, our framework can be endowed with any similarity measure defined on two bidimensional images. In this work, we use SAD for the monomodal heart dataset and MI for the multimodal brain dataset.

Regularization Terms

We define two different regularization terms, one regularizing the plane selection and the other one the in-plane deformation. The first regularization term penalizes the average distance between the nodes $i, j \in V^P$ and the plane corresponding to the neighboring one. If $D_\pi(\boldsymbol{p})$ indicates the point-to-plane distance between the point $\boldsymbol{p}$ and the plane π, we define the regularization term e^P as the average of these distances for two neighboring points i, j and their corresponding planes:

$$e^P_{i,j}(\boldsymbol{l_i^P}, \boldsymbol{l_j^P}) = \frac{1}{2}(D_{\pi_j}(\boldsymbol{p_i}') + D_{\pi_i}(\boldsymbol{p_j}')). \tag{6}$$

where $\boldsymbol{p_i}'$ and $\boldsymbol{p_j}'$ are the positions after applying label $\boldsymbol{l_i^P}$, $\boldsymbol{l_j^P}$ to $\boldsymbol{p_i}$, $\boldsymbol{p_j}$ respectively. This value is 0 when both points lie the same plane.

The second regularization term controls the in-plane deformation and is defined between nodes i and j included in V_I. We use a distance preserving approach which is symmetric, based on the ratio between the current position of the control points $\boldsymbol{p_i}, \boldsymbol{p_j}$ and their original position $\boldsymbol{p_{o,i}}, \boldsymbol{p_{o,j}}$:

$$\psi_{i,j}(\boldsymbol{l_i^I}, \boldsymbol{l_j^I}) = \frac{|| \, (\boldsymbol{p_i} + \boldsymbol{l_i^I}) - (\boldsymbol{p_j} + \boldsymbol{l_j^I}) \, ||}{|| \, (\boldsymbol{p_{o,i}}) - (\boldsymbol{p_{o,j}}) \, ||}. \tag{7}$$

Once defined ψ_{ij}, we need our regularizer to fulfill two conditions: first, we want it to be symmetric with respect to the displacement of the points, i.e. to penalize with the same cost whenever the control points are closer or more distant; second, we need the energy to be zero when the points are preserving distances and bigger than zero otherwise. The following regularization term fulfills both conditions for a couple of nodes $i, j \in V^I$ labeled with labels $\boldsymbol{l_i^I}, \boldsymbol{l_j^I}$:

$$e^I_{i,j}(\boldsymbol{l_i^I}, \boldsymbol{l_j^I}) = (1 - \psi_{i,j}(\boldsymbol{l_i^I}, \boldsymbol{l_j^I}))^2 + (1 - \psi_{i,j}(\boldsymbol{l_i^I}, \boldsymbol{l_j^I})^{-1})^2. \tag{8}$$

Note that both types of pairwise terms are not sub-modular since we include the current position of the points (which can be arbitrary) in their formulation and therefore sub-modularity constraint is not fulfilled.

Implementation Details

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 best possible 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.

The pairwise graphical model is optimized using the Loopy Belief Propagation algorithm (other discrete optimization algorithms can be used as well) implemented in the OpenGM2 library [12]. In [7], we used FastPD [14] instead of Loopy Belief Propagation for optimizing our pairwise model, which is among the most efficient optimization algorithms. However, due to its construction (lifting of the duality gap minimization) FastPD requires in general (towards optimizing complexity) an equal number of labels for all nodes which is an issue in our setting given the different dimensionality of the graph spaces (3d and 2d). Furthermore, while it can converge to a minimum even for non-submodular graphs, it is known that the quality of the linear programming (LP) relaxation is far from being satisfied and therefore the solution itself might be a very bad local minimum. Message passing methods like Loopy Belief Propagation do not inherit the computational constraints of FastPD while it is known (at least experimentally) that do good job as well even with highly non-submodular pairwise functions.

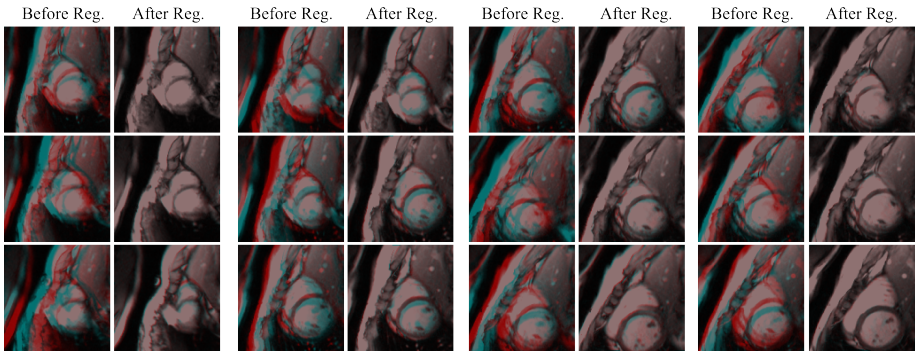


Fig. 2 12 registration cases of the same sequence, before and after registration. The overlapping images (in light blue we show the source image and in red the target) showed before registration corresponds to the source image and a slice taken from the volume at the initial position. The overlapping after registration corresponds to the deformed source image and the slice taken from the volume at the estimated plane position.

3 Validation & Results Discussion

We validate our method in two different scenarios and we compare the results with our previous method [7]. The first one corresponds to a monomodal sequence of 2D MRI images randomly extracted from a 3D MRI temporal series of a beating heart. The second one is a multimodal brain dataset formed by 2D US images and 3D CT extracted from [20].

In order to compare both methods in a fair way, we exhaustively tested different parameter configurations (empirically for every dataset) on a grid of discretized values, and we took the best combination for each method.

3.1 Heart Dataset

The MRI heart dataset consists of ten sequences of twenty bidimensional MRI slices each one, that are registered with a MRI volume, giving a total of 200 registration cases. In order to generate them, as it was described in [7], we took a temporal series of 20 MRI volumes of a beating heart, and we extracted ten random trajectories of twenty slices I_i each one (one slice for every volume M_i). Starting from a random initial rotation $R_0 = (R_{x_0}, R_{y_0}, R_{z_0})$ and translation $T_0 = (T_{x_0}, T_{y_0}, T_{z_0})$, we extracted a 2D slice I_0 from the initial volume M_0. In every sequence, the position of slice I_i was generated adding Gaussian noise to the position of slice I_{i-1} with $\sigma_r = 3°$ and $\sigma_t = 5\ mm$ to every translation (T_x, T_y, T_z) and rotation $(R_x,\ R_y,\ R_z)$ parameters respectively. It gives maximum distances of about 25 mm between the current and its succeeding slice. The MRI resolution was $192 \times 192 \times 11$ and the voxel size was $1.25 \times 1.25 \times 8\ mm^3$.

For every sequence, we initialize the registration adding the same noise (with the same parameters than before) to the ground truth. During the reg-

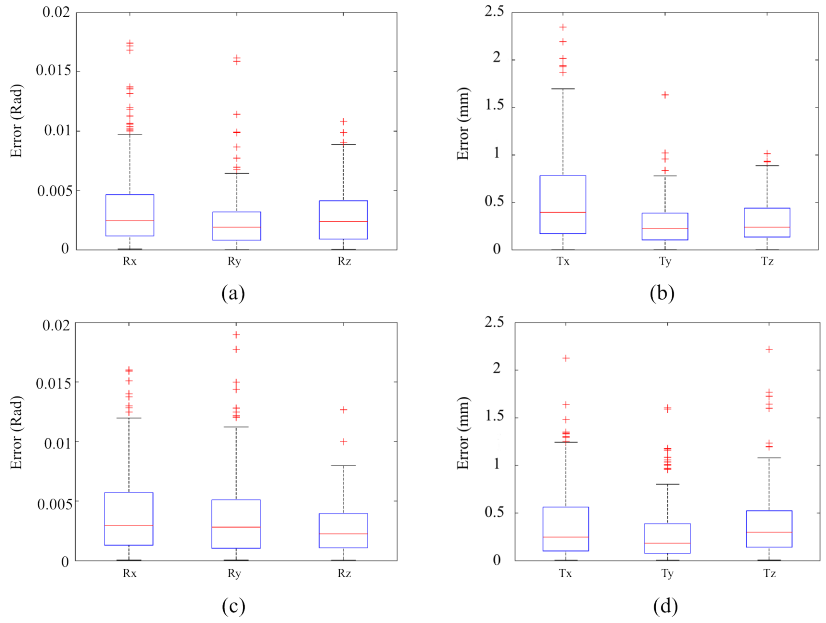


Fig. 3 Comparison of the error estimation for plane parameters (R_x, R_y, R_z) and (T_x, T_y, T_z) for our decoupled method (figures (a) and (b)) and the overparameterized approach presented by [7] (figures (c) and (d)). For presentation clarity, three outliers between 0.02 and 0.05 *rad* as well as one at 4 *mm* have been removed at Figures (c) and (b) respectively.

istration process, given two consecutive slices of the same sequence, the estimated transformation for slice I_i was used as initialization for the registration of slice I_{i+1}.

		R_x	R_y	R_z	T_x	T_y	T_z
Decoupled Method	Mean	0.0036	0.0024	0.0029	0.5403	0.2713	0.2966
	SD	0.0034	0.0024	0.0024	0.4914	0.2296	0.2236
Overparameterized Method [7]	Mean	0.0051	0.0051	0.0031	0.4164	0.2874	0.4847
	SD	0.0122	0.0134	0.0051	0.4720	0.2976	1.1546

Table 1 Error estimation for plane parameters (R_x, R_y, R_z) and (T_x, T_y, T_z) for our decoupled method and the previous overparameterized approach presented in [7].

Figure 2 shows the overlapping between the source image and the corresponding target plane, before and after registration, for 12 cases of one sequence. As we can observe in a qualitative way, the overlapping increases after registration.

Figure 3 compares our results in a quantitative way with the ones obtained using our previous method. We measure the error between the estimated transformation parameters and the ground truth. The mean error was (0.0036, 0.0024, 0.0029) *rad* for rotation and (0.5403, 0.2713, 0.2966) *mm* for translation parameters, with a standard deviation of (0.0034, 0.0024, 0.0024) *rad*

and (0.4914, 0.2296, 0.2236) *mm* respectively. The average running time was around 60 seconds for every registration case. Using the method presented in [7], we obtained (0.0051, 0.0051, 0.0031) *rad* and (0.4164, 0.2874, 0.4847) *mm* for rotation and translation parameters error, and standard deviation equal to (0.0122, 0.0134, 0.0051) *rad* and (0.4720, 0.2976, 1.1546) *mm*. Results are presented in Table 1. Every registration case took around 220 seconds (almost 3.5 times more than our method). As we can see, the quality of the results was preserved (and improved in some cases) while the computational time was reduced approximately 3.5 times (keeping equivalent grid and label space sizes, sampling patch size and number of algorithm iterations).

Validation of in-plane deformation was performed over 20 registration cases, deforming an initial segmentation of the left endocardium using the estimated deformation field T_{D_i}. We measure the average DICE coefficient between the segmentations, before and after deforming the initial one, to measure the impact of the deformation in the registration process. The average DICE before deformation was 0.858 and after registration was 0.907, showing that our method can capture in-plane deformations and select the correct plane at the same time.

Common parameters used for both methods were 3 grid levels, 5 iterations per level, initial control point distance of 40 *mm* and minimum sampling patch size of 20 *mm*. In case of the decoupled model we use $\gamma = 1$, $\beta = 0.2$, $\alpha = 0.8$, 41 labels in the plane label space and 91 labels in the deformations label space. In case of the overparameterized model we use 13122 labels and $\alpha = 0.9$ (for a complete understanding of these parameters refer to [7]). We run the experiments on an Intel Xeon W3670 with 6 Cores, 64bits and 16GB of RAM.

3.2 Brain Dataset

The brain dataset consists of a pre-operative brain MRI volume (voxel size of $0.5 \times 0.5 \times 0.5$ mm^3 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 [20]. The size of the US images was 48×38 *mm* and the pixel resolution 0.3×0.3 *mm*. The ventricles were manually segmented by specialists in both modalities and used to calculate DICE coefficient and Contour Mean Distance (CMD) to evaluate and compare the quality of the results. Initializations were done following the same methodology that we described for the Heart Dataset (Section 3.1).

Figure 4 summarizes the average DICE and CMD coefficients for each series. It shows that, using our decoupled method, the mean DICE increases after the registration process an average of 0.0405, a little bit more than the 0.0380 obtained with [7] method. Regarding the CMD, the average decrement for our method is 0.3654 *mm* while for the other one is 0.3943 *mm*. Even if our new method performs better in average, we can observe that results are almost equivalent in terms of DICE and CMD. However, there is a big difference in terms of computing time: while our method is taking around 3 *min* per reg-

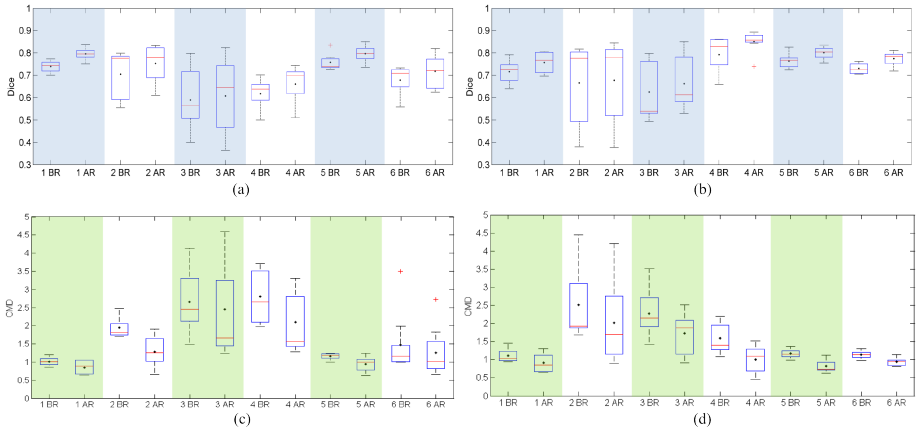


Fig. 4 The figures show a quantitative comparison of the two methods, before (BR) and after (AR) registration for the 6 sequences of brain data. Figures (a) and (c) show results for our decoupled method (DICE and CMD respectively) while figures (b) and (d) show results for the overparameterized approach presented in [7] (DICE and CMD respectively).

istration case, the overparameterized method takes around 10 *min* running in the same computer using the same configuration. To perform the experiments with both methods, we used the same configuration given by 3 grid levels, initial control point distance of 8 *mm*, 4 iterations per level and minimum sampling patch size of 13 *mm*. In case of the decoupled model, we set $\gamma = 1$, $\beta = 0.05$, $\alpha = 0.2$, 41 labels in the plane label space and 91 labels in the deformations label space. For the overparameterized method we set $\alpha = 0.8$ and 6174 labels. We run the experiments in the same Intel Xeon W3670 with 6 Cores, 64bits and 16GB of RAM used for the heart dataset.

3.3 Discussion & Comparison With Other Methods

As we have shown, our method is able to achieve state of the art results while decreasing the computational time when we compare to another MRF based method (namely [7]). In the monomodal case we reduce it from around 3.5 *min* to 1 *min* while in the multimodal one we go from 10 *min* to 3 *min*, giving a time factor reduction of about 3 times.

The main strength of the proposed formulation is the linear complexity of the inference process with respect to the product of the label spaces. This allows to go even further for challenging cases (brain tumor removal) where precision is required to substantially increase the label space. This is not the case for the approach presented in [7] due to the complexity of the label space.

An interesting point to discuss about is the 5-fold improvement in the standar deviation error of parameter T_z that we obtain with the new method. In [7], the justification for the poor performance of the method when estimating T_z was told to be that image resolution in z axis was lower than in x and y. We

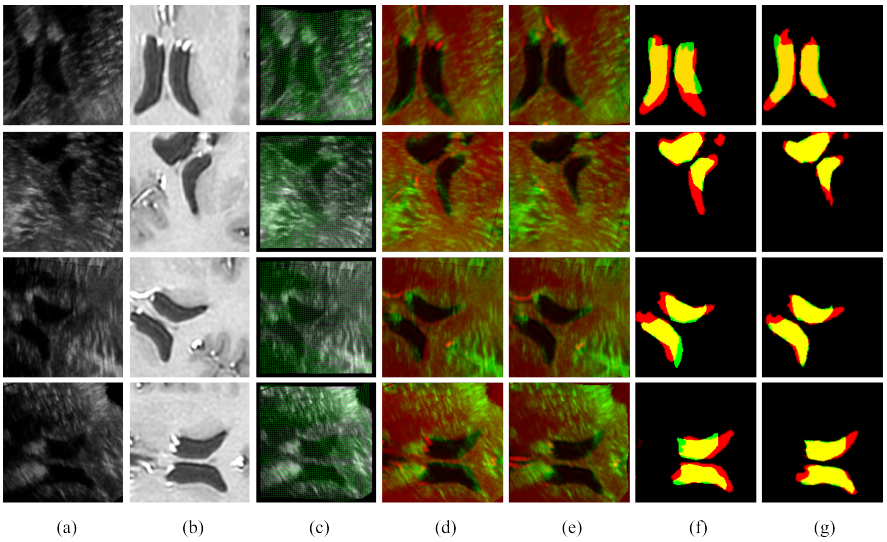


Fig. 5 Results for one slice from four of the six brain sequences (each row correspond to a different sequence). (a) Source 2D Ultrasound image. (b) Slice extracted from the MRI corresponding to the initial position of the plane. (c) Deformed source image overlapped with the estimated deformation field. (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.

think that the new algorithm is less sensitive to image resolution anisotropy mainly because of the different way we explore the plane-selection label space by allowing a deeper exploration when decoupling it without exponentially increasing the amount of labels.

It is important to remark that both, the decoupled and overparameterized methods, are highly dependent on the initialization given for the first slice of the sequence. Since these algorithms optimize the energy based on a limited search space (determined by the label space), if the solution is not reachable from the intial position using the current label space, the algorithm will fail. Another factor that is crucial for the success of the algorithm is the similarity measure used to decide whether or not two patches coming from different images correspond to the same anatomical structure. The study of different similarity measures is outside the scope of this paper; however, note that in order to use the method in other image modalities, it will be necesary to choose an accurate similarity measure and calibrate the parameters accordingly.

Comparison with other methods in the field of slice-to-volume registration is a complicated task, mainly because of the lack of public datasets. Here we include some of the results reported by other state of the art methods for their own datasets, in terms of accuracy and/or performance. In [10] for example, authors report a mean Target Registration Error (TRE) lower than 1 *mm* when estimating rigid transformations in a monomodal MRI dataset of

prostate images (for a pixel size of $1.5 \times 1.5 \times 3$ *mm*). Random initializations were generated by modifying the ground truth position with displacements of 10 *mm* and rotations of 10deg maximum. The Matlab implementation of their algorithm took between 36 *sec* and 107 *sec* depending on the algorithm configuration. In [23], authors tested on a multimodal dataset formed by 2D ultrasound and CT volumes of the heart. They report errors around 1.56 ± 0.78 *mm* when estimating rigid transformations on CT images with 0.6 *mm* isotropic resolution, using initializations with uniformly random shifts in the range -5 to 5 *mm*. They achieve quasi real time performance with execution times around 4 *sec*. Another interesting example to compare with is the multi-slice to volume registration case that tackles [25] applying it to MRI-guided transperineal prostate biopsy. Authors report that deformable registrations were accurate to within 2 *mm* in images with a slice spacing of 3.6 *mm*. The execution time for the complete deformable registration algorithm is about 30 *sec*. Even if it is not possible to do a fair comparison mainly because of the lack of standard benchmarks, by observing these examples we can clearly remark that our results are in the state of the art level. Moreover, visual assessment on the obtained results seems to confirm that these are satisfactory in the context of a clinical setting.

In terms of complexity, it is interesting to remark the difference with respect to our previous method. The optimization complexity/difficulty heavily depends on the maximum number of label combinations that the pairwise cliques can take (this is the bottle neck for most optimization algorithms). In this perspective, the complexity of the overparameterized model is given by $O(|L|^2)$, where $|L|$ is the cardinality (number of labels) of the label space. In our new approach, we introduce two label spaces L_1 and L_2 that decouple the previous one. To give an idea about the reduction in the complexity of our new model, let us say that $|L| = |L_1.L_2|$. Because of the way in which we construct our decoupled graph (as it is indicated in Figure 1), it is straightforward to show that the complexity of the new model reduces now to $O(\max(|L_1|, |L_2|)^2)$. Therefore, because of the decoupling strategy, the complexity of the model reduces to the square of the cardinality of the biggest label space (instead of being quadratic in the cardinalities of the joint space), with a slight increase of the graphical model connectivity. Consequently, while the number of nodes augment linearly, the number of labels is decreased in a quadratic order.

4 Conclusions

We presented a new method to perform slice-to-volume registration based on a decoupled model that associates two local graphs to the plane selection and the in-plane deformations while imposing consistency through direct connections between the corresponding nodes. In order to solve this problem, we seek the plane and the in-plane deformation that best matches our energy function. It is important to remark that we just look for the in-plane deformations given the nature of the problems we are trying to solve (mainly image fusion for

IGS), where it is not useful to find out-of-the-plane deformations at least for visualization purposes, even if they can exist.

As we have shown in the previous section, our method achieves state of the art results while decreasing substantially the time of computation when it is compared to our previous MRF based method that uses a unique high dimensional label space [7]. It confirms our initial hypothesis, meaning that decoupling the graphical model and labeling it using two lower dimensional label spaces, we can achieve the same results while reducing the complexity of our method.

We have also shown that the method is robust with respect to the type of images we are registering. Since slice-to-volume registration has multiple applications, other problems are under investigation (it should be noted that such a task is complex due to the complete absences of public ground truth). To this end, two clinical scenarios are currently under investigation, the first refers to liver tumor resection guidance, while the second to US guidance during prostate biopsy through fusion of intra-operative ultrasound and pre-operative CT/MR.

In order to improve the quality of the results, specially in multimodal cases, feature engineering must be considered. Future work includes adapting and using features specifically designed for multimodal registration such as the LC^2 presented in [8] and the MIND descriptor presented in [2]. Furthermore, energy regularizers inspired on precise biophysical modeling and tissue properties could lead to accuracy improvements as well. The underlying idea is to adapt the "smoothness" constraint of the deformation model by explicitly taking into account organ specific motion/deformation constraints like for example in the context of liver biopsies or brain tumor ablation.

Finally, we are investigating new methods to improve the parameters estimation procedure. Energy parameters estimation based on machine learning techniques [13] have to be considered as a future work if we want to exploit at the maximum level the potential of the proposed method.

5 Funding

This research was partially supported by European Research Council Starting Grant Diocles (ERC-STG-259112).

6 Conflict of Interest

The authors declare that they have no conflict of interest.

7 Ethical approval

– This article does not contain any studies with animals performed by any of the authors.

– All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. International Journal of Computer Vision **92**(1), 1–31 (2011)
2. Bardera, A., Feixas, M., Boada, I., Sbert, M.: High-dimensional normalized mutual information for image registration using random lines. In: J. Pluim, B. Likar, F. Gerritsen (eds.) Biomedical Image Registration, *Lecture Notes in Computer Science*, vol. 4057, pp. 264–271. Springer Berlin Heidelberg (2006)
3. Birkfellner, W., Figl, M., Kettenbach, J., Hummel, J., Homolka, P., Schernthaner, R., Nau, T., Bergmann, H.: Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images. Medical Physics **34**(1), 246 (2007). DOI 10.1118/1.2401661
4. Birkfellner, W., Hummel, J., Wilson, E., Cleary, K.: Tracking devices. In: Image-Guided Interventions, pp. 23–44. Springer (2008)
5. Chandler, A.G., Pinder, R.J., Netsch, T., Schnabel, J.A., Hawkes, D.J., Hill, D.L., Razavi, R.: Correction of misaligned slices in multi-slice mr cardiac examinations by using slice-to-volume registration. In: Journal of Cardiovascular Magnetic Resonance2008, 10:13 (2008)
6. Dalvi, R., Abugharbieh, R.: Fast feature based multi slice to volume registration using phase congruency. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp. 5390–5393. IEEE (2008)
7. Ferrante, E., Paragios, N.: Non-rigid 2d-3d medical image registration using markov random fields. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 163–170. Springer (2013)
8. Fuerst, B., Wein, W., Muller, M., Navab, N.: Automatic ultrasoundmri registration for neurosurgery using the 2d and 3d lc2 metric. Medical Image Analysis **18**(8), 1312 – 1319 (2014). Special Issue on the 2013 Conference on Medical Image Computing and Computer Assisted Intervention
9. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), 721–741 (1984)
10. Gill, S., Abolmaesumi, P., Vikal, S., Mousavi, P., Fichtinger, G.: Intraoperative prostate tracking with slice-to-volume registration in mri pp. 154–158 (2008)
11. Glocker, B., Sotiras, A., Komodakis, N., Paragios, N.: Deformable medical image registration: setting the state of the art with discrete methods. Annu Rev Biomed Eng **13**, 219–244 (2011). DOI 10.1146/annurev-bioeng-071910-124649
12. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Lellmann, J., Komodakis, N., Rother, C.: A comparative study of modern inference techniques for discrete energy minimization problem (2013)
13. Komodakis, N.: Efficient training for pairwise or higher order crfs via dual decomposition. In: CVPR, pp. 1841–1848 (2011)
14. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic mrfs. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pp. 1–8. IEEE (2007)
15. Kotsas, P., Dodd, T.: A review of methods for 2d/3d registration. WASET Conference Paris 14-16 November (2011)
16. Lee, K., Kwon, D., Yun, I., Lee, S.: Deformable 3d volume registration using efficient mrfs model with decomposed nodes. In: British Machine Vision Conference, pp. 1–10 (2008)

17. Mahapatra, D., Sun, Y.: Nonrigid registration of dynamic renal mr images using a saliency based mrf model. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008 pp. 771–779 (2008)
18. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3d/2d registration methods for image-guided interventions. Medical image analysis (2012)
19. Marks, L., Young, S., Natarajan, S.: Mri–ultrasound fusion for guidance of targeted prostate biopsy. Current opinion in urology **23**(1), 43 (2013)
20. Mercier, L., Del Maestro, R.F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.L.: Online database of clinical mr and ultrasound images of brain tumors. Medical Physics **39**, 3253 (2012)
21. Osechinskiy, S., Kruggel, F.: Slice-to-volume nonrigid registration of histological sections to mr images of the human brain. Anatomy Research International **2011** (2010)
22. Penney, G., Blackall, J., Hayashi, D., Sabharwal, T., Adam, A., Hawkes, D.: Overview of an ultrasound to ct or mr registration system for use in thermal ablation of liver metastases. In: Proc. Medical Image Understanding and Analysis, vol. 1, p. 6568. Citeseer (2001)
23. San José Estépar, R., Westin, C., Vosburgh, K.: Towards real time 2d to 3d registration for ultrasound-guided endoscopic and laparoscopic procedures. International journal of computer assisted radiology and surgery **4**(6), 549–560 (2009)
24. Shekhovtsov, A., Kovtun, I., Hlaváč, V.: Efficient mrf deformation model for non-rigid image matching. Comput. Vis. Image Underst. **112**(1), 91–99 (2008). DOI 10.1016/j.cviu.2008.06.006. URL http://dx.doi.org/10.1016/j.cviu.2008.06.006
25. Xu, H., Lasso, A., Fedorov, A., Tuncali, K., Tempany, C., Fichtinger, G.: Multi-slice-to-volume registration for mri-guided transperineal prostate biopsy. International journal of computer assisted radiology and surgery pp. 1–10 (2014)
26. Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Kamen, A., Paragios, N., Navab, N.: Linear intensity-based image registration by markov random fields and discrete optimization. Medical Image Analysis **14**(4), 550–562 (2010). DOI 10.1016/j.media.2010.04.003. URL http://dx.doi.org/10.1016/j.media.2010.04.003

List of Figures

List of Tables

Bibliography

[1] Pedro Aguiar, Eric P Xing, Mário Figueiredo, Noah A Smith, and André Martins. An augmented lagrangian approach to constrained map inference. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 169–176, 2011. (Cited on pages 66 and 67.)

[2] Karteek Alahari, Pushmeet Kohli, and Philip H S Torr. Reduce, Reuse & Recycle Efficiently Solving Multi-Label MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. (Cited on page 22.)

[3] Cevdet Aykanat, Enver Kayaaslan, Ali Pinar, and Ümit Çatalyürek. Hypergraph Partitioning through Vertex Separators on Graphs. 2011. (Cited on page 25.)

[4] Isaac N Bankman and Serban Morcovescu. Handbook of medical imaging. processing and analysis. *Medical Physics*, 29(1):107–107, 2002. (Cited on page 6.)

[5] Daniel I Barnea and Harvey F Silverman. A class of algorithms for fast digital image registration. *Computers, IEEE Transactions on*, 100(2):179–186, 1972. (Cited on page 35.)

[6] Dhruv Batra, Andrew C. Gallagher, Devi Parikh, and Tsuhan Chen. Beyond trees: MRF inference via outer-planar decomposition. *Computer Vision and Pattern Recognition*, pages 2496–2503, June 2010. (Cited on page 22.)

[7] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006. (Cited on page 34.)

[8] Jeffrey S Beis and David G Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1000–1006. IEEE, 1997. (Cited on page 35.)

[9] Claude Berge and Edward Minieka. *Graphs and hypergraphs*, volume 7. North-Holland publishing company Amsterdam, 1973. (Cited on page 12.)

[10] P.J. Besl and H.D. McKay. A method for registration of 3-d shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2):239 –256, feb 1992. (Cited on page 35.)

[11] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. (Cited on page 43.)

[12] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient nd image segmentation. *International journal of computer vision*, 70(2):109–131, 2006. (Cited on page 19.)

[13] Yuril Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001. (Cited on page 19.)

[14] Bernhard Brendel, Susanne Winter, Andreas Rick, Martin Stockheim, and Helmut Ermert. Registration of 3d ct and ultrasound datasets of the spine using bone structures. *Computer Aided Surgery*, 7(3):146–155, 2002. (Cited on page 6.)

[15] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992. (Cited on page 33.)

[16] Samuel Rota Bulò and Marcello Pelillo. A Game-Theoretic Approach to Hypergraph Clustering. *Advances in Neural Information Processing Systems*, pages 1–9, 2009. (Cited on page 25.)

[17] Samuel Rota Bulò and Marcello Pelillo. Probabilistic Clustering Using the Baum-Eagon Inequality. *International Conference on Pattern Recognition*, pages 1429–1432, 2010. (Cited on page 20.)

[18] Andrew E Caldwell, Andrew B Kahng, and Igor L Markov. Design and implementation of move-based heuristics for VLSI hypergraph partitioning. *Journal of Experimental Algorithmics*, 5(2000):5–es, 2000. (Cited on page 25.)

[19] A. Can, C.V. Stewart, B. Roysam, and H.L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):347 –364, mar 2002. (Cited on page 35.)

[20] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986. (Cited on page 34.)

[21] Yongchoel Choi and Seungyong Lee. Injectivity conditions of 2d and 3d uniform cubic b-spline functions. *Graphical models*, 62(6):411–427, 2000. (Cited on page 31.)

[22] Gary E Christensen, Sarang C Joshi, Michael Miller, et al. Volumetric transformation of brain anatomy. *Medical Imaging, IEEE Transactions on*, 16(6):864–877, 1997. (Cited on page 61.)

[23] Gary E Christensen, Richard D Rabbitt, Michael Miller, et al. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10):1435–1447, 1996. (Cited on pages 61 and 62.)

[24] Martha M Coselmon, James M Balter, Daniel L McShan, and Marc L Kessler. Mutual information based ct registration of the lung at exhale and inhale breathing states using thin-plate splines. *Medical physics*, 31(11):2942–2948, 2004. (Cited on page 6.)

[25] Christopher M Cyr, Ahmed F Kamal, Thomas B Sebastian, and Benjamin B Kimia. 2d-3d registration based on shape matching. In *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*, pages 198–203. IEEE, 2000. (Cited on page 75.)

[26] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005. (Cited on page 34.)

[27] Rupin Dalvi and Rafeef Abugharbieh. Fast feature based multi slice to volume registration using phase congruency. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 5390–5393. IEEE, 2008. (Cited on page 75.)

[28] George B Dantzig and Philip Wolfe. Decomposition principle for linear programs. *Operations research*, 8(1):101–111, 1960. (Cited on page 43.)

[29] Jérôme Declerck, Jacques Feldmar, Michael L Goris, and Fabienne Betting. Automatic registration and alignment on a template of cardiac stress and rest reoriented spect images. *Medical Imaging, IEEE Transactions on*, 16(6):727–737, 1997. (Cited on page 31.)

[30] Scott L Delp, David S Stulberg, Brian Davies, Frederic Picard, and Francois Leitner. Computer assisted knee replacement. *Clinical orthopaedics and related research*, 354:49–56, 1998. (Cited on page 6.)

[31] Olivier Duchenne, Francis Bach, Inso Kweon, and Jean Ponce. A Tensor-Based Algorithm for High-Order Graph Matching. *PAMI*, (2), May 2011. (Cited on page 23.)

[32] 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*, 4(6):549–560, 2009. (Cited on page 75.)

[33] VD Fachinotti, AA Anca, and A Cardona. A method for the solution of certain problems in least squares. *Int J Numer Method Biomed Eng*, 27(4):595–607, 2011. (Cited on page 62.)

[34] Vivien Fecamp, Aristeidis Sotiras, and Nikos Paragios. Modular linear iconic matching using higher order graphs. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1097–1101. IEEE, 2015. (Cited on page 55.)

[35] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012. (Cited on page 6.)

[36] Thomas S Ferguson. An inconsistent maximum likelihood estimate. *Journal of the American Statistical Association*, 77(380):831–834, 1982. (Cited on page 62.)

[37] 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*, pages 163–170. Springer, 2013. (Cited on pages 76, 79, 80 and 109.)

[38] Bernd Fischer and Jan Modersitzki. Curvature based image registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85, 2003. (Cited on page 62.)

[39] Bernd Fischer and Jan Modersitzki. Ill-posed medicine—an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008. (Cited on page 6.)

[40] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. (Cited on page 33.)

[41] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. (Cited on page 35.)

[42] Leila MG Fonseca and Max HM Costa. Automatic registration of satellite images. In *Computer Graphics and Image Processing, 1997. Proceedings., X Brazilian Symposium on*, pages 219–226. IEEE, 1997. (Cited on page 34.)

[43] Lester R Ford and Delbert R Fulkerson. Maximal flow through a network. *Canadian journal of Mathematics*, 8(3):399–404, 1956. (Cited on page 17.)

[44] Dongshan Fu and Gopinath Kuduvalli. Fiducial-less tracking with non-rigid image registration, 2008. US Patent 7,327,865. (Cited on page 75.)

[45] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976. (Cited on page 67.)

[46] Andrew C. Gallagher, Dhruv Batra, and Devi Parikh. Inference for order reduction in Markov random fields. *Computer Vision and Pattern Recognition*, pages 1857–1864, 2011. (Cited on pages 21 and 24.)

[47] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6:721–741, 1984. (Cited on pages 14 and 36.)

[48] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984. (Cited on page 76.)

[49] David T Gering, Arya Nabavi, Ron Kikinis, Noby Hata, Lauren J O'Donnell, W Eric L Grimson, Ferenc A Jolesz, Peter M Black, and William M Wells. An integrated visualization system for surgical planning and guidance using image fusion and an open mr. *Journal of Magnetic Resonance Imaging*, 13(6):967–975, 2001. (Cited on page 6.)

[50] David T Gering, Arya Nabavi, Ron Kikinis, Noby Hata, Lauren J O'Donnell, W Eric L Grimson, Ferenc A Jolesz, Peter M Black, and

William M Wells. An integrated visualization system for surgical planning and guidance using image fusion and an open mr. *Journal of Magnetic Resonance Imaging*, 13(6):967–975, 2001. (Cited on page 6.)

[51] Sean Gill, Purang Abolmaesumi, Siddharth Vikal, Parvin Mousavi, and Gabor Fichtinger. Intraoperative prostate tracking with slice-to-volume registration in mr. *SMIT (08 2008)*, pages 154–158, 2008. (Cited on page 75.)

[52] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through MRFs and efficient linear programming. *Medical image analysis*, 12(6):731–41, dec 2008. (Cited on page 70.)

[53] R Glowinski and A Marrocco. Sur l'approximation par el ements nis d'ordre un, et lan r esolution par p enalisation-dualit e, d'une classe de probl emes de dirichlet non lin eaires. *Rev. Fr. Autom. Inf. Rech. Oper. Anal. Num er*, pages 41–76. (Cited on page 67.)

[54] Gerhard W Goerres, Cyrill Burger, Michael R Schwitter, Thai-Nia H Heidelberg, Burkhardt Seifert, and Gustav K von Schulthess. Pet/ct of the abdomen: optimizing the patient breathing pattern. *European radiology*, 13(4):734–739, 2003. (Cited on page 6.)

[55] A Ardeshir Goshtasby. *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005. (Cited on page 6.)

[56] Ardeshir Goshtasby, George C Stockman, and Carl V Page. A region-based approach to digital image registration with subpixel accuracy. *Geoscience and Remote Sensing, IEEE Transactions on*, (3):390–399, 1986. (Cited on page 34.)

[57] Stefan Growe and Ralf Tönjes. A knowledge based approach to automatic image registration. In *Image Processing, 1997. Proceedings., International Conference on*, volume 3, pages 228–231. IEEE, 1997. (Cited on page 34.)

[58] Yujun Guo, Radhika Sivaramakrishna, Cheng-Chang Lu, Jasjit S Suri, and Swamy Laxminarayan. Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing*, 44(1-2):15–26, 2006. (Cited on page 6.)

[59] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer, 1988. (Cited on page 34.)

[60] Magnus R Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969. (Cited on page 67.)

[61] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in medicine and biology*, 46(3):R1, 2001. (Cited on page 6.)

[62] Yuan C Hsieh, David M McKeown, and Frederic P Perlant. Performance evaluation of scene registration and stereo matching for artographic feature extraction. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2):214–238, 1992. (Cited on page 34.)

[63] Xishi Huang, Jing Ren, Gerard Guiraudon, Derek Boughner, and Terry M Peters. Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation. *Medical Imaging, IEEE Transactions on*, 28(11):1802–1814, 2009. (Cited on page 6.)

[64] Yuchi Huang, Qingshan Liu, and Dimitris Metaxas. Video object segmentation by hypergraph cut. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1738–1745, 2009. (Cited on page 23.)

[65] Ragnar Bang Huseby, Ole Martin Halck, and Rune Solberg. A model-based approach for geometrical correction of optical satellite images. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International*, volume 1, pages 330–332. IEEE, 1999. (Cited on page 35.)

[66] Luis Ibanez, William Schroeder, Lydia Ng, and Josh Cates. The itk software guide. 2003. (Cited on page 6.)

[67] Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2993–3000, 2009. (Cited on pages 21 and 24.)

[68] Hiroshi Ishikawa. Higher-order gradient descent by fusion-move graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 568–574, Kyoto, Japan, 2009. (Cited on pages 21 and 24.)

[69] Hiroshi Ishikawa. Transformation of general binary MRF minimization to the first-order case. *PAMI*, 33(6):1234–49, June 2011. (Cited on pages 21 and 24.)

[70] Satoru Iwata and Lisa Fleischer. A Combinatorial , Strongly Polynomial-Time Algorithm for Minimizing Submodular Functions. *Journal of the ACM*, 48(4):1–17, 2001. (Cited on page 19.)

[71] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002. (Cited on page 6.)

[72] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2):825–841, oct 2002. (Cited on page 36.)

[73] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156, 2001. (Cited on page 6.)

[74] Grand Roman Joldes, Adam Wittek, and Karol Miller. Real-time nonlinear finite element computations on gpu–application to neurosurgical simulation. *Computer methods in applied mechanics and engineering*, 199(49):3305–3314, 2010. (Cited on page 6.)

[75] Michael I Jordan. Graphical models. *Statistical Science*, pages 140–155, 2004. (Cited on page 23.)

[76] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004. (Cited on page 6.)

[77] Fredrik Kahl and Peter Strandmark. Generalized roof duality for pseudo-boolean optimization. In *ICCV*, number 1, pages 255–262. Centre for Mathematical Sciences, Lund University, Sweden, IEEE, 2011. (Cited on pages 21 and 24.)

[78] Jorg H Kappes, Bjoern Andres, Fred Hamprecht, Christoph Schnorr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X Kausler, Jan Lellmann, Nikos Komodakis, et al. A comparative study of modern inference techniques for discrete energy minimization problems. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1328–1335. IEEE, 2013. (Cited on page 79.)

[79] Bayesteh G Kashef and Alexander A Sawetauk. A survey of new techniques for image registration and mapping. In *27th Annual Techincal Symposium*, pages 222–239. International Society for Optics and Photonics, 1984. (Cited on page 33.)

[80] Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *NeuroImage*, 46(3):786 – 802, 2009. (Cited on page 6.)

[81] Stefan Klein, Marius Staring, Keelin Murphy, Max Viergever, Josien PW Pluim, et al. Elastix: a toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*, 29(1):196–205, 2010. (Cited on page 6.)

[82] Stefan Klein, Marius Staring, Keelin Murphy, Max a Viergever, and Josien P W Pluim. Elastix: a Toolbox for Intensity-Based Medical Image Registration. *IEEE transactions on medical imaging*, 29(1):196–205, January 2010. (Cited on page 53.)

[83] Pushmeet Kohli and Philip HS Torr. Dynamic graph cuts for efficient inference in markov random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2079–2088, 2007. (Cited on page 19.)

[84] E. Kokiopoulou and P. Frossard. Minimum distance between pattern transformation manifolds: Algorithm and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(7):1225 –1238, july 2009. (Cited on page 36.)

[85] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006. (Cited on page 22.)

[86] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1568–1583, 2006. (Cited on page 62.)

[87] Vladimir Kolmogorov and Carsten Rother. Minimizing nonsubmodular functions with graph cuts - a review. *PAMI*, 29(7):1274–9, July 2007. (Cited on page 19.)

[88] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–59, February 2004. (Cited on page 19.)

[89] Nikos Komodakis. Efficient training for pairwise or higher order CRFs via dual decomposition. *Computer Vision and Pattern Recognition*, pages 1841–1848, June 2011. (Cited on page 85.)

[90] Nikos Komodakis and Nikos Paragios. Beyond Loose LP-relaxations : Optimizing MRFs by Repairing Cycles. *ECCV 08 Proceedings of the 10th European Conference on Computer Vision*, 5304(2):806–820, 2008. (Cited on page 22.)

[91] Nikos Komodakis and Nikos Paragios. Beyond Pairwise Energies : Efficient Optimization for Higher-order MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2985–2992, 2009. (Cited on page 22.)

[92] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF energy minimization and beyond via dual decomposition. *PAMI*, 33(3):531–52, March 2011. (Cited on pages 43 and 66.)

[93] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios. Fast, Approximately Optimal Solutions for Single and Dynamic MRFs. *IEEE Conference on Computer Vision and Pattern Recognition*, D(2):1–8, 2007. (Cited on pages 21 and 68.)

[94] Alexander Kubias, Frank Deinzer, Tobias Feldmann, and Dietrich Paulus. Extended global optimization strategy for rigid 2d/3d image registration. In *Computer Analysis of Images and Patterns*, pages 759–767. Springer, 2007. (Cited on page 75.)

[95] Jan Kybic and Michael Unser. Fast parametric elastic image registration. *Image Processing, IEEE Transactions on*, 12(11):1427–1442, 2003. (Cited on page 31.)

[96] Victor Lempitsky and Andrew Blake. LogCut - Efficient Graph Cut Optimization for Markov Random Fields. *Complexity*, C:1–8, 2007. (Cited on page 19.)

[97] JP Lewis. Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123, 1995. (Cited on page 27.)

[98] Stan Li, Josef Kittler, and Maria Petrou. Matching and recognition of road networks from aerial images. In *Computer Vision—ECCV*, pages 857–861. Springer, 1992. (Cited on page 34.)

[99] Hans-Andrea Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 21(1):28–41, 2004. (Cited on page 23.)

[100] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999. (Cited on page 34.)

[101] Timo Mäkelä, Patrick Clarysse, Outi Sipilä, Nicoleta Pauna, Quoc Cuong Pham, Toivo Katila, and Isabelle E Magnin. A review of cardiac image registration methods. *Medical Imaging, IEEE Transactions on*, 21(9):1011–1021, 2002. (Cited on page 6.)

[102] Primoz Markelj, D Tomaževič, Bostjan Likar, and F Pernuš. A review of 3d/2d registration methods for image-guided interventions. *Medical image analysis*, 16(3):642–661, 2012. (Cited on page 75.)

[103] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001. (Cited on page 34.)

[104] Marc Modat, David M Cash, Pankaj Daga, Gawin P Winston, John S Duncan, and Sébastien Ourselin. A symmetric block-matching framework for global registration. In *SPIE Medical Imaging*, pages 90341D–90341D. International Society for Optics and Photonics, 2014. (Cited on page 53.)

[105] Jan Modersitzki. *Numerical methods for image registration*. Oxford university press, 2003. (Cited on page 61.)

[106] Andrew Y Ng, Michael B Jordan, and Yair Weiss. On Spectral Clustering: Analysis and an algorithm. In T G Dietterich, S Becker, and Z Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, volume 14 of *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press, 2001. (Cited on page 20.)

[107] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990. (Cited on page 8.)

[108] Sergey Osechinskiy and Frithjof Kruggel. Slice-to-volume nonrigid registration of histological sections to mr images of the human brain. *Anatomy research international*, 2011, 2010. (Cited on page 75.)

[109] S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block matching: A general framework to improve robustness of rigid registration of medical images. In A.M. DiGioia and S. Delp, editors, *Third International Conference on Medical Robotics, Imaging And Computer Assisted Surgery (MICCAI 2000)*, volume 1935 of *Lectures Notes in Computer Science*, pages 557–566, Pittsburgh, Penn, USA, octobre 11-14 2000. Springer. (Cited on page 52.)

[110] S Ourselin, A Roche, G Subsol, X Pennec, and N Ayache. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing*, 19(1-2):25–31, 2001. (Cited on page 36.)

[111] Marcello Pelillo and Samuel Rota Bulò. What is a Cluster ? Perspectives from Game Theory. *NIPS Workshop on Clustering*, 2009. (Cited on page 25.)

[112] Mark R Pickering, Abdullah A Muhit, Jennie M Scarvell, and Paul N Smith. A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 5821–5824. IEEE, 2009. (Cited on page 76.)

[113] Steve Pieper, Michael Halle, and Ron Kikinis. 3d slicer. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 632–635. IEEE, 2004. (Cited on page 6.)

[114] Josien PW Pluim, JB Antoine Maintz, Max Viergever, et al. F-information measures in medical image registration. *Medical Imaging, IEEE Transactions on*, 23(12):1508–1516, 2004. (Cited on page 6.)

[115] Brian Potetz. Efficient Belief Propagation for Vision Using Linear Constraint Nodes.pdf. In *Computer Vision and Pattern Recognition*, number 1. Ieee, 2007. (Cited on page 17.)

[116] Michael JD Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162, 1964. (Cited on page 62.)

[117] Michael JD Powell. *" A method for non-linear constraints in minimization problems"*. UKAEA, 1967. (Cited on page 67.)

[118] W.K. Pratt. *Digital Image Processing, 2nd ed.* Wiley, New York, 1991. (Cited on page 35.)

[119] Srikumar Ramalingam, Pushmeet Kohli, and Karteek Alahari. Exact inference in multi-label crfs with higher order cliques. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:1–8, 2008. (Cited on pages 21 and 24.)

[120] Phillip A. Regalia and Kofidis Eleftherios. The Higher-Order Power Method Revisited. In *ICASSP '00 Proceedings of the Acoustics, Speech, and Signal Processing, 2000. on IEEE International Conference*, pages 2709–2712, 2000. (Cited on page 23.)

[121] Nicola Ritter, Robyn Owens, James Cooper, Robert H Eikelboom, and Paul P Van Saarloos. Registration of stereo and temporal images of the retina. *Medical Imaging, IEEE Transactions on*, 18(5):404–418, 1999. (Cited on page 6.)

[122] Alexis Roche, Xavier Pennec, Grégoire Malandain, and Nicholas Ayache. Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information. *Medical Imaging, IEEE Transactions on*, 20(10):1038–1049, 2001. (Cited on page 55.)

[123] Karl Rohr. *Landmark-based image analysis: using geometric and intensity models*, volume 21. Springer Science & Business Media, 2001. (Cited on page 34.)

[124] Carsten Rother, Pushmeet Kohli, Wei Feng, and Jiaya Jia. Minimizing sparse higher order energy functions of discrete variables. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1382–1389, June 2009. (Cited on pages 21 and 24.)

[125] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3):309–314, 2004. (Cited on page 19.)

[126] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky, and Martin Szummer. Optimizing Binary MRFs via Extended Roof Duality. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. (Cited on page 19.)

[127] 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*, 18(8):712–21, August 1999. (Cited on pages 31 and 64.)

[128] Daniel Rueckert, Paul Aljabar, Rolf A Heckemann, Joseph V Hajnal, and Alexander Hammers. Diffeomorphic registration using b-splines. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 702–709. Springer, 2006. (Cited on page 31.)

[129] Harpreet S Sawhney and Rakesh Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(3):235–243, 1999. (Cited on page 36.)

[130] Thomas W Sederberg and Scott R Parry. Free-form deformation of solid geometric models. In *ACM SIGGRAPH computer graphics*, volume 20, pages 151–160. ACM, 1986. (Cited on page 31.)

[131] R. Shekhar and V. Zagrodsky. Mutual information-based rigid and non-rigid registration of ultrasound volumes. *Medical Imaging, IEEE Transactions on*, 21(1):9 –22, jan. 2002. (Cited on pages 36 and 53.)

[132] Raj Shekhar, Vivek Walimbe, Shanker Raja, Vladimir Zagrodsky, Mangesh Kanvinde, Guiyun Wu, and Bohdan Bybel. Automated 3-dimensional elastic registration of whole-body pet and ct from separate or combined scanners. *Journal of Nuclear Medicine*, 46(9):1488–1496, 2005. (Cited on page 6.)

[133] Alexander Shekhovtsov and Ivan Kovtun. Efficient MRF Deformation Model for Non-Rigid Image Matching. *CVPR*, pages 0–5, 2007. (Cited on page 63.)

[134] Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *Medical Imaging, IEEE Transactions on*, 21(11):1421–1439, 2002. (Cited on page 62.)

[135] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, 2000. (Cited on page 20.)

[136] Piotr J Slomka and Richard P Baum. Multimodality image registration with software: state-of-the-art. *European journal of nuclear medicine and molecular imaging*, 36(1):44–55, 2009. (Cited on page 6.)

[137] Stephen Smith and Michael Brady. Susan—a new approach to low level image processing. *International journal of computer vision*, 23(1):45–78, 1997. (Cited on page 34.)

[138] Aristeidis Sotiras. *Discrete Image Registration : a Hybrid Paradigm.* PhD thesis, École Centrale Paris, 2011. (Cited on page 61.)

[139] Aristeidis Sotiras, Davatzikos Christos, and Nikos Paragios. Deformable Medical Image Registration: A Survey. Research Report RR-7919, INRIA, Sep 2012. (Cited on page 6.)

[140] Charles V Stewart, Chia-Ling Tsai, and Badrinath Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *Medical Imaging, IEEE Transactions on*, 22(11):1379–1394, 2003. (Cited on page 6.)

[141] C.V. Stewart, Chia-Ling Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *Medical Imaging, IEEE Transactions on*, 22(11):1379 –1394, nov. 2003. (Cited on page 35.)

[142] George Stockman, Steven Kopstein, and Sanford Benett. Matching images to models for registration and object detection via clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (3):229–241, 1982. (Cited on page 34.)

[143] J-P Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, 2(3):243–260, 1998. (Cited on page 61.)

[144] Philip H S Torr, Pushmeet Kohli, and M. Pawan Kumar. P3 & Beyond : Solving Energies with Higher Order Cliques. *Computer Vision and Pattern Recognition*, pages 1–8, 2007. (Cited on page 19.)

[145] Philip H S Torr, Pushmeet Kohli, and Ladický L'ubor. Robust Higher Order Potentials for Enforcing Label Consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009. (Cited on pages 23 and 24.)

[146] P.H.S. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138 – 156, 2000. (Cited on page 35.)

[147] Antonio Tristán-Vega, Gonzalo Vegas-Sánchez-Ferrero, and Santiago Aja-Fernández. Local similarity measures for demons-like registration algorithms. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 1087–1090. IEEE, 2008. (Cited on page 62.)

[148] Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998. (Cited on page 62.)

[149] Nicholas J Tustison and Brian B Avants. Explicit b-spline regularization in diffeomorphic image registration. *Frontiers in neuroinformatics*, 7, 2013. (Cited on page 6.)

[150] Nicholas J Tustison, Philip A Cook, Arno Klein, Gang Song, Sandhitsu R Das, Jeffrey T Duda, Benjamin M Kandel, Niels van Strien, James R Stone, James C Gee, et al. Large-scale evaluation of ants and freesurfer cortical thickness measurements. *Neuroimage*, 99:166–179, 2014. (Cited on page 6.)

[151] Marcel van Herk, Allison Bruce, AP Guus Kroes, Tarek Shouman, Adriaan Touw, and Joos V Lebesque. Quantification of organ motion during conformal radiotherapy of the prostate by three dimensional image registration. *International Journal of Radiation Oncology* Biology* Physics*, 33(5):1311–1320, 1995. (Cited on page 6.)

[152] Vibhav Vineet and PJ Narayanan. Cuda cuts: Fast graph cuts on the gpu. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008. (Cited on page 19.)

[153] Paul Viola and William M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997. (Cited on pages 27, 36 and 38.)

[154] Nenad Vujovic and Dragana Brzakovic. Establishing the correspondence between control points in pairs of mammographic images. *Image Processing, IEEE Transactions on*, 6(10):1388–1399, 1997. (Cited on page 34.)

[155] MJ Wainwright, TS Jaakkola, and AS Willsky. MAP estimation via agreement on trees: message-passing and linear programming. *Information Theory, IEEE*, 51(October 2002):2005, 2005. (Cited on page 22.)

[156] 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*, 12(5):577–585, 2008. (Cited on pages 28, 29 and 55.)

[157] Wolfgang Wein, Alexander Ladikos, Bernhard Fuerst, Amit Shah, Kanishka Sharma, and Nassir Navab. Global registration of ultrasound to mri using the lc2 metric for enabling neurosurgical guidance. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 34–41. Springer, 2013. (Cited on pages 28 and 29.)

[158] Tomáš Werner. High-arity Interactions , Polyhedral Relaxations , and Cutting Plane Algorithm for Soft Constraint Optimisation (MAP-MRF). *Computing*, 215078(June):1–8, 2008. (Cited on pages 23 and 25.)

[159] Jay West, J Michael Fitzpatrick, Matthew Y Wang, Benoit M Dawant, Calvin R Maurer Jr, Robert M Kessler, Robert J Maciunas, Christian Barillot, Didier Lemoine, Andre Collignon, et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of computer assisted tomography*, 21(4):554–568, 1997. (Cited on page 6.)

[160] Bo Xiang, Chaohui Wang, Jean-Francois Deux, Alain Rahmouni, and Nikos Paragios. Tagged cardiac mr image segmentation using boundary & regional-support and graph-based deformable priors. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1706–1711. IEEE, 2011. (Cited on page 6.)

[161] Michael A Yassa and Craig EL Stark. A quantitative evaluation of cross-participant registration techniques for mri studies of the medial temporal lobe. *Neuroimage*, 44(2):319–327, 2009. (Cited on page 6.)

[162] Ron Zass and Amnon Shashua. Probabilistic graph and hypergraph matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 9:1–8, 2008. (Cited on page 24.)

[163] Yefeng Zheng, Adrian Barbu, Bogdan Georgescu, Michael Scheuering, and Dorin Comaniciu. Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. *Medical Imaging, IEEE Transactions on*, 27(11):1668–1681, 2008. (Cited on page 36.)

[164] Yefeng Zheng, Bogdan Georgescu, and Dorin Comaniciu. Marginal space learning for efficient detection of 2d/3d anatomical structures in medical images. In *Information Processing in Medical Imaging*, pages 411–422. Springer, 2009. (Cited on page 36.)

[165] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with Hypergraphs: Clustering, Classification, and Embedding. *Advances in Neural Information Processing Systems 19*, 19(Figure 1):1601–1608, 2007. (Cited on page 23.)

[166] 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*, 14(4):550–562, 2010. (Cited on pages 36, 53 and 54.)

[167] Djemel Ziou, Salvatore Tabbone, et al. Edge detection techniques-an overview. *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, 8:537–559, 1998. (Cited on page 34.)

[168] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003. (Cited on page 33.)