Méthodes variationnelles pour la segmentation d'images à partir de modèles : applications en imagerie médicale

Raphaël Prevost



UNIVERSITÉ PARIS-DAUPHINE
ECOLE DOCTORALE DE DAUPHINE

Méthodes variationnelles pour la segmentation d'images à partir de modèles – applications en imagerie médicale

---

Thèse soumise pour l'obtention du titre de

Docteur en Sciences - Spécialité Mathématiques Appliquées

(Arrêté du 7 Août 2006)

présentée par

Raphael PREVOST

le 21 Octobre 2013 devant le jury composé de

Rapporteurs:	**Daniel RUECKERT** Professeur, Imperial College London, UK
	Anthony YEZZI Professeur, Georgia Institute of Technology, USA
Examinateurs:	**Nicholas AYACHE** Directeur de recherche, INRIA Asclepios, France
	Gabriel PEYRE Chargé de recherche, Université Paris-Dauphine, France
Directeurs:	**Roberto ARDON** Senior scientist, Philips Research, France
	Laurent D. COHEN Directeur de recherche, Université Paris-Dauphine, France



UNIVERSITÉ PARIS-DAUPHINE
ECOLE DOCTORALE DE DAUPHINE

Variational methods for model-based image segmentation – applications in medical imaging

Thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy in Science - Applied Mathematics

(Decree of August 7th, 2006)

defended by

Raphael PREVOST

on the 21st October 2013 in front of the jury composed of

Reviewers:	**Daniel RUECKERT** Professor, Imperial College London, UK
	Anthony YEZZI Professor, Georgia Institute of Technology, USA
Examinators:	**Nicholas AYACHE** Research Director, INRIA Asclepios, France
	Gabriel PEYRE Researcher, Université Paris-Dauphine, France
Advisors:	**Roberto ARDON** Senior Scientist, Philips Research, France
	Laurent D. COHEN Research Director, Université Paris-Dauphine, France

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs.

Abstract

Within the wide field of medical imaging research, image segmentation is one of the earliest topics but still active. Retrieving the shape of an organ is indeed of high interest for diagnosis, therapy planning or medical research in general. While automated methods are desirable for the sake of reproducibility, they are difficult to design in a generic and robust way. This thesis focuses on model-based methods, which achieve a good trade-off between customization (models carry prior knowledge on the target structure) and genericity (which is important for research efficiency).

The goal of this thesis is to build an efficient segmentation framework that is able to leverage all kinds of external information, *i.e.* pre-segmented databases via statistical learning, other images from the same patient via co-segmentation and user input via live interactions with the algorithm. This work is based on the implicit template deformation framework, which is a variational method relying on an implicit representation of shapes.

After improving the mathematical formulation and the optimization process of this approach, we show its potential on challenging clinical problems. Then, we introduce throughout the remainder of the thesis different generalizations aimed at enriching the model. All the presented extensions are independent but complementary. This thesis therefore eventually transforms the implicit template deformation algorithm into a powerful segmentation framework that is able to (i) learn from a database useful information about both the shape and appearance of the target organ, (ii) use simultaneously multiple images from the same patient. Special care is taken to preserve the most interesting properties of template deformation such as computational efficiency, topology preservation and the ability to take into account user interactions.

The diversity of the clinical applications addressed across this thesis (such as kidney segmentation in ultrasound, contrast-enhanced ultrasound and CT 3D images or myocardium segmentation in ultrasound and MR images) shows the genericity and the effectiveness of the proposed contributions.

Keywords: image segmentation, medical imaging, model-based methods, variational methods, statistical learning, kidney, myocardium, ultrasound, contrast-enhanced ultrasound.

Résumé (French)

Parmi le très large champ de l'imagerie médicale, la segmentation d'image fut l'une des premières tâches étudiées mais constitue encore aujourd'hui un sujet de recherche actif. L'extraction de la forme d'un organe est en effet d'une utilité majeure pour un clinicien afin d'effectuer certains diagnostics, préparer une opération ou simplement la recherche médicale en général. L'automatisation des algorithmes de segmentation est de plus en plus demandée pour garantir la reproductibilité des résultats, mais difficile à atteindre de façon robuste et générique. Le sujet de cette thèse est relié aux méthodes de segmentation à partir de modèles. De telles méthodes présentent un bon compromis entre adaptabilité (un modèle apporte des informations spécifiques à l'application visée) et généricité (qui est importante pour la réutilisabilité des outils et donc l'efficacité de la recherche).

Le but de cette thèse est de construire une méthode de segmentation efficace et capable d'exploiter toutes sortes d'informations externes: des bases de données d'images annotées via l'apprentissage statistique, d'autres images du même patient via une approche de co-segmentation et enfin des interactions avec l'utilisateurs pour guider l'algorithme en temps réel. Ce travail repose sur l'algorithme de segmentation par déformation de modèle implicite, qui est une méthode variationnelle basée sur une représentation implicite des formes.

Après avoir amélioré la formulation (ainsi que son optimisation) de cette approache du point de vie mathématique, nous démontrons son potentiel en l'appliquant à des problèmes cliniques difficiles. Nous introduisons ensuite dans le reste du manuscrit différentes généralisations visant à enrichir le modèle utilisé en incluant des informations sur la forme ou l'apparence de l'organe à segmenter. Toutes ces extensions sont indépendantes mais complémentaires. Cette thèse transforme donc finalement l'algorithme de segmentation par déformation de modèle implicite en une approche puissante et complète qui est capable de (i) apprendre, à partir d'une base de données, des informations utiles à la segmentation sur l'organe cible, (ii) utiliser simultanément plusieurs images du même patient. Une attention particulière est portée à la préservation des propriétés intéressantes de l'algorithme initial, à savoir l'efficacité algorithmique, la préservation de topologie du modèle ainsi que la capacité à prendre en compte des interactions de l'utilisateur.

La diversité des applications cliniques traitées à travers cette thèse (par exemple la segmentation de rein en images 3D d'échographie, échographie de contraste et scanner ou la segmentation de myocarde en images d'échographie ou par résonance magnétique) démontre la généricité et l'efficacité des contributions présentées.

Mots-clés: segmentation d'image, imagerie médicale, méthodes basées modèles, méthodes variationnelles, apprentissage statistique, rein, myocarde, échographie, échographie de contraste.

Contents

Notations

d	image dimension, $d = 2$ for 2D images, $d = 3$ for 3D volumes.
Ω	image domain, an open subset of $\mathbb{R}^d$, $\Omega \subset \mathbb{R}^d$.
Ω_0	template domain, an open subset of $\mathbb{R}^d$, $\Omega \subset \mathbb{R}^d$.
I	an image, an integrable function $I : \Omega \to \mathbb{R}^d$.
$\mathbf{x}$	a point in a domain, $\mathbf{x} \in \Omega$ or $\mathbf{x} \in \Omega_0$.
$\mathbf{y}$	another point in a domain, $\mathbf{y} \in \Omega$ or $\mathbf{y} \in \Omega_0$.
$\mathbf{u}$	vector field, usually defined in the template domain, $\mathbf{u} : \Omega_0 \to \mathbb{R}^d$.
$\mathbf{n}$	outward pointing unit normal field of the boundary Γ.
D	differential operator.
∇	gradient operator in $\mathbb{R}^d$.
div	divergence operator in $\mathbb{R}^d$.
Δ	Laplacian operator in $\mathbb{R}^d$, such that $\Delta = \mathrm{div}(\nabla)$.
H	Heaviside step function.
δ	Dirac distribution.
ϕ	implicit function, positive in the foreground.
$\circ$	functional composition, i.e. $f \circ g(\mathbf{x}) = f\big(g(\mathbf{x})\big)$.
$*$	convolution, i.e. $[f * g](\mathbf{x}) = \displaystyle\int_\Omega f(\mathbf{x})\ g(\mathbf{x}-\mathbf{y})\ d\mathbf{y}$.
K_σ	d-dimensional Gaussian functions of scale σ.
I_σ	image I filtered by a Gaussian function of scale σ, $I_\sigma = K_\sigma * I$.
$I_{\sigma,M}$	image I filtered by a Gaussian function of scale σ via a normalized. convolution with the mask M.
$I_{\sigma,M}$	image I filtered by a Gaussian function of scale σ via a normalized.
$\mathbb{1}_P$	indicator function of the predicate P; returns 1 when P holds, 0 otherwise.

List of Figures

Introduction (French)

English speakers are invited to go directly to Chapter 1 which is a translated and extended version of this introduction.

Segmentation d'images médicales

Depuis la découverte des rayons X par Wilhelm Röntgen en 1895, l'imagerie médicale est un domaine de recherche très actif. L'invention et le développement de diverses modalités d'imagerie (échographie, résonance magnétique, scanner) ont permis de nombreuses avancées dans la recherche médicale, mais ont également révolutionné la pratique clinique. Ces différentes techniques procurent aux cliniciens des informations anatomiques et fonctionnelles sur le corps humains. La constante augmentation du volume de ces données rend nécessaire le développement d'outils informatiques de traitement d'images.

Parmi les différentes tâches de traitement d'images médicales, la segmentation d'organe fut l'une des premières et des plus importantes. Pouvoir récupérer la forme d'un organe est très intéressant d'un point de vue clinique et a de nombreuses applications: mesure précise du volume, estimation de statistiques sur la forme, guidage d'algorithmes de recalage d'images, planning d'opérations chirurgicales.

Pour être utilisé en pratique clinique, un algorithme de segmentation doit être le plus efficace (avoir un temps de calcul raisonnable) et le plus automatique (pour des raisons de reproductibilité) possible. Cependant en pratique il est très difficile de construire de tels algorithmes à cause de différents facteurs impondérables tels que les artéfacts d'acquisition, les effets de volume partiel, et la présence potentielle mais imprévisible de pathologies. Ces manques et corruptions d'information peuvent être en partie compensés par l'utilisation d'informations *a priori* sur l'organe à segmenter. Les méthodes de segmentation basées sur l'utilisation de modèles, sur lesquelles est bâtie cette thèse, représentent un bon compromis entre adaptabilité (un modèle apporte des informations spécifiques à l'application visée) et généricité (qui est importante pour la réutilisabilité des outils et donc l'efficacité de la recherche).

De façon générale, de nombreuses informations externes peuvent être utiles pour guider un algorithme de segmentation. Par example, dans beaucoup de protocoles cliniques, le patient subit plusieurs examens: des images provenant de différentes modalités sont

donc disponibles. En outre, de plus en plus de bases de données d'images annotées sont accessibles publiquement. Elles représentent une précieuse source d'information qu'il est possible d'exploiter en estimant (en *apprenant*) des statistiques, notamment sur la forme ou l'apparence de l'organe à segmenter.

L'objectif général de cette thèse est de présenter une méthode de segmentation capable d'exploiter toutes les sources d'informations possibles. Notre travail s'appuie sur l'algorithme de *déformation de modèle implicite*, récemment décrit dans [Mory, 2011] mais basé sur des travaux antérieurs. Cette approche est particulièrement intéressante par la flexibilité de sa formulation ainsi que son efficacité algorithmique. La déformation de modèle implicite, ainsi que toutes les extensions qui seront proposées dans ce manuscrit, sont des *méthodes variationelles* basées sur une *représentation implicite des formes*. Nous rappelons ces notions dans la prochaine section.

Méthodes variationnelles et segmentation par fonctions implicites

Les méthodes variationnelles reposent sur le calcul des variations et supposent donc que les variables considérées sont continues et différentiables. Une image I n'est donc pas simplement un ensemble discret de pixels mais l'observation d'une fonction continue $I : \Omega \rightarrow \mathbb{R}$ où Ω est un sous-ensemble de $\mathbb{R}^d$ (avec $d = 2$ pour une image, 3 pour un volume).

Segmenter une image I consiste alors à trouver la sous-région de Ω qui contient l'organe d'intérêt. Il existe différentes manières de représenter une sous-région (ou de façon équivalente sa frontière); notons $\mathbb{S}$ l'ensemble des segmentations admissibles pour la représentation choisie.

La segmentation ϕ^* est définie comme le minimum d'un certain critère:

$$\phi^* = \arg\min_{\phi \in \mathbb{S}} E(\phi, I) \tag{1}$$

où E, communément appelé *coût* ou *énergie*, est une fonctionnelle qui quantifie la qualité de la segmentation (plus E est petit, meilleure est la segmentation). Une méthode variationnelle consiste à résoudre un tel problème via le calcul des variations. Sous certaines conditions, un extremum de E est en effet caractérisé par une annulation de sa dérivée. L'approche la plus utilisée est alors de réaliser une descente de gradient, c'est-à-dire faire évoluer ϕ itérativement dans la direction opposée du gradient de l'énergie E. Chaque choix pour $\mathbb{S}$ et E aboutit à une méthode différente; le Chapitre 2 décrit un grand nombre d'entre elles.

Cette thèse repose sur la méthode de déformation de modèle implicite. Dans cette approche, les formes sont représentées implicitement par des fonctions $\phi : \Omega \rightarrow \mathbb{R}$ définies en tout point de l'espace, et à valeurs réelles. La sous-région correspondant à la segmentation est alors l'ensemble des points où la fonction ϕ est positive. En supposant ce domaine borné, l'ensemble de niveau zéro de ϕ définit alors une (ou plusieurs) hypersurface fermée(s). Cette frontière représente la bordure de la segmentation. Cette représentation implicite des formes possède plusieurs avantages: elle peut être utilisée quelle que soit la dimension, et

elle peut naturellement représenter plusieurs objets avec des topologies différentes.

La notion de *déformation de modèle* intervient dans le choix de l'espace des fonctions implicites admissibles $\mathbb{S}$. Au lieu de chercher une fonction ϕ arbitraire, nous définissons celle-ci par rapport à une autre fonction implicite initiale $\phi_0 : \Omega \to \mathbb{R}$ qui servira d'a priori de forme. L'ensemble $\mathbb{S}$ est alors défini de la façon suivante:

$$\mathbb{S} = \{\phi : \Omega \to \mathbb{R} \text{ tel que } \phi = \phi_0 \circ \psi \text{ avec } \psi : \Omega \to \Omega\} . \tag{2}$$

La transformation spatiale ψ, qui devient l'inconnue du problème de minimisation, doit alors être contrainte pour contrôler l'écart de la segmentation par rapport au modèle de forme ϕ_0 et ainsi obtenir un résultat réaliste. Cette contrainte est introduite en ajoutant un terme de pénalisation sur ψ dans l'énergie E. Plus de détails sur cette méthode seront fournis dans la suite du manuscrit. Le plan du reste de la thèse est décrit ci-dessous.

Plan et contributions

A la suite de cette introduction, étendue et traduite en anglais dans le Chapitre 1, nous présentons dans le Chapitre 2 un état de l'art des principales méthodes de segmentation d'images médicales. En particulier, celui-ci comporte une description exhaustive de l'algorithme de *déformation de modèle implicite*, ainsi que ses liens avec les autres approches de segmentation. Le reste du manuscrit décrit les différentes contributions et est organisé comme suit:

- le **Chapitre 3** est une analyse technique de la formulation mathématique de l'algorithme de déformation de modèle implicite. Nous soulevons et résolvons quelques problèmes théoriques dans la formulation et sa minimisation proposée dans [Mory, 2011].
- le **Chapitre 4** démontre le potentiel de l'algorithme de déformation de modèle implicite en l'appliquant au problème de segmentation de rein dans diverses modalités. Pour chacune de ces modalités, nous proposons une méthode automatique pour
  - initialiser le modèle ϕ_0 en tant qu'ellipsoïde. Pour cela, nous utilisons soit une forêt aléatoire de régression pour apprendre et prédire la taille et la position du rein, soit un détecteur d'ellipsoïde robuste basé sur une nouvelle méthode variationnelle;
  - définir l'énergie E après avoir appris une forêt aléatoire qui prédit en chaque pixel la probabilité du pixel d'appartenir au rein. En particulier, nous proposons d'utiliser une stratégie similaire à l'*auto-contexte* [Tu & Bai, 2010] pour prendre en compte la structure spatiale durant l'apprentissage de l'apparence complexe du rein en échographie.
- le **Chapitre 5** introduit une méthode générique qui permet l'utilisation de plusieurs images au sein d'une grande classe d'algorithmes de segmentation. Les images n'ont pas besoin d'être initialement alignées car leur recalage est estimé simultanément à leur segmentation, d'où le nom de *co-segmentation et recalage joints*.

- le **Chapitre 6** décrit comment apprendre et exploiter dans notre méthode de segmentation des informations sur la variabilité anatomique d'un organe, à partir d'une base de données de formes pré-segmentées. En une unique approche, nous apprenons à la fois:
  - le meilleur modèle ϕ_0 qui doit être déformé pour segmenter de nouvelles images. Ce modèle optimal est défini comme la moyenne des formes de la base de données, au sens d'une distance adaptée à l'algorithme de déformation de modèle implicite;
  - un sous-espace de déformations communément observées dans la base de données. Durant la segmentation d'une nouvelle image, nous pouvons alors décomposer la transformation ψ en deux parties: une qui appartient à ce sous-espace privilégié, et l'autre qui n'en fait pas partie. Un nouveau terme de régularisation, qui n'agit que sur cette seconde partie, est introduit.

  Ces améliorations permettent de robustifier l'algorithme tout en maintenant son efficacité algorithmique.

- le **Chapitre 7** généralise la déformation de modèle implicite en introduisant la notion de *modèle étiqueté*. Des sous-régions sont définies dans le référentiel du modèle ϕ_0; chaque sous-région a sa propre énergie basée image E. Différentes parties du modèle peuvent donc être attirées par différentes caractéristiques de l'image. Ces régions peuvent soit être pré-définies manuellement, soit apprises via un processus automatique décrit dans le chapitre.

- le **Chapitre 8** conclut ce manuscrit en résumant les différentes contributions et en discutant de potentielles perspectives.

Toutes les contributions techniques sont indépendantes mais complémentaires. N'importe quelle combinaison de ces améliorations peut donc être utilisée tout en préservant l'efficacité algorithmique de la méthode standard.

D'un point de vue clinique, cette thèse propose une solution aux problèmes suivants:

- **segmentation de rein en CT:** Même si de nombreuses autres méthodes ont déjà été proposées, aucune d'entre elles n'est aussi générique que la nôtre. L'approche proposée est rapide, robuste et est capable de traiter n'importe quelle image CT (qu'importe le champ de vue ou la phase de contraste). Elle est évaluée sur une grande base de données composée de sujets sains et de patients atteints de cancer du rein. Nous montrons ainsi comment l'apprentissage via des forêts aléatoires permet de construire une méthode automatique malgré la grande variabilité des images.

- **segmentation de rein en US et CEUS:** Notre travail constitue la première tentative vers un algorithme complètement automatique pour résoudre ce problème particulièrement difficile.

- **co-segmentation de rein et recalage en US et CEUS:** Nous développons dans le Chapitre 5 une méthode générique pour effectuer simultanément une co-segmentation et un recalage de plusieurs images. Appliquée aux images de rein en échographie (US) et échographie de contraste (CEUS), elle permet d'obtenir une meilleure segmentation de l'organe ainsi qu'un recalage (et donc une fusion possible) des deux images.
- **stabilisation de séquences CT de perfusion:** La méthode de co-segmentation peut être appliquée dans d'autres contextes, en particulier sur des séquences $3D+t$ d'images CT avec injection de produit de contraste. En co-segmentant le rein dans toutes les images, nous sommes capables d'estimer son mouvement. Nous pouvons alors compenser son mouvement et obtenir une stabilisation de la séquence qui surpasse l'état de l'art en terme de précision de recalage et d'estimation de paramètres pharmacocinétiques.

D'autres applications, comme la segmentation de myocarde en imagerie par résonance magnétique (Chapitre 6) et en échographie (Chapitre 7), sont également présentées. Nous insistons cependant sur le fait que toutes les techniques présentées dans ce manuscrit sont génériques et peuvent donc être utilisées dans de nombreux autres problèmes.

Chapter 1

Introduction

Contents

1.1 Segmentation of medical images

Since the discovery of the interesting properties of X-rays by Wilhelm Röntgen in 1895, medical imaging has been and still is an extremely active field of research. The invention and development of ultrasound (US), computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET) imaging – among others – not only constitute major breakthroughs in medical research, but also have changed day-to-day clinical practice.

By enabling a non-invasive visualization of the internal human body, each of those modalities provide clinicians with a lot of valuable information concerning either its anatomy or its functioning. The increasing number of data lead to the development of computer-based processing tools. While the first tools were composed of low-level algorithms, computational capacities have greatly increased in time and the methods have evolved accordingly. Medical imaging research has now its own community, stemming from various fields such as physics, signal and image processing, mathematics, computer science, statistical learning and naturally medicine.

One of the earliest but still open and active topics in medical image analysis is image segmentation. Segmentation of medical images can relate to slightly different tasks such as delineating the contours of an organ or extracting its internal structures but also classifying healthy tissues versus tumours. Here, we will mainly focus on the first aforementioned aspect. Apart from the simple 3D visualization of the shape of an organ, its segmentation has various clinical applications. First, it yields a precise estimate on the organ volume, which is usually

of interest from a diagnosis perspective. More generally, shape analyses provide useful information towards many clinical decisions. On a different level, segmentations can also be exploited to guide image registration methods. Since the development of computer-aided surgery, recent applications also appear in therapy planning and guiding.

Although a high number of interactive segmentation tools are available, there is still a need for automation and efficiency. The time of physicists is precious and cannot be wasted in long waits for computations or tedious interactions with the algorithm, especially during a surgical procedure. Moreover, reproducibility is not guaranteed as soon as a method is not fully automated. In order to compare any set of results or perform a statistical analysis, the segmentation method should indeed be consistent and thus reproducible.

In practice however, it is extremely difficult to design automated and robust segmentation algorithms, especially for medical applications where the challenges are numerous. Depending on the acquisition conditions, images may present artifacts (due to a patient's motion for instance) and be corrupted by a powerful noise. The low resolution of images can also be a source of errors, especially because of the partial volume effect. Finally, pathologies such as lesions may appear unpredictably and trouble a method that has been designed and tested on healthy subjects.

In order to be robust enough to be integrated in clinical routine, a method should be tailored to a particular medical problem. Yet generic approaches are required for software re-usability and therefore research efficiency. We believe that model-based methods, which are the topic of this thesis, achieve a good trade-off between the desired level of unification and customization. Models can carry prior knowledge on the target structure and therefore be specific to the application; conversely, the way of using such models to actually segment the image can stay generic.

To help a segmentation process, it is sensible to use as much external information as possible. In some clinical settings, the patient undergoes several exams and images from different modalities are available. It is important to be able to take into account such valuable information. Furthermore, there is nowadays an increasing number of public databases of labeled images. They represent another precious source of information that should be leveraged, for instance by estimating (or *learning*) statistics on the shape or the appearance of an organ.

The general objective of this thesis is to present a segmentation framework that is able to exploit such kinds of information. Our work is based on the *implicit template deformation* segmentation method, recently proposed in [Mory, 2011] but pioneered by earlier work [Huang et al., 2004; Saddi et al., 2007]. This approach and all its extensions that will be introduced in this manuscript are *variational methods* based on an *implicit representation of shapes*. In the remainder of the chapter, we briefly recall these notions before stating more precisely our contributions.

1.2 Variational image segmentation and implicit template deformation

Variational methods rely on the calculus of variations and therefore assume that variables are continuous and differentiable. Instead of being a discrete set of pixels, an image I is rather modeled as a differentiable function $I : \Omega \to \mathbb{R}$, where Ω is a subset of $\mathbb{R}^d$ (with $d = 2$ or 3 typically).

Segmenting an organ in the image I consists in finding the sub-region of Ω in which this organ lies. There exist various mathematical means of describing this sub-region or, equivalently, its boundary. Let us denote $\mathbb{S}$ the set of admissible segmentations ϕ for the chosen representation. The best segmentation ϕ^* is then defined as a minimum point of a particular criterion:

$$\phi^* = \arg\min_{\phi \in \mathbb{S}} E(\phi, I) \tag{1.1}$$

where E is a cost functional that quantifies how good a segmentation ϕ is for the image I (low E mean good segmentations). For a physical interpretation of this functional is often possible, it can also be called an *energy*. The specificity of variational methods is that (1.1) is minimized using the calculus of variations. Under some assumptions, an extremum of the functional E is indeed characterized by a point where its derivative with respect to ϕ vanishes. The most straightforward approach to find such a point is to perform a gradient descent, *i.e.* letting ϕ iteratively evolve in the opposite direction of the gradient of E.

Each choice for $\mathbb{S}$ and E in (1.1) yields a different method and Chapter 2 describes a large number of them. This thesis explores the potential of the *implicit template deformation* framework, for which we briefly describes such choices:

Image-based energy From the image I, two functions $r_{int}, r_{ext} : \Omega \to \mathbb{R}^+$ must be estimated. They should give at each point $\mathbf{x}$ a cost for $\mathbf{x}$ to be classified as respectively inside or outside the segmentation. For instance, r_{int} is supposed to have low values inside the organ of interest and high values elsewhere. The functional E is then defined as the sum of pointwise costs over the whole image: the contribution of each pixel $\mathbf{x}$ will be $r_{int}(\mathbf{x})$ if $\mathbf{x}$ is inside the segmentation ϕ and $r_{ext}(\mathbf{x})$ otherwise, *i.e.*

$$E(\phi, I) = \int_{\text{inside } \phi} r_{int}(\mathbf{x})\, d\mathbf{x} + \int_{\text{outside } \phi} r_{ext}(\mathbf{x})\, d\mathbf{x} \ . \tag{1.2}$$

Segmentation with an implicit model Any closed hypersurface Γ (*i.e.* a curve in 2D, a surface in 3D) or a set of closed hypersurfaces can be represented as the level-set (typically, the zero level-set) of an implicit function $\phi : \Omega \to \mathbb{R}$ defined on the whole ambient space Ω.

$$\Gamma = \{\mathbf{x} \in \Omega \text{ such that } \phi(\mathbf{x}) = 0\} = \phi^{-1}(0) \tag{1.3}$$

As illustrated in Figure 1.1, the zero level-set of ϕ defines a boundary between a closed

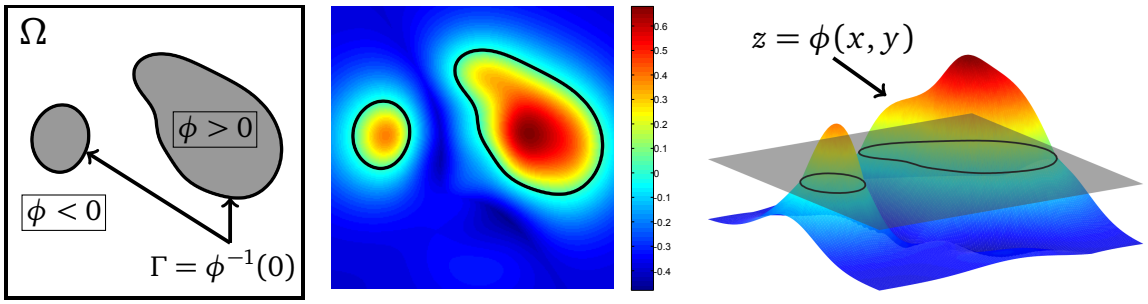


Figure 1.1: Representation of a set of regions via an implicit function. (Left) Definition of the regions defined by the boundary Γ. (Middle) Implicit function ϕ whose zero level-set is Γ. (Right) 3D visualization of the implicit function ϕ.

region (the domain where ϕ is positive) and the rest of the space. This representation will be chosen as it offers several advantages, in particular a good trade-off between flexibility and consistency. First it can be easily extended to any number of dimensions. Furthermore, implicit functions are naturally able to represent objects with different topologies. The explicit segmentation surface can be efficiently retrieved, *e.g.* as a triangular mesh with the *marching cubes* algorithm [Lorensen & Cline, 1987].

The notion of *model* comes from the choice of the space $\mathbb{S}$ of admissible implicit functions. The segmentation ϕ is defined as a deformed version of an initial implicit function $\phi_0 : \Omega \to \mathbb{R}$ that will act as a shape prior. The set $\mathbb{S}$ is thus defined as:

$$\mathbb{S} = \{\phi : \Omega \to \mathbb{R} \text{ such that } \phi = \phi_0 \circ \psi \text{ with } \psi : \Omega \to \Omega\} \,. \tag{1.4}$$

The transformation ψ should naturally be constrained to control the deviation of the segmentation from the shape prior ϕ_0 and have realistic results. This is done by adding a penalization term on ψ to the functional in (1.2), but we defer such details to the comprehensive description of the framework in the next chapter.

1.3 Technical contributions and outline

Following this introductory chapter on the challenges and objectives of this thesis, **Chapter 2** lists and describes the most popular classes of segmentation algorithms. In particular, it comprises an in-depth description of the implicit template deformation framework, around which this thesis revolves. The rest of the manuscript presents the different contributions of this thesis; its outline is detailed hereafter.

- **Chapter 3** consists in a technical discussion on the implicit template deformation framework.

  The formulation of the problem proposed in [Mory, 2011] is analyzed from a mathematical point of view and some theoretical pitfalls concerning the regularity of the solution

and the convergence of the algorithm are pointed out. By modifying the regularization term, we fix these mathematical difficulties while maintaining the computational efficiency of the method. Furthermore, we propose an alternative optimization scheme to ensure that the transformation ψ applied to the model is diffeomorphic. This is a very interesting property since it guarantees that the obtained segmentation shares the same topology as the template ϕ_0;

- **Chapter 4** demonstrates the high potential of the implicit template deformation framework on clinical applications. We address the problem of kidney segmentation in CT, US and contrast-enhanced US 3D images. For each modality, we provide a fully automated method to:
  - initialize the model ϕ_0 as an ellipsoid. This is done either by using regression random forests to learn and predict the kidney's position and size, or by running a robust ellipsoid detector based on a novel variational method;
  - define the two error functions r_{int} and r_{ext}. We actually learn a classifier, also based on random forests, that computes at each voxel the probability to be inside the kidney. We develop a learning strategy that takes into account spatial structure and consistency, and therefore yields a much better estimate of the probability.

  We finally investigate the benefits of enabling the user to guide the segmentation interactively, and we find out that only a few clicks are required to correct cases of failures.

- **Chapter 5** introduces a framework that enables the use of multiple images within a segmentation method. The images do not need to be aligned as their registration is estimated simultaneously to the segmentation, hence the name of *joint co-segmentation and registration*. The genericity of the approach is two-fold:
  - it can be applied to a vast class of variational methods, in particular the implicit template deformation and the robust ellipsoid detection;
  - it can be used in various clinical settings (*e.g.* multi-modal organ segmentation, motion tracking and so on). We detail and quantitatively evaluate our method on some of them.

- **Chapter 6** describes how we can learn and exploit anatomical prior knowledge on an organ from a database of pre-segmented shapes. Within a single approach, we are able to learn both:
  - the best template ϕ_0 that should be deformed to segment new images. We define the optimal template as the mean over the shape database, with respect to a distance that is tailored for the implicit template deformation algorithm;
  - a subspace of deformations that are common in the dataset. When segmenting a new image, we are now able to decompose the transformation ψ into two parts:

one that lies in the learnt subspace and one that does not. This allows us to penalize and constrain only the latter part.

The proposed improvements make the algorithm much more robust while maintaining its computational efficiency.

- **Chapter 7** generalizes the implicit template deformation by introducing the concept of *tagged* templates. Subregions are defined in the referential of the template ϕ_0. Each of these regions has its own r_{int} and r_{ext} functions, *i.e.* different parts of the model can be attracted by different image features. The subregions can be either set manually, or learnt via an automated process that we describe in this chapter.

- **Chapter 8** eventually concludes the manuscript with a synthesis of its contributions and a discussion on potential future work.

The main technical contributions of this thesis are summarized in Figure 1.2. In addition to the formulation of template deformation functional itself, they cover the whole *environment* of the algorithm, *i.e.* what is not actually part of the framework but plays a key role on it: its initialization and optimization process, but also the exploitation of a database by statistical learning, the use of multiple images from the same patient or user interactions.

The different contributions are all independent but complementary. Any combination of the proposed improvements can therefore be used. Moreover, it should be noted that even if they were developed for the implicit template deformation framework, most of them rely on generic ideas that could be applied to a wide variety of segmentation approaches.

1.4 Clinical contributions of this thesis

Throughout this thesis, we show the applicability and the potential to solve clinical problems of all aforementioned technical contributions.

In Chapter 4 of this thesis, we propose different workflows for automated (and interactive in case of failures) kidney segmentation in CT, US and contrast-enhanced US (CEUS) 3D images. A short background on the acquisition of such images is available in Appendix A.

While some previous work have addressed the problem of kidney segmentation in CT volumes, none of them are as generic as our method. The proposed approach is fast, robust and able to deal with images acquired with any field of view, and with any contrast phase. It is thoroughly evaluated on a large and challenging dataset of both healthy subjects and patients with renal cancer. We thus show how statistical learning via random forests enables to design an automated method despite the large variety of image settings.

Regarding kidney segmentation in US and contrast-enhanced US images, our work constitutes the first fully-automated proposal to solve this particularly difficult task. In particular, the auto-context strategy introduced in Section 4.4 is able to capture the complex appearance

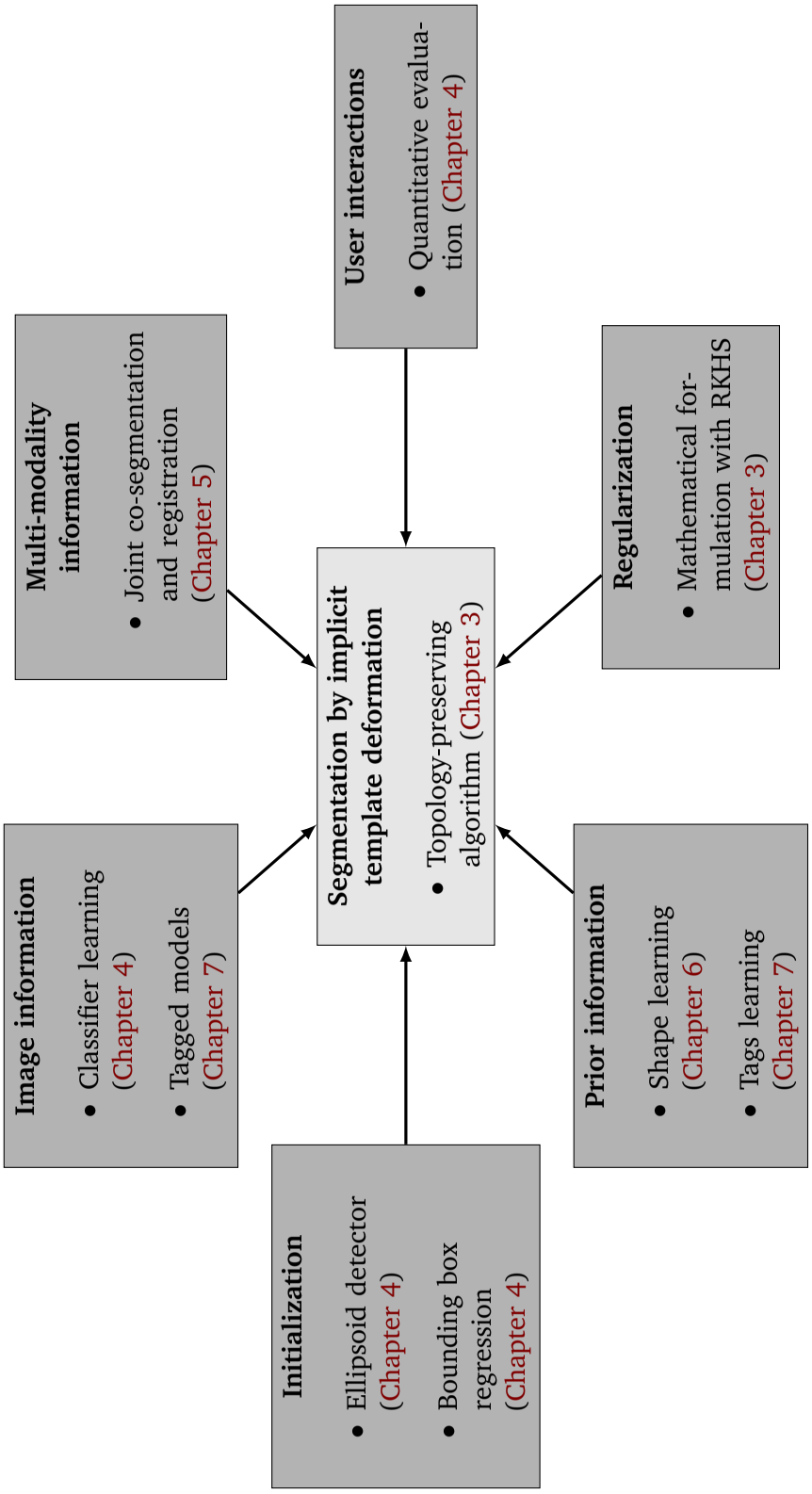


Figure 1.2: Technical contributions of this thesis around the *implicit template deformation* framework.

of kidneys in B-mode images.

We then extend our segmentation methods in Chapter 5 for co-segmentation and registration of both US and CEUS images. This yields even better segmentation and paves the way for new applications including fusion of US and CEUS 3D images, or registration of pre/post therapy images to control the effectiveness of the treatment.

We also apply this joint co-segmentation and registration framework to stabilize a $3D + t$ sequence of free-breathing perfusion CT images of the kidney. Our method outperforms the current state-of-the-art in terms of pharmacokinetical parameters estimations.

Finally, other clinical applications are also addressed in this thesis. More particularly, myocardium segmentation is addressed in Chapter 6 (for MR images) and Chapter 7 (for US images). Our shape learning approach manages to preserve the very thin and intricate shape of the myocardium, while tagged models are able to capture its complex appearance in US images.

Chapter 2

Technical background on image segmentation

Contents

Abstract

This chapter consists in a generic review of the different approaches for medical image segmentation. From low-level to region-based methods, we describe both discrete and continuous approaches. We then discuss more elaborate algorithms based on deformable and statistical models or on atlas propagation. Finally, we provide an in-depth description of the implicit template deformation framework which is at the core of this thesis, and point out relations to aforementioned approaches.

Résumé

Ce chapitre consiste en une revue générale des différentes approches utilisées pour la segmentation d'images médicales. Nous décrivons à la fois des méthodes bas niveau et globales, basées sur des formulation discrètes ou continues. Nous présentons ensuite des algorithmes plus élaborés, comme ceux basés sur des modèles déformables et statistiques, ou sur de la propagation d'atlas. Finalement, nous proposons une description en détail de la méthode de segmentation par déformation de modèle implicite, qui est au centre de cette thèse. Nous discutons également les liens de cette méthode hybride avec les autres approches mentionées.

In this chapter, we draw a short review of different approaches that are commonly used for image segmentation. The purpose of this section is not to provide an exhaustive list of all existing algorithms but to build a very generic taxonomy of the various classes of methods. More specific descriptions of state-of-the-art will be provided throughout this thesis in the adequate sections.

2.1 Definition and notations

Image segmentation is the process of assigning a label to each pixel[1] $\mathbf{x}$ of an image $I : \Omega \to \mathbb{R}$ in order to simplify its representation or extract useful information. Assuming that the number of labels L is finite, we can represent each of them via an integer and the set of possible labels is $[\![0, L-1]\!]$. The goal is therefore to build a function $f : \Omega \to [\![0, L-1]\!]$ (or more generally a relaxed version $f : \Omega \to \mathbb{R}$) that will correctly predict the label of each pixel.

For instance in medical imaging, a common task is to extract the shape of an organ. In such applications, each pixel $\mathbf{x}$ should be classified at either as belonging to the structure of interest ($f(\mathbf{x}) = 1$) or part of the background ($f(\mathbf{x}) = 0$).

To infer such a classification, a set of *features* $x \in \mathbb{R}^F$ is computed at each pixel. Each entry of x represents an information supposedly useful for the segmentation: the intensity of the image at the given point, or more generally the response of a filter. For the sake of simplicity, f will therefore denote either a function over Ω or $\mathbb{R}^F$.

[1] In 3D images, *pixels* are also referred to as *voxels*.

2.2 Low-level methods

Low-level methods aim at assigning the label of each pixel using information from a very local scale (the pixel itself or its neighborhood). Such methods are often efficient and simple to implement. Yet as they do not exploit contextual information, they usually lack robustness and are very sensible to noise or image information ambiguities. They are therefore used most of the time as a first step or an input to more elaborate approaches.

2.2.1 Thresholding

Image thresholding is the most straightforward way of performing image segmentation. The decision f is directly inferred by comparing the intensity of the pixel to one or two thresholds:

$$f(x) = \begin{cases} 1 & \text{if } I(\mathbf{x}) \leq t \\ 0 & \text{otherwise} \end{cases} \quad \text{or} \quad f(\mathbf{x}) = \begin{cases} 1 & \text{if } I(\mathbf{x}) \geq t_1 \text{ and } I(\mathbf{x}) \leq t_2 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Such rules are particularly adapted to extract objects that have a clear and definite contrast with their background, or more generally when the intensities inside and outside the objects of interest respectively lie in disjoint intervals.

The thresholds are parameters set by the user but they can be also automatically estimated using some criterion. The most well-known method [Otsu, 1975], named after Nobuyuki Otsu, consists in selecting the threshold such that the two intra-class variance is minimal (or equivalently the inter-class variance is maximal).

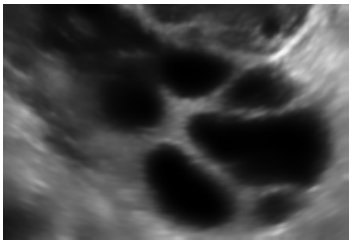
Sometimes a single threshold cannot represent the whole object. This is especially the case when a global change of illumination occur or when the background is heterogeneous. Adaptive thresholding approaches [Sezgin et al., 2004] have been developped to partially fill this pitfall. For instance in [Sauvola & Pietikäinen, 2000], the threshold t is a function of the local mean $\mu(\mathbf{x})$ and standard-deviation $\sigma(\mathbf{x})$ of the image in a small window:

$$t(\mathbf{x}) = \mu(\mathbf{x})\left(1 \pm \frac{1}{2}\left(\frac{\sigma(\mathbf{x})}{R} - 1\right)\right) \tag{2.2}$$

where R is a parameter depending on the image dynamics. Figure 2.1 shows a segmentation result with a fixed and adaptive threshold, on an ultrasound image of an ovary with stimulated follicles.

2.2.2 Clustering approaches

Image segmentation can also be considered as automatically partitioning an image into meaningful subsets, which can be closely related to clustering techniques developed in the field of statistical analysis. They aim at exploring some dataset in order to split it into different disjoint and coherent groups. In image processing, the notion of coherence between two pixels can be defined via various features such as their color, intensity, texture or location.



Original image

Fixed threshold
[Otsu, 1975]

Adaptive threshold
[Sauvola & Pietikäinen, 2000]

Figure 2.1: Segmentation of stimulated follicles in an ultrasound image of an ovary by image thresholding.

The literature on clustering is extremely vast but most of the proposed methods rely on the same approach. They try to explain the data repartition through a statistical model whose parameters are estimated by finding the maximum likelihood estimate of the observed data points $(x_n)_{n=1..N}$. Such problems are generally solved via the *Expectation Maximization* approach [Sundberg, 1974; Dempster et al., 1977] which is composed of two steps, alternatively performed until convergence:

1. Expectation step *(E-step)* – the parameters of each cluster are updated given the current estimates of the assignments.

2. Maximization step *(M-step)* – every point is assigned to the cluster that presents the maximum likelihood.

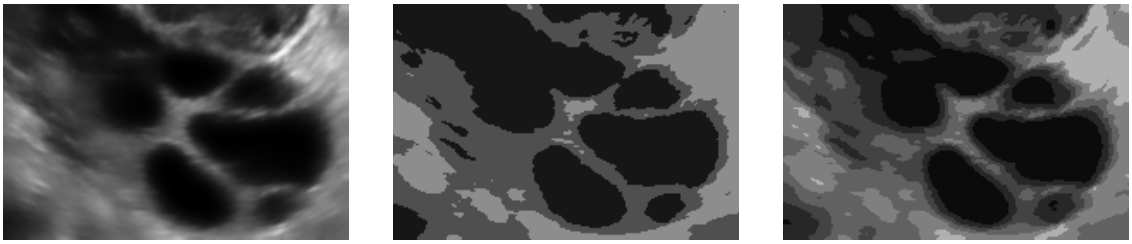
Note that the number of clusters is supposed to be known and their initial guesses have to be provided to the algorithm. As the likelihood criterion is generally non convex, this process usually results in a local optimum that greatly depends on the initialization.

One of the most simple and well-known algorithm is the *k-means*, originally proposed in [MacQueen et al., 1967]. In this framework, the L clusters $(C_l)_l$ are represented by their centers $(\mathbf{c}_l)_l$ which are found as the minimum of the following function

$$E_{kmeans}((\mathbf{c}_l)_l, f) = \sum_{n=1}^{N} \|x_n - \mathbf{c}_{f(x_n)}\|^2 \tag{2.3}$$

where $f : \Omega \to [\![0, L-1]\!]$ is the assignment function that associates each point of the database (*i.e.* each pixel) to the index of its cluster. The choice of the l_2-norm indicates that points are assumed to spread around each center cluster with an isotropic Gaussian distribution. Minimization with respect to $(\mathbf{c}_k)_k$ is performed by averaging the position of all points belonging to each of the clusters, while minimization with respect to f consists in assigning at every point the cluster whose center is the closest. These computations respectively represent the E-step and M-step described above. Figure 2.2 shows results of a k-means clustering run on the image intensities.

The hard cluster assignment f made at each iteration of the k-means algorithms can be



Original image | k-means with 3 clusters | k-means with 6 clusters

Figure 2.2: Clustering of an ultrasound image of an ovary by k-means on the intensities, with various number of clusters.

relaxed to a probability assignment $f : \Omega \to [0,1]^L$ instead of $[\![0, L-1]\!]$. The new algorithm, usually called *fuzzy k-means* [Bezdek, 1981], is more robust (points at the boundaries of several clusters have a weaker contribution to the parameters estimates) and more sensible from a statistical point of view. Another standard generalization is the use of a *Gaussian mixture model*: the clusters are then characterized by both their centers and their covariance matrix. The algorithm is thus able to capture more complex structures but the dependency on the initial guesses is often inevitably increased. Other variants of the approach include the *Expectation Conditional Maximization* [Meng & Rubin, 1993], *Generalized Expectation Maximization* [Neal & Hinton, 1998] or *α-Expectation Maximization* [Matsuyama, 2003].

In all these EM-based algorithms, the number of clusters L has a great influence on the result (see Figure 2.2 with k-means). If L cannot be guessed beforehand, one solution is to run several times the clustering algorithm with various number of clusters and decide afterwards with criteria that are derived from the information theory [Sugar & James, 2003].

Conversely, other approaches do not require L as an input. For instance, the *mean-shift* algorithm [Fukunaga & Hostetler, 1975] does not assume a particular mixture of parametric distributions but only a non-parametric density. By successively shifting each data point towards the barycenter of its neighborhood, the method tends to naturally accumulate subsets of points at the modes of the underlying distribution. Therefore it automatically and intrinsically estimates the number of detected clusters.

All in all, clustering methods perform well to segment multiple objects in images, especially when no or few prior information is available. Yet this is not the case in a context of organ segmentation. Even if the number of clusters is set to 2, there is no guarantee that clustering methods will yield a proper isolation of the organ of interest. If the background is heterogeneous, the clustering algorithm may split it instead of separating it from the target object. When prior information can be used (*i.e.* we know how the object is supposed to look like or an annotated database is available), it is more sensible to explicit model the problem as a classification task between the object of interest and the background.

2.2.3 Supervised classifiers

Supervised classification, unlike unsupervised clustering, is able to leverage some information extracted from a database about the target object via statistical learning [Hastie et al., 2001]. The training database is a set of pairs $\{(x_n, y_n)\}_{n=1...N}$ of pixel features $x_n \in \mathbb{R}^F$ and their associated label ℓ_n which are extracted from a collection of annotated images.

Kernel density estimators

The simplest way to guess the label of a new pixel is to predict the same label as the most similar point in the database [Cover & Hart, 1967]. To increase the robustness of the decision, one may rely not only on one point from the database but on the *k nearest neighbors*. The k different predictions are then aggregated by averaging or majority voting. Such approaches somehow aim at estimating the label expectation (conditionally to the pixel's features) and can therefore be thought of as kernel density estimators. Actually, the kernel corresponding to the *k nearest neighbors* is a uniform ball.

Although it yields very interesting theoretical properties when the size of the training goes to the infinity, it faces problems in practice when the number of samples is fixed. When working in very high dimensions, the whole space is very sparsely sampled and two closest points may actually be separated by a very large distance and consequently have different labels [Beyer et al., 1999].

Optimized classifiers

Learning the function f means that it should be chosen in agreement with the database. It is therefore often defined as the solution of an optimization problem:

$$f^* = \arg\min_{f \in \mathbb{F}} \sum_{n=1}^{N} \mathscr{L}(f(x_n), y_n) \tag{2.4}$$

where $\mathscr{L}$ is a *loss-function*, *i.e.* a function $\mathscr{L} : \mathbb{R} \times [\![0, L-1]\!] \to \mathbb{R}^+$ which is used to assess the quality of a predicted label $f(x_n)$ compared to the real one y_n. Different choices of $\mathscr{L}$ span a large variety of classifiers.

When $\mathscr{L}$ is simply the ℓ_2-distance, (2.4) comes down to the standard least-square regression [Fisher, 1922]. This choice leads to simple computations, especially when f is a linear function of the features, but lacks robustness. When there are only two possible outputs (which is typically the case in segmentation), another very standard choice is the *logistic loss* [Hosmer & Lemeshow, 1989] defined as $\mathscr{L}(z, y) = ln(1 + e^{-zy})$. when $y \in \{-1, 1\}$. The plateaus of this function saturate the contributions of outliers and therefore yield a better robustness. The very popular Support Vector Machines [Cortes & Vapnik, 1995] rely on the *hinge loss* $\mathscr{L}(z, y) = \max(0, 1 - zy)$. The rationale comes from the notion of maximum margin: the classifying function f should separate the training examples from the two labels with a margin as large as possible. Indeed, because of the unitary offset, the hinge loss

penalizes correct but uncertain predictions of the training data.

Naturally, every function such that $\forall n, f(x_n) = \ell_n$ minimizes the loss-function without necessarily capturing any useful information (the function may predict arbitrary values outside the training set). For f to extrapolate reasonably well to unseen data, it must present a certain smoothness that matches the regularity of the data space. This regularity can be enforced either by restricting the space $\mathbb{F}$ of possible functions or by adding a penalization term in (2.4).

A common and convenient[1] choice for $\mathbb{F}$ is the set of linear functions $\{f \text{ s.t. } f(x) = w^T x, w \in \mathbb{R}^F\}$. Even if this may seem extremely restrictive, data are usually linearly separable in high dimensions. Besides one may still include artificial non-linearities by creating new features such as x_1^2 or $x_1 x_2$, at the price of an increased number of dimensions and therefore computational burden. When this strategy is not effective or tractable, kernel methods are a good way to deal with non-linearities [Shawe-Taylor & Cristianini, 2004]. They consist in embedding the data in an infinite dimensional Euclidean space defined through a kernel function (*e.g.* Gaussian). The key advantage is that such methods do not need to compute the actual mapping of the points; only the inner products between the data points are required. Kernel methods are compatible with regression, support vector machines and other standard approaches.

Once the space of functions chosen, it is possible to add a specific term to the optimization problem that will penalize non-smooth functions. It is often set as a norm on the parameterizing variable. For example in the linear case described above, over-fitting can be reduced by adding the term $\|w\|_p$ to (2.4). A vast number of different settings were studied. When $p = 2$, the problem is called *ridge regression* (if the loss function is the ℓ_2-distance) while it is denoted *LASSO* when $p = 1$. A combination of the two penalization is an *elastic net* [Zou & Hastie, 2005].

From weak to strong classifiers

In 1990, Shapire showed in a seminal paper [Schapire, 1990] that classifiers yielding robust but not precise predictions[2] (also called *weak classifiers*) had as much potential as classifiers with a high accuracy on all but a small number of points. His work pioneered the idea that multiple simple classifiers could be exploited to produce a stronger one. This idea of *boosting* weak classifiers has been applied to a large number of approaches. One of the first and most famous is *AdaBoost* (for Adaptive Boosting), presented in [Freund & Schapire, 1995]. It consists in building iteratively a linear combination of weak classifiers. At each iteration, a new weak classifier is added to the linear combination. The important step is that all training data are reweighted so that points that were misclassified at the previous iterations have more importance. The downside of this method is its increased sensibility to outliers.

Artificial neural networks (*e.g. perceptron* [Rosenblatt, 1958]) also rely on the similar

[1] in terms of computations and interpretation.
[2] only slightly better than randomness.

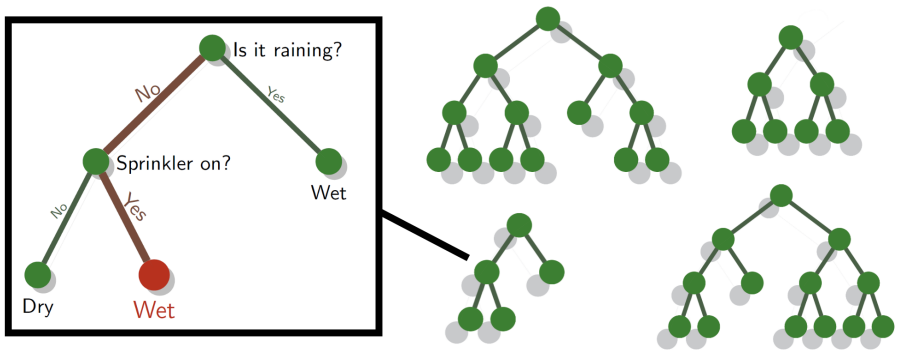


Figure 2.3: Decision forest answering the question: *Is the grass wet ?* Each tree produces an answer based on a chain of simple rules. The different decisions are then fused by averaging or majority voting. (Source: R. Cuingnet)

idea that classification tasks should be performed by combining the results of small units, just as in *decision trees* learning [Quinlan, 1986].

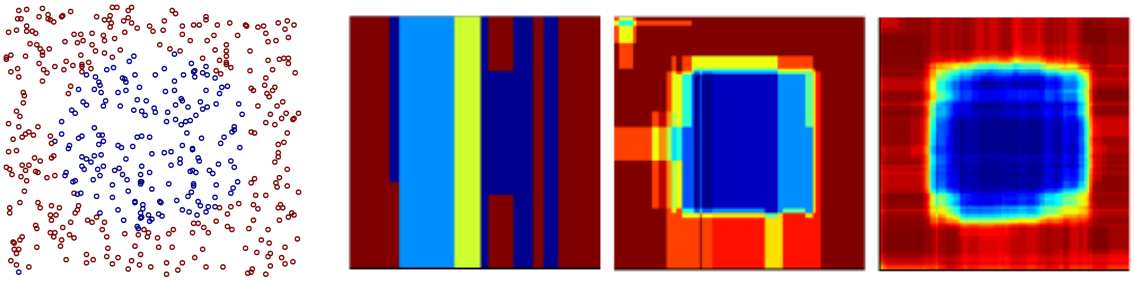
Decision trees and random forests

Although introduced some time ago [Breiman, 2001], *decision trees* and in particular *random forests* have recently gained interest in both the computer vision and medical imaging community. Since we will use this approach several times in this thesis, we provide hereafter a short description of their training and testing. We refer the interested reader to [Criminisi & Shotton, 2013] for a thorough description of random forests.

A *decision tree* is a tree[1] (usually binary) used to produce decisions. Each internal node (*split*) is a test (*i.e.* a weak classifier) on the input data whose answer is supposed to bring some clues towards the decision, while terminal nodes (*leaves*) contain possible answers.

Testing In order to take a decision from a new data point, we submit it to the *root* node. Depending on the answer of this first test, the data is sent to one of the children of the root node. Another test is then applied and the data is once again sent to a child. This process goes on until a terminal node is reached; the decision predicted by the tree is then the value of the leaf. Figure 2.3 (left) illustrates a decision tree that answers to the question *Is the grass wet ?* from simple questions. The first question directly checks whether it is raining (in which case, the answer is definitely positive). If not, another question is needed to give a reliable answer. In this case, an answer (wet/dry) is directly associated to each leaf. However, leaves can also contain more elaborated predictors: since points that fall in the same leaf are supposed to be very similar, simple but robust regressors can be used. For a binary classification in $\{0, 1\}$, a possible solution is to assign to each leaf the proportion of positive points in the training subset that fell into that leaf.

[1] undirected graph without loop, organized in a hierarchical manner



Training set (2 features) blue (-1) vs. red (1)

Random Forests classification with 1, 5 and 50 trees with outputs in $[-1; 1]$

Figure 2.4: Random forests with various numbers of trees for binary classification on a 2D toy example. The weak classifiers at each node are simple thresholds on one of the two coordinates. When the forest has multiple trees, the output is the average of all trees outputs.

Training The idea is therefore to split the features spaces into subregions within which it is easier to predict the result. Training a decision tree means using an annotated database to find the best possible settings for its nodes, *i.e.* training the weak classifiers at each internal node and the regressors at each terminal node. At each node, the parameters of the weak classifier (*e.g.* thresholding or linear classifier) are optimized so that they split the dataset into two subsets that are more organized. Mathematically, this means that the information gain (defined as the difference between the entropy of the set and the sum of the entropies of the subsets) should be maximal. The dataset is then divided between the two children according to the result of the learnt classifier; each child learns its subsequent classifier on its assigned subset, and so on. The process stops when (i) the information gain is too low, (ii) the tree reaches a maximum depth, (iii) there are too few samples in the propagated subset. In such a case, the current node becomes a leaf and a regressor is learnt on the supposedly purified dataset.

From trees to forests A *decision forest* is a set of decision trees. In order to make a decision with a forest, the input data is submitted to all trees of the forest. Each tree produces an answer, and the results are then aggregated (for example by averaging or majority voting) into a unique decision. Naturally, this approach is meaningful only when the trees are different. Actually, the more uncorrelated they are, the more effective the forest. The idea behind *random forests* is to introduce randomness during the learning of the trees, which can be done by several ways. First, each tree is trained on a random subset of the original database. Second, at each node, only a randomly chosen subset of features are considered to choose the best weak classifier. Such an approach is a particular case of a more generic idea called *bootstrap aggregating* or *bagging* [Breiman, 1996]. Figure 2.4 shows on a 2D synthetic example the importance of using multiple and uncorrelated trees. Note that even if the weak classifiers are simple thresholds along one of the dimensions, the tree structure is able to capture non-linearities.

Random forests are thus a powerful and efficient non-linear classifier: a prediction only

requires the computation of a single path from the root to a leaf in each tree, in which simple tests must be evaluated at each level. They are able to deal with large databases (because of the bootstrap strategy, it only requires a subset of the data) and unbalanced classes (thanks to the hierarchical structure of the trees). This is particularly useful for segmentation applications, for which we have a lot of pixels belonging to the target organ (and not just one per image) and even more negative examples.

In conclusion, supervised classifiers are useful for segmentation since they are able to discover from a database some potentially complex rules to predict the label of the pixel. However, in order to segment an object in an image, all pixels have to be classified. For a $128 \times 128 \times 128$ volume, this represents more than 2 million predictions. This high computational cost should be taken into account when choosing the kind of features and classifier used.

Besides, no matter how sophisticated the classifier may be, it is bound to fail when there is too much noise or when it is simply not possible to infer the label from a local point of view. To cope with such problems, it is sensible to enforce spatial consistency and therefore to look directly for objects as connected groups of pixels.

2.3 Objects as discrete groups of pixels

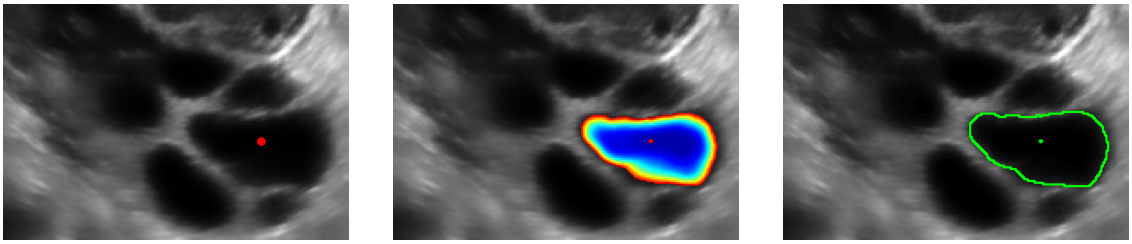
The following methods consider objects as connected sets of pixels, which provides the spatial coherence that was lacking to low-level methods.

2.3.1 Region growing

Region growing is one of the earliest region-based proposed method for image segmentation [Adams & Bischof, 1994] in which the segmentation is iteratively built pixel by pixel. The region stems from an initial point $\mathbf{x}_0$, often called the *seed* point, that is required as input. Then, every pixel in the neighborhood of the current segmentation is visited and added to the region if it is in agreement with it. This step is repeated until the segmentation does not grow anymore. The notion of agreement with the current region can be defined through statistical tests on the pixel intensity. For instance, the pixel $\mathbf{x}$ is added if

$$\frac{|I(\mathbf{x}) - \mu|}{\sigma} \leq \varepsilon \tag{2.5}$$

where μ and σ are respectively the mean and the variance of the intensities in the current segmentation estimate. This algorithm can be thought of as a particular case of *region merging* [Nock & Nielsen, 2004] when the possibly merging regions are the current segmentation and each of its neighboring pixels. One interesting property of region growing is that it always returns a result as a connected set, which is often desirable for segmentation of medical images. Another advantage is that even when decision rules are more complex than (2.5), efficient implementations (even if 3D) can often be derived. The counterpart



Seed point | Arrival time map T | Final segmentation

Figure 2.5: Segmentation of a follicle in an ultrasound image by region growing with fast marching.

of this procedure being so simple is that it is difficult to cast in a variational framework in which an energy functional is actually minimized.

Another means of performing region growing from a seed is to solve an eikonal equation. The eikonal equation is a non-linear partial differential equation simulating a wave propagation starting from the seed point $\mathbf{x}_0$. It can be written in the following form

$$|\nabla T(\mathbf{x})|\, F(\mathbf{x}) = 1 \qquad \text{with} \qquad T(\mathbf{x}_0) = 0 \tag{2.6}$$

where $T(\mathbf{x})$ is the arrival time of the wave front at point $\mathbf{x}$. The function F represents the space-varying propagation speed and must be chosen carefully. It should therefore be high in areas that are considered to be inside the target structure. For instance, if one wants to segment a dark object (as in Figure 2.5), a possible choice is

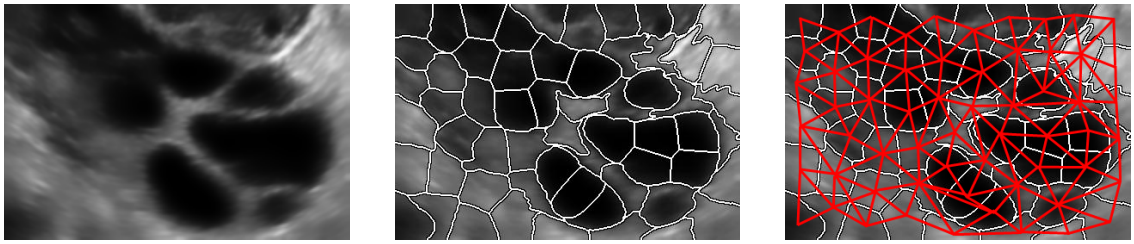
$$F(\mathbf{x}) = \frac{1}{I(\mathbf{x})^2 + \varepsilon^2} + \gamma \tag{2.7}$$

with $\gamma > 0$ a parameter controlling the isotropy of the propagation and ε is a small positive scalar. An efficient way to obtain the numerical solution of (2.6) is the *Fast Marching* method [Malladi & Sethian, 1998]. The segmented object is then defined as the set of locations visited by the front before a certain time t_{final}:

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) < t_{final} \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

This formulation is mathematically more appealing (but more restrictive) than the original formulation, since the segmentation is defined as a ball for an image-based norm. An example of segmentation with this approach is illustrated in Figure 2.5.

Region growing methods are suited for interactive segmentation as they require the seed point as an input and are particularly scalable in 3D applications. However, they suffer from several problems. As points are progressively and independently added, it is



Original image SLIC superpixels Corresponding graph

Figure 2.6: Partitioning of an ultrasound image into SLIC superpixels [Achanta et al., 2010] and its corresponding graph.

not straightforward to take into account priors on the shape of the segmented object [Rose et al., 2007]. Because of this lack of global information, region growing is prone to leaking in regions where the image information is ambiguous. The choice of the stopping criterion is also particularly tricky, although it could be eluded by asking a second input point (when the propagation reaches this point, the algorithm stops). Region growing approaches are therefore particularly interesting to segment objects that have a small, simple and rounded shape such as tumours or cysts [Prevost et al., 2012a] in medical images.

2.3.2 Superpixels

In order to alleviate the computational burden of advanced image processing algorithms, one possible solution is to downsample the images. However, when decimating the image naively, meaningful information is lost (especially edges and texture). The idea behind *superpixels* is to merge neighboring pixels that share some similarities, so that they form small uniform regions. Naturally, the problem of finding superpixels in an image is closely related to clustering (see Section 2.2.2) but additional constraints apply, *e.g.* superpixels are supposed to be connected and compact regions.

Although particularly simple, the *Simple Linear Iterative Clustering* (SLIC) method [Achanta et al., 2010] appears as the state-of-the-art method thanks to both its effectiveness and efficiency. The idea is actually very similar to the *k-means* algorithm. First the superpixel centers are initialized on a regular grid and moved to the closest point with a minimum image gradient norm (such locations are more likely to be good superpixel centers). Then for each superpixel, the closest pixels in a region of interest are added (as in region growing, the notion of closeness is defined with respect to both spatial and color distances) and the cluster center is re-estimated. This alternate process is run until convergence. An example of SLIC superpixels is given in Figure 2.6. The main features of the image are preserved while the number of nodes has greatly decreased (4000 pixels versus 70 superpixels).

Superpixels partitioning can be coupled with a subsequent analysis to provide a segmentation, for example a graph-based approach [Schick et al., 2012].

2.3.3 Graph-based approaches

Images can be considered as a weighted undirected graph $G = (V, E, w)$ such that each node $p \in V$ is a (super)pixel and edges in E exist when two (super-)pixels are neighbors. Most of the time, the weight $w(p, q)$ of each edge (p, q) is a distance based on the (super-)pixels features.

Graphical models

Graphical models consider graphs as a probabilistic model that represents the conditional dependence (the weighted edges) between random variables (the nodes). An image is thus considered as a particular realization of a field of random variables somehow linked together [Perez et al., 1998]. Several variants of random fields exist (*e.g.* conditional random fields, Gaussian random fields), each of them assuming a different probability model, but the most commonly encountered in the computer vision literature is the Markov random field (MRF) [Li, 2009].

Originally exploited in statistical physics, the Markovian hypothesis states that although there might be a global mutual dependency between all variables, it only comes from a combination of local interactions. Basically if two nodes are not neighbors, then they are independent given the rest of the nodes. This assumption allows to express the global probability of the random field as a product over all possible cliques of the graph. Within such a formulation, one can enforce spatial consistency: nodes belonging to the same clique have a higher probability to have similar values.

To use such a framework for image segmentation (*i.e.* to find a label partition in $[\![0, L-1]\!]$), one has to consider each pixel i of an image as a realization of a random variable X_i and attach to it another *hidden* random variable $F_i \in [\![0, L-1]\!]$. The realization f of this discrete random variable will represent the sought label map. After defining the relationship between F and X, one also has to choose the interaction model: for example in the Potts model, a penalization is introduced every time two neighboring pixels have different label values. The problem to solve is finally formulated as follows. Given an observation x (*i.e.* the image) of the random variable X, a segmentation can be achieved by retrieving the most probable configuration of the hidden variable F:

$$f^* = \arg\max_f \mathbb{P}_{F|X}(f|x) \tag{2.9}$$

This approach, called *maximum a posteriori* (MAP) estimation, aims at finding the mode of the probability distribution $F|X$. However solving (2.9) is extremely challenging and for some time, people resorted to simulated annealing, which is a very slow algorithm. In the general case, message passing algorithms such as *belief propagation* proved useful at providing a local optimum [Felzenszwalb & Huttenlocher, 2006]. Although some recent progress has been made by introducing dual techniques [Komodakis et al., 2011], optimization remains challenging when the size of the data is large. However, under certain assumptions of submodularity, another faster way of solving this problem is available. Indeed, casting it into

a *graph-cut* formulation [Boykov et al., 2001] provides significantly better estimates of the solution. Graph-based approaches have been very popular since this link has been unveiled.

Graph partitioning

When an image is considered as a graph, binary segmentation can be thought of as its splitting into two subgraphs (one containing the pixels inside the segmentation and the other with the rest). The unknown is therefore the set of edges that should be removed in order to obtain two disjoint partitions, which is called a *graph cut* [Boykov & Veksler, 2006].

We consider an augmented graph $G' = (V', E')$ such that the set of nodes V' includes all the nodes of the initial graph V and two special terminal nodes: the source s and the sink t. These two nodes are connected to every node of the image. To sum up, we have

$$V' = V \cup \{s, t\} \quad \text{and} \quad E' = E \cup \{(s, p), p \in V\} \cup \{(t, p), p \in V\} \tag{2.10}$$

as illustrated in Figure 2.7. An admissible s/t-cut is a splitting of G' into two subgraphs S and T such that $s \in S$ and $t \in T$. This means that every node $p \in V$ is either linked to s (the label f_p is equal to 1) or p (the label f_p is equal to 0) but never to both. The segmentation $\{p \in V,\ f_p = 1\}$ is then defined as the set of pixels whose nodes belong to S. The best s/t-cut is defined as the s/t-cut with the minimum cost which is defined as the sum of all edges that are removed:

$$\sum_{(p,q) \in E',\ p \in S,\ q \in T} w(p, q) \tag{2.11}$$

In this sum to be minimized, there are two kinds of edges:

- the ones with a terminal node (a subset of $E' \setminus E$), that exploit the image information. For any node p, $w(s, p) > w(t, p)$ when the image information seems to indicate that pixel p is inside the target object. This weight can be determined for instance by any low-level method presented in Section 2.2.

- the ones involving two internal nodes (a subset of E) which bring spatial regularization. In the Potts model described above, $w(p, q) = \mathbb{1}_{f_p \neq f_q}$ which is non-zero when p and q have different labels.

In [Ford & Fulkerson, 1956], an equivalence is made between the min-cut problem and the problem of maximizing the flow through a network. To understand the idea, consider the graph as a pipe network in which the source sends water towards the sink. Each edge represents a pipe whose capacity is the weight of the edge. When the network flow is maximal, some pipes will be saturated. The min-cut/max-flow equivalence states that the set of such pipes defines the minimal s/t-cut of the graph. An efficient algorithm to solve this problem is proposed in [Boykov & Kolmogorov, 2004].

The genericity of the graph-cuts attracted a lot of research work and a number of papers aimed at drawing links with other segmentation methods such as Random walkers [Grady, 2006] or Watershed transforms [Couprie et al., 2011]. Another advantage of the formulation

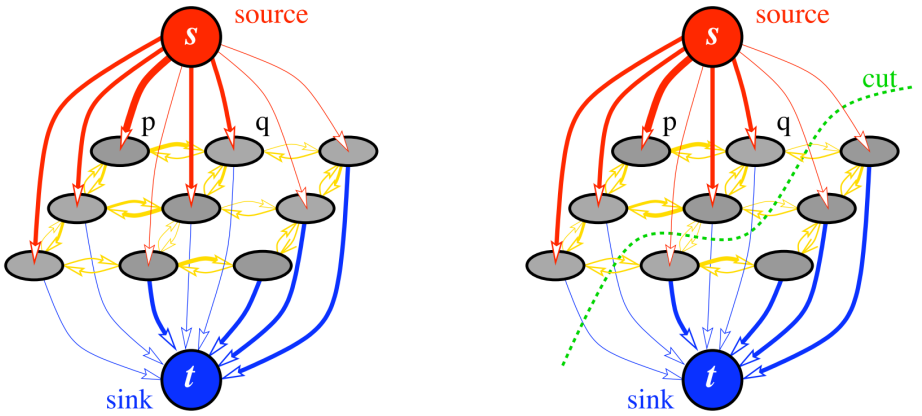


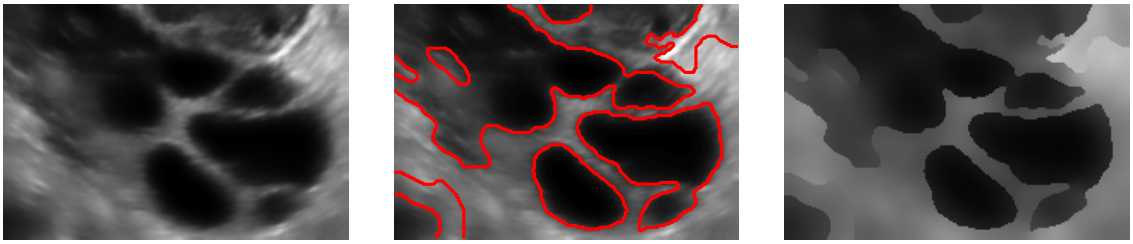
Figure 2.7: Illustration of a graph-cut. (Left) Graph representation where thickness of the arrows indicates similarities. All nodes are linked to artificial nodes (source and sink). (Right) Exemple of a cut on the graph realizing a partition of the nodes between the source and the sink. (Source: [Boykov & Veksler, 2006]).

is that it can be naturally used to perform interactive segmentation [Boykov & Jolly, 2001; Rother et al., 2004]. The weight between the pixels which are indicated by the user to lie inside the target object and the source are simply set to the infinity.

However, graph-cuts present several pitfalls. First, they have a shrinking bias: they tend to produce small and isolated segmentation. They are therefore more suited for segmentation of rather large objects. The *normalized cuts*, proposed in [Shi & Malik, 2000], allows to limit this bias but still suffer from it. Second, optimization algorithms are very computationally expensive, especially in terms of memory. While this is not a problem for 2D images, it is a major issue when dealing with 3D images as the graph does not fit in the memory of standard computers. A possible strategy to alleviate this pitfall is to work on a downsampled volume or reduce the considered neighborhood size. Both solutions however tend to introduce block artifacts in the results (the coarse discretization becomes visible on the surface of the segmentation). This metrication error (”blockiness“) is not unique to graph-cuts but inherent to all discrete methods. Such artifacts however do not appear with methods based on a continuous formulation.

2.4 Region-based variational methods

Unlike in discrete approaches, an image can be seen as a discretized version of a continuous function $I : \Omega \rightarrow \mathbb{R}$. Instead of working from the pixel point of view from the start, it is possible to design a whole method in a continuous setting that will be discretized only at the implementation level. This perspective allows useful optimization tools to find the solutions of optimization problems, in particular functional derivatives.



Original image | Edges Γ | Piecewise-smooth u

Figure 2.8: Piecewise-smooth approximation of an ultrasound image using the Mumford-Shah framework.

2.4.1 The Mumford-Shah piecewise-smooth approximation

Following a similar idea than super-pixels partitioning (see Section 2.3.2), the *Mumford-Shah functional* aims at characterizing the best splitting of an image $I : \Omega \to \mathbb{R}$ into coherent subregions [Mumford & Shah, 1989]. The idea is that an image can be approximated by a piecewise-smooth function $u : \Omega \to \mathbb{R}$. Smooth parts represent objects while discontinuities are their edges. Solving the Mumford-Shah problem means finding the best approximation u as well at its set of discontinuities $\Gamma \subset \Omega$. The D-dimensional energy to be minimized reads

$$E_{MS}(u,\Gamma) = \int_{\Omega} (I(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x} + \mu \int_{\Omega\setminus\Gamma} \|\nabla u(\mathbf{x})\|^2 \, d\mathbf{x} + \lambda \, \mathscr{H}^{D-1}(\Gamma) \tag{2.12}$$

where $\mathscr{H}^{D-1}(\Gamma)$ is the $(D-1)$-dimensional Haussdorff measure of Γ, *i.e.* the perimeter of Γ (length in 2D or surface area in 3D). λ and γ are positive parameters to be tuned. In (2.12), the first term constrains u to be an approximation of I whereas the second one enforces its piecewise-smoothness: strong variations are penalized everywhere outside the edges Γ. The third term behaves as a regularization on Γ as we expect boundaries to be relatively smooth. It also excludes the trivial solution of defining an edge around each pixel. An example of the Mumford-Shah approach on the ultrasound image is provided in Figure 2.8. Main edges are detected and the image is correctly approximated. Yet, the method tends to merge regions when the boundary is too fuzzy.

The discontinuous nature of the unknown Γ makes the minimization of (2.12) very challenging. An interesting result was published by Ambrosio and Tortorelli in 1992. In [Ambrosio & Tortorelli, 1990], they provided an approximation of this regularization term, by using a fuzzy edge indicator $v : \Omega \to [0,1]$ instead of a subset $\Gamma \subset \Omega$. If the support of v is a narrow band with a small enough width ε near Γ, then it is possible to approximate the Haussdorff measure:

$$\mathscr{H}^{D-1}(\Gamma) \approx \int_{\Omega} \left(\frac{v(\mathbf{x})^2}{4\,\varepsilon} + \varepsilon \|\nabla v(\mathbf{x})\|^2 \right) d\mathbf{x} \tag{2.13}$$

which consequently yields an approximation of the Mumford-Shah energy:

$$E_{MS,\varepsilon}(u,v) = \int_\Omega (I(x)-u(\mathbf{x}))^2 \, d\mathbf{x} \; + \; \mu \int_\Omega (1-v(\mathbf{x}))^2 \, \|\nabla u(\mathbf{x})\|^2 \, d\mathbf{x} \\ + \lambda \int_\Omega \left(\frac{v(\mathbf{x})^2}{4\,\varepsilon} + \; \varepsilon \|\nabla v(\mathbf{x})\|^2 \right) d\mathbf{x} \tag{2.14}$$

In the second term, the integral over $\Omega \setminus \Gamma$ is replaced by an integral over the whole domain Ω, weighted by the function $(1-v)^2$. The last term is replaced with the approximation of (2.13). The remarkable property is that $E_{MS,\varepsilon}(u,v)$ converges towards $E_{MS}(u,\Gamma)$ when $\varepsilon \to 0$. This relaxation enabled the use of more efficient algorithms: solving a series of continuous problems is easier than solving the original one. However the choice of ε in tricky in practice and numerical stability issues occur when ε becomes too small.

Research on improving the resolution of the Mumford-Shah formulation is still active. Since the relaxation of Ambrosio and Tortorelli, other algorithms have been introduced and developed [Pock et al., 2009; Grady & Alvino, 2009] but still have a prohibitive cost to process 3D images. Various simplifications of (2.12) have been studied to alleviate its computational burden.

The piecewise-constant model is the particular case of $\mu \to \infty$. In such a scenario, the gradient of the optimal u must vanish almost everywhere (except on Γ): u is a piecewise-constant function. Denoting $(\Omega_i)_{i=1\ldots N}$ the set of subregions generated by Γ, u is completely parametrized by N scalar $(c_i)_{i=1\ldots N}$ (its value in each of the N subregions) given Γ. The functional then reads

$$E_{MSc}((c_i)_i,\Gamma) = \sum_{i=1}^{N} \int_{\Omega_i} (I(\mathbf{x})-c_i(\mathbf{x}))^2 \, d\mathbf{x} + \lambda \, \mathscr{H}^{D-1}(\Gamma) \, . \tag{2.15}$$

When Γ is fixed, one can easily see (via standard calculus of variation) that the optimal value for c_i^* is the average intensity inside the region Ω_i. Indeed, the calculus of variation yields

$$c_i^* = \frac{\int_{\Omega_i} I(\mathbf{x}) \, d\mathbf{x}}{\int_{\Omega_i} d\mathbf{x}} \, . \tag{2.16}$$

On the other hand, the best set of contours Γ^* is a bit more difficult to find. Although it would be possible to parametrize Γ with a set of points and let it evolve following the energy gradient, this assumes that the initial curves (surfaces in 3D) in Γ have the correct topology. In practice, it is hard to guess it beforehand. When Γ is restricted to be a set of *closed* curves/surfaces in 3D, another representation – with an *implicit* function – is available. This hypothesis usually holds in medical applications, especially for organ or tumour segmentation. Chan and Vese exploited this approach to solve the Mumford-Shah problem [Chan & Vese, 2001b] with closed surfaces. In the following subsection, we give more details on this approach for a simplified case of two-phase partitioning.

2.4.2 Region partition with an implicit function

In their seminal paper [Chan & Vese, 2001a], Chan and Vese propose to solve (2.15) with two regions (the foreground and the background). Let $\Omega_1 \subset \Omega$ be the foreground region, $\Omega_2 = \Omega \backslash \Omega_1$ the background region and $\text{Per}(\Omega_1) = \mathscr{H}^{D-1}(\Gamma)$ the perimeter of their common boundary Γ. The minimized energy reads

$$E_{CV}(\Omega_1, c_1, c_2) = \int_{\Omega_1} (I(\mathbf{x}) - c_1)^2 \, d\mathbf{x} + \int_{\Omega \backslash \Omega_1} (I(\mathbf{x}) - c_2)^2 \, d\mathbf{x} + \lambda \, \text{Per}(\Omega_1) \tag{2.17}$$

The breakthrough idea of [Chan & Vese, 2001a] does not concern the energy itself, but the way it is minimized. When the boundary of Ω_1 (denoted Γ) is bounded and closed, it is possible to represent Ω_1 as the positive domain of an implicit function $\phi : \Omega \to \mathbb{R}$. This means in particular that $\Gamma = \phi^{-1}(0)$. With the *Heaviside step function* $H : \mathbb{R} \to \{0, 1\}$ defined by

$$H(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.18}$$

the characteristic function of the foreground Ω_1 is equal to $\mathbf{x} \mapsto f(\mathbf{x}) = H\big(\phi(\mathbf{x})\big)$ while the characteristic function of the background is $\mathbf{x} \mapsto H(-\phi(\mathbf{x})) = 1 - H(\phi(\mathbf{x}))$. The derivative (in the sense of distributions) of the step function H is the Dirac distribution δ. The perimeter of Ω_1 can therefore be expressed with respect to the implicit function:

$$\text{Per}(\Omega_1) = \int_\Omega \|\nabla H(\phi)\| = \int_\Omega \delta(\phi) \|\nabla \phi\| \tag{2.19}$$

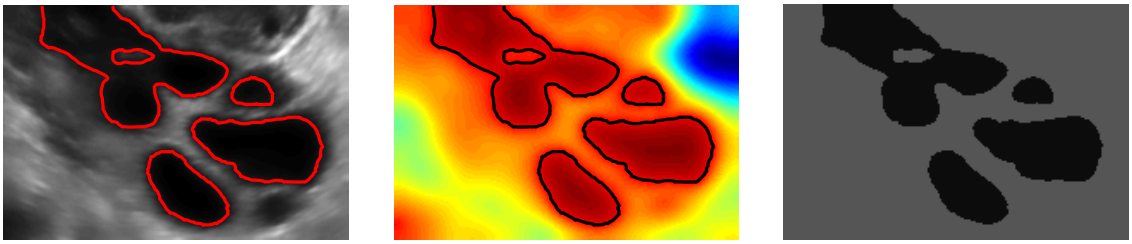
Altogether, E_{CV} can be written as the following expression

$$E_{CV}(\phi, c_1, c_2) = \int_\Omega H(\phi)(I - c_1)^2 + \int_\Omega H(-\phi)(I - c_2)^2 + \lambda \int_\Omega \delta(\phi) \|\nabla \phi\| \tag{2.20}$$

The new unknown variable is the function ϕ, and the three integrals are now defined over the whole fixed domain Ω. The advantage of such a formulation is the simplification of the computation of the functional derivatives with respect to ϕ. The energy E_{CV} is indeed minimized with respect to ϕ with a gradient descent evolution, driven by the following equation:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[-(I - c_1)^2 + (I - c_2)^2 \; + \lambda \, \text{div}\left(\frac{\nabla \phi}{\|\nabla \phi\|} \right) \right] \tag{2.21}$$

The first two terms in the brackets of (2.21) move the contour in a direction that depends whether $I(\mathbf{x})$ is closest to c_1 or c_2. The third term has a smoothing behaviour on the contour represented by the implicit function. Without image forces, it would remove all convexities so that the zero level-set is a shrinking circle. Note that, due to the $\delta(\phi)$ factor, the evolution is driven by forces that are only localized at the boundary of the current



Segmented image | Implicit function ϕ (zero level-set in black) | Piecewise-constant approximation

Figure 2.9: Piecewise-constant approximation of an ultrasound image using the Chan-Vese framework.

segmentation. This enables fast computations of the level-set evolution. On the downside, it makes the optimization problem non-convex since only local information is taken into account. Consequently, the initialization has to be closed to the desired solution. In [Mory & Ardon, 2007], a relaxed version of (2.20) is shown to be convex, which eludes the dependency on the initialization.

Unfortunately, as such, the Chan and Vese formulation is not really usable on medical images: the piecewise-constant assumption is too restrictive does not hold in practice (see Figure 2.9). Nevertheless, it faced a great success in the computer vision community and inspired a lot of research work that aimed at extending it such as the *region competition* approach.

2.4.3 Two-phase region competition

As pointed out in [Zhu & Yuille, 1996], the term $(I(\mathbf{x})-c_i)^2$ in the equations above has a statistical meaning. It can be seen as a negative log-likelihood of the pixel $\mathbf{x}$ to be part of the i-th region if the intensity distribution in this region is Gaussian with mean c_i and unitary standard-deviation. In the more general case of a Gaussian distribution with standard-deviation σ_i, we have

$$\mathbb{P}(\mathbf{x}\in\Omega_i|c_i,\sigma_i) = e^{-\frac{(I(\mathbf{x})-c_i)^2}{2\sigma_i^2}} \Rightarrow -\log\mathbb{P}(\mathbf{x}\in\Omega_i|c_i,\sigma_i) = \frac{1}{2\sigma_i^2}(I(\mathbf{x})-c_i)^2 + \log(\sigma_i) \quad (2.22)$$

The parameters of the intensity distribution model $\boldsymbol{\alpha}_i = (c_i,\sigma_i)$ can be either fixed or regularly updated. Based on (2.22), we can generalize (2.20) with any probability distribution to the

two-phase region competition functional:

$$E_{RC}(\phi, \boldsymbol{\alpha}_{int}, \boldsymbol{\alpha}_{ext}) = \lambda \mathscr{R}(\phi) + \int_{\Omega} H(\phi(\mathbf{x})) \underbrace{\left[-\log \mathbb{P}(\mathbf{x} \in \Omega_1 | \boldsymbol{\alpha}_1)\right]}_{r_{int}(\mathbf{x}; \boldsymbol{\alpha}_{int})} d\mathbf{x} + \int_{\Omega} H(-\phi(\mathbf{x})) \underbrace{\left[-\log \mathbb{P}(\mathbf{x} \in \Omega_2 | \boldsymbol{\alpha}_2)\right]}_{r_{ext}(\mathbf{x}; \boldsymbol{\alpha}_{ext})} d\mathbf{x} \tag{2.23}$$

where $\mathscr{R}$ is any regularization term on ϕ. The integrands r_{int} and r_{ext} act as classification error functions: r_{int} (resp. r_{ext}) is supposed to be high at points that are likely to be outside (resp. inside) the target object.

To better understand the behaviour of the region competition approach, note that the energy in (2.23) can be rewritten as

$$E_{RC}(\phi, \boldsymbol{\alpha}_{int}, \boldsymbol{\alpha}_{ext}) = \lambda\, \mathscr{R}(\phi) + \int_{\Omega} H(\phi(\mathbf{x})) \left[r_{int}(\mathbf{x}; \boldsymbol{\alpha}_{int}) - r_{ext}(\mathbf{x}; \boldsymbol{\alpha}_{ext}) \right] d\mathbf{x} + \int_{\Omega} r_{ext}(\mathbf{x}; \boldsymbol{\alpha}_{ext})\, d\mathbf{x} \tag{2.24}$$

When r_{int} and r_{ext} are fixed, the optimization with respect to ϕ is slightly simpler:

$$\min_{\phi} E_{RC}(\phi) = \min_{\phi} \left\{ \lambda\, \mathscr{R}(\phi) + \int_{\Omega} H(\phi(\mathbf{x})) \left[r_{int}(\mathbf{x}) - r_{ext}(\mathbf{x}) \right] d\mathbf{x} \right\} \tag{2.25}$$

We point out that the optimal implicit function only depends on the difference

$$r(\mathbf{x}) \;=\; r_{int}(\mathbf{x}) \;-\; r_{ext}(\mathbf{x})\,, \tag{2.26}$$

hence the name of *region competition*. Indeed the interior and exterior regions are competing at each pixel of the boundary and, putting aside the effect of regularization, the one with the lowest classification error "conquers" the point.

Table 2.1 gathers the most common choices for r_{int} and r_{ext}. The first three rows refer to global appearance models, while the two next are localized estimates of the intensity distributions (which is more flexible but less robust). Finally, the last row is entitled *Gradient Flux*. This setting is particularly interesting[1] and we detail it in the following remark. More details about the rest of Table 2.1 is available in [Mory, 2011].

Remark 2.1 (Gradient Flux)**.** *Although the formulation* (2.23) *of the region competition approach seems to be limited to region homogeneity measures, edge-based criteria can also be embedded. While region-based terms are usually more robust and alleviate the dependency on the initial conditions, they may yield not so precise results. Conversely, taking into account edge information brings accuracy.*

[1] it will be used several times in this thesis

	Classification error function $r_{int}(\mathbf{x}, \boldsymbol{\alpha}_{int})$ / $r_{ext}(\mathbf{x}, \boldsymbol{\alpha}_{ext})$	Update of parameters $\boldsymbol{\alpha}_{int}$ / $\boldsymbol{\alpha}_{ext}$
Piecewise-Constant	$r_i = (I(\mathbf{x}) - c_i)^2$	c_i = mean
Gaussian	$r_i = \frac{1}{2\sigma_{i^2}} \left(I(\mathbf{x}) - c_i\right)^2 + \log(\sigma_i)$	(c_i, σ_i^2) = (mean, var)
Non-Parametric	$r_i = -\log\left(p_i(I(\mathbf{x}))\right)$	p_i = Parzen PDF
Piecewise-Smooth	$r_i = \int_\Omega K(\mathbf{x}-\mathbf{y})(I(\mathbf{x}) - s_i(\mathbf{y}))^2 d\mathbf{y}$	s_i = Norm. Conv.
Local Non-Parametric	$r_i = -\int_\Omega K(\mathbf{x}-\mathbf{y}) \log p_i(I(\mathbf{x}), \mathbf{y}) d\mathbf{y}$	p_i = Local PDF
Gradient Flux	$r_{int} = \pm\Delta I(\mathbf{x}) \qquad r_{ext} = 0$	N/A

Table 2.1: List of global/local region statistics and flux maximization. Depending on the target anatomical structure, one model can be chosen from this list, optionally with different settings for the foreground and the background regions. For local models, K typically denotes a Gaussian kernel. (Source: [Mory, 2011])

As pointed out in [Vasilevskiy & Siddiqi, 2002], a possible criterion is the image gradient through the region boundary. The idea is that when the boundary of the segmentation matches the edge of an image, its outward normal $\vec{\mathbf{n}}$ *is in the same direction than the gradient of the image (see Figure 2.10.a). This can be measured by the flux* $\mathbf{F}$ *defined as:*

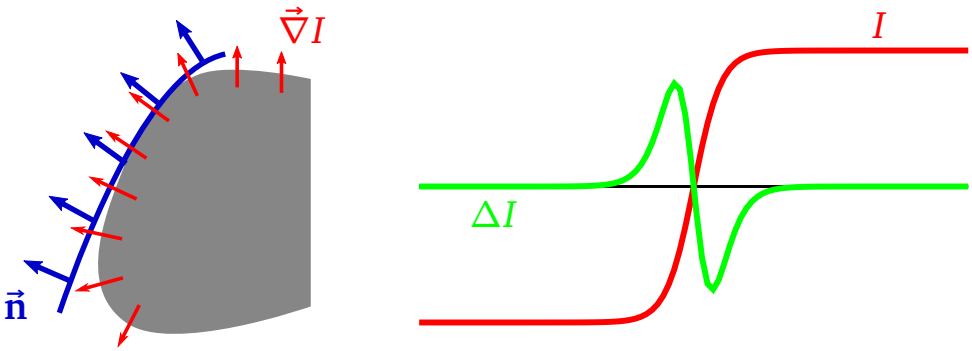
$$\mathbf{F}(\Omega_1) = \int_{\partial\Omega_i} \left\langle \vec{\nabla} I_\sigma(\mathbf{x}), \vec{\mathbf{n}}(\mathbf{x}) \right\rangle dS(\mathbf{x}) \tag{2.27}$$

where $I_\sigma = K_\sigma * I$ *is the image filtered with a Gaussian kernel. The filtering regularizes the image before derivating it but also diffuses the edge information so that the capture range is widened.*

A more convenient way of expressing (2.27) *can be obtained by transforming the boundary interval into a volumic one. This is done by applying the Stokes' theorem (or divergence theorem) and gives*

$$\begin{aligned} \mathbf{F}(\Omega_1) &= \int_{\partial\Omega_1} \left\langle \vec{\nabla} I_\sigma, \vec{\mathbf{n}} \right\rangle dS(\mathbf{x}) = \int_{\Omega_1} div\left(\vec{\nabla} I_\sigma(\mathbf{x})\right)\, d\mathbf{x} \\ &= \int_{\Omega_1} \Delta I_\sigma(\mathbf{x})\, d\mathbf{x} \\ \mathbf{F}(\Omega_1) &= \int_{\Omega} H(\phi(\mathbf{x}))\, \Delta I_\sigma(\mathbf{x})\, d\mathbf{x} \end{aligned} \tag{2.28}$$

where Δ *is the Laplacian operator The tight connexion of gradient flux with the zero-crossings of the Laplacian was also discussed in [Kimmel & Bruckstein, 2003]. As shown in Figure 2.10.b, the Laplacian shows a very strong variation around the edge location. This explains why the edge-based criterion improves the accuracy of the segmentation: if the segmentation boundary is*



(a) (b)

Figure 2.10: Gradient flux maximization. (a) When the boundary matches the edge of the image, its outward normal $\vec{\mathbf{n}}$ is in the same direction than the gradient of the image $\vec{\nabla} I$. (b) The laplacian ΔI crosses zero at the edge location and shows a steep change of sign right before and after it.

slightly shifted, the value of the Laplacian significantly changes. However, as it vanishes quickly around the edge, it has a relatively low capture range.

The last line in (2.28) *draws the link with* (2.25)*. It is therefore possible to use the gradient flux criterion by setting:*

$$r_{int}(\mathbf{x}) = \pm \Delta I_\sigma(\mathbf{x}) \qquad r_{ext}(\mathbf{x}) = 0 \qquad \boldsymbol{\alpha} = \{\emptyset\} \tag{2.29}$$

The sign of ΔI_σ must be chosen in advance to perform either minimization or maximization of $\mathbf{F}$*, which depends on the expected contrast of the target structure. When the target object is brighter than its surrounding, the flux should be minimized. Therefore the correct settings would be $r_{int} = -\Delta I_\sigma$ for the foreground and $r_{ext} = 0$.*

Remark 2.2. *The very flexible* region competition *framework will be used in the implicit template deformation approach described in Section 2.7 which is at the core of this thesis.*

We have seen how various kinds of intensity priors on the target object appearance can be used. However in all the presented approaches so far, the shape was not really constrained (apart from the regularization term in the energies). In the next section, we review some model-based approaches that are able to carry more prior information about the sought shape.

2.5 Deformable models

The transition from low-level methods to deformable models has been a tremendous breakthrough in numerous fields of computer vision and medical imaging such as pattern recognition, tracking and naturally image segmentation. Deformable models have pioneered and inspired a huge number of papers and played an important role in the development of the

most advanced current techniques. Many surveys on these approaches are available, such as [McInerney & Terzopoulos, 1996] and [Blake & Isard, 1998].

2.5.1 Active contours

One of the most influential deformable models was introduced by Kass *et al.* and called *snakes* [Kass et al., 1988]. A snake is a hypersurface $\mathscr{C}$ (curve in 2D, surface in 3D) that deforms in order to match the strong edges of an image. It minimizes the following energy

$$E_{AC}(\mathscr{C}) = \int_0^1 P(\mathscr{C}(s))ds + \alpha \int_0^1 \left\| \frac{\partial \mathscr{C}}{\partial s} \right\|^2 ds + \beta \int_0^1 \left\| \frac{\partial^2 \mathscr{C}}{\partial s^2} \right\|^2 ds \tag{2.30}$$

where $s \in [0;1]$ is the normalized curvilinear coordinate of the curve. The first term is called the *external* energy and depends on the image: P is supposed to be low at high-gradient points (moving on an edge is not penalized). A straightforward choice is therefore:

$$P(\mathbf{x}) = -\|\vec{\nabla} I_\sigma(\mathbf{x})\|^2 . \tag{2.31}$$

The two other terms are the *internal* energy of the curve and only depend on its shape. By penalizing the length and the curvature of $\mathscr{C}$, they enforce regularity.

To find a minimizer of 2.30, we start from an initial curve $\mathscr{C}_0$ and let it evolve towards an Euler-Lagrange state of E_{AC}. We therefore define $\mathscr{C}$ as a function of an artificial time t such that $\mathscr{C}(0) = \mathscr{C}_0$ and

$$\frac{\partial \mathscr{C}}{\partial t} = -\nabla P(\mathscr{C}) + \alpha \frac{\partial^2 \mathscr{C}}{\partial s^2} - \beta \frac{\partial^4 \mathscr{C}}{\partial s^4} \tag{2.32}$$

The two internal and external information can be gathered in a single term. Instead of minimizing both an image-based potential and the Euclidean length of the curve, one can minimize a modified length with an image-based metric:

$$E_{GAC}(\mathscr{C}) = \int_0^1 g\left(\mathscr{C}(s)\right) \left\| \frac{\partial \mathscr{C}}{\partial s} \right\| ds \tag{2.33}$$

where the function g plays a similar role to P in (2.33). It should be positive though, so a possible choice is

$$g(\mathbf{x}) = \frac{1}{1 + \|\vec{\nabla} I_\sigma(\mathbf{x})\|^2} . \tag{2.34}$$

This variant is referred to as *geodesic* [Caselles et al., 1997] or *geometric* [Kichenassamy et al., 1995] active contours. The great advantage of this approach is that the energy is now independent from the parametrization of the contour. When $\mathscr{C}$ is an open curve with fixed end-points, a very efficient procedure to minimize (2.33) was proposed in [Cohen & Kimmel, 1997].

Active contours can be generalized to take into account more prior knowledge. In [Cre-

mers et al., 2007], Cremers *et al.* draw a survey on how to integrate priors about appearance, motion and shape in the level-set active contour framework which we will now describe.

Representation of active contours

To minimize Energy (2.30), one has to define a parametrization for $\mathscr{C}$. We briefly describe the main ones here below and refer the interested reader to [Montagnat et al., 2001] for an exhaustive review.

In the original paper of snakes [Kass et al., 1988], the differential operators in (2.30) are discretized using a finite difference scheme. In 1993, Cohen and Cohen improved both the numerical stability and the efficiency of the method by adopting a finite elements strategy [Cohen & Cohen, 1993], which allowed them to deal with 3D medical data. Other interpolation schemes between control points were then proposed, such as B-splines [Brigger et al., 2000].

To avoid the numerical issues encountered with the explicit representation of the contours and handle topology changes, level-sets methods have been proposed. Unlike the standard snakes, implicit functions are able to represent topology-changing (but necessarily closed) curves. The segmenting curve is defined as the zero level-set of an implicit function: $\mathscr{C} = \phi^{-1}(0)$. All the required quantities can be computed according to ϕ: for instance, the inward unit normal is given at each point by $\frac{\vec{\nabla}\phi}{\|\vec{\nabla}\phi\|}$. The evolution equation (2.32) then becomes

$$\frac{\partial \phi}{\partial t} = g \; div\left(\frac{\vec{\nabla}\phi}{\|\vec{\nabla}\phi\|}\right)\|\vec{\nabla}\phi\| + \left\langle \vec{\nabla} g, \vec{\nabla}\phi \right\rangle \tag{2.35}$$

The parabolic nature of this partial differential equation makes it difficult to solve, and numerical instabilities occur when the time step is too large. Moreover, all level-sets of ϕ (not only the zero level-set) have a tendency to get close to each other near the object boundaries. Therefore $\|\vec{\nabla}\phi\|$ becomes arbitrarily large at these locations, which introduces numerical issues. It is therefore compulsory to periodically re-initialize the implicit function ϕ as the signed distance function of $\mathscr{C}$.

Deformable templates were a first step towards statistical methods [Yuille et al., 1992]. A deformable template is a set of simple geometric elements that can roughly represent a more complex shape. For example an eye model is composed by two parabolic sections (the eye boundary) and a circle (the iris). The rationale is that penalizing the first or second derivatives of the contour does not always correspond to the desired prior: when possible, the segmentation should be more constrained. With deformable models, the number of degrees of liberty is greatly decreased since each optimized parameter (*e.g.* the radius of the circle) has a global influence on the shape.

In the next subsection, we present other model-based approaches that use statistical learning to avoid the tedious and manual construction of those deformable templates.

2.5.2 Statistical models

Statistical methods have been developed to leverage the growing number of available databases in order to build shape priors. They rely on learning, from a set of pre-segmented images, the most likely shape and its variations to constrain the segmentation.

Active shape models

One of the most popular work on this topic was proposed by Cootes and Taylor in their seminal paper on *active shape models* (ASM) [Cootes et al., 1995]. It relies on the point distribution model of shapes: the idea is to represent each shape by a set of landmarks that are present in all the shapes of the database. These landmarks are usually chosen as a set of relevant anatomical features (see top row of Figure 2.11) or simply the vertices positions on 3D meshes. Let $\mathbf{x}_{i,n}$ denote the position of the i-th landmark in the n-th shape of the database. A *mean* shape can be obtained by simply averaging the features positions of all the shapes, *i.e.*

$$\bar{\mathbf{x}}_i = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_{i,n} \tag{2.36}$$

which represents the barycenter of the shape distribution. Assuming that all the shapes from the database are scattered around $\bar{\mathbf{x}}$ according to a Gaussian distribution, it is possible to capture the information about the principal axis of variations by performing a *principal component analysis* (PCA) on the centered data. An eigen-decomposition of the covariance matrix S defined as

$$S_{ij} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_{i,n} - \bar{\mathbf{x}}_i)\,(\mathbf{x}_{j,n} - \bar{\mathbf{x}}_j) \tag{2.37}$$

yields N eigenvectors $(\mathbf{X}_{i,k})_{k=1\ldots N}$ associated to N eigenvalues $(\lambda_k)_{k=1\ldots N}$. The first eigenvector (with the highest eigenvalue) is the direction that explains the most variability in the dataset. The second eigenvector is the second axis of variability that is orthogonal to the first one, and so on.

New realistic shapes (*i.e.* in agreement with the database) can then be generated from affine combinations:

$$\forall i, \quad \mathbf{x}_i[\mathbf{w}] = \bar{\mathbf{x}}_i + \sum_{k=1}^{N}\mathbf{w}_k\mathbf{X}_{i,k} \tag{2.38}$$

where each weight $\mathbf{w}_k$ is supposed to lie in $[-3\lambda_k; 3\lambda_k]$ according to the Gaussian hypothesis. To restrain the model to the main variations and neglect small perturbations that can be due to noise, the sum in (2.38) can be limited to the first $M < N$ modes of variation. For any new shape, it is thus possible to model its probability by

$$\mathbb{P}(\mathbf{w}) = e^{-\sum_{k=1}^{M}\frac{\mathbf{w}_k^2}{2\lambda_k}} \tag{2.39}$$

which can be used as a prior in a segmentation framework.

By constraining the space of possible shapes, segmentation approaches using the ASM often yield more robust results that always look realistic. The counterpart of this very strong constraint is the loss of accuracy (the correct boundary cannot be exactly represented by a linear combination of modes), so a refinement step is usually needed. A comprehensive and detailed review on the use of active shape models for medical imaging is available in [Heimann & Meinzer, 2009].

Implicit models

Just like active shape models, *implicit models* consist in learning a mean shape and its modes of variation. The difference lies in the representation of shapes: they are not considered as sets of points but as implicit functions [Leventon et al., 2000a; Tsai et al., 2003]. Each shape is represented by the signed distance function $\phi : \Omega \to \mathbb{R}$ to its boundary. The statistical analysis (*e.g.* PCA) is then performed on the set $\{\phi_i\}_i$ considered as vectors.

Once the mean and modes are estimated, it is possible to embed them within the geodesic active contours [Leventon et al., 2000a] or the region competition framework [Tsai et al., 2003] to improve segmentation results. As in the active shape model, the segmentation is thus constrained to be the sum of the mean and a linear combination of eigenmodes, whose weights are penalized according to their associated eigenvalues.

The great advantage of this method is that, even though it requires a pre-alignment of shapes, the problem of finding exact correspondences from points to points is eluded. On the downside, the approach suffers from a mathematical pitfall: signed distance functions are not stable by linear combination. For instance, the average shape obtained is not a true distance function. This can be a critical issue when dealing with thin shapes. A solution to this problem was proposed in [Rousson & Paragios, 2002], in which the average implicit function is explicitly constrained to be a distance function, at the price of a greatly increased computational burden.

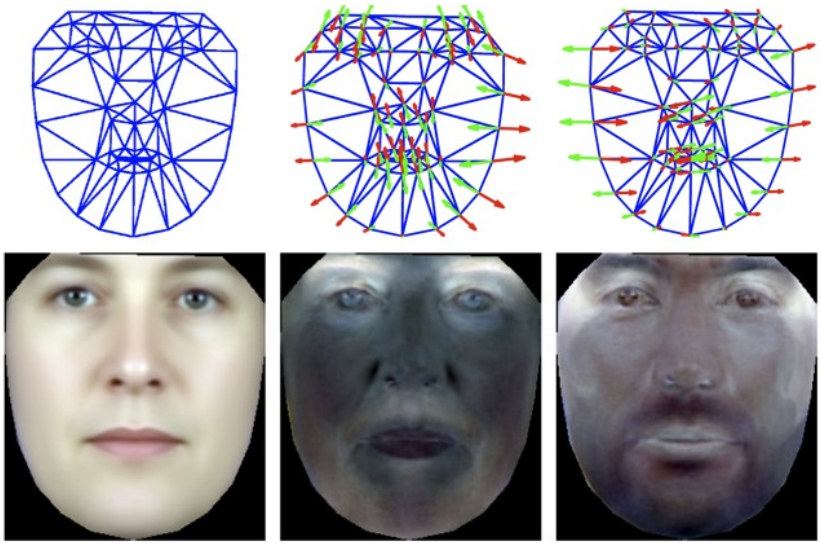
Even if the implicit function stays a distance function, the topology of the represented shape may change. Whether this is a good thing or not depends on the application. For organ segmentation in medical images it is often not desirable, as the topology of the target structure is often fixed and therefore considered as a prior.

Remark 2.3. *Chapter 6 of this manuscript is dedicated to the description of a new shape learning approach. Both the active shape model and the implicit model will be compared to the proposed method in the experiments part.*

Learning appearance with shape

When a database of segmented images is available, it might be interesting to learn not only the shape but also the appearance of the structure of interest.

This is the underlying idea of *active appearance models* (AAM) which are a generalization of the active shape models [Cootes et al., 2001]. First an active shape model is learnt. Each image is then warped into the referential of the mean shape and a PCA is applied to this



(a) (b) (c)

Figure 2.11: Active Appearance Model [Cootes et al., 2001]: mean (a) and two principal modes of variations (b-c) in both shape (top) and appearance (bottom) are learnt. (Source: [Katsamanis et al., 2009])

set of images (considered as vectors in which every entry is a pixel). This step produces the mean and main variations of the image appearance. However, by doing so the shape and appearance are learnt separately while there might be some correlation between them. Therefore, an additional PCA is applied to the concatenation of the vectors representing shape and appearance. Figure 2.11 shows the average and two main modes of variation of both shape and appearance of faces.

The segmentation of a new image consists in finding the best approximation of the object of interest (in terms of shape and appearance) as a linear combination of modes. For instance, the following function can be minimized with respect to the modes weights $\mathbf{w}$:

$$\int_{\Omega} \left(I(\mathbf{x}) - \left(A_0 + \sum_{k=1}^{M} \mathbf{w}_k \, A_k \right) \circ \phi(\mathbf{x}; \mathbf{w}) \right)^2 d\mathbf{x} \tag{2.40}$$

where A_0 is the average appearance, $(A_k)_k$ the appearance eigenmodes. The function $\phi(\cdot; \mathbf{w})$ is the geometric transformation such that each landmark $\mathbf{x}_0$ of the model is warped to the point $\mathbf{x}_0 + \sum_{k=1}^{M} \mathbf{w}_k \mathbf{X}_{0,k}$ (and interpolated elsewhere).

For the appearance to be correctly captured, the sampling of the landmarks should be dense inside the object. For example in Figure 2.11, it would not have been possible to capture correct modes solely with the points on the boundary of the faces. Moreover, since (2.40) is an integral over the whole image, AAM require much more computations than standard algorithms using ASM. On the other hand, AAM methods are usually much more robust. Extensively exploited for face recognition [Tan et al., 2006], the AAM have also been successfully applied to medical problems such as cardiac segmentation [Mitchell

et al., 2002]. This shows that global information on the appearance is useful for clinical applications, hence the development of atlas-based methods.

2.6 Atlas-based segmentation

Following a similar perspective than the active appearance model, atlas-based approaches use contextual information by exploiting the global appearance of the image. Originally proposed in a simplified version for brain structures segmentation in MR images [Sandor & Leahy, 1997], they have now been extended to other kinds of applications such as cardiac segmentation [Isgum et al., 2009] or abdominal organs segmentation [Wolz et al., 2012] in CT images.

The basic idea is illustrated in Figure 2.12. Assume that there is an available segmented image (the *atlas* or *template*). To segment a new image, the principle is to register this atlas to the image with an iconic (intensity-based) method. Then, the obtained transformation is applied to the label map, which yields another label map in the referential of the new image, hence its segmentation. The great advantage is that, provided the transformation is an homeomorphism, the topology of the atlas labels is preserved.

Naturally, the result greatly depends on the chosen template. There might be a high inter-subject variability and when the patient present some abnormalities (lesions, uncommon structures, etc.), the registration algorithm may fail at providing an accurate transformation. One solution to ease the registration process is to select the atlas as the image in a database that is the most similar to the image to be segmented [Commowick & Malandain, 2007]. However, this requires to perform a lot of comparisons and is therefore not suited when the database is too large.

Another strategy to improve the robustness of the approach is to build a *mean* template from the database. By doing so, individual particularities are diluted while the common parts are captured. Several methods have been proposed to construct such an average atlas, but they all more or less consist in registering all images to a common referential and then averaging them. For instance, an elegant and unbiased method based on the notion of Karcher mean is proposed in [Joshi et al., 2004].

Sometimes, the images from the database may be organized in clusters and a single average atlas cannot properly capture the information. It has been proven that using multiple templates yields more robust results [Heckemann et al., 2006]. In this method, all templates are registered to the new image. Each of them then propagates its label information. Finally, the predictions are fused by voting approaches. The standard way is to use STAPLE [Warfield et al., 2004] but other more elaborate strategies have been introduced to help preserving the topology of the structures that can be lost during the fusion step [Cardoso et al., 2012]. We refer the interested reader to [Cabezas et al., 2011] for a review of different atlas-based strategies.

Atlas-based approaches have recently attracted a lot of research interest. In several

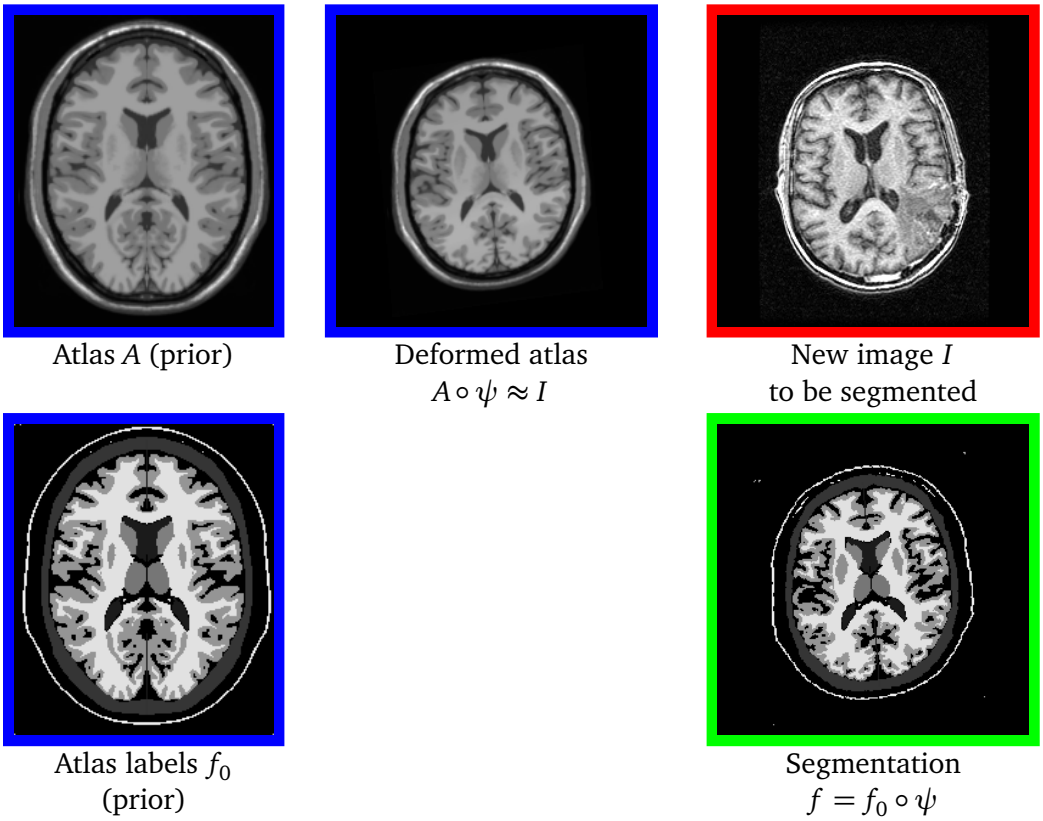


Figure 2.12: Atlas-based segmentation. A labeled atlas is deformed towards the new image to be segmented. The atlas labels are then warped using the same transformation ψ, which yields a segmentation of the new image. (Source: O. Commowick)

segmentation challenges, they are considered as the state-of-the-art method. They are indeed quite robust and, since any number of labels can be propagated, particularly suited to the segmentation of numerous structures. However, they assume that any image can be mapped with a geometric transformation to an atlas. While this is a reasonable assumption for standardized modalities (CT or MR), it does not hold for ultrasound images in which few contextual information is available. Furthermore, they heavily rely on the registration approach adopted and the regularization enforced to the transformation. Therefore, their result might sometimes lack accuracy. Last but not least, atlas-based methods require numerous iconic registrations that have an extremely high computational burden. Indeed, the authors of [Wolz et al., 2012] report a computational time of 3 hours on a computer with eight Intel Xeon cores clocked at 3GHz and 32GB RAM.

The *implicit template deformation* framework, described in the next section, can be considered as a hybrid method between region competition (see Section 2.4.3) and atlas-based segmentation. By exploiting a binary template, it benefits both from the flexibility and efficiency of the region competition approach and from the robustness and shape prior of an atlas-based method.

2.7 Segmentation by implicit template deformation

As this thesis revolves around the implicit template deformation algorithm, we now focus on this framework and present in this section a thorough description of it. This section is largely inspired by the Ph.D. thesis of Benoit Mory [Mory, 2011] and the paper published at the MICCAI 2012 conference [Mory et al., 2012].

2.7.1 Implicit template deformation and region competition

The implicit template deformation framework is based on a region-based segmentation variational formulation (see (2.25) in Section 2.4.3). With an implicit representation ϕ, positive in the foreground, the minimization problem involves a regularization $\mathscr{R}(\phi)$ and a volume integral measuring the classification error, of the form:

$$\min_{\phi\in\mathbb{S}} \Big\{ E(\phi) = \int_\Omega H(\phi(\mathbf{x}))\; r(\mathbf{x})\; d\mathbf{x} + \lambda\; \mathscr{R}(\phi) \Big\} \qquad (2.41)$$
$$\text{with} \quad r(\mathbf{x}) = r_{int}(\mathbf{x}) - r_{ext}(\mathbf{x})$$

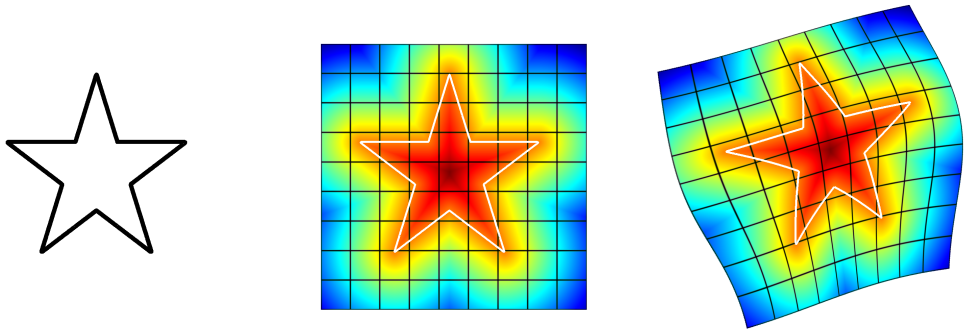
in the supervised case (*i.e.* $r_{int/ext}$ are fixed). The parameter λ is a positive scalar that controls the regularity of the solution. The set $\mathbb{S}$ is the set of implicit functions defined on Ω and real-valued.

Within such a framework, the compliance to a shape prior is often embedded via the regularization term $\mathscr{R}(\phi)$. For instance in [Rousson & Paragios, 2002], $\mathscr{R}$ is defined as the L_2-distance between the segmentation ϕ and the signed distance function of a predefined shape. In this approach, the set of admissible segmentations must thus be restricted to distance functions, *i.e.* $\mathbb{S} = \{\phi : \Omega \to \mathbb{R} \text{ such that } \|\nabla\phi\| = 1\}$. Unfortunately, this technique does not guarantee that the obtained segmentation ϕ shares the same topology of the shape prior. Moreover, for the segmentation ϕ to have smooth boundaries, $\mathscr{R}$ also includes a penalization term on their perimeter. Small or pointy details, even if they are part of the shape prior, are therefore necessarily smoothed out.

In order to avoid such pitfalls, an alternative approach is to rather deform a template shape, assumed here to be defined in its own referential Ω_0, with a geometric transformation $\psi : \Omega \to \Omega_0$. The *template* (or *model*) is implicitly represented by a function $\phi_0 : \Omega_0 \to \mathbb{R}$ and the set of admissible segmenting functions is then defined as

$$\mathbb{S} = \{\phi : \Omega \to \mathbb{R} \text{ such that } \phi = \phi_0 \circ \psi \text{ with } \psi : \Omega \to \Omega_0\}\ . \qquad (2.42)$$

The unknown of the optimization problem (2.41) thus becomes the transformation ψ. The principle of *template deformation* is illustrated in Figure 2.13 where a star-shaped template in 2D is deformed with a regular transformation ψ. Note how the segmentation maintains the sharp arms of the star prior, even after a smooth transformation. Another



(a) star-shaped prior (b) implicit template ϕ_0 (c) deformed $\phi = \phi_0 \circ \psi$

Figure 2.13: Implicit template deformation: the implicit function is sought as a deformed initial template, which allows to keep sharp details. (Source: [Mory, 2011])

advantage of (2.42) is that, unlike standard level-set methods, there is no need to periodically re-initialize the implicit function as a signed distance function: with adequate constraints on ψ, the resulting implicit function ϕ will not present arbitrarily sharp gradients.

The problem of region-based image segmentation by implicit template deformation therefore reads

$$\min_{\psi \in \Psi} \left\{ \int_\Omega H(\phi_0 \circ \psi(\mathbf{x}))\, r(\mathbf{x})\, d\mathbf{x} + \lambda\, \mathscr{R}(\psi) \right\} \tag{2.43}$$

The regularization term $\mathscr{R}$ is naturally transferred to ψ: the transformation ψ should indeed be constrained, otherwise the shape prior would not have any effect. This can be done by choosing adequately the set Ψ and the regularization term $\mathscr{R}$. For instance, a great advantage of the approach is its ability to preserve the topology of the implicit prior ϕ_0 if ψ is a homeomorphism. This point will be thoroughly discussed in Section 3.2, while we rather focus now on the choice of the set Ψ of admissible transformations.

2.7.2 A shape-based template deformation model

Different transformation models were proposed in the literature. In [An & Chen, 2007], Ψ was the set of similarities (rotation, translation and global scaling), which excluded local deformations. Conversely in [Saddi et al., 2007], the authors used a non-rigid transformation that is regularized with a diffeomorphic fluid model with no global constraint, which may result in a very large deviation from the shape prior. Somphone *et al.* defined the deformations based on finite elements with partition of unity to achieve a compromise between global and local transformations [Somphone et al., 2008]. Finally, Huang and Metaxas have proposed to use Free Form Deformations [Rueckert et al., 1999] in the *Metamorphs* algorithm [Huang & Metaxas, 2008].

Here we focus on the transformation model that was specifically exploited in [Mory et al., 2012], which relies on an algebraic definition of shapes as in [Soatto & Yezzi, 2002; Yezzi & Soatto, 2003].

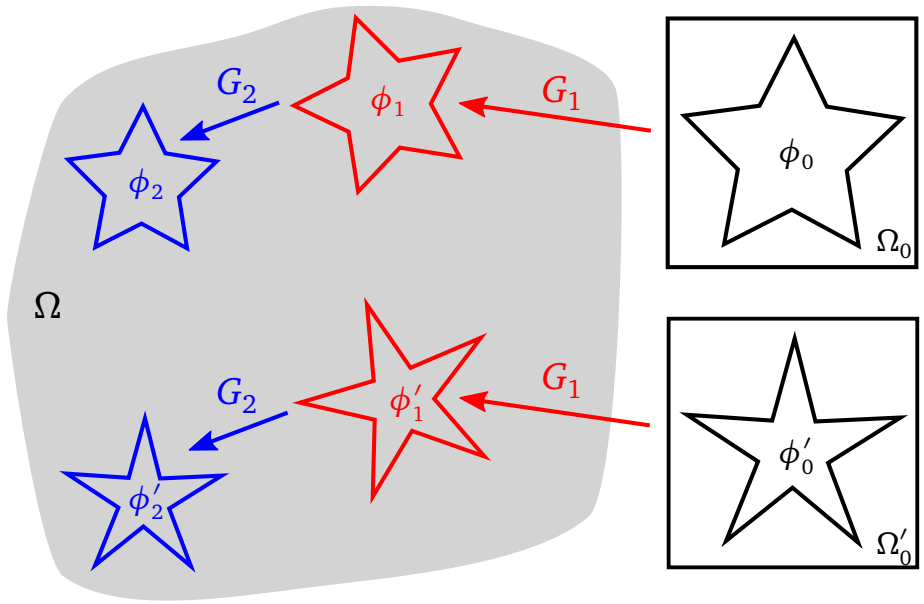


Figure 2.14: Image and shape referentials in $\mathbb{R}^2$. The shape priors are provided in centered and normalized referentials Ω_0 and Ω'_0. (Top) Shapes from the orbit of the prior ϕ_0: ϕ_1 and ϕ_2 can be exactly matched to ϕ_0 with global transforms from $\mathbb{G}$. (Bottom) Shapes lying in another orbit, this time generated from ϕ'_0. The two shapes ϕ_0 and ϕ'_0 are different since they cannot be matched with a transformation in $\mathbb{G}$.

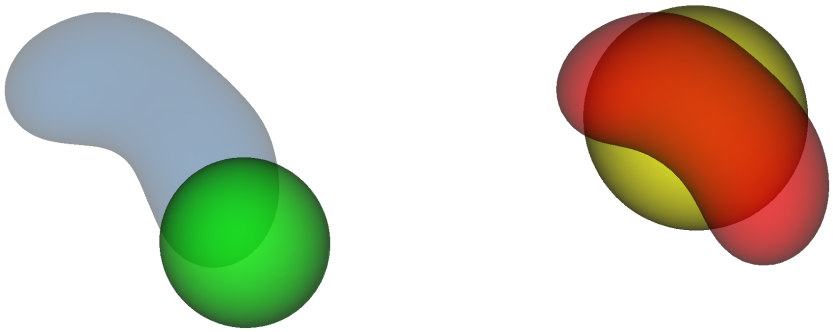
An algebraic definition of shapes

The *shape* of an object is a geometrical representation of the portion of the space occupied by this object (or equivalently a representation of its boundary). Usually, the concept of shape does not comprehend any notion on location, orientation or size. The shape prior ϕ_0 in (2.42) is thus defined in an abstract referential $\Omega_0 \subset \mathbb{R}^d$ that is centered and normalized. For this shape to represent a real object, it should be mapped to the image referential $\Omega \subset \mathbb{R}^d$ via a global transformation $G \in \mathbb{G}$ that does include the information on the object's position, orientation and scale. The set $\mathbb{G}$ of such transformations represents the possible *poses* of the shape in the image.

As transformations from $\mathbb{R}^d$ to $\mathbb{R}^d$, the elements of this set form a *group* $(\mathbb{G}, \circ)$ with the functional composition operation. Considering the *group action* of $\mathbb{G}$ on $\mathbb{S}$, we define an implicit shape as an *orbit* with respect to this action, *i.e.* a class of the equivalence relation $R_{\mathbb{G}}$ defined on $\mathbb{S}$ by:

$$\forall (\phi_1, \phi_2) \in \mathbb{S}^2, \quad \phi_1 \, R_{\mathbb{G}} \, \phi_2 \iff \exists \, G \in \mathbb{G}, \; \phi_2 = \phi_1 \circ G \tag{2.44}$$

Thus, the orbits form a partition of $\mathbb{S}$, meaning that any closed surface belongs to one and only one shape; this partition – and hence the shape space – depends on the choice of $\mathbb{G}$. We consider groups of global parametric transforms, typically similarities.



(a) implicit template $\{\phi_0 = 0\}$ (green) and target shape (blue)

(b) optimal pose $\{\phi_0 \circ G = 0\}$ (yellow) and deformation $\{\phi_0 \circ L \circ G = 0\}$ (red)

Figure 2.15: Transformation decomposition into pose and deformation. (Source: [Mory, 2011])

This notion of shape is illustrated in Figure 2.14. Since ϕ_1 and ϕ_2 can be exactly matched to ϕ_0 with transforms G_1 and $G_2 \circ G_1$ from $\mathbb{G}$, they belong to its orbit and represent the same shape. On the bottom row, ϕ_1' and ϕ_2' lie in another orbit. Since they cannot be exactly matched to ϕ_0 with transforms from $\mathbb{G}$, they represent different shapes. We also point out the difference between the referentials of the models Ω_0 and Ω_0' and the image referential Ω. While in practice both are subsets of $\mathbb{R}^d$, we prefer to keep this distinction when defining variables (either with respect to the image or the model) for the sake of clarity.

An orbit-based transformation model

The notion of orbits is embedded in the template deformation model by explicitly separating the *pose* of the prior and its *shape change*. This is done by considering the following set of transformations:

$$\Psi = \left\{ \psi : \Omega \to \Omega_0 = L \circ G \text{ such that } \begin{array}{l} G : \Omega \to \Omega_0,\ G \in \mathbb{G} \\ L : \Omega_0 \to \Omega_0 \end{array} \right\} \tag{2.45}$$

where L is a locally defined transformation inducing the shape deformation. This definition, relying on the composition of two different kinds of transformations, is depicted in Figure 2.15 for a synthetic example in 3D. The template representing a sphere (in green in Figure 2.15.a), is transformed to reach a target bean shape. This transformation is two-fold: a global scaling and translation G (resulting in the yellow sphere in Figure 2.15.b) and a non-rigid deformation L to actually match the target shape (in red in Figure 2.15.b).

In other words, the deformation L acting in the model referential Ω_0 changes the shape of the template and thus *selects an orbit* in $\mathbb{S}$. Conversely, G allows the segmenting implicit function to *move within this orbit* but preserves the shape of the template. More details on the parametrization of both transformations will be given in the next subsection.

The great benefit of the decomposition in (2.45) is the possibility to define the regularization term $\mathscr{R}(\psi)$ intrinsically to the shape.

A pose-invariant shape term

The decomposition $\psi = L \circ G$ indeed allows to express the shape prior term independently from the pose, *i.e.* $\mathscr{R}(\psi) = \mathscr{R}(L)$. The purpose of $\mathscr{R}$ is to control the deviation *in terms of shape* of the segmentation ϕ from the prior shape ϕ_0. To that end, the magnitude of the induced deformation is penalized via the L^2-norm, *i.e.*

$$\mathscr{R}(L) = \frac{1}{2}\|L - \mathbf{Id}\|_2^2 \tag{2.46}$$

When the image information is ambiguous, this term dominates and locally attracts L towards the identity. This means that in such uncertain regions, the segmentation will merely be an interpolation driven by the shape prior.

Note also that since only the shape deformation is penalized, all global components of the transformation ψ (such as translations) will naturally tend to be included in G rather than L.

As defined in (2.46), the regularization term does not impose any regularity on L. In [Mory, 2011], smoothness is rather directly embedded in the parametrization of L, which we discuss now.

2.7.3 Parametrization and numerical details

This subsection presents the chosen parametrization for both transformations and their optimization with respect to the energy E in (2.41).

Pose $G : \Omega \to \Omega_0$ is chosen, as mentioned above, as a parametric transform such as a similarity. Its purpose is to globally register the template with the target object in the image, so that the deformation required to actually match it is minimal. From an implementation point of view, G is therefore represented by a matrix in homogeneous coordinates defined by 7 parameters $\mathbf{p} = \{\mathbf{p}_i\}_{i=1\cdots7}$ and noted $G_{\mathbf{p}}$.

Deformation $L : \Omega_0 \to \Omega_0$ is a free-form deformation [Rueckert et al., 1999] expressed using a displacement field $\mathbf{u} \in L^2(\Omega_0, \mathbb{R}^d)$ in the template referential Ω_0 as

$$L = \mathbf{Id} + \mathbf{u} \tag{2.47}$$

This model corresponds to what is commonly called the *small deformations* framework, which comes from the fact that L may lose its invertibility when the deformation field $\mathbf{u}$ becomes large (more details will be given in Chapter 3).

Issues also occur when $\mathbf{u}$ is not smooth enough. It would have been possible to include additional terms on the derivatives of $\mathbf{u}$ in $\mathscr{R}(\psi)$, but this comes with a computational

price. Instead, the field $\mathbf{u}$ is directly defined as a filtered version of an integrable unknown displacement field $\mathbf{v} \in L^2(\Omega_0, \mathbb{R}^d)$,

$$\mathbf{u}(\mathbf{x}) = \big[K_\sigma * \mathbf{v}\big](\mathbf{x}) = \int_{\Omega_0} K_\sigma(\mathbf{x}-\mathbf{y})\ \mathbf{v}(\mathbf{y})\ d\mathbf{y} \tag{2.48}$$

where K_σ is a Gaussian kernel of scale σ and the unknown becomes the field $\mathbf{v}$. The rationale of this deformation model is similar to the *Demons* image registration algorithm [Thirion, 1998]. It allows to enforce very efficiently long-range regularity using recursive implementations of the Gaussian filtering.

The scale σ in (2.48) is the only parameter and corresponds to the spatial extent of the smoothness. Since L is defined in the template referential Ω_0, it is easy to tune this parameter which is intrinsic to the shape prior (independently from its pose).

Remark 2.4. *In [Mory, 2011], the obtained deformation L is claimed to be regular by construction. In Chapter 3, we will see that this statement is not so true and should be moderated. We will also introduce a new definition of the displacement field (based on the Gaussian reproducing kernel Hilbert space) that solves such theoretical problems.*

The sought transformation $\psi = \psi_{\mathbf{p},\mathbf{v}}$ depends, as seen above, on a set of parameters $\mathbf{p}$ that defines the global transformation G and a hidden displacement field $\mathbf{v}$ encoding the transformation L. Gathering (2.43), (2.45) and (2.46) yields the global expression of implicit template deformation:

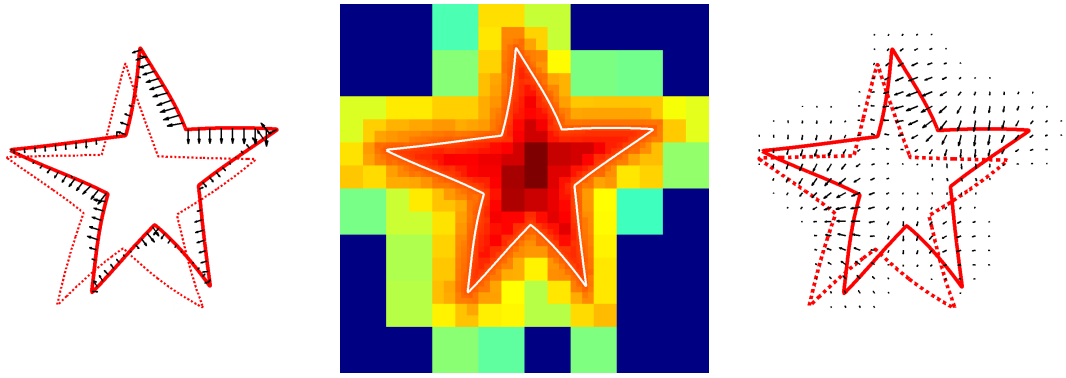
$$\begin{array}{ll} \min\limits_{\mathbf{p},\mathbf{v}} & \left\{ E(\psi_{\mathbf{p},\mathbf{v}}) = \displaystyle\int_\Omega H(\phi_0 \circ \psi_{\mathbf{p},\mathbf{v}}(\mathbf{x}))\ r(\mathbf{x})\ d\mathbf{x} + \frac{\lambda}{2}\|L_\mathbf{v} - \mathbf{Id}\|_2^2 \right\} \\ \text{with} & r(\mathbf{x}) = r_{int}(\mathbf{x}) - r_{ext}(\mathbf{x}) \\ \text{and} & \psi_{\mathbf{p},\mathbf{v}} = L_\mathbf{v} \circ G_\mathbf{p} \quad \text{such that} \quad L_\mathbf{v} = \mathbf{Id} + \mathbf{u} = \mathbf{Id} + K_\sigma * \mathbf{v} \end{array} \tag{2.49}$$

Numerical optimization

Starting from an initialization, the values of the unknowns $\mathbf{p}$ and $\mathbf{v}$ are updated iteratively through a *gradient descent* scheme with a fixed time-step Δt:

$$\left\{ \begin{array}{rcll} \mathbf{p}_i & \leftarrow & \mathbf{p}_i & - \Delta t\ \nabla_{\mathbf{p}_i} E \\ \mathbf{v} & \leftarrow & \mathbf{v} & - \Delta t\ \nabla_\mathbf{v} E \end{array} \right. \tag{2.50}$$

Remark 2.5. *Problem (2.49) is highly non-convex, so the gradient descent will end up in a local minimum that depends on the initialization (which should therefore be carefully chosen). This is often considered as a "bad" property. However in most applications it is possible to build an algorithm that provides a good enough initialization, whereas designing an energy whose global optimum correspond exactly to the organ of interest can be particularly challenging.*



(a) surface/pointwise forces (b) coarse-to-fine $\phi_0 \circ L$ (c) convolved deformation

Figure 2.16: Fast gradient computation with coarse-to-fine distance function warping and convolutions. (a) Forces are only computed on a reduced number of points. (b) Such points are efficiently found by a coarse-to-fine strategy based on octrees. (c) Forces are diffused in the whole space by fast filtering techniques. (Source: [Mory, 2011])

The expression of the gradients with respect to p_i and $\mathbf{v}$ reads

$$\begin{cases} \nabla_{\mathbf{p}_i} E &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\, r \circ G^{-1} \mathscr{A}_i \, |J_{G^{-1}}| \\ \nabla_{\mathbf{v}} E &= K_\sigma * \Big[\underbrace{\lambda \mathbf{u}}_{\text{shape}} + \underbrace{\delta(\phi_0 \circ L)\, r \circ G^{-1} \, |J_{G^{-1}}|}_{\text{image force}} \quad \nabla \phi_0 \circ L \Big] \end{cases} \tag{2.51}$$

where

$$\mathscr{A}_i(\mathbf{x}) = \left\langle \nabla \phi_0 \circ L(\mathbf{x}), \left(\mathbf{I} + \mathbf{J}_\mathbf{u} \circ G\right) \frac{\partial G}{\partial p_i} \circ G^{-1}(\mathbf{x}) \right\rangle \tag{2.52}$$

with $\mathbf{I}$ the Identity matrix and $\mathbf{J}_\mathbf{u}$ is the Jacobian matrix of $\mathbf{u}$. The details of the derivation of such equations are provided in Appendix B.

The gradient equations in (2.51) are integrals of surface terms. In most papers (*e.g.* [Chan & Vese, 2001a]), such terms are numerically computed by approximating the Dirac distribution δ by a C^∞ function δ_ε defined as

$$\forall a \in \mathbb{R},\ \delta_\varepsilon(a) = \frac{\varepsilon}{\pi\,(a^2 + \varepsilon^2)}\,. \tag{2.53}$$

Note that this approximation has a non-compact support, so an integral over the whole volume (or at least in a narrow band of the surface) must be computed. Moreover, the selection of the small parameter ε does influence the accuracy of the segmentation result. Therefore we adopt a different strategy which exploits the generalized scaling property of the Dirac distribution that is recalled below.

Proposition 2.1 (Generalized scaling property). *Let $\Omega \subset \mathbb{R}^d$, $f : \Omega \to \mathbb{R}$ a continuous function*

and $\phi : \Omega \to \mathbb{R}$ a Lipschitz-continuous function such that almost every of its level-sets is a smooth hypersurface. Then the following equality holds

$$\int_{\Omega} \delta(\phi(\mathbf{x}))\, f(\mathbf{x})\, d\mathbf{x} = \int_{\phi(\mathbf{x})=0} \frac{f(\mathbf{x})}{\left\|\nabla\phi(\mathbf{x})\right\|}\, d\mathbf{x} \quad . \tag{2.54}$$

This property enables to transform each integral of (2.51) into a surface integral. In concrete terms, we run the marching cubes algorithm [Lorensen & Cline, 1987] on $\phi_0 \circ L$ to obtain a small set of points that are subsequently used to discretize the surface integrals. The computational price to pay is the division by the norm of $\left\|\nabla(\phi_0 \circ L)\right\|$ as required in (2.54). However, if we choose the initial template ϕ_0 as a distance function, this term is approximately equal to 1 and can thus be neglected. This process yields therefore faster and more precise results than using (2.53).

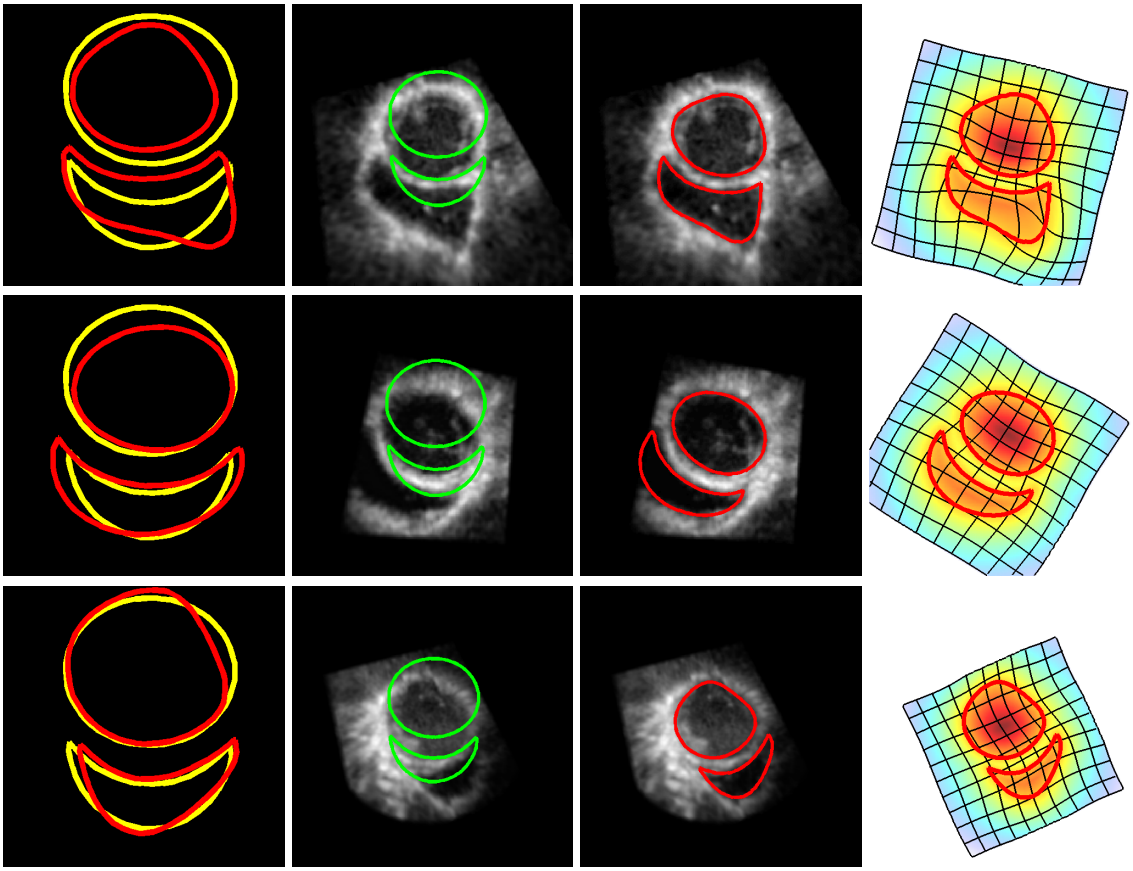
Remark 2.6. *Note that* (2.51) *does not really depend on the values of the implicit function, except its zero level-set of course and the direction of its spatial gradient, which represents the normal of the represented shape*[1]*. Furthermore, all gradients are only supported by the zero level-set of the implicit function, that is to say the surface of the object itself. Consequently, we can draw here an analogy with the notion of* shape gradient *that has been developed in a number of works (e.g. [Aubert et al., 2003]), and the use of an implicit representation is just a convenient implementation of a more generic approach.*

The gradient equations in (2.51) can actually be computed very efficiently. Interpolating $\phi_0 \circ L$ and $\nabla\phi_0 \circ L$ over the whole domain Ω_0 would be extremely time-consuming. However, since such terms are multiplied either by $\delta(\phi_0 \circ L)$ or $\delta_{G(\mathbf{x}_k)}$, the warped gradient field $\nabla\phi_0 \circ L$ is only needed on the set $\left\{\phi_0 \circ L = 0\right\}$ and at a limited number of points $\{\mathbf{x}_k\}$ (see Figure 2.16.a).

Moreover, the warped template $\phi_0 \circ L$ is only necessary near its zero level set. When ϕ_0 is a distance function, a coarse-to-fine approach using octrees can be implemented. At each level a decision is made to further refine the cell depending on the distance measure (Figure 2.16.b), drastically dropping complexity. Finally, because of the definition of L in (2.48), the difussion of image and point-wise forces to the whole space boils down to a convolution with a Gaussian kernel (Figure 2.16.c). In practice, an optimized 3D implementation supports up to 100 time steps per second when discretizing Ω_0 with a grid with 48^3 points. As the displacement field is supposedly smooth, it can indeed be stored at a very low resolution.

Remark 2.7. *The choice of a gradient descent with a fixed time-step as an optimization process may seem naive, and line search approaches [Nocedal & Wright, 1999] for example could decrease the number of iterations needed for convergence. However, line search approaches require multiple computations of the energy. This would involve the computation of volume integrals while gradients are only surface integrals. Therefore, in practice, it is more efficient to*

[1] This normal is originally only defined at the zero level-set of the implicit function, but we can extend this interpretation to the whole space.



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

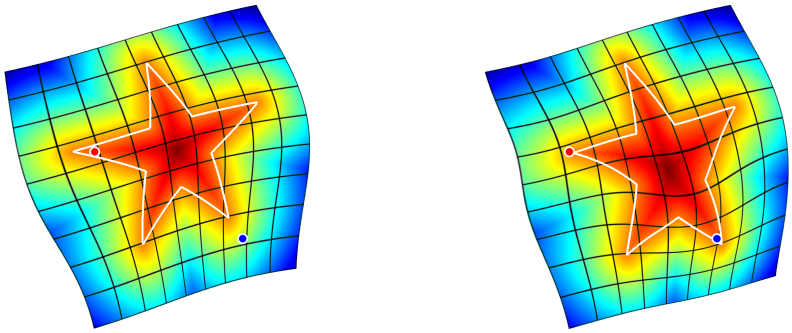
Figure 2.17: Segmentation of cardiac ultrasound images in short axis view by maximization of the gradient flux, *i.e.* $r = -\Delta I$. The same prior template, shown in yellow in (a), is used for the three cases. (a) template referential, in yellow ϕ_0, in red $\phi_0 \circ L$. (b) initialization (green). (c) segmentation result (red). (d) visualization of the warped template in the image referential - the deformed grid represents the transform $\psi = L \circ G$. (Source: [Mory, 2011])

perform more iterations of a gradient descent with a sub-optimal step than to spend a lot of time finding the optimal one.

Figure 2.17 presents some segmentation results in 2D short-axis ultrasound images of the heart. The same template ϕ_0 (representing both the left and right ventricles) is deformed in three different cases to provide a satisfying segmentation, even if the boundaries are sometimes ambiguous. The algorithm runs in a few seconds, depending on the initial model placement (left here to the user) and the chosen intensity models.

2.7.4 Including user interactions

Additional control and robustness can be obtained by offering to the user the possibility to indicate whether some specific points shall lie inside or outside the surface to extract.



(a) $\phi = \phi_0 \circ \psi$ violating constraints

(b) $\phi = \phi_0 \circ \psi$ satisfying constraints

Figure 2.18: User constraints as inside/outside labeled points. Blue (resp. red) dots indicate points that should lie inside (resp. outside) the final segmentation. (Source: [Mory, 2011])

Denoting $\{\mathbf{x}_k\}_{k\in[\![1,K]\!]} \subset \Omega^K$ these labeled points, user input can be translated into N constraints on the sign of the transformed template $\phi = \phi_0 \circ \psi$, at $\{\mathbf{x}_k\}$:

$$\forall k \in [\![1,K]\!], \quad \gamma_k \phi_0 \circ \psi(\mathbf{x}_k) \geq 0 \tag{2.55}$$

where $\gamma_k = 1$ (resp. -1) for inside (resp. outside) points. We point out that forcing the segmentation $\phi^{-1}(0)$ to pass through a specific point $\mathbf{x}$ can be done by adding at this point both inside/outside constraints. The implicit function shall be both positive and negative (and therefore null) at this point. Figure 2.18 illustrates the constraints induced by an inside point (in blue) and an outside point (in red) on the deformation of the star-shaped object of Figure 2.13.

Gathering (2.49) and (2.55) yields the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{p},\mathbf{v}} \quad & \left\{ E(\psi_{\mathbf{p},\mathbf{v}}) = \int_\Omega H(\phi_0 \circ \psi_{\mathbf{p},\mathbf{v}}(\mathbf{x}))\, r(\mathbf{x})\, d\mathbf{x} + \frac{\lambda}{2} \|L_\mathbf{v} - \mathbf{Id}\|_2^2 \right\} \\ \text{subject to} \quad & \gamma_k \phi_0 \circ \psi(\mathbf{x}_k) \geq 0, \quad \forall k \in \{1,\dots,K\} \\ \text{with} \quad & \psi_{\mathbf{p},\mathbf{v}} = L_\mathbf{v} \circ G_\mathbf{p} \quad \text{and} \quad L_\mathbf{v} = \mathbf{Id} + \mathbf{u} = \mathbf{Id} + K_\sigma * \mathbf{v} \end{aligned} \tag{2.56}$$

Since $E(\psi_{\mathbf{p},\mathbf{v}})$ is a non-convex functional to be minimized under a set of non-linear constraints, no specifically tailored algorithms are available. Yet we require an efficient optimization scheme, in order to preserve the ability of real-time feedback. For this matter, we follow a general Augmented Lagrangian methodology [Nocedal & Wright, 1999] consisting in defining an equivalent unconstrained problem that can be locally minimized by gradient descent. Note that although the sought transformation here is $\psi = \psi_{\mathbf{p},\mathbf{v}}$ as recalled in (2.56), the following strategy generalizes to any transformation model.

The constrained problem (2.56) can equivalently be written as an unconstrained mini-

mization problem of the form

$$\min_{\psi_{\mathbf{p},\mathbf{v}}} \left\{ \tilde{E}(\psi_{\mathbf{p},\mathbf{v}}) = \max_{\boldsymbol{\alpha}\geq 0} \left\{ E(\psi_{\mathbf{p},\mathbf{v}}) - \sum_{k=1}^{K} \alpha_k c_k(\psi_{\mathbf{p},\mathbf{v}}) \right\} \right\} \text{ with } c_k(\psi_{\mathbf{p},\mathbf{v}}) = \gamma_k \phi_0 \circ \psi_{\mathbf{p},\mathbf{v}}(\mathbf{x}_k) \tag{2.57}$$

where α_k is the Lagrange multiplier associated to the k^{th} constraint. (2.57) has the same set of solutions as the original problem (2.56): if $\psi_{\mathbf{p},\mathbf{v}}$ satisfies all the constraints c_k (*i.e.* $\psi_{\mathbf{p},\mathbf{v}}$ is admissible), then $\tilde{E}(\psi_{\mathbf{p},\mathbf{v}}) = E(\psi_{\mathbf{p},\mathbf{v}})$, otherwise $\tilde{E}(\psi_{\mathbf{p},\mathbf{v}})$ is infinite. Since $\tilde{E}$ jumps from finite to infinite values at the boundary of the feasible set, a more practical minimization requires to introduce a smooth approximation $\hat{E}$. Within an iterative process, in order to constrain the maximizers $\boldsymbol{\alpha} = \{\alpha_k\}_{k=1,\cdots,K}$ to finite values, one can explicitly introduce at each iteration a quadratic penalty parameter μ and a set of Lagrange multipliers $\boldsymbol{\alpha}^j$ (at the j^{th} iteration) to define

$$\hat{E}_\mu(\psi_{\mathbf{p},\mathbf{v}}, \boldsymbol{\alpha}^j) = \max_{\boldsymbol{\alpha}\geq 0} \left\{ E(\psi_{\mathbf{p},\mathbf{v}}) - \sum_{k=1}^{K} \alpha_k c_k(\psi_{\mathbf{p},\mathbf{v}}) - \frac{1}{2\mu} \sum_{k=1}^{K} \left(\alpha_k - \alpha_k^j\right)^2 \right\} \tag{2.58}$$

The maximizing Lagrange multipliers associated to each constraint $c_k(\psi_{\mathbf{p},\mathbf{v}})$ can then be found as functions of previously estimated values:

$$\alpha_k^{j+1} = \begin{cases} 0 & \text{if } \alpha_k^j - \mu c_k(\psi_{\mathbf{p},\mathbf{v}}) \leq 0 \\ \alpha_k^j - \mu c_k(\psi_{\mathbf{p},\mathbf{v}}) & \text{otherwise.} \end{cases} \tag{2.59}$$

Substituting (2.59) in (2.58) yields the expression of the smooth approximation $\hat{E}$:

$$\hat{E}_\mu(\psi_{\mathbf{p},\mathbf{v}}, \alpha^j) = E(\psi_{\mathbf{p},\mathbf{v}}) + \sum_{k=1}^{K} F_\mu\left(c_k(\psi_{\mathbf{p},\mathbf{v}}), \alpha_k^j\right) \tag{2.60}$$

$$\text{with} \quad F_\mu(a,b) = \begin{cases} -ab + \dfrac{\mu}{2}a^2 & \text{if } \mu a \leq b \\ -\dfrac{1}{2\mu}b^2 & \text{otherwise.} \end{cases} \tag{2.61}$$

The alternate scheme described in Algorithm 1, in which the penalty parameter μ is gradually increased, provides a local minimizer of (2.56) that eventually satisfies the user constraints.

Remark 2.8. *In Algorithm 1, it is also possible to skip the loop on μ, i.e. set a fixed value. This can be interpreted as turning the hard constraints into soft ones. Although convergence is not guaranteed anymore, this does produce reasonable results in practice and allows to greatly increase the speed of the method.*

The gradient descent evolution equations are obtained by applying standard calculus of

Algorithm 1: Augmented Lagrangian Scheme For Inequality Constraints.

given starting penalty parameter μ^0, and $\alpha^0 = 0$,
repeat
  choose $\mu^t > \mu^{t-1}$,
  repeat
    (1) $\psi_{\mathbf{p},\mathbf{v}}$ being fixed, update the Lagrange multipliers $\boldsymbol{\alpha}^{j+1}$ (2.59)
    (2) $\boldsymbol{\alpha}^j$ being fixed, update $\psi_{\mathbf{p},\mathbf{v}}$ by minimizing $\hat{E}_{\mu^t}(\psi_{\mathbf{p},\mathbf{v}}, \boldsymbol{\alpha}^j)$ (2.60)
  until *convergence*;
until *a local minimum of* $E(\psi_{\mathbf{p},\mathbf{v}})$ *satisfying* $\forall k, c_k(\psi_{\mathbf{p},\mathbf{v}}) \geq 0$ *is found*;

variations. With the same notations as (2.51), the new gradient equations for $\mathbf{p}_i$ and $\mathbf{v}$ reads

$$\begin{cases} \nabla_{\mathbf{p}_i} \hat{E}_\mu &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\; r \circ G^{-1} \mathscr{A}_i \; |J_{G^{-1}}| \;+\; \sum_{k=1}^{K} b_k \mathscr{A}_i \circ G(\mathbf{x}_k) \\ \nabla_{\mathbf{v}} \hat{E}_\mu &= K_\sigma * \Bigg[\underbrace{\lambda \mathbf{u}}_{\text{shape}} + \underbrace{\Big(\delta(\phi_0 \circ L)\; r \circ G^{-1} \; |J_{G^{-1}}|}_{\text{image force}} + \underbrace{\sum_{k=1}^{K} b_k \delta_{G(\mathbf{x}_k)}\Big)}_{\text{constraints}} \nabla \phi_0 \circ L \Bigg] \end{cases} \tag{2.62}$$

where the new terms

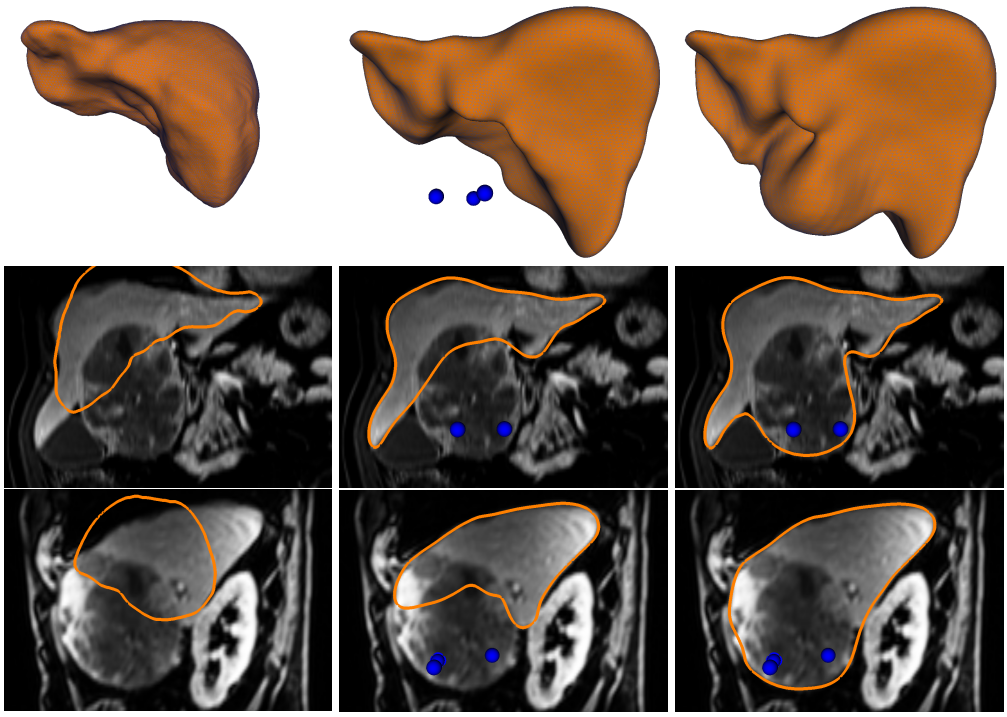
$$b_k = \gamma_k \frac{\partial F_\mu}{\partial a}\left(c_k, \alpha_k^j\right) = \begin{cases} \gamma_k \left(\mu c_k - \alpha_k^j\right) & \text{if } \mu c_k \leq \alpha_k^j \\ 0 & \text{otherwise.} \end{cases} \tag{2.63}$$

come from the Lagrangian formulation of the user constraints and $\mathscr{A}_i$ denotes the same quantity as in (2.52).

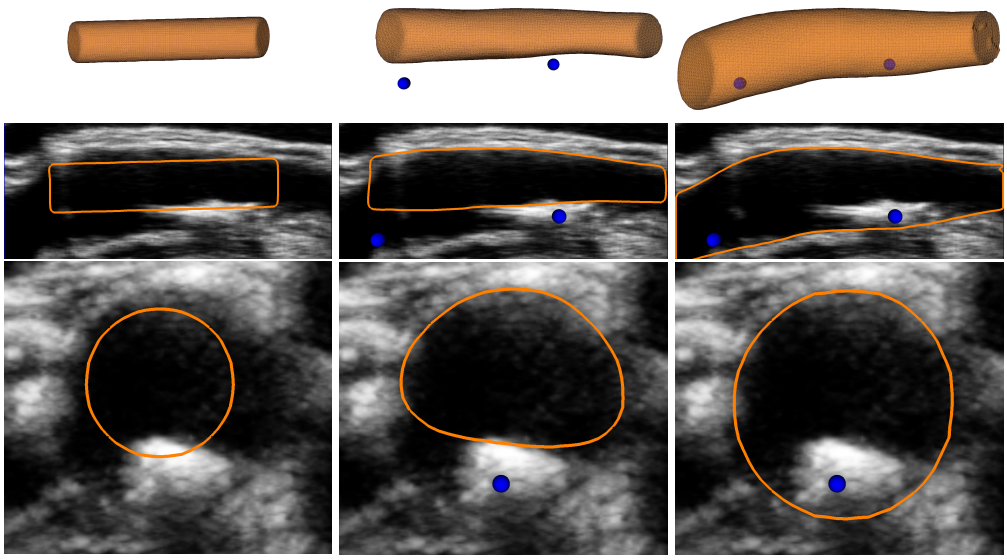
Note that only pointwise terms are added from (2.51) to (2.62). Therefore, the computational efficiency that was discussed in the previous section is maintained. This is a paramount property of the approach since it must fast enough to provide a real-time display of the segmentation evolution.

Figure 2.19 illustrates the influence of user interactions for segmentation in two clinical settings. In the first case, we show a 3D magnetic resonance image of a patient's liver with a very large and protuberant tumour. Even if the image quality is fair, the segmentation is hindered by the presence of this unpredictable lesion. With a few clicks, the user is able to either include or exclude the tumour, depending on the application. The second image is an ultrasound volume of the carotid, in which a synthetic tubular template is deformed to segment the artery wall. The obtained solution under-segments the artery because of the presence of a sclerotic plaque. Two user inputs enables to correct the segmentation so that the plaque is included.

Remark 2.9. *A quantitative evaluation of user interactions (how many clicks are needed and how well do they improve the segmentation) will be provided in Section 4.3.4 in the context of*



(a) initial template (b) unconstrained (c) with 3 constraints



(d) initial template (e) unconstrained (f) with 2 constraints

Figure 2.19: Interactive segmentation on two examples. (Top) MRI of the liver for preoperative resection planning. A 3D template of the liver is deformed to extract most of the liver tissue. Three inside user constraints are required to include a severe and unpredictable lesion. (Bottom) 3D Ultrasound image of the carotid for atherosclerosis assessment. A synthetic tubular template deforms towards the artery wall. Two user inputs allow to correct for under-segmentation due to a sclerotic plaque. (Source: [Mory, 2011])

kidney segmentation in contrast-enhanced ultrasound 3D images.

*
* *

To sum up, the *implicit template deformation* framework combines many approaches that were recalled throughout the different sections of this review. By registering a binary template to a low-level classification map of the voxels (Section 2.2), it benefits from the flexibility and efficiency of the region competition framework (Section 2.4) while imposing a shape prior as in deformable models (Section 2.5) and exploiting robustness of atlas-based methods (Section 2.6). Furthermore, as a variational approach, it does not suffer from the issues of discrete methods (Section 2.3).

In the remainder of this thesis, we will present both technical improvements (Chapter 3) and clinical applications (Chapter 4) of this segmentation framework. We will then generalize it to a joint co-segmentation/registration algorithm (Chapter 5). Finally, we will propose a method to learn shape (Chapter 6) or appearance (Chapter 7) variability from a database, and use it within this algorithm.

Chapter 3

On the template deformation functional and its minimization

Contents

Abstract

This chapter is a technical discussion on the formulation of the implicit template deformation method. We analyze it from a mathematical point of view and point out some theoretical pitfalls concerning the regularity of the solution and the convergence of the algorithm. By modifying the regularization term, we fix these mathematical difficulties while maintaining the computational efficiency of the method. Furthermore, we propose an alternative optimization scheme to ensure that the transformation applied to the model is diffeomorphic, which is a paramount property since it guarantees that the obtained segmentation shares the same topology as the template.

Résumé

Ce chapitre est une discussion technique sur la méthode de déformation de modèle implicite. Nous analysons sa formulation d'un point de vue mathématique et montrons qu'elle présente des écueils théoriques à propos de la régularité de la solution et de la convergence de l'algorithme. En définissant le champ de déformation dans un espace de Hilbert à noyau reproduisant gaussien et en modifiant le terme de régularisation, nous résolvons ces problèmes tout en préservant l'efficacité algorithmique de la méthode. Dans un second temps, nous proposons un schéma d'optimisation modifié, de façon à contraindre la déformation obtenue à être un difféomorphisme. C'est une propriété particulièrement importante car elle garantie que la topologie du modèle de forme sera préservée dans la segmentation.

This thesis is structured around the segmentation algorithm of implicit template deformation. In this short chapter, we present some contributions on the mathematical formulation of the framework and its optimization.

In Section 2.7, we recalled the principles of the implicit template deformation framework as it was presented in the paper [Mory et al., 2012]. This formulation however suffers from two theoretical flaws that we discuss and correct in Section 3.1. First, the definition of the deformation field (2.48) in Section 2.7 inducing the deformation (i) does not asymptotically provide any guarantee of smoothness and (ii) leads to an ill-posed optimization problem. Second, the optimization process via a gradient descent described in Section 2.7.3 does not necessarily produce diffeomorphic transformations. Yet in most medical applications, the object to be segmented has a fixed topology (is it a simply connected shape ? what is its Euler characteristic ?). This major information can be included in the initial template and should be taken into account during the segmentation. We propose in Section 3.2 an alternate optimization scheme that do guarantee the preservation of the template topology.

3.1 A modified formulation based on Reproducing Kernel Hilbert Spaces

We first come back to the formulation of the implicit template deformation framework, and more specifically on the definition and the regularization of the local deformation L. The formulation from Section 2.7.2:

$$\begin{array}{ll} \min\limits_{L,G} & \left\{ \int_\Omega H(\phi_0 \circ L \circ G)\; r(\mathbf{x})\; d\mathbf{x} + \dfrac{\lambda}{2}\|L - \mathbf{Id}\|_2^2 \right\} \\ \text{with} & L = \mathbf{Id} + \mathbf{u} = \mathbf{Id} + K_\sigma * \mathbf{v} \end{array} \tag{3.1}$$

leads to an ill-posed optimization problem, as detailed in the two subsections hereafter.

We fix this theoretical pitfall in Section 3.1.3 using the Reproducing Kernel Hilbert Spaces (RKHS) formalism.

3.1.1 Non-guaranteed smoothness

We first define $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ the subset of $L^2(\Omega_0, \mathbb{R}^d)$ as

$$L^2_\sigma(\Omega_0, \mathbb{R}^d) = \{\mathbf{u} \text{ s.t. } \mathbf{u} = K_\sigma * \mathbf{v},\ \mathbf{v} \in L^2(\Omega_0, \mathbb{R}^d)\}\ . \tag{3.2}$$

This is the space in which the displacement field in Eq (3.1) is sought. The convolution with the Gaussian kernel K_σ is supposed to enforce a smoothness on the displacement fields in $L^2_\sigma(\Omega_0, \mathbb{R}^d)$. However, the regularization term in Eq (3.1) is based on the L^2-norm and $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is actually not complete for such a norm, but rather dense in $L^2(\Omega_0, \mathbb{R}^d)$.

Proposition 3.1. *$L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is not complete for the L^2-norm.*

Proof. $L^2(\Omega_0, \mathbb{R}^d)$ is a Hilbert space so its complete subspaces are its closed subspaces. To prove that $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is not closed, it is thus sufficient to show that $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is dense in $L^2(\Omega_0, \mathbb{R}^d)$ (*i.e.* the orthogonal of $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ in $L^2(\Omega_0, \mathbb{R}^d)$ is $\{0\}$) and that $L^2_\sigma(\Omega_0, \mathbb{R}^d) \neq L^2(\Omega_0, \mathbb{R}^d)$. The latter condition is easily proven by considering the function K_σ which is in $L^2(\Omega_0, \mathbb{R}^d)$ but not in $L^2_\sigma(\Omega_0, \mathbb{R}^d)$. To show the first point, let us consider u an element of $L^2_\sigma(\Omega_0, \mathbb{R}^d)^\perp$. By definition, we have $\forall\ \mathbf{v} \in L^2_\sigma(\Omega_0, \mathbb{R}^d)$, $\langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\Omega_0, \mathbb{R}^d)} = 0$. As $\mathbf{v}$ belongs to $L^2_\sigma(\Omega_0, \mathbb{R}^d)$, there exists $\mathbf{w}$ in $L^2(\Omega_0, \mathbb{R}^d)$ such that $\mathbf{v} = K_\sigma * \mathbf{w}$, which yields $\left\langle \mathbf{u}, K_\sigma * \mathbf{w} \right\rangle_{L^2(\Omega_0, \mathbb{R}^d)} = 0$. Thanks to the symmetry of K_σ, we have $\left\langle K_\sigma * \mathbf{u}, \mathbf{w} \right\rangle_{L^2(\Omega_0, \mathbb{R}^d)} = 0$. This equality holds for every $\mathbf{w}$ in the dual of $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ and therefore in particular for every $\mathbf{w}$ in $L^2(\Omega_0, \mathbb{R}^d)$. Finally, $K_\sigma * \mathbf{u} = 0$, which implies $\mathbf{u} = 0$ because of the injectivity of K_σ. □

This means that a sequence of functions in $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ can converge towards an outside element (in $L^2(\Omega_0, \mathbb{R}^d) \backslash L^2_\sigma(\Omega_0, \mathbb{R}^d)$): the gradient descent on $\mathbf{u}$ may ultimately result in highly irregular displacement fields.

3.1.2 An ill-posed problem

The situation is actually even worse : finding a true minimum for the energy in Eq (3.1) is not possible. Recall that, as $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is not closed, $\mathbf{v}^* \in L^2_\sigma(\Omega_0, \mathbb{R}^d)$ is an extremum of E if and only if $\nabla_{\mathbf{v}} E(\mathbf{v}^*) = 0$.

In our case, the gradient of the implicit template deformation energy with respect to $\mathbf{v}$ is expressed as follows

$$\nabla_{\mathbf{v}} E = |J_{G^{-1}}|\ K_\sigma * \left\{ \delta(\phi_0 \circ L)\ .\ \nabla \phi_0 \circ L\ .\ r \circ G^{-1} \right\} + \lambda K_\sigma * (K_\sigma * \mathbf{v})\ . \tag{3.3}$$

With the injectivity of K_σ, the condition of optimality is equivalent to

$$|J_{G^{-1}}|\ .\ \delta(\phi_0 \circ L)\ .\ \nabla \phi_0 \circ L\ .\ r \circ G^{-1} = -\lambda K_\sigma * \mathbf{v} \tag{3.4}$$

We notice that the right-hand side term belongs necessarily to $L^2_\sigma(\Omega_0, \mathbb{R}^d)$ whereas the left-hand side does not. Indeed, as a sum of Dirac masses, it is a distribution whose support is a surface. Therefore, the equation is usually not feasible (unless the trivial case of $\mathbf{v} = 0$ and $r = 0$ at the zer level-set of ϕ_0) and the equilibrium cannot be reached.

Why it seems to work in practice

With all these limitations, one may wonder why the algorithm did work in practice. We expose here some possible answers, and some intuitions on what the algorithm actually does.

- **Change of metric for the gradient descent:** As we pointed out previously, the solution $\mathbf{v}^*$ is not necessarily smooth. However, defining the displacement field as a filtered version of another one reduces to performing a gradient descent with another metric (recall that a gradient is defined up to the choice of metric). Indeed, for any $\mathbf{u}$ in $L^2(\Omega_0, \mathbb{R}^d)$ we have

$$dE(\mathbf{u}) = \left\langle |J_{G^{-1}}| \, \delta(\phi_0 \circ (\mathbf{Id} + \mathbf{u})) \, . \, \nabla \phi_0 \circ (\mathbf{Id} + \mathbf{u}) \, . \, r \circ G^{-1} + \lambda \, \mathbf{u} \, , \, \delta \mathbf{u} \right\rangle_{L^2(\Omega_0, \mathbb{R}^d)} \quad (3.5)$$

  Using a new scalar product $< \mathbf{u} \, , \, \mathbf{v} >_\sigma = < K_\sigma^{-1} * \mathbf{u} \, , \, \mathbf{v} >_{L^2(\Omega_0, \mathbb{R}^d)}$ yields

$$dE(\mathbf{u}) = \left\langle |J_{G^{-1}}| K_\sigma * \left\{ \delta(\phi_0 \circ (\mathbf{Id} + \mathbf{u})) \, . \, \nabla \phi_0 \circ (\mathbf{Id} + \mathbf{u}) \, . \, r \circ G^{-1} \right\} + \lambda K_\sigma * \mathbf{u} \, , \, \delta \mathbf{u} \right\rangle_\sigma \quad (3.6)$$

  Therefore the gradient of E with respect to the scalar product $< . \, , \, . >_\sigma$ has exactly the same expression as in Equation (3.3), *i.e.* the one that is actually used in [Mory, 2011]. This demonstrates that defining $\mathbf{u} = K_\sigma * \mathbf{v}$ does not change the energy (and in particular its minimizers). What is impacted is just the path of the gradient descent: The high-frequencies are removed so that the algorithm is more robust and the number of iterations needed to observe a non-smooth vector field is very large. Note also that the algorithm may also converge towards different local minima, and in particular smoother ones if there is any. Similar ideas were exploited in [Sundaramoorthi et al., 2007] and [Charpiat et al., 2007]. In both papers, the authors defined a shape gradient with respect to the Sobolev H^1-metric instead of the usual L^2-norm, which also allowed to favour smooth/coherent displacements.

- **Numerical discretization :** The second point is the non reachability of the local minima. Actually, because of the numerical discretization, the kernel K_σ that is applied is not invertible. Therefore, (3.4) does not hold and we only have

$$|J_{G^{-1}}| \, . \, K_\sigma * \left\{ \delta(\phi_0 \circ L) \, . \, \nabla \phi_0 \circ L \, . \, r \circ G^{-1} \right\} = -\lambda \, K_\sigma * \mathbf{u} \, . \quad (3.7)$$

  Both terms belong to $L^2(\Omega_0, \mathbb{R}^d)$ and there is therefore no reason for this equation to be impossible: an equilibrium may be indeed reachable.

Another related consequence of the numerical discretization is that $\mathbf{v}$ has a guaranteed minimum smoothness, depending on the sampling frequency and the interpolation strategy.

3.1.3 A new formulation with RKHS

The theoretical issues we encounter with the current formulation come from the choice of the regularization norm. We therefore propose another formulation, very close to the existing one, to cope with those problems. As in the standard formulation, we will choose a norm that penalizes large values of the gradient. To do so, we present a convenient framework based on Reproducing Kernel Hilbert Spaces (RKHS) that generalizes that idea and makes the formulation well-posed from the theoretical point of view.

Reproducing Kernel Hilbert Spaces

Let U be a Hilbert space of $\mathbb{R}^d$-valued functions on Ω_0. It is a reproducing kernel Hilbert space if, for any $\mathbf{x} \in \Omega_0$, the linear mapping

$$\delta_{\mathbf{x}} : \mathbf{u} \in U \rightarrow \mathbf{u}(\mathbf{x}) \in \mathbb{R}^d$$

is continuous. In such a case, for any $\mathbf{a} \in \mathbb{R}^d$, the function $\delta_{\mathbf{x}}^{\mathbf{a}} : \mathbf{u} \mapsto \mathbf{a}^T\mathbf{u}(\mathbf{x})$ is a continuous linear form on U. By the Riesz representation theorem, there exists a unique $k_{\mathbf{x}}^{\mathbf{a}} \in U$ such that

$$\forall \mathbf{u} \in U,\ \delta_{\mathbf{x}}^{\mathbf{a}}(\mathbf{u}) \ = \ \mathbf{a}^T\mathbf{u}(\mathbf{x}) \ = \ \left\langle k_{\mathbf{x}}^{\mathbf{a}}, \mathbf{u} \right\rangle_U \ . \tag{3.8}$$

Note that the mapping $\mathbf{a} \mapsto k_{\mathbf{x}}^{\mathbf{a}}$ is linear from $\mathbb{R}^d$ to U and consequently, for any $\mathbf{y} \in \mathbb{R}^d$, the mapping $\mathbf{a} \mapsto k_{\mathbf{x}}^{\mathbf{a}}(\mathbf{y})$ is linear from $\mathbb{R}^d$ to $\mathbb{R}^d$. We can thus define a function $K : \Omega_0 \times \Omega_0 \rightarrow \mathbb{R}^d$ such that $k_{\mathbf{x}}^{\mathbf{a}}(\mathbf{y}) = K(\mathbf{x},\mathbf{y})\mathbf{a}$ or equivalently $k_{\mathbf{x}}^{\mathbf{a}} = K(\mathbf{x},\cdot)\mathbf{a}$. This function is called the *reproducing kernel* of U due to the following *reproducing property*:

$$\forall (\mathbf{a},\mathbf{b}) \in \mathbb{R}^d \times \mathbb{R}^d,\ \langle K(\mathbf{x},\cdot)\mathbf{a}, K(\mathbf{y},\cdot)\mathbf{b}\rangle_U = \left\langle k_{\mathbf{x}}^{\mathbf{a}}, K(\mathbf{y},\cdot)\right\rangle = \mathbf{a}^T K(\mathbf{x},\mathbf{y})\mathbf{b} \tag{3.9}$$

It can be easily shown that such a kernel is symmetric and positive definite (under some assumptions). We refer the reader to [Aronszajn, 1950] for a thorough description of RKHS.

A modified formulation of implicit template deformation

We still express the local deformation as $L = \mathbf{Id} + \mathbf{u}$, but now $\mathbf{u}$ lies in the RKHS U defined[1] by the Gaussian kernel K_σ on $L^2(\Omega_0, \mathbb{R}^d)$. The operator $K : \mathbf{u} \mapsto K_\sigma * \mathbf{u}$ and its inverse K^{-1} allow to go back and forth between U to its dual space U^*. For each $\mathbf{u} \in U$, there exists a unique $\mathbf{v} \in U^*$ such that $\mathbf{u} = K\mathbf{v}$ or $\mathbf{v} = K^{-1}\mathbf{u}$. We have $U \subset L^2(\Omega_0, \mathbb{R}^d) \subset U^*$ so the vector

[1] According to the Moore-Aronszajn theorem, any RKHS is uniquely defined by its symmetric positive definite kernel and vice versa.

field $\mathbf{v}$ can be highly irregular as it is not even necessarily in $L^2(\Omega_0, \mathbb{R}^d)$. The appropriate norm on U is $\|\mathbf{u}\|_U^2 = \langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\Omega_0,\mathbb{R}^d)} = \left\langle \mathbf{u}, K^{-1}\mathbf{u} \right\rangle_{L^2(\Omega_0,\mathbb{R}^d)} = \langle K\mathbf{v}, \mathbf{v} \rangle_{L^2(\Omega_0,\mathbb{R}^d)}$.

The new template deformation energy then reads

$$\begin{array}{ll} \min\limits_{L,G} & \left\{ E_U(L,G) = \displaystyle\int_\Omega H(\phi_0 \circ L \circ G)\, r(\mathbf{x})\, d\mathbf{x} + \frac{\lambda}{2}\|L - \mathbf{Id}\|_U^2 \right\} \\ \text{with} & L = \mathbf{Id} + \mathbf{u} \quad \text{and } \|\mathbf{u}\|_U^2 = \left\langle \mathbf{u}, K_\sigma^{-1}\mathbf{u} \right\rangle_{L^2(\Omega_0,\mathbb{R}^d)} \end{array} \tag{3.10}$$

We can now compute the small variation $dE(\mathbf{u})$ of E with respect to the variation $\delta\mathbf{u}$:

$$dE_U(\mathbf{u}) = \left\langle |J_{G^{-1}}|\ \delta(\phi_0 \circ L)\ .\ \nabla\phi_0 \circ L\ .\ r \circ G^{-1}\ ,\ \delta\mathbf{u} \right\rangle_{L^2(\Omega_0,\mathbb{R}^d)} + \lambda \left\langle K^{-1}\mathbf{u}, \delta\mathbf{u} \right\rangle_{L^2(\Omega_0,\mathbb{R}^d)} \tag{3.11}$$

This gradient is computed with respect to the L^2-norm but what interests us for the gradient descent is the variation in the sense of the U scalar product, which is

$$dE_U(\mathbf{u}) = \left\langle K_\sigma * |J_{G^{-1}}|\ \delta(\phi_0 \circ L)\ .\ \nabla\phi_0 \circ L\ .\ r \circ G^{-1}\ ,\ \delta\mathbf{u} \right\rangle_U + \lambda \left\langle \mathbf{u}, \delta\mathbf{u} \right\rangle_U \tag{3.12}$$

and the gradient finally reads

$$\nabla_{\mathbf{u}} E_U = K_\sigma * \left\{ |J_{G^{-1}}|\ \delta(\phi_0 \circ L)\ .\ \nabla\phi_0 \circ L\ .\ r \circ G^{-1} \right\} + \lambda\mathbf{u}\,. \tag{3.13}$$

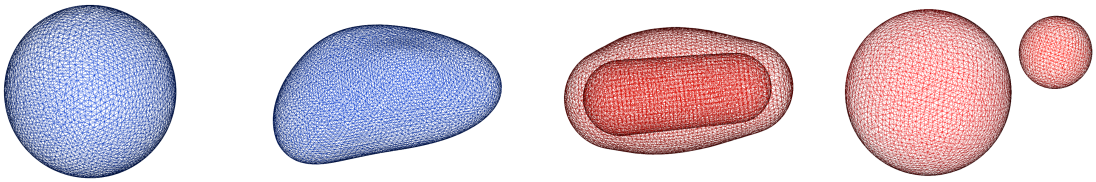
The dual deformation field $\mathbf{v} \in U^*$ is therefore only a hidden variable that we do not need to store or even compute during the gradient descent.

Remark 3.1. *Even if we do not need to compute the dual variable $\mathbf{v} \in U^*$, it is interesting to note that the optimality condition on $\mathbf{v}$ is*

$$|J_{G^{-1}}|\ .\ \delta(\Phi_0 \circ L)\ .\ \nabla\phi_0 \circ L\ .\ r \circ G^{-1} = -\lambda\ \mathbf{v}\,. \tag{3.14}$$

In particular, it means that unlike the previous formulation, an equilibrium is reachable if the left-hand side of Eq. (3.14) *belongs to the dual space U^*. To this end, it is sufficient to impose higher-order regularity assumptions on the RKHS U, which are usually satisfied for standard kernels (such as the Gaussian kernel).*

This formulation based on RKHS is very close to the previous one. Actually, the only difference is that one Gaussian filtering is removed in the computation of the gradient $\nabla_{\mathbf{u}} E_U$. The novel formulation is therefore not only theoretically sounder, but also more efficient computationally. From now on, we will always use this new model and, for presentation purposes, every mention to energy E should be understood as energy E_U.



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

Figure 3.1: Topology preservation. The sphere (a) can be mapped with a homeomorphism to the shape (b) but not to the hollow shape (c) or two spheres (d).

3.2 Ensuring topology preservation with diffeomorphisms

3.2.1 Diffeomorphic optmization process

In most medical applications, the organ of interest has (although possibly a complex shape) a fixed topology. This should be taken into account when deforming a shape. In order to preserve the topology, a space deformation $L : \Omega_0 \to \Omega_0$ requires two important properties:

- the deformation must not generate folds; distinct points should be mapped to distinct points. In other words, L should be injective.
- no hole must be created; every point of the domain should have a pre-image. In other words, L should be onto.

Homeomorphisms are therefore suitable candidates as deformations (see Figure 3.1).

Definition 3.1. *A **homeomorphism** of Ω_0 is a continuous bijection $\psi : \Omega_0 \to \Omega_0$ such that its inverse ψ^{-1} is continuous.*

However, one often wants to include additional regularity constraints on the considered deformations. Most people thus rather use the concept of diffeomorphisms instead of homeomorphisms.

Definition 3.2. *A **diffeomorphism** of Ω_0 is a continuously differentiable homeomorphism $\psi : \Omega_0 \to \Omega_0$ such that its inverse ψ^{-1} is continuously differentiable.*

It is easy to see that diffeomorphisms are stable by function composition. Actually, they form a group that we denote $(\mathrm{Diff}(\Omega_0), \circ)$. When manipulating and combining diffeomorphisms, it is therefore natural to use the composition operator. Conversely, adding diffeomorphisms has no geometrical meaning and the sum of two diffeomorphisms is not necessarily one. In light of this remark, one may question the optimization process of implicit template deformation by gradient descent, that was proposed in [Mory et al., 2012], *i.e.*

$$\begin{cases} L_{(0)} & = \mathbf{Id} \\ L_{(t+1)} & = L_{(t)} - \Delta t \ \nabla_{\mathbf{u}} E \end{cases} \tag{3.15}$$

or equivalently

$$\begin{cases} \mathbf{u}_{(0)} & = \mathbf{0} \\ \mathbf{u}_{(t+1)} & = \mathbf{u}_{(t)} - \Delta t \ \nabla_{\mathbf{u}} E \end{cases} \tag{3.16}$$

Indeed this process represents an additive construction of the deformation L and is commonly referred in the registration community as the *small deformation* model [Ashburner et al., 2007]. This name comes from the fact that L preserves the topology only if the magnitude of the added displacement field is low. In our application, the template may undergo relatively large deformations; there is therefore no guarantee for L to be diffeomorphic. To avoid this issue, we propose to use a different update, that is rather based on composition :

$$\begin{cases} L_{(0)} & = \mathbf{Id} \\ L_{(t+1)} & = L_{(t)} \circ (\mathbf{Id} - \Delta t \ \nabla_{\mathbf{u}} E) \end{cases} \tag{3.17}$$

which means for the displacement field $\mathbf{u}$:

$$\begin{cases} \mathbf{u}_{(0)} & = \mathbf{0} \\ \mathbf{u}_{(t+1)} & = \mathbf{u}_{(t)} \circ (\mathbf{Id} - \Delta t \ \nabla_{\mathbf{u}} E) - \Delta t \ \nabla_{\mathbf{u}} E \end{cases} \tag{3.18}$$

This procedure can be interpreted as immediately propagating the template by the gradient vector field and performing the next iteration with $\phi_0 \leftarrow \phi_0 \circ (\mathbf{Id} - \Delta t \ \nabla_{\mathbf{u}} E)$ and $L \leftarrow \mathbf{Id}$. Naturally, the regularization term on L intrinsically keeps track of the total deformation underwent by the template, as it constrains it towards the identity. More details and links to other approaches are provided in the next subsection.

Such a process has the great advantage to produce diffeomorphic transformations, as Proposition 3.2 will state. To prove it, we first need the following lemma.

Lemma 3.1. *Let $\mathbf{u} \in C_0^1(\Omega_0, \Omega_0)$ such that $\mathbf{u}$ and $D\mathbf{u}$ vanishes at infinity. There exists $\varepsilon_0 > 0$ such that $\forall \varepsilon < \varepsilon_0$, $(\mathbf{Id} + \varepsilon \ \mathbf{u})$ is a diffeomorphism.*

Proof. In this proof, that is inspired by [Younes, 2010], we denote $\psi_\varepsilon : \Omega_0 \to \Omega_0$ the transformation $(\mathbf{Id} + \varepsilon \ \mathbf{u})$. Given the condition on $\mathbf{u}$, ψ_ε is naturally continuous and continuously differentiable.
Let us first show that ψ_ε is one-to-one by considering $(\mathbf{x}, \mathbf{x}') \in \Omega_0^2$ such that $\psi_\varepsilon(\mathbf{x}) = \psi_\varepsilon(\mathbf{x}')$. Then

$$|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}')| = \left| \frac{(\psi_\varepsilon(\mathbf{x}) - \mathbf{x})}{\varepsilon} - \frac{(\psi_\varepsilon(\mathbf{x}') - \mathbf{x}')}{\varepsilon} \right| = \frac{1}{\varepsilon} |\mathbf{x} - \mathbf{x}'| \ . \tag{3.19}$$

Since $D\mathbf{u}$ is continuous and vanishes at infinity, it is bounded and there exists a constant $C > 0$ such that $|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}')| \leq C|\mathbf{x} - \mathbf{x}'|$. This yields

$$|\mathbf{x} - \mathbf{x}'| \leq \varepsilon C |\mathbf{x} - \mathbf{x}'| \tag{3.20}$$

From now on, we assume that $\varepsilon_0 \leq \dfrac{1}{C}$, which implies $\varepsilon C < 1$ and then $\mathbf{x} = \mathbf{x}'$. ψ_ε is therefore an injective transformation.

We then prove that ψ_ε is surjective. For all $\mathbf{y} \in \Omega_0$, we denote $f_\mathbf{y} : \Omega_0 \to \Omega_0$ the function defined by $f_\mathbf{y}(\mathbf{z}) = -\varepsilon\mathbf{u}(\mathbf{y}+\mathbf{z})$. Then we have for all $(\mathbf{z}, \mathbf{z}') \in \Omega_0^2$

$$|f_\mathbf{y}(\mathbf{z}) - f_\mathbf{y}(\mathbf{z}')| = \varepsilon|\mathbf{u}(\mathbf{y}+\mathbf{z}) - \mathbf{u}(\mathbf{y}+\mathbf{z}')| \leq \varepsilon C|\mathbf{z}-\mathbf{z}'| \tag{3.21}$$

This implies that $f_\mathbf{y}$ is contractant as $\varepsilon C < 1$ and the fixed-point theorem states that there exists $\mathbf{z}_\mathbf{y}$ such that $f_\mathbf{y}(\mathbf{z}_\mathbf{y}) = -\varepsilon\mathbf{u}(\mathbf{y}+\mathbf{z}_\mathbf{y}) = \mathbf{z}_\mathbf{y}$. Thus

$$\psi_\varepsilon(\mathbf{y}+\mathbf{z}_y) = \mathbf{y} + \mathbf{z}_\mathbf{y} + \varepsilon\mathbf{u}(\mathbf{y}+\mathbf{z}_y) = \mathbf{y}\,. \tag{3.22}$$

Every $\mathbf{y}$ is the image by ψ_ε of a point in Ω_0, therefore ψ_ε is onto.
The next step is to prove that ψ_ε^{-1} is continuous: consider $(\mathbf{y}, \mathbf{y}') \in \Omega_0^2$ and their associated fixed points $(\mathbf{z}_\mathbf{y}, \mathbf{z}_{\mathbf{y}'}) \in \Omega_0^2$. We have

$$\begin{aligned}
|\mathbf{z}_\mathbf{y} - \mathbf{z}_{\mathbf{y}'}| &= \varepsilon|\mathbf{u}(\mathbf{y}+\mathbf{z}_\mathbf{y}) - \mathbf{u}(\mathbf{y}'+\mathbf{z}_{\mathbf{y}'})| \\
&\leq \varepsilon C|(\mathbf{y}+\mathbf{z}_\mathbf{y}) - (\mathbf{y}'+\mathbf{z}_{\mathbf{y}'})| \\
&\leq \varepsilon C\left(|\mathbf{y}-\mathbf{y}'| + |\mathbf{z}_\mathbf{y} - \mathbf{z}_{\mathbf{y}'}|\right) \\
&\leq \frac{\varepsilon C}{1-\varepsilon C}|\mathbf{y}-\mathbf{y}'|\,.
\end{aligned} \tag{3.23}$$

Besides, Eq. (3.22) implies that

$$\begin{aligned}
|\psi_\varepsilon^{-1}(\mathbf{y}) - \psi_\varepsilon^{-1}(\mathbf{y}')| &= |(\mathbf{y}+\mathbf{z}_\mathbf{y}) - (\mathbf{y}'+\mathbf{z}_{\mathbf{y}'})| \\
&\leq |\mathbf{y}-\mathbf{y}'| + |\mathbf{z}_\mathbf{y} - \mathbf{z}_{\mathbf{y}'}|\,,
\end{aligned} \tag{3.24}$$

which yields, combined with (3.23),

$$|\psi_\varepsilon^{-1}(\mathbf{y}) - \psi_\varepsilon^{-1}(\mathbf{y}')| \leq \frac{1}{1-\varepsilon C}|\mathbf{y}-\mathbf{y}'|\,, \tag{3.25}$$

hence the continuity of ψ_ε^{-1} and it only remains to prove its differentiability (and the continuity of its derivative). For this, it is sufficient to show that $D\psi_\varepsilon$ is nowhere zero. Suppose there exists a $\mathbf{x}_0$ where $D\psi_\varepsilon$ vanishes:

$$D\psi_\varepsilon(\mathbf{x}_0) = 0 \iff \mathbf{1} + \varepsilon D\mathbf{u}(\mathbf{x}_0) = 0 \iff D\mathbf{u}(\mathbf{x}_0) = -\frac{1}{\varepsilon}\mathbf{1} \tag{3.26}$$

Yet $D\mathbf{u}$ is bounded by the constant C such that $\varepsilon C < 1$ so (3.26) does not admit any solution. This concludes the proof of the diffeomorphic property of ψ_ε. □

Proposition 3.2. *After any fixed number of iterations, the optimization process described in* (3.17) *produces a diffeomorphic transformation for a sufficiently small time step* Δt.

Proof. We prove this proposition by mathematical induction. $L_{(0)}$ is the Identity transformation and therefore diffeomorphic. We now assume that $L_{(t)}$ is a diffeomorphism. For a sufficiently small time step Δt, $(\mathbf{Id} - \Delta t\ \nabla_\mathbf{u} E)$ is also a diffeomorphism (according to

Lemma 3.1). Diffeomorphisms are stable by composition so $L_{(t+1)}$ is also diffeomorphic. □

The proposed optimization process will therefore preserve the topology of the template with little computational overhead. Comparing (3.15) and (3.17) shows that only one additional interpolation is needed for the update of the deformation. However, it is not a gradient descent anymore. This means that the evolution direction will not maximize the energy decrease and more iterations may be needed. Actually, the energy E is not even guaranteed to decrease but based on our experiments, it always does in practice.

Remark 3.2. *In the remainder of this thesis, every time that we use a gradient descent on a variable that is supposed to be a diffeomorphism, this must be understood as the composition sense* (3.17) *rather than the usual one* (3.15).

Another benefit of this modified process is the opportunity to build recursively an estimate $\tilde{L}$ of the inverse transformation, which may be needed in some advanced computations:

$$\begin{cases} \tilde{L}_{(0)} & = \mathbf{Id} \\ \tilde{L}_{(t+1)} & = (\mathbf{Id} + \Delta t \ \nabla_{\mathbf{u}} E) \circ \tilde{L}_{(t)} \end{cases} \tag{3.27}$$

These equations are directly obtained by inverting Eq. (3.17) and assuming that Δt is small so that $(\mathbf{Id} - \Delta t \ \nabla_{\mathbf{u}} E)^{-1} \approx (\mathbf{Id} + \Delta t \ \nabla_{\mathbf{u}} E)$.

Remark 3.3. *Because of numerical stability issues, the process described in* (3.27) *may slightly drift in time. To avoid this pitfall, one can regularly correct the estimation by minimizing the energy* $\|\tilde{L} \circ L - \mathbf{Id}\|^2$ *by a gradient descent initialized with the estimate given by* (3.27). *Since this estimate is really close, very few iterations (in practice less than 5) are needed to prevent the drift.*

Figure 3.2 shows an example of segmentation with both the additive (3.15) and the compositive (3.17) update rules. The initial template is set as a circle. This experiment highlights the effect of the composition update towards the topology preservation of the model. Despite the highly ambiguous image, the segmentation stays in a single connected component. Conversely, the additive process manages to deform the circle so severely that it actually becomes two disconnected shapes.

A discussion on the benefits of diffeomorphisms composition over addition is also available in [Vercauteren et al., 2008] for registration applications.

3.2.2 Links with other works

To better understand the behaviour of the proposed optimization scheme, we draw in this section some links with other works. A similar update scheme was proposed in [Saddi et al., 2007] in the context of template deformation. However, their minimized energy had no regularization term: the magnitude of the deformation underwent by the template was not penalized at all and the required smoothness was enforced by a filtering of the gradient vector field that had no energetic meaning. As pointed out in [Christensen et al., 1996;

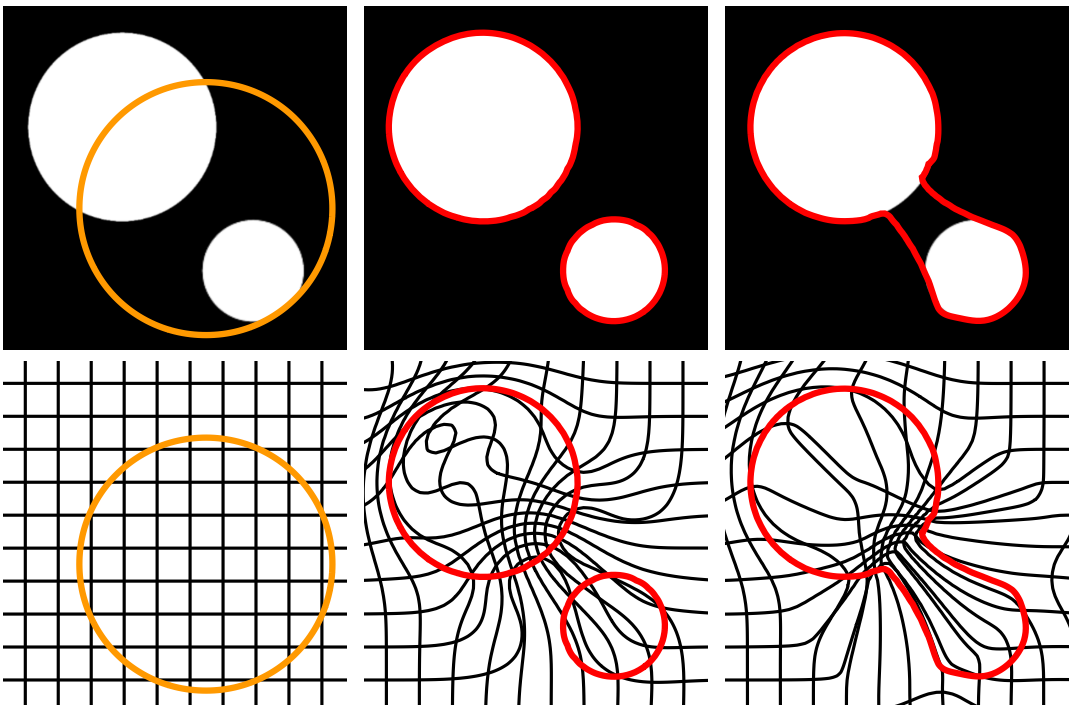


Figure 3.2: Comparison of the two optimization processes. (a) Initialization of the segmentation with a circle template. (b) Segmentation with the additive update: the deformation is not diffeomorphic and topology is lost. (c) Segmentation with the composition update: the deformation stays diffeomorphic and topology is preserved.

Chefd'hotel et al., 2002], the compositive scheme can be related to the following transport equation :

$$\frac{\partial L_{(t)}}{\partial t} = -DL_{(t)}\ \nabla_{\mathbf{u}} E \tag{3.28}$$

Indeed, if we discretize Eq. (3.28) in time with an upwind scheme, we obtain

$$\frac{L_{(t+1)} - L_{(t)}}{\Delta t} = -DL_{(t)}\ \nabla_{\mathbf{u}} E \tag{3.29}$$

$$L_{(t+1)} = L_{(t)} - \Delta t\ \nabla_{\mathbf{u}} E \tag{3.30}$$

This expression can be approximated, using a first-order Taylor expansion, by:

$$L_{(t+1)} = L_{(t)} \circ (\mathbf{Id} - \Delta t\ \nabla_{\mathbf{u}} E)\ , \tag{3.31}$$

which is the actual composition update that we proposed in (3.17). It can be shown [Chefd'hotel et al., 2002] that this defines a consistent, first-order accurate step forward operator for the approximation of the transport equation. The template thus "flows" in time with respect to the vector field given by the gradient of the energy.

We may also relate this process to the popular framework of large deformation diffeomorphic metric mapping (LDDMM) introduced by Beg *et al.* [Beg et al., 2005], but mainly used in registration applications. In order to register two images I_0 and I_1, the LDDMM framework consists (with our notations) in solving the following problem

$$\begin{cases} \min\limits_{(\mathbf{u}_{(t)})_t} \left\{ \|I_0 \circ L_{(t=1)} - I_1\|^2 + \dfrac{\lambda}{2} \displaystyle\int_0^1 \|\mathbf{u}_{(t)}\|_U^2 \, dt \right\} \\ \text{with } \dfrac{\partial L_{(t)}}{\partial t} = \mathbf{u}_{(t)} \circ L_{(t)} \text{ and the initial condition } L_{(t=0)} = \mathbf{Id} \end{cases} \tag{3.32}$$

The first term in the minimized energy is specific to registration and depends on a transformation L. To define it, a fictive time $t \in [0,1]$ in introduced and discretized. Deformation fields of each those time step $(\mathbf{u}_{(t)})_t$ are updated iteratively with various strategies [Hart et al., 2009; Vialard et al., 2012]. Thanks to the second energy term, the transformation $L_{(1)}$ obtained by integrating the deformation fields is a geodesic in the space of diffeomorphisms (under some assumptions, see [Dupuis et al., 1998] for example). Two main differences with our approach can be noticed :

- Instead of penalizing the norm of each instantaneous velocity field, we penalize the global resulting transformation. This roughly represents an interchanging of the time-integration and the norm.
- The fictive time t can be seen as our iteration counter. The difference is that at time t, we cannot change the past estimates of the deformation fields. Our optimization process is thus somehow similar to a greedy optimization algorithm in the LDDMM framework.

Therefore, compared to LDDMM methods, our method loses the geodesic property of the transformation. While this is not a problem for segmentation applications, it might have a negative effect when such transformations are used within a learning process. However, the LDDMM framework is computationally extremely demanding and still intractable for real-time applications such as interactive segmentation.

A slightly more simple approach, named the *log-Euclidean framework*, was proposed in [Arsigny et al., 2006a,b] and developed in [Ashburner et al., 2007; Vercauteren et al., 2008] for registration applications. This framework is similar to (3.32) but the vector field $\mathbf{u}_{(t)} \equiv \mathbf{u}$ is constant over unit time. The diffeomorphism L is thus defined as the *exponential* of this vector field : $L = \exp(\mathbf{u})$. Although far less computationally demanding than LDDMM, this approach suffers from an important pitfall. Indeed, expressing the whole deformation with a single static vector field is a strong constraint, and some diffeomorphisms cannot be expressed as the exponential of any vector field. Using this strategy in the template deformation framework may consequently result in inaccurate segmentations.

*
* *

This chapter was a mathematical discussion on the implicit template deformation framework and the minimization of its functional. Although we pointed out some pitfalls in the original formulation [Mory, 2011], we showed that they can be easily fixed by slightly modifying the formulation and using the Gaussian reproducing kernel Hilbert space.

We also proposed a different optimization process that enforces the transformation L to be a diffeomorphism. This is a paramount property since it guarantees that the segmentation shares the same topology as the initial template. Compared to other works, our approach achieves a satisfying trade-off between flexibility and efficiency.

Chapter 4

Applications of the template deformation algorithm to kidney segmentation

Contents

Abstract

This chapter presents several applications of the implicit template deformation algorithm for kidney segmentation. After providing a short background on this organ and its functioning, we propose various automated pipelines to segment kidneys in CT, CEUS and US 3D images. Our approaches are composed of two steps. First we detect the kidney by estimating an ellipsoid, which is done (depending on the modality) either by a regression forest or a novel variational method that we introduce in this chapter. This ellipsoid is then deformed with the implicit template deformation algorithm, using an image-based term learnt with a random forest. Furthermore, we prove the great benefits of taking into account user interactions (in CEUS images) and adopting an auto-context strategy to learn the classification forests (in US images). All applications are evaluated on significant databases.

Résumé

Ce chapitre présente plusieurs applications de l'algorithme de déformation de modèle implicite pour la segmentation de rein. Après une brève introduction sur cet organe et son fonctionnement, nous proposons diverses approches automatiques pour la segmentation de rein dans des images 3D CT, CEUS et US. Toutes ces méthodes sont composées de deux étapes. Dans un premier temps, le rein est détecté dans l'image en tant qu'ellipsoïde. Cette détection se fait, selon la modalité, soit via une forêt de régression, soit par une nouvelle méthode variationnelle. Cet ellipsoïde est ensuite déformé avec l'algorithme de déformation de modèle implicite, selon un terme image appris par une forêt aléatoire. Dans ce chapitre, nous prouvons également l'importance de prendre en compte des interactions utilisateurs (dans les images CEUS) et d'adopter une stratégie appelée "auto-context" pour l'apprentissage de la forêt de classification (dans les images US). Toutes les applications décrites sont évaluées sur de conséquentes bases de données.

This chapter is dedicated to the use of implicit template deformation in various clinical settings. We show its efficacy and efficiency in the context of kidney segmentation in 3D CT images (Section 4.2), 3D CEUS images (Section 4.3) and 3D US images (Section 4.4). The image-based term should naturally be chosen according to the organ appearance in the each modality and will then be different in the three cases. When this appearance is difficult to model explicitly, we will resort to machine learning techniques to perform voxel classification (as in Sections 4.2 and 4.4). Besides, as a non-convex variational formulation, the segmentation algorithm produces results that depend on the initialization. The initialization process must therefore be carefully designed. In the following, we will use either regression approaches (Section 4.2) or robust estimation algorithms (Section 4.3 and 4.4) to find a suitable initial template ϕ_0 to be deformed. We first provide some clinical background on

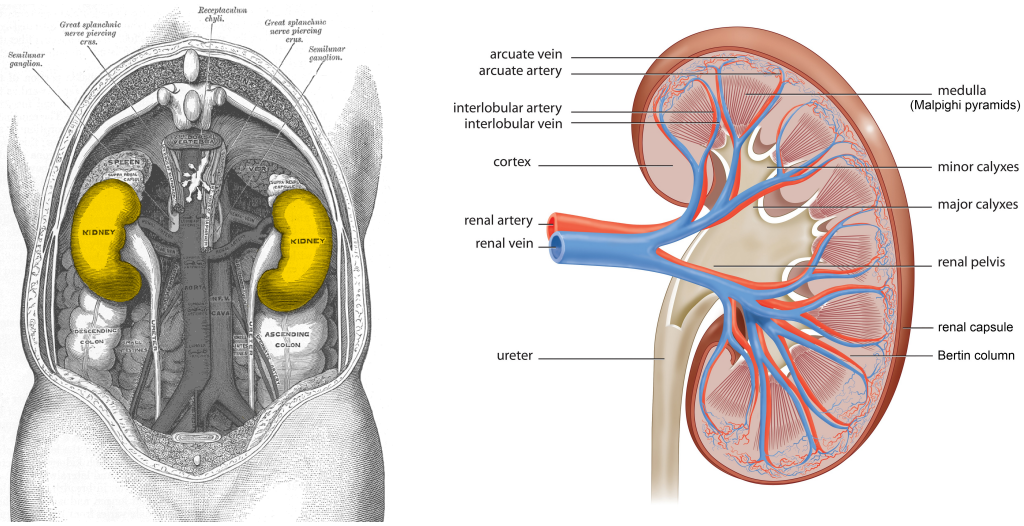


Figure 4.1: (Left) Location of the kidneys. (Right) Anatomy of the kidney.

the kidney in Section 4.1, and explain why the model ϕ_0 will be chosen as an ellipsoid for kidney segmentation.

4.1 Clinical background on the kidney

In this section, we present some general clinical information on the kidney, which is the organ of interest throughout the whole chapter.

4.1.1 Anatomy of the kidney

Kidneys are vital organs present in many species, that are most often paired and located in the abdomen (see Figure 4.1). They play an essential part of the urinary system by serving the body as a natural filter of the blood and removing wastes which are diverted to the urinary bladder. They also have an important role in the regulation of the blood pressure or the maintenance of acid-base balance. Kidneys receive blood from the paired renal arteries, and drain it towards the renal veins.

The kidney is a convex organ with a bean-shaped structure. It measures approximately 12 cm in length, 6 cm in width and 3 cm in thickness. It is important to keep such dimensions in mind as they can be a diagnosis indicator, since dysfunctional kidneys tend to atrophy. The concave surface (the renal hilum) is the point at which the renal artery enters the organ, and the renal vein and ureter leave. The kidney is enclosed by a tough fibrous tissue (the renal capsule), which is itself surrounded by peripheral fat. Kidneys are therefore rather rigid organs: for a given patient, although they may move from one image to another depending on the breathing state, they do not really deform.

The inner structure of the kidney can be divided into two main parts (see Figure 4.1), respectively located in the exterior and interior part:

Parenchyma The parenchyma is composed of both the medullary and the cortical zone. The medulla is the set of all Malpighi pyramids. Conversely, the cortex is defined as the union of the superficial layer of the kidney and the Bertin columns that separate each pair of pyramids. The parenchyma contain mostly nephrons which are the basic structural and functional units of the kidney that actually filter the blood. They produce the urine that is drained by a single collecting duct that join the Malpighi pyramids.

Sinus The sinus is the most inner part of the kidney. It is composed of minor and major calyxes that receive the urine from the pyramids and empty it in the renal pelvis. The renal pelvis goes out of the kidney end becomes the ureter that finally reaches the bladder.

4.1.2 Pathologies of the kidney

Although it is possible to live with only one kidney, it is considered as a vital organ. When its function becomes really insufficient, physicists resort to renal transplantation or dialysis. The lifespan of the patient is then strongly shortened.

Among the different renal pathologies (*e.g.* renal tumours and cysts, acute kidney injury, urinary tract afflictions), cancers receive a lot of attention. Around 210,000 new cases of kidney cancer are diagnosed in the world each year [Lindblad & Adami, 2002], which represents 2% of all cancers. There are several types of kidney cancer, but most cases are *renal cell cancers* (RCC) that are sometimes also called *renal adenocarcinoma* or *renal cell carcinoma*. This type of cancer develops from a cell in a kidney tubule which becomes cancerous. In time the tumour may grow through the wall of the kidney and invade nearby tissues and organs such as the muscles around the spine, the liver, blood vessels, etc. Some cells may break off into the lymph channels or bloodstream. The cancer may then spread to lymph nodes or spread to other areas of the body (metastases). Though anyone can develop RCC, the risk increases with age. Moreover, smoking and obesity are also potential risk factors.

RCC can be divided into several sub-types of cancers. Classification of tumours can be done by performing a biopsy but also by non-invasive means thanks to medical imaging (usually by studying how a contrast agent diffuses in the lesion). Such taxonomy is paramount as it allows to predict how aggressive the tumours are and how they respond to various medical treatments. While the most effective strategy is to perform a *nephrotomy* or *nephrectomy* (partial or complete removal of the kidney), this constitutes a heavy surgical operation to which not all patients are candidates. Another common treatment is the radiotherapy, which uses high energy beams to irradiate canerous tissue and thus kill cancerous cells. This however comes with secondary effects and risks associated to the radiations.

When the tumour site is not too large, minimally invasive procedures such as *radiofrequency ablation* (RFA) are proposed to patients that cannot undergo an open surgery. RFA consists in inserting a needle through the patient's skin and into the tumour. A probe, that is placed inside this needle, generates a high-frequency electrical current, which induces a local heating. The cells in the neighborhood of the probe tip are thus destroyed by heat (see

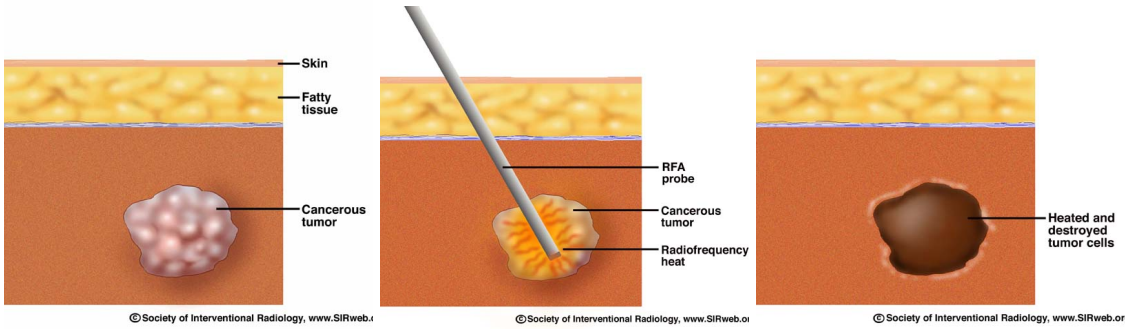


Figure 4.2: Radiofrequency ablation of a tumour: the cancer cells are destroyed by heat generated by a probe. The operation is thus minimally invasive. (Source: Society of Interventional Radiology)

Figure 4.2). This technique appears very promising as it presents major advantages [Kwan & Matsumoto, 2007] such as a reduced perioperative morbidity, a shortened hospital stay, a faster recovery and a better preservation of the renal parenchyma. However, the needle insertion requires a very high accuracy so radiologists are assisted by an imaging modality. Most of the time, they have to use an US system to image the kidney (and more particularly the tumour) in real-time while they insert the needle. Moreover, the complete destruction of cancer cells is hard to guarantee. The range of effectiveness depends on the heat diffusion and therefore on the ablation duration. A trade-off between complete tumour ablation and healthy tissue preservation is difficult to achieve. Therefore, it is paramount to check afterwards whether there is a recurrence or not. For such follow-up purposes, medical imaging will be once again extremely valuable.

4.1.3 Kidneys in medical images

Various imaging modalities are used to assess the kidneys function or for the diagnosis and follow-up of renal lesions (see Figure 4.3). In most modalities, one may distinguish three main areas of the kidney that have different appearances: the cortical zone, the medulla and the sinus. This particular structure generally helps distinguishing the kidney from surrounding tissues in images.

CT (and more specifically contrast-enhanced CT) has been considered as a gold-standard for imaging the kidney for quite a long time [Semelka et al., 1993]. However, the nephrotoxicity of the iodinated contrast media is a problem. Consequently, MR imaging is often performed in cases of impaired renal function or for patients allergic to the iodinated contrast agent. Yet this technique has some limitations including low accessibility and movement artifacts in poorly collaborating patients. Moreover, a major concern has been raised regarding nephrogenic systemic fibrosis which may be associated with the use of gadolinium contrast agents in patients with chronic renal failure.

Conversely, conventional (US) and contrast-enhanced ultrasound (CEUS) are a good alternative as they are inexpensive, easy to perform, non-nephrotoxic and non-invasive

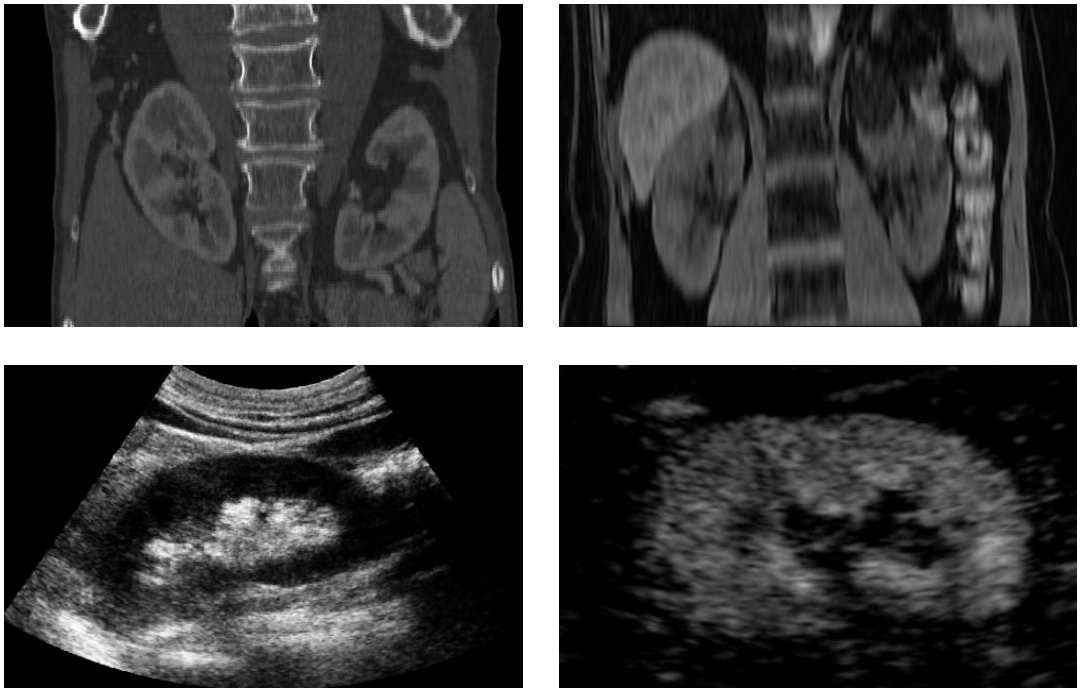


Figure 4.3: Imaging of the kidney in various modalities. From left to right, top to bottom: CT, MR, US and CEUS data. (Source: Prof. Jean-Michel Correas)

imaging techniques. Their benefits for lesion assessment have already been proven in a number of works [Albrecht et al., 2004; Setola et al., 2007]. They however suffer from a usually poor image quality and their user-dependency. Automated and robust tools to assist the radiologist during diagnosis or therapy would therefore be extremely valuable.

In this chapter, we propose different workflows for automated (and interactive in case of failures) kidney segmentation in CT, US and CEUS images. This constitutes a precise way for the clinician to assess the renal volume, which is relevant during diagnosis. We then extend these methods in Chapter 5 for co-segmentation and registration of such images. This yields even better segmentation and paves the way for new applications including fusion of US and CEUS 3D images, registration of pre/post therapy images to control the effectiveness of the treatment or sequence stabilization to estimate perfusion pharmacokinetical parameters.

4.1.4 Estimating the kidney as an ellipsoid

As described in Section 2.7, the implicit template deformation algorithm consists in minimizing by gradient descent a non-convex energy. Therefore, the result will be a local minimum of this energy. This minimum depends on the algorithm initialization and in particular on the choice of ϕ_0, which we discuss in this section.

Figure 4.4 shows an overview of the variety among kidneys. One can notice that an ellipsoid could roughly represent each of the "bean-like" kidney shapes. Actually, this approximation is widespread in clinical practice (*e.g.* [Bakker et al., 1999; O'Neill et al.,

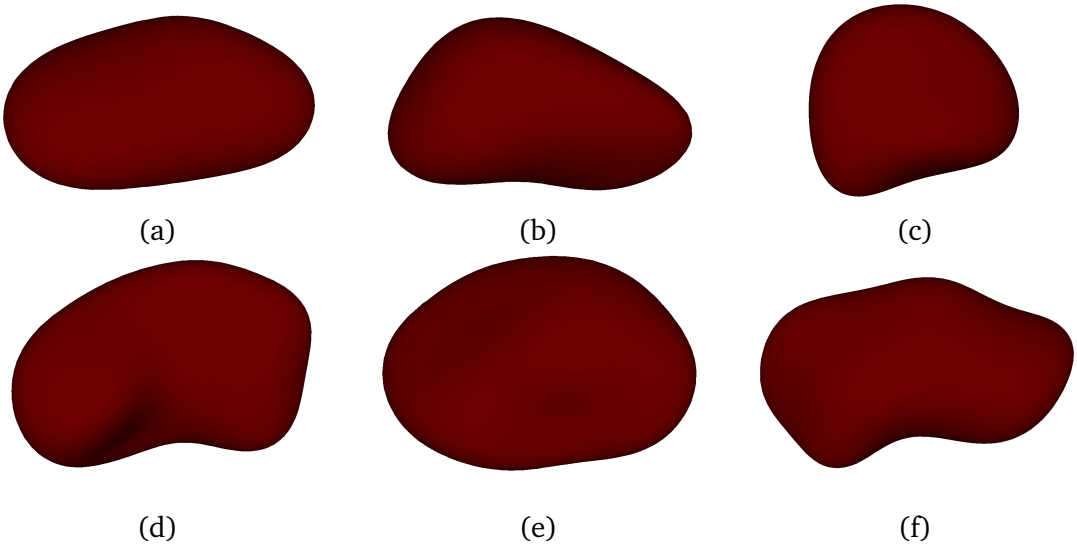


Figure 4.4: A variety of kidney shapes from six different patients (a-f). Every shape can be roughly represented by an ellipsoid.

2005]) when clinicians want to assess the renal volume. Instead of segmenting the complete organ, they draw the 3 main axis of the kidney and use the ellipsoid volume formula, *i.e.*

$$\text{Volume} = \frac{4\pi}{3 \times 2^3} \times \text{length} \times \text{width} \times \text{depth} \,. \tag{4.1}$$

We will use the same approximation to initialize the implicit template deformation algorithm. Several reasons advocate for the choice of ϕ_0 as an ellipsoid (over a mean kidney model, for example). First, ellipsoids are simple mathematical objects and it is generally easier to find such shapes in an image: dedicated algorithms can be used. Moreover, their symmetry allows (i) to elude the question of whether the target organ is a left or right kidney and (ii) can be exploited to reduce computational times. Finally, using a simpler shape allows a gain of robustness while the loss of precision is not detrimental as the template only serves as a shape prior.

As we represent shapes implicitly, we have to define an implicit function $\phi_{\mathscr{E}} : \Omega \to \mathbb{R}$ for each ellipsoid $\mathscr{E}$. While the signed Euclidean distance function to $\mathscr{E}$ is a natural choice, we rather use the algebraic distance that can be directly parametrized by a point $\mathbf{c} \in \mathbb{R}^d$ and a positive-definite matrix $\mathbf{M} \in M_d(\mathbb{R})$:

$$\phi_{\mathscr{E}}(\mathbf{x}) = 1 - (\mathbf{x}-\mathbf{c})^T \, \mathbf{M} \, (\mathbf{x}-\mathbf{c}) \tag{4.2}$$

With this representation, there is a one-to-one correspondence between an ellipsoid $\mathscr{E}$ and its center/matrix $(\mathbf{c}, \mathbf{M})$. We shall thus denote its implicit function either $\phi_{\mathscr{E}}$ or $\phi_{\mathbf{c},\mathbf{M}}$.

In each of the following sections (that correspond to different clinical applications), we will describe how we can find the initial ellipsoid $\mathscr{E}$ to be deformed (or equivalently its center

and matrix).

4.2 Kidney segmentation in 3D CT images

This work has been done in collaboration with Rémi Cuingnet and presented in the MICCAI 2012 conference [Cuingnet et al., 2012].

4.2.1 Introduction

In this section, we address the problem of kidney segmentation in 3D CT images. CT images analysis is an active field of research, and a number of papers in the literature report good segmentation results in automated renal segmentation [Tsagaan et al., 2002; Spiegel et al., 2009; Li et al., 2011; Khalifa et al., 2011]. Thus it constitutes a first interesting application to test the potential of the template deformation framework.

When designing an automated image processing algorithm on CT images, one often relies on their somehow standardized acquisition process :

- Because image intensities have a physical meaning in CT, they are consistent from one image to another. It is therefore generally easier to design image-based forces to be used for segmentation.
- The patient's position is most of the time fixed: they lie on their back upon the table of the CT scanner. The orientation can therefore be assumed as known.

However there are still other uncertainties that are often eluded in the aforementioned literature: what is the field of view ? has a contrast agent has been injected ? if yes, what contrast phase is it ? Figure 4.5 shows the variability of such images within a clinical database. Our goal is to build a pipeline that is generic enough to account for all the different cases.

Let us first briefly review the previous works concerning kidney segmentation in CT images. In [Spiegel et al., 2009] and [Li et al., 2011], the authors used the Active Shape Model framework to learn the kidney mean shape and principal modes of variation, in order to constrain the segmentation. Recently Khalifa *et al.* proposed a level-set approach, based on a new force combining shape and intensity priors as well as spatial interactions, which showed promising results [Khalifa et al., 2011]. However, all of them were assessed on small datasets (41, 17 and 20 volumes in [Spiegel et al., 2009], [Li et al., 2011] and [Khalifa et al., 2011] respectively). Moreover, those algorithms are either based on a manual initialization, or tested on images already cropped around the kidney. A fully automatic method has already been introduced in [Tsagaan et al., 2002], but their detection of the region of interest presents limitations. First, it relies on hard geometrical constraints, which requires knowledge on the field of view. Then, a rough search is done by template matching, which is not robust to pathologies or variations in kidney orientation.

Here, we propose a fast and completely automatic method to detect and segment both kidneys in any kind of CT image: acquired at different contrast phases (or without contrast)

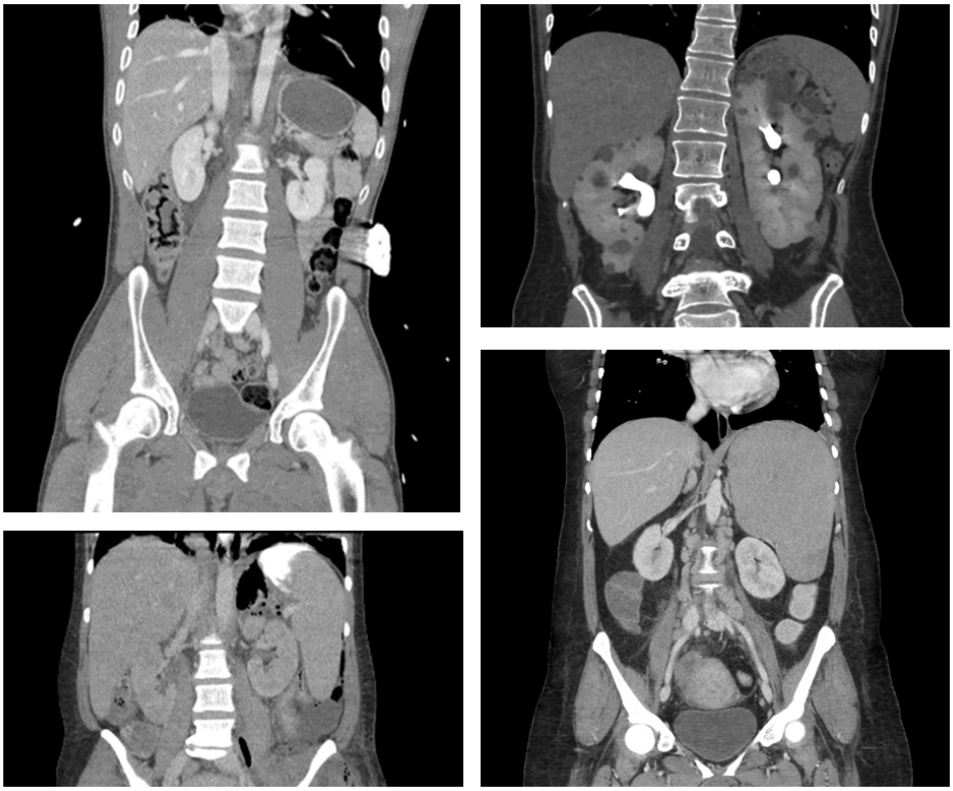


Figure 4.5: Slices of different CT volumes showing the high variability of the images from the database. Note the various fields of view, the different contrast phases and the possible presence of pathologies.

with various fields of view, from both healthy subjects and patients with kidney tumors. Both the kidney localization (thus the template ϕ_0 initialization) and the construction of the image-based term r used for the segmentation will be learnt via random forests.

4.2.2 Bounding box detection via regression forests

This subsection presents a fast and reliable estimation of the kidneys' locations. Various approaches for anatomy detection and localization have been proposed in the literature. We propose a regression-based method in two steps. The whole image is first used to provide an estimate of the region of interest which is then refined using local information only.

Background on organ detection

Registration-based approaches using labeled atlases (*e.g.* [Fenchel et al., 2008]) have often been used for this problem. However such approaches are subject to registration errors due to inter individual variability. The robustness of the registration step can be improved by using multi-atlas or multi-template techniques [Isgum et al., 2009] but at the cost of an increase in computation time.

More recently, supervised learning methods have been used for this detection problem to better take into account interindividual variability. Most *classification-based* detection algorithms consist in constructing a classifier whose role is to predict from local features to which organ a voxel belongs. However, by considering only local features, such approaches do not benefit from anatomical contextual information. To overcome this shortcoming, Criminisi *et al.* used a generalization of Haar features that models contextual information [Criminisi et al., 2009]. Instead of classifying each voxel, some authors consider the detection problem as finding a vector of parameters describing the organ locations. Such parameters can describe contour line or surface of an organ [Georgescu et al., 2005] or more simply bounding boxes around the different organs of interest [Zheng et al., 2008]. The role of the classifier is then to predict whether a set of parameters is correct or not. Zheng *et al.* used a greedy approach [Zheng et al., 2008] to avoid searching the whole parameter space, which is intractable.

Zhou *et al.* showed that finding a set of continuous parameters from an image is by definition a multiple-output regression problem [Zhou et al., 2005]. More precisely, they proposed a boosting ridge regression to detect and localize the left ventricle in cardiac ultrasound 2D images. *Regression-based* techniques do not require an exhaustive search of parameters. Other regressors such as regression forests and random ferns have also been proposed [Criminisi et al., 2011b; Pauly et al., 2011]. Note that voting-based approach such as the Generalized Hough Transform (GHT) can be also considered as regression [Ballard, 1981; Ecabert et al., 2008].

In the following, we consider regression forests to simultaneously detect both kidneys. Regression forests [Breiman, 2001; Criminisi et al., 2011a] are particularly well adapted to this problem in clinical routine since, thanks to their tree structures, they allow very fast testing with nonlinear regression functions. Since there is no explicit regularization, random forests require a large number of training samples to avoid overfitting the training data. Here, this is not a limitation since the training samples are the voxels of the training CT scans.

Coarse localization of the kidneys

We consider the detection step as the problem of finding bounding boxes around both kidneys. First, we find a coarse positioning based on contextual information adapting the approach proposed in [Criminisi et al., 2011b]. Then, the position of each box is refined based on local information.

Each bounding box is parameterized by a vector in $\mathbb{R}^6$ (two points in 3D). A random forest is trained on CT scans with known kidney bounding boxes to predict for each voxel the relative position and size of the kidneys. Since CT intensities (expressed in Hounsfield units) have direct physical meaning, as explained in [Criminisi et al., 2011b], the features used are the mean intensities over displaced, asymmetric cuboid regions. To allow a much faster training, we used residual sum of squares (RSS) instead of the information gain for the node optimization in the training stage. Note that optimizing the RSS comes to minimizing the

trace of the covariance matrix at each node instead of its determinant. We did not notice any differences in term of prediction accuracy.

This step gives a first estimate of the kidneys' positions and sizes. By construction, the relative estimated position of the left and right kidneys are strongly correlated. Such a correlation ensures coherent results but may not reflect the whole possible interindividual variability. This might be critical when the number of subjects in the training set is low. Increasing the number of subjects in the training set could overcome this shortcoming. However obtaining ground truth often turns out to be a tedious task, even if the user interaction can be reduced using active learning techniques [Iglesias et al., 2011]. To overcome this shortcoming, we propose a refinement step of the bounding boxes that relaxes the correlation between the two kidneys' position.

Refinement of the region of interest

This step consists in refining the left and right kidneys' positions based on local information only. The constraints between the kidneys' relative positions are relaxed by treating both kidneys independently. For each kidney, a regression forest is trained to predict, from every voxel located in its neighborhood, the relative position of the kidney's center. We used the same training set as in the previous step. The features used for this step are, for each voxel, its intensity and its gradient magnitude, as well its neighbors'.

For testing, only the voxels in the neighborhood of the center of the bounding box predicted by the first step are considered. As depicted in Figure 4.6.b, each voxel $\mathbf{x}$ then votes for a location $\hat{\mathbf{c}}_{\mathbf{x}}$ of the kidney's center. For robustness sake, the final location is defined as a median estimate:

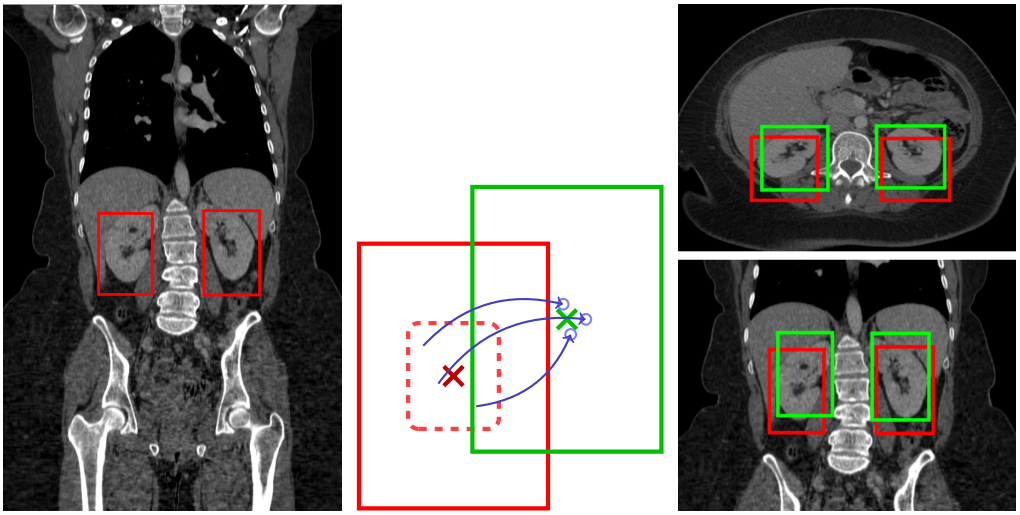
$$\hat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^3}{\arg\min} \sum_{i=1}^{K} \|\mathbf{c} - \hat{\mathbf{c}}_{\mathbf{x}_i}\|_1 \tag{4.3}$$

where $\left(\hat{\mathbf{c}}_{\mathbf{x}_i}\right)_{1\dots K}$ are the K votes with the highest probability. The final bounding box is then translated accordingly.

To ensure stability, this refinement step is constrained to very small displacements and is iterated until convergence. This can be considered as a cascaded pose regression similar to [Dollar et al., 2010]. An illustration of the kidney detection is given in Figure 4.6 while quantitative results are reported in Section 4.2.5.

4.2.3 Estimation of the image-based term using classification forests

Even when the image is cropped to a region around the kidney, its segmentation remains a challenging task: (i) kidneys are composed of different tissues (cortex, medulla, sinus) resulting in different image intensities, (ii) surrounding organs may touch the kidney without a clear boundary, (iii) the contrast phase of the CT image is unknown. For all these reasons, it is not possible to solely rely on the image intensity, and we rather use it simultaneously with other features.



(a) (b) (c)

Figure 4.6: Illustration of the kidney detection on a CT volume. (a) Initial bounding boxes detected using global contextual information. (b) Refinement step: voxels near the center of the initial bounding box (red) vote for its new center, using only local information. (c) Comparison between the initial (red) and refined (green) bounding box.

In addition to regression, random forests can also be used to perform classification [Breiman, 2001; Criminisi et al., 2011a]. We trained a random forest classifier to predict, for each voxel $\mathbf{x}$ of the previously detected bounding box, the probability $p_{CT}(\mathbf{x})$ of belonging to a kidney. This random forest combines different image features: intensity and first/second order derivatives of the voxel and its neighbors. Decision stumps were used as weak classifiers and the impurity criterion was the Gini index [Breiman, 2001; Criminisi et al., 2011a]. Such probability maps are shown in Figures 4.7.a and 4.7.c. Independently from the contrast-phase, the whole kidney tissues are enhanced, whereas the confusing adjacent structures are removed.

4.2.4 Settings for implicit template deformation

At this stage, we have a probability map p_{CT} cropped around the region of interest, that indicates at each voxel $\mathbf{x}$ the probability to be inside the kidney. We will find a coarse estimate of the kidney's position, size and orientation using simple statistics on this map. The template ϕ_0 is defined as the implicit function $\phi_{\mathbf{c}_\mathscr{E},\mathbf{M}_\mathscr{E}}$ of the ellipsoid such that the point

$$\mathbf{c}_\mathscr{E} = \frac{1}{\int_\Omega p_{CT}(\mathbf{x})\, d\mathbf{x}} \int_\Omega p_{CT}(\mathbf{x})\, \mathbf{x}\, d\mathbf{x} \tag{4.4}$$

is the barycenter of the pixels from the probability map P. $\mathbf{M}_{\mathscr{E}}$ is then proportional to the inverse of the weighted covariance matrix

$$\mathbf{M}_{\mathscr{E}}^{-1} = \frac{\mu}{\int_\Omega p_{CT}(\mathbf{x})\, d\mathbf{x}} \int_\Omega p_{CT}(\mathbf{x})\, (\mathbf{x}-\mathbf{c}_{\mathscr{E}})(\mathbf{x}-\mathbf{c}_{\mathscr{E}})^T\, d\mathbf{x}\,. \tag{4.5}$$

Such ellipsoids are shown in yellow for two cases in Figure 4.7.a and Figure 4.7.c.

Before running the template deformation, we also need to define an image-based term r_{int} (resp r_{ext}) that describes the interior appearance of the kidney (resp. its surrounding). Given the probability map p_{CT}, a natural choice would be to set

$$r_{int}(\mathbf{x}) = (1-p_{CT}(\mathbf{x})) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = p_{CT}(\mathbf{x})\,. \tag{4.6}$$

This definition would give a paramount importance to the level-set $\{p_{CT} = 0.5\}$. Indeed, recall that the segmentation only depends on the difference $r = r_{int} - r_{ext}$. While our probability estimation does return high values inside the kidney and low values outside, there is no particular guarantee that the cut-off value will be at 0.5 exactly. We rather want that the segmentation maximizes the gradient flux of this probability map across its surface, and therefore define

$$r_{int}(\mathbf{x}) = -\Delta p_{CT}(\mathbf{x}) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = 0\,. \tag{4.7}$$

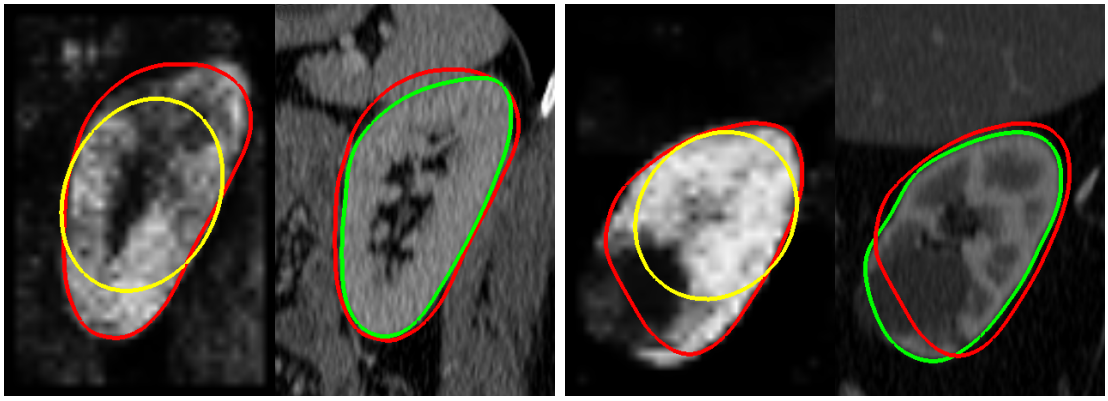
More details on the use of the Laplacian Δ were given in Remark 2.1 (see Section 2.4.3).

Results of template deformation with this setting are displayed in red in Figure 4.7. We notice that the segmentation is globally satisfying but clearly lacks precision. This is due to the fact that the probability map p_{CT} is estimated using large-scale features. Thus p_{CT} can only be low-resoluted and have a poor precision, especially near the kidney boundaries. To overcome this problem, the last iterations are performed at a finer scale directly on the input image: $r_{int} = -\Delta I$. The final results are shown in green in Figure 4.7.b and Figure 4.7.d.

4.2.5 Material and results

The validation was performed on a representative clinical dataset of 233 CT volumes from 89 subjects including diseased patients. The scans were contrast-enhanced or not and with various fields of view and spatial resolutions. They have between 33 and 973 (with a mean of 260) 512×512 slices with slice resolutions ranging from 0.5 to 1 mm and an interslice resolution ranging from 0.5 to 3.0 mm. 16% of the kidneys were slighlty truncated, but were nevertheless included in the evaluation to keep it clinically representative. The database was split into a training set of 54 volumes from 26 randomly selected patients, and a testing set composed of the other 179 volumes from 63 patients.

The proposed algorithm used 3 regression forests and 2 classification forests. Each forest was composed of 7 trees with a maximum tree depth $d = 15$ and a minimal node



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

Figure 4.7: Illustration of the two-step kidney segmentation on two cases: (a-b) non-contrasted volume of a healthy patient, (c-d) contrast-enhanced image of a kidney with a tumor. The kidney probability maps (a) and (c) are learned with a random forest, and used to coarsely segment the kidney (red) by deforming an initial ellipsoid (yellow). The segmentation is then refined (green) using the original volumes (b) and (d).

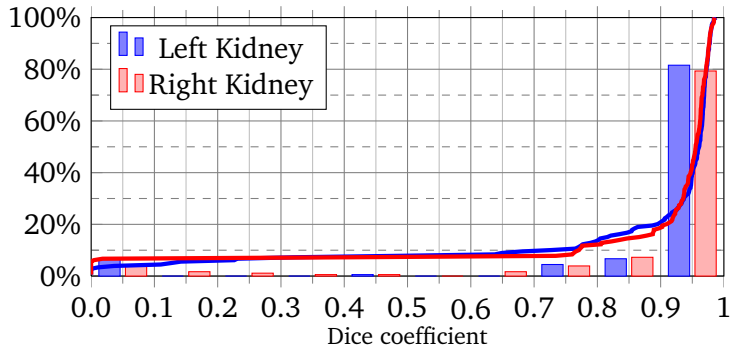
size $n = 100$. We did not notice a high sensitivity of the results to these parameters value. The whole training procedure lasts ~ 5 hours. Times are indicated for an optimized C++ implementation on a standard computer (Intel Core i5 2.67 Ghz, 4GB RAM).

Results of kidney detection

Detection errors were defined as the absolute difference between predicted and true wall positions averaged over all the bounding box sides. The distance between the predicted bounding box center and the ground truth was also used to assess the detection accuracy. These results are given in Table 4.1 and compared to those reported in [Criminisi et al., 2011b]. The refinement step, for a low extra time cost, greatly increases the accuracy of the bounding box detection (e.g. the median center error is divided by 3).

Detection	**Walls error** (mm)		**Center error** (mm)		**Time** (s)
	Left	Right	Left	Right	Left+Right
Baseline	17 ± 17 (13)	19 ± 18 (12)	–	–	–
Coarse	12 ± 7 (10)	13 ± 6 (11)	23 ± 14 (20)	26 ± 13 (23)	2.1 ± 0.5 (2.0)
Refined	7 ± 10 (5)	7 ± 6 (6)	11 ± 18 (6)	10 ± 12 (7)	2.8 ± 1.7 (2.4)

Table 4.1: Detection results reported as Mean ± Standard-deviation (Median), compared between the baseline method [Criminisi et al., 2011b] and the two steps of the proposed approach.



Dice	Left	Right
1st quartile	0.93	0.93
median	0.96	0.96
3rd quartile	0.97	0.97
maximum	0.99	0.99

Figure 4.8: Distribution of the Dice coefficient between the ground truth and the automatically segmented kidneys. Red and blue lines show the cumulative distribution.

Results of kidney segmentation

The total execution time for the detection of and segmentation of both kidneys is around 10 s. The results of the automatic segmentation including the detection step were evaluated using the Dice coefficient between the segmentation result S and the ground truth GT, defined as

$$Dice(S, GT) = 2\,\frac{\mathrm{Vol}(S \cap GT)}{\mathrm{Vol}(S) + \mathrm{Vol}(GT)} \tag{4.8}$$

where $\mathrm{Vol}(X)$ denotes the volume of a region X. The Dice coefficient is 1 when the segmentation S is perfect (equal to the ground truth GT) and 0 when S and GT do not overlap at all.

Figure 4.8 shows the histograms of the scores for both kidneys. 80 % of the kidneys were correctly detected and segmented (Dice > 0.90). The bounding box detection failed in only 6% of the cases (which resulted in Dice coefficients < 0.65). Note that there is not a meaningful difference between the results on the left and right kidney. Segmentation results on images with various acquisition conditions are also reported in Figure 4.9.

Discussion

This section presented a fully automatic method to detect and segment both kidneys in any CT volume using random regression and classification forests. Regression forests were used to estimate the kidneys' positions. A classification forest was then used to obtain a probability map of each kidney. The segmentation was carried out with an implicit template deformation algorithm. The full automation and the execution time are compatible with clinical routine. Results show that our method provides an accurate segmentation in 80% of the cases despite the highly heterogeneous database. Remaining cases were mostly due to pathological kidneys not represented in the training set. Such cases could be quickly corrected by the clinician, since the template deformation framework can take into account user interactions (see Section 2.7.4). We also emphasize the genericity of our framework, that could easily be extended to other organs: it has been for example successfully applied

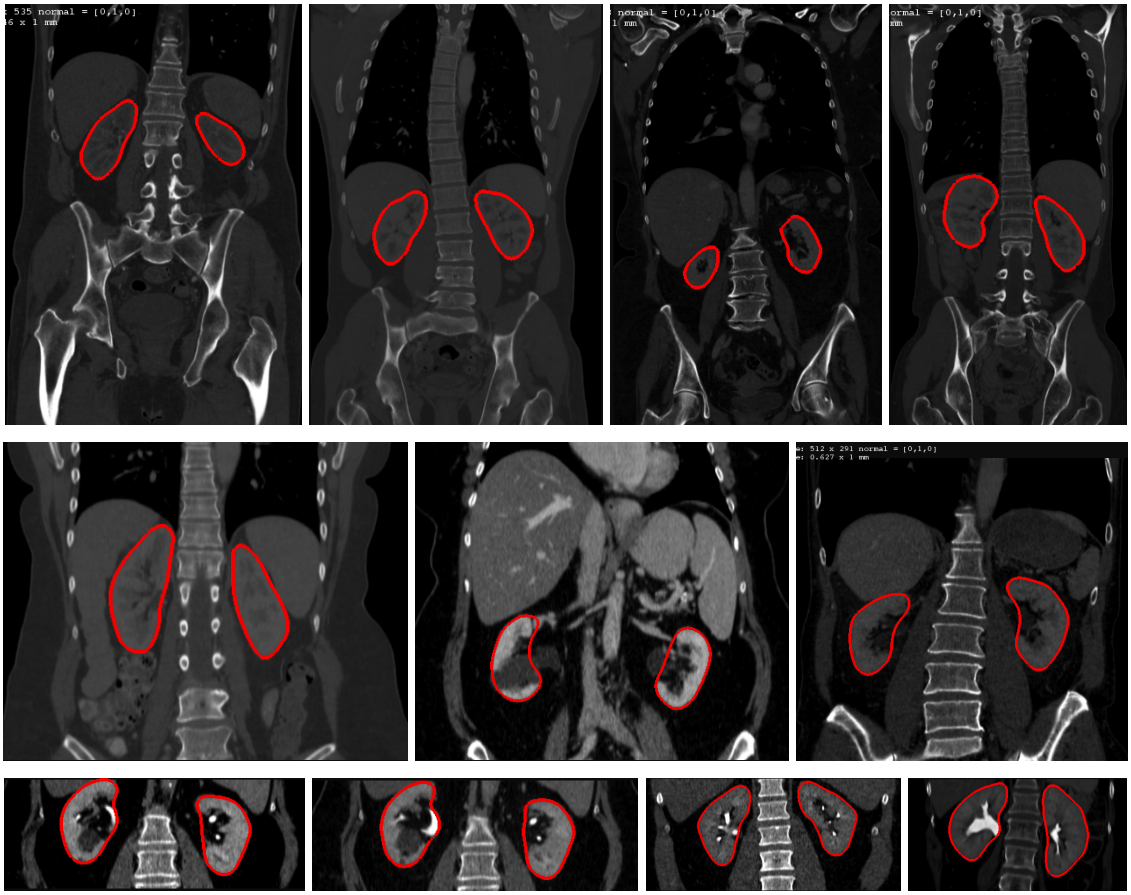


Figure 4.9: Examples of automatic kidney segmentation results (red) on a subset of the testing database with various acquisition conditions.

in [Gauriau et al., 2013] for liver segmentation.

In the next section, we address the problem of kidney segmentation in 3D contrast-enhanced ultrasound (CEUS) images which is, as we will see, much more challenging.

4.3 Kidney segmentation in 3D CEUS images

This section is an extended version of the paper presented in the IEEE ISBI 2012 conference [Prevost et al., 2012c]. It results from a collaboration with Prof. Jean-Michel Correas (Hôpital Necker, Paris, France) who kindly provided us with clinical data.

4.3.1 Introduction

This section is dedicated to the kidney segmentation in another kind of volumes : 3D contrast-enhanced ultrasound images. Contrast-enhanced ultrasound (CEUS) consists in acquiring an ultrasound image after injecting in the patient's blood a contrast agent made of gas-filled

microbubbles. Since those bubbles have a different acoustic response from the tissues, they can be isolated and images showing only the blood flow can be generated [Albrecht et al., 2004]. This modality is particularly valuable for visual assessment of the functioning of highly vascularized organs like kidneys. Figure 4.10 shows two examples of both conventional ultrasound and contrast-enhanced ultrasound images of the kidney. After a few seconds, the kidney is completely filled with microbubbles injected in the blood and is therefore highlighted in the enhanced image (compared to its fatty surrounding that produces almost no signal). More information about this modality is available in Appendix A.

Unfortunately the usually poor quality of CEUS images makes any automated analysis challenging: in addition to having powerful speckle noise, the image is extremely grainy and almost binary as a result of ultrasound interactions with individual bubbles. Unlike in conventional B-mode ultrasound (US) [Noble & Boukerroui, 2006], very few segmentation methods of 3D CEUS images have been reported in the literature. Among them, Gasnier *et al.* introduced an interactive approach to segment and analyze tumors in this modality [Gasnier et al., 2010]. However, their framework was specific to lesion segmentation, just as the automatic methods proposed in [Kissi et al., 2004; Prevost et al., 2012a]. In [Ma et al., 2009], an automated algorithm is developed to segment the heart left ventricle. This method, although applicable to other organs, does not provide any natural way to refine or correct the result interactively. Besides, it has been designed for images acquired with a particular transducer, producing sparse rotated slices instead of a whole 3D volume. Our approach will therefore constitute the first attempt at segmenting fully-automatically the kidney in CEUS images.

The problem of kidney segmentation in CEUS images is quite different than in CT images, and we will use a different strategy than in Section 4.2. Indeed, the field of view is much narrower in US than in CT so a regression approach is not necessary. However, the acquisition is not standardized so the kidney is in an arbitrary orientation: instead of detecting a bounding box, we will directly aim at detecting an ellipsoid with an adequate and original method (see Section 4.3.2). Furthermore, contrary to CT images, intensities in CEUS images are less reliable so we will use an edge-based term (see Section 4.3.3) instead of learning a probability map.

4.3.2 Initialization via a robust ellipsoid estimation

Since voxels with high intensities are likely to be inside the kidney, the detection problem in CEUS images can be initially reduced to finding the smallest ellipsoid encompassing most of the hyperechoic voxels. A large number of methods (e.g. Hough transforms [Guil & Zapata, 1997; McLaughlin, 1998]) has already been proposed to detect ellipses in images [Wong et al., 2012]. However their extension to 3D, though possible, are usually computationally expensive mainly because of the number of parameters to estimate (9 for a 3D ellipsoid). Furthermore, they do not explicitly use the fact that only one ellipsoid is present in the image. On the other hand, statistical approaches like robust Minimum Volume Ellipsoid (MVE) estimators [Van Aelst & Rousseeuw, 2009] are better suited but require prior knowledge on the proportion of outliers (here the noise, artifacts or neighboring structures), which

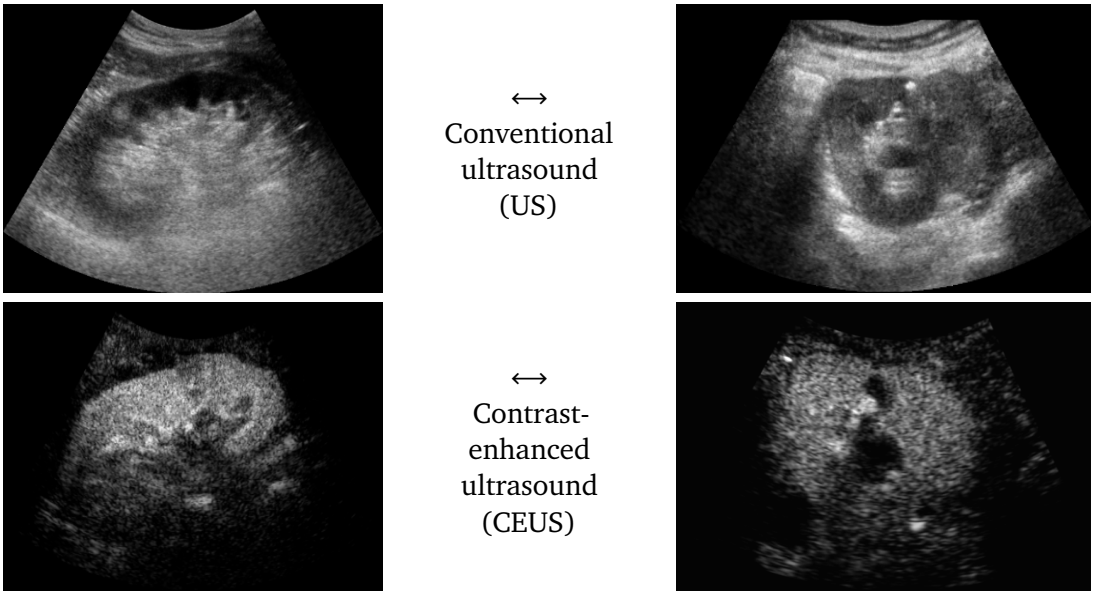


Figure 4.10: Slices of conventional (US) and contrast-enhanced ultrasound (CEUS) 3D images of the kidney for two different patients (left and right).

may vary from one image to another and is thus not available. We therefore propose an original variational framework, that is robust and fast, to estimate the best ellipsoid in an image $I : \Omega \subset \mathbb{R}^d \to \mathbb{R}^+$. As we will see, it can be viewed as a generalization of the weighted barycenter/covariance matrix estimation proposed in Section 4.2.4.

A variational framework for robust ellipsoid estimation

In the considered framework, we recall that an ellipsoid is implicitly represented using an implicit function $\phi : \Omega \to \mathbb{R}$ that can be parametrized by the center of the ellipsoid $\mathbf{c} \in \mathbb{R}^d$ and its sizes and orientations encoded by a $d \times d$ positive-definite matrix $\mathbf{M}$ as in (4.2). The detection method should be robust to outliers, *i.e.* bright voxels coming from noise, artifacts or other neighboring structures. Excluding those outliers is done by estimating a weighting function $w : \Omega \to [0, 1]$ that provides a confidence score for any point $\mathbf{x}$ to be an inlier. The ellipsoid estimation is then formulated as an energy minimization problem with respect to $\mathbf{c}$, $\mathbf{M}$ and w:

$$\min_{\mathbf{c},\mathbf{M},w}\{E_{det}(\mathbf{c},\mathbf{M},w) = -\int_\Omega \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x})\, w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x} + \mu.\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right).\left(\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x}\right)\} \tag{4.9}$$

$$\text{with} \quad \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x}) = 1-(\mathbf{x}-\mathbf{c})^T\,\mathbf{M}\,(\mathbf{x}-\mathbf{c})$$

$$\text{and} \quad \mathrm{Vol}(\phi_{\mathbf{c},\mathbf{M}}) = \mathrm{Vol}(\mathbf{M}) = \frac{4\pi}{3}\sqrt{\det \mathbf{M}^{-1}} \quad \text{the ellipsoid volume.}$$

The ellipsoid detection energy E_{det} is composed of two terms:

- a **data-fidelity term**: The first term is an integral over the whole image domain Ω of the product $\phi_{\mathbf{c},\mathbf{M}}$ by wI. Note that wI is highly positive at voxels that have a high intensity but are not outliers. To minimize the energy, such voxels must therefore be included inside the ellipsoid *i.e.* where ϕ is positive.

- a **regularization term**: The second term penalizes the volume of the ellipsoid $\mathrm{Vol}(\mathbf{M})$ with respect to the domain volume $|\Omega|$. The logarithm provides a statistical interpretation of the problem and eases the minimization of the energy, as will be seen hereafter. It is normalized by $\int wI$ and weighted by a trade-off parameter $\mu > 0$. High values of μ therefore favor smaller ellipsoids.

Numerical optimization

This ellipsoid estimation process can be viewed as robustly fitting a Gaussian distribution to the bright and meaningful pixels of the image by minimizing its negative log-likelihood. Therefore E_{det} has a statistical meaning and minimizers $(\mathbf{c}^*,\mathbf{M}^*,w^*)$ of E_{det} have a closed form, as detailed in the following proposition.

Proposition 4.1. *Let $\mathbf{c}\in\mathbb{R}^d$, $\mathbf{M}\in M_d(\mathbb{R})$ and $w:\Omega\to[0,1]$. The two following assertions hold:*
(i) When w is fixed, minimizers $(\mathbf{c}^*,\mathbf{M}^*)$ of $E_{det}(\cdot,\cdot,w)$ are given by :

$$\mathbf{c}^* = \frac{\displaystyle\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, \mathbf{x}\, d\mathbf{x}}{\displaystyle\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x}} \tag{4.10}$$

$$\mathbf{M}^* = \frac{\displaystyle\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T\, d\mathbf{x}}{\displaystyle\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x}} \tag{4.11}$$

(ii) When $\mathbf{c}$ *and* $\mathbf{M}$ *are fixed, minimizer* w^* *of* $E_{det}(\mathbf{c},\mathbf{M},\cdot)$ *is given by :*

$$w^*(\mathbf{x}) = \mathbb{1}_{\left\{\phi_{\mathbf{c},\mathbf{M}} - \mu \log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right) \geq 0\right\}}(\mathbf{x}) \quad . \tag{4.12}$$

Proof. We begin the proof by considering the minimization E_{det} with respect to the ellipsoid center. We have

$$\min_{\mathbf{c}\in\mathbb{R}^n} E_{det}(\mathbf{c},\mathbf{M},w) = \min_{\mathbf{c}\in\mathbb{R}^n} -\int_\Omega \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x})\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x} \tag{4.13}$$

$$= \min_{\mathbf{c}\in\mathbb{R}^n} \int_\Omega (\mathbf{x}-\mathbf{c})^T\, \mathbf{M}\,(\mathbf{x}-\mathbf{c})\;\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x} \tag{4.14}$$

The right member of (4.14) is a second-order polynomial expression in $\mathbf{c}$ that admits a unique minimum $\mathbf{c}^*$. Setting its derivative to zero yields

$$-\int_\Omega 2\;\mathbf{M}\,(\mathbf{x}-\mathbf{c}^*)\;\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x} = 0 \tag{4.15}$$

$$\Leftrightarrow \int_\Omega 2\;\mathbf{M}\mathbf{c}^*\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x} = \int_\Omega 2\;\mathbf{M}\mathbf{x}\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x} \tag{4.16}$$

Since $\mathbf{M}$ is positive-definite (and therefore invertible), we then obtain the first result :

$$\mathbf{c}^* = \frac{\displaystyle\int_\Omega w(\mathbf{x})\; I(\mathbf{x})\; \mathbf{x}\; d\mathbf{x}}{\displaystyle\int_\Omega w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x}} \;. \tag{4.17}$$

As for the minimization with respect to $\mathbf{M}$ when $\mathbf{c}$ is fixed, we have:

$$\mathbf{M}^* = \underset{\mathbf{M}\in M_d(\mathbb{R})}{\arg\min} -\int_\Omega \left(1-(\mathbf{x}-\mathbf{c})^T\,\mathbf{M}\,(\mathbf{x}-\mathbf{c})\right)\; I(\mathbf{x})\; w(\mathbf{x})\; d\mathbf{x}$$

$$+\mu.\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right).\left(\int_\Omega w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x}\right) \tag{4.18}$$

$$\tag{4.19}$$

which simplifies in

$$\mathbf{M}^* = \underset{\mathbf{M}\in M_d(\mathbb{R})}{\arg\min} +\int_\Omega (\mathbf{x}-\mathbf{c})^T\,\mathbf{M}\,(\mathbf{x}-\mathbf{c})\;\; w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x}$$

$$-\mu\;\frac{1}{2}\;\log(\det\mathbf{M})\;\left(\int_\Omega w(\mathbf{x})\; I(\mathbf{x})\; d\mathbf{x}\right) \tag{4.20}$$

We then recall that

$$\frac{d(\det \mathbf{M})}{d\mathbf{M}} = \det \mathbf{M}.(\mathbf{M}^{-1})^T \quad \text{therefore} \quad \frac{d(\log(\det \mathbf{M}))}{d\mathbf{M}} = (\mathbf{M}^{-1})^T, \tag{4.21}$$

giving the following expression for the derivative

$$\frac{dE_{det}}{d\mathbf{M}} = \int (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T \; w(\mathbf{x}) \; I(\mathbf{x}) \; d\mathbf{x} - \frac{\mu}{2}(\mathbf{M}^{-1})^T \int w(\mathbf{x}) \; I(\mathbf{x}) \; d\mathbf{x} \tag{4.22}$$

The first-order optimality condition can be easily expressed as follows :

$$\frac{dE_{det}}{d\mathbf{M}}(\mathbf{M}^*) = 0 \Leftrightarrow (\mathbf{M}^*)^{-1} \int w(\mathbf{x}) \; I(\mathbf{x}) \; d\mathbf{x} = \frac{2}{\mu}\left(\int (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T \; w(\mathbf{x}) \; I(\mathbf{x}) \; d\mathbf{x} \right)^T \tag{4.23}$$

The right-hand side of this equation is a symmetric matrix so it is equal to its transpose. We therefore obtain the result

$$\mathbf{M}^* = \frac{\mu}{2}\left(\frac{1}{\int w(\mathbf{x}) I(\mathbf{x}) d\mathbf{x}} \int (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T w(\mathbf{x}) I(\mathbf{x}) d\mathbf{x} \right)^{-1} \tag{4.24}$$

Finally, rewriting the definition of w^* as

$$\begin{aligned} w^* &= \min_{w:\Omega \to [0,1]} E_{det}(\mathbf{c}, \mathbf{M}, w) \\ &= \min_{w:\Omega \to [0,1]} \int \left(\phi_{\mathbf{c},\mathbf{M}} - \mu \log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|} \right) \right) w(\mathbf{x}) I(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{4.25}$$

emphasizes the linearity of E_{det} with respect to w which is by definition restricted to $[0,1]$. Besides, the minimization can be performed pointwise since all contributions are decoupled. Therefore, at every voxel $\mathbf{x}$ the minimizer $w^*(\mathbf{x})$ is equal to 0 or 1, depending only on the sign of the multiplying factor. Since I has a constant positive sign, it can be dropped off the condition, hence the final result. □

Proposition 4.1 states that $\mathbf{c}^*$ is the barycenter of all voxels $\mathbf{x}$ weighted by $w(\mathbf{x}) \; I(\mathbf{x})$ while $\mathbf{M}^*$ is the inverse of the covariance matrix (up to a constant multiplier) of the same data. Furthermore, w^* is the indicator of the current ellipsoid estimation which has been dilated proportionately to μ. Its purpose is to remove the contribution of the points which are far away from the current ellipsoid and may disturb its refinement.

The weighting function w is initialized to 1 everywhere. Minimization of E_{det} is then performed with an alternate iterative scheme that successively updates the variables $\mathbf{c}$, $\mathbf{M}$ and w, as summarized in Algorithm 2. As the energy E_{det} decreases at each step, the algorithm is guaranteed to converge. We use a criterion based on the norm of the difference between two successive estimates of $\mathbf{c}$ and $\mathbf{M}$ to stop the process. In practice, few iterations are required for convergence (*e.g.* less than 10) and the total computational time is less than a second for a 3D image.

Algorithm 2: Robust ellipsoid detection algorithm

initialization $\forall\ \mathbf{x} \in \Omega,\ w(\mathbf{x}) \leftarrow 1$
repeat
  // Estimation of center $\mathbf{c}$ and matrix $\mathbf{M}$
  $$\mathbf{c} \leftarrow \frac{1}{\int_\Omega wI} \int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, \mathbf{x}\, d\mathbf{x}$$
  $$\mathbf{M}^{-1} \leftarrow \frac{2}{\mu \int_\Omega wI} \int_\Omega w(\mathbf{x})\, I(\mathbf{x})\ (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T\, d\mathbf{x}$$
  // Update of the weighting function w for each $\mathbf{x} \in \Omega$
  if $(\mathbf{x}-\mathbf{c})^T\, \mathbf{M}\, (\mathbf{x}-\mathbf{c}) \leq 1 - \mu \log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right)$ **then**
    $w(\mathbf{x}) \leftarrow 1$
  else
    $w(\mathbf{x}) \leftarrow 0$
until *convergence of* $\mathbf{c}$ *and* $\mathbf{M}$;

Remark 4.1. *The choice of μ in (4.9) is paramount as it controls the number of points taken into account for the ellipsoid matrix estimation. To find a suitable value, let us consider an ideal case of an image I_0 in which there is one white ellipsoid ($I_0 = 1$) on a black background ($I_0 = 0$), whose implicit function is $\phi_{\mathbf{c}_0,\mathbf{M}_0}$. We also assume that the confidence weight is $w \equiv 1$ everywhere. Then the matrix estimated by our approach would be*

$$\begin{aligned}\mathbf{M}^* &= \arg\min_{\mathbf{M}}\ E_{det}(\mathbf{c}_0, \mathbf{M}, \mathbf{1}) \\ &= \left[\frac{2}{\mu}\frac{1}{\int_\Omega I_0(\mathbf{x})\, d\mathbf{x}} \int_\Omega I_0(\mathbf{x})\ \big(\mathbf{x}-\mathbf{c}_0\big)\big(\mathbf{x}-\mathbf{c}_0\big)^T\, d\mathbf{x}\right]^{-1}\end{aligned} \tag{4.26}$$

Using the fact that $I_0 = \mathbb{1}_{\{1-(\mathbf{x}-\mathbf{c}_0)^T\mathbf{M}_0(\mathbf{x}-\mathbf{c}_0)\geq 0\}}$ is the indicator of the ellipsoid yields

$$\mathbf{M}^* = \left[\frac{2}{\mu}\frac{1}{\mathrm{Vol}(\mathbf{M}_0)} \int_{\{1-(\mathbf{x}-\mathbf{c}_0)^T\mathbf{M}_0(\mathbf{x}-\mathbf{c}_0)\geq 0\}} \big(\mathbf{x}-\mathbf{c}_0\big)\big(\mathbf{x}-\mathbf{c}_0\big)^T\, d\mathbf{x}\right]^{-1} \tag{4.27}$$

After a variable substitution $\mathbf{x} \leftarrow \mathbf{M}_0^{1/2}(\mathbf{x}-\mathbf{c}_0)$, this expression becomes

$$\mathbf{M}^* = \left[\frac{2}{\mu}\frac{\det(\mathbf{M}_0^{-1/2})}{\mathrm{Vol}(\mathbf{M}_0)}\mathbf{M}_0^{-1/2} \int_{\{\|\mathbf{x}\|\leq 1\}} \mathbf{x}\mathbf{x}^T d\mathbf{x}\ \ \mathbf{M}_0^{-1/2}\right]^{-1} \tag{4.28}$$

With $\text{Vol}(\mathbf{M}_0) = \frac{4\pi}{3}\sqrt{\det(\mathbf{M}_0^{-1})} = \frac{4\pi}{3}\det(\mathbf{M}_0^{-1/2})$*, we then obtain*

$$\mathbf{M}^* = \left[\frac{2}{\mu}\,\frac{3}{4\pi}\,\mathbf{M}_0^{-1/2}\int_{\{\|\mathbf{x}\|\leq 1\}}\mathbf{x}\mathbf{x}^T d\mathbf{x}\;\;\mathbf{M}_0^{-1/2}\right]^{-1} \tag{4.29}$$

Note that the integral $\int_{\{\|\mathbf{x}\|\leq 1\}}\mathbf{x}\mathbf{x}^T d\mathbf{x}$ *denotes the covariance matrix of a 3D unit ball, which is actually a scalar matrix that can be easily computed*

$$\int_{\{\|\mathbf{x}\|\leq 1\}}\mathbf{x}\mathbf{x}^T d\mathbf{x} = \begin{pmatrix} 2\pi\,\frac{2}{3}\,\frac{1}{5} & 0 & 0 \\ 0 & 2\pi\,\frac{2}{3}\,\frac{1}{5} & 0 \\ 0 & 0 & 2\pi\,\frac{2}{3}\,\frac{1}{5} \end{pmatrix} = \frac{4\pi}{15}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.30}$$

Combining (4.29) *and* (4.30) *leads to*

$$\mathbf{M}^* = \left[\frac{2}{\mu}\left(\frac{1}{5}\,\mathbf{M}_0^{-1}\right)\right]^{-1} \tag{4.31}$$

which yields the following relationship between $\mathbf{M}^*$ *and* $\mathbf{M}_0$ *:*

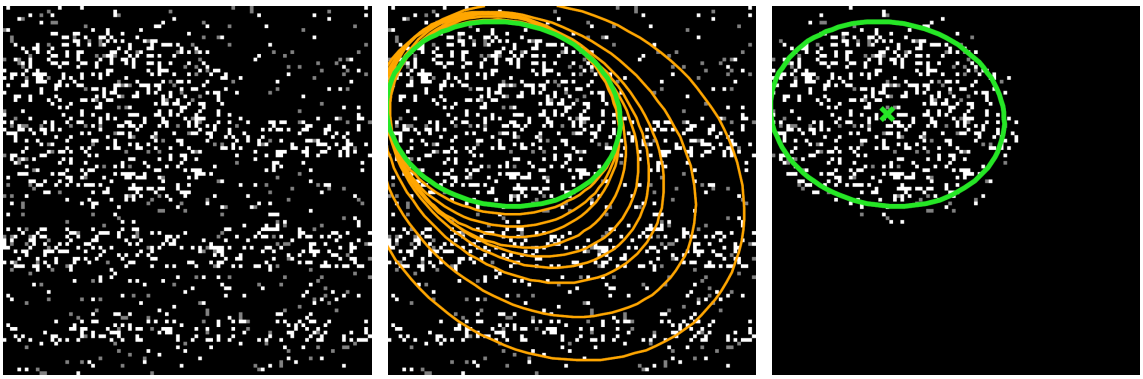
$$\mathbf{M}^* = \frac{5}{2}\,\mu\,\mathbf{M}_0 \tag{4.32}$$

This shows that the exact solution $\mathbf{M}_0$ *is retrieved for* $\mu = \frac{2}{5}$*. This value actually depends on the dimension of* Ω*. Here we assumed* $\Omega \subset \mathbb{R}^3$ *but for 2D images, the optimal value would rather be* $\mu = \frac{1}{2}$*. In practice, we choose values close to* $\frac{2}{5}$ *(or* $\frac{1}{2}$ *in 2D) depending on the image quality and in particular the level of noise.*

Figure 4.11 illustrates such a process for a synthetic 2D image. On the left part is shown the original image, which is very noisy and with confusing neighboring structures. In Figure 4.11.b, the different ellipses obtained along the iterations are displayed. The first ellipse is the largest one: all pixels contribute since $w \equiv 1$ everywhere, which results in an over-estimation. Nevertheless, far points are then progressively eliminated via the weighting function w until the algorithm converges towards the good solution. We also present results on real CEUS data in Figure 4.12. The estimated ellipsoids are not perfectly accurate but robust and close enough to be used as an initialization for a more elaborate segmentation algorithm.

4.3.3 Settings for implicit template deformation

In ultrasound images, defining a proper image-based term is more difficult than in CT because intensities are not consistent from one image to another. We will only make the assumption that the kidney is much brighter than its surrounding. This advocates for the choice of an edge-based term, therefore based on the Laplacian of the image. Yet some pre-processing steps are required.



(a) (b) (c)

Figure 4.11: (a) Original 2D synthetic image, corrupted by salt-and-pepper noise. (b) Evolution of the ellipse along the iterations (orange) and final result (green). (c) Ellipse contour and center superimposed on the product wI at convergence, which represents the pixels actually taken into account to estimate the final ellipse.

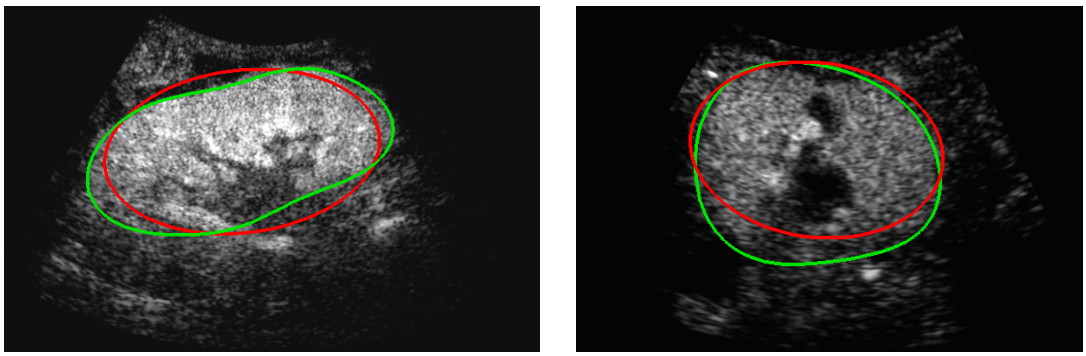
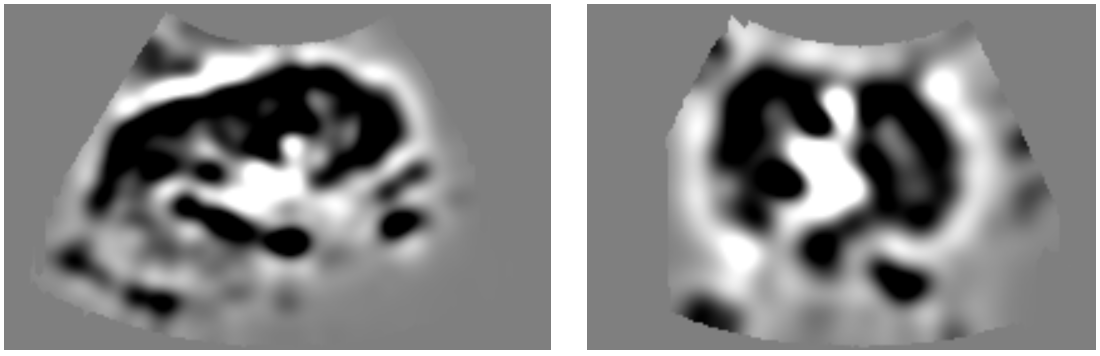


Figure 4.12: Results of the ellipsoid detection (red) compared to the ground truth (green), on slices of the volumes shown in Figure 4.10.

First, the ultrasound field of view should be detected since its boundary may create artificial edges. To do so, we use a very simple approach: on each 2D slice of the volume, we fit two symmetric lines in the left and right side of the image so that they exclude a maximum of pixels with zero intensity but no other pixel. These two lines intersect at a point that can be considered as the beam formation center. The top and bottom part of the field of view are then finally estimated as circle portions (whose center is the aforementioned point) according to the same criterion as for the lines. We denote $M : \Omega \to \{0, 1\}$ the indicator function of the detected field of view, that will be exploited in the next step.

Given the high level of noise, we then filter the image in order to retain only the most relevant edges. The smoothing operation is performed using a Gaussian kernel K_σ and also shows the benefits of diffusing the edge information far away. This diffusion is however biased near the boundaries of the field of view, as exterior pixels (with no information) are mistakingly taken into account. In order to prevent this pitfall, we use a so-called *normalized*



(a) (b)

Figure 4.13: Slices of the masked Laplacian-based image term r for the two images presented in Figure 4.10. Black (resp. white) intensities represent negative (resp. positive) values.

convolution [Knutsson & Westin, 1993]. The smoothed image, masked by M, is defined as :

$$I_{\sigma,M} = \frac{(M.I) * K_\sigma}{M * K_\sigma} \tag{4.33}$$

The division by the smoothed mask has no effect in the center of the field of view, but compensates the signal decrease near its boundary. Note that $I_{\sigma,M}$ in (4.33) is not defined far outside the mask so we simply set its value to zero. The region terms are finally defined as

$$r_{int}(\mathbf{x}) = -M(\mathbf{x})\ .\ \Delta I_{\sigma,M}(\mathbf{x}) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = 0\ . \tag{4.34}$$

Examples of such r_{int} for real data are given in Figure 4.13. With this definition and the initialization of ϕ_0 as aforementioned, we are ready to use the template deformation algorithm to get an automatic and more precise segmentation.

4.3.4 Material and results

Our database is composed of 64 CEUS volumes acquired from 35 different patients, via an iU22 ultrasound system (Philips, The Netherlands). It must be noted that the majority of the patients suffered from either renal tumors or cysts. Therefore the datasets can be considered as particularly challenging. Some healthy kidneys were also included in order to have a clinically representative database. Images were acquired using different probes, namely V6-2 and X6-1 (Philips, The Netherlands) US probes, with various fields of view. The volumes size was $512 \times 510 \times 256$ voxels with different spatial resolutions ($0.25 \times 0.25 \times 0.55$ mm in average). The acquisition protocol was as follows: first, the clinician scouted for the patient's kidney using conventional US and acquired a US volume (that will be used further in this thesis). Then, 2.4 mL of Sonovue (Bracco, Italy) contrast agent were injected to the patient and a CEUS acquisition was performed after a few seconds.

The completely automatic pipeline had a computational time of around 5 seconds on a standard computer (Intel Core i5 2.67 Ghz, 4GB RAM).

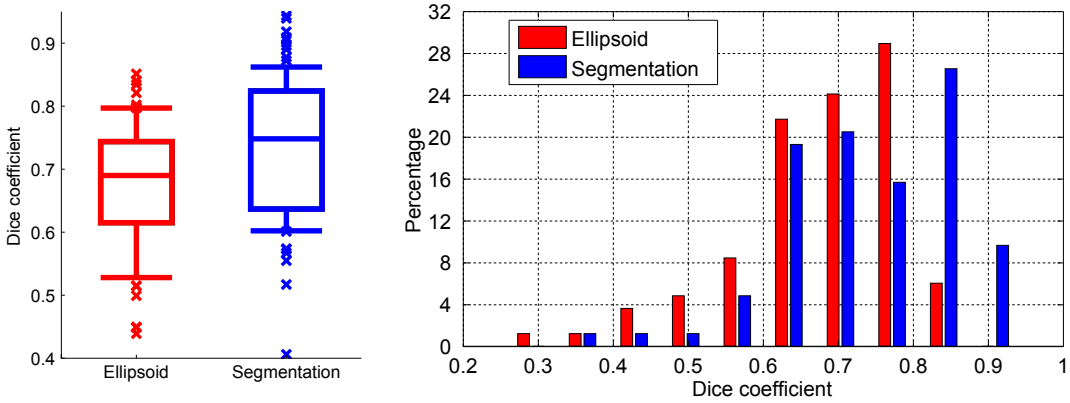


Figure 4.14: Kidney detection (red) and segmentation (blue) results in 3D CEUS images in terms of Dice coefficients shown as boxplots (left) and histograms (right). Boxplots show respectively the first decile, the first quartile, the median, the third quartile and the ninth decile. Extreme points are shown separately.

Results of automatic kidney segmentation

Quantitative results are reported in Figure 4.14. The overall median Dice coefficient is 0.69 for the detection and 0.76 for the segmentation and 25% of the database have a very satisfying segmentation (Dice coefficient higher than 0.85), given the very poor image quality and the presence of pathologies. Figure 4.15 shows the obtained result for the two cases introduced in Figure 4.10. The segmentations are very similar to the ground truth and can be considered as satisfying. Some cases are however more difficult (*e.g.* Figure 4.16) and will require additional information. Indeed ultrasound shadows or kidney pathologies makes the image information unreliable and thus hinder the segmentation algorithm. It is therefore important to provide the clinician a way to guide or correct the segmentation easily and with a real-time feedback. As recalled in Section 2.7.4, this can be done easily within the implicit template deformation framework.

Influence of user interactions on kidney segmentation in CEUS

Validation of the user interactions has been performed on a subset of 21 CEUS volumes from 21 different patients of our database. For each case, the automatic segmentation has been performed and its result was refined with user interactions from an expert. Figure 4.17 reports the Dice coefficients obtained as a function of the number of clicks. The score gradually increases as the user interacts with the algorithm and rapidly converges: most of the time, less than 3 clicks are needed for a fairly precise result (Dice ≥ 0.9). The ground truth may not exactly be reached because of the high intra-operator variability. The results also show that even when the initialization produces a low score, very few interactions can improve a lot the segmentation. The influence of user interactions is illustrated in Figure 4.16, where we show results on a difficult case. The patient has a lot of renal cysts that are anechogenic and hinders the automatic segmentation. With 3 clicks, the segmentation is much closer to

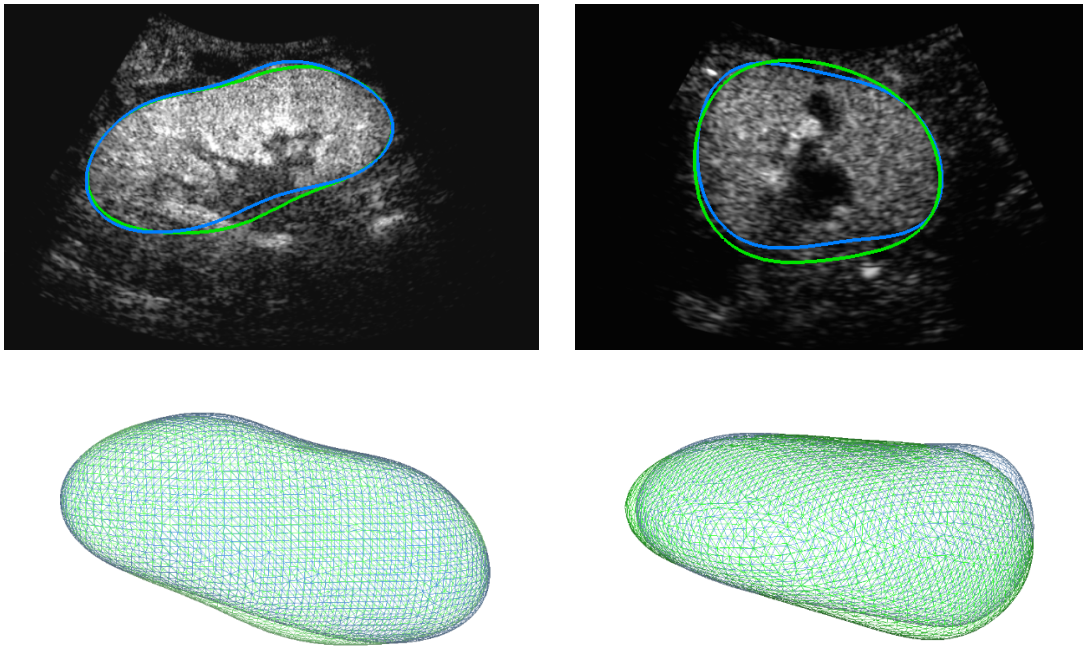


Figure 4.15: Result of the automatic segmentation (blue) compared to the ground truth (green) on a particular slice (top) and in 3D (bottom).

the ground truth.

Conclusion

In this section, we presented another application of the template deformation algorithm, this time in 3D CEUS images. In order to initialize it, we proposed a novel ellipsoid estimation method based on a robust and efficient variational approach. This yields a fully automated workflow for kidney segmentation which most of the time returns a satisfying result but may fail when the image information is too ambiguous (because of the noise, shadows or pathologies). We showed that in such cases, the segmentation can be easily corrected by an expert user in only a few clicks. In the next section, we apply a similar workflow to segment kidneys in B-mode ultrasound images.

4.4 Kidney segmentation in 3D US images

Kidney segmentation in conventional ultrasound 3D images is an important step towards the diagnosis of some diseases. It allows a much more precise estimation of the renal volume than the standard ellipsoid method (see Section 4.1.4) Since B-mode ultrasound acquisitions are more common in clinical routine, the problem of kidney segmentation in US has received more attention from the community in conventional US images compared to CEUS images.

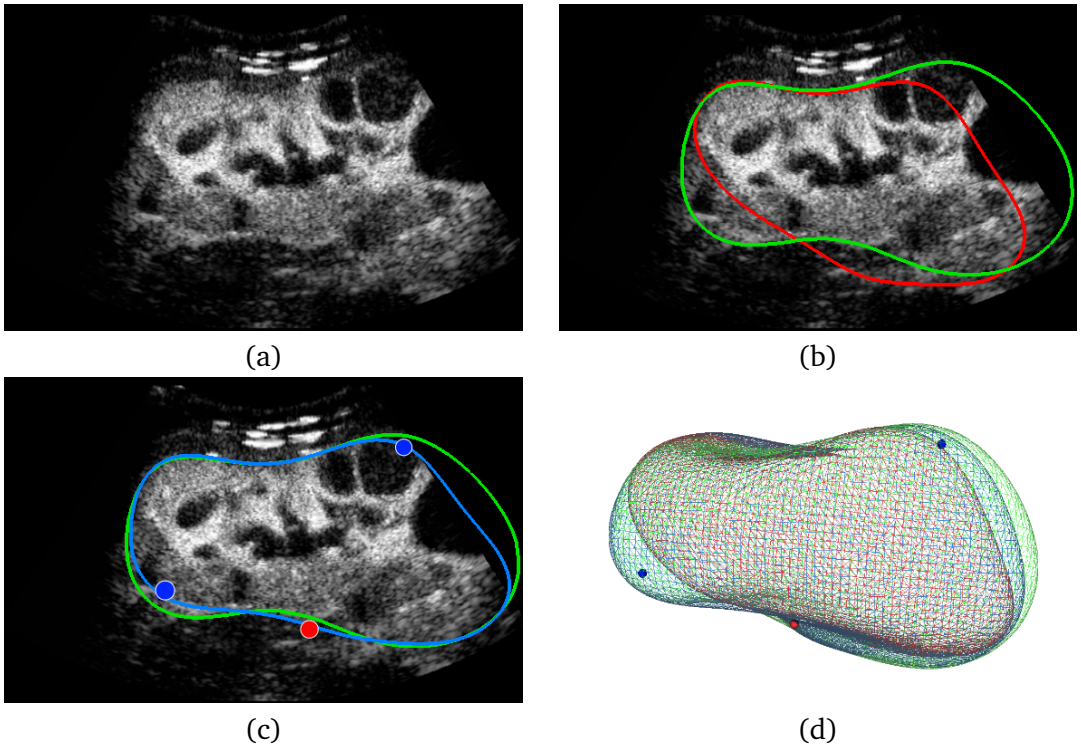


Figure 4.16: Example of a segmentation with user interactions. (a) Slice of the original CEUS volume. (b) Comparison of the ground truth (green) and automatic segmentation (red). (c) Corrected segmentation (blue) with two inside points (blue dots) and one outside point (red dot). (d) 3D visualization of the ground truth (green), the automatic (red) and corrected (blue) segmentation, with constraint points.

However, the vast majority of previous literature addressed the problem of kidney segmentation in 2D images [Xie et al., 2005; Wu & Sun, 2006; Raja et al., 2007; Booth et al., 2006; Mendoza et al., 2013] whose relevance for renal volume measurement can be questioned. Furthermore, most of them depend on user interactions, which also raises the issue of standardization and repeatability. Raja *et al.* introduced an interactive method based on higher order spline interpolation of contour points given by the user [Raja et al., 2007]. In [Xie et al., 2005], a mean kidney shape is put manually in the image and a segmentation is subsequently performed using an active appearance model [Cootes et al., 2001] on texture image features. This textural information was also exploited in [Wu & Sun, 2006], but this time followed by a MAP estimation and used within a deformable model. However it also required an initialization as an ellipsoid given by the user. Recently, Mendoza *et al.* proposed to reduce the dependency on the information provided by the user [Mendoza et al., 2013]. Indeed their method only requires the input of two points approximately at the extremities of the kidney major axis. The only attempt at providing a fully automated workflow was described in [Booth et al., 2006]. In their work, an ellipsoid is automatically fitted to the points of highest gradient magnitude. This simple rule is unfortunately not robust enough in practice, especially for 3D images, when the variability of the dataset increases.

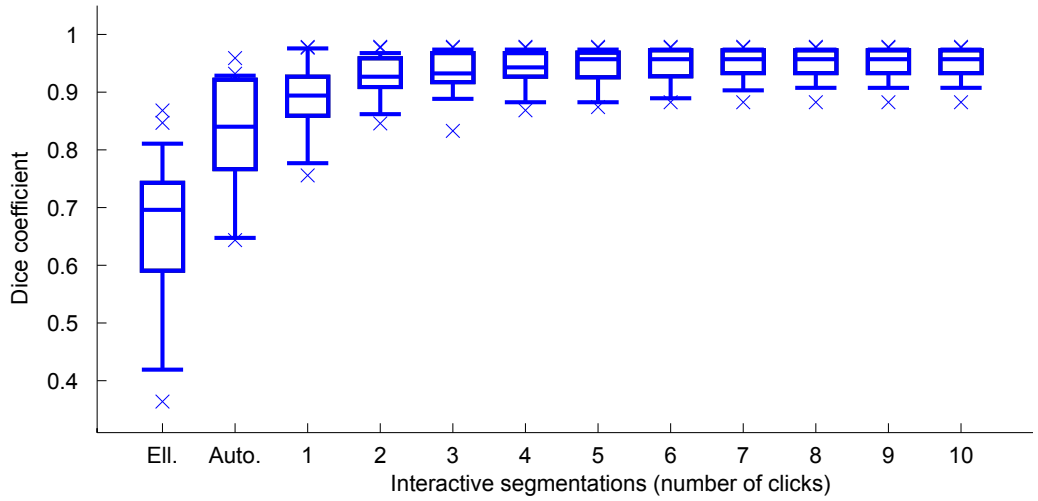


Figure 4.17: Boxplots of the Dice coefficient between the ground truth and the segmentation at different steps of the proposed algorithm. Boxplots show respectively the first decile, the first quartile, the median, the third quartile and the ninth decile. Extreme points are shown separately.

Prior work of kidney segmentation in 3D US images is limited to [Martin-Fernandez & Alberola-Lopez, 2005], in which an active contour is coupled with a Markov random field. Once again though, the initialization is a manually drawn contour.

In this section about kidney segmentation in 3D US images, we combine some ideas from both Sections 4.2 and 4.3 to build an automatic workflow. According to the review above, this represents the first proposal towards a fully automated pipeline for this application. Moreover, the two-step strategy for learning kidney appearance using the auto-context framework with random forests classifiers constitutes another contribution of this work.

4.4.1 Estimation of the image-based term using auto-context classification forests

In CEUS images, bright areas indicate the presence of contrast agent which is mainly localized in the kidney. This is why we directly used the image intensity as a voxel probability to be inside the kidney. However in conventional US images, this does not hold and we need to transform the image into a more elaborate kidney probability map as in CT.

The kidney appearance has a much higher variability in US images, although its structure is consistent: kidneys are always composed of a bright sinus surrounded by a darker parenchyma (see Figure 4.10 or 4.19). As intensity itself is not reliable enough, we chose to combine multiple image features using decision forests [Breiman, 2001] to obtain a class posterior map p_{US}, as in Section 4.2.3. Recent work [Payet & Todorovic, 2010; Montillo et al., 2011; Kontschieder et al., 2012; Glocker et al., 2012; Zikic et al., 2012] proved that

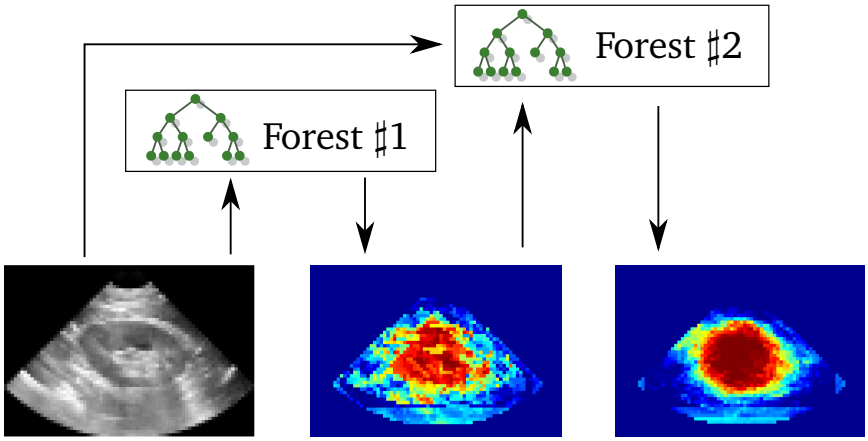


Figure 4.18: Principle of the kidney probability estimation in US images. Two random forests classifiers are chained in order to include contextual information and therefore spatial coherence.

adding contextual information allows to improve spatial consistency and thus classification performance. In our application, spatial structure is indeed paramount due to the particular structure sinus/parenchyma. These two regions have very a different appearance so learning them directly as a whole organ is difficult.

Here we propose to exploit the kidney structure in a simple yet efficient way. Similarly to the *auto-context* framework [Tu & Bai, 2010], contextual information is included using two classifiers in cascade. A first classification (kidney vs background) is performed in each voxel using a decision forest. Then we use these class posterior probabilities as additional input of a second random forest that will give the final kidney probability p_{US} (see Figure 4.18). Note that in the original auto-context paper, new such classification layers are added until the probability estimation converges. In our experiments, we restricted ourselves to two layers for practical reasons: classifying every voxel of the volume is quite computationally demanding. Furthermore, based on the obtained results for our application, we expect little further improvement from adding another classifier to the chain. Some example results of the proposed learning strategy are reported in Figures 4.18 and 4.19.

The features used for the first decision forest were the intensity of the image and its Laplacian at the considered voxel as well as at its neighbors' within a $7 \times 7 \times 7$ local patch, at three different scales ($\sigma = 2, 4, 6$ mm). Since intensities in US do not have a physical meaning, they were normalized in each patch. For the second forest, we added the estimated class posterior as additional channels. Each forest was composed of 10 trees with maximum depth 15.

4.4.2 Settings for implicit template deformation

For this application, the template ϕ_0 is set to the ellipsoid detected on the probability map p_{US} with our robust ellipsoid estimation algorithm (see Section 4.3.2).

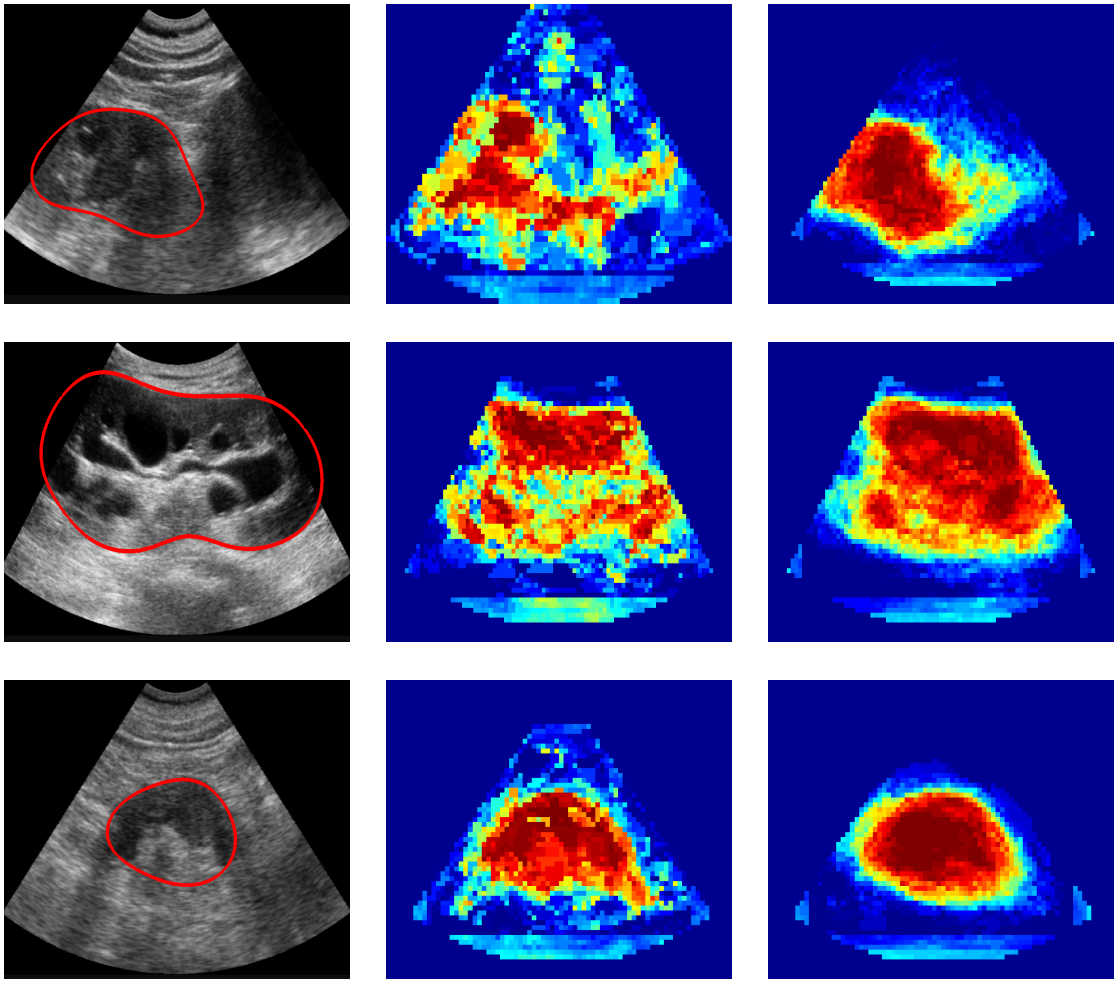


Figure 4.19: Kidney appearance in US images (the kidney boundary is denoted in red). (Left) Original images showing the high variability of the database. (Middle) Kidney probability given by the first classifier. (Right) Final kidney probablity p_{US}.

The region term for the template deformation algorithm is defined as the flux of kidney probability, as in the CT case:

$$r_{int}(\mathbf{x}) = -\Delta p_{US}(\mathbf{x}) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = 0\ . \tag{4.35}$$

Since estimated probabilities suffer from similar accuracy problems near the kidney boundaries, the last iterations are also performed on the input image: $r_{int} = \Delta I_{\sigma,M}$ (near its boundary, kidneys have darker intensities than the background). As indicated in (4.33), $I_{\sigma,M}$ denotes the image smoothed with a Gaussian kernel via a normalized convolution with M the mask of the field of view. The other settings are similar to the CEUS application.

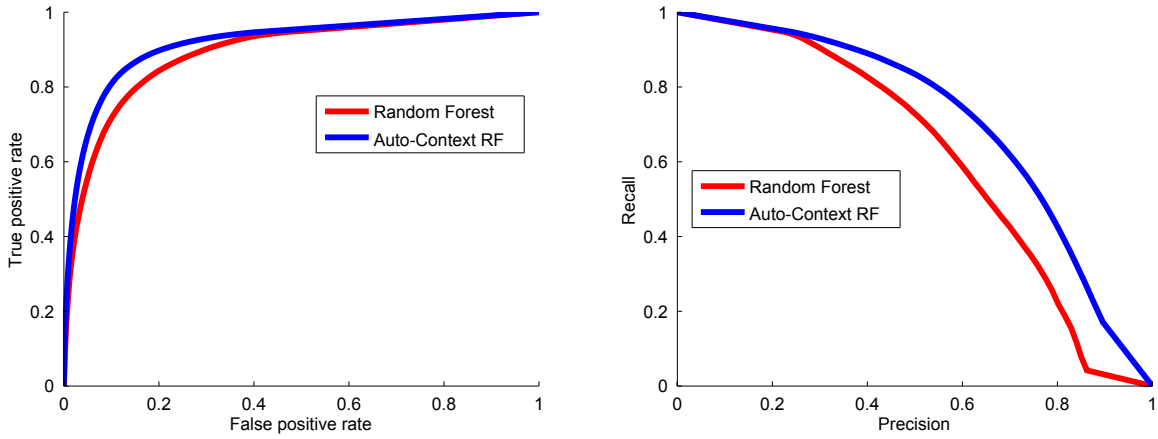


Figure 4.20: Comparison of classification results for the single decision forest and the auto-context approach. (Left) ROC Curve. (Right) Precision-Recall curve.

4.4.3 Material and results

The database used for the validation is the same as in Section 4.3.4, that is to say 64 US volumes acquired from 35 different patients. Once again, we highlight the difficulty of this database that is mostly composed of diseased patients. For all the experiments, a two-fold cross-validation strategy has been performed: the patient database was randomly split into two parts of approximately the same size. One served as a training set while the other was the testing set, and vice versa.

Compared to the workflow proposed in CEUS images, an additional step of probability computation was added. This step is performed in around 10 seconds, which raises the total computation time to roughly 15 seconds. The probability estimation is therefore the bottleneck from a computational point of view. This was expected as it requires the classification of every voxel of the image. Even if we work at a lower resolution than the native one, the images cannot be cropped as in the CT case (Section 4.2.3).

Results of kidney probability estimation

First, we present some experiments on the probability estimations to highlight the benefits of our two-step strategy. Figure 4.20 shows the ROC and Precision-Recall curves computed (i) by the first decision forest and (ii) using the auto-context approach with another forest in cascade. We also compare in Figure 4.21 the Dice coefficients obtained by merely thresholding the probability maps. The latter strategy provides better kidney probabilities with respect to all reported statistics. In particular, Dice coefficients are significantly improved, with a p-value $< 10^{-4}$ for the Wilcoxon signed-rank test [Wilcoxon, 1945]. Indeed, taking into account structural information helps for instance distinguishing the kidney sinus from the background or the parenchyma from shadows, and allows a more spatially coherent classification (see Figure 4.19).

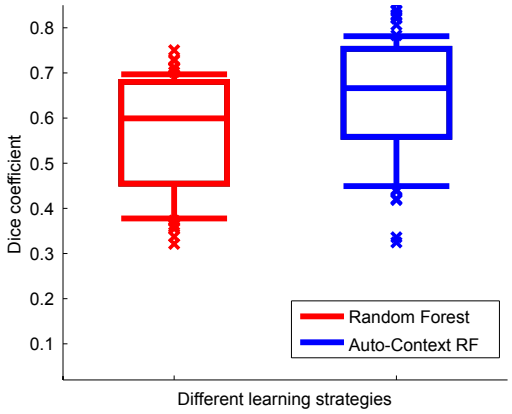


Figure 4.21: Comparison of Dice coefficients obtained by thresholding the probability maps from the single decision forest (red) and the auto-context approach (blue).

Results of kidney segmentation

Some segmentations examples are given in Figures 4.23, 4.24 and 4.25, while quantitative results are reported in Figure 4.22. The overall median Dice coefficient is 0.62 for the detection and 0.72 for the segmentation. These statistics are similar to what we reported for CEUS images (see Figure 4.14 in Section 4.3.4) but their variability is higher (standard-deviation of 0.15 in US versus 0.11 in CEUS). In particular, the first decile and quartile are much lower than in CEUS. This can be explained by the voxel classification step that sometimes mistakes the kidney with other structures, artifacts or pathologies (as in Figure 4.24, in which the segmentation is hindered by the presence of the tumour). For such cases, the ellipsoid is not correctly estimated and the subsequent segmentation is bound to fail. In a number of other challenging cases (like the one reported in Figure 4.25), our classifier shows a tendency to over-estimate the kidney, which results in detection and therefore segmentation leaks. Yet a visual inspection of many such cases reveals that the true kidney boundaries can be spotted by a human observer (which was not really the case for most failures in CEUS images). This indicates that there is still work to do in the probability estimation step to extract more relevant information. However, when the contrast is sufficient, we are able to retrieve a satisfying segmentation (see example in Figure 4.23).

Discussion

In this section, we adapted our frameworks to build a fully automatic workflow for kidney segmentation in 3D US images. To that end, we designed an original learning strategy via auto-context random forests to estimate each voxel probability to lie inside the kidney. Both the robust ellipsoid detection algorithm and the implicit template deformation are then run on this probability map. The final segmentation results obtained are somehow a bit disappointing. Failures are mainly due to the probability estimation that is still not reliable enough to provide a good ellipsoid initialization. Although we believe that some issues could be at

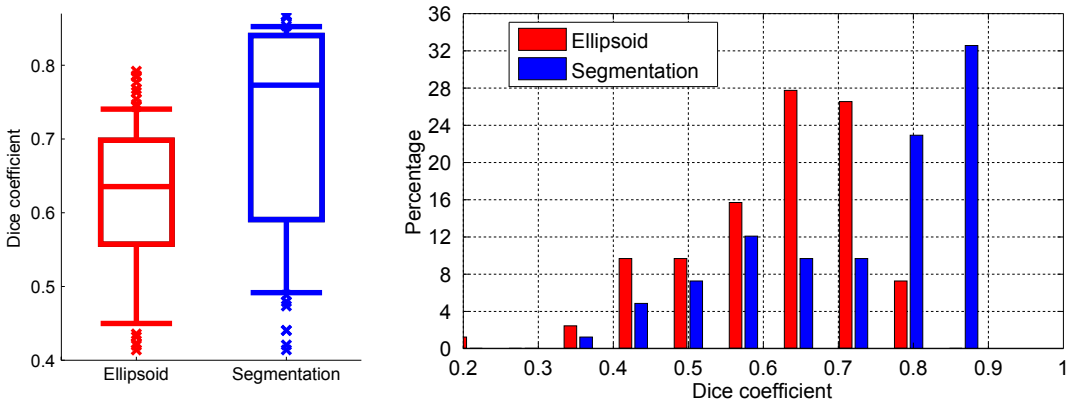


Figure 4.22: Kidney detection (red) and segmentation (blue) results in 3D US images in terms of Dice coefficients shown as boxplots (left) and histograms (right). Boxplots show respectively the first decile, the first quartile, the median, the third quartile and the ninth decile. Extreme points are shown separately.

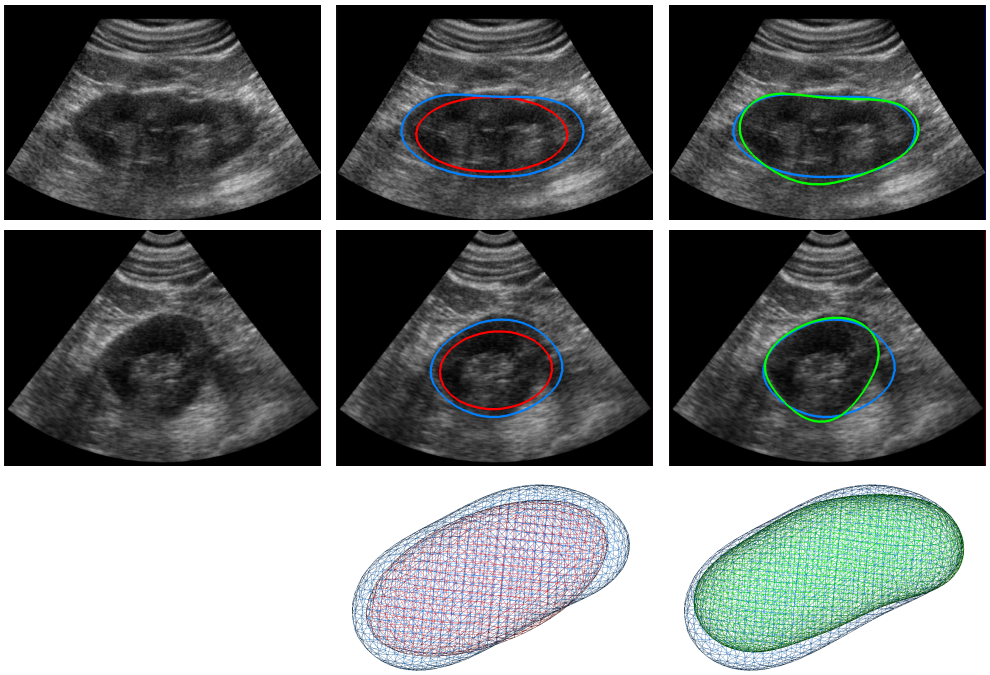


Figure 4.23: Results on a first US volume shown on two orthogonal slices (top and middle) and in 3D (bottom). The ellipsoid detection is shown (red) and the segmentation (blue) are compared to the ground truth (green).

least limited by complexing the pipeline (*e.g.* starting with a much smaller ellipsoid and using criteria on the image gradient during the segmentation), we may question the very choice of a voxelwise approach. More global approaches, such as template matching [Brunelli, 2009] or Hough transforms [Ballard, 1981; Ecabert et al., 2008], might actually be more successful

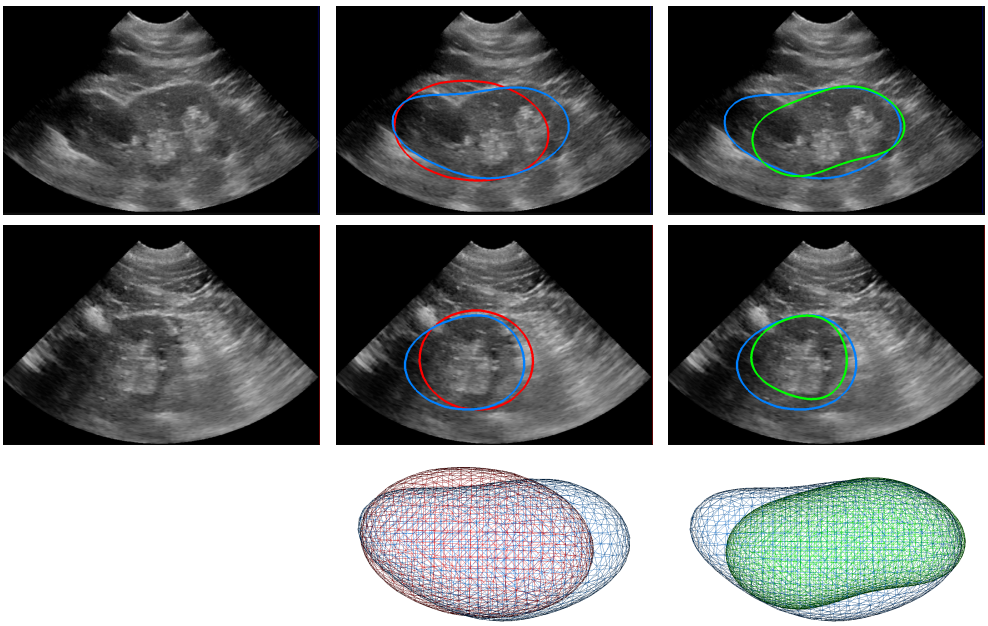


Figure 4.24: Results on a second US volume shown on two orthogonal slices (top and middle) and in 3D (bottom). The ellipsoid detection is shown (red) and the segmentation (blue) are compared to the ground truth (green).

in finding an initial ellipsoid in US images. Robust and automated kidney detection in B-mode images therefore remains an open problem that is still currently under investigation.

*
* *

Overall, we have shown the great potential of the implicit template deformation framework in the different sections of this chapter. When combined with an appropriate initialization method and image-based term, this approach achieves satisfying segmentation results even in difficult settings. In the remainder of this thesis, we will generalize and extend its formulation in various ways in order to enrich the model and incorporate even more information. For instance, the next chapter describes a generic method to exploit multiple images at the same time.

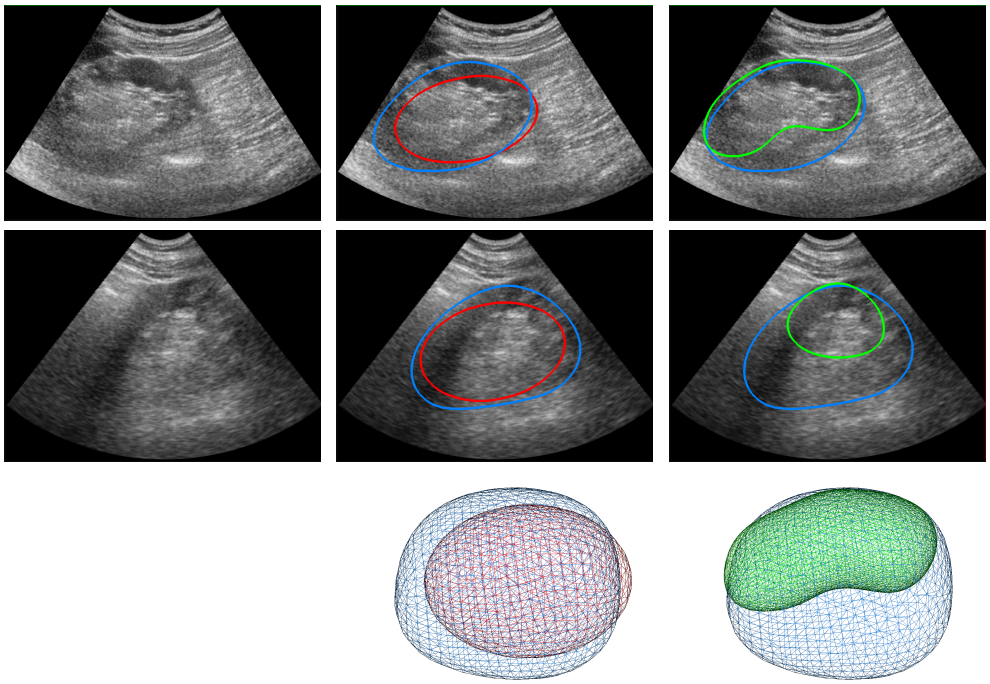


Figure 4.25: Results on a third US volume shown on two orthogonal slices (top and middle) and in 3D (bottom). The ellipsoid detection is shown (red) and the segmentation (blue) are compared to the ground truth (green).

Chapter 5

A joint co-segmentation and registration framework

Contents

Abstract

In this chapter, we propose to exploit simultaneously the information coming from multiple images into a given segmentation process. To that end, we introduce a generic framework for joint co-segmentation and registration that seeks objects having the same shape (but not necessarily the same appearance) in several images. From this framework, we derive both an ellipsoid co-detection and an implicit template co-deformation algorithm. Two clinical applications are presented: kidney segmentation in US and CEUS images, and registration of a 3D+t sequence of abdominal perfusion CT images. In both cases, our method improves on the state-of-the-art techniques.

Résumé

Dans ce chapitre, nous proposons d'utiliser simultanément l'information provenant de plusieurs images au sein d'un algorithme de segmentation donné. Pour cela, nous introduisons une approche générique pour réaliser conjointement une co-segmentation et un recalage d'images. Notre méthode consiste à chercher un objet qui a la même forme (mais pas nécessairement la même apparence) dans les différentes images. Cette idée est appliquée à la fois à l'algorithme de détection d'ellipsoïde et à la déformation de modèle implicite. Nous présentons enfin deux applications cliniques: la segmentation de rein dans des images US et CEUS, et le recalage de séquences 3D+t d'images CT de perfusion. Dans les deux cas, notre approche produit de meilleurs résultats que l'état de l'art actuel.

In the previous chapter, we presented diverse workflows for automated kidney segmentation in various modalities. Among these applications, contrast-enhanced ultrasound was the most challenging. For a number of cases, the image term was highly ambiguous (due to pathologies, artifacts or shadows), which lead to failed segmentations. In such situations, we showed that a few interactions from the clinician were sufficient to correct the result. Allowing the user to interact with the segmentation was a way to provide the algorithm with additional external information. Yet alternative sources of information, for example other images from the same patient, could also be used to improve the segmentation.

This chapter is dedicated to the development of a framework that enables the use of multiple images within a segmentation process, hence the name of *co-segmentation*. Since most of the time the images are not aligned, the problem of registration should also be addressed. We will couple the two problems to solve them simultaneously. The genericity of the proposed co-segmentation/registration approach is twofold:

- it can be applied to any variational algorithm. In Section 5.2, we present its application to two different frameworks: robust ellipsoid detection and implicit template deformation;

- it can be used in a wide variety of clinical settings. It was motivated by the application of kidney segmentation in ultrasound images (Section 5.2), in which information from both US and CEUS images is used to provide an improved segmentation. However, it can also be exploited from a different perspective, to perform sequence stabilization. This idea is presented in Section 5.3 to register a full 3D+t sequence of dynamic contrast-enhanced CT images.

Another benefit of our framework is that it most of the time yields little computational overhead thanks to its simple formulation, which is the topic of the first section of this chapter.

5.1 A generic framework for joint co-segmentation and registration

This section presents a generic approach to extend any variational segmentation algorithm to a joint co-segmentation and registration method.

5.1.1 Introduction and related work

Appearance-based or shape-based segmentation

In the computer vision community, the term of *co-segmentation* denotes the task of finding an object of interest in a collection of images. When dealing with 2D natural images (photographies), the consistency of the object's appearance (*i.e.* its color histogram) through the different images can be leveraged to obtain an unsupervised segmentation [Rubio et al., 2012; Vicente et al., 2010]. Yet objects undergo a projective transformation and a change of point of view may result in a high variation of the object silhouette: shape consistency is difficult to exploit. To circumvent this problem, Toshev *et al.* proposed to use 3D models of the target object to help its segmentation in a video sequence [Toshev et al., 2009]. They thus aim at retrieving its pose in each frame so that its projections match the segmentations. However the 3D model is considered as an initial prior and does not accurately segment the object in the images.

Conversely, in most medical applications we have access to three-dimensional data and the problem of pose observation is eluded. We can therefore give more importance to the shape of the object itself. On the other hand, we want to be able to use images from different acquisition modalities, so the structure to be segmented cannot be assumed to have the same appearance in all images. The settings are therefore very different (see Figure 5.1) and standard methods from computer vision (*e.g.* [Hochbaum & Singh, 2009]) are not adequate for clinical problems.

In 2008, Riklin *et al.* presented a shape-based mutual segmentation [Riklin-Raviv et al., 2008]. In order to segment a couple of images, they use two implicit functions ϕ_1 and ϕ_2 that are constrained to each other (up to a projective transformation T) via a pseudo-distance term in the energy:

$$\int_\Omega \left|H(\phi_1(\mathbf{x})) - H(\phi_2(T(\mathbf{x})))\right|^2 \, d\mathbf{x} \tag{5.1}$$

Since the constraint between the two implicit functions is soft, ϕ_1 and ϕ_2 do not represent the same shape. In particular, the two segmentations do not necessarily have the same topology. In most medical applications, this lack of topology consistency is undesirable. However, the idea of linking the two images via the transformation T remains interesting and was exploited in a number of registration-based co-segmentation methods.

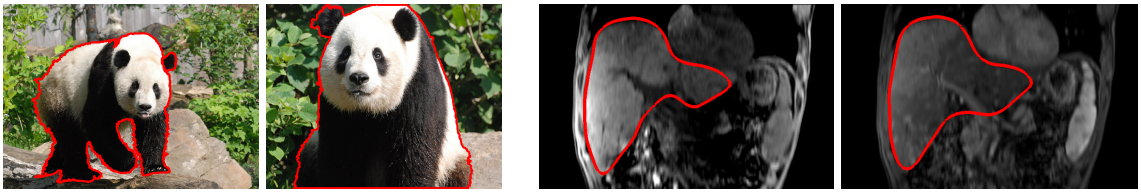


Figure 5.1: Two notions of co-segmentation : appearance consistency versus shape consistency. (Left) Co-segmentation of two different pandas with similar appearance – taken from iCoSeg dataset [CMU-Cornell, 2010]. (Right) Co-segmentation of the same liver with two different appearances – MR images acquired with different protocols.

Co-segmentation and registration

Although segmentation and registration are often seen as two separate problems, several approaches have already been proposed to perform them simultaneously. Most of them rely on an iconic registration guiding the segmentation (*e.g.* [Wang & Vemuri, 2005; Pohl et al., 2006; Lu & Duncan, 2012]). Yet they assume that the segmentation is known in one of the images, which is not the case in the applications of co-segmentation that we will consider. Moreover, in several multimodal settings, iconic registration might be bound to fail since visible structures do not always correspond to each other (for example in US/CEUS images).

Instead of registering the images themselves, Wyatt and Noble developed a *maximum-a-posteriori* formulation to perform registration on label maps resulting from a Markov random field segmentation step [Wyatt & Noble, 2002]. However no shape model is enforced and noise or misclassifications may degrade the results.

In 2003, Yezzi *et al.* introduced a variational framework to perform co-segmentation with active contours [Yezzi et al., 2003]. Two closed hypersurfaces Γ_1 and Γ_2 are sought as minima of a sum of region competition energies :

$$\left(\int_{inside\ \Gamma_1} r_{int,1}(\mathbf{x})\ d\mathbf{x} + \int_{outside\ \Gamma_1} r_{ext,1}(\mathbf{x})\ d\mathbf{x}\right) + \left(\int_{outside\ \Gamma_2} r_{int,2}(\mathbf{x})\ d\mathbf{x} + \int_{outside\ \Gamma_2} r_{ext,2}(\mathbf{x})\ d\mathbf{x}\right) \quad (5.2)$$

The two problems are coupled by the constraint $\Gamma_2 = g(\Gamma_1)$ where $g : \Omega \to \Omega$ is a global transformation. Both Γ_1 and g (which fully determines Γ_2) are unknown and simultaneously estimated. Another way of understanding the underlying idea is to consider this method as a feature-based registration in which the features are actually the segmented contours. This approach does not require the exact segmentation in one of the images as a prior and, unlike [Riklin-Raviv et al., 2008], guarantee a shape consistency between the different images. It is therefore more adapted to our needs.

The co-segmentation framework that we present in the next subsection is inspired by [Yezzi et al., 2003]. We rely on the same idea that one single shape should segment multiple images. However we apply it to different segmentation frameworks and in particular implicit approaches instead of parametric active contours. We also present it in a more generic formulation. Finally, we present various clinical applications (including multimodal segmentation or time sequences stabilization) on which the method is thoroughly evaluated.

5.1.2 A variational formulation for joint co-segmentation and registration

We consider here any variational algorithm that consists in minimizing an energy of the following form

$$E_I(\phi) = \int_\Omega f(\phi(\mathbf{x}))\, r_I(\mathbf{x})\, d\mathbf{x} + \lambda\, \mathscr{R}(\phi) \tag{5.3}$$

where f is a real-valued function and $r_I(\mathbf{x})$ denotes a pointwise score that is negative when $\mathbf{x}$ probably belongs to the target object in the image I, and positive when it does not. This is a standard setting in which the optimal implicit function ϕ must achieve a trade-off between an image-based term and a regularization term $\mathscr{R}$. For example, the seminal method of active contours without borders [Chan & Vese, 2001a] falls in this framework with $f = H$ the Heaviside function and $r_I(\mathbf{x}) = (I(\mathbf{x}) - c_{int})^2 - (I(\mathbf{x}) - c_{ext})^2$ with c_{int} and c_{ext} denoting mean intensities inside and outside the target object.

We are interested in the case where a pair of images $I_1 : \Omega \to \mathbb{R}$ and $I_2 : \Omega \to \mathbb{R}$ containing the same object are available. For instance, in medical imaging I_1 and I_2 can be two images of the same organ acquired from different modalities or at different times. If those images were perfectly aligned with respect to the target organ, the energy in (5.3) can be straightforwardly generalized to perform co-segmentation :

$$E_{I_1,I_2}(\phi) = \frac{1}{2}\int_\Omega f(\phi(\mathbf{x}))\, (r_{I_1}(\mathbf{x}) + r_{I_2}(\mathbf{x}))\, d\mathbf{x} + \lambda\, \mathscr{R}(\phi)\,. \tag{5.4}$$

Unfortunately, such an assumption never holds in medical applications[1]. A more realistic hypothesis is to assume that the target object, segmented by ϕ, is not deformed between the two acquisitions, but only undergoes an unknown global transformation G_r as in [Yezzi et al., 2003]. The co-segmentation energy thus reads:

$$E_{I_1,I_2}(\phi, G_r) = \frac{1}{2}\int_\Omega f(\phi(\mathbf{x}))\, r_{I_1}(\mathbf{x})\, d\mathbf{x} + \frac{1}{2}\int_\Omega f(\phi \circ G_r(\mathbf{x}))\, r_{I_2}(\mathbf{x})\, d\mathbf{x} + \lambda\, \mathscr{R}(\phi) \tag{5.5}$$

Note that, after a variable substitution, it can be equivalently written

$$E_{I_1,I_2}(\phi, G_r) = \frac{1}{2}\int_\Omega f(\phi(\mathbf{x}))\, (r_{I_1}(\mathbf{x}) + r_{I_2} \circ G_r^{-1}(\mathbf{x})\, |J_{G_r}|)\, d\mathbf{x} + \lambda\, \mathscr{R}(\phi)\,. \tag{5.6}$$

Minimizing E_{I_1,I_2} with respect to ϕ and G_r simultaneously can be therefore interpreted as performing jointly segmentation (via ϕ) and global registration (via G_r). We thus couple the two problems of co-segmentation and registration within a common framework, *i.e.* a single energy minimization problem. This generalizes a more common co-segmentation approach (*e.g.* [Han et al., 2011; Zagrodsky et al., 2005]) where the images are first aligned in a preprocessing step.

[1] unless for very specific modalities, such as 2D US-CEUS acquisitions (see Appendix A.2)

Remark 5.1. *Should one of the images be easier to segment than the other, multipliers can be introduced in 5.6 to weigh the contributions of r_{I_1} and r_{I_2}. However, if the difficulty gap between the two images is really large (for example using a CT image versus an US image), a probably more robust strategy would be to first segment the easy image independently and then use this result as prior to segment the more challenging one.*

Remark 5.2. *The proposed formulation is not fully symmetric since G_r depends on r_{I_2} but not on r_{I_1}. One could fix this issue using two transformations $G_{r,1}$ and $G_{r,2}$ that respectively act on I_1 and I_2, and the registering transformation would then be $G_{r,1} \circ G_{r,2}^{-1}$. However for the sake of simplicity we use only one transformation G_r, which alleviates the ambiguity that would be introduced if ϕ, $G_{r,1}$ and $G_{r,2}$ were sought simultaneously.*

Remark 5.3. *For now, we only considered two images but all equations can be generalized straightforwardly to an arbitrary number of images, as exploited in Section 5.3. If N images $(I_n)_{n=1\dots N}$ are available, we introduce N transformations $(G_n)_{n=1\dots N}$ (with $G_1 = \mathbf{Id}$) and minimize the following energy:*

$$E_{(I_n)_n}(\phi, (G_n)_n) = \frac{1}{N} \sum_{n=1}^{N} \int_\Omega f(\phi \circ G_n(\mathbf{x}))\, r_{I_n}(\mathbf{x})\, d\mathbf{x} + \lambda\, \mathscr{R}(\phi)\,. \tag{5.7}$$

with respect to ϕ and transformations $(G_n)_{n=2\dots N}$.

The choice of the space of transformations to which G_r belongs is paramount and obviously depends on the clinical application. Whether G_r is a rigid transformation, a similarity or an affine transformation will therefore be discussed later. This choice is actually a trade-off between the precision of the registration (how well can this type of transformation approximate the true motion of the structure of interest between the two images ?) and the easiness of optimization (adding more degrees of freedom increases the risk of falling into local minima and the dependency on the initialization).

We also mention that the same approach could be used with G_r as a non-global transformation that does change the shape of the object from one image to another. However this would completely decouple the two segmentations, and the benefits of co-segmentation would then be lost. Although this issue could be fixed by adding a penalty term on G_r in the energy, other problems remain. For instance, this would greatly increase the sensitivity to the initialization since the coupling between the two segmentations would be loosened. Consequently, we will not investigate further this idea and stick with affine (or more restricted) transformations.

Now that we have laid the principles of our co-segmentation framework, we will detail its various applications in the remainder of this chapter.

5.2 Ellipsoid co-estimation and template co-deformation – Application to kidney segmentation in 3D US and 3D CEUS images

This section is an extended version of the paper presented at the IPMI 2013 conference [Prevost et al., 2013b] and the filed patent [Prevost et al., 2012b]. It results from a collaboration with Prof. Jean-Michel Correas (Hôpital Necker, Paris, France) who kindly provided us with clinical data.

5.2.1 Motivation

In Section 4.3, we proposed a method to detect and segment kidneys in 3D CEUS images. While we provided an automated pipeline, failures were reported in several cases and user interactions were needed to obtain a satisfying result. Yet, because of shadowing effects, pathologies and restricted field of view, parts of the kidney may be hardly visible in the image. In such cases, even expert users may have difficulties delineating the boundary of the organ by solely relying on the CEUS images.

Additional information should be taken into account in order to improve the segmentation. A first idea would be to rely on gold-standard modalities such as CT imaging. Indeed we designed in Section 4.2 a robust and precise algorithm for automated kidney detection and segmentation in any CT image. The segmentation thus obtained could be used as a model to segment the CEUS image. However, CT imaging is not always part of the acquisition protocol, especially for screening purposes. Relying on the presence of a CT image is a strong assumption that will not hold in a large number of cases.

Conversely, in clinical routine every CEUS acquisition can be preceded by a conventional US acquisition. Indeed, before injecting the contrast agent, the clinician has to scout for the kidney with B-mode imaging. Unlike CT, we know that such an acquisition will always be available. Consequently it could be used to complement the CEUS image and thus cope with missing and corrupted information. For both US and CEUS segmentation are equally challenging, we will address them simultaneously by performing kidney co-segmentation in the two images.

As simultaneous acquisition of US and CEUS is not possible on current 3D imaging systems, the two images are acquired subsequently in arbitrary referentials and therefore need to be aligned. However standard iconic registration methods are not adapted since visible structures (apart from the kidney itself) are completely different in US and CEUS for physical reasons. Co-segmentation shall therefore help registration, just as registration helps co-segmentation.

For all these reasons, the co-segmentation approach proposed in the previous section is clearly adequate (see Figure 5.2). In the following, we will apply it to extend the two-step framework (a robust ellipsoid estimation that will be used as an initialization of the implicit template deformation) of Section 4.3 to improve the segmentation. As a side outcome, we will also get a global registration of the CEUS and US images, which may also be useful for clinical purposes.

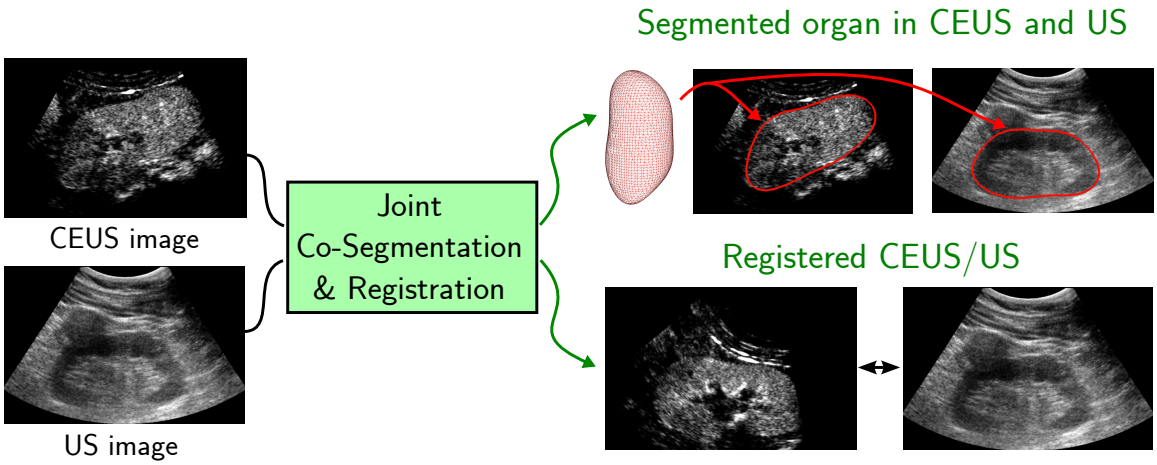


Figure 5.2: Joint co-segmentation and registration. Given a pair of non-aligned US and CEUS image of the kidney, the proposed method aims at segmenting the organ in both images as well as estimating a rigid transformation between them.

Since kidneys are surrounded by the fibrous renal capsule, they do not deform much. The transformation G_r between the US and the CEUS acquisition can therefore be assumed to be rigid.

5.2.2 Robust ellipsoid co-detection

We will first try to detect the kidney in the pair of US/CEUS images by estimating an ellipsoid that will fit in both images (up to a rigid transformation). We first recall the robust ellipsoid detection energy that was introduced in Section 4.3.2

$$\min_{\mathbf{c},\mathbf{M},w} \left\{ E_{det}(\mathbf{c},\mathbf{M},w) = -\int_\Omega \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x})\, w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x} + \mu.\log\left(\frac{\text{Vol}(\mathbf{M})}{|\Omega|}\right).\left(\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x}\right) \right\} \tag{5.8}$$

$$\text{with} \quad \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x}) = 1-(\mathbf{x}-\mathbf{c})^T\,\mathbf{M}\,(\mathbf{x}-\mathbf{c})$$

$$\text{and} \quad \text{Vol}(\mathbf{M}) = \frac{4\pi}{3}\sqrt{\det \mathbf{M}^{-1}} \quad \text{the ellipsoid volume.}$$

This formulation falls into the framework described in (5.3) with :

- $f = Id$ the identity function;
- $r_I = -wI$ the image intensities weighted by a function w to be estimated;
- $\mathscr{R}(\phi_{\mathbf{c},\mathbf{M}}) = \mathscr{R}(\mathbf{M}) = \mu.\log\left(\frac{\text{Vol}(\mathbf{M})}{|\Omega|}\right).\left(\int_\Omega w(\mathbf{x})\, I(\mathbf{x})\, d\mathbf{x}\right)$ which penalizes the volume of the ellipsoid.

Expanding this algorithm to another image I_2 requires the introduction of another weighting function w_2. Following (5.5), we can now define the co-detection energy as

$$\begin{aligned} E_{co-det}(\mathbf{c},\mathbf{M},w_1,w_2,G_r) = &-\int_\Omega \phi_{\mathbf{c},\mathbf{M}}(\mathbf{x})\; w_1(\mathbf{x})\; I_1(\mathbf{x})\; d\mathbf{x} \\ &-\int_\Omega \phi_{\mathbf{c},\mathbf{M}}\circ G_r(\mathbf{x})\; w_2(\mathbf{x})\; I_2(\mathbf{x})\; d\mathbf{x} \\ &+\mu\left(\int_\Omega \big(w_1(\mathbf{x})\; I_1(\mathbf{x}) + w_2(\mathbf{x})\; I_2(\mathbf{x})\big)\, d\mathbf{x}\right)\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right) \\ \text{with}\quad \mathrm{Vol}(\mathbf{M}) = &\frac{4\pi}{3}\sqrt{\det \mathbf{M}^{-1}} \quad \text{the ellipsoid volume.} \end{aligned} \tag{5.9}$$

To facilitate the resolution of such a problem, G_r - as a rigid transformation - can be decomposed into a rotation and a translation. We can therefore equivalently write the energy as a function of the ellipsoid center $\mathbf{c}_2$ in the second image and the rotation matrix $\mathbf{R}$:

$$\begin{aligned} E_{co-det}(\mathbf{c}_i,w_i,\mathbf{R},\mathbf{M}) = &-\int_\Omega \phi_{\mathbf{c}_1,\mathbf{M}}(\mathbf{x})\; w_1(\mathbf{x})\; I_1(\mathbf{x})\; d\mathbf{x} \\ &-\int_\Omega \phi_{\mathbf{c}_2,\mathbf{R}^T\mathbf{M}\mathbf{R}}(\mathbf{x})\; w_2(\mathbf{x})\; I_2(\mathbf{x})\; d\mathbf{x} \\ &+\mu\left(\int_\Omega \big(w_1(\mathbf{x})\; I_1(\mathbf{x}) + w_2(\mathbf{x})\; I_2(\mathbf{x})\big)\right)\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right). \end{aligned} \tag{5.10}$$

Minimization of such functional is done in an alternate three-step process:

1. The statistical interpretation still holds for the ellipsoids centers and matrix: minimizers c_1^* and c_2^* are weighted centroids while minimizer $\mathbf{M}^*$ is proportional to the inverse of the weighted covariance matrix of pixels coming from both images registered.

2. The unknown matrix $\mathbf{R}$ accounts for the possible rotation between the two images and can be parametrized by a vector of angles $\Theta \in \mathbb{R}^3$. A gradient descent is performed at each iteration to minimize the energy with respect to Θ. The energy gradient with respect to the i-th parameter of $\Theta \in \mathbb{R}^3$ reads

$$\nabla_{\Theta_i} E_{co-det} = \int_\Omega (\mathbf{x}-\mathbf{c}_2)^T \left(\frac{\partial \mathbf{R}}{\partial \Theta_i}^T \mathbf{M}\mathbf{R} + \mathbf{R}^T\mathbf{M}\frac{\partial \mathbf{R}}{\partial \Theta_i}\right)(\mathbf{x}-\mathbf{c}_2)\; w_2(\mathbf{x})\; I_2(\mathbf{x})\; d\mathbf{x} \tag{5.11}$$

3. The weights w_1 and w_2 are finally updated as indicator functions (up to a slight dilation) of the current ellipsoid estimates in each image.

The complete minimization strategy is summarized in Algorithm 3. This algorithm is computationally efficient : closed-form solutions are available (except for $\mathbf{R}$) and the process, though iterative, usually converges in very few iterations.

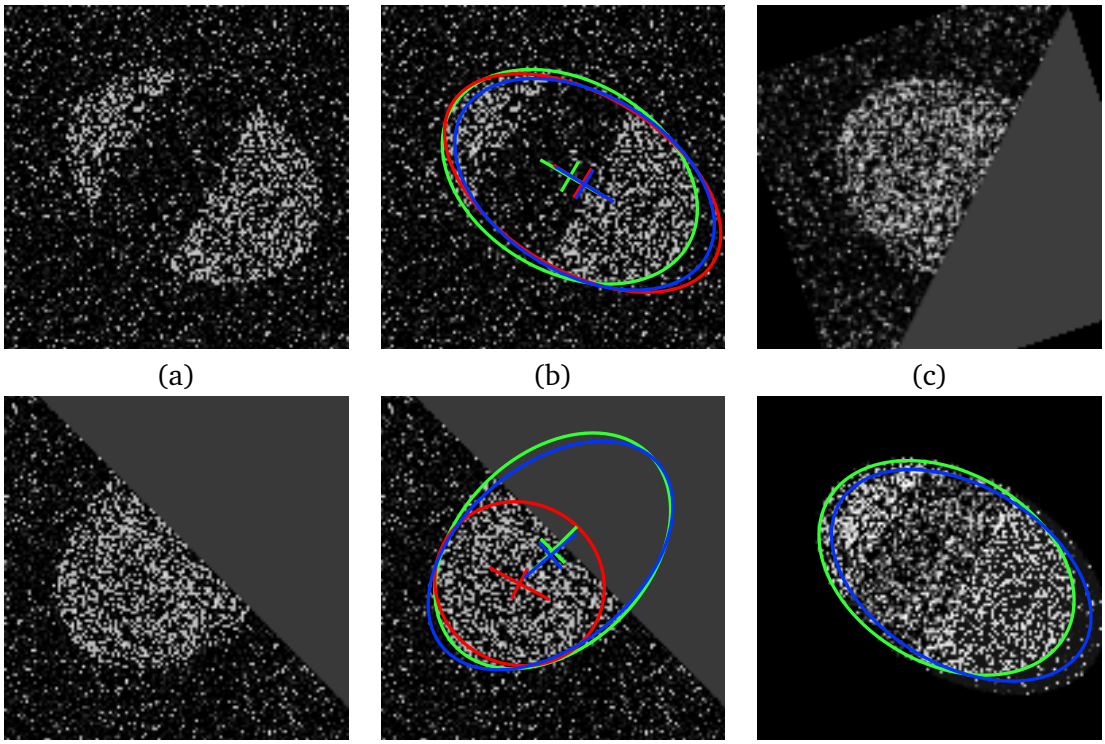


Figure 5.3: Ellipse detection on two synthetic images I_1 (a) and I_2 (d). Detected ellipses with their center and main axes are shown in (b) and (e) for independent ellipse detection (red) and proposed method for co-detection (blue), compared to the ground truth (green). (c) Second image registered with the estimated transform G_r^{-1}. (f) Combination of image terms $w_1 I_1 + (w_2 I_2) \circ G_r^{-1}$ used for ellipse estimation at convergence.

Figure 5.3 shows an example of ellipse co-detection in synthetic images, where the probability of belonging to the target object is the image intensity. Despite the noise, the simulated shadow and the reduced field-of-view effect, the co-detection algorithm provides a good estimate on the ellipse position, size and orientation in both images. Conversely, an independent detection would slightly overestimate the ellipsoid size in the first image and be completely mistaken (for both its size and orientation) in the second one.

5.2.3 Kidney co-segmentation with template deformation

Implicit template deformation, as previously described in Section 2.7, consists in minimizing the following energy:

$$\min_{L,G} \left\{ E_{seg}(L,G) = \int_\Omega H(\phi_0 \circ L \circ G)\, r(\mathbf{x})\, d\mathbf{x} + \frac{\lambda}{2} \|L - \mathbf{Id}\|_U^2 \right\} \tag{5.12}$$

Consequently, it is also part of the framework defined in (5.3) with :

Algorithm 3: Robust ellipsoid co-detection algorithm

initialization $\forall\ \mathbf{x}\in\Omega,\ w_1(\mathbf{x})\leftarrow 1,\ w_2(\mathbf{x})\leftarrow 1,\ \mathbf{R}\leftarrow \mathbf{Id}$

repeat

// Estimation of centers $\mathbf{c}_1$ and $\mathbf{c}_2$ and matrix $\mathbf{M}$

$$\mathbf{c}_1 \leftarrow \frac{1}{\int_\Omega w_1 I_1}\int_\Omega w_1(\mathbf{x})\ I_1(\mathbf{x})\ \mathbf{x}\ d\mathbf{x}$$

$$\mathbf{c}_2 \leftarrow \frac{1}{\int_\Omega w_2 I_2}\int_\Omega w_2(\mathbf{x})\ I_2(\mathbf{x})\ \mathbf{x}\ d\mathbf{x}$$

$$\mathbf{M}^{-1} \leftarrow \frac{2}{\mu\int_\Omega w_1 I_1 + w_2 I_2}\left(\int_\Omega w_1(\mathbf{x})\ I_1(\mathbf{x})\ \big(\mathbf{x}-\mathbf{c}_1\big)\big(\mathbf{x}-\mathbf{c}_1\big)^T d\mathbf{x} + \int_\Omega w_2(\mathbf{x})\ I_2(\mathbf{x})\ \mathbf{R}\big(\mathbf{x}-\mathbf{c}_2\big)\big(\mathbf{x}-\mathbf{c}_2\big)^T\mathbf{R}^T d\mathbf{x}\right)$$

// Update of the rotation matrix by gradient descent with time step Δt

repeat

$\mathbf{R}(\Theta)\leftarrow \mathbf{R}\big(\Theta-\Delta t\ \nabla_\Theta E_{co-det}(\Theta)\big)$

until *convergence*;

// Update of the weighting functions w_1 and w_2 for each $\mathbf{x}\in\Omega$

if $(\mathbf{x}-\mathbf{c})^T\mathbf{M}(\mathbf{x}-\mathbf{c})\le 1-\mu\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right)$ **then** $w_1(\mathbf{x})\leftarrow 1$ **else** $w_1(\mathbf{x})\leftarrow 0$

if $\big(\mathbf{x}-\mathbf{c}_2\big)^T\mathbf{R}^T\mathbf{M}\mathbf{R}\big(\mathbf{x}-\mathbf{c}_2\big)\le 1-\mu\log\left(\frac{\mathrm{Vol}(\mathbf{M})}{|\Omega|}\right)$ **then** $w_2(\mathbf{x})\leftarrow 1$ **else** $w_2(\mathbf{x})\leftarrow 0$

until *convergence*;

- $f=H$ the Heaviside step function;
- $r_I = r_{int} - r_{ext}$ the difference between the classification error functions;
- $\mathscr{R}(\phi_0\circ\psi)=\mathscr{R}(L)=\dfrac{1}{2}\|L-\mathbf{Id}\|_U^2$ which penalizes the amplitude and the irregularities of the deformation L.

We can therefore extend it to a co-segmentation method using (5.5) by considering the following functional

$$\begin{aligned} E_{co-seg}(\phi_0\circ L\circ G, G_r) &= E_{co-seg}(L, G, G_r)\\ E_{co-seg}(\phi_0\circ L\circ G, G_r) &= +\frac{1}{2}\int_\Omega H(\phi_0\circ L\circ G(\mathbf{x}))\ r_1(\mathbf{x})\ d\mathbf{x}\\ &\quad +\frac{1}{2}\int_\Omega H(\phi_0\circ L\circ G\circ G_r(\mathbf{x}))\ r_2(\mathbf{x})\ d\mathbf{x}\\ &\quad +\frac{\lambda}{2}\|L-Id\|_U^2\ . \end{aligned} \tag{5.13}$$

The energy E_{co-seg} is then minimized with respect to $\mathbf{p}$ (resp. $\mathbf{p}_r$) the parameters of G (resp. G_r) and each component of the vector field[1] $\mathbf{u}$, through a gradient descent with the following expressions:

$$\begin{cases} \nabla_{\mathbf{p}_{ri}} E_{co-seg} &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\; r_2 \circ G_r^{-1} \circ G^{-1}\; |J_{G_r^{-1}}||J_{G^{-1}}| \\ & \qquad \left\langle \nabla\phi_0 \circ L, (\mathbf{Id}+\mathbf{J_u})\; G \circ \frac{\partial G_r}{\partial \mathbf{p}_{ri}} \circ G_r^{-1} \circ G^{-1} \right\rangle \\ \nabla_{\mathbf{p}_i} E_{co-seg} &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\; |J_{G^{-1}}| \left\langle \nabla\phi_0 \circ L,\; (\mathbf{Id}+\mathbf{J_u})\; \frac{\partial G}{\partial \mathbf{p}_i} \circ G^{-1}\; r_1 \circ G^{-1} \right. \\ & \qquad \left. + (\mathbf{Id}+\mathbf{J_u})\; \frac{\partial G}{\partial \mathbf{p}_i} \circ G^{-1}\; r_2 \circ G_r^{-1} \circ G^{-1}\; |J_{G_r^{-1}}| \right\rangle \\ \nabla_{\mathbf{u}} E_{co-seg} &= K_\sigma * \left[\delta(\phi_0 \circ L)\; \nabla\phi_0 \circ L\; \left(r_1 \circ G^{-1} + r_2 \circ G_r^{-1} \circ G^{-1} |J_{G_r^{-1}}| \right) |J_{G^{-1}}| \right] \\ & \qquad + \lambda \mathbf{u} \end{cases} \tag{5.14}$$

These equations are very similar to the standard implicit template deformation (see Section 2.7.3). The computational overhead for segmentation is linear with respect to the number of considered images. Registering transformation G_r only depends on the zero level-set of the implicit function and does not require computations over the whole images.

Results on synthetic images are reported in Figures 5.4 and 5.5. Figure 5.4 shows that co-segmentation allows to be more robust to ambiguous image terms than independent segmentation, even if the two images are not initially aligned and the segmentation initialization (arbitrarily chosen here) is not perfect. The estimated registration between the two images is presented in Figure 5.5. One can see that the sum of the two registered images is far less ambiguous, which explains why our co-segmentation approach works better.

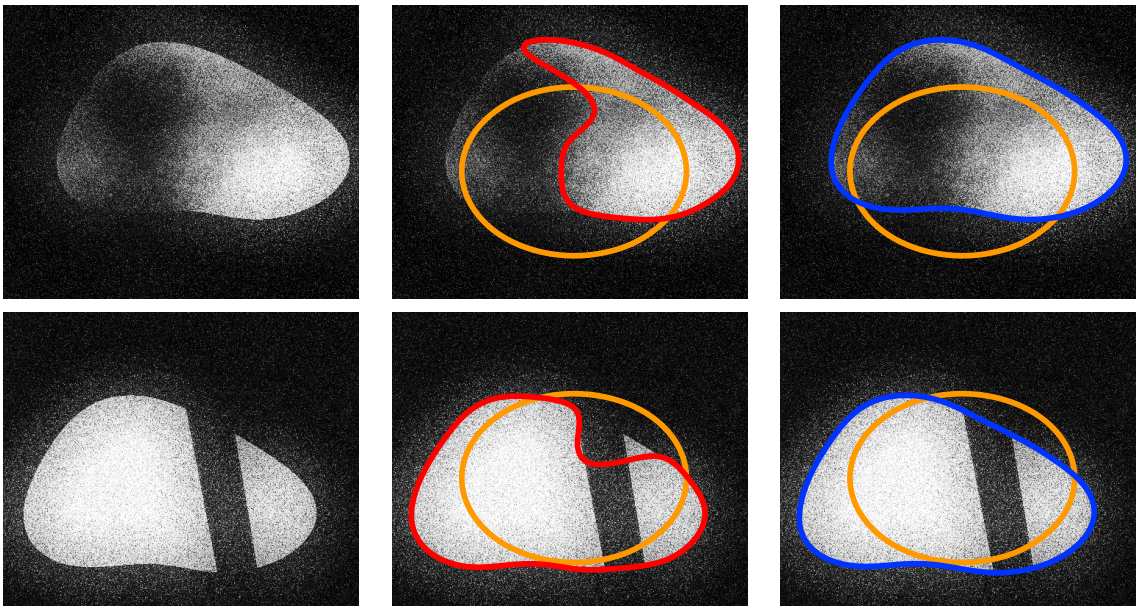
In the application of kidney segmentation in US and CEUS images, the template ϕ_0 is defined as the implicit representation of the detected ellipsoid $\phi_{\mathbf{c}_1,\mathbf{M}}$. G and L are initially set to the identity while G_r is initialized with the previously estimated registering transformation: $G_r(\mathbf{x}) = \mathbf{R}\,(\mathbf{x}+\mathbf{c}_1-\mathbf{c}_2)$.

5.2.4 Material and results

Our database is composed of 64 pairs of CEUS and US volumes acquired from 35 different patients (more details were given in Section 4.3.4). The proposed method was implemented in C++ and the average overall computational time was around 20 seconds on a standard computer (Intel Core i5 2.67 Ghz, 4GB RAM). Note that performing the segmentation independently in the two images would have taken roughly the same time.

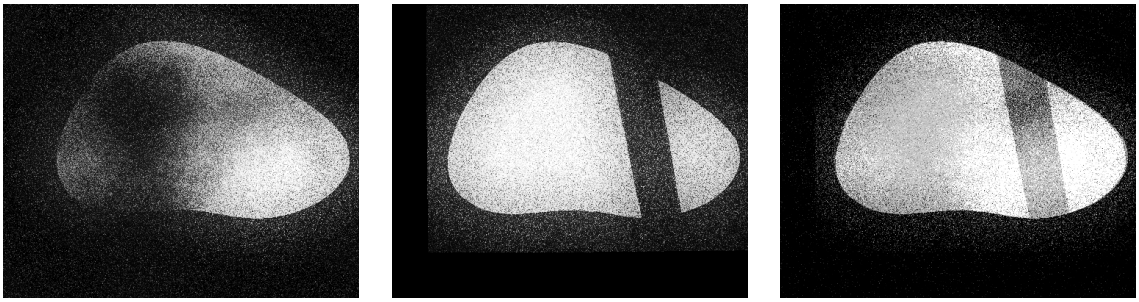
The choice of the image-based terms is exactly the same as in Sections 4.3 and 4.4, *i.e.*

[1] Recall that the deformation is defined as $L = \mathbf{Id} + \mathbf{u}$.



(a) (b) (c)

Figure 5.4: Co-segmentation of two synthetic images (top/bottom) with implicit template deformation. (a) Original images. (b) Independent segmentation (red) compared to the arbitrary initialization (yellow). (c) Co-segmentation (blue) compared to the initialization (yellow).

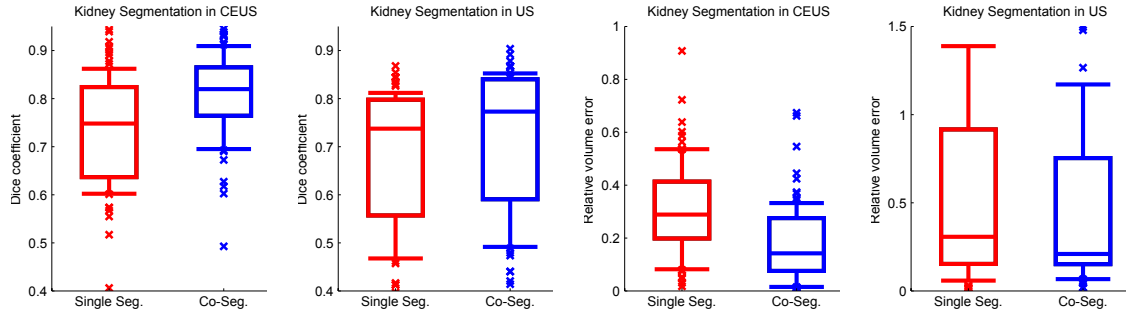


(a) (b) (c)

Figure 5.5: Registration of the two images from Figure 5.4 estimated by co-segmentation. (a) First original image. (b) Second image, registered with transformation G_r. (c) Sum of the two registered images $I_1 + I_2 \circ G_r^{-1}$.

- **kidney detection** we set $r_{I_{CEUS}} = -w_1 \, I_{CEUS}$ for CEUS images and $r_{I_{US}} = -w_2 \, p_{US}$ for US images.
- **kidney segmentation** we set $r_{I_{CEUS}} = -M \, . \, \Delta \, I_{CEUS,\sigma,M}$ for the CEUS images and $r_{I_{US}} = -\Delta p_{US}$ for US images, where M is the mask of the detected field of view in each image. Following the strategy in Section 4.4, we perform a refinement step at the end by executing the algorithm a second time with $r_{I_{US}} = M \, . \, \Delta I_{US,\sigma,M}$.

In all CEUS/US couples, we performed (i) co-detection then co-segmentation and (ii) independent detection then segmentation in each image (see Chapter 4). Validation was



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

Figure 5.6: Boxplots of segmentation results for kidney segmentation in US and CEUS images, in terms of Dice coefficients (a-b) and relative volume error (c-d). The proposed co-segmentation compares favorably to independent segmentation, with a p-value $< 10^{-4}$ for a Wilcoxon signed-rank test.

performed by comparing the segmentation results to the ground truth in both US and CEUS images. Dice coefficients and relative errors on the measured kidney volume are reported in Figure 5.6. Using simultaneously the complementary information from US and CEUS images significantly improves the segmentation accuracy in both modalities. More specifically, the median Dice coefficient is increased from 0.74 to 0.81 in CEUS (p-value $< 10^{-4}$ for a Wilcoxon signed-rank test [Wilcoxon, 1945]) and 0.73 to 0.78 in US (p-value $< 10^{-4}$). Furthermore, the proposed approach provides more reliable clinical information as the median error on the kidney volume is almost divided by two in CEUS (29% versus 15%) and in US (25% versus 13%). Figure 5.7 shows the joint co-segmentation and registration results for one case. Independent segmentation fails in both US and CEUS images because of the kidney lesion (indicated by the yellow arrow), that looks like the background in CEUS but like the kidney in US. Conversely, the proposed co-segmentation manages to overcome this difficulty by combining information from the two modalities. Furthermore, for this example, one can assess the estimated registration by comparing the location of the lesion in the two modalities. Results on another case were also displayed in Figure 5.2.

The results prove that using multiple images allows to improve the segmentation by implicit template deformation. In different clinical settings, we will see that the co-segmentation framework can have other applications, as described in the next section.

5.3 Sequence stabilization via kidney co-segmentation – Application to the registration of free-breathing 3D+t abdominal DCE-CT sequences

This section is an extended version of the paper presented at MICCAI 2013 conference [Prevost et al., 2013c]. It results from a collaboration with Prof. Olivier Lucidarme (Hôpital La Pitié-Salpétrière, Paris, France) who kindly provided us with clinical data.

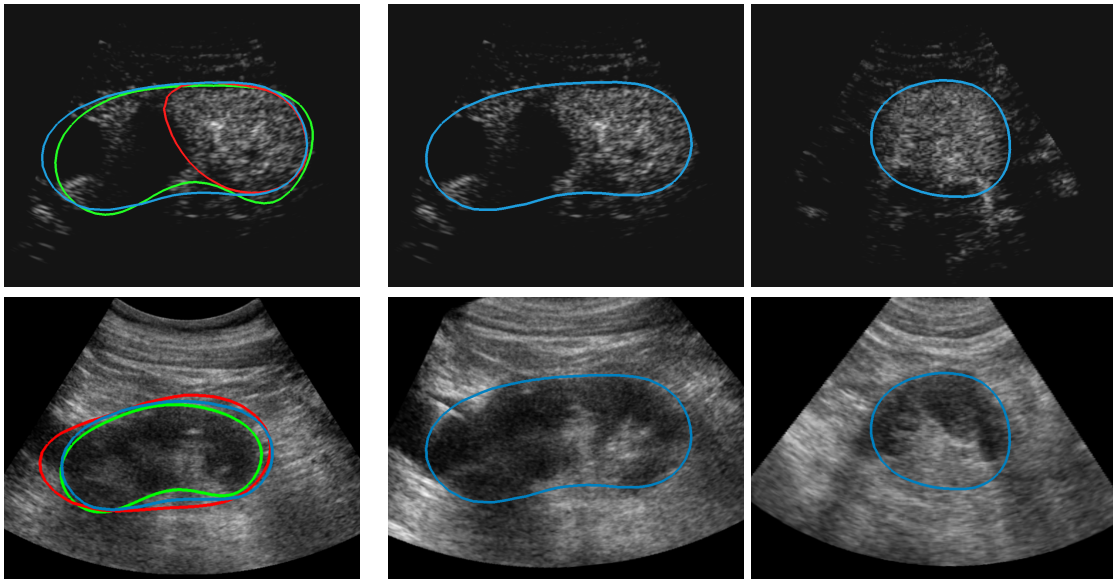


Figure 5.7: Example of joint co-segmentation and registration for a CEUS (top) and a US (bottom) images. (Left) Comparison of independent segmentations (red) and the proposed co-segmentation (blue) with respect to the ground truths (green). (Middle, Right) Two views of the registered volumes that can be assessed by considering the position of the lesion (yellow arrow).

5.3.1 Introduction

Clinical context

Dynamic contrast-enhanced (DCE) or perfusion imaging consists in acquiring a time sequence of images after a contrast agent injection. The diffusion of this contrast agent implies temporal changes in tissue attenuation and thus in image intensities. These variations, interpreted through pharmacokinetic model fitting, yield valuable information to assess the tissues functioning [Kambadakone & Sahani, 2009]. This technique is particularly used for oncologic applications, such as renal tumours follow-up, as the estimated parameters yield valuable information on healthy tissues and lesions.

Perfusion images can be acquired by MRI or CT systems. Here, we focus on DCE-CT sequences as it presents several advantages over DCE-MRI [Kambadakone & Sahani, 2009]. First, there is a linear relation between contrast agent concentration and image intensities (Hounsfield units), which simplifies pharmacokinetic models fitting. Second, CT is a cheaper and more widespread modality than MRI. However, this modality generates ionizing radiations that may harm the patient. In order to limit those risks, such acquisitions are performed in a very limited field of view with a reduced dose and a low frame-rate. This results in small and low-resolution volumes (both spatially and temporally) that are noisier than static CT acquisitions (see examples in Figure 5.8).

To capture the full dynamics of the contrast diffusion, a perfusion protocol lasts several minutes. Because of the patient's breathing, the volumes of the sequence are not aligned. A

given voxel therefore corresponds to different anatomical locations across the sequence. The major challenge in parameter estimation is therefore to design a robust registration method, that cannot use temporal consistency because of the low frame-rate.

This problem can be eluded through gated acquisitions [Li et al., 2012] or breath-holding protocols [Koh et al., 2009] but in both cases, this reduces the number of available acquisitions and hinders the pharmacokinetic parameters estimation. We aim at simplifying the protocol by letting patients breathe freely, which requires a registration of the whole sequence.

Related work

Regarding free-breathing DCE-CT registration, previous work is limited - to the best of our knowledge - to [Romain et al., 2012] in which a registration by block-matching is proposed, with a modified entropy-based similarity measure. This shall be considered as the baseline method.

Yet several methods have already been proposed to register DCE-MR sequences. Non-rigid iconic registration methods [Sance et al., 2007; Linguraru et al., 2009; Zöllner et al., 2009] are computationally demanding and rely on too specific similarity criteria. Standard choices such as mutual information are not effective in DCE-CT sequences [Romain et al., 2012]. Bhushan *et al.* used the pharmacokinetic model fitting error as registration criterion for DCE-MR, thus coupling the two tasks of sequence stabilization and parameter estimation [Bhushan et al., 2011]. However, the latter is a highly non-convex problem: including additional unknowns (namely the pharmacokinetic parameters) increases even further the dependency to the initialization.

Some methods rather use a segmentation of the organ of interest to guide the registration [Sun et al., 2004; Song et al., 2006; El-Baz et al., 2006]. Because of contrast diffusion, edge information is indeed more robust than region-based terms. In all such previous works though, segmentation and registration processes are performed sequentially. Yet they are inter-dependent (and equally challenging in DCE-CT images). Here, we propose a method to address both tasks simultaneously via co-segmentation.

Indeed, in this application, we are interested in compensating the motion only in the neighborhood of the kidney, *i.e.* a region of interest that includes the renal tumour. Our main assumption will be that the motion of this region of interest is the same as the kidney's. Since the field of view of the acquired image is very limited, the kidney does not entirely lie in one single acquisition. This advocates for our co-segmentation approach which is capable of exploiting the information from multiple images.

5.3.2 Sequence stabilization by kidney co-segmentation

Multi-images implicit template deformation

As summarized in Figure 5.8, we will use the template co-deformation framework to segment the kidney in every frame. This step can somehow be considered as a novel model-based tracking algorithm. The estimated transformations from the common shape to each image

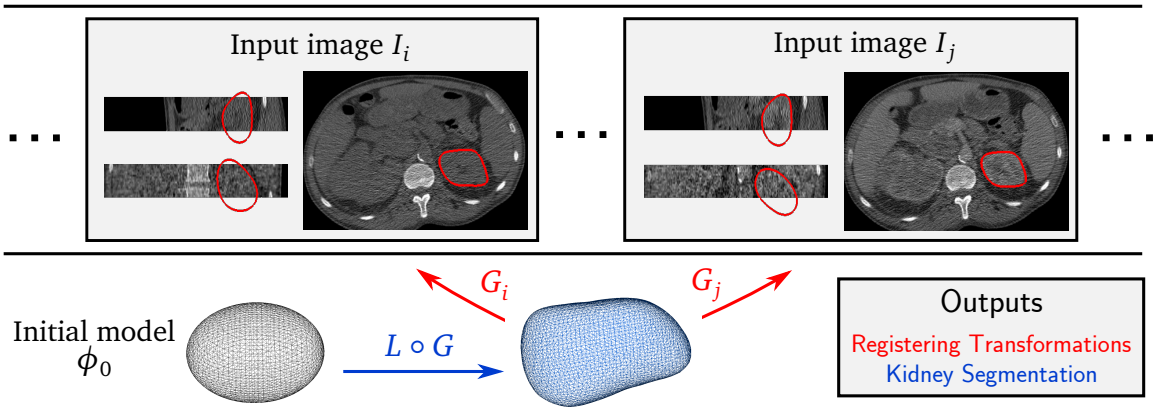


Figure 5.8: Tracking via co-segmentation of the kidney in a sequence of volumes $(I_i)_i$. The segmentation is performed by template deformation $(\phi_0 \circ L \circ G)$ using the whole sequence, while stabilizing transformations $(G_i)_i$ are simultaneously estimated.

will then be used to register all frames (*i.e.* stabilize the sequence) into a reference.

The optimization problem to solve reads

$$E_{co-seg}(L, G, (G_i)_i) = \frac{1}{N}\sum_{i=1}^{N}\int_{\Omega} H(\phi_0 \circ L \circ G \circ G_i(\mathbf{x}))\; r_i(\mathbf{x})\; d\mathbf{x} + \frac{\lambda}{2}\|L - Id\|_U^2 \;. \tag{5.15}$$

where $G_1 = \mathbf{Id}$ and the other G_i are additional unknowns. Minimization of (5.15) is performed by gradient descent, through the following equations:

$$\begin{cases} \nabla_{\mathbf{p}_i} E_{co-seg} &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\; r_i \circ G_i^{-1} \circ G^{-1}\; |J_{G_i^{-1}}||J_{G^{-1}}| \\ & \qquad \left\langle \nabla\phi_0 \circ L, (\mathbf{Id} + \mathbf{J_u})\; G \circ \dfrac{\partial G_i}{\partial \mathbf{p}_i} \circ G_i^{-1} \circ G^{-1} \right\rangle \\ \nabla_{\mathbf{p}} E_{co-seg} &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L)\; |J_{G^{-1}}| \\ & \qquad \left\langle \nabla\phi_0 \circ L, \sum_{i=1}^{N} (\mathbf{Id} + \mathbf{J_u})\; \dfrac{\partial G}{\partial \mathbf{p}} \circ G^{-1}\; r_i \circ G_i^{-1} \circ G^{-1}\; |J_{G_i^{-1}}| \right\rangle \\ \nabla_{\mathbf{u}} E_{co-seg} &= K_\sigma * \left[\delta(\phi_0 \circ L)\; \nabla\phi_0 \circ L \left(\sum_{i=1}^{N} r_i \circ G_i^{-1} \circ G^{-1} |J_{G_i^{-1}}| \right) |J_{G^{-1}}| \right] + \lambda \mathbf{u} \end{cases} \tag{5.16}$$

with the deformation defined as $L = Id + \mathbf{u}$. The model ϕ_0 was set to an ellipsoid inside the kidney in the first frame. The ellipsoid sizes and orientations were fixed in advance and we therefore only required a click inside the kidney in the first frame. We believe however that even this single interaction could be avoided by automating the kidney detection (*e.g.* [Cuingnet et al., 2012]). All registering transformations $(G_i)_i$ were initialized to the identity: no pre-registration was performed. The choice of the image-based term r_i

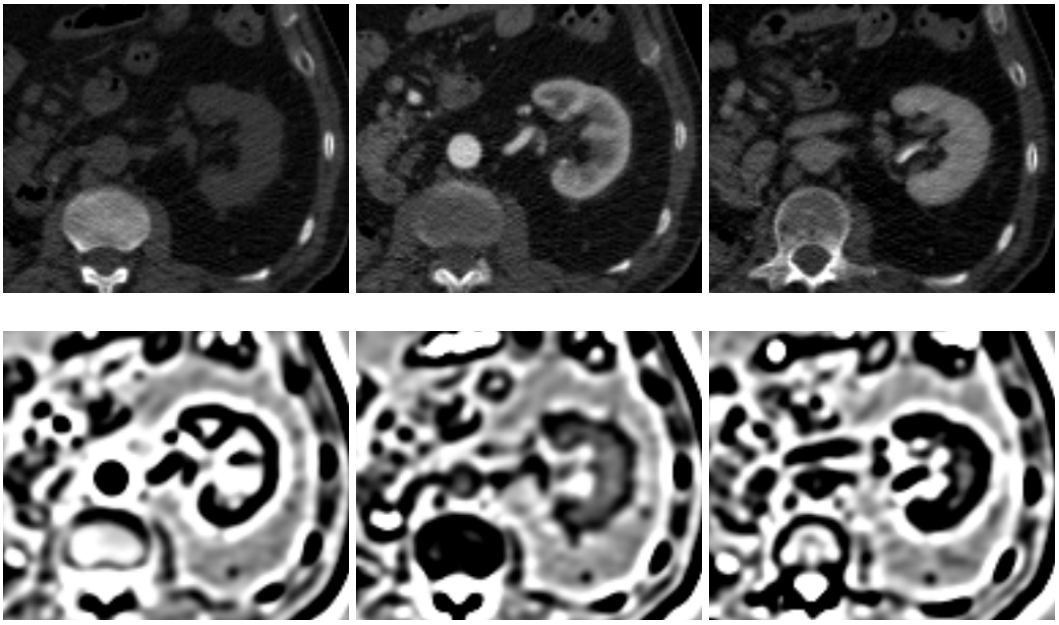


Figure 5.9: 2D crops of a sequence at three different acquisition times: original images I_i (top) and Laplacian-based image terms r_i (bottom). Regardless of the contrast phase, r_i is negative in the kidney inner boundary and positive outside.

corresponding the i^{th} image is detailed hereinafter.

Choice of the image-based term

The choice of the image-based terms $(r_i)_i$ is paramount for the segmentation. A common choice, suggested by maximum likelihood principles, is $r_i(\mathbf{x}) = \log \frac{p_{ext}(I_i(\mathbf{x}))}{p_{int}(I_i(\mathbf{x}))}$ if intensities distributions are known inside (p_{int}) and outside (p_{ext}) the target object. However image intensities vary along the sequence because of contrast agent injection and kidney's appearance may change even between two successive acquisitions. It is therefore not robust to use such intensity models. We rather rely on edge information by only assuming that in every image I_i of the sequence, the kidney is brighter than its surrounding. This assumption is based on the fact that kidneys are highly vascularized organs whose contrast uptake is very early. The selected criterion to be minimized is the image gradient flux through the segmentation surface (see Remark 2.1 from Section 2.4.3) in image I_i

$$r_{i,int}(\mathbf{x}) = -\Delta I_{i,\sigma}(\mathbf{x}) \qquad \text{and} \qquad r_{i,ext}(\mathbf{x}) = 0\ . \tag{5.17}$$

Figure 5.9 shows this term for various images of a sequence, with σ set to 3mm. After saturation, one can see that it describes effectively and consistently the interior of the kidney, regardless of the contrast phase.

5.3.3 Material and results

The experiments are based on 15 3D+t sequences coming from six different patients with renal tumours. The data were acquired on a Brilliance iCT 256 (Philips Healthcare, The Netherlands) scanner. For each patient, a dynamic CT protocol of perfusion was used. 66 volumes were acquired per sequence (48 volumes every 2.5 seconds then 18 volumes every 10 seconds). The patients were asked to breathe normally during the whole exam. Typical image size was $512 \times 512 \times 22$ voxels with a spatial resolution of $0.68 \times 0.68 \times 2.5$ mm^3. Our algorithm, implemented in C++, processes a whole 3D+t sequence in 30 seconds on a standard computer (Intel Core i5 2.67 GHz with 4GB RAM). Such a small computational time is possible because the registration is driven by the segmentation and thus only requires computations on the boundary of the segmented organ.

Evaluation of registration

An example of a sequence before and after registration is given in Figure 5.12. We first assess the quality of the registration by computing error measurements on landmarks. In our application, we are particularly interested in the region near the renal lesion; we thus choose the lesion as a landmark to evaluate the registration. For every sequence, the lesion has been manually segmented in each frame. If the registration were perfect, the segmentation (after compensation by the motion estimated via the kidney) would be stable along the frames. We thus evaluate our registration by computing the Dice coefficient between the lesion in each frame and the lesion in the reference frame. Figures 5.10 shows this score, in comparison with the original sequence and the sequence registered by the block-matching method of [Romain et al., 2012]. Our method globally provides more precise and robust registration in the area of the tumour as the obtained Dice coefficients have both a higher mean and a lower variance. Furthermore, the motion of center of the lesion, which was of 6.6mm (median over the datasets) in the original sequences, is reduced to 1.6mm after our stabilization, which is below the axial resolution. Detailed boxplots are reported in Figure 5.11. A Wilcoxon signed-rank test [Wilcoxon, 1945] on those scores shows that our approach outperforms the baseline with a p-value $< 10^{-4}$. The very few failures observed in Sequence 10 (see Figures 5.10 and 5.11) were due to a large motion that made our method fall into a local minimum, and could be corrected by performing a global pre-registration in the axial direction.

Evaluation on parametric images

Our method is further evaluated by comparing parametric images estimated from the registered sequence. The chosen pharmacokinetic model, named the Tofts model, is a simple yet effective way to understand the contrast dynamics [Tofts et al., 1999]. Actually, it is often considered as the standard model for DCE sequences [Sourbron & Buckley, 2011]. In each voxel, two parameters (θ_1, θ_2) are obtained by fitting the contrast concentration $C_{tissue}(t)$ - which is proportional to the time-intensity curve $I(t)$ - to the solution of the following

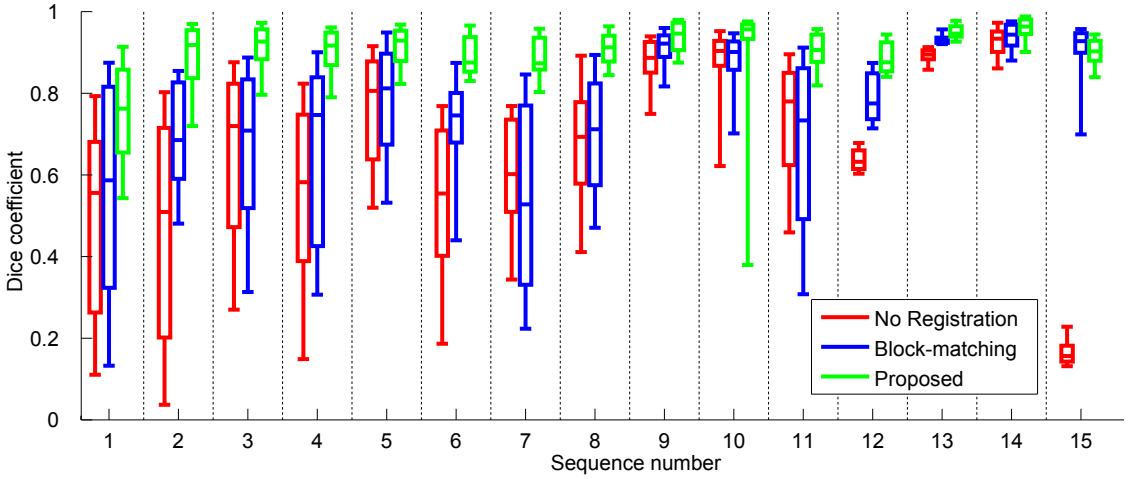


Figure 5.10: Boxplots of lesions Dice coefficients from original sequences (red), sequences registered with entropy-based block matching [Romain et al., 2012] (blue) and the proposed method (green).

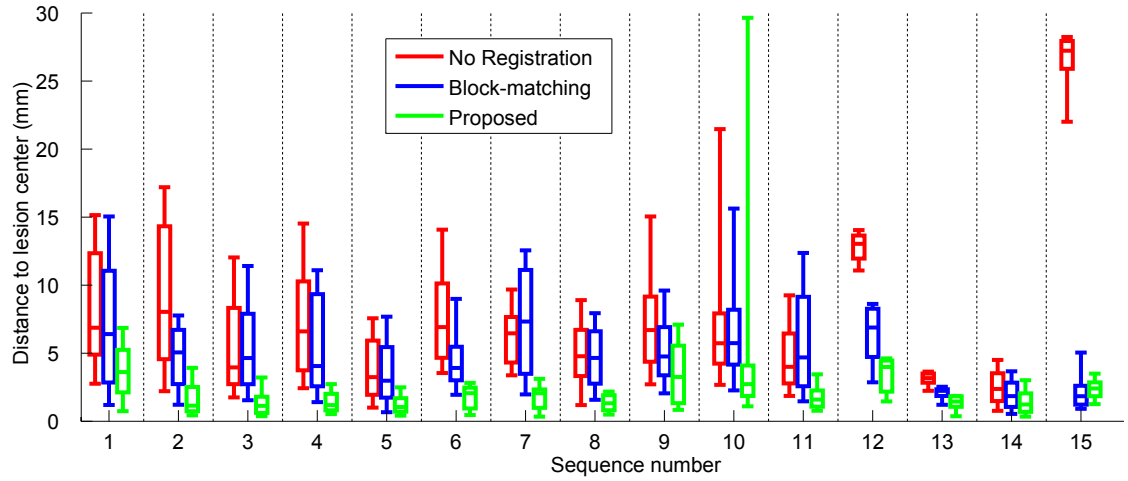


Figure 5.11: Boxplots of distance to lesion centers from original sequences (red), sequences registered with entropy-based block matching [Romain et al., 2012] (blue) and the proposed method (green).

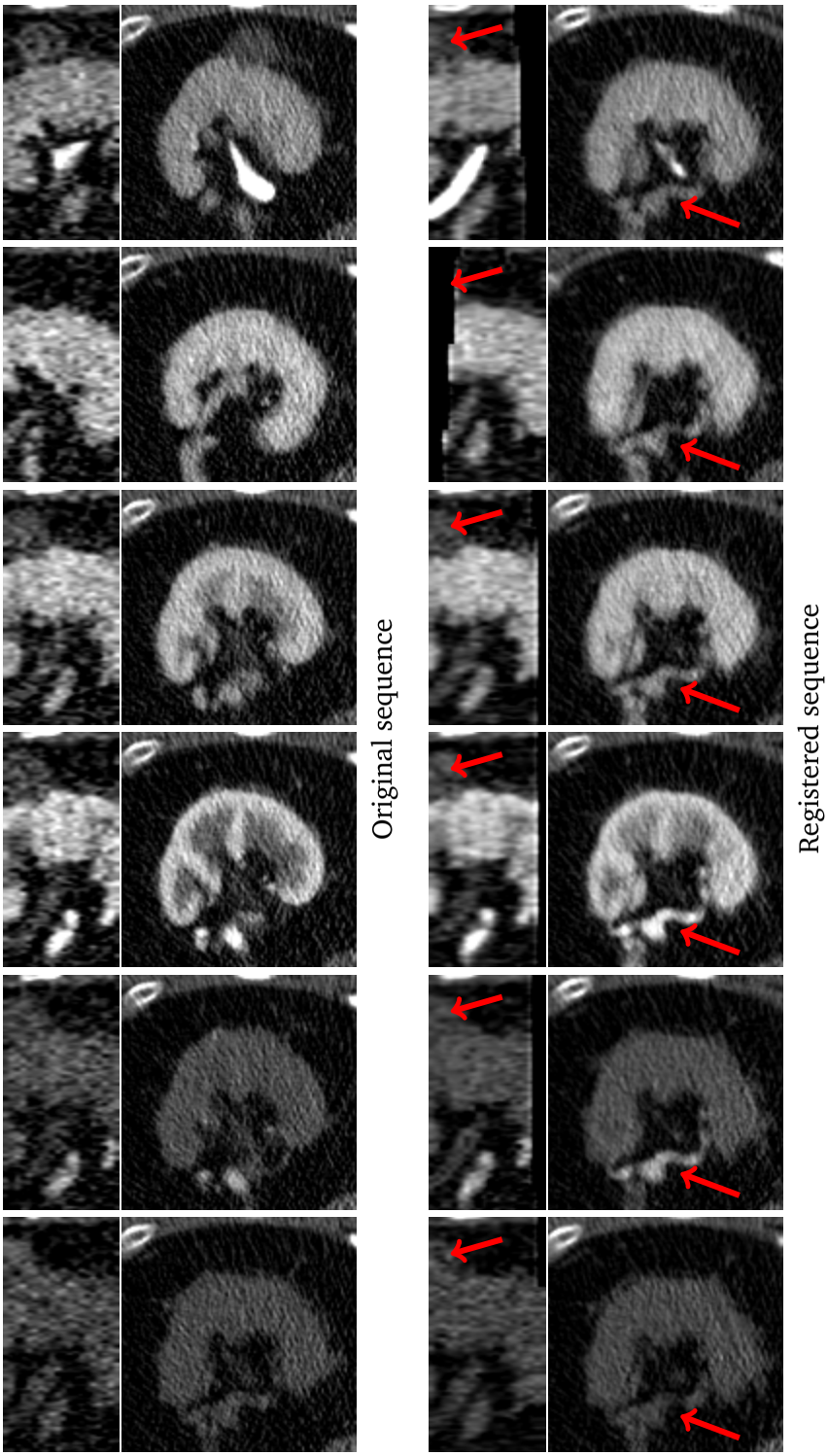


Figure 5.12: Crops of coronal and axial slices between an original sequence (top) and the same sequence registered with our method (bottom) along the acquisition times. Note the stabilization of the small blood vessel and the lesion (arrows).

differential equation

$$\frac{dC_{tissue}(t)}{dt} = \theta_1\, C_{aorta}(t) - \theta_2\, C_{tissue}(t). \tag{5.18}$$

Depending on the tissue, θ_1 represents either the blood flow (in high permeability approximation) or the permeability surface area (in low permeability approximation) whereas θ_2 can be interpreted as the same quantity divided by the extra-vascular extracellular volume. More details on the model are available in [Tofts et al., 1999]. The contrast concentration in the aorta, denoted by C_{aorta}, was modeled as the sum of two sigmoid functions and a Gaussian function:

$$C_{aorta}(t) = \frac{\alpha_1}{1+e^{-\alpha_2(t-\alpha_3)}} + \frac{\alpha_4}{1+e^{-\alpha_5(t-\alpha_6)}} + \frac{\alpha_7}{\sqrt{2\pi}\alpha_8} e^{-\frac{(t-\alpha_9)^2}{2\alpha_8^2}} \tag{5.19}$$

This model was chosen empirically; having a closed-form expression for $C_{aorta}(t)$ simplifies the pharmacokinetic parameters (θ_1, θ_2) estimation. The parameters $(\alpha_i)_{i=1\ldots9}$ in 5.19 were estimated for each sequence within a region of interest inside the aorta.

The sum of squared errors is minimized using the Levenberg-Marquardt [Marquardt, 1963] algorithm. In each voxel, we compute the residual fitting error at convergence, which quantifies how much the time-intensity curve deviates from the model. These are reported in Figure 5.13, which shows an improvement over the baseline method in every tested sequence but one. This is illustrated by the example given in Figure 5.14. Our method provides much smoother curves in both region of interests, which improves the reliability of the subsequent parameter estimation. As the true parameters are unknown, we can only assess their value visually. In the case reported in Figure 5.15, the lesion is much better distinguished in the parametric image obtained with our registration, which tends to prove the pertinence of our approach. The inner structures of the kidney are also more precisely delineated.

5.3.4 Discussion

In this section, we proposed a fast, automatic and robust method to register 3D+t DCE-CT sequences. To be able to cope with contrast uptake, our approach relies on a segmentation of the organ of interest, rather than intensity-based similarity criteria. This segmentation is simultaneously estimated, within a co-segmentation framework. Experiments showed that it provides better results than the state-of-the-art both quantitatively in terms of registration and qualitatively in terms of pharmacokinetic parameters estimation. The only few failures happened for a sequence (see Sequence 10 in Figure 5.10) in which an extremely large tumour hinders the kidney segmentation in less than 10% of the frames.

The proposed approach is generic and can be extended to other organs. For the kidney, a rigid transformation was enough to capture the movement of the region of interest. Other organs may undergo a different type of motion, such as affine transformations. This can be taken into account within our framework by adapting $(G_i)_i$ transformations. Another

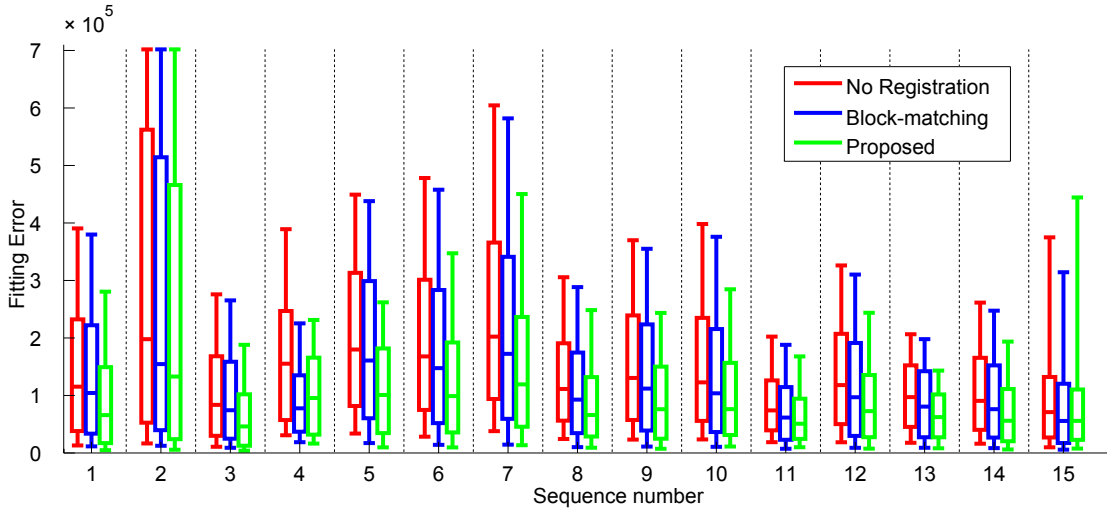


Figure 5.13: Boxplots of pixelwise fitting errors (sum of squared differences with the model) for each sequence, estimated from original sequences (red), sequences registered with entropy-based block matching [Romain et al., 2012] (blue) and the proposed method (green).

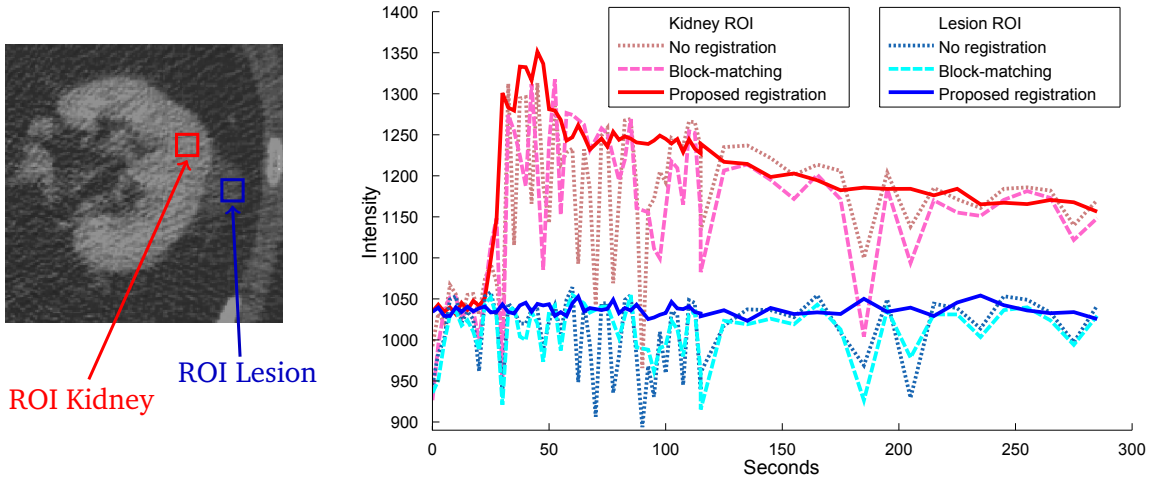


Figure 5.14: Time-intensity curves in two different regions of interest of a sequence (one inside a kidney colmun, another one in a cyst) for (a) the unregistered sequence and the sequence registered with (b) entropy-based block matching [Romain et al., 2012] and (c) the proposed method.

interesting idea would be to launch the co-segmentation on multiple organs and then blend each of the organ-based estimated transformation to create a poly-rigid or poly-affine motion [Arsigny et al., 2005].

The co-segmentation could also have been directly applied to the tumour instead of the organ, but the definition of the image-based classification terms $(r_i)_i$ would have been challenging and a pre-registration (*e.g.* via the block matching approach of [Romain et al., 2012]) would probably have been necessary.

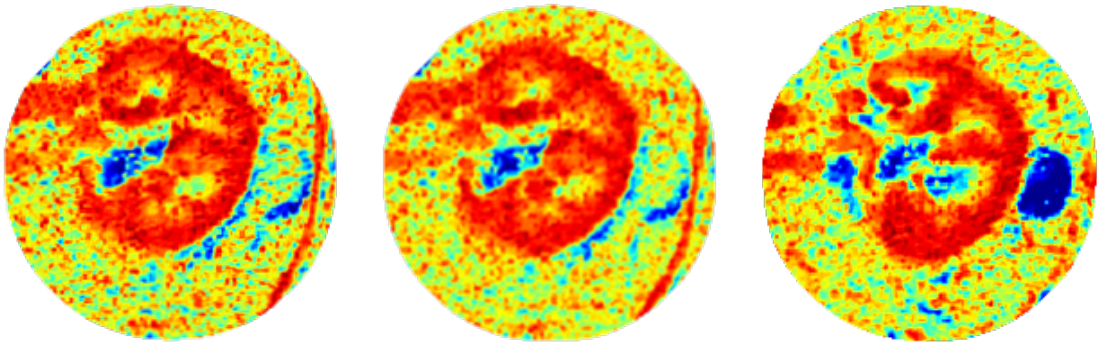


Figure 5.15: Maps of parameter θ_2 estimated on unregistered sequences (left) and sequences registered with entropy-based block matching [Romain et al., 2012] (center) and the proposed method (right). Note that the lesion is much more visible in our parametric image.

In our experiments, the acquisitions frequency was so low (at least 2.5 seconds between two successive frames) that no temporal coherence was enforced. For other applications, temporal consistency can however be useful. This could be achieved by adding extra terms in the energy to constrain the transformations $(G_i)_i$ and is currently under investigation.

*
* *

We have presented in this chapter a framework for joint co-segmentation and registration, which allows to take into account within a given segmentation algorithm either multi-modal or longitudinal information. In order to show its genericity, we applied this approach to two algorithms (namely a robust ellipsoid detection method and implicit template deformation) in two very different clinical settings. In both cases, we achieved improvements over the state-of-the-art methods.

The co-segmentation with the implicit template deformation framework will be used again and generalized in Chapter 7 with the introduction of *tagged* templates.

Chapter 6

Incorporating shape variability in implicit template deformation

Contents

Abstract

In this chapter, we propose a method to learn and use prior knowledge on shape variability in the implicit template deformation framework. This shape prior is learnt via an original and dedicated process in which both an optimal template and principal modes of variations are estimated from a collection of shapes. This learning strategy does not require one-to-one correspondences between shape sample points and is not biased by a pre-alignment of the training shapes. We then generalize the implicit template deformation formulation to automatically select the most plausible deformation as a shape prior. This novel framework maintains the two main properties of implicit template deformation: topology preservation and computational efficiency. Our approach can be applied to any organ with a

possibly complex shape but fixed topology. We validate our method on myocardium segmentation from cardiac magnetic resonance short-axis images and demonstrate segmentation improvement over standard template deformation.

Résumé

Dans ce chapitre, nous proposons une méthode pour apprendre et incorporer dans la méthode de déformation de modèle implicite des informations concernant la variabilité de forme de l'organe à segmenter. Ces informations sont apprises via un algorithme original et dédié au sein duquel sont estimés, à partir d'une base de données de formes, à la fois un nouveau modèle optimal et des modes principaux de variations. Cet apprentissage ne nécessite pas de correspondance point à point entre les formes et n'est pas biaisé par le choix du pré-alignement des formes. Nous généralisons ensuite la formulation de la méthode de déformation de modèle implicite pour prendre en compte ces informations tout en gardant les propriétés essentielles de la méthode: préservation de la topologie du modèle et effiacité algorithmique. Le principe est de sélectionner automatiquement un a priori de forme, conforme à la variabilité observée dans la base de données, adapté à l'image segmentée. Nous montrons enfin que notre méthode, appliquée à la segmentation du myocarde dans des images de résonance magnétique, produit de meilleures segmentations que l'approche classique.

Implicit template deformation is a model-based segmentation framework that consists in deforming an initial template to segment an image. When one applies it to a particular clinical problem, the first step is to choose an adequate template. Indeed as we are working with diffeomorphisms, the template must have the same topology as the organ that we want to segment. But it also needs to be "close" to the target object, since the magnitude of the deformation is penalized. While it is always possible to build a synthetic template (*e.g.* an ellipsoid for the kidney in Section 4.1.4), one feels that this choice is probably suboptimal in other applications. The purpose of this chapter is to answer the following questions: how can we use a database to design an optimal (in some sense) template ? can we learn the shape variability from this database so that we can take it into account within the deformation penalization ?

This chapter is organized as follows: Section 6.1 introduces the problem and lists some of the related work available in the literature. In Section 6.2, we describe an original learning process that is tailored to the implicit template deformation framework. The learnt statistics will be used in a generalized formulation of the segmentation algorithm introduced in Section 6.3. Validation proving the benefits of our approach are provided in Section 6.4 in the application of myocardium segmentation in 2D MR images. Finally, discussion on potential improvements concludes the chapter in Section 6.5.

The main contents of this chapter were presented at the MICCAI 2013 conference [Prevost et al., 2013a].

6.1 Introduction

6.1.1 Motivation

Model-based methods are particularly effective and popular in medical image segmentation. Among them, implicit template deformation has recently been used in various applications for its interesting properties (computational efficiency, topology preservation, compatibility with user interactions). As previously detailed, this variational method consists in seeking a segmenting implicit function as a deformed implicit template. This template, acting as a shape prior, is therefore of paramount importance. However, in previous works the initial template was either set as a synthetic model (*e.g.* an ellipsoid for a kidney) or as a segmented organ from a single arbitrary image [Saddi et al., 2007]. Despite the consensus that learning shape priors is a powerful approach to improve robustness [Cremers et al., 2007; Heimann & Meinzer, 2009], this has never been proposed in the context of segmentation by implicit template deformation. As public databases are developed and become available, it is important to think about how we can exploit them to validate but also improve our algorithms.

Here we aim at (i) learning statistics from a database of shapes (*i.e.* the most likely shape and the main variations around it) on organs that present a possibly complex shapes but a consistent topology, (ii) proposing a method to exploit such learnt information within the segmentation framework of implicit template deformation. Naturally, our approaches have to maintain the interesting characteristics of implicit template deformation, namely the computational efficiency and the topology preservation. These two properties are usually incompatible but we are notwithstanding able to guarantee both by generalizing the formulation of the implicit template deformation. Combination of learnt shapes will be used not to directly segment the images but rather within the regularization term. Thus they will act as a shape prior that is automatically updated during the segmentation.

6.1.2 Related work on shape learning

The shape learning literature being considerably large, we point out here only well-known or closely related techniques.

In the early and popular active shape model [Cootes et al., 1995], objects are represented by an explicit parameterization of their boundary vertices. Statistics (mean shape and variations) are computed on these vertices coordinates, thus a suitable one-to-one vertices correspondence is needed across the database. This correspondence can be complicated to obtain, either tedious when relying on manually labeled points or lacking robustness when automatically obtained (e.g. [Besl & McKay, 1992]). Due to boundary self-intersections, shape topology may also be lost.

Implicit methods [Leventon et al., 2000a; Rousson & Paragios, 2002; Tsai et al., 2003; Cremers et al., 2003a] represent objects through the signed distance functions to their boundary to estimate statistics. Although vertices correspondence is no longer needed during the learning step, this representation is still inadequate for topology preservation. In [Rousson & Paragios, 2002], Rousson and Paragios built a probabilistic model in order to estimate a mean implicit function ϕ_m (and an associated variance at each pixel) from a dataset of signed distance functions. Unlike most previous works, they constrained ϕ_m to be not any implicit function but a true distance function, *i.e.* $\|\nabla\phi_m\| = 1$. Although more satisfying, this approach requires complex optimization schemes and the constraint is only enforced during the learning (and not the subsequent segmentations). Furthermore, it is still inadequate for topology preservation.

Finally, closely related with this paper, a third class of methods uses statistics on diffeomorphisms of implicit shape representations [Arsigny et al., 2006a; Vaillant et al., 2004] or on currents [Durrleman, 2010]. While they present elegant and appealing theoretical properties and do preserve shape topology, they are also much more computationally expensive. Most of their applications therefore lie in offline shape analysis and they were not used for segmentation purposes (apart from atlas-based segmentation methods [Khan et al., 2008], which are not compatible with real-time or user-interactions).

Here, we propose an approach that is closely related to this third class of methods, since implicit template deformation consists in seeking a space transformation. We thus introduce a dedicated learning approach by using the template deformation energy as a pre-metric in the shapes space. This idea was inspired by the seminal paper of Joshi *et al.* , in which they construct an unbiased mean template by minimizing a sum of distances to a database [Joshi et al., 2004]. As they were motivated by registration applications, they worked directly on images. When applied to shapes, this approach yields a co-segmentation process (sharing some ideas with [Yezzi & Soatto, 2003] to a certain extent), within which an optimal shape is estimated (see Figure 6.1). However, we go further than both papers by learning (and subsequently exploiting) also the variability of the shape around this mean.

6.2 A learning process dedicated to template deformation

Given a training set of variables $(X_n)_{n=1\ldots N} \subset \mathbb{S}^N$, one can define its *mean* (more precisely its Fréchet-mean or Karcher-mean [Karcher, 1977]) as the solution of the following problem:

$$\bar{X} = \arg\min_{X\in\mathbb{S}} \frac{1}{N}\sum_{n=1}^{N} d^2(X, X_n) \tag{6.1}$$

This definition therefore depends both on the space $\mathbb{S}$ that is used to represent shapes and the distance d that defines their similarity. We will use these two degrees of freedom to build a mean shape that is tailored for the implicit template deformation.

6.2.1 A dedicated estimation of a mean model

The first choice concerns the space of shapes $\mathbb{S}$. Shapes can be represented by different types of objects (*i.e.* vertices [Cootes et al., 1995], implicit functions [Tsai et al., 2003], deformations [Joshi et al., 2004], currents [Durrleman, 2010], etc.). Our goal here is to estimate a model for the implicit template deformation framework, so we will choose an implicit representation. However, we would like to specify and fix the topology of the considered shapes. This information will be given by an initial implicit template ϕ_0, on which the space of admissible shapes will depend:

$$\mathbb{S}_{\phi_0} = \left\{\phi : \Omega \to \mathbb{R} \text{ such that } \phi = \phi_0 \circ L \text{ with } L \in \text{Diff}(\Omega)\right\} \tag{6.2}$$

which can be thought of as the *orbit* of ϕ_0 in the set of shapes. Note that such a space is stable under any diffeomorphism. Its dependency on ϕ_0 is rather low (except the topology) as it is identical to any $\mathbb{S}_\phi$ such that $\phi \in \mathbb{S}_{\phi_0}$. For the sake of simplicity, in the following we will omit the index and denote this space $\mathbb{S}$.

In order to estimate statistics in $\mathbb{S}$, we then define a metric-like function in this space which should be related to our segmentation framework. To do so, we point out that any shape $\phi_1 \in \mathbb{S}$ can be warped to another shape $\phi_2 \in \mathbb{S}$ via implicit template deformation. Indeed, we can segment an image representing ϕ_2 using ϕ_1 as template. With the notations of (3.10), we simply have to set $\phi_0 := \phi_1$, $r_{int}^{\phi_2} := \max(-\phi_2, 0)$ and $r_{ext}^{\phi_2} := \max(\phi_2, 0)$. The definition of $r_{int}^{\phi_2}$ and $r_{ext}^{\phi_2}$ is not unique and we could have selected other functions that represents the interior and the exterior of ϕ_2. The rationale behind this particular choice is that the difference $r_{int}^{\phi_2} - r_{ext}^{\phi_2}$ is then equal to $-\phi_2$. This leads to a tailored definition of shape dissimilarity d^2.

Definition 6.1. *The shape dissimilarity from shape ϕ_1 to shape ϕ_2 is defined as*

$$d^2(\phi_1, \phi_2) = \min_{\substack{L \in Diff(\Omega) \\ G \in \mathbb{G}}} \begin{cases} & \displaystyle\int_\Omega H(\phi_1 \circ L \circ G)\ \max(-\phi_2, 0) \\ + & \displaystyle\int_\Omega (1 - H(\phi_1 \circ L \circ G))\ \max(\phi_2, 0)\ , \\ + & \dfrac{\lambda}{2}\|L - Id\|_U^2 \end{cases}$$

or equivalently

$$d^2(\phi_1, \phi_2) = C(\phi_2) + \min_{\substack{L \in Diff(\Omega) \\ \mathcal{G} \in \mathbb{G}}} -\int_\Omega H(\phi_1 \circ L \circ \mathcal{G})\ \phi_2 + \frac{\lambda}{2}\|L - Id\|_U^2\ , \tag{6.3}$$

where $C(\phi_2)$ is a constant that only depends on ϕ_2.

In this definition, we recall that the U-norm represents the natural norm in the Gaussian reproducing kernel Hilbert space (see Section 3.1 for more details).

Remark 6.1. *The shape constraint parameter λ should be chosen carefully, since a too high value will prevent ϕ_1 to be exactly matched to ϕ_2 and the learning will be biased. In practice however, it is not difficult to find a suitable value.*

This shape dissimilarity measure is not a distance but a pre-metric, as it is not symmetric and does not verify triangular inequality. The lack of symmetry is directly inherited from the segmentation process itself as the template ϕ_0 has a very particular role. Triangular inequality does not either appear as an important property for our application. Cremers *et al.* discussed these properties in [Cremers & Soatto, 2003] and point out that defining a true distance between implicit shapes is still an open problem. But anyway, this function does measure a closeness between two shapes and we can still use it define our dedicated notion of mean shape.

Definition 6.2. *An implicit function $\bar{\phi}$ is a mean of the set $\{\phi_n\}_{n=1..N}$ (in the sense of implicit template deformation) if it is a local minimum of the shape dissimilarity to all the elements of this set, i.e.*

$$\bar{\phi} = \arg\min_{\phi \in \mathbb{S}} \frac{1}{N} \sum_{n=1}^{N} d^2(\phi, \phi_n) \, . \tag{6.4}$$

It is important to note that we seek the mean shape $\bar{\phi}$ as an element of $\mathbb{S}$. Indeed, in our application the mean shape has to preserve the topology of the training shapes. This means that there exists $L \in \mathrm{Diff}(\Omega)$ such that $\bar{\phi} = \phi_0 \circ L$. The mean shape expression can thus be reformulated as

$$\bar{\phi} = \phi_0 \circ \left\{ \arg\min_{L \in \mathrm{Diff}(\Omega)} \frac{1}{N} \sum_{n=1}^{N} d^2(\phi_0 \circ L, \phi_n) \right\} \, . \tag{6.5}$$

Expanding the segmentation costs and neglecting constant terms in Equation (6.5) yields the following optimization problem to solve:

$$\min_{\substack{L \in \mathrm{Diff}(\Omega) \\ (L_n)_n \in \mathrm{Diff}(\Omega)^N \\ (G_n)_n \in \mathbb{G}^N}} E_{learn} = -\sum_{n=1}^{N} \int_{\Omega} H(\phi_0 \circ L \circ L_n \circ G_n) \, \phi_n + \sum_{n=1}^{N} \frac{\lambda}{2} \|L_n - \mathbf{Id}\|_U^2 \, . \tag{6.6}$$

This can be interpreted as segmenting simultaneously all training shapes $(\phi_n)_n$ starting from ϕ_0 while estimating an optimal common intermediate shape $\phi_0 \circ L$ (see Figure 6.1). In (6.6), the energy E_{learn} is minimized with respect to three kinds of variables

- the global transformations $(G_n)_n$, called the **poses**, that register all shapes to ϕ_0 with translation, rotation and scaling. As they are part of the optimization process, they do not bias the learning, as a fixed pre-alignement (*e.g.* [Rousson & Paragios, 2002; Tsai et al., 2003]) would do.

- the **common deformation** L, which includes the common parts of the deformations from ϕ_0 to all the training shapes;

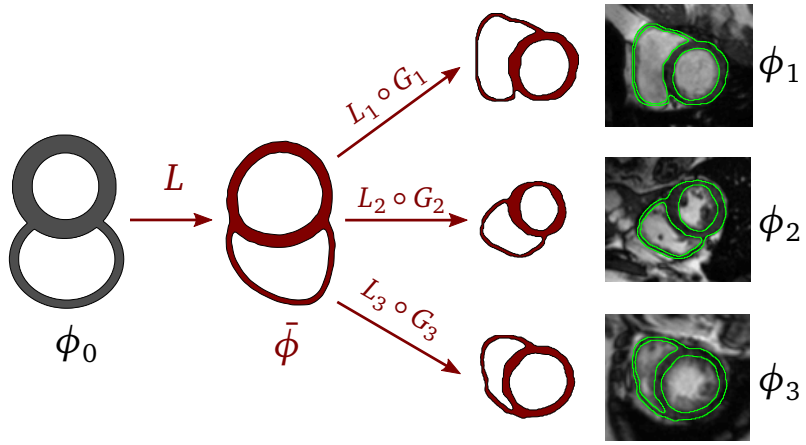


Figure 6.1: Given an initial synthetic shape ϕ_0, a set of shapes $\{\phi_n\}_n$ is simultaneously segmented via implicit template deformation while an intermediate mean shape $\bar{\phi} = \phi_0 \circ L$ is estimated. The topology of ϕ_0 is preserved during the process.

- the local deformations $(L_n)_n$, called the **residual deformations**, are the residual components of the deformations from $\phi_0 \circ L$ to ϕ_n. Unlike L, their magnitude is penalized so that any deformation which is common to all the training set will be preferably included in L.

The optimal common deformation L^* can be used to define the optimal shape (in the sense of the segmentation algorithm) as $\bar{\phi} = \phi_0 \circ L^*$. This shape globally minimizes the magnitude of residual deformations to each shape of the dataset. Note that the magnitude of L is not penalized so the choice of ϕ_0 defines the topology of $\bar{\phi}$ but does not affects it further (modulo the smoothness enforced to L). In our experiments, running a second time the learning process with $\phi_0 := \bar{\phi}$ did not alter the results.

Remark 6.2. *We assumed that the set of training shapes $\{\phi_n\}_n$ was a subset of $\mathbb{S}_{\phi_0}$. It is a very natural hypothesis and ϕ_0 should be chosen accordingly. However, those training shapes will probably come from manual annotations of images and, as such, be prone to errors. As a consequence, it may occur that some training shapes do not have the correct topology. This does not question the soundness of our learning because, in such cases, we will implicit learn the "closest shapes" with the topology of ϕ_0.*

Details on the resolution of (6.6) are provided in the next subsection.

6.2.2 Numerical optimization

Problem (6.6) presents some similarities with the co-segmentation proposed in the previous chapter (see Section 5.2.3). It is therefore minimized similarly, with a gradient descent simultaneously on each of the unknowns. The gradient directions with respect to $\mathbf{p}_{n,i}$ (the i-th parameter of G_n), the common deformation L and the residual deformations L_n are given by the following equations[1]:

[1]They are obtained with standard calculus of variation, but we omit the tedious details here.

$$\begin{cases} \nabla_{\mathbf{p}_{n,i}} E_{learn} &= \displaystyle\int_{\Omega_0} \delta(\phi_0 \circ L \circ L_n)\ \phi_n \circ G_n^{-1} \left| J_{G_n}^{-1} \right| \\ & \qquad \left\langle \nabla \phi_0 \circ L \circ L_n,\ DL \circ L_n \,.\, DL_n \circ \dfrac{\partial G_n}{\partial \mathbf{p}_{n,i}} \circ G_n^{-1} \right\rangle \\ \nabla_L E_{learn} &= K_\sigma * \left[\displaystyle\sum_{n=1}^{N} \delta(\phi_0 \circ L)\ \phi_n \circ G_n^{-1} \circ L_n^{-1} \,.\, \left| J_{L_n \circ G_n}^{-1} \right| \,.\, DL \,.\, \nabla \phi_0 \circ L \right] \\ \nabla_{L_n} E_{learn} &= K_\sigma * \left[\delta(\phi_0 \circ L \circ L_n)\ \phi_n \circ G_n^{-1} \,.\, \left| J_{G_n}^{-1} \right| \,.\, DL_n \,.\, \nabla \phi_0 \circ L \circ L_n \right] + \lambda (L_n - \mathbf{Id}) \end{cases} \tag{6.7}$$

The first gradient – with respect to the poses – is used in a standard gradient procedure, while the two others – with respect to the common and residual deformations – are exploited in a topolgy-preserving optimization scheme (see (3.17) in Section 3.2).

All these integrands actually have a very small support (the zero level-set of an implicit function), which makes the computations fast. Moreover, the three kinds of gradients have a lot of terms in common.

Remark 6.3. *Some terms depends on inverses of diffeomorphisms L_n^{-1}. As mentioned in Section 3.2, these deformations are built iteratively and simultaneously to the direct transformations L_n.*

Remark 6.4. *Although the computations needed for the training process are relatively fast, there is a high memory requirement (especially in 3D) since a high number of implicit functions and deformation fields have to be stored simultaneously. A possible solution would be to use a stochastic gradient descent [Bottou, 1998], i.e. at each iteration only consider a randomly chosen subset of the training set.*

6.2.3 Building a space of deformation priors

As seen in the previous subsection, minimization of (6.6) yields a mean shape. However, the optimal residual deformations $(L_n^*)_n$ are also available and can be used to capture further information on the variability of the training shapes.

We build a space of principal deformations $\mathbb{L}$ to constrain future segmentation of new images. Similarly to [Rueckert et al., 2001], a principal component analysis (PCA) [Jolliffe, 1986] is applied to the residual deformations to find a suitable parametrization of such a space. The goal of this analysis is to find a reduced number of orthogonal vectors that maximize the explained variance of the residual deformations. This is accomplished by first computing the mean residual deformation

$$\bar{\ell} = \frac{1}{N} \sum_{n=1}^{N} L_n^* \tag{6.8}$$

and then performing a singular value decomposition (SVD) of the sample covariance matrix

$$S = \frac{1}{N-1} \sum_{n=1}^{N} (L_n^* - \bar{\ell})(L_n^* - \bar{\ell})^T \ . \tag{6.9}$$

Any deformation ℓ in agreement with the variability of the training data can then be approximated by a linear combination of the offset $\bar{\ell}$ and $(\ell_k)_{k=1..M}$ the first M singular vectors of S. It is parametrized by the vectors of its weights $w \in \mathbb{R}^M$:

$$\ell \approx \ell[w] = \bar{\ell} + \sum_{k=1}^{M} w_k \, \ell_k \ . \tag{6.10}$$

We denote $\mathbb{L}$ the set of such transformations. M can be empirically chosen using the distribution of the modes' eigenvalues. To each singular vector ℓ_k corresponds a singular value λ_k that represents the amount of variance of the residual deformations that is explained with this mode of variation.

Remark 6.5. *Note that even if the PCA is applied to the residual deformations, $\bar{\ell}$ is non-null (though with a very small magnitude) because it denotes a mean with respect to a different metric than L^*.*

The space of diffeomorphisms is not stable under linear combinations. There is therefore no guarantee that an element of $\mathbb{L}$ is actually a diffeomorphism. Nevertheless, as shown in the next section, it is possible to use this space indirectly in a topology-preserving segmentation framework.

6.3 Generalized implicit template deformation

The previously estimated statistical information can be used to robustify and improve future segmentations. In order to incorporate such information in the segmentation process, we propose a generalization of the implicit template deformation framework.

6.3.1 An improved formulation for segmentation

A first improvement is achieved by replacing the original template ϕ_0 by the mean template $\bar{\phi} = \phi_0 \circ L^*$. Secondly, the estimation of the deformation can also be enhanced by using the space of principal deformations $\mathbb{L}$. In most previous works [Cootes et al., 1995; Leventon et al., 2000a; Tsai et al., 2003], the learnt variable is expressed as a linear combination of modes. When dealing with deformations, this does not guarantee topological preservation. Therefore, we rather use such linear combinations indirectly. More specifically, we modify the regularization term so that the diffeomorphism L is constrained with respect to the set $\mathbb{L}$ instead of the identity (see Figure 6.2). Thus, only deformations that cannot be explained

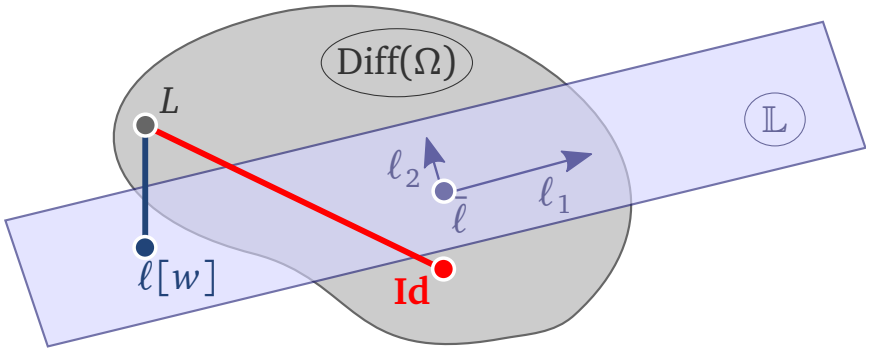


Figure 6.2: Comparison of the penalization of the deformation L with the standard regularization term towards the Identity (red) and the novel term towards the space $\mathbb{L}$ (blue) that is centered in $\bar{\ell}$ and spanned by the modes $(\ell_k)_k$. The new prior $\ell[w]$ is the projection of L onto the set $\mathbb{L}$. Note that all transformations of $\mathbb{L}$ are not diffeomorphisms but L is constrained to be one.

through the learnt space $\mathbb{L}$ are penalized. The new segmentation energy therefore reads

$$E_{seg}(L,G,w) = \int_\Omega H(\bar{\phi}\circ L\circ G)\; r_{int} + (1-H(\bar{\phi}\circ L\circ G))\; r_{ext} + \frac{\lambda}{2}\|L-\ell[w]\|_U^2 \;. \quad (6.11)$$

This represents a generalization of the standard template deformation formulation. The novel regularization term can be interpreted as a shape prior that depends on the image. Thus, even if the target organ has a high variability around the mean, we can learn it in order to automatically select the most plausible shape that is implicitly used to constrain the segmentation.

A related approach was proposed in [Rousson et al., 2004] with implicit functions. In this paper, the authors defined the regularization term as a distance between the segmenting implicit function and a linear combination of implicit modes previously learnt. However our method presents a major advantage over theirs. Indeed, as described in Section 3.2, we are able to let the deformation L evolve while preserving its diffeomorphic properties (and therefore maintaining the topology of the template ϕ_0). Conversely, it is not possible to easily enforce such a constraint into the evolution of an implicit function.

6.3.2 Numerical optimization

Minimization of (6.11) can be performed with a two-step alternate scheme :

Update of the segmentation With $\ell[w]$ fixed, the energy is minimized through a gradient descent-like scheme on L and G (see Sections 2.7.3 and 3.2).

Update of the shape prior With L and G fixed, the update of $\ell[w]$ can be seen as a projection of L onto $\mathbb{L}$. Indeed the energy comes down to a simple quadratic function, whose

minimizers are obtained by solving a simple linear system, as stated by the following proposition.

Proposition 6.1. *The minimum of $E_{seg}(L, G, \cdot)$ is reached at $w^* \in \mathbb{R}^M$ such that*

$$Aw^* = b_L \tag{6.12}$$

where A is a $M \times M$ matrix whose entries are $(A_{ij}) =< \ell_i, \ell_j >_U$ while $b_L \in \mathbb{R}^M$ is defined by $(b_{L,i}) =< L - \bar{\ell}, \ell_i >_U$.

Proof. With L and G fixed, the minimization problem comes down to

$$\arg\min_{w \in \mathbb{R}^M} E_{seg}(L, G, w) = \arg\min_{w \in \mathbb{R}^M} \|L - \bar{\ell} - \sum_{k=1}^{M} w_k \ell_k\|_U^2 \,. \tag{6.13}$$

Setting to zero its derivative with respect to the weight of the mode k_0 yields

$$0 = \frac{d}{dw_{k_0}} \left\langle L - \bar{\ell} - \sum_{k=1}^{M} w_k \ell_k,\ L - \bar{\ell} - \sum_{k=1}^{M} w_k \ell_k \right\rangle_U (w^*) \,, \tag{6.14}$$

$$0 = \left\langle L - \bar{\ell} - \sum_{k=1}^{M} w_k^* \ell_k,\ \ell_{k_0} \right\rangle_U \,, \tag{6.15}$$

$$\sum_{k=1}^{M} w_k^* \langle \ell_k,\ \ell_{k_0} \rangle_U = \langle L - \bar{\ell},\ \ell_{k_0} \rangle_U \,. \tag{6.16}$$

Each k_0 yields a linear equation in w, hence the result. □

Note that the matrix A is quite small so this system is very easy to solve. Actually, we can even pre-compute the inverse of A since it only involves learnt variables.

However we may simplify this solution further by making some hypotheses. Recall that by construction via the PCA, the $(\ell_k)_k$ are L^2-orthogonal. If we assume that they are also nearly U-orthogonal, then the matrix M is diagonal and the solutions are given by:

$$\forall k \in \{1, ..., M\},\ w_k^* = \frac{\langle L - \bar{\ell}, \ell_k \rangle_U}{\langle \ell_k, \ell_k \rangle_U} \,. \tag{6.17}$$

The values of w_k^* are subsequently clipped in the interval $[-3\sqrt{\lambda_k}; 3\sqrt{\lambda_k}]$, as these bounds represent the lengths of the semi-axes of the ellipsoid estimated by the principal component analysis. Any deformation obtained with weights beyond this interval are not in agreement with the training set and thus should not be considered as possible priors. Other possibilities of computing these weights will be mentioned in the Discussion section of this chapter.

To sum up, the first step is similar to the standard implicit template deformation formulation, and the second one is straightforward. Therefore, the computational efficiency of the method is maintained while topology preservation is still guaranteed.

6.4 Application: Myocardium segmentation in 2D MR images

We validated our method in the context of myocardium analysis and segmentation in cardiac short-axis 2D cine-MR images. Quantitative assessment on the heart muscle is critical for diagnosis or therapy planning. This task is particularly challenging for model-based approaches because of the complex topology of the target object, *i.e.* a band around left and right ventricles.

6.4.1 Material

Our dataset is composed of 245 MR images coming from 61 different patients (for each case, several slices in the z-direction are available). The acquisitions have been synchronised so that each heart is in the same cardiac phase. The typical images size was 256×256 with resolution 1.56×1.56 mm. In every image, a myocardium segmentation has been manually performed by a radiologist. Based on the geometric information, we set for our method the scale of the deformation field σ to 10 mm. The initial synthetic template ϕ_0 used is shown in Figure 6.1. Global transformations are sought in the set $\mathbb{G}$ of similarities (accouting for translation, rotation and isotropic scaling). The dataset was randomly split into a training set including 120 images from 30 patients and a testing set composed of the remaining 125 images coming from 31 patients.

6.4.2 Experiments on the learnt information

Synthetic experiments

First we conducted controlled experiments to assess quantitatively the estimation of the mean model. Random myocardium shapes were generated by applying random deformation fields to an original myocardium. We aim at recovering this original shape by estimating a mean model from subsets of these synthetically generated shapes. The efficiency of a learning process is evaluated by computing the Dice coefficient between the ground truth and the estimated mean shape. To avoid randomness bias, the experiments have been performed 100 times and the results averaged.

We reported in Figure 6.3 a comparison of three fully automatic methods using this metric: the implicit shape model proposed in [Tsai et al., 2003], the active shape framework [Cootes et al., 1995] with point correspondances estimated by ICP [Besl & McKay, 1992] and the proposed method. For any number of used shapes, our method provided statistically significantly better estimates of the original shape than the two others. These results can be better understood with Figure 6.4 showing the spatial localization of the errors. Indeed the implicit method fails to recover the entire muscle around the right ventricle: working

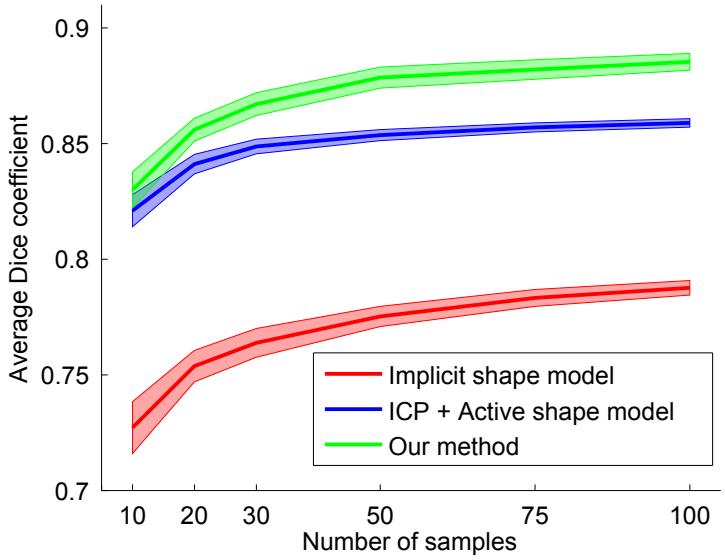


Figure 6.3: Dice coefficients (averaged over the 100 experiments) between the estimated mean model and the original model as a function of number of samples using implicit shape model [Tsai et al., 2003], ICP [Besl & McKay, 1992] + active shape model [Cootes et al., 1995] and the proposed method. Bands around the curves delineate the 95% confidence interval.

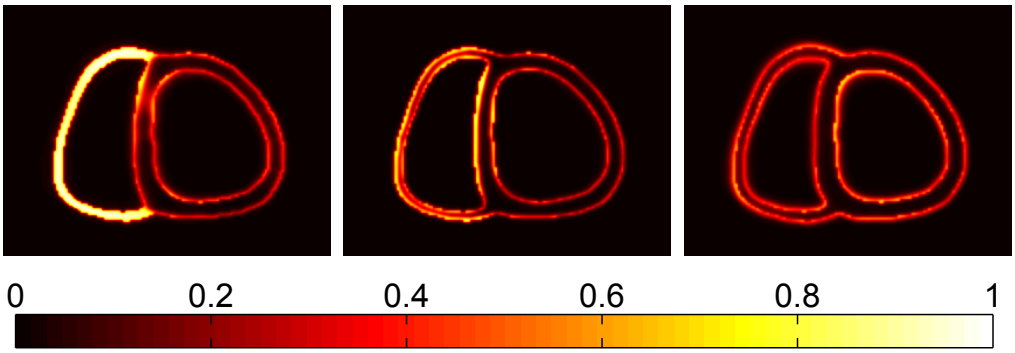


Figure 6.4: Repartition of errors on the estimated model using implicit shape model [Tsai et al., 2003] (left), ICP [Besl & McKay, 1992] + active shape model [Cootes et al., 1995] (middle) and the proposed method (right). Color indicates the pixelwise empirical probability of bad classification (inside vs outisde the shape).

directly on signed distance function is not adapted to thin structures. This area also causes high errors for the explicit method, which retrieves but underestimates this part of the band. Conversely, errors for our method are lower and more evenly distributed.

Mean model and principal modes

We now provide a qualitative comparison between the different approaches on learnt information from clinical data. The initial dataset was randomly split into a training set including 120 images from 30 patients and a testing set composed of 125 images coming from the remaining 31 patients.

The mean shape and first two modes of variation are shown in Figure 6.5 for each method. As expected from the results of previous subsection, the implicit method fails at recovering the true topology of the mean shape, but also with the first modes of variations. The explicit method performs better and provides a reasonable mean model. However, the modes of variation are not satisfying because very irregular and difficult to interpret. On the other hand, the results obtained with our method show a much better preservation of the topology and seem more realistic. Moving along the first principal components changes the relative size of the ventricles. This corresponds to the variability observed when moving on the axial direction of a given heart. This variation was expected because the training set include several slices of each heart. The second principal component controls the global anisotropic scaling of the hearts, which seems to rather represent an inter-subject variability. Such variations were not taken into account by the global poses because their scaling were isotropic.

6.4.3 Validation of the improved segmentation

We finally evaluate how learnt information improves segmentation via implicit template deformation of unseen images. Myocardiums have been segmented in test images using (i) the synthetic model ϕ_0 as template, (ii) the estimated mean model $\bar{\phi}$ as template, (iii) the new deformation model-based regularization term in addition to the mean model $\bar{\phi}$ (with 5 modes).

The image-based classification functions r_{int} and r_{ext} were set to

$$r_{int}(\mathbf{x}) = -\log(p_{int}(\mathbf{x})) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = -\log(p_{ext}(\mathbf{x}))\ . \tag{6.18}$$

where p_{int} and p_{ext} are of intensity probability distributions inside and outside the myocardium (estimated from the training datasets). For the intensities to be comparable, all the images were normalized beforehand.

Performance of each algorithm is quantified using Dice coefficients between the segmentation and the expert ground truth. Results on the whole testing set are summarized in Figure 6.6. Changing the template from ϕ_0 to the learnt $\bar{\phi}$ makes the algorithm more robust as the minimum Dice coefficient greatly increases (from 0.46 to 0.69). Modifying the regularization term by taking into account the deformation model further raises it 0.86. The proposed method globally enhances the algorithm on most images of the test database as the median goes from 0.85 for the baseline method to 0.92 with our modifications. These improvements are statistically significant with a p-value $< 10^{-4}$ for a Wilcoxon signed-rank

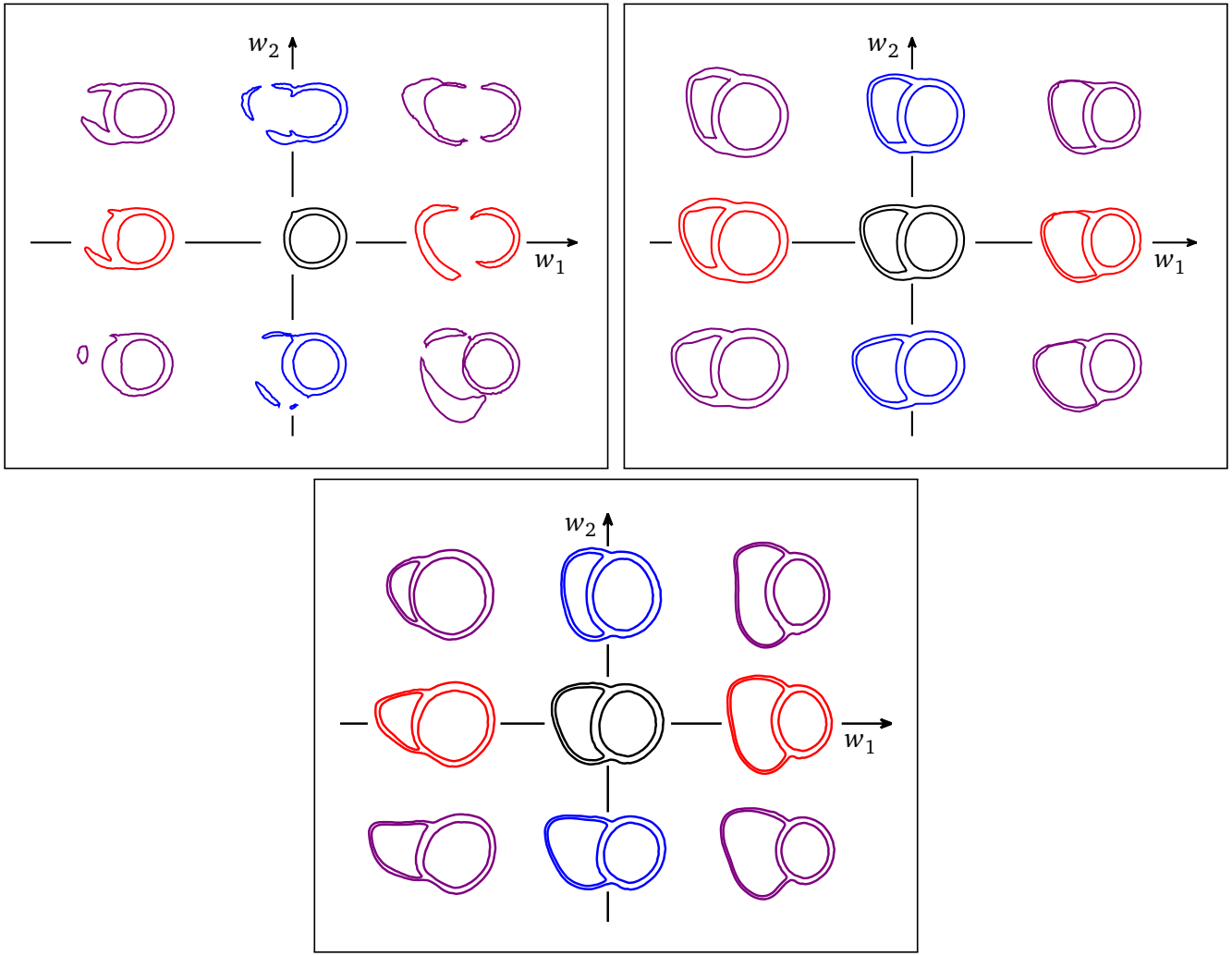


Figure 6.5: Mean model and first two modes of the variation of the myocardium learnt on the training dataset using implicit shape model [Tsai et al., 2003] (left), ICP [Besl & McKay, 1992] + active shape model [Cootes et al., 1995] (right) and the proposed method (bottom). For our approach, the visualized shapes are the zero level-sets of $\bar{\phi} \circ (\bar{\ell} + w_1\, \ell_1 + w_2\, \ell_2)$.

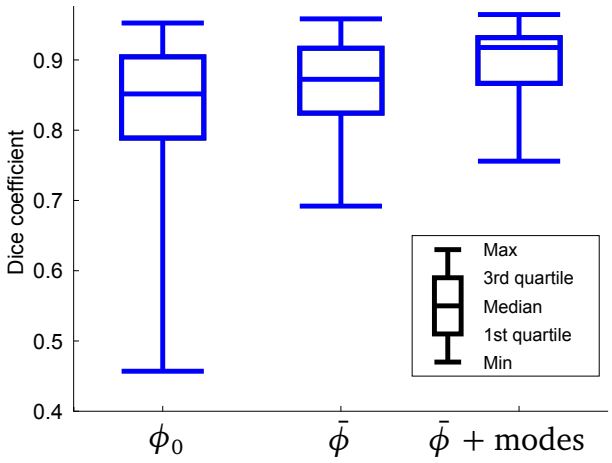


Figure 6.6: Boxplot of the Dice coefficients for myocardium segmentation in MR images via implicit template deformation with synthetic model ϕ_0 (left), mean model $\bar{\phi}$ (middle), mean model $\bar{\phi}$ and deformation modes (right).

test [Wilcoxon, 1945].

Figure 6.7 shows segmentation results in three different cases, for the classical regularization term with two values of the shape constraint parameter $\lambda \in \{1, 2\}$ and the new model-based regularization term. In all settings, the template was the mean model $\bar{\phi}$. Consider Case ♯1 (first row). Since the image term is reliable, a satisfying result is obtained with a small shape constraint. However, the myocardium deviates significantly from the mean shape: using a too strong constraint ($\lambda = 2$) prevents the algorithm to converge towards the right solution. Conversely in Case ♯2, the image information is much more ambiguous. This provokes some leaks (*e.g.* in papillary muscles of the left ventricle) with $\lambda = 1$, which shows there is no fixed value that allows a good segmentation in the first two cases. Yet by introducing the new regularization (fourth column), likely deformations are not penalized. This widens the capture range while still maintaining a high constraint on the shape and therefore avoiding unrealistic leaks. Finally, Case ♯3 illustrates that our method may also improve the result even if no λ was originally successful.

6.5 Discussion

We discuss herebelow some limitations of the current work and propose several ideas for future work.

From 2D to 3D Although we first proved the potential of our approach on a 2D application, it should be noted that the whole method can be directly extended to 3D shapes, thanks to the implicit representation of shapes. Another advantage of our approach is that it does not require point-to-point correspondences between shapes, which can be particularly challenging to obtain for three-dimensional shapes. We are thus currently investigating 3D applications, such as learning the shape variability of the liver.

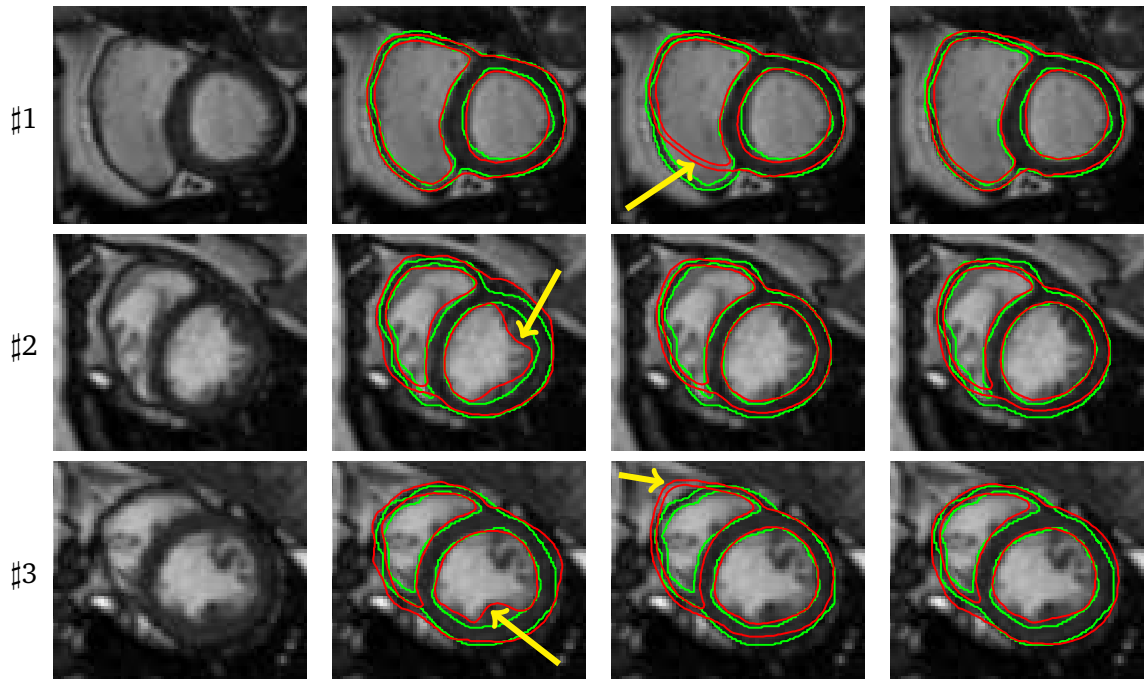


Figure 6.7: Segmentation results (red) of different cases versus ground truths (green). Main failures are highlighted by yellow arrows. (a) Original images, (b,c) Standard method with small ($\lambda = 1$) and high ($\lambda = 2$) shape constraint, (d) Proposed method.

Penalizing the weights of the modes In this chapter, we proposed to replace the regularization term on the deformation L from

$$\|L - Id\|^2 \qquad \text{to} \qquad \min_w \|L - \ell[w]\|^2 \ \text{ with } \ell[w] \in \mathbb{L} \tag{6.19}$$

which basically consists in removing the penalization on all deformations in the affine space $\mathbb{L}$. The rationale was that $\mathbb{L}$ is composed of deformations in agreement with the database. However the weights w should be not too large: with a Gaussian assumption, the training shapes are supposed to have their weight w_k in $[-3\sqrt{\lambda_k}; 3\sqrt{\lambda_k}]$. Our solution was to clip each weight into this interval, but there is a more elegant approach. We could decompose the original term $\|L - Id\|_U^2$ into

$$\min_{w \in \mathbb{R}^M} \|L - \ell[w]\|^2 + \|\ell[w] - \bar{\ell}\|^2 + \|\bar{\ell} - Id\|^2 \tag{6.20}$$

where the first term is the same as in (6.19). The third term is constant and can therefore be discarded in the minimization. Finally, the second term penalizes the distance between $\ell[w]$ and $\bar{\ell}$ that are both deformations in $\mathbb{L}$. Instead of using the standard L^2- or U-norms, we can exploit the assumption of Gaussian distribution

around $\bar{\ell}$ and define a Mahalanobis distance in $\mathbb{L}$ as follows:

$$\|\ell[w]-\bar{\ell}\|^2 = \nu \sum_{k=1}^{M} \frac{w_k^2}{\lambda_k^2} \tag{6.21}$$

where ν is a normalization factor. This simply means that modes with high variance are less penalized. Minimization remains easy as (6.17) becomes

$$\forall k \in \{1,...,M\},\ w_k^* = \frac{\left\langle L-\bar{\ell}, \ell_k \right\rangle_U}{\left\langle \ell_k, \ell_k \right\rangle_U + \nu \frac{1}{\lambda_k}} \tag{6.22}$$

The weights of the modes are slightly shifted down, according to the variance of the corresponding mode.

Choice of the dimension reductions method In order to build the space $\mathbb{L}$, we used as a dimension reduction approach the principal component analysis (PCA) since it is a standard and easy-to-implement method. However we might try other methods such as independent component analysis (ICA) [Comon, 1994] which, unlike PCA, does not aim at capturing the largest variance with orthogonal vectors but with decorrelated ones.

To understand the potential benefits of ICA over PCA, consider the following simple "thought experiment": in all the training shapes, only two disjoint zones Ω_A and Ω_B differ from the mean model. These zones are the same in each shape but they vary independently from each other. PCA will capture all the changes in a single mode, while ICA needs two modes (one for each region). Now let us imagine the segmentation of a corrupted image in which there is no information in Ω_A. With both methods, the weights of the shape prior will solely be determined by the information available in Ω_B. However, since the two regions are in the same PCA mode, *the shape prior will also change in* Ω_A if we use the PCA approach. This is clearly not desirable since the two variations were independent: there is no reason for Ω_A to influence Ω_B. The correct behaviour (*i.e.* the prior in Ω_A should be only the mean model *unless there exists some statistical correlation with other regions*) is obtained with the ICA modes.

Learning on logarithms We mentioned earlier that diffeomorphisms are not stable under linear combinations. This was the reason why we did not use directly the modes of the PCA in the segmentation step. However, we may question the learning step itself: is it really sensible to perform a PCA in a space that is not linear ? The answer is probably negative from a purely theoretical point of view.

One possible solution would be to perform the PCA on the logarithms of the residual deformations $(L_i^*)_i$, since the logarithm of a diffeomorphism is a vector field that does lie in a linear space [Arsigny et al., 2006a]. Preliminary experiments that we are currently conducting however suggest that both learning and segmentation results do not significantly change when working on logarithms, whereas the efficiency of the

algorithm is lost.

*
* *

In this chapter, we have presented an approach to include organ shape variability in the implicit template deformation framework. The training process is automatic, does not require landmarks correspondence and relies upon a definition of shape dissimilarity that is directly derived from the implicit template deformation functional. We also proposed a generalization of the original segmentation algorithm in which the shape prior is automatically adapted to the image to be segmented during the deformation process while still maintaining both the computational efficiency and the topology preservation of the method.

All in all, this approach is very promising and was proven to be both effective and efficient on the addressed clinical application. It also paves the way for numerous further investigations: for instance in the next chapter (more precisely in Section 7.4), we go beyond and learn not only the shape variability but also the local appearance of the organ of interest.

Chapter 7

Tagged models for implicit template deformation

Contents

Abstract

This chapter generalizes the implicit template deformation framework by introducing the notion of tagged models. A tagged model is a model in which different subregions are defined. In each of these subregions, a specific image-based term can be used. This framework therefore allows to greatly enrich the method by including more prior knowledge in the model. The tags can be either set manually or automatically learnt via a process that we describe in this chapter. Various applications are presented and validation is performed in the context of myocardium segmentation in US images.

Résumé

Ce chapitre généralise l'algorithme de déformation de modèle implicite en introduisant la notion de modèle étiqueté. Un modèle étiqueté est un modèle dans lequel sont définies plusieurs sous régions. A chaque sous-région est associée son propre terme image: différentes parties du modèle peuvent être attirées par différentes caractéristiques de l'image. Cette formulation permet donc d'enrichir considérablement le modèle en incluant encore plus d'information a priori. Les étiquettes peuvent être soit définies manuellement, soit apprises via un processus décrit dans ce chapitre. Plusieurs applications sont présentées et une validation est effectuée dans le contexte de la segmentation de myocarde dans des images US.

In this chapter, we present a method to extend the implicit template deformation algorithm to a multi-region formulation. The goal is to enrich the model used to segment an image by splitting the domain of this model in different areas with a particular image-based energy in each of them. We call *tagged model* a model with such a subdivision attached to it. Whereas a global appearance model was used in the standard template deformation framework, tagged models enable to exploit different image features depending on the regions thus defined. The proposed extension also allows to vary the strength of the shape prior according to such areas.

We first introduce the problem and review related work in Section 7.1. Then we describe the new formulation in Section 7.2 and apply it to the problem of liver and vessel segmentation in 3D ultrasound images. In Section 7.3, we show the benefits of combining it with the co-segmentation framework that was presented in Chapter 5. Finally, Section 7.4 presents a method to automatically learn the labeling of the model from a database and shows its benefits in the context of myocardium segmentation in US images.

7.1 Introduction

7.1.1 Motivation

In the standard implicit template deformation framework, a shape prior is used through an initial implicit function and two image-based functions r_{int} and r_{ext} are built to guide the segmentation as indicated in (2.41) from Section 2.7: the pixels $\mathbf{x}$ such that $r_{int}(\mathbf{x}) < r_{ext}(\mathbf{x})$ have a higher probability to be inside the target object. The interior and exterior priors are defined globally, which makes the segmentation of some images difficult (especially in cases where the contrast of the object with respect to its background is inverted, like in Figure 7.1). When the initialization of the segmentation is very close to the target object, it is possible to estimate r_{int} and r_{ext} locally and update their values along the segmentation process. However r_{int} and r_{ext} are necessarily defined in the referential of the image rather than the referential of the model ϕ_0. It is therefore impossible to include prior spatial information

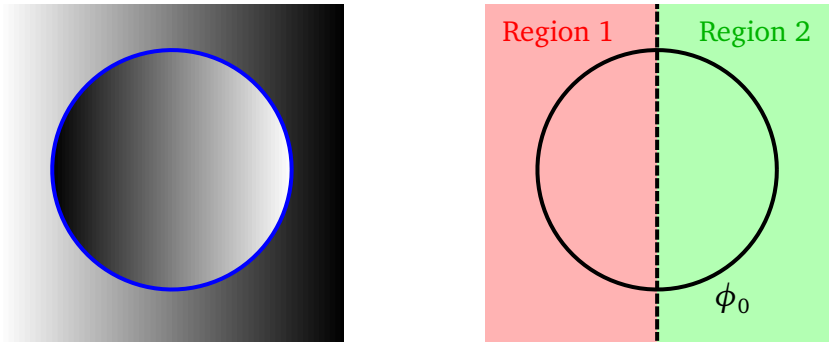


Figure 7.1: (Left) Example of an object (in blue) with a varying contrast: darker on the left, brighter on the right. We would like to use this information as a prior in the model. (Right) Tagged template ϕ_0 required to correctly segment the image, split into two regions in which different r_{int} and r_{ext} functions will be defined.

on these image-based terms. For instance in Figure 7.1, knowing *beforehand* that we are looking for a circle whose left part is darker but right part brighter would greatly ease the segmentation and alleviate the dependency on the initialization.

To enable the use of such priors in the model, we want to be able to label different regions of the template, and assign each of these areas different r_{int} and r_{ext} terms. In the right hand-side of Figure 7.1, we show how such regions should be defined on this synthetic case. We call such a model with regions labels a *tagged* model (or template). This approach has two main benefits:

- by defining different couples r_{int}/r_{ext}, we can exploit different image features, more dedicated to each part of the model;
- by varying the absolute value of $r_{int} - r_{ext}$, we can modify the trade-off between the image-based term and the regularization (shape prior) term. This enables us, for instance, to define some areas where the object/background contrast is low and the image information should be ignored (or at least less trusted than the model).

7.1.2 Related work

Most segmentation methods, even if they use a shape prior, assume that the appearance can be globally defined (either by a single intensity model or with edge-based terms). In this subsection, we report some of the methods that go beyond and propose a more elaborate appearance model, *i.e.* using different image features in particular regions of the model or spatially varying the strength of the prior. How such information is included in the segmentation obviously depends on the model itself.

In Active Shape Models [Cootes et al., 1995], shapes are represented via a set of landmarks and it is natural to learn a pointwise appearance prior, *e.g.* the intensity profile along the normal of each point [Van Ginneken et al., 2002; Spiegel et al., 2009]. Another more global approach is to learn the complete appearance of the object and its main variations as in the

Active Appearance Models [Cootes et al., 2001] or the method described in [Yang & Duncan, 2004]. The segmentation of an unseen image is then done by simultaneously searching for the object boundary and an approximation of its appearance by a linear combination of learnt modes, which tends to provide more robust results but is computationally much more expensive.

Level-set approaches were also extended to take into account variations in the appearance prior. In [Leventon et al., 2000b], the appearance model is a function of the distance to the shape boundary. While this allows a better representation than a global appearance prior, it is still too constrained for many medical applications. Cremers *et al.* proposed to introduce a labeling function (defined at each voxel) that balances the contribution of the data-fidelity term with respect to the regularization term [Cremers et al., 2003b]. Thus they are able to incorporate a notion of "trust" in arbitrary regions but they cannot vary the intensity model among such regions. Actually, this can be seen as a simplified case of the approach that we present in this chapter.

Finally, an extension of the MetaMorph [Huang & Metaxas, 2008] framework (one of the closest approaches to implicit template deformation) was proposed in [Huang et al., 2004]. This work aims at learning jointly the shape and the appearance of the organ to be segmented. However, this results in a mutual information registration problem that needs to be solved at test time, which would be too time-consuming in 3D for our target applications.

7.2 Tagged template deformation

In this part, we introduce and describe the notion of *tagged* template. Two different (but eventually very close) formulations are proposed, depending on the kind of image features used: region-based criteria in Section 7.2.1 and edge-based terms in Section 7.2.2. An application of the latter is presented in Section 7.2.3 in the context of liver segmentation in ultrasound images.

7.2.1 Generic formulation for regional terms

In the standard formulation, we defined two image-based classification error functions $r_{int} : \Omega \rightarrow \mathbb{R}$ and $r_{ext} : \Omega \rightarrow \mathbb{R}$, such that $r_{int}(\mathbf{x}) < r_{ext}(\mathbf{x})$ at locations $\mathbf{x}$ that have a higher probability to be inside the target object than outside it.

From now on, we assume that we have a set of N_r such functions: $(r_{int,k})_{k=1...N_r}$ and $(r_{ext,k})_{k=1...N_r}$. Instead of encoding the appearance of the whole target object, each of them can be specialized to describe a particular region of the structure to be segmented. For example, if we assume a mean intensity inside (resp. outside) two different regions $c_{int,1}$ and $c_{int,2}$ (resp. $c_{ext,1}$ and $c_{ext,2}$), then these terms could be $r_{int,1}(\mathbf{x}) = (I(\mathbf{x}) - c_{int,1})^2$, $r_{ext,1}(\mathbf{x}) = (I(\mathbf{x}) - c_{ext,1})^2$, $r_{int,2}(x) = (I(\mathbf{x}) - c_{int,2})^2$ and $r_{ext,2}(\mathbf{x}) = (I(\mathbf{x}) - c_{ext,2})^2$.

In order to use them adequately, we have to label different regions of the space. This is done through N_r different functions on $\Omega \rightarrow \mathbb{R}^+$, namely $(t_k)_{k=1...N_r}$, which will weight the contributions of each region terms: r_{int} and r_{ext} should be respectively replaced by

$$r_{int}(\mathbf{x}) \to \sum_{k=1}^{N_r} t_k(\mathbf{x})\ r_{int,k}(\mathbf{x}) = \left\langle \vec{t}(\mathbf{x}),\ \vec{r_{int}}(\mathbf{x}) \right\rangle \tag{7.1}$$

$$r_{ext}(\mathbf{x}) \to \sum_{k=1}^{N_r} t_k(\mathbf{x})\ r_{ext,k}(\mathbf{x}) = \left\langle \vec{t}(\mathbf{x}),\ \vec{r_{ext}}(\mathbf{x}) \right\rangle \tag{7.2}$$

with the vectorial notations:

$$\vec{t}(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) \\ \vdots \\ t_{N_r}(\mathbf{x}) \end{pmatrix},\ \vec{r_{int}}(\mathbf{x}) = \begin{pmatrix} r_{int,1}(\mathbf{x}) \\ \vdots \\ r_{int,N_r}(\mathbf{x}) \end{pmatrix} \text{ and } \vec{r_{ext}}(\mathbf{x}) = \begin{pmatrix} r_{ext,1}(\mathbf{x}) \\ \vdots \\ r_{ext,N_r}(\mathbf{x}) \end{pmatrix}$$

However, we do not want to label the space of the image (because we do not know beforehand which regions of the image correspond to different parts of the model), but the space of the model itself (in which we actually have prior information). Therefore we rather define N_r tags functions $T_k : \Omega_0 \to \mathbb{R}^+$ in the referential Ω_0 and warp them via the same transformation that is applied to the template ϕ_0, which means

$$\forall k \in [\![1, N_r]\!], \quad \vec{t} = \vec{T} \circ L \circ G \tag{7.3}$$

or in vectorial notation

$$\vec{t} = \vec{T} \circ L \circ G \tag{7.4}$$

The new implicit template deformation energy is then

$$\begin{aligned} E_{tags}(L, G) = &\int_\Omega H(\phi_0 \circ L \circ G(\mathbf{x})) \sum_{k=1}^{N_r} \left[T_k \circ L \circ G(\mathbf{x}) \left(r_{int,k}(\mathbf{x}) - r_{ext,k}(\mathbf{x}) \right) \right] d\mathbf{x} \\ &+ \int_\Omega \sum_{k=1}^{N_r} T_k \circ L \circ G(\mathbf{x})\ r_{ext,k}(\mathbf{x})\ d\mathbf{x} \\ &+ \frac{\lambda}{2} \|L - \mathbf{Id}\|_U^2 \end{aligned} \tag{7.5}$$

In each voxel $\mathbf{x}$, the norm of the vector $\vec{T}(\mathbf{x})$ indicates the general confidence in the data while tuning its repartition along the different entries of the vector allows to favour particular features with respect to others.

Remark 7.1. *When there is only one couple r_{int}/r_{ext} and $T_1 \equiv 1$ everywhere, we retrieve the standard implicit template deformation formulation.*

Note that the second term now depends on L and G and thus cannot be dismissed anymore during the optimization. This will induce an additional complexity that we discuss in the next subsection.

Numerical details

The energy in (7.5) is minimized with a gradient descent scheme. The derivative of E_{tags} with respect to a parameter p_i of the pose G reads

$$\begin{aligned}\nabla_{\mathbf{p}_i} E_{tags} = |J_{G^{-1}}| \Bigg(& \int_\Omega \delta(\phi_0 \circ L) \left\langle \nabla\phi_0 \circ L \ , \ \frac{\partial}{\partial \mathbf{p}_i}(L \circ G) \circ G^{-1} \right\rangle \left\langle \vec{T} \circ L, \ (\vec{r_{int}} - \vec{r_{ext}}) \circ G^{-1} \right\rangle \\ & + \int_\Omega H(\phi_0 \circ L) \left\langle D\vec{T} \circ L \ . \ \frac{\partial}{\partial p_i}(L \circ G) \circ G^{-1}, \ (\vec{r_{int}} - \vec{r_{ext}}) \circ G^{-1} \right\rangle \\ & + \int_\Omega \left\langle D\vec{T} \circ L \ . \ \frac{\partial}{\partial p_i}(L \circ G) \circ G^{-1}, \ \vec{r_{ext}} \circ G^{-1} \right\rangle \Bigg) \end{aligned} \tag{7.6}$$

while its derivative with respect to the deformation field is

$$\begin{aligned}\nabla_{\mathbf{u}} E_{tags} = \ |J_{G^{-1}}| \ K_\sigma * \Big(& \delta(\phi_0 \circ L) \left\langle \vec{T} \circ L, \ (\vec{r_{int}} - \vec{r_{ext}}) \circ G^{-1} \right\rangle \nabla\phi_0 \circ L \\ & + H(\phi_0 \circ L) \ (D\vec{T} \circ L)^T \ (\vec{r_{int}} - \vec{r_{ext}}) \circ G^{-1} \\ & + (D\vec{T} \circ L)^T \ \vec{r_{ext}} \circ G^{-1} \Big) . \end{aligned} \tag{7.7}$$

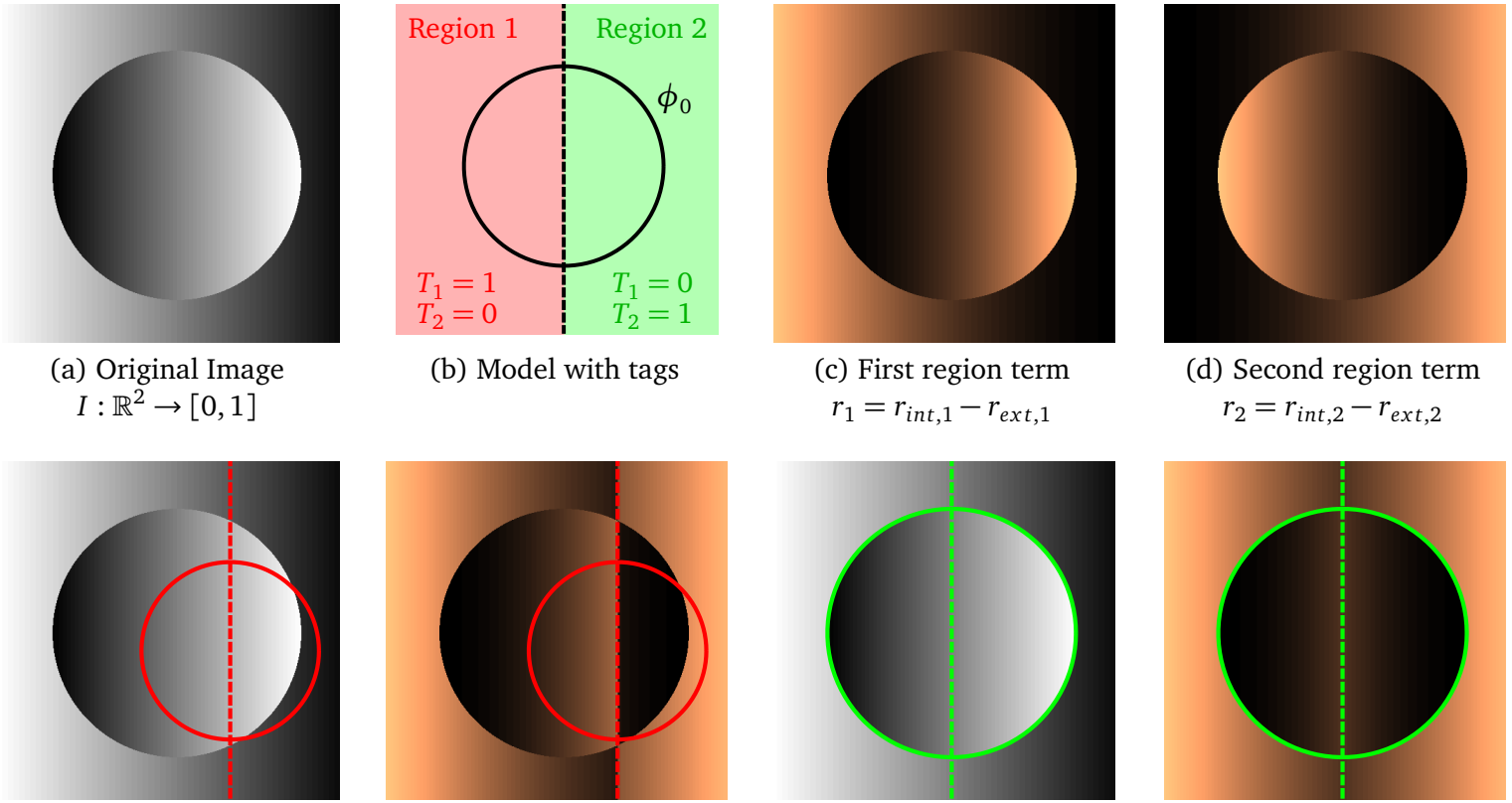
Such expressions require more computations than the standard template deformation framework. For instance in (7.6), two new integral terms (with a large support since their integrand is not multiplied by $\delta(\phi_0 \circ L \circ G)$) appear and potentially decrease the computational efficiency of the method. However, we point out that they both depend on the derivative of the vector fo tags functions $\vec{T}$, on which we have some control. If we choose the functions $\vec{T}$ adequately (for instance piecewise constant), it can have a very sparse gradient. Such knowledge may be used to speed up the computations.

Figure 7.2 describes the segmentation of a synthetic image with a tagged model deformation. The two couples $r_{int,1}/r_{ext,1}$ (Figure 7.2.b) and $r_{int,2}/r_{ext,2}$ (Figure 7.2.c) cannot independently characterize the global object appearance with respect to the background. However, by using a tagged model, it is possible to build a composite image-based term that correctly describes the interior of the object of interest (Figure 7.2.f).

7.2.2 Formulation for edge-based terms

In a number of applications, we are mainly interested to specify the contrast that the object should have compared to its neighborhood. As described in Remark 2.1 from Section 2.4.3, the image-based criterion is then the gradient flux through the segmentation boundary, which is closely related to the Laplacian of the image [Kimmel & Bruckstein, 2003]. By setting the image-based term r_{int} to either ΔI or $-\Delta I$ and r_{ext} to 0, it is possible to use such prior in the implicit template deformation framework. In the previous chapters of this thesis, we used several times this approach.

However, it was not feasible to specify in each part of the model a particular direction of the gradient (*i.e.* whether a part is darker or brighter than its surrounding). The goal of this



(a) Original Image $I : \mathbb{R}^2 \to [0,1]$

(b) Model with tags

(c) First region term $r_1 = r_{int,1} - r_{ext,1}$

(d) Second region term $r_2 = r_{int,2} - r_{ext,2}$

(e) Segmentation and tags boundary (initialization)

(f) Tagged image-based term $T_1\ r_1 + T_2\ r_2$ (initialization)

(g) Segmentation and tags boundary (at convergence)

(h) Tagged image-based term $T_1 \circ L \circ G\ r_1 + T_2 \circ L \circ G\ r_2$ (at convergence)

Figure 7.2: Tagged template deformation on the synthetic image (a) with the tagged model (b). (c) First image-based term $r_{int,1} - r_{ext,1} = I^2 - (I-1)^2$. (d) Second image-based term $r_{int,2} - r_{ext,2} = (I-1)^2 - I^2$. (d) Initialization of the model (plain line) and tags separation (dotted line). (f) Composite image-based term using tags at initialization. (g) Segmentation (plain line) and tags separation (dotted line) at convergence. (h) Composite image-based term using tags at convergence. Black (resp. copper) indicates low (resp. high) values.

section is to use tagged models to encode such an information.

We assume that two tags functions $T_1 : \Omega_0 \to \mathbb{R}^+$ and $T_2 : \Omega_0 \to \mathbb{R}^+$ are attached to the model ϕ_0. These two functions defines locations where the model should be brighter (respectively darker) than the background. The gradient-flux criterion of 2.27 with tags then becomes

$$\begin{aligned} E_{edge}(L,G) = &\int_{(\phi_0 \circ L \circ G)^{-1}(0)} T_1 \circ L \circ G(\mathbf{x}) \, \langle \vec{\nabla} I(\mathbf{x}) \, , \vec{\mathbf{n}}(\mathbf{x}) \rangle \, d\mathbf{x} \\ &- \int_{(\phi_0 \circ L \circ G)^{-1}(0)} T_2 \circ L \circ G(\mathbf{x}) \, \langle \vec{\nabla} I(\mathbf{x}) \, , \vec{\mathbf{n}}(\mathbf{x}) \rangle \, d\mathbf{x} \end{aligned} \tag{7.8}$$

Note that this energy only depends on the difference $T_1 - T_2$ that we represent as a single function $T : \Omega_0 \to \mathbb{R}$ (with both positive and negative values). The energy then reads

$$E_{edge}(L,G) = \int_{(\phi_0 \circ L \circ G)^{-1}(0)} T \circ L \circ G(\mathbf{x}) \, \langle \vec{\nabla} I(\mathbf{x}) \, , \vec{\mathbf{n}}(\mathbf{x}) \rangle \, d\mathbf{x} \tag{7.9}$$

In order to obtain a more convenient expression for the subsequent calculus of variation, we apply the divergence theorem which yields the following expression:

$$E_{edge}(L,G) = \int_{\Omega} H(\phi_0 \circ L \circ G(\mathbf{x})) \, . \, div\left(T \circ L \circ G(\mathbf{x}) \, \vec{\nabla} I(\mathbf{x})\right) \, d\mathbf{x} \tag{7.10}$$

Recalling that for any $a \in \mathbb{R}$ and $\vec{u} \in \mathbb{R}^d$,

$$div(a\vec{u}) \; = \; a \, div(\vec{u}) \; + \; \left\langle \vec{\nabla} a, \vec{u} \right\rangle, \tag{7.11}$$

we can rewrite the energy as

$$E_{edge}(L,G) = \int_{\Omega} H(\phi_0 \circ L \circ G(\mathbf{x})) \, \left(T \circ L \circ G(\mathbf{x}) \, \Delta I(\mathbf{x}) \; + \; \left\langle \vec{\nabla} T \circ L \circ G(\mathbf{x}), \vec{\nabla} I(\mathbf{x}) \right\rangle \right) \, d\mathbf{x} \tag{7.12}$$

whose integrand, unlike the standard template deformation (without tags), is composed of two terms.

In (7.12), the first term falls exactly the region-based framework presented in the previous section. It corresponds to using either ΔI or $-\Delta I$ as r_{int} in different parts of the model and setting r_{ext} to 0 everywhere; its interpretation is therefore simple to grasp. The segmentation is mostly driven by this first term. On the other hand, the second term usually has a lower influence and acts as a correction term that takes into account the impact of the tags variation on the energy. Although it was omitted in [Kimmel & Bruckstein, 2003], it may be useful to obtain a really precise segmentation at the boundaries of the subregions defined by the tags (*i.e.* where the gradient of T is not null). We discuss here below its implications in terms of computational cost.

Remark 7.2. *Contrary to the region-based approach of tagged models (see Section 7.2.1), this edge-based formulation only depends on the values of the tags on the zero level-set of the template ϕ_0. While this is easy to see in (7.8), it might seem paradoxical when looking at (7.12) because of the second term including $\vec{\nabla} T$. Actually, this term only cancels out the artificial dependency on T outside the template boundary introduced by the first term.*

Numerical details

The derivative of E_{edge} with respect to a parameter p_i of the pose G reads

$$\begin{aligned}\nabla_{\mathbf{p}_i} E_{edge} = \int_\Omega \delta(\phi_0 \circ L \circ G) \left\langle \nabla\phi_0 \circ L \circ G \ , \ \frac{\partial}{\partial p_i}(L \circ G) \right\rangle \\ \big(T \circ L \circ G \ . \ \Delta I \ + \ \langle DT \circ L \circ G, \vec{\nabla} I \rangle\big) \\ + \int_\Omega H(\phi_0 \circ L \circ G) \left(\left\langle DT \circ L \circ G \ , \frac{\partial}{\partial \mathbf{p}_i}(L \circ G) \right\rangle \Delta I + \right. \\ \left. \left\langle D^2 T \circ L \circ G \ . \frac{\partial}{\partial p_i}(L \circ G), \vec{\nabla} I \right\rangle \right)\end{aligned} \tag{7.13}$$

while its derivative with respect to the deformation field is

$$\begin{aligned}\nabla_{\mathbf{u}} E_{edge} = \int_\Omega \delta(\phi_0 \circ L \circ G) \quad \nabla\phi_0 \circ L \circ G \big(T \circ L \circ G \ . \ \Delta I \ + \ \langle DT \circ L \circ G, \vec{\nabla} I \rangle\big) \\ + \int_\Omega H(\phi_0 \circ L \circ G) \qquad \big(DT \circ L \circ G \ \Delta I \ + \ D^2 T \circ L \circ G \ \vec{\nabla} I\big) \, .\end{aligned} \tag{7.14}$$

In each of these gradient equations, two additional terms have to be computed compared to (7.6) and (7.7); on the upside though, the terms with r_{ext} vanish.

The first new term is the integral of $\delta(\phi_0 \circ L \circ G)\langle \vec{\nabla} T \circ L \circ G, \vec{\nabla} I \rangle$. This function has a very small support as it is the intersection of the zero level-set of the deformed implicit function and the set of points where the tag T vary. It is therefore very fast to estimate (and almost negligible anyway).

On the other hand, the two other new integrands are defined in the whole volume represented by $H(\phi_0 \circ L \circ G)$ and thus potentially represent a computational burden. However, we once again point out that they both depend on the derivative (first or second order) of the tag function T. If defined adequately (*i.e.* with a sparse gradient), we will be able to compute these terms from a small number of contributions.

For all these reasons, the efficiency of the algorithm is not really lost. An application of this approach is presented in the next subsection.

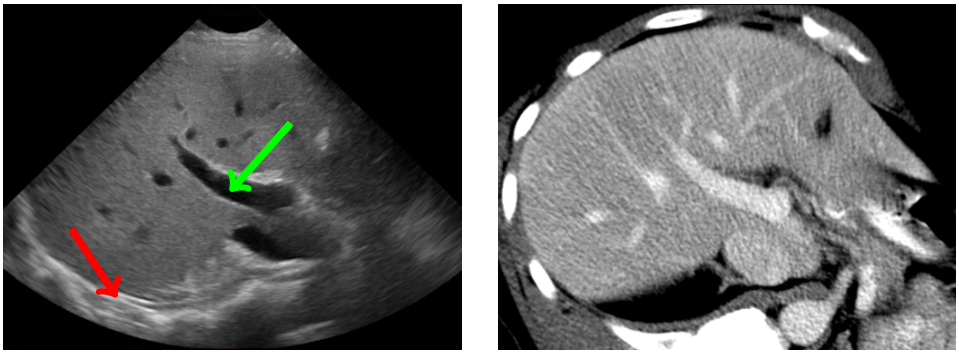


Figure 7.3: Aligned images of the liver and its inner vessels in US (left) and contrast-enhanced CT (right). In ultrasound, the liver has a varying contrast: it is brighter than the blood vessel (green arrow) but darker than the diaphragm (red).

7.2.3 Application: Liver and vessel segmentation in 3D US images

In this subsection, we consider the following clinical setting. A patient suffers from a pathology related to the liver (for example a tumour) which was diagnosed using a contrast-enhanced CT acquisition. The recommended treatment is a minimally invasive intervention, taking place in a room equipped with an ultrasound system (that is used to monitor the intervention in real time) but no CT scanner. However, the information shown in US images is somehow limited: the field of view is narrow (the liver does not fit in a single image) and the image quality is usually poor. During such an intervention, it would be extremely valuable to be able to display at all times some pre-op CT annotated data (*e.g.* tumor shape and location) aligned with the ultrasound data (see Figure 7.3).

Here we use the method described in Section 7.2.2 to find an alignment transformation between the pre-op CT and the US image, which is indeed an active field of research [Wein et al., 2008]. The proposed approach is based on liver segmentation by template deformation. The key idea is to transform the liver from the CT image so that it segments correctly the US image. The estimated transformation can then be propagated to the CT image and the annotated data, in order to embed them in the US display.

The liver is not a rigid organ so it may undergo severe deformations depending on the patient's position, breathing state or even the pressure of the transducer on the patient's skin. The ultrasound data is not unequivocal enough to capture such deformations. Relying solely on the liver boundary is therefore not robust enough and additional landmarks have to be taken into account. The liver vascular trees are good candidates since they are visible in both modalities.

Thanks to the implicit representation of shapes, we can design a template that includes both the liver and its vascular tree. Nevertheless, it remains difficult to build beforehand a single couple r_{int}/r_{ext} of image-based classification function: as illustrated in Figure 7.3, the liver has a varying contrast with respect to its surrounding. It is indeed brighter than the blood vessels but darker than the diaphragm. The different parts of the liver/vessel model should be attracted to different image features, hence the usefulness of tagging the model.

Application of the tagged implicit template deformation

For each patient, both an US and a CT volume were available. The model was built using segmentations automatically extracted from the CT image. The liver model ϕ_l was computed using the method from [Gauriau et al., 2013] while the vessels ϕ_v were extracted via an approach similar to [Bauer et al., 2010]. In order to use simultaneously the liver boundary and the inner vessels, we set ϕ_0 as a composite model:

$$\phi_0(\mathbf{x}) = \max(\phi_l(\mathbf{x}), -\phi_v(\mathbf{x})) \,. \tag{7.15}$$

Figures 7.4.a-c illustrate the construction of this extruded model.

We followed the approach of Section 7.2.2 to define the image-based term and the tags: for the intensity model inside the liver in US images is rather complex, we choose to rely on edge information. Indeed, the diaphragm (which is adjacent to the liver) generates a very bright signal, while the main vessels of the liver also have a high contrast (but they are darker than their surroundings). We therefore defined the classification term as follows

$$r_{int}(\mathbf{x}) = \Delta I_\sigma(\mathbf{x}) \qquad \text{and} \qquad r_{ext}(\mathbf{x}) = 0 \tag{7.16}$$

and a single tag function $T : \Omega_0 \to \{-1, 1\}$ is used to only indicate the sign of the Laplacian. As indicated in Figure 7.4.d, near the vessels the model should be brighter than its exterior so the tag is set to -1. Conversely, the liver outer boundary should be attracted to brighter structures so the tag should be 1.

In order to automatically initialize the position of the model ϕ_0, the vessel segmentation algorithm is also executed on the US image. The two segmentations (CT and US) are then globally registered by analyzing their main branches and bifurcations. This provides a good estimate of the transformation between the CT and the US image, and can serve as an initialization for G.

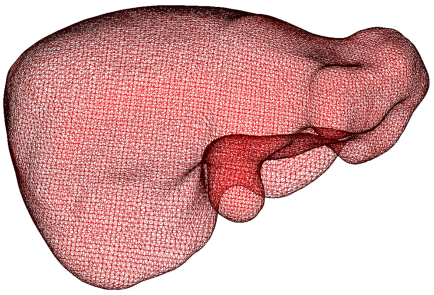
Since we are interested in globally aligning the CT and US, we only optimize the global transformation G. This allows to both make the algorithm more robust and greatly decrease the computational time[1]. Because of the potentially severe deformations the liver may undergo between the two acquisitions, the pose G is sought as an affine transformation. In other words, the deformation L is fixed to the Identity while we run a gradient descent on the parameters of G.

Experiments

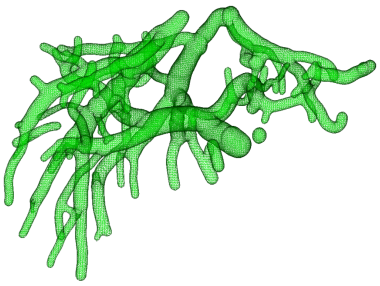
We present hereafter qualitative results on two different cases to show the benefits of the proposed approach.

The first case is illustrated in Figure 7.5. In the top row, the initialization (in orange) is not far from the expected result. However, the liver model is too large: because of the

[1] Due to the choice of the complex template with small vessels, the resolution has to be very high, hence an increased computational burden.

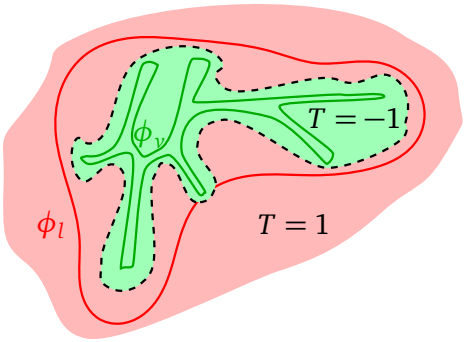


(a) Liver template ϕ_l



(b) Vessel template ϕ_v

(c) Composite model ϕ_0



(d) Tags definition

Figure 7.4: Construction of the model ϕ_0 for liver segmentation in US images (c) composed of a liver model (a) and a vessel model (b) that both come from automated segmentation in a CT image of the same patient. Tags are defined according to the scheme in (d).

transducer pressure, the liver is compressed (note that the template goes beyond the skin limit at the top of the image). Furthermore, it does not fit exactly against the diaphragm. If the model only includes the liver boundary (middle row), the segmentation matches the main edges (such as the diaphragm) but still leaks in the top of the image. However, by using a tagged model (bottom row) which includes both the liver boundary and the vessels, we are able to retrieve a more satisfying segmentation. Blood vessels anchor the model and prevent it from drifting away while still allowing it to capture boundary information.

Figure 7.6 shows the results obtained on a more challenging dataset. Indeed the estimation of the transformation to initialize the algorithm is less accurate and the vessel tree is not as dense as in the first case. The segmentation using the standard algorithm with solely the liver boundary is presented in the middle row. Most strong edges are matched, especially the tip of the liver has been dragged down. However, the estimated transformation is not realistic and deformed too much the liver (note in the 3D visualization how the shape was compressed). Even if the segmentation seems reasonable where the boundary is clear, the model is likely to extrapolate poorly far away. As demonstrated in the bottom row, the vessels (though not abundant) once again act as powerful landmarks and the mixed model yields a better segmentation.

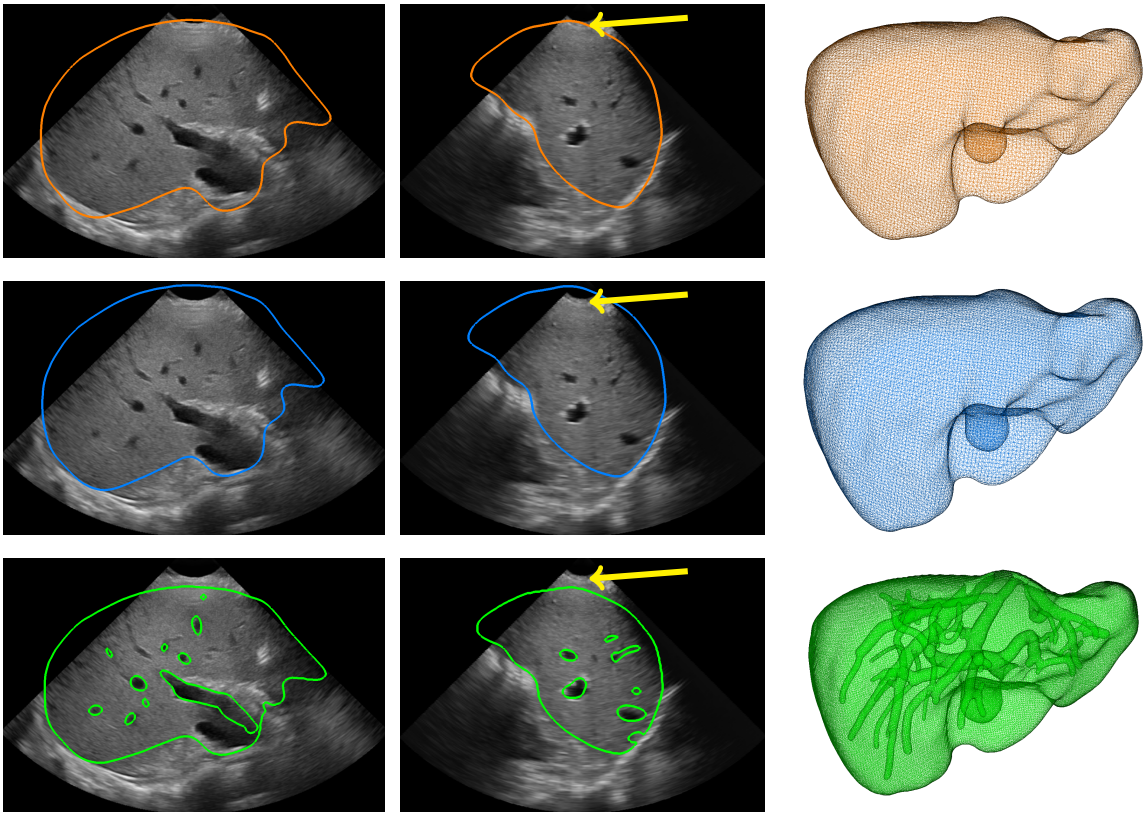


Figure 7.5: Liver tagged segmentation in US images on a first case. The first two columns show two orthogonal slices while the third one presents a 3D visualization of the segmentation. (Top) Initialization obtained from the segmentation of a CT image. (Middle) Segmentation obtained only with liver boundary. (Bottom) Segmentation obtained with the mixed (liver boundary and inner vessels) tagged model.

7.3 Coupling tagged models and co-segmentation

The different enhancements of the implicit template deformation algorithm presented throughout this thesis are independent and can therefore be combined. In this section, we propose an approach to segment a given object in multiple images as in Chapter 5. The difference is that, thanks to the tagged template deformation introduced in Section 7.2, we are able to specify which parts of the model should be attracted by an image and which parts by another one. We show a proof of this concept for a new clinical application, namely the assessment of *abdominal aortic aneurysms* in both conventional and contrast-enhanced ultrasound 3D images.

7.3.1 Tagged co-segmentation with implicit template deformation

The aim is to segment a particular structure from a shape prior ϕ_0. Two images I_1 and I_2 are available[1] and some parts of the model are known to be more visible in one particular

[1]The extension to a higher number of images is straightforward though.

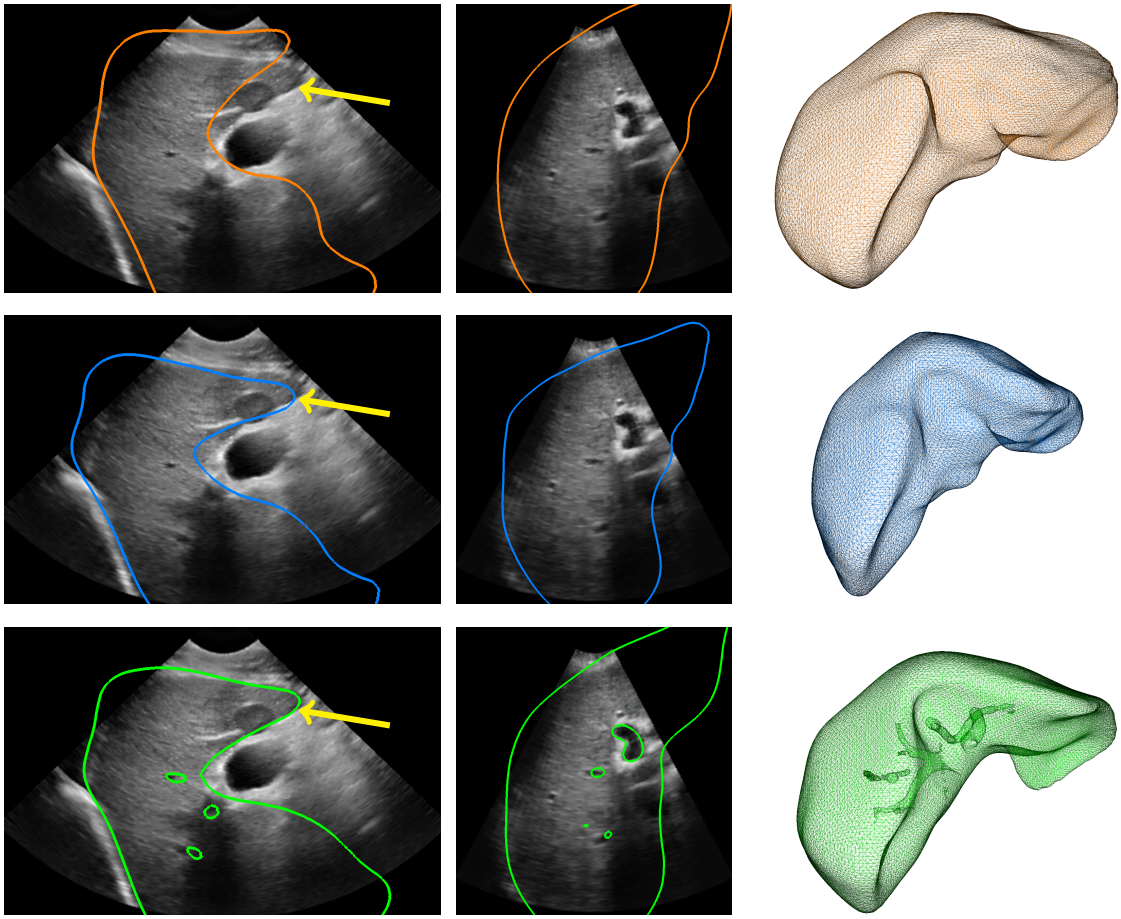


Figure 7.6: Liver tagged segmentation in US images on a second case. The first two columns show two orthogonal slices while the third one presents a 3D visualization of the segmentation. (Top) Initialization obtained from the segmentation of a CT image. (Middle) Segmentation obtained only with liver boundary. (Bottom) Segmentation obtained with the mixed (liver boundary and inner vessels) tagged model.

image. Such information will be incorporated via the tags framework presented earlier in this chapter.

We thus define $T_1 : \Omega_0 \to [0,1]$ a *tag function* defined over the domain of the model ϕ_0 that indicates the confidence that we grant to the first image. For instance, $T_1(\mathbf{x}) = 1$ if the first image helps computing the transformation applied to ϕ_0 at point $\mathbf{x}$ because this part of the model is "visible". Conversely, $T_1(\mathbf{x}) = 0$ when the first image should not be taken into account (lacking or misleading information). We also define a similar function $T_2 : \Omega_0 \to [0,1]$ for the second image.

Finally, we still assume that the two images are not necessarily aligned and the motion between their acquisition should be accounted for. As in Section 5.1, we denote this transformation G_r. The joint co-segmentation and registration framework can then be extended by including the new weighting terms T_1 and T_2. This yields the following energy:

$$
\begin{aligned}
E_{tags-co-seg}(L, G, G_r) = &+ \frac{1}{2}\int_{\Omega} H(\phi_0 \circ L \circ G(\mathbf{x}))\; T_1 \circ L \circ G(\mathbf{x})\; r_{I_1}(\mathbf{x})\; d\mathbf{x} \\
&+ \frac{1}{2}\int_{\Omega} H(\phi_0 \circ L \circ G \circ G_r(\mathbf{x}))\; T_2 \circ L \circ G \circ G_r(\mathbf{x})\; r_{I_2}(\mathbf{x})\; d\mathbf{x} \\
&+ \frac{\lambda}{2}\|L - Id\|_U^2 \quad . \qquad (7.17)
\end{aligned}
$$

The tags T_1 and T_2 are transported in each image domain via $L \circ G$ and $L \circ G \circ G_r$ respectively, and they are used to favour the contribution from one image or the other, depending on the domain parts.

This represents a generalization of both formulations (5.13) and (7.5). For example, (5.13) is retrieved when $T_1 \equiv 1$ and $T_2 \equiv 1$, while (7.5) is retrieved with $T_2 \equiv 0$. The discussion of Section 7.2.1 on the computational details still holds, except the new dependency on G_r that is introduced.

Remark 7.3. *The idea of using tags with co-segmentation is very generic and could be applied to other algorithms than implicit template deformation.*

In the next subsection, we present some early results derived from this new formulation in the clinical context of abdominal aortic aneurysms in US and CEUS images.

7.3.2 Application: Abdominal aortic aneurysms segmentation in 3D US and 3D CEUS images

Clinical problem

Abdominal aortic aneurysms, also known as AAA, are the most common form of aortic aneurysms. They are clinically defined as a local dilation (of more than 50%) of the abdominal aorta and are very often situated at the level or below the kidneys. Apart from occasional pain, the aneurysms in themselves do not cause any particular symptom. Yet they present a very high risk of complication: because of the swelling, the vessel wall loses its resistance and may rupture. In such cases, a large quantity of blood leaks into the abdominal cavity which often leads to the patient's death in several minutes. This pathology draws a number of research studies as its prevalence reaches 5% in elderly subjects [Lindholt et al., 2008] while its mortality risk goes up to 80% (in case of rupture).

When an AAA is found, the patient therefore undergoes a close surveillance and the size of the aneurysm should be regularly reported. To monitor its evolution, clinicians resort to medical imaging. While CT has been considered the gold-standard modality, the repeated contrast injections and radiation absorptions may harm the patient. Conversely, recent studies tend to prove the accuracy of the measures performed on ultrasound images that do not present such drawbacks [Long et al., 2012]. Figure 7.7 shows both a conventional and a contrast-enhanced ultrasound acquisition of an aneurysm. In the first image, one can see

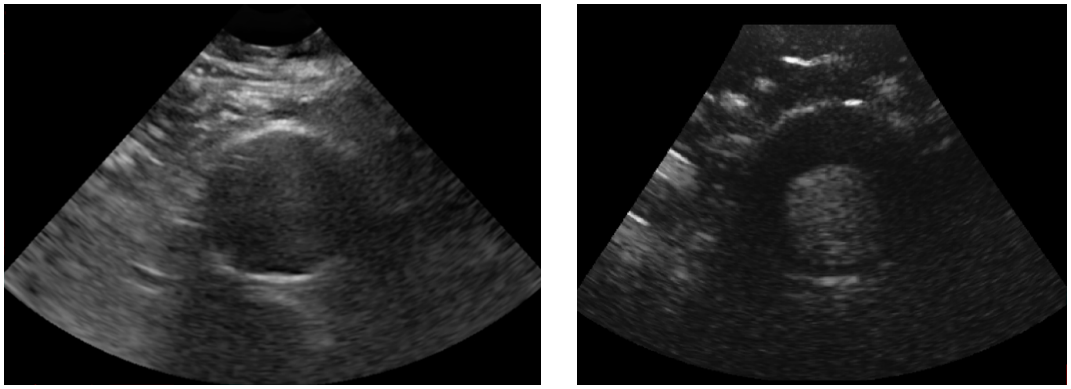


Figure 7.7: Slices from ultrasound volumes of an abdominal aortic aneurysm. The vessel boundary is visible in the conventional US image (left) while the actual lumen is visible in the contrast-enhanced image (right).

clearly the vessel exterior wall but not its inner structure (the lumen). Conversely, in the second image the lumen of the vessel is visible[1] unlike the outer boundary.

To assess the severity of the aneurysm, it would be interesting to measure the width of the vessel wall. One may segment sequentially the vessel boundary in the US volume and the lumen in the CEUS volume. However, there is no guarantee that these images are aligned and this is too strong an assumption: a registration of the two volumes is compulsory. Just like with the kidney segmentation (see Section 5.2), the registration helps the segmentation and vice versa. This motivates our tagged co-segmentation framework.

Application of the tagged co-segmentation

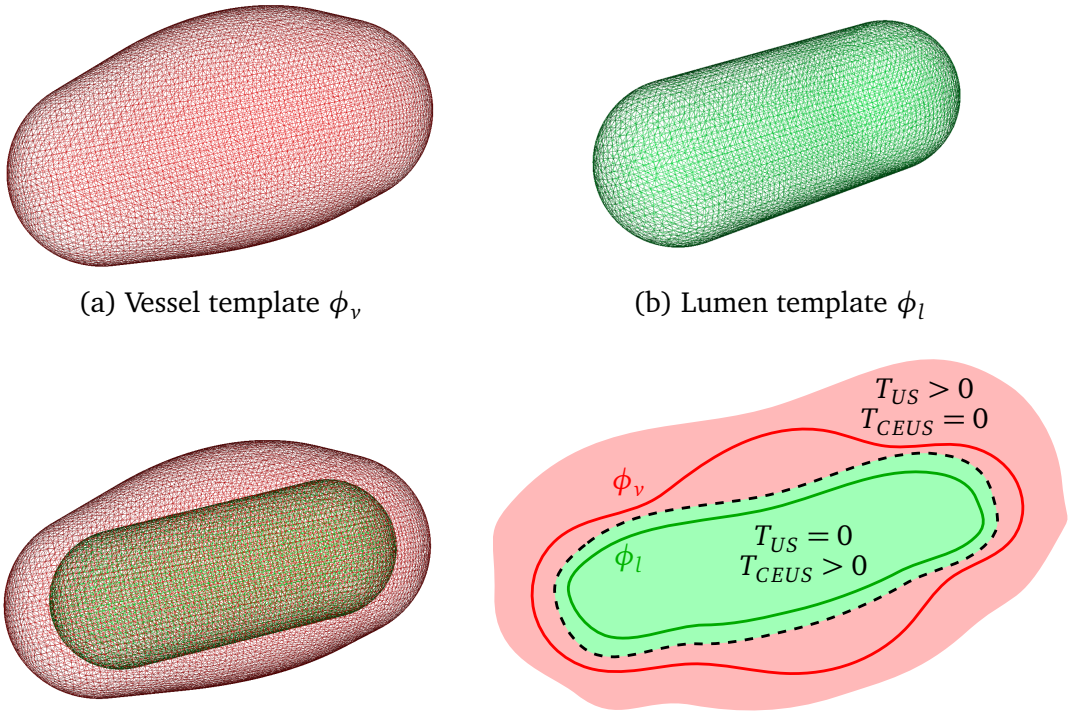
To apply the template deformation algorithm, we must first choose the initial template ϕ_0. As depicted in Figure 7.8, ϕ_0 is constructed as the difference of two *generalized cylinders* [Ulupinar & Nevatia, 1995]. We ask the user to click on several points on the centerline of the aorta in the US images and to associate a radius to each of these points. This allows to build a rough estimate of the vessel outer boundary whose implicit function is denoted ϕ_v (see Figure 7.8.a). To build the model for the lumen, we repeat the same process but only consider the extremal points with a reduced radius (see Figure 7.8.b), yielding the implicit function ϕ_l. The template ϕ_0 is finally defined as

$$\phi_0(\mathbf{x}) = \max(\phi_v(\mathbf{x}), -\phi_l(\mathbf{x})) \tag{7.18}$$

which is the vessel model extruded from the lumen model (see Figure 7.8.c). There are therefore two separate components in the zero level-set of ϕ_0, which leads to a natural definition of the tagged regions as illustrated in Figure 7.8.d.

To define the image-based terms, we rely on the fact that in both US and CEUS images the vessel interior is darker than its exterior (*i.e.* the vessel surroundings in US and the lumen

[1] Thanks to the contrast injection, what one sees in the CEUS image is exactly where the blood is.



(a) Vessel template ϕ_v (b) Lumen template ϕ_l

(c) Composite model ϕ_0 (d) Tags definition

Figure 7.8: Construction of the model ϕ_0 for abdominal aortic aneurysm (c) composed of a generalized cylinder for the vessel wall (a) and another one for the lumen (b). Tags are defined according to the scheme in (d).

in CEUS). Consequently, we set the image-based terms to the Laplacian of the images at a given scale and with a given sign:

$$\begin{array}{lll} r_{int}^{US}(\mathbf{x}) = \Delta I_{US,\sigma}(\mathbf{x}) & \text{and} & r_{ext}^{US}(\mathbf{x}) = 0 \\ r_{int}^{CEUS}(\mathbf{x}) = \Delta I_{CEUS,\sigma}(\mathbf{x}) & \text{and} & r_{ext}^{CEUS}(\mathbf{x}) = 0 \end{array} \tag{7.19}$$

Finally, the registration transform G_r is initialized to the Identity (we thus assume that neither the probe or the patient have moved significantly in the time interval between the two acquisitions). We then minimize the energy in 7.17 by gradient descent with respect to the deformation L and the transformations G and G_r G_r.

Results on the dataset from Figure 7.7 are reported hereafter. The segmentation in both images is presented in Figure 7.9. The initial model ϕ_0 has undergone a deformation to adapt to both images: note how the outer boundary corresponds to the US image while the inner boundary correctly segments the lumen visible in the CEUS image. The 3D visualization allows to compute useful statistics to assess the severity of the aneurysm. Note that the optimal transformation G_r can be used to register the US image to the CEUS one, which is shown in Figure 7.10. Note how the vessel boundaries (also a bit visible in the CEUS image for this case) are aligned. This paves the way for image fusion techniques that would be extremely valuable for the clinician.

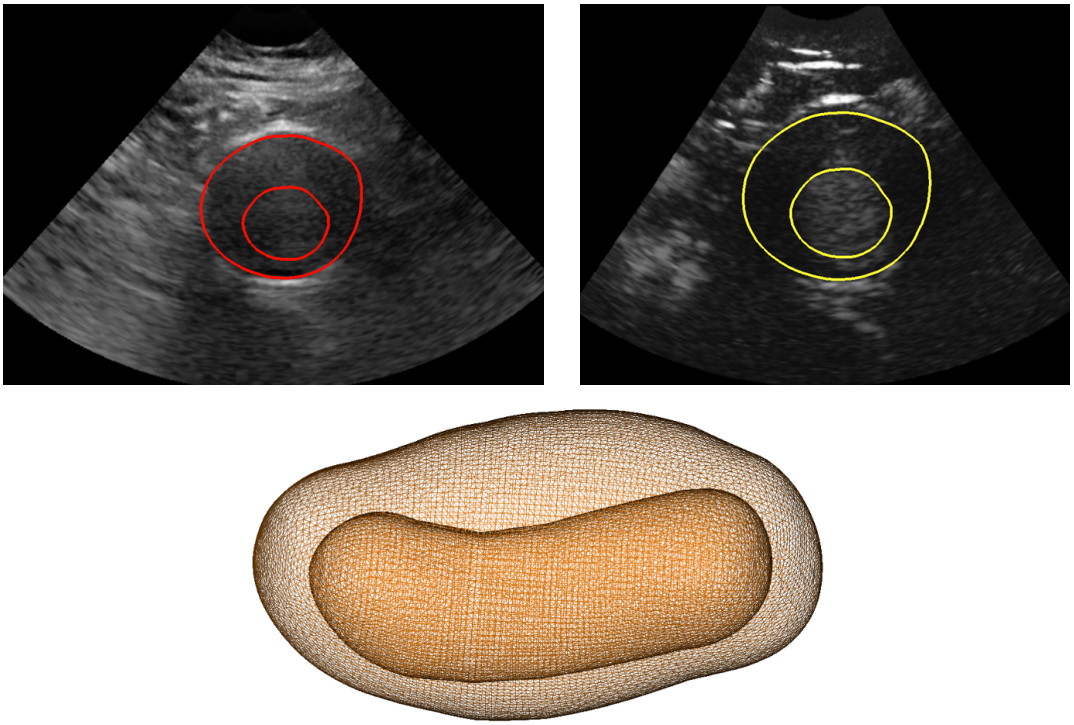


Figure 7.9: Tagged co-segmentation of the vessel wall and its lumen in US (left) and CEUS (right) volumes. (Bottom) Deformed model $\phi_0 \circ L$, visualized in 3D.

We recall that this is only a proof of concept and a more elaborate validation on different cases is deferred as future work.

7.4 Tags learning. Application: Myocardium segmentation in 2D US images

In all the applications presented so far, since the considered models were composite, the definition of the tags was quite natural and could be easily done *ad hoc*. In some applications however, it is not so easy to define them and when an image database is available, we would like to leverage it and learn the relevant subregions. This is the purpose of this section which describes a learning process to automatically uncover such regions. We present an application in the context of myocardium segmentation in US 2D images.

Remark 7.4. *In this section we focus on the tags learning for the edge-based formulation (see Section 7.2.2), i.e. we try to find which contrast there is (if any) at each part of the model. However, it would have been also possible to learn tags for the region-based formulation of Section 7.2.1.*

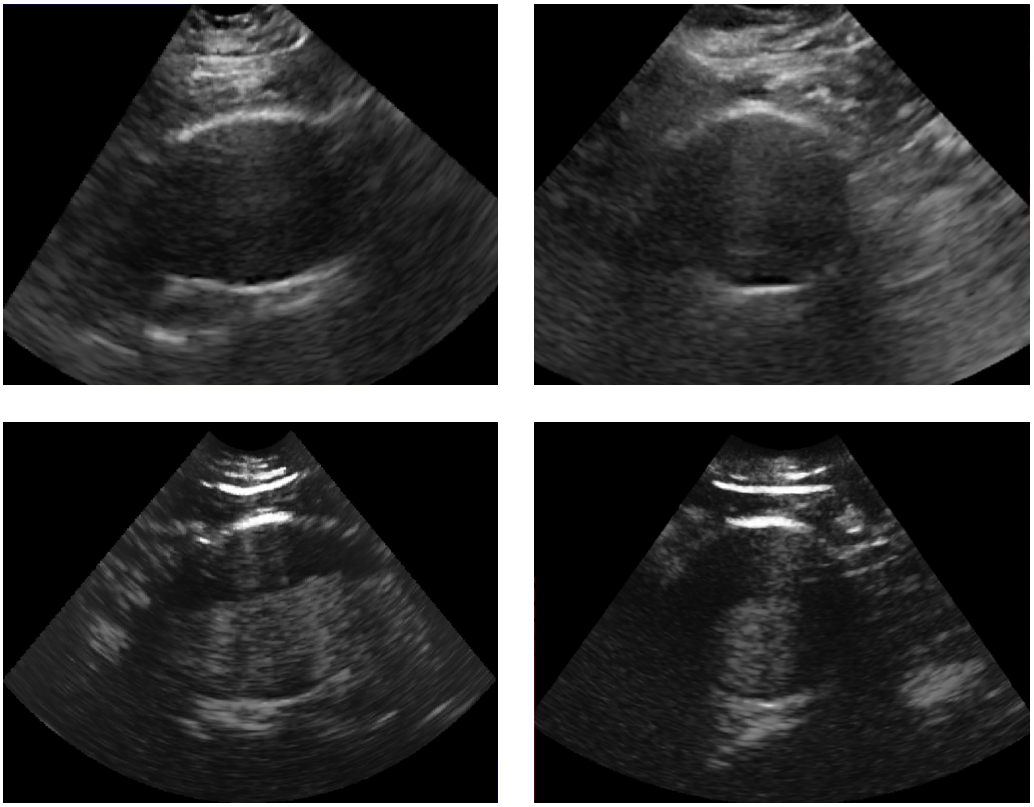


Figure 7.10: Two views of the US (top) and CEUS (bottom) volumes registered images based on the proposed co-segmentation.

7.4.1 Clinical problem

Ultrasound imaging is widely used to diagnose heart problems and understand heart diseases. The aim of an echocardiography exam is to provide the cardiologist both qualitative information (*e.g.* visualization of the heart motion) but also quantitative measures (*e.g.* strain, ejection fraction). Such statistics are often based on a manual delineation of the myocardium, and therefore sensitive to the high inter- and intra-operator variability. Automation is needed for consistency and reproducibility. Myocardium segmentation in US images is indeed currently an active field of research [Dietenbeck et al., 2012; Qin et al., 2013].

In this section, we are interested in segmenting the myocardium around the left ventricle, as indicated in Figure 7.11. The difficulty comes from the varying appearance of the myocardium. In Figure 7.11, we show some identifiable regions of the myocardium with different appearances. Such regions are consistent across different images; consequently, prior knowledge about them should be incorporated into the segmentation model, hence our tagged template deformation framework.

Although we could try to build manually the subdivision of the model, we will propose in this section to learn it automatically from a database of pre-segmented images.

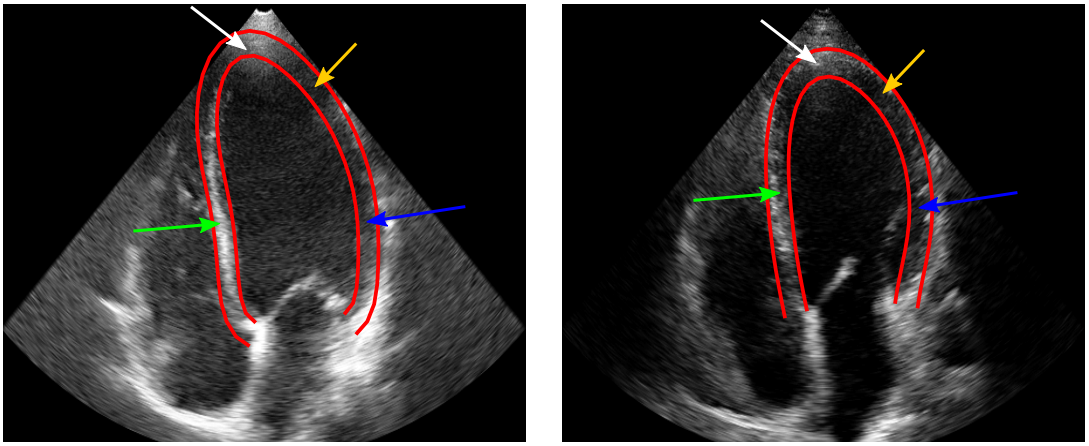


Figure 7.11: Two four-chamber US images of different hearts, with the target segmentation in red around the left ventricle. Note the difference of appearance between the septum (green arrow) and the opposite portion of the myocardium (blue arrow). The apex (white arrow) has an ambiguous definition, and some regions have no clear boundaries (yellow arrow).

7.4.2 Learning tags from a database

We assume that we have a set of images $(\tilde{I}_n)_{n=1\dots N}$ that are all registered together, for example in Ω_0 the referential of the model ϕ_0. The registration does not need to be precise everywhere but in the neighborhood of the zero level-set of ϕ_0. For example, if the shape learning method proposed in Chapter 6 was performed, the minimizers of (6.6), in particular L_n^* and G_n^*, are available. Then the transformation $L_n^* \circ G_n^*$ is adequate to register the original image I_n in the model referential and we can define $\tilde{I}_n = I_n \circ G_n^{*-1} \circ L_n^{*-1}$.

From this set of images, we will learn what contrast (bright-to-dark or dark-to-bright) each point of the model presents. The goal is thus to build a tag function $T : \Omega \to [-1, 1]$ that can be used in the edge-based formulation of tagged template deformation presented in Section 7.2.2. To that end, we define a function $S : \Omega_0 \to \mathbb{R}$ as the mean scalar product between each image and the normalized gradient of the implicit model:

$$S(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\left\langle \frac{\vec{\nabla}\phi_0(\mathbf{x})}{|\vec{\nabla}\phi_0(\mathbf{x})|},\ \vec{\nabla}I_{n,\sigma}(\mathbf{x})\right\rangle = \left\langle \frac{\vec{\nabla}\phi_0(\mathbf{x})}{|\vec{\nabla}\phi_0(\mathbf{x})|},\ \frac{1}{N}\sum_{n=1}^{N}\vec{\nabla}I_{n,\sigma}(\mathbf{x})\right\rangle \tag{7.20}$$

where $I_{n,\sigma}$ is the image I_n filtered by a Gaussian filter of standard-deviation σ (whose purpose is to avoid contributions from noise and too weak edges). We are mainly interested in the values of S on the zero level-set of ϕ_0. At such points, $\frac{\vec{\nabla}\phi_0(\mathbf{x})}{|\vec{\nabla}\phi_0(\mathbf{x})|}$ represents the inward unit normal of the hypersurface represented by ϕ_0. If the point $\mathbf{x}$ belongs to an edge of the image $\tilde{I}_n$ that follows the boundary of the model, then the image gradient $\vec{\nabla}I_{n,\sigma}(\mathbf{x})$ will be collinear to the normal and their scalar product will be high (in absolute value). Therefore $S(\mathbf{x})$ will have a large amplitude where there is a consistent edge across the images

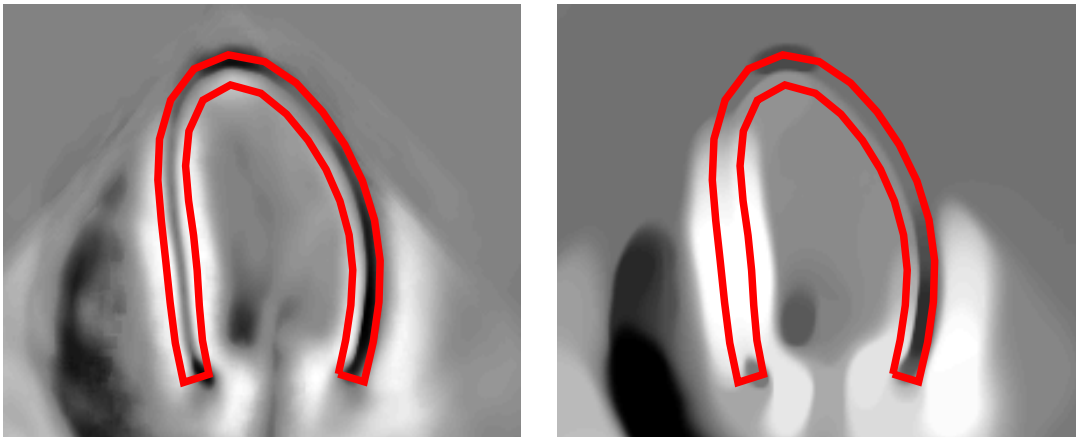


Figure 7.12: Tags learning for myocardium segmentation in US in the referential of the mean model (red). Black represents -1, grey 0 and white 1. (Left) Mean scalar product map S. (Right) Tags T^* obtained after total-variation regularization of S.

of the database (positive for bright-to-dark edges and negative for dark-to-bright edges). Conversely, at points where the interior of the model is sometimes brighter and sometimes darker than its exterior (*i.e.* image edges that are not reliable), S will be close to zero. It will also vanish when the model boundary crosses a perpendicular edge (which is also not relevant). At this point, we see that S would be a good candidate as a tag function.

However, S is smooth (but not piecewise-constant) and we mentioned in Section 7.2.2 that the complexity of the method was linked to the support of the gradient of the tag function. We therefore regularize S and reduce the support of its gradient, and actually define the tags as

$$T^*(\mathbf{x}) = \arg\min_T \int_{\Omega_0} \left(\frac{1}{2} \|T(\mathbf{x}) - S(\mathbf{x})\|^2 + \nu \|\vec{\nabla} T(\mathbf{x})\| \right) d\mathbf{x} \tag{7.21}$$

which is the usually called a *total-variation denoising* of S. Indeed the L_1-norm of the gradient, *i.e.* the total variation norm, has the interesting property of favoring piecewise-constant functions. Problem (7.21) is solved with the method described in [Chambolle & Pock, 2011]. The result T^* is shown in Figure 7.12. The most significant and consistent edges (in the region of the septum for instance) are detected. The pixels at the apex are also clustered into a sub-region, but with lower confidence. Furthermore, we notice a tag inversion between the inner and outer boundary at the apex and the bottom-right of the model. Others areas (*e.g.* at each part of the apex) are completely neglected: the segmentation will solely be interpolated by the shape prior without taking the image into account.

	Dice coefficient	Mean absolute distance (mm)	Maximum distance (mm)
Initialization (3 clicks)	0.59 (0.17)	4.87 (1.53)	15.17 (3.65)
Segmentation with standard model	0.59 (0.08)	5.10 (0.49)	16.65 (3.48)
Segmentation with tagged model	**0.77** (0.06)	**3.15** (0.88)	**9.76** (2.29)

Table 7.1: Results for the myocardium segmentation averaged over the 42 images. Figures in brackets indicate the standard-deviation of the metric.

7.4.3 Material and results

Our approach is validated on a set of 42 images coming from 14 subjects (both healthy volunteers and patients). The considered images are 2D long-axis, taken from a 4-chamber view of the heart. We asked a clinician to click on 3 points in each image within the myocardium: one at the apex and one at each valve. These points are used to initialize the position and size of the mean model (see left image in Figure 7.13).

The mean model is estimated via the shape learning process described in Chapter 6 of this thesis. The image-based term is the tagged Laplacian, as indicated in Section 7.2.2. Finally, we use the three points indicated by the clinician as constraints during the optimization. As detailed in Section 2.7.4, it is indeed possible to indicate some points that should lie inside or outside the segmentation. We therefore define these points as inner constraints.

Due to the reduced number of available images, validation has been performed with a leave-one-out strategy: for each patient, we used the information learnt from the other 13 patients. We evaluate our approach by computing for each image (i) the mean absolute distance, (ii) the maximum distance and (iii) the Dice coefficient, between the segmentation and the ground truth. The results are summarized in Table 7.1. For comparison purposes, we also indicate the scores obtained with the initial contour (placed with 3 points) and with the baseline method (constant positive tags).

An average Dice coefficient of 0.77 may seem low at first sight but it should be noted that we are segmenting a very thin object. Consequently, any little shift induces a large decrease of the Dice coefficient. This number should not be directly compared to the 0.96 average coefficient reported in [Dietenbeck et al., 2012]; in this article, the authors do not compute the Dice coefficient on the band itself but on its convex hull, which significantly (but artificially) raises the values obtained. The distance-based metrics (3.15 for the mean absolute distance and 9.76 for the maximum distance) are also higher than [Dietenbeck et al., 2012], namely 1.18 and 4.41. However, their method assumes that 6 points *on the contours*, while we only need 3 points *inside* the myocardium. Finally, their validation database was solely composed of healthy subjects. Images from patients with pathologies are more difficult to segment since the learning is less reliable.

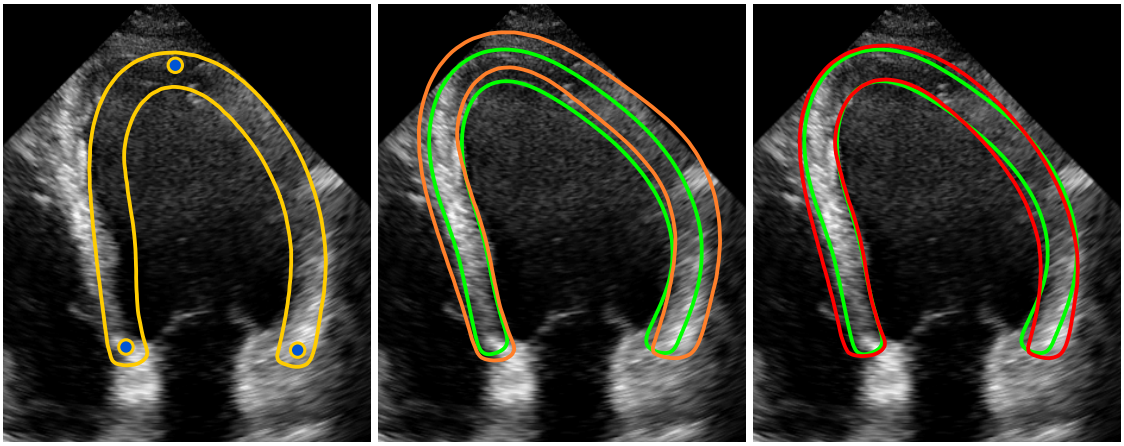


Figure 7.13: Myocardium segmentation in US images compared to the ground truth in green. (Left) Initialization of the mean model with 3 points. (Middle) Segmentation with the standard template deformation approach in orange. (Right) Segmentation with the tagged model in red.

We also show visually the benefits of the tagged template over the baseline algorithm in Figure 7.13. As the standard template can only take into account gradient information in a single direction, it may segment correctly the septum but then cannot capture the correct boundary at the apex. Furthermore, it takes too much into account the image information at some points of the model (typically on both sides of the apex). Conversely, the segmentation obtained with the tagged model has a better behaviour and is much closer to the ground truth.

Further results of segmentations with the tagged model are also reported in Figure 7.14. All segmentations seem reasonable and sensible (given the challenging image quality and the lack of clear boundaries) even if they sometimes deviate from the ground truth. We also point out that the only feature exploited in these experiments was the image gradient flux at a single scale. We believe that the results could be improved by taking into account multiple and more elaborate features, which is possible in the proposed framework. Another area of improvement would be the use of temporal information as in [Qin et al., 2013], for instance via our co-segmentation approach described in Chapter 5.

*
* *

By introducing tagged models, we have greatly enriched the prior information that is exploited in the template deformation framework. This extension widens the scope of potential clinical applications of this segmentation method; we indeed showed that major improvements were achieved over the standard approach in the context of myocardium segmentation in US images. This new framework paves the way for multi-organ segmentation: several organs can be represented by an implicit function, each of them being tagged in

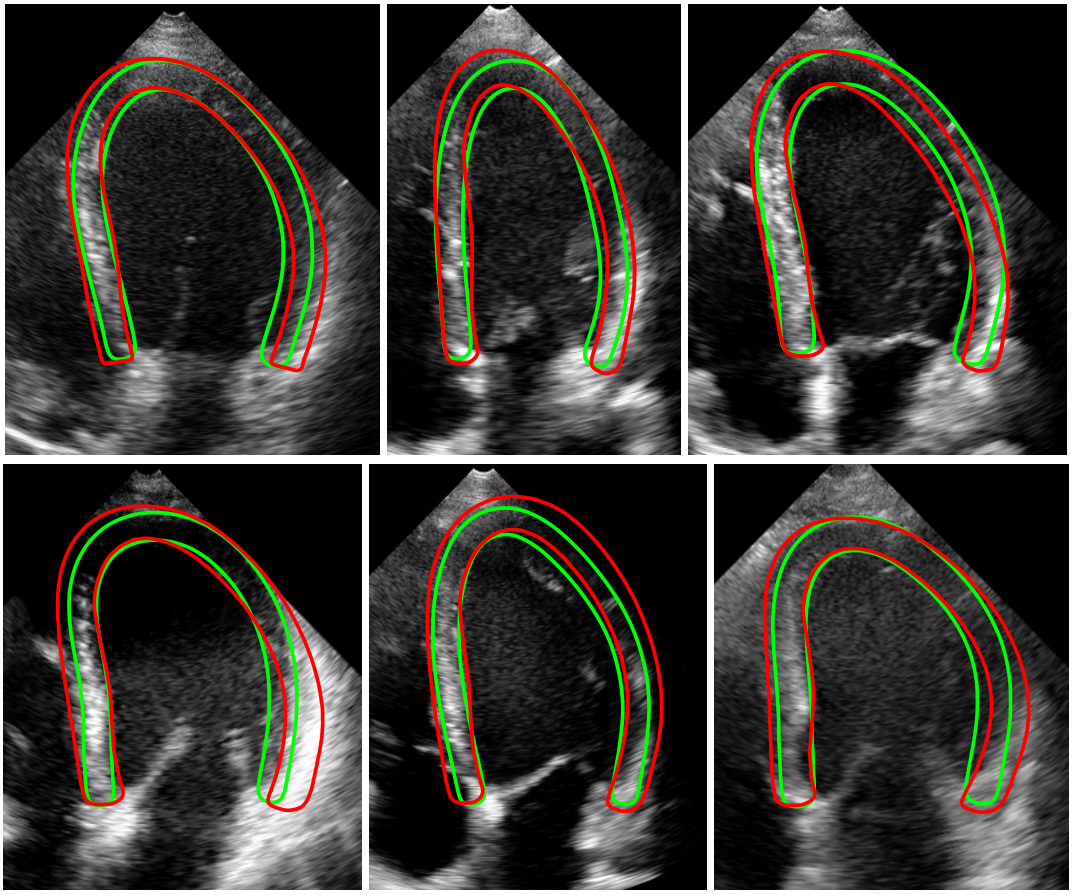


Figure 7.14: Myocardium segmentation results (red) compared to the ground truth (green) on various images representing the overall quality.

order to be attached to a dedicated image-based energy. It therefore represents a further step towards the robustness of atlas-based methods, with a much more efficient method though.

Chapter 8

Conclusion

Contents

This thesis was dedicated to the implicit template deformation segmentation framework and its applications. We have introduced several extensions of this approach, and presented other contributions that are related to it. In this conclusion, we summarize them and discuss potential future work.

8.1 Summary of the contributions

The technical contributions presented throughout this thesis can be clustered in the following categories.

On the implicit template deformation functional and its minimization The first contributions deal with the formulation itself of implicit template deformation. In the original functional [Mory, 2011], the smoothness of the deformation field was enforced by a Gaussian smoothing but, under careful mathematical analysis, this process turned out to be dubious. By defining the deformation field in a Gaussian reproducing kernel Hilbert space, we fixed the mathematical pitfalls appearing in the minimization of the functional.

Furthermore, we proposed an alternative minimization process that constrains ψ to be a diffeomorphism. The great benefit of this improvement is the topology preservation of the model ϕ_0, which is most of the time desirable in medical applications.

However, the algorithm still converges to a minimum that is local, so the initialization remains paramount. We proposed two methods to initialize automatically and robustly the model in an image. The first one consists in learning and regressing via random forests the position and size of an organ based on the contextual information. As the

bounding box of the organ is predicted (and not its particular shape), it can be applied for any organ but requires a standardized acquisition as in CT or MR imaging. The second one is an efficient variational approach for ellipsoid robust detection which only requires as an input a voxelwise probability to be inside the target object. It was therefore more indicated for (CE)US images.

Enhancing the model The notion of *model* in the original template deformation was limited, in the sense that the only information available was a fixed prior shape. We proposed two generalizations that enrich the model: (i) by incorporating information on shape variability, we enabled the algorithm to automatically select a shape prior adapted to the image being segmented, (ii) by splitting the model in different regions via a tag function, we are able to embed more information about the appearance of the target object and therefore make the algorithm more robust.

Statistical learning We have also shown throughout this thesis how to take advantage of databases of pre-segmented images. First, we proposed to learn the two functions r_{int} and r_{ext} as voxelwise classifiers with random forests. We then improved this approach in Section 4.4 by using an auto-context strategy that enforces spatial structure and consistency in the prediction maps.

In the context of tagged models (see Chapter 7), we also showed the possibility to learn – within the model – subregions that are attracted by different kinds of edges.

Finally, we also learnt information about the shape variability of the organ of interest. By computing a mean over a database of shapes defined with a dedicated distance, we constructed a shape template that is tailored to our algorithm. We even used further statistics by estimating (and then exploiting for the segmentation of unseen images) the main modes of variations of the deformations. The remarkable properties of this approach is its computational efficiency and the topology preservation of the initial model.

Exploiting external information Apart from large databases, other kinds of external information can be used to improve the segmentation. We studied the benefits of enabling the user to correct or refine the segmentation interactively. On the particular application of kidney segmentation in CEUS images (see Chapter 4), we show that a great improvement can be achieved in challenging images with just a few clicks.

We also presented a very generic framework for joint co-segmentation and registration of images. Although it may be adapted in various settings (*e.g.* motion estimation, tracking, etc.), its main interest lie in a multi-modal or longitudinal segmentation setting: by using multiple images of the same object, the segmentation can be greatly improved. This idea was applied to both the ellipsoid detection algorithm and the implicit template deformation framework.

All these contributions can be used together. We gave such an example in Section 7.3 with the tagged co-segmentation, but we could imagine other interesting combinations. For

instance, user interactions can be easily taken into account within the co-segmentation of multiple images.

From a clinical point of view, the main contributions of this thesis include:

- a fully automated pipeline for kidney segmentation in CT images with any field of view and any contrast phase. To our knowledge, this was the first approach evaluated on such a challenging clinical database;
- a fully automated pipeline for kidney segmentation in 3D US and 3D CEUS images. No such attempt was reported in the previous literature;
- the simultaneous use of US and CEUS images in a co-segmentation framework. It yields both an improved segmentation result and a registration of the US and CEUS images;
- a method to stabilize 3D+t sequences of free-breathing abdominal perfusion CT images, that proved to outperform the current state-of-the-art in terms of registration accuracy and pharmacokinetical parameters estimation.

Naturally, there is still much research to pursue as we have just unveiled the true potential of the implicit template deformation framework.

8.2 Future work

All the extensions that we have introduced in this manuscript are paving the way for various developments that we discuss hereinafter.

Coupling shape and appearance learning We proposed two approaches to learn useful information about respectively the shape and appearance of an organ. However these methods are performed independently. Just like in the Active Appearance Model [Cootes et al., 2001], we could find a way to learn them simultaneously and thus also capture their correlation.

Tagged models for a spatially-varying regularization The introduction of tagged models was motivated by the need of specializing the image-based term in different parts of the shape prior. However, we could also define different regularizations (*e.g.* using different widths for the Gaussian kernel defining the deformation field) in subregions of the model. This idea is somehow related with the registration method proposed [Schmah et al., 2013] and could use a similar formulation.

Towards multi-organ segmentation In the presented applications of the tags, the model stood for a single organ. Yet implicit functions are able to represent multiple disconnected objects. A natural extension is thus to perform multi-organ segmentation: each organ has its own tag (and therefore its own image-based term). This would

be a step forward towards the robustness of atlas-based segmentation approaches while benefiting from the algorithmic efficiency of the implicit template deformation framework.

Other clinical applications In this thesis, we focused on kidney and cardiac applications but the genericity of all our approaches enables their exploitation in a large number of clinical settings. For instance, we are currently investigating liver segmentation [Gauriau et al., 2013] in various modalities.

List of Publications and Patents

The work done during this Ph.D. lead to the following publications and patents.

Peer-Reviewed Conference Papers

- R. Prevost, B. Romain, B. Mory, O. Lucidarme, R. Cuingnet, L.D. Cohen, R. Ardon. "Registration of 3D+T Abdominal Perfusion CT Images via Co-Segmentation". *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2013)*.

- R. Prevost, B. Mory, R. Cuingnet, L.D. Cohen, R. Ardon. "Incorporating Shape Variability in Image Segmentation by Implicit Template Deformation". *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2013)*.

- R. Gauriau, R. Cuingnet, R. Prevost, B. Mory, R. Ardon, D. Lesage, I. Bloch. "A Generic, Robust and Fully-Automatic Workflow for 3D CT Liver Segmentation". *5th International MICCAI Workshop on Abdominal Imaging: Computational and Clinical Applications (MICCAI ABDI 2013)*.

- R. Prevost, R. Cuingnet, B. Mory, J.-M. Correas, L.D. Cohen, R. Ardon. "Joint Co-Segmentation and Registration of 3D Ultrasound Images". *Information Processing in Medical Imaging (IPMI 2013)*.

- R. Cuingnet, O. Somphone, B. Mory, R. Prevost, M. Yaqub, R. Napolitano, A. Papageorghiou, D. Roundhill, J.A. Noble, R. Ardon. "Where is my Baby? A Fast Fetal Head Auto-Alignment in 3D-Ultrasound". *IEEE International Symposium on Biomedical Imaging (ISBI 2013)*.

- R. Cuingnet, R. Prevost, B. Mory, D. Lesage, L.D. Cohen, R. Ardon. "Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests". *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2012)*.

- B. Mory, O. Somphone, R. Prevost, R. Ardon. "Real-Time 3D Image Segmentation by User-Constrained Template Deformation". *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2012)*.

- R. Prevost, B. Mory, J.-M. Correas, L.D. Cohen, R. Ardon. "Kidney detection and real-time segmentation in 3D contrast-enhanced ultrasound images". *IEEE International Symposium on Biomedical Imaging (ISBI 2012)*.
- G. Pizaine, R. Prevost, E. Angelini, I. Bloch, S. Makram-Ebeid. "Segmentation-free and multiscale-free extraction of medial information using Gradient Vector Flow". *IEEE International Symposium on Biomedical Imaging (ISBI 2012)*.
- R. Prevost, L.D. Cohen, J.-M. Correas, R. Ardon. "Automatic detection and segmentation of renal lesions in 3D contrast-enhanced ultrasound images". *SPIE Medical Imaging 2012*.
- B. Mory, O. Somphone, R. Prevost, R. Ardon. "Template Deformation with User Constraints for Live 3D Interactive Surface Extraction". *MICCAI 2011 Workshop on Mesh Processing in Medical Image Analysis (MICCAI Meshmed 2011)*.

Book Chapters

- R. Prevost, B. Mory, R. Cuingnet, J.-M. Correas, L.D. Cohen, R. Ardon. "Kidney Detection and Segmentation in Contrast-Enhanced Ultrasound 3D Images". *Abdomen and Thoracic Imaging - An Engineering and Clinical Perspective (Springer)*

Journal Articles

- B. Zhang, S. Makram-Ebeid, R. Prevost, G. Pizaine. "Fast Solver for Some Computational Imaging Problems: A Regularized Weighted Least-Squares Approach". *Submitted, 2013*

Patents

- R. Prevost, C. Dufour, B. Mory, R. Ardon. "Segmentation of a large object from multiple 3D partial views". *European Application 13305228.2, 2012*.
- R. Prevost, B. Mory, R. Ardon. "Coupled segmentation in 3D conventional ultrasound and contrast-enhanced ultrasound images". *European Application 12306033.7, 2012*.

Bibliography

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). Slic superpixels. *École Polytechnique Fédérale de Lausssanne (EPFL), Tech. Rep*, 149300. 26

Adams, R. & Bischof, L. (1994). Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6), 641–647. 24

Albrecht, T. et al. (2004). Guidelines for the use of contrast agents in ultrasound. *Ultraschall Med*, 25(4), 249–256. 78, 89, 209

Aldrich, J. E. (2007). Basic physics of ultrasound imaging. *Critical care medicine*, 35(5), S131–S137. 206

Ambrosio, L. & Tortorelli, V. M. (1990). Approximation of functional depending on jumps by elliptic functional via t-convergence. *Communications on Pure and Applied Mathematics*, 43(8), 999–1036. 30

An, J. H. & Chen, Y. (2007). Region based image segmentation using a modified mumford-shah algorithm. In *SSVM* (pp. 733–742). 45

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc*, 68(3), 337–404. 63

Arsigny, V., Commowick, O., Pennec, X., & Ayache, N. (2006a). A log-euclidean framework for statistics on diffeomorphisms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006* (pp. 924–931). Springer. 70, 136, 150

Arsigny, V., Commowick, O., Pennec, X., & Ayache, N. (2006b). A log-euclidean polyaffine framework for locally rigid or affine registration. In *Biomedical Image Registration* (pp. 120–127). Springer. 70

Arsigny, V., Pennec, X., Ayache, N., et al. (2005). Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations–application to the registration of histological slices. *Medical image analysis*, 9(6), 507–523. 131

Ashburner, J. et al. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1), 95–113. 66, 70

Aubert, G., Barlaud, M., Faugeras, O., & Jehan-Besson, S. (2003). Image segmentation using active contours: Calculus of variations or shape gradients? *SIAM Journal on Applied Mathematics*, 63(6), 2128–2154. 51

Bakker, J., Olree, M., Kaatee, R., de Lange, E. E., Moons, K. G., Beutler, J. J., & Beek, F. J. (1999). Renal Volume Measurements: Accuracy and Repeatability of US Compared with That of MR Imaging. *Radiology*, 211(3), 623–628. 78

Ballard, D. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2), 111–22. 82, 106

Bauer, C., Pock, T., Sorantin, E., Bischof, H., & Beichel, R. (2010). Segmentation of interwoven 3D tubular tree structures utilizing shape priors and graph cuts. *Medical image analysis*, 14(2), 172–184. 163

Beg, M. F., Miller, M. I., Trouvé, A., & Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2), 139–157. 70

Besl, P. & McKay, N. (1992). A method for registration of 3D shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2), 239–256. 135, 144, 145, 147

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? In *Database Theory – ICDT'99* (pp. 217–235). Springer. 20

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers. 19

Bhushan, M., Schnabel, J., Risser, L., Heinrich, M., Brady, J., & Jenkinson, M. (2011). Motion correction and parameter estimation in DCE-MRI sequences: application to colorectal cancer. *Proceedings of MICCAI 2011*, (pp. 476–483). 124

Blake, A. & Isard, M. (1998). *Active shape models*. Springer. 37

Blomley, M. & Cosgrove, D. (1997). Microbubble echo-enhancers: a new direction for ultrasound? *Lancet*, 349(9069), 1855–1856. 207

Booth, B., Patel, V., Lou, E., Le, L., & Li, X. (2006). Towards medical ultrasound image segmentation with limited prior knowledge. In *Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 4th* (pp. 488–93). 100

Bottou, L. (1998). Online learning and stochastic approximations. *On-line learning in neural networks*, 17, 9. 140

Boykov, Y. & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9), 1124–1137. 28

Boykov, Y. & Veksler, O. (2006). Graph cuts in vision and graphics: Theories and applications. In *Handbook of mathematical models in computer vision* (pp. 79–96). Springer. 28, 29

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11), 1222–1239. 28

Boykov, Y. Y. & Jolly, M.-P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1 (pp. 105–112).: IEEE. 29

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140. 23

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. 22, 82, 84, 101

Brigger, P., Hoeg, J., & Unser, M. (2000). B-spline snakes: a flexible tool for parametric contour detection. *Image Processing, IEEE Transactions on*, 9(9), 1484–1496. 38

Brunelli, R. (2009). *Template matching techniques in computer vision: theory and practice*. Wiley. 106

Burns, P. (1997). Overview of echo-enhanced vascular ultrasound imaging for clinical diagnosis in neurosonology. *Journal of neuroimaging: official journal of the American Society of Neuroimaging*, 7, S2. 207

Buzug, T. M. (2008). *Computed tomography: from photon statistics to modern cone-beam CT*. Springer. 201

Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., & Bach Cuadra, M. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, 104(3), e158–e177. 42

Cardoso, M. J., Winston, G., Modat, M., Keihaninejad, S., Duncan, J., & Ourselin, S. (2012). Geodesic shape-based averaging. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012* (pp. 26–33). Springer. 42

Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic active contours. *International journal of computer vision*, 22(1), 61–79. 37

Chambolle, A. & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1), 120–145. 173

Chan, T. & Vese, L. (2001a). Active contours without edges. *IEEE TIP*, 10(2), 266–77. 32, 50, 113

Chan, T. F. & Vese, L. A. (2001b). A level set algorithm for minimizing the mumford-shah functional in image processing. In *Variational and Level Set Methods in Computer Vision, 2001. Proceedings. IEEE Workshop on* (pp. 161–168).: IEEE. 31

Charpiat, G., Maurel, P., Pons, J.-P., Keriven, R., & Faugeras, O. (2007). Generalized gradients: Priors on minimization flows. *International Journal of Computer Vision*, 73(3), 325–344. 62

Chefd'hotel, C., Hermosillo, G., & Faugeras, O. (2002). Flows of diffeomorphisms for multimodal image registration. In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on* (pp. 753–756).: IEEE. 69

Christensen, G. E., Rabbitt, R. D., & Miller, M. I. (1996). Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10), 1435–1447. 68

CMU-Cornell (2010). iCoseg database – http://http://chenlab.ece.cornell.edu/projects/touch-coseg/. 112

Cohen, L. D. & Cohen, I. (1993). Finite-element methods for active contour models and balloons for 2-D and 3-D images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11), 1131–1147. 38

Cohen, L. D. & Kimmel, R. (1997). Global minimum for active contour models: A minimal path approach. *International journal of computer vision*, 24(1), 57–78. 37

Commowick, O. & Malandain, G. (2007). Efficient selection of the most similar image in a database for critical structures segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007* (pp. 203–210). Springer. 42

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3), 287–314. 150

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6), 681–685. 40, 41, 100, 156, 179

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1), 38–59. 39, 135, 137, 141, 144, 145, 147, 155

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297. 20

Cosgrove, D. (2006). Developments in ultrasound. *Imaging*, 18(2), 82. 201

Cosgrove, D., Blomley, M., Jayaram, V., & Nihoyannopoulos, P. (1998). Echo-enhancing (contrast) agents. *Ultrasound Quarterly*, 14(2), 66. 207

Couprie, C., Grady, L., Najman, L., & Talbot, H. (2011). Power watershed: A unifying graph-based optimization framework. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(7), 1384–1399. 28

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1), 21–27. 20

Cremers, D., Kohlberger, T., & Schnorr, C. (2003a). Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9), 1929–1943. 136

Cremers, D., Rousson, M., & Deriche, R. (2007). A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision*, 72(2), 195–215. 37, 135

Cremers, D. & Soatto, S. (2003). A pseudo-distance for shape priors in level set segmentation. In *2nd IEEE workshop on variational, geometric and level set methods in computer vision* (pp. 169–176).: Citeseer. 138

Cremers, D., Sochen, N., & Schnörr, C. (2003b). Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *Scale Space Methods in Computer Vision* (pp. 388–400).: Springer. 156

Criminisi, A. & Shotton, J. (2013). *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Publishing Company, Incorporated. 22

Criminisi, A., Shotton, J., & Bucciarelli, S. (2009). Decision forests with long-range spatial context for organ localization in ct volumes. In *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*. 82

Criminisi, A., Shotton, J., & Konukoglu, E. (2011a). *Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*. Technical report, Microsoft Research. 82, 84

Criminisi, A., Shotton, J., Robertson, D., & Konukoglu, E. (2011b). Regression forests for efficient anatomy detection and localization in CT studies. *MICCAI Workshop on Medical Computer Vision*, (pp. 106–17). 82, 86

Cuingnet, R., Prevost, R., Lesage, D., Cohen, L. D., Mory, B., & Ardon, R. (2012). Automatic detection and segmentation of kidneys in 3D CT images using random forests. In *MICCAI*, volume 7512 of *LNCS* (pp. 66–74). Springer. 80, 125

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 1–38). 18

Dietenbeck, T., Alessandrini, M., Barbosa, D., D'hooge, J., Friboulet, D., & Bernard, O. (2012). Detection of the whole myocardium in 2D-echocardiography for multiple orientations using a geometrically constrained level-set. *Medical Image Analysis*, 16(2), 386–401. 171, 174

Dollar, P., Welinder, P., & Perona, P. (2010). Cascaded pose regression. In *CVPR* (pp. 1078–85).: IEEE. 83

Dupuis, P., Grenander, U., & Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, 56(3), 587. 70

Durrleman, S. (2010). Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. *These de sciences (PhD thesis), Université de Nice-Sophia Antipolis (March 2010)*. 136, 137

Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M., Vembar, M., Olszewski, M., Subramanyan, K., Lavi, G., et al. (2008). Automatic model-based segmentation of the heart in CT images. *IEEE TMI*, 27(9), 1189–201. 82, 106

Edelman, R. R., Hesselink, J. R., & Zlatkin, M. B. (1996). *Clinical magnetic resonance imaging*, volume 2. WB Saunders Philadelphia. 201

El-Baz, A., Fahmi, R., Yuksel, S., Farag, A., Miller, W., El-Ghar, M., & Eldiasty, T. (2006). A new CAD system for the evaluation of kidney diseases using DCE-MRI. *Proceedings of MICCAI 2006*, (pp. 446–453). 124

Feldman, M. K., Katyal, S., & Blackwood, M. S. (2009). US Artifacts. *Radiographics*, 29(4), 1179–1189. 206

Felzenszwalb, P. F. & Huttenlocher, D. P. (2006). Efficient belief propagation for early vision. *International journal of computer vision*, 70(1), 41–54. 27

Fenchel, M., Thesen, S., & Schilling, A. (2008). Automatic labeling of anatomical structures in MR FastView images using a statistical atlas. In *MICCAI*, volume 5241 of *LNCS* (pp. 576–584). 81

Fenster, A., Downey, D., & Cardinal, H. (2001). Three-dimensional ultrasound imaging. *Physics in medicine and biology*, 46, R67. 204

Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, 85(4), 597–612. 20

Ford, L. R. & Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 8(3), 399–404. 28

Freund, Y. & Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory* (pp. 23–37).: Springer. 21

Fukunaga, K. & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1), 32–40. 19

Gasnier, A. (2010). *Validation of Quantitative 3D Ultrasound for the Assessment and Follow-Up of Tumor Radiofrequency Ablation*. Technical report, Philips Research Medisys. 201

Gasnier, A., Ardon, R., Ciofolo-Veit, C., Leen, E., & Correas, J. (2010). Assessing tumour vascularity with 3D contrast-enhanced ultrasound: a new semi-automated segmentation framework. In *Proceedings of IEEE ISBI 2010* (pp. 300–303). 89

Gauriau, R., Cuingnet, R., Prevost, R., Mory, B., Ardon, R., Lesage, D., & Bloch, I. (2013). A generic, robust and fully-automatic workflow for 3D CT liver segmentation. In *MICCAI Workshop on Abdominal Imaging (ABDI 2013)*. 88, 163, 180

Georgescu, B., Zhou, X., Comaniciu, D., & Gupta, A. (2005). Database-guided segmentation of anatomical structures with complex appearance. In *CVPR*, volume 2 (pp. 429–36).: IEEE. 82

Glocker, B., Pauly, O., Konukoglu, E., & Criminisi, A. (2012). Joint classification-regression forests for spatially structured multi-object segmentation. In *ECCV*, volume 7575 of *LNCS* (pp. 870–81). Springer. 101

Grady, L. (2006). Random walks for image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(11), 1768–1783. 28

Grady, L. & Alvino, C. V. (2009). The piecewise smooth mumford–shah functional on an arbitrary graph. *Image Processing, IEEE Transactions on*, 18(11), 2547–2561. 31

Guil, N. & Zapata, E. (1997). Lower order circle and ellipse Hough transform. *Pattern Recognition*, 30(10), 1729 – 1744. 89

Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., & Wu, X. (2011). Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In *IPMI* (pp. 245–56). 113

Hart, G. L., Zach, C., & Niethammer, M. (2009). An optimal control approach for deformable registration. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on* (pp. 9–16).: IEEE. 70

Harvey, C., Pilcher, J., Eckersley, R., Blomley, M., & Cosgrove, D. (2002). Advances in ultrasound. *Clinical Radiology*, 57(3), 157–177. 201

Hastie, T., Tibshirani, R., & Friedman, J. J. H. (2001). *The elements of statistical learning*, volume 1. Springer New York. 20

Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., & Hammers, A. (2006). Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1), 115–126. 42

Heimann, T. & Meinzer, H.-P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Medical image analysis*, 13(4), 543. 40, 135

Hochbaum, D. S. & Singh, V. (2009). An efficient algorithm for co-segmentation. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 269–276).: IEEE. 111

Hosmer, D. & Lemeshow, S. (1989). Applied logistic regression. *Wiley Series in Probability and mathematical statistics. Applied Probability and Statistics*. 20

Hsu, P., Prager, R., Gee, A., & Treece, G. (2008). Real-time freehand 3D ultrasound calibration. *Ultrasound in Medicine and Biology*, 34(2), 239–251. 204

Huang, X., Li, Z., & Metaxas, D. (2004). Learning coupled prior shape and appearance models for segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004* (pp. 60–69). Springer. 8, 156

Huang, X. & Metaxas, D. (2008). Metamorphs: Deformable shape and appearance models. *IEEE Trans. PAMI*, 30(8), 1444–1459. 45, 156

Iglesias, J., Konukoglu, E., Montillo, A., Tu, Z., & Criminisi, A. (2011). Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In *IPMI* (pp. 25–36).: Springer. 83

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M. A., & van Ginneken, B. (2009). Multi-atlas-based segmentation with local decision fusion – application to cardiac and aortic segmentation in ct scans. *Medical Imaging, IEEE Transactions on*, 28(7), 1000–1010. 42, 81

Jolliffe, I. T. (1986). *Principal component analysis*, volume 487. Springer-Verlag New York. 140

Joshi, S. et al. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(1), 151. 42, 136, 137

Kambadakone, A. & Sahani, D. (2009). Body perfusion CT: technique, clinical applications, and advances. *Radiologic clinics of North America*, 47(1), 161–178. 123

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5), 509–541. 136

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4), 321–331. 37, 38

Katsamanis, A., Papandreou, G., & Maragos, P. (2009). Face active appearance modeling and speech acoustic information to recover articulation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(3), 411–422. 41

Khalifa, F., Elnakib, A., Beache, G., Gimelöarb, G., El-Ghar, M., Ouseph, R., Sokhadze, G., Manning, S., McClure, P., & El-Baz, A. (2011). 3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function. In *MICCAI*, volume 6893 of *LNCS* (pp. 587–94).: Springer. 80

Khan, A. R., Wang, L., & Beg, M. F. (2008). Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping. *NeuroImage*, 41(3), 735–746. 136

Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., & Yezzi, A. (1995). Gradient flows and geometric active contour models. In *Computer Vision, 1995. Proceedings., Fifth International Conference on* (pp. 810–815).: IEEE. 37

Kimmel, R. & Bruckstein, A. M. (2003). Regularized Laplacian zero crossings as optimal edge integrators. *International Journal of Computer Vision*, 53(3), 225–243. 35, 158, 160

Kissi, A., Cormier, S., Pourcelot, L., Bleuzen, A., & Tranquart, E. (2004). Contrast enhanced ultrasound image segmentation based on fuzzy competitive clustering and anisotropic diffusion. In *IEEE IEMBS 2004*, volume 1 (pp. 1613 –1615). 89

Knutsson, H. & Westin, C.-F. (1993). Normalized and differential convolution. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on* (pp. 515–523).: IEEE. 97

Koh, T., Thng, C., Hartono, S., Lee, P., Choo, S., Poon, D., Toh, H., & Bisdas, S. (2009). Dynamic contrast-enhanced CT imaging of hepatocellular carcinoma in cirrhosis: feasibility of a prolonged dual-phase imaging protocol with tracer kinetics modeling. *European radiology*, 19(5), 1184–1196. 124

Komodakis, N., Paragios, N., & Tziritas, G. (2011). MRF energy minimization and beyond via dual decomposition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3), 531–552. 27

Kontschieder, P., Bulò, S., Criminisi, A., Kohli, P., Pelillo, M., & Bischof, H. (2012). Context-sensitive decision forests for object detection. In *Proceedings of NIPS* (pp. 440–8). 101

Kwan, K. & Matsumoto, E. (2007). Radiofrequency ablation and cryoablation of renal tumours. *Current Oncology*, 14(1), 34. 77

Leventon, M., Grimson, W., & Faugeras, O. (2000a). Statistical shape influence in geodesic active contours. In *Proceedings of CVPR 2000* (pp. 316–323). 40, 136, 141

Leventon, M. E., Faugeras, O., Grimson, W. E. L., & Wells III, W. M. (2000b). Level set based segmentation with intensity and curvature priors. In *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on* (pp. 4–11).: IEEE. 156

Li, S. Z. (2009). *Markov random field modeling in image analysis*. Springer. 27

Li, X., Chen, X., Yao, J., Zhang, X., & Tian, J. (2011). Renal cortex segmentation using optimal surface search with novel graph construction. In *MICCAI*, volume 6893 of *LNCS* (pp. 387–94). Springer Berlin / Heidelberg. 80

Li, Z., Caan, M., Ziech, M., Stoker, J., van Vliet, L., & Vos, F. (2012). 3D non-rigid motion correction of free-breathing abdominal DCE-MRI data. *Abdominal Imaging. Computational and Clinical Applications*, (pp. 44–50). 124

Lindblad, P. & Adami, H.-O. (2002). *Kidney cancer*. Oxford University Press, Oxford, UK. 76

Lindholt, J. S., Norman, P., et al. (2008). Screening for abdominal aortic aneurysm reduces overall mortality in men. A meta-analysis of the mid-and long-term effects of screening for abdominal aortic aneurysms. *European journal of vascular and endovascular surgery: the official journal of the European Society for Vascular Surgery*, 36(2), 167. 167

Lindner, J. R. (2004). Microbubbles in medical imaging: current applications and future directions. *Nature Reviews Drug Discovery*, 3(6), 527–533. 207

Linguraru, M., Yao, J., Gautam, R., Peterson, J., Li, Z., Marston Linehan, W., & Summers, R. (2009). Renal tumor quantification and classification in contrast-enhanced abdominal CT. *Pattern recognition*, 42(6), 1149–1161. 124

Long, A., Rouet, L., Lindholt, J. S., & Allaire, E. (2012). Measuring the maximum diameter of native abdominal aortic aneurysms: Review and critical analysis. *European Journal of Vascular and Endovascular Surgery*, 43(5), 515–524. 167

Lorensen, W. E. & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. In *ACM Siggraph Computer Graphics*, volume 21 (pp. 163–169).: ACM. 10, 51

Lu, C. & Duncan, J. (2012). A coupled segmentation and registration framework for medical image analysis using robust point matching and active shape model. In *IEEE Workshop on MMBIA* (pp. 129–36). 112

Ma, M., Stralen, M., Reiber, J., Bosch, J., & Lelieveldt, B. (2009). Left ventricle segmentation from contrast enhanced fast rotating ultrasound images using three dimensional active shape models. In *Proceedings of FIMH 2009* (pp. 295–302). 89

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(281-297), 14. 18

Malladi, R. & Sethian, J. A. (1998). A real-time algorithm for medical shape recovery. In *Computer Vision, 1998. Sixth International Conference on* (pp. 304–310).: IEEE. 25

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2), 431–441. 130

Martin-Fernandez, M. & Alberola-Lopez, C. (2005). An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours. *MedIA*, 9(1), 1–23. 101

Matsuyama, Y. (2003). The α-em algorithm: Surrogate likelihood maximization using α-logarithmic information measures. *Information Theory, IEEE Transactions on*, 49(3), 692–706. 19

McInerney, T. & Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2), 91–108. 37

McLaughlin, R. A. (1998). Randomized Hough transform: improved ellipse detection with comparison. *Pattern Recognition Letters*, 19(3), 299–305. 89

Mendoza, C. S., Kang, X., Safdar, N., Myers, E., Peters, C. A., & Linguraru, M. G. (2013). Kidney segmentation in ultrasound via genetic initialization and active shape models with rotation correction. In *Proceedings of ISBI*, LNCS: IEEE. 100

Meng, X.-L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2), 267–278. 19

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P., van der Geest, R. J., Reiber, J. H., & Sonka, M. (2002). 3D active appearance models: segmentation of cardiac MR and ultrasound images. *Medical Imaging, IEEE Transactions on*, 21(9), 1167–1178. 41

Montagnat, J., Delingette, H., & Ayache, N. (2001). A review of deformable surfaces: topology, geometry and deformation. *Image and vision computing*, 19(14), 1023–1040. 38

Montillo, A., Shotton, J., Winn, J., Iglesias, J., Metaxas, D., & Criminisi, A. (2011). Entangled decision forests and their application for semantic segmentation of CT images. In *Proceedings of IPMI*, volume 6801 of *LNCS* (pp. 184–96).: Springer. 101

Mory, B. (2011). *Interactive Segmentation of 3D Medical Images with Implicit Surfaces*. PhD thesis, STI, Lausanne. 2, 3, 8, 10, 34, 35, 44, 45, 47, 48, 49, 50, 52, 53, 56, 62, 71, 177, 211

Mory, B. & Ardon, R. (2007). Fuzzy region competition: a convex two-phase segmentation framework. In *Scale Space and Variational Methods in Computer Vision* (pp. 214–226). Springer. 33

Mory, B., Somphone, O., Prevost, R., & Ardon, R. (2012). Real-time 3D image segmentation by user-constrained template deformation. In *MICCAI*, volume 7510 of *LNCS* (pp. 561–8). Springer. 44, 45, 60, 65

Mumford, D. & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5), 577–685. 30

Neal, R. M. & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models* (pp. 355–368). Springer. 19

Noble, J. A. & Boukerroui, D. (2006). Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging*, 25(8), 987–1010. 89

Nocedal, J. & Wright, S. J. (1999). *Numerical optimization*. Springer. 51, 53

Nock, R. & Nielsen, F. (2004). Statistical region merging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(11), 1452–1458. 24

O'Neill, W. C., Robbin, M. L., Bae, K. T., Grantham, J. J., Chapman, A. B., Guay-Woodford, L. M., Torres, V. E., King, B. F., Wetzel, L. H., Thompson, P. A., et al. (2005). Sonographic assessment of the severity and progression of autosomal dominant polycystic kidney disease: the consortium of renal imaging studies in polycystic kidney disease (crisp). *American journal of kidney diseases*, 46(6), 1058–1064. 78

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23–27. 17, 18

Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A., Nekolla, S., & Navab, N. (2011). Fast multiple organ detection and localization in whole-body MR Dixon sequences. In *MICCAI*, volume 6893 of *LNCS* (pp. 239–47).: Springer. 82

Payet, N. & Todorovic, S. (2010). 2-Random Forest Random Field. In *Proceedings of NIPS 2010*. 101

Perez, P. et al. (1998). Markov random fields and images. *CWI quarterly*, 11(4), 413–437. 27

Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2009). An algorithm for minimizing the Mumford-Shah functional. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 1133–1140).: IEEE. 31

Pohl, K., Fisher, J., Grimson, W., Kikinis, R., & Wells, W. (2006). A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1), 228–39. 112

Poon, T. & Rohling, R. (2005). Comparison of calibration methods for spatial tracking of a 3D ultrasound probe. *Ultrasound in medicine & biology*, 31(8), 1095. 204

Prevost, R., Cohen, L., Correas, J., & Ardon, R. (2012a). Automatic detection and segmentation of renal lesions in 3D contrast-enhanced ultrasound images. *Proc. of SPIE Vol*, 8314, 83141D–1. 26, 89

Prevost, R., Cuingnet, R., Mory, B., Cohen, L. D., & Ardon, R. (2013a). Incorporating Shape Variability in Image Segmentation by Implicit Template Deformation. In *MICCAI*, volume to appear of *LNCS*. Springer. 135

Prevost, R., Cuingnet, R., Mory, B., Correas, J.-M., Cohen, L. D., & Ardon, R. (2013b). Joint Co-segmentation and Registration of 3D Ultrasound Images. In *Information Processing in Medical Imaging* (pp. 268–279). Springer. 115

Prevost, R., Mory, B., & Ardon, R. (2012b). Coupled segmentation in 3D conventional ultrasound and contrast-enhanced ultrasound images. European Patent Application 12306033.7. 115

Prevost, R., Mory, B., Correas, J.-M., Cohen, L. D., & Ardon, R. (2012c). Kidney detection and real-time segmentation in 3D contrast-enhanced ultrasound images. In *Proceedings of IEEE ISBI* (pp. 1559–62). 88

Prevost, R., Romain, B., Cuingnet, R., Mory, B., Rouet, L., Lucidarme, O., Cohen, L. D., & Ardon, R. (2013c). Registration of Free-Breathing 3D+t Abdominal Perfusion CT Images via Co-Segmentation. In *MICCAI*, volume to appear of *LNCS*. Springer. 122

Qin, X., Cong, Z., Halig, L. V., & Fei, B. (2013). Automatic segmentation for right ventricle on ultrasound images using sparse matrix transform and level set. In *SPIE Medical Imaging* (pp. 86690Q–86690Q).: International Society for Optics and Photonics. 171, 175

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81–106. 22

Raja, K. B., Madheswaran, M., & Thyagarajah, K. (2007). A general segmentation scheme for contouring kidney region in ultrasound kidney images using improved higher order spline interpolation. *International Journal of Biological and Medical Sciences*, 2(2), 81–8. 100

Riklin-Raviv, T., Sochen, N., & Kiryati, N. (2008). Shape-based mutual segmentation. *International Journal of Computer Vision*, 79(3), 231–245. 111, 112

Romain, B., Letort, V., Lucidarme, O., d'Alche Buc, F., & Rouet, L. (2012). Registration of free-breathing abdominal 3D contrast-enhanced CT. In *MICCAI Workshop on Abdominal Imaging*, volume 7601 (pp. 274–282).: Springer. 124, 127, 128, 131, 132

Rose, J.-L., Revol-Muller, C., Almajdub, M., Chereul, E., & Odet, C. (2007). Shape prior integrated in an automated 3D region growing method. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 1 (pp. I–53).: IEEE. 26

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386. 21

Rother, C., Kolmogorov, V., & Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23 (pp. 309–314).: ACM. 29

Rousson, M. & Paragios, N. (2002). Shape priors for level set representations. *Proceedings of ECCV 2002*, (pp. 416–418). 40, 44, 136, 138

Rousson, M., Paragios, N., & Deriche, R. (2004). Implicit active shape models for 3D segmentation in MR imaging. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004* (pp. 209–216). Springer. 142

Rubio, J. C., Serrat, J., López, A., & Paragios, N. (2012). Unsupervised co-segmentation through region matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 749–756).: IEEE. 111

Rueckert, D., Frangi, A., & Schnabel, J. (2001). Automatic construction of 3D statistical deformation models using non-rigid registration. *Proceedings of MICCAI 2001*, (pp. 77–84). 140

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., & Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on*, 18(8), 712–721. 45, 48

Rumack, C., Wilson, S., & Charboneau, W. (2005). Diagnostic ultrasound. *Mosby, Philadelphia, PA*. 207

Saddi, K., Chefd'hotel, C., Rousson, M., & Cheriet, F. (2007). Region-based segmentation via non-rigid template matching. *Proceedings of ICCV*, 0, 1–7. 8, 45, 68, 135

Sance, R., Ledesma-Carbayo, M., Lundervold, A., & Santos, A. (2007). Alignment of 3D DCE-MRI abdominal series for optimal quantification of kidney function. In *Proceedings of IEEE ISPA 2007* (pp. 413–417). 124

Sandor, S. & Leahy, R. (1997). Surface-based labeling of cortical anatomy using a deformable atlas. *Medical Imaging, IEEE Transactions on*, 16(1), 41–54. 42

Sauvola, J. & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, 33(2), 225–236. 17, 18

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2), 197–227. 21

Schick, A., Bauml, M., & Stiefelhagen, R. (2012). Improving foreground segmentations with probabilistic superpixel markov random fields. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on* (pp. 27–31).: IEEE. 26

Schmah, T., Risser, L., & Vialard, F.-X. (2013). Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In *To appear in MICCAI 2013 Proceedings*. Springer. 179

Semelka, R. C., Patrick Shoenut, J., Magro, C. M., Kroeker, M. A., Macmahon, R., & Greenberg, H. M. (1993). Renal cancer staging: Comparison of contrast-enhanced ct and gadolinium-enhanced fat-suppressed spin-echo and gradient-echo mr imaging. *Journal of Magnetic Resonance Imaging*, 3(4), 597–602. 77

Setola, S., Catalano, O., Sandomenico, F., & Siani, A. (2007). Contrast-enhanced sonography of the kidney. *Abdominal Imaging*, 32(1), 21–28. 78

Sezgin, M. et al. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1), 146–168. 17

Shawe-Taylor, J. & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge university press. 21

Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8), 888–905. 29

Soatto, S. & Yezzi, A. J. (2002). Deforming motion , shape average and the joint registration and segmentation of images. *Computer Vision Eccv 2002 Pt Iii*, 2352(3), 32–47. 45

Somphone, O., Mory, B., Makram-Ebeid, S., & Cohen, L. D. (2008). Prior-based piecewise-smooth segmentation by template competitive deformation using partitions of unity. In *ECCV*. 45

Song, T., Lee, V., Rusinek, H., Wong, S., & Laine, A. (2006). Integrated four dimensional registration and segmentation of dynamic renal MR images. *Proceedings of MICCAI 2006*, (pp. 758–765). 124

Sourbron, S. P. & Buckley, D. L. (2011). On the scope and interpretation of the tofts models for dce-mri. *Magnetic Resonance in Medicine*, 66(3), 735–745. 127

Spiegel, M., Hahn, D. A., Daum, V., Wasza, J., & Hornegger, J. (2009). Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. *Comput Med Imaging Graph*, 33(1), 29–39. 80, 155

Sugar, C. A. & James, G. M. (2003). Finding the number of clusters in a dataset. *Journal of the American Statistical Association*, 98(463). 19

Sun, Y., Jolly, M.-P., & Moura, J. (2004). Contrast-invariant registration of cardiac and renal MR perfusion images. In *Proceedings of MICCAI 2004* (pp. 903–910). Springer. 124

Sundaramoorthi, G., Yezzi, A., & Mennucci, A. C. (2007). Sobolev active contours. *International Journal of Computer Vision*, 73(3), 345–366. 62

Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, (pp. 49–58). 18

Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9), 1725–1745. 41

Thirion, J. (1998). Image Matching as a Diffusion Process: An analogy with Maxwell's demons. *Medical I*, 2, 243–260. 49

Tofts, P. S., Brix, G., Buckley, D. L., Evelhoch, J. L., et al. (1999). Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3), 223–232. 127, 130

Toshev, A., Makadia, A., & Daniilidis, K. (2009). Shape-based object recognition in videos using 3D synthetic object models. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 288–295).: IEEE. 111

Treece, G., Gee, A., Prager, R., Cash, C., & Berman, L. (2003). High-definition freehand 3D ultrasound. *Ultrasound in medicine & biology*, 29(4), 529–546. 204

Tsagaan, B., Shimizu, A., Kobatake, H., & Miyakawa, K. (2002). An automated segmentation method of kidney using statistical information. In *MICCAI*, volume 2488 of *LNCS* (pp. 556–63). Springer Berlin / Heidelberg. 80

Tsai, A., Yezzi Jr, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., & Willsky, A. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Transactions on*, 22(2), 137–154. 40, 136, 137, 138, 141, 144, 145, 147

Tu, Z. & Bai, X. (2010). Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE TPAMI*, 32(10), 1744–57. 3, 102

Ulupinar, F. & Nevatia, R. (1995). Shape from contour: Straight homogeneous generalized cylinders and constant cross section generalized cylinders. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2), 120–135. 168

Vaillant, M., Miller, M., Younes, L., & Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23(1), 161. 136

Van Aelst, S. & Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 71–82. 89

Van Ginneken, B., Frangi, A. F., Staal, J. J., ter Haar Romeny, B. M., & Viergever, M. A. (2002). Active shape model segmentation with optimal features. *medical Imaging, IEEE Transactions on*, 21(8), 924–933. 155

Vasilevskiy, A. & Siddiqi, K. (2002). Flux maximizing geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12), 1565–1578. 35

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., et al. (2008). Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*, 45(1). 68, 70

Vialard, F.-X., Risser, L., Rueckert, D., & Cotter, C. J. (2012). Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2), 229–241. 70

Vicente, S., Kolmogorov, V., & Rother, C. (2010). Cosegmentation revisited: Models and optimization. In *Proceedings of ECCV*, volume 6312 (pp. 465–79). Springer. 111

Wang, F. & Vemuri, B. (2005). Simultaneous registration and segmentation of anatomical structures from brain MRI. In *Proceedings of MICCAI*, volume 3749 of *LNCS* (pp. 17–25). Springer. 112

Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7), 903–921. 42

Wein, W., Brunke, S., Khamene, A., Callstrom, M. R., & Navab, N. (2008). Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical image analysis*, 12(5), 577. 162

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6), 80–83. 104, 122, 127, 148

Wolz, R., Chu, C., Misawa, K., Mori, K., & Rueckert, D. (2012). Multi-organ abdominal ct segmentation using hierarchically weighted subject-specific atlases. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012* (pp. 10–17). Springer. 42, 43

Wong, C., Lin, S., Ren, T., & Kwok, N. (2012). A survey on ellipse detection methods. In *Industrial Electronics (ISIE), 2012 IEEE International Symposium on* (pp. 1105–1110).: IEEE. 89

Wu, C.-H. & Sun, Y.-N. (2006). Segmentation of kidney from ultrasound B-mode images with texture-based classification. *Comput Methods Programs Biomed*, 84(2), 114–23. 100

Wyatt, P. & Noble, J. (2002). MAP MRF joint segmentation and registration. In *Proceedings of MICCAI*, volume 2488 of *LNCS* (pp. 580–7). Springer. 112

Xie, J., Jiang, Y., & Tsui, H. (2005). Segmentation of kidney from ultrasound images based on texture and shape priors. *IEEE TMI*, 24(1), 45–57. 100

Yang, J. & Duncan, J. S. (2004). 3d image segmentation of deformable objects with joint shape-intensity prior models using level sets. *Medical Image Analysis*, 8(3), 285–294. 156

Yezzi, A., Zöllei, L., & Kapur, T. (2003). A variational framework for integrating segmentation and registration through active contours. *MedIA*, 7(2), 171–85. 112, 113

Yezzi, A. J. & Soatto, S. (2003). Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2), 153–167. 45, 136

Younes, L. (2010). *Shapes and diffeomorphisms*, volume 171. Springer. 66

Yu, J., Tan, J., & Wang, Y. (2010). Ultrasound speckle reduction by a susan-controlled anisotropic diffusion method. *Pattern Recognition*, 43(9), 3083–3092. 205

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *International journal of computer vision*, 8(2), 99–111. 38

Zagrodsky, V., Walimbe, V., Castro-Pareja, C., Qin, J. X., Song, J.-M., & Shekhar, R. (2005). Registration-assisted segmentation of real-time 3D echocardiographic data using deformable models. *Medical Imaging, IEEE Transactions on*, 24(9), 1089–1099. 113

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., & Comaniciu, D. (2008). Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE TMI*, 27(11), 1668–81. 82

Zhou, S., Georgescu, B., Zhou, X., & Comaniciu, D. (2005). Image based regression using boosting method. In *ICCV*, volume 1 (pp. 541–48).: IEEE. 82

Zhu, S. C. & Yuille, A. (1996). Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(9), 884–900. 33

Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., & Price, S. (2012). Decision Forests for Tissue-specific Segmentation of High-grade Gliomas in Multi-channel MR. In *MICCAI*, volume 7512 of *LNCS* (pp. 369–76). Springer. 101

Zöllner, F., Sance, R., Rogelj, P., Ledesma-Carbayo, M., Rørvik, J., Santos, A., & Lundervold, A. (2009). Assessment of 3D DCE-MRI of the kidneys using non-rigid image registration and segmentation of voxel time courses. *Computerized Medical Imaging and Graphics*, 33(3), 171–181. 124

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. 21

Appendix A

A short background on imaging modalities

In this subsection, we describe the principles of the main imaging modalities that will be used in the thesis, namely ultrasound (US) and contrast-enhanced ultrasound (CEUS). The purpose is to give an insight on how such images are generated, in order to understand what one actually sees (or not) and why image quality may be degraded.

Note that Computed Tomography (CT) and Magnetic Resonance Imaging (MR) are also considered in some applications of the presented work. However, for the sake of concision, we will limit our discussion to ultrasound image acquisition. We refer the interested reader to, for example, [Buzug, 2008] for a thorough description of CT imaging or to [Edelman et al., 1996] for MR imaging.

Some parts of this section are inspired by a technical report [Gasnier, 2010].

A.1 Ultrasound (US)

Ultrasound has seen rapid developments in recent years: the progressive introduction of grey scale (1970), real time (1975), multi-element arrays (1975), pulsed Doppler (1970) and subsequently colour Doppler (1985), in addition to the migration to digital systems (1985) have extended its reliability and ease of use [Harvey et al., 2002] (see Figure A.1). Radical innovations include the development of contrast agents for ultrasound, of elastography and the use of ultrasound in therapy, both as a means of heat-coagulating tissue (high intensity focused ultrasound) and as a way to improve drug delivery [Cosgrove, 2006]. Ultrasonography is widely used in medicine to perform diagnosis or to guide interventional procedures, thanks to its effectiveness for imaging soft tissues of the body. It presents unique properties that are particularly valuable: it is cheap, portable, completely harmless for the patient (no radiation) and it allows real-time visualization.

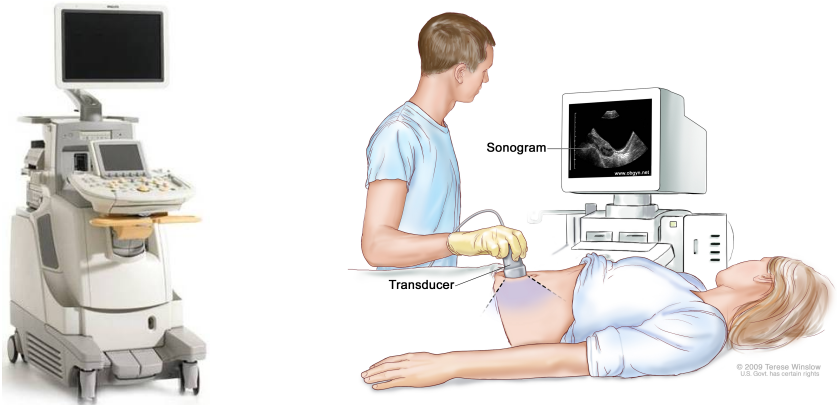


Figure A.1: The Philips iU22 ultrasound scanning machine that was used to acquire all ultrasound images used in this thesis.

A.1.1 Ultrasound imaging technique

The creation of a US image relies upon the physical properties of sound propagation in matter and its interactions with interfaces. A *sound* is a travelling wave, *i.e.* an oscillation of pressure transmitted through a solid, liquid, or gas, composed of frequencies within the range of hearing. An *ultrasound* is a vibration similar to a sound with a frequency above human hearing (20 kHz). Typical sonographic scanners operate in the frequency range of 2 to 18 megahertz, hundreds of times greater than the limit of human hearing. The choice of frequency is a trade-off between spatial resolution of the image and imaging depth: lower frequencies produce less resolution but image deeper into the body. Superficial structures such as muscles, tendons, testes, breast and the neonatal brain are imaged at a higher frequency (7-18 MHz), which provides better axial and lateral resolution. Deeper structures such as liver and kidney are imaged at a lower frequency (1-6 MHz) with lower axial and lateral resolution but greater penetration.

The actual creation of an image is done in three steps: producing a sound wave, receiving echoes, and interpreting those echoes.

Producing a sound wave A sound wave is typically produced by a piezoelectric transducer encased in a probe. The PZT (Pb Zr Ti) ceramic converts mecanical energy into electric energy and viceversa. As the ceramic receives strong, short electrical pulses, the PZT element vibrates and produces ultrasound waves at the desired frequency (see Figure A.2).

Materials on the face of the transducer enable the wave to be transmitted efficiently into the body. In addition, a water-based gel is placed between the patient's skin and the probe.

The sound is focused either by the shape of the transducer, a lens in front of the transducer, or a complex set of control pulses from the ultrasound scanner machine. This focusing produces an arc-shaped sound wave from the face of the transducer. The wave travels into the body and comes into focus at a desired depth.

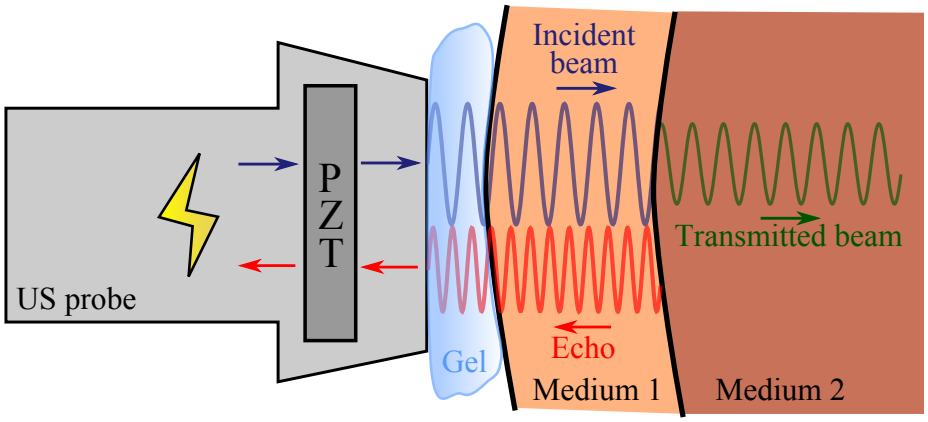


Figure A.2: Electric pulses are transformed by the piezoelectric transducer (PZT) into an ultrasound wave propagating in the patient's body thanks to the gel. The incident wave (blue) travels in the first medium until an interface is reached. An echo (red) reflects back to the origin while a part of the beam continues its propagation in the second medium (green). When the echo reaches the transducer, it is converted back to an electric pulse. The system therefore detects each interface between two different media.

Receiving the echoes The sound wave is partially reflected from the layers between different tissues. Specifically, sound is reflected anywhere there are density changes in the body such as at the surface of small structures in organs. Some of the reflections return to the transducer (see Figure A.2).

The return sound wave vibrates the piezo-electric ceramic, that turns the vibrations back into electrical pulses. That signal is then processed and transformed into a digital image by the ultrasound scan.

Forming the image Each pulse transmitted by the transducer explores a volume that can be assimilated to a line. Whenever a sound wave encounters a material with a different density (acoustical impedance), part of the sound wave is reflected back to the probe and is detected as an echo. Knowing the time taken by this echo to travel back to the transducer as well as the speed of sound in tissues, the ultrasound scan can determine the depth of the scatterer, *i.e.* the tissue interface causing the echo.

Morever knowing the strength of the echo and the attenuation of the pulse depending of the depth, the ultrasound scan determines the brightness of that point in the image (from black for a weak echo to white for a strong echo). With that information, a greyscale digital image can be created from the received signal with various representations.

A.1.2 From 1D to 3D Ultrasound

Several modes of representations for ultrasound signals are used in medical imaging.

A-mode The A-mode is the earliest type of ultrasound. A single transducer scans a line through the body with the strength of the reflected sound plotted as a function of its depth.

M-mode In this mode, a series of pulses are emitted very quickly which allows to acquire a time sequence of A-mode signals (for example). The echo amplitude is often represented on a 2D image as a function of the depth and time. This enables the evaluation of rapidly moving structures such as cardiac valves or chamber walls.

B-mode A 2D real-time image is built by concatenation of a series of scan lines that browse a plane through the body. Variation in displayed intensities reflect signals of different amplitudes. This is the standard way of acquiring ultrasound images and is often referred to as conventional US imaging.

Doppler This mode exploits the Doppler effect to detect and measure moving structures, such as a blood flow. In color Doppler, the velocity measurements are overlaid in color on a B-mode image.

In order to acquire even more information, three generations of 3D ultrasound systems have been developed. Initially, volume images were acquired by freehand sweeping a 2D US probe [Treece et al., 2003; Poon & Rohling, 2005; Hsu et al., 2008]. They can now be automatically acquired thanks to the advent of mechanical probes applying a mechanical sweep of 2D probe elements. However, since the mechanical scanning is slow, it is difficult to make 3D images of moving tissues. More recently, electronic probes using a 2D matrix of elements are available to acquire volumes at once through electronic scanning of the acoustic field [Fenster et al., 2001]. These can acquire volumes much faster (see Figure A.3) but at a cost of a reduced spatial resolution. 3D ultrasound offers a better knowledge of heterogeneous objects by providing volume information. Moreover it improves acquisition reproducibility by lessening the importance of the acquisition plan and thus the user dependence.

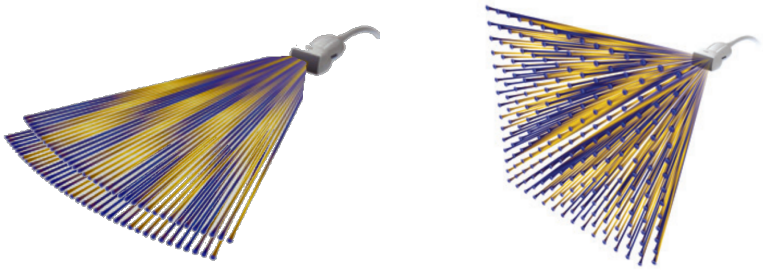


Figure A.3: Probes for 3D ultrasound acquisition. (Left) A 2D probe that acquires a volume by applying a mechanical sweep. (Right) A matrix probe that directly acquires 3D images.

A.1.3 Ultrasound artifacts

The generation process described hereabove explains the different sources of artifacts that may be a source of confusion when interpreting US images (see Figure A.4). They typically occur when one of the following physical assumptions, upon which relies the system relies, does not hold:

Single reflection The US system assumes that every echo returned to the tranducer comes from a single reflection and its depth only depends from its returning time. However when two interfaces are close to each other, undesirable interactions happen. Figure A.4.a shows a case of *reverberation* artifact: the emitted ultrasound is repeatedly reflected back and forth two highly reflective surfaces. This produces multiple and equally spaced fake interfaces in the final image. Another configuration is depicted in Figure A.4.b, where a *mirror-image* of a structure is created because of indirect reflections.

Origin of the echos All echos received by the transducer are supposed to be reflections from the emitted ultrasounds. Yet this incident wave may cause resonant vibrations when some fluid is trapped between air bubbles (Figure A.4.c). Such structure therefore emits a continuous sound that is interpreted as a long vertical interface, which is denoted as a *ring-down effect*.

Constant speed The sonographic instrument assumes that the acoustic velocity is constant at 1540 $m.s^{-1}$ within the whole scanned volume. Yet it does depend on the acoustical impedance of the material and several reflections occuring at the same depth may have a different returning time. This produces *speed displacement* artifacts, often seen in the diaphragm in liver imaging (Figure A.4.d).

Constant attenuation When travelling in a medium, a wave amplitude naturally decreases as its path lengthens. The ultrasound systems takes this into account by automatically enhancing echos that arrive later. By doing so, it assumes that the attenuation is only a function of the depth. In the case when the ultrasound wave crosses a particularly strong or weak attenuator (compared to its surrounding), there may be under- or over-compensation of its echo amplitude. This results in darker or brighter regions distal to the attenuator (Figure A.4.e,f), which are called *attenuation* artifacts.

Finally, although it is not really an artifact, the *speckle noise* should be mentioned as it is quite powerful in US images. Speckle is a random pattern in an image due to sub-resolution scatterers that radiate coherently. While it may be useful for tracking applications (since it is temporally coherent), it degrades the overall image quality and hinders diagnoses. Its correction is an active field of research (*e.g.* [Yu et al., 2010]) but is still considered as an open problem.

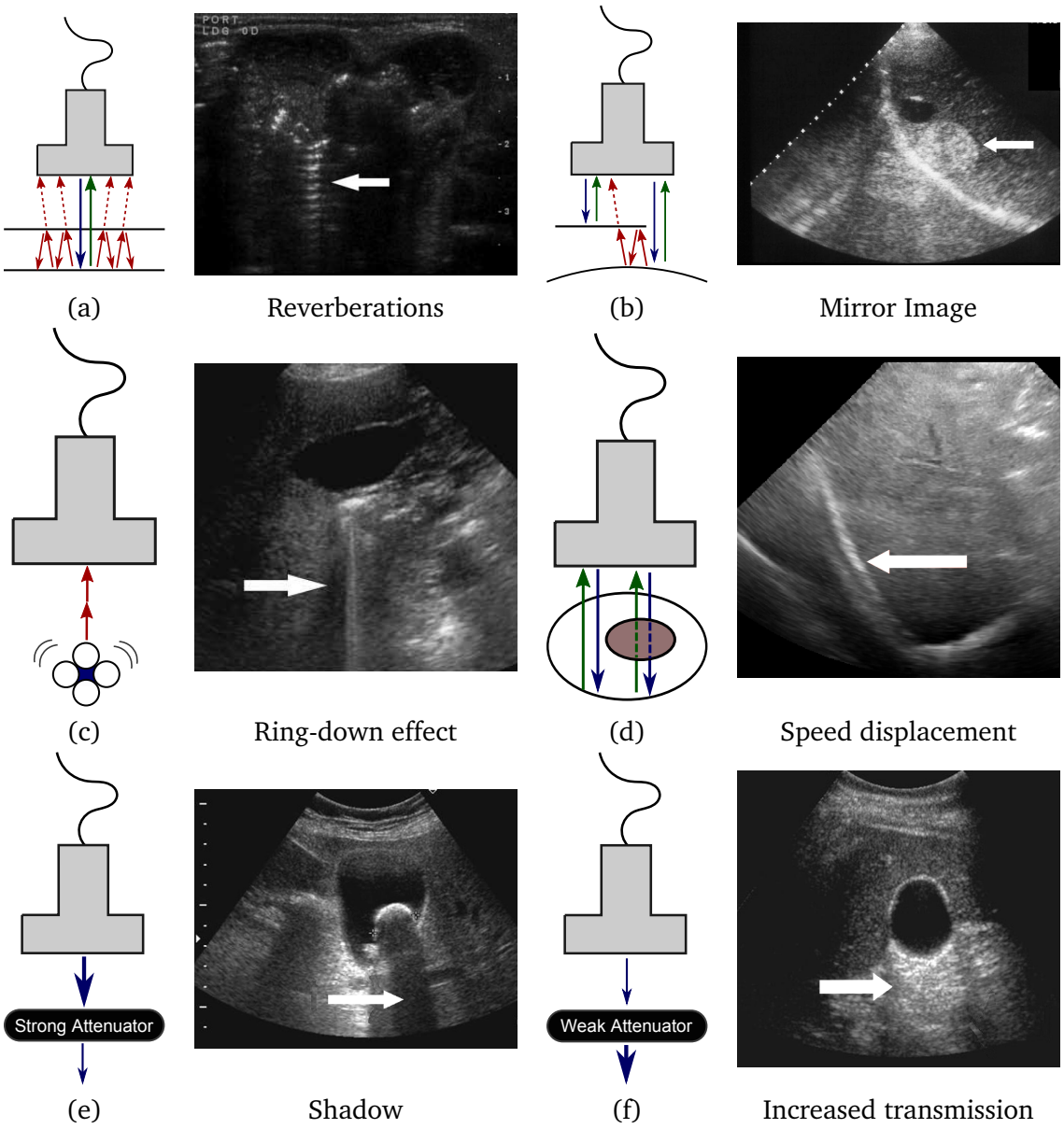


Figure A.4: Different artifacts observable in US images when the physical hypotheses assumed by the scanning maching do not hold. Incident waves are represented in blue, expected echos in green and undesired ones in red. (Source: [Aldrich, 2007; Feldman et al., 2009])

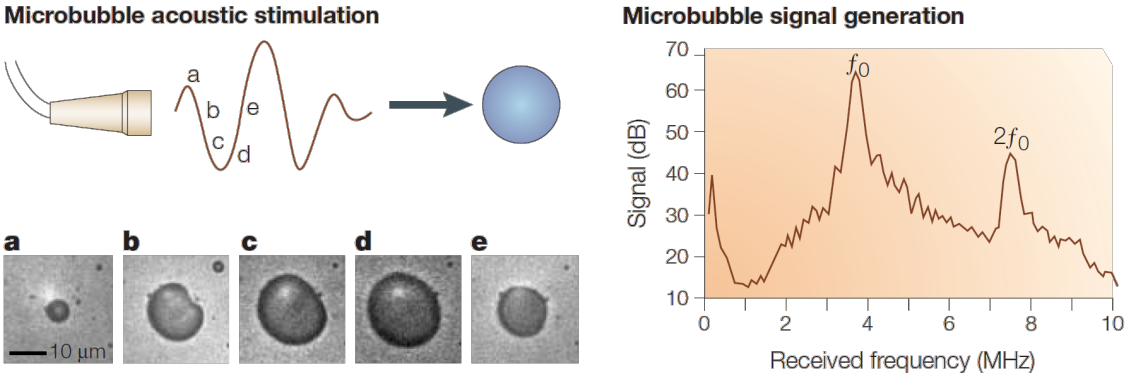


Figure A.5: (Left) Volumetric oscillation of a microbubble during exposure to ultrasound. The microbubble images were acquired under a light microscope each $330ns$. (Right) Frequency versus amplitude data from microbubbles demonstrating returning signal both at the fundamental f_0 and second harmonic $2f_0$ frequencies. (Source: [Lindner, 2004]).

A.2 Contrast-Enhanced Ultrasound (CEUS)

Intravenously injected contrast agents are widely used in most imaging modalities to visualize blood flow in the microcirculation and larger vasculatures. Perfusion assessment is indeed an important clinical differentiator, for instance in the context of lesion characterization. In ultrasound imaging, blood flow visualization can be achieved with Doppler-based techniques [Rumack et al., 2005]. However only high velocity flows ($> cm.s^{-1}$) can be detected this way. Another way of making a flow "visible" in ultrasound images is to inject scatterers into the patient's blood.

A.2.1 Microbubbles as contrast agent

The idea of injecting gas-filled microbubbles (typically 3 μm in diameter) as contrast agent is quite recent [Burns, 1997; Blomley & Cosgrove, 1997]. It is based on the fact that gas compressibility enables microbubbles to be an efficient scatterer, while their size allows them to circulate freely in most vessels [Cosgrove et al., 1998]. Contrary to contrast agents in CT which are filtered by the kidneys and may harm them (*i.e.* they are nephrotoxic), microbubbles used for contrast-enhanced ultrasound are completely safe for the patient. Indeed, the gas they contain is released through the lungs while their membrane is eliminated via natural metabolism processes. Another difference with CT or MR contrast agents is thus that they do not diffuse into the extracellular space.

Microbubbles work by resonating in response to an ultrasound beam, rapidly contracting and expanding in response to the pressure changes of the sound wave (see Figure A.5). Due to their complex oscillation, microbubbles produces a non-linear echo that allows them to be enhanced separately from the tissue echoes. Indeed if a pulse is transmitted at a frequency f, a microbubble will emit echoes at the fundamental frequency f_0 as well as harmonic signals, *i.e.* it emits echoes at the frequencies $2f_0$ and so on (Figure A.5).

A.2.2 Non-linear imaging methods

To retrieve only the signal emitted by the contrast agents, the first idea was to filter the second harmonic component of the backscattered echoes, which was believed to originate solely from the microbubbles. However it has been later found that the tissues also have a significant harmonic response and the separation was therefore not optimal. Besides this restricted the available bandwidth as the fundamental signal should not overlap with the harmonics. To overcome such limitations, several non-linear imaging methods were proposed.

Pulse Inversion imaging Rather than filtering out the fundamental signal, Pulse Inversion (PI) imaging uses a subtraction technique to cancel the fundamental echoes. Thus PI can separate the fundamental component of the bubble echoes from the harmonic even when they overlap. Two pulses are transmitted, the first being a normal pulse and the second being an inverted replica of the first one. Wherever there is a positive pressure on the first pulse there is an equal negative pressure on the second. These are then added and all stationary linear targets cancel. As PI cancels the fundamental component, it enhances the second harmonic component of the microbubble response when the backscattered signals are added together (see Figure A.6).

Power Modulation imaging Power Modulation (PM) imaging is an alternative to change the amplitude of successive pulses. PM detects the differential nonlinear response generated from two different excitations. A low amplitude pulse is transmitted to estimate the linear response of a target volume. Then a slightly higher amplitude pulse is transmitted to elicit a nonlinear response from the target volume. Upon reception, the lower amplitude is rescaled by the factor between transmit pulses and subtracted. The resulting difference at the fundamental frequency represents energy which has leaked out of the first pulse into the higher harmonics (see Figure A.7). Because it generally uses lower frequencies, Power Modulation has a better penetration but lower spatial resolution than Pulse Inversion.

Contrast Pulse Sequence imaging The first two principles can be combined simultaneously: two pulses are generated with both different amplitude and scale. This method is called Contrast Pulse Sequence, or simply Pulse Inversion with Power Modulation. Contrast Pulse Sequence reaches a compromise between the two in terms of resolution and penetration. Another advantage lies in its ability to detect nonlinearities at both fundamental and second harmonic frequencies.

Note that one of the pulse can still be used to form a conventional B-mode image of the observed organ. From a physical point of view, it is therefore possible to acquire *at the same time* a conventional and a contrast-enhanced ultrasound image. However, due to hardware limitations, this functionality is only available for 2D acquisitions at the time of writing.

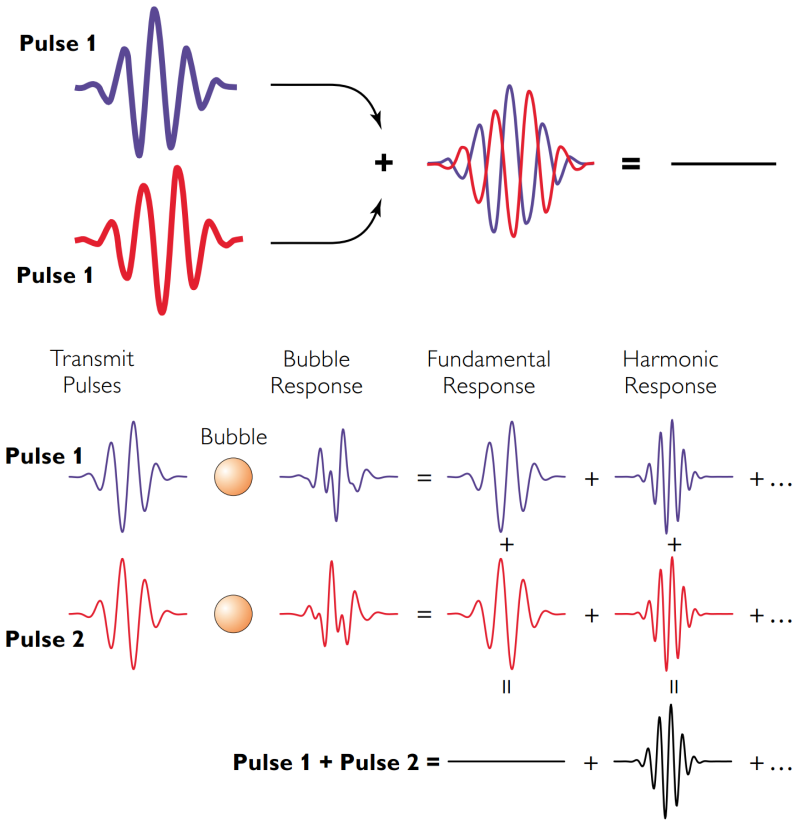


Figure A.6: By adding two echoes from inverted pulses, Pulse Inversion consists in cancelling fundamental echoes and leaving only the nonlinear components generated by microbubbles.

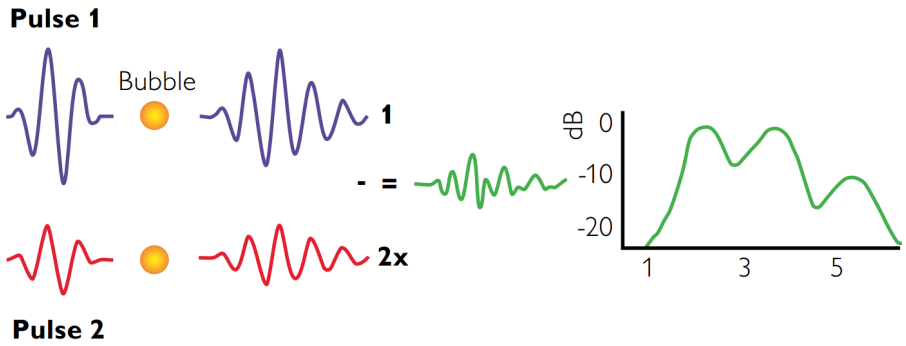


Figure A.7: Power Modulation consists in sending two pulses with various amplitude and then rescaling and subtracting the responses, to retain only the nonlinear energy.

A.2.3 Applications of CEUS imaging

Although solely licensed for echocardiography in the United States, contrast agents are used in a wide variety of applications in the rest of the world [Albrecht et al., 2004]. It is still developping and it even tends to replace CT or MR examinations in specific contexts.

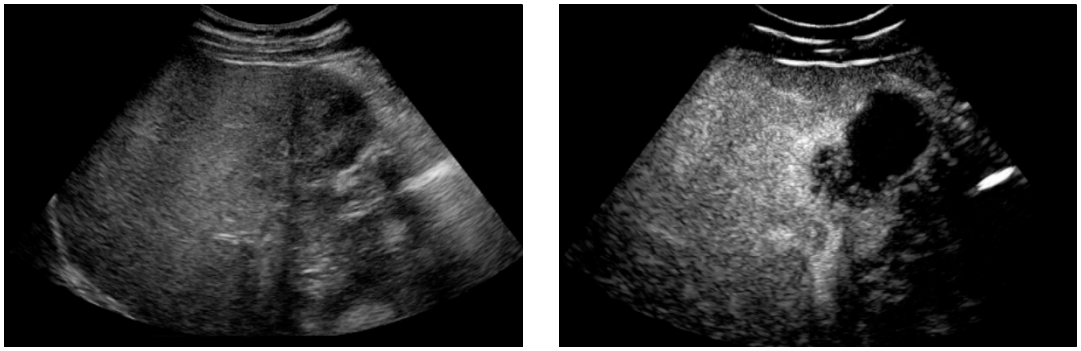


Figure A.8: Metastatis visualization in the liver with conventional ultrasound (left) and contrast-enhanced ultrasound (right). The lesion is much more visible in the latter image.

One of its most important benefits lies in the detection and characterisation of lesions in abdominal organs, such as the liver, the kidney or the pancreas. On a related subject, contrast agents are also used for breast ultrasound for it improves the differential diagnosis of solid masses. Patients with strokes can also benefit from this modality as it enables hemorrhage detection in the brain. Furthermore, therapy planning and guidance are also improved with the use of contrast-enhanced ultrasound, since target lesions are much more visible: Figure A.8 shows an example of liver metastatis with and without injection of microbubbles. Finally, thanks to the temporal resolution of current ultrasound imaging systems, it it possible to do perfusion imaging, *i.e.* estimating pharmacokinetic parameters from the diffusion of the contrast agent along the time.

Appendix B

Gradient equations for the implicit template deformation functional

In this appendix, we detail the calculus of variation that lead to the gradient equations of the implicit template deformation in (2.51). We limit our calculations to the standard formulation that was introduced in [Mory, 2011]. All the other generalizations proposed throughout the thesis can be obtained by a very similar approach.

The equation below recalls the expression of the energy of implicit template deformation, that must be minimized with respect to the pose transformation G and the deformation L (more details are available in Section 2.7):

$$\min_{L,G} \quad \left\{ E(L,G) = \int_\Omega H(\phi_0 \circ L \circ G(\mathbf{x}))\; r(\mathbf{x})\; d\mathbf{x} + \frac{\lambda}{2}\|L - \mathbf{Id}\|_2^2 \right\} \quad \text{with} \quad L = \mathbf{Id} + \mathbf{u} = \mathbf{Id} + K_\sigma * \mathbf{v} \tag{B.1}$$

The two following sections, we compute the gradient of this energy with respect to the displacement field and the parameters of the pose respectively.

B.1 Gradient with respect to the deformation field

The deformation $L : \Omega_0 \to \Omega_0$ is parametrized by the displacement field $\mathbf{v} \in L^2(\Omega_0, \mathbb{R}^d)$ such that

$$L_\mathbf{v} = \mathbf{Id} + K_\sigma * \mathbf{v} \tag{B.2}$$

which yields, if we denote $\mathbf{v}_t = \mathbf{v} + t.\eta$ a perturbation of $\mathbf{v}$ in the direction $\eta \in L^2(\Omega_0, \mathbb{R}^d)$,

$$\frac{\partial L}{\partial t} = \lim_{t \to 0} \frac{L_{\mathbf{v}_t} - L_{\mathbf{v}_t}}{t} = K_\sigma * \eta \tag{B.3}$$

We then compute $D_\eta E$ the Gateaux-derivative of E with respect to $\mathbf{v}$ in the direction η, and try to express is as a scalar product to exhibit its gradient. To that end, we first apply a

variable substitution to the image-based term in (B.1)

$$E = |J_{G^{-1}}| \int_{G(\Omega)} H(\phi_0 \circ L)\; r \circ G^{-1}(\mathbf{x})\; d\mathbf{x} + \frac{\lambda}{2}\|L - \mathbf{Id}\|_2^2 \tag{B.4}$$

where $|J_{G^{-1}}|$ is the determinant of the Jacobian matrix of G. Expanding the L^2-norm and assuming that $G(\Omega) \subset \Omega_0$ gives

$$E = \int_{\Omega_0} \left(H(\phi_0 \circ L)\; r \circ G^{-1}(\mathbf{x})\; |J_{G^{-1}}| + \frac{\lambda}{2}\; (K_\sigma * \mathbf{v})^2(\mathbf{x}) \right) d\mathbf{x} \tag{B.5}$$

We now use the chain rule to derive this expression of the energy:

$$\begin{aligned} D_\eta E = \int_{\Omega_0} \Big(& \delta(\phi_0 \circ L(\mathbf{x})) \left\langle \nabla\phi_0 \circ L(\mathbf{x}),\; K_\sigma * \eta(\mathbf{x}) \right\rangle\; r \circ G^{-1}(\mathbf{x})\; |J_{G^{-1}}| \\ & + \lambda\; \left\langle K_\sigma * \eta(\mathbf{x}),\; K_\sigma * \mathbf{v}(\mathbf{x}) \right\rangle \Big)\; d\mathbf{x} \end{aligned} \tag{B.6}$$

In this equation, the term $K_\sigma * \eta(\mathbf{x})$ can be factorized, which yields

$$D_\eta E = \int_{\Omega_0} \left\langle K_\sigma * \eta(\mathbf{x}),\; \delta(\phi_0 \circ L)\; \nabla\phi_0 \circ L(\mathbf{x})\; r \circ G^{-1}(\mathbf{x})\; |J_{G^{-1}}| + \lambda\; K_\sigma * \mathbf{v}(\mathbf{x}) \right\rangle\; d\mathbf{x} \tag{B.7}$$

We then express $K_\sigma * \eta(\mathbf{x})$ as the integral $\int_{\Omega_0} K_\sigma(\mathbf{x}-\mathbf{y})\; \eta(\mathbf{y})\; d\mathbf{y}$, and change the order of integration of $\mathbf{x}$ and $\mathbf{y}$. We thus obtain

$$D_\eta E = \int_{\mathbf{y}\in\Omega_0} \left\langle \eta(\mathbf{y}),\; \int_{\mathbf{x}\in\Omega_0} K_\sigma(\mathbf{x}-\mathbf{y})\; \left[\delta(\phi_0 \circ L(\mathbf{x}))\; \nabla\phi_0 \circ L(\mathbf{x})\; r \circ G^{-1}(\mathbf{x})\; |J_{G^{-1}}| + \lambda\; K_\sigma * \mathbf{v}(\mathbf{x}) \right] d\mathbf{x} \right\rangle d\mathbf{y} \tag{B.8}$$

As such, (B.8) shows that the gradient of E with respect to $\mathbf{v}$, *i.e.* the direction of η that maximizes the energy variation $D_\eta E$, is given by

$$\nabla_{\mathbf{v}} E(\mathbf{y}) = \int_{\mathbf{x}\in\Omega_0} K_\sigma(\mathbf{x}-\mathbf{y})\; \left[\delta(\phi_0 \circ L(\mathbf{x}))\; \nabla\phi_0 \circ L(\mathbf{x})\; r \circ G^{-1}(\mathbf{x})\; |J_{G^{-1}}| + \lambda\; K_\sigma * \mathbf{v}(\mathbf{x}) \right] d\mathbf{x} \tag{B.9}$$

Finally, we point out that K_σ is a symmetric Gaussian kernel so $K_\sigma(\mathbf{x}-\mathbf{y}) = K_\sigma(\mathbf{y}-\mathbf{x})$. We are then able to express $\nabla_{\mathbf{v}} E$ as a convolution :

$$\nabla_{\mathbf{v}} E = K_\sigma * \left[\delta(\phi_0 \circ L)\; .\; \nabla\phi_0 \circ L\; .\; r \circ G^{-1}\; |J_{G^{-1}}| + \lambda\; K_\sigma * \mathbf{v} \right] \tag{B.10}$$

which is the desired result.

B.2 Gradient with respect to the pose

The pose $G : \Omega \to \Omega_0$ is parametrized with a finite set of parameters $\mathbf{p}$. For example if G is a similarity, $\mathbf{p}$ is an element of $\mathbb{R}^7$ and represents the three directions of translation, the three angles of rotation and the scaling factor. Therefore we need to compute the derivatives of E with respect to each entry of $\mathbf{p}$, that we denote $\mathbf{p}_i$ for $i \in [\![1, 7]\!]$.

As in the previous section, we apply the chain rule to (B.1), which yields the following expression for the energy gradient

$$\frac{\partial E}{\partial \mathbf{p}_i} = \int_\Omega \delta(\phi_0 \circ L \circ G(\mathbf{x})) \left\langle \nabla \phi_0 \circ L \circ G(\mathbf{x}) \, , \, \frac{\partial}{\partial \mathbf{p}_i}(L \circ G)(\mathbf{x}) \right\rangle r(\mathbf{x}) \, d\mathbf{x} \tag{B.11}$$

We now point out that

$$\frac{\partial}{\partial \mathbf{p}_i}(L \circ G)(\mathbf{x}) = D_L \circ G(\mathbf{x}) \, . \, \frac{\partial G}{\partial \mathbf{p}_i}(\mathbf{x}) = (\mathbf{I} + \mathbf{J}_\mathbf{u}) \circ G(\mathbf{x}) \, . \, \frac{\partial G}{\partial \mathbf{p}_i}(\mathbf{x}) \tag{B.12}$$

where $\mathbf{I}$ is the Identity matrix in $M_d(\mathbb{R})$ and $\mathbf{J}_\mathbf{u}(\mathbf{x})$ is the Jacobian matrix of the displacement field $\mathbf{u}$ at point $\mathbf{x}$. The transformation derivative $\frac{\partial G}{\partial \mathbf{p}_i}$ is also a matrix in $M_d(\mathbb{R})$. In (B.12), dots represent the standard matrix product. In particular, we have

$$\frac{\partial}{\partial \mathbf{p}_i}(L \circ G) \circ G^{-1}(\mathbf{x}) = (\mathbf{I} + \mathbf{J}_\mathbf{u})(\mathbf{x}) \, . \, \left(\frac{\partial G}{\partial \mathbf{p}_i} \circ G^{-1} \right)(\mathbf{x}) \tag{B.13}$$

which, combined with (B.11), gives after the variable substitution $\mathbf{x} \leftarrow G(\mathbf{x})$ the expression

$$\frac{\partial E}{\partial \mathbf{p}_i} = \int_\Omega \delta(\phi_0 \circ L(\mathbf{x})) \left\langle \nabla \phi_0 \circ L(\mathbf{x}) \, , \, (\mathbf{I} + \mathbf{J}_\mathbf{u})(\mathbf{x}) \, . \, \left(\frac{\partial G}{\partial \mathbf{p}_i} \circ G^{-1} \right)(\mathbf{x}) \right\rangle r \circ G^{-1}(\mathbf{x}) \, |J_{G^{-1}}| \, d\mathbf{x} \tag{B.14}$$

Finally, the gradient of the energy E with respect to the parameter $\mathbf{p}_i$ is given by

$$\nabla_{\mathbf{p}_i} E = \int_\Omega \delta(\phi_0 \circ L) \left\langle \nabla \phi_0 \circ L \, , \, (\mathbf{I} + \mathbf{J}_\mathbf{u}) \, . \, \frac{\partial G}{\partial \mathbf{p}_i} \circ G^{-1} \right\rangle r \circ G^{-1} \, |J_{G^{-1}}| \tag{B.15}$$

which corresponds to (2.51).