ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM
ParisTech

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Signal et Images »

présentée et soutenue publiquement par

Romane GAURIAU

le 22 juin 2015

Méthodes multi-organes rapides avec a priori de forme pour la localisation et la segmentation en imagerie médicale 3D

Directeur de thèse : **Isabelle BLOCH**
Co-encadrement de la thèse : **David LESAGE**

Jury
M. Hervé DELINGETTE, Directeur de recherche, INRIA, Sophia-Antipolis   Rapporteur
M. Daniel RUECKERT, Professeur, Imperial College, Londres, Grande-Bretagne   Rapporteur
Mme Catherine ADAMSBAUM, Professeur, APHP CHU Le Kremlin Bicêtre, Université Paris Sud   Examinateur
M. Olivier BERNARD, Maître de conférence, INSA Creatis, Lyon   Examinateur
M. Jean-Philippe THIRAN, Professeur associé, EPFL, Lausanne, Suisse   Examinateur

T
H
È
S
E

EDITE - ED 130

T H E S I S

submitted in fulfillement of the requirements for the degree of

Doctor of Philosophy of TELECOM ParisTech

with specialty « Signal and Images »

defended by

Romane GAURIAU

on June 22nd, 2015

Shape-based approaches for fast multi-organ localization and segmentation in 3D medical images

Supervisor : **Isabelle BLOCH**
Co-supervisor : **David LESAGE**

Jury

Mr Hervé DELINGETTE, Research director, INRIA, Sophia-Antipolis		Reviewer
Mr Daniel RUECKERT, Professor, Imperial College, London, UK		Reviewer
Mrs Catherine ADAMSBAUM, Professor, APHP CHU Le Kremlin Bicêtre, Université Paris Sud		Examiner
Mr Olivier BERNARD, Researcher, INSA Creatis, Lyon		Examiner
Mr Jean-Philippe THIRAN, Associate professor, EPFL, Lausanne, Swiss		Examiner

T H È S E

TELECOM ParisTech

école de l'Institut Mines-Télécom - membre de ParisTech

ABSTRACT

With the ubiquity of imaging in medical applications (diagnostic, treatment follow-up, surgery planning...), image processing algorithms have become of primary importance. Algorithms help clinicians extract critical information more quickly and more reliably from increasingly large and complex acquisitions. In this context, anatomy localization and segmentation is a crucial component in modern clinical workflows. Due to particularly high requirements in terms of robustness, accuracy and speed, designing such tools remains a challenging task.

In this work, we propose a complete pipeline for the segmentation of multiple organs in medical images. The method is generic, it can be applied to varying numbers of organs, on different imaging modalities. Our approach consists of three components: (i) an automatic localization algorithm, (ii) an automatic segmentation algorithm, (iii) a framework for interactive corrections. We present these components as a coherent processing chain, although each block could easily be used independently of the others.

To fulfill clinical requirements, we focus on robust and efficient solutions. Our anatomy localization method is based on a cascade of Random Regression Forests (Cuingnet et al., 2012). One key originality of our work is the use of shape priors for each organ (thanks to probabilistic atlases). Combined with the evaluation of the trained regression forests, they result in shape-consistent confidence maps for each organ instead of simple bounding boxes. Our segmentation method extends the implicit template deformation framework of Mory et al. (2012) to multiple organs. The proposed formulation builds on the versatility of the original approach and introduces new non-overlapping constraints and contrast-invariant forces. This makes our approach a fully automatic, robust and efficient method for the coherent segmentation of multiple structures. In the case of imperfect segmentation results, it is crucial to enable clinicians to correct them easily. We show that our automatic segmentation framework can be extended with simple user-driven constraints to allow for intuitive interactive corrections. We believe that this final component is key towards the applicability of our pipeline in actual clinical routine.

Each of our algorithmic components has been evaluated on large clinical databases. We illustrate their use on CT, MRI and US data and present a user study gathering the feedback of medical imaging experts. The results demonstrate the interest in our method and its potential for clinical use.

RÉSUMÉ COURT

Avec l'utilisation de plus en plus répandue de l'imagerie dans la pratique médicale (diagnostic, suivi, planification d'intervention, etc.), le développement d'algorithmes d'analyse d'images est devenu primordial. Ces algorithmes permettent aux cliniciens d'analyser et d'interpréter plus facilement et plus rapidement des données de plus en plus complexes. Dans ce contexte, la localisation et la segmentation de structures anatomiques sont devenues des composants critiques dans les processus cliniques modernes. La conception de tels outils pour répondre aux exigences de robustesse, précision et rapidité demeure cependant un réel défi technique.

Ce travail propose une méthode complète pour la segmentation de plusieurs organes dans des images médicales. Cette méthode, générique et pouvant être appliquée à un nombre varié de structures et dans différentes modalités d'imagerie, est constituée de trois composants : (i) un algorithme de localisation automatique, (ii) un algorithme de segmentation, (iii) un outil de correction interactive. Ces différentes parties peuvent s'enchaîner aisément pour former un outil complet et cohérent, mais peuvent aussi bien être utilisées indépendemment.

L'accent a été mis sur des méthodes robustes et efficaces afin de répondre aux exigences cliniques. Notre méthode de localisation s'appuie sur une cascade de régression par forêts aléatoires (Cuingnet et al., 2012). Elle introduit l'utilisation d'informations a priori de forme, spécifiques à chaque organe (grâce à des atlas probabilistes) pour des résultats plus cohérents avec la réalité anatomique. Notre méthode de segmentation étend la méthode de segmentation par modèle implicite (Mory et al., 2012) à plusieurs modèles. La formulation proposée permet d'obtenir des déformations cohérentes, notamment en introduisant des contraintes de non recouvrement entre les modèles déformés. En s'appuyant sur des forces images polyvalentes, l'approche proposée se montre robuste et performante pour la segmentation de multiples structures. Toute méthode automatique n'est cependant jamais parfaite. Afin que le clinicien garde la main sur le résultat final, nous proposons d'enrichir la formulation précédente avec des contraintes fournies par l'utilisateur. Une optimisation localisée permet d'obtenir un outil facile à utiliser et au comportement intuitif. Ce dernier composant est crucial pour que notre outil soit réellement utilisable en pratique.

Chacun de ces trois composants a été évalué sur plusieurs grandes bases de données cliniques (en tomodensitométrie, imagerie par résonance magnétique et ultrasons). Une étude avec des utilisateurs nous a aussi permis de recueillir des retours positifs de plusieurs experts en imagerie médicale. Les différents résultats présentés dans ce manuscrit montrent l'intérêt de notre méthode et son potentiel pour une utilisation clinique.

RÉSUMÉ LONG

CONTEXTE ET PRÉSENTATION

Introduction

Ce document vise à donner un résumé en français cette thèse. Le contenu de chaque chapitre est synthétisé afin de mettre en avant les différentes contributions développées dans la thèse. Des liens renvoient vers les différentes sections et figures de la thèse.

Contexte

L'imagerie médicale est aujourd'hui un élément fondamental dans la pratique médicale. Cette technique permet aux cliniciens d'obtenir une représentation de l'intérieur du corps, laquelle donne des informations clés pour le diagnostic, la planification de traitement ou d'opérations par exemple. Différents types d'imagerie sont aujourd'hui disponibles et permettent de répondre à différents besoins. On distingue deux types de technologies d'imagerie : l'imagerie fonctionnelle et l'imagerie structurelle. La première donne des informations sur des aspects physiologiques (par exemple l'imagerie tomographique par émission de positron, TEP) tandis que la deuxième donne des informations sur les tissus et l'anatomie (par exemple la tomo-densitométrie, TDM, ou l'imagerie par résonance magnétique, IRM). Dans ce travail nous abordons ce deuxième type d'imagerie.

La compréhension et l'interprétation des images médicales demande une grande expérience et un haut niveau d'expertise. En effet, les acquisitions contiennent une information complexe et variée, plus ou moins difficile à analyser suivant :

- la variabilité intrinsèque à la modalité d'imagerie (contraste, bruit, résolution, dimension, biais, champ de vue etc.),

- la variabilité inter-patient (par exemple la taille, la forme, l'apparence de l'organe, voir figure 1.1),

- la variabilité intra-patient (par exemple les organes mous peuvent bouger ou changer d'aspect avec la respiration ou avec le temps).

Depuis plusieurs décades, ces différents défis ont conduit au développement de méthodes assistées par ordinateur pour la simplification et l'accélération de la compréhension et de l'interprétation des images médicales. L'analyse

d'image est aujourd'hui un champ de recherche à part entière avec sa propre communauté de chercheurs. En particulier, deux sujets très actifs concernent la localisation et la segmentation (ou délinéation) de structures anatomiques. La localisation consiste à trouver la position de la structure dans l'image, tandis que la segmentation consiste à en trouver les contours. Cette information est précieuse pour les médecins et permet de multiples applications, de la visualisation à la quantification (par exemple le calcul de volume). Les récents progrès dans l'automatisation de ces tâches ont par ailleurs ouvert de nouvelles possibilités dans la pratique clinique. Encore plus récemment, l'extension de telles méthodes à plusieurs structures ou organes permet aujourd'hui de générer des modèles anatomiques spécifiques à chaque patient. Cela permet à la fois d'appréhender l'anatomie dans sa globalité et de contribuer à de nouvelles procédures médicales (par exemple en chirurgie laparoscopique).

Cependant, malgré les avancées dans ce domaine, l'automatisation de la localisation et segmentation multi-organe reste un champ très actif, car il y a encore de nombreux défis à surmonter pour prendre en compte les variabilités énumérées précédemment.

Les principaux requis pour qu'un outil de segmentation soit utilisable en contexte clinique sont les suivants :

- la robustesse : l'algorithme doit être utilisable dans des conditions variées (liées à la modalité et au patient),

- la précision : l'algorithme doit donner des résultats précis et exploitables,

- la rapidité : le temps de calcul doit être court,

- facilité d'utilisation : l'algorithme doit requérir le moins d'intervention possible de l'utilisateur.

Le travail présenté dans cette thèse tente de répondre à ces critères. Nous avons développé une chaîne complète et modulaire pour la localisation et segmentation de plusieurs organes. La solution proposée est générique, robuste et cohérente. Chaque composant peut aussi être adapté en fonction de l'application et la modalité. Un grand soin a été porté pour garder la solution aussi simple et générique que possible afin de répondre aux contraintes de temps de calcul nécessaires pour une utilisation clinique.

Description du contenu du manuscrit

Le manuscrit est structuré autour du design de la chaîne complète de segmentation multi-organe, de la localisation aux corrections interactives. Cette chaîne est composée de deux modules principaux qui sont indépendants :

- une approche de localisation reposant sur une cascade de forêts aléatoires de régression avec informations de forme a priori,

- une approche de segmentation multi-modale fondée sur la déformation de plusieurs modèles et pouvant donc être utilisée :
  - en mode automatique ou semi-automatique,
  - en mode interactif, pour la correction de segmentations existantes.

Les composants proposés peuvent être utilisés indépendamment ou former un ensemble cohérent pour la segmentation automatique. Le résultat de localisation, fourni sous la forme de cartes de probabilités dédiées à chaque organe, sont assez précises et informatives pour être utilisées telles quelles (par exemple pour la visualisation de données). Le module de segmentation peut aussi être utilisé à différentes fins. Il peut fournir des résultats de segmentation complètement automatique ou intégrer des interactions de l'utilisateur pour la correction de plusieurs contours.

La chaîne complète est illustrée sur la figure 1.2 du manuscrit.

Plusieurs annexes sont aussi ajoutées à la fin du document. Elles sont principalement liées au différentes collaborations réalisées durant la thèse. L'annexe A présente nos travaux préliminaires pour la segmentation du foie en TDM 3D. Ce travail a été réalisé en tout début de thèse. Il a ouvert la voie vers les méthodes développées par la suite. L'annexe B présente des résultats sur la segmentation automatique du foie en IRM 3D. Cette méthode exploite notre travail sur la localisation et met en valeur cette contribution. L'annexe C décrit une autre application de l'algorithme de localisation dans des images ultrasonores 3D. Il montre que notre approche peut être utilisée avec d'autres types de modalités et différentes structures anatomiques. L'annexe D donne des détails sur les calculs variationnels des chapitres 4 et 5. L'annexe E présente une méthode pour la construction de modèle moyen d'organe s'appuyant sur la représentation de forme implicite.

ETAT DE L'ART

Ce chapitre vise à donner les principaux fondements et un aperçu bibliographique pour comprendre le contexte méthodologique de nos travaux. La section 2.2 présente un aperçu des différentes méthodes d'apprentissage supervisé avec un accent particulier mis sur les forêts de décision. En particulier nous décrivons les méthodes récentes adaptées à la localisation fondées sur des variantes des forêts de décision.

Bibliographie sur les méthodes de localisation multi-structures

La section 2.3 présente les principales méthodes proposées jusqu'à aujourd'hui pour la localisation de plusieurs structures en imagerie médicale. Nous catégorisons ces méthodes en quatre types différents :

- les méthodes exploitant l'image,

- les méthodes fondées sur des atlas,

- les méthodes par graphes,

- les méthodes par apprentissage.

Chacun de ces types de méthodes peut être potentiellement utilisé suivant le type d'application. Le choix va surtout dépendre de la variabilité inhérente au type d'imagerie, la qualité des images, la variabilité du type de structures à détecter et la reproductibilité. Les méthodes exploitant l'image sont suffisantes pour des problèmes plutôt simples. Les méthodes par atlas sont plutôt performantes dans le cas de très nombreuses structures (les structures du cerveau par exemple). Les méthodes par apprentissage sont plutôt dédiées à des problèmes avec une grande variabilité d'entrée. La bibliographie proposée dans le manuscrit met en avant l'avantage des méthodes par apprentissage pour la localisation multi-objets. Elles permettent une bonne généralisation par l'exemple et donnent des solutions performantes. Plus spécifiquement, les méthodes par régression semblent donner des solutions rapides. Dans notre travail, nous proposons donc d'explorer ce type d'approches pour obtenir un bon compromis entre robustesse et temps d'exécution.

Bibliographie sur les méthodes de segmentation multi-organes

La section 2.4 fait le bilan des différentes méthodes existantes pour la segmentation multi-objets en imagerie médicale. Nous catégorisons ces méthodes en cinq groupes :

- les méthodes exploitant l'image,

- les méthodes fondées sur des atlas,

- les méthodes par graphes,

- les méthodes par apprentissage,

- les méthodes par déformation de modèle.

Les méthodes existantes de segmentation multi-organes sont généralement des généralisations de méthodes mono-organe. Cependant l'aperçu donné dans la section 2.4 montre que l'extension au multi-organe n'est pas toujours évidente. La modélisation des relations inter-organes (non recouvrement, inclusion, adjacence) apparaît comme le point critique de ces méthodes. Le but est en effet de pouvoir produire un résultat cohérent.

La littérature à ce sujet est encore aujourd'hui relativement parcimonieuse mais malgré tout en rapide évolution. L'incorporation d'information de haut

niveau paraît être de première importance pour obtenir des méthodes robustes et précises. Avec le développement de bases de données de plus en plus larges, ce type d'information devient de plus en plus accessible. La difficulté réside maintenant plutôt dans l'extraction et la modélisation de ces informations de haut niveau et leur incorporation dans des méthodes existantes tout en gardant flexibilité et efficacité. Notre travail tente de répondre à de tels critères.

LOCALISATION AVEC INFORMATION A PRIORI DE FORME

Le chapitre 3 du manuscrit présente une nouvelle méthode pour la localisation de plusieurs organes dans des images médicales 3D. La méthode proposée donne des résultats rapides, robustes et précis démontrés par une évaluation quantitative.

Principe de la méthode

L'approche est fondée sur une cascade de forêts aléatoires de régression avec une stratégie globale à locale comme dans les travaux de Criminisi et al. (2013); Cuingnet et al. (2012). Nous proposons d'améliorer la robustesse et la précision de cette approche en introduisant de l'information de forme a priori grâce à l'exploitation d'atlas probabilistes spécifiques à chaque organe. La figure 3.1 résume le déroulement du processus proposé.

La méthode fonctionne en deux étapes globale et locale et fat intervenir des méthodes d'apprentissage supervisé. Dans un premier temps les données d'apprentissage sont utilisées pour construire des atlas probabilistes pour chaque organe. La méthode, décrite en section 3.2.3, vise principalement à aligner et moyenner des masques binaires de segmentation manuelle.

Les données d'apprentissage sont ensuite utilisées pour apprendre plusieurs régresseurs en suivant la méthode proposée par (Cuingnet et al., 2012). L'idée est d'apprendre une forêt d'arbres aléatoires pour prédire simultanément la position des boîtes englobantes de chaque organe à partir de descripteurs calculés dans l'image. Ces descripteurs sont une généralisation des descripteurs de Haar et correspondent à des différences d'intensité dans deux boîtes tirées aléatoirement (voir figure 3.2). Le principe est ensuite de faire voter différents voxels de l'image pour la position des organes et de moyenner ces votes pour estimer la position finale. Dans nos travaux, nous proposons de combiner la distribution de ces votes avec les atlas probabilistes de chaque organe. Chaque vote est combiné avec l'atlas probabiliste et accumulé dans une carte. Le résultat est une carte de confiance pour chaque organe qui encode la probabilité pour un voxel de l'image d'appartenir à l'organe en question. Ce principe est illustré à la figure 3.4. Une solution est par ailleurs proposée pour calculer rapidement ces cartes.

Cette méthode est ensuite utilisée dans une approche en cascade en suivant les travaux de (Cuingnet et al., 2012). Dans un premier temps tous les organes sont localisés grâce à une unique forêt aléatoire qui exploite des caractéristiques à une échelle plutôt grande. Une première localisation grossière peut alors être obtenue et les cartes de confiances sont calculées. Dans un second temps il s'agit d'utiliser des forêts aléatoires dédiées à chaque organes et correspondant à de l'information plus localisée. Ceci permet de raffiner la localisation.

Evaluation

La méthode est évaluée sur une base de donnée assez variée de 130 volumes TDM et pour la localisation de 6 organes abdominaux (le foie, les deux reins, la rate, la vésicule biliaire et l'estomac). Une étude étendue des différents paramètres d'apprentissage et de test est tout d'abord réalisée. Celle-ci montre une bonne robustesse à des variations même moyennes de ces paramètres. L'influence de ces différents paramètres est aussi étudiée.

Par la suite, la précision et la robustesse de l'algorithme sont évalués et démontrés. En comparant ces résultats avec deux méthodes existantes Criminisi et al. (2013); Cuingnet et al. (2012) nous démontrons une amélioration significative de performance pour un temps de calcul qui reste très raisonnable. En effet, la localisation de ces 6 organes abdominaux est réalisée en environ 5 secondes.

Une partie de ce travail a été publiée dans Gauriau et al. (2014) et une version plus complète a été publiée dans Gauriau et al. (2015).

SEGMENTATION MULTI-ORGANES AVEC DÉFORMATION DE MODÈLES

Le chapitre 4 du manuscrit présente une nouvelle formulation pour la segmentation de plusieurs organes par déformation de modèles. L'approche s'inspire du travail de (Mory et al., 2012) sur la déformation de modèle de forme implicite et dédié à la segmentation mono-organe. La section 4.2 pose les bases de ce travail. L'idée est d'optimiser la transformation qui déforme un modèle de forme fixe pour correspondre à l'image en suivant des forces calculées à partir de l'image. L'efficacité de cette méthode repose sur différents aspects : la décomposition de la transformation en une transformation globale (la pose) et un champ de déformations locales, la régularisation et le lissage de ces déformation locales et une régularisation permettant de contrôler la déviation par rapport au modèle. Par ailleurs, une optimisation efficace permet d'obtenir des temps de calcul de l'ordre de la seconde.

Principe

En section 4.3 nous montrons comment dériver une nouvelle formulation pour la segmentation multi-organes. La formulation est par ailleurs enrichie avec une contrainte de non recouvrement entre les différents organes voisins. Cette formulation peut être résumée par l'équation suivante :

$$
E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_{\Omega} H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2
$$
$$
\text{subject to } \forall (i,j) \in [\![1, N]\!]^2, i < j,
$$
$$
C(\mathcal{G}_i, \mathcal{G}_j) = \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x}))d\mathbf{x} = 0
$$

(0.1)

où N est le nombre d'objets, Ω est le domaine de l'image, $\mathcal{G}_1, ..., \mathcal{G}_N$ correspondent aux transformations de pose pour chaque objet, $\mathcal{L}$ correspond à la déformation locale commune à tous les objets, H est la fonction de Heaviside, ϕ_n est la fonction implicite de l'objet n, f_n est le terme de force de l'objet n, λ est une constante, Id est la fonction identité.

La dernière ligne de cette équation correspond à la contrainte de non recouvrement entre les différents objets déformés. Elle pénalise le recouvrement de toute paire d'objets.

L'idée sous-jacente de cette approche est d'assigner une transformation de pose pour chaque objet afin que chacun soit positionné par rapport aux autres en répondant aux contraintes de non recouvrement (voir figure 4.2). Les déformations locales sont opérées par un champ de déformation unique ($\mathcal{L}$) qui permet de déformer tous les organes simultanément et de manière cohérente.

Cette approche est résolue numériquement en utilisant la méthode du Lagrangien augmenté pour se ramener à un problème non contraint. La méthode du gradient est alors utilisée pour résoudre le problème.

Des exemples jouets sont tout d'abord donnés en 2D. Ils montrent l'intérêt d'introduire cette contrainte de non recouvrement.

Proposition d'un terme image robuste et invariant au changement de contraste

En section 4.4 nous définissons un terme de force invariant par rapport aux contrastes de l'image. Ce terme générique repose sur des informations d'intensité et de contours et représenté par l'équation 4.33. Cette équation est composée de deux termes. Le premier fait intervenir des probabilités d'intensité calculées à l'intérieur et à l'extérieur de l'objet. Ce terme dépend donc des contours de l'objet. Le deuxième terme s'inspire des travaux de (Kimmel and Bruckstein, 2003). Il fait intervenir un terme de signe correspondant au sens du contraste et un terme correspondant au Laplacien de l'image calculé le long des lignes

de niveaux de l'image. Nous montrons avec quelques exemples jouets que ce terme permet d'être utilisé sans se préoccuper du sens du contraste et qu'il est robuste au bruit.

Application à la segmentation d'organes en imagerie TDM

Enfin, en section 4.5 nous proposons d'appliquer notre nouvelle approche à la segmentation automatique d'organes abdominaux en TDM 3D. L'initialisation des modèles est réalisée en utilisant la méthode de localisation présentée au chapitre 3. La chaîne de segmentation automatique est illustrée à la figure 4.15.

Une base de données de 156 volumes TDM est utilisée pour évaluer l'algorithme pour la segmentation du foie, des deux reins, de la rate et de la vésicule biliaire. Les figures 4.18 et 4.19 montrent les distributions des coefficients de Dice et de distance moyenne à la vérité terrain pour chaque organe et après chaque étape de l'algorithme. Les expériences montrent un bon compromis entre précision et rapidité. En environ 30 secondes, 5 organes sont segmentés avec une distance moyenne de 3, 2mm (médiane : 2mm). Les figures 4.21 et 4.22 montrent des exemples de résultats.

Une partie du contenu de ce chapitre a été publiée dans (Gauriau et al., 2015a).

SEGMENTATION INTERACTIVE AVEC DÉFORMATION DE MODÈLES

Ce chapitre vise à enrichir la formulation de segmentation multi-organes présentée au chapitre précédent pour prendre en compte des interactions d'un utilisateur. Le but est en particulier de permettre à un utilisateur de corriger des résultats de segmentation existants, le résultat de segmentation automatique par exemple. Le type de corrections se veut le plus simple possible pour l'utilisateur. La formulation par déformation de modèle implicite permet d'obtenir des déformations cohérentes entre les différents organes. La section 5.1 présente une introduction à ce chapitre. En particulier la section 5.1.2 donne une étude bibliographique sur la segmentation interactive. Cette étude montre que la frontière entre segmentation semi-automatique et interactive est parfois ténue. Nous distinguons différents types d'approches :

- segmentation manuelle guidée (l'utilisateur trace les contours manuellement mais le logiciel peut lui proposer des contours plus pertinents),

- par optimisation globale à partir d'une initialisation venant de l'utilisateur (ex : *graph-cut*, *fast marching* etc.),

- par segmentation séquentielle avec entrée utilisateur (l'utilisateur propose une première initialisation, l'algorithme propose un résultat, puis l'utilisateur peut raffiner son initialisation et ainsi de suite),

- corrections localisées (les interactions de l'utilisateur ont un impact local uniquement).

Cet état de l'art permet en particulier de souligner le peu de méthodes existantes adaptées au multi-objets, en particulier en 3D. Par ailleurs les méthodes dédiées à la correction de segmentations existantes sont aussi quasiment inexistantes.

La section 5.1.3 présente la méthode de (Mory et al., 2012) qui introduit des interaction de l'utilisateur dans la formulation mono-organe de déformation de modèle implicite.

La section 5.2 montre comment, en s'inspirant des travaux précédents, il est possible d'ajouter ces interactions de l'utilisateur sous forme de points de contraintes dans la formulation présentée au chapitre 4. Le principe est illustré à la figure 5.2 : transformations globales et transformation locale sont optimisées simultanément pour prendre en compte l'entrée de l'utilisateur alors que la contrainte de non recouvrement est aussi respectée. Le schéma d'optimisation numérique est présenté. Il repose en particulier sur l'utilisation de la méthode du Lagrangien augmenté pour transformer la formulation sous contrainte en un problème non contraint. En 5.2.3 une nouvelle équation de l'énergie est aussi présentée pour permettre de rendre les contraintes utilisateur localisées. Des exemples sur des images médicales 3D sont ensuite donnés.

Enfin, la section 5.3 donne une application : la correction de quatre organes segmentés automatiquement dans des images abdominales TDM 3D. En particulier, une étude utilisateur est menée. Dix experts ont expérimenté notre outil et ont segmenté dix volumes où les segmentations automatiques étaient de précision variable. Toutes les interactions ont été enregistrées et les experts étaient amenés à remplir un questionnaire de satisfaction à la fin de l'expérience. Ils devaient notamment évaluer la facilité de prise en main, la qualité des corrections et le temps de correction. Cette étude a permis d'avoir des retours sur la prise en main de notre outil et sa facilité d'utilisation. Les résultats sont positifs et très encourageants. La figure 5.7 donne la distance moyenne à la vérité terrain et le coefficient de Dice en fonction du temps de correction. En moins de deux minutes de correction, la précision moyenne sur les quatre organes est inférieure à 2mm. La figure 5.10 montre les résultats du questionnaire d'évaluation. La figure 5.9 donne des exemples de résultats de correction en quelques cliques.

CONCLUSION

Dans cette thèse nous avons présenté une chaîne complète et cohérente pour la localisation, segmentation et correction de plusieurs organes dans des images médicales 3D. Une évaluation poussée sur une base de donnée conséquente a permis de montrer la robustesse, la précision et la rapidité de chaque composant

de notre méthode. Au vu de l'enthousiasme des retours des experts qui ont évalué notre outil, nous sommes en mesure de dire que celui-ci a un véritable potentiel pour une utilisation clinique.

D'un point de vue méthodologique, notre travail apporte des contributions aussi bien dans le domaine de la localisation que de la segmentation. Tout d'abord, nous avons proposé une approche par apprentissage pour la localisation de plusieurs structures anatomiques. Le résultat est donné sous la forme de cartes de probabilités pour chaque organe. Cette méthode étend le travail de (Criminisi et al., 2013) sur la localisation de structures anatomiques par régression et le travail de Cuingnet et al. (2012) sur son extension avec une approche en cascade. Notre méthode propose d'introduire l'utilisation d'information a priori de forme pour une localisation plus précise. Les cartes de confiance calculées peuvent avoir différentes applications. Dans ce travail elles sont utilisées pour initialiser et guider de futures étapes de segmentation.

Nous avons aussi proposé une nouvelle approche de segmentation multi-organes. Notre approche étend et enrichit les travaux de (Mory et al., 2012) sur la segmentation par déformation de modèle implicite de forme. L'originalité de notre solution est d'exploiter la décomposition de la transformation optimisée en une transformation de pose et des déformations locales. Chaque objet possède sa propre transformation globale afin d'être positionné correctement par rapport aux autres objets. Les déformations locales sont opérées par un champ de déformation non-rigide mais régulier, commun à tous les objets. Des contraintes de non recouvrement sont par ailleurs introduites afin d'assurer la cohérence des différents contours de segmentation. Nous avons aussi détaillé différentes optimisation algorithmiques permettant d'obtenir une exécution rapide. Nous avons aussi défini un terme de force générique reposant sur les intensités et les informations de gradient. Nous montrons que nous pouvons utiliser les cartes de confiance pour enrichir et contraindre le terme de force. En utilisant les mêmes paramètres pour chaque organe nous obtenons déjà des résultats robustes et satisfaisants dans des conditions assez variées. Les forces peuvent cependant facilement être modifiées et adaptées au type d'application voulu.

Enfin nous avons démontré que notre méthode de segmentation peut être étendue avec l'intégration d'entrées utilisateur pour la correction. Une nouvelle formulation de l'énergie a été proposée afin d'assurer un comportement intuitif (en particulier, l'aspect local des corrections). Les contraintes utilisateur sont incorporées de façon assez naturelle et cohérente dans notre formulation. En particulier, notre approche permet d'obtenir des déformations cohérentes entre les organes voisins : les organes se déforment ensemble en fonction des différentes interactions de l'utilisateur.

En conclusion, nous avons développé un outil puissant et générique répondant aux principaux besoins cliniques : robustesse, précision, rapidité et facilité

d'utilisation. L'application clinique présentée dans ce travail démontre le potentiel de notre approche :

- l'application à différentes modalités,

- l'utilisation de une à plusieurs structures,

- l'utilisation en mode automatique ou interactif.

Enfin ce travail montre l'intérêt d'introduire de la connaissance a priori en terme de robustesse et précision, que ce soit grâce à l'apprentissage ou a de l'information a priori de forme.

RÉSUMÉ DES CONTRIBUTIONS

Une chaîne complète pour la segmentation automatique et la correction de plusieurs organes :

- à la fois générique et modulaire,

- robuste, précis et efficace,

- facile d'utilisation.

Une nouvelle méthode de localisation multi-organes :

- robuste grâce à l'exploitation des relations spatiales, d'information contextuelle et de forme,

- donnant des résultats spécifiques à chaque organe sous la forme de cartes de confiance,

- fournissant un résultat pouvant être utilisé dans différentes applications.

Une nouvelle approche de segmentation multi-organes :

- robuste grâce à une formulation par déformation de modèle, à l'exploitation d'information de forme et de forces invariantes aux changements de contraste,

- des résultats cohérents assurés par des contraintes de non recouvrement,

- des corrections utilisateur locales et intuitives.

Perspectives

À la fin de chaque chapitre du manuscrit sont déjà mentionnées de possibles améliorations pour chaque méthode. Nous proposons ici des commentaires plus généraux.

AMÉLIORER LA LOCALISATION DE PETITES STRUCTURES Nos travaux ont montré la faiblesse de notre approche de localisation pour les petites structures comme la vésicule biliaire. Cela semble être lié à la taille de la structure étant donné que la distance moyenne à la vérité terrain reste généralement assez faible. En partant des régions obtenues à partir des cartes, des méthodes de classification s'appuyant sur des caractéristiques d'apparence pourraient certainement permettre de raffiner ces résultats.

AMÉLIORER LA COHÉRENCE DES CARTES EN TERME DE FORME Dans ce travail, l'information a priori de forme est incluse dans nos modèles à l'aide de modèles moyens de forme. Cependant les variations de forme pourraient certainement aussi être exploitées. Les travaux de Zhou and Comaniciu (2007), Cootes et al. (2012), Chen and Zheng (2014) sur la régression de forme pourraient être explorés afin d'obtenir des carte de confiance soulignant mieux la variabilité inter-patient. L'intégration de variations de formes dans la méthode de segmentation pourrait aussi être étudiée comme dans les travaux de Prevost et al. (2013b). La construction de modèles de variations joints est aussi une piste intéressante (par exemple un foie assez imposant a de grande chance d'influencer la position et la taille du rein droit voisin).

EXPLOITER ENCORE MIEUX LES CARTES DE CONFIANCE Les cartes de confiance contiennent une information riche qui peut sûrement être encore exploitée. L'incertitude qu'elles encodent pourraient être utilisée pour contrebalancer l'influence des forces. En termes d'application, elles pourraient aussi être utilisées pour améliorer l'expérience de l'utilisateur : focalisation automatique sur l'organe voulu, personnalisation de logiciel (proposition automatique d'outils adaptés en fonction de l'organe survolé par la souris), amélioration de rendu 3D etc.

EXPLOITER LES CARTES DE FORCE Les cartes de force contiennent aussi une information intéressante qui pourrait être utilisée pour mesurer l'incertitude de la segmentation finale. Cette information pourrait par exemple être utilisée pour indiquer à l'utilisateur les zones où la segmentation finale est incertaine et a de grandes chances de requérir des corrections.

Le calcul des forces pourrait aussi bénéficier de l'introduction de connaissance a priori sur l'apparence des organes, en particulier avec l'utilisation de

méthodes d'apprentissage. Cela permettrait probablement d'améliorer la robustesse de l'algorithme de segmentation. Cela permettrait aussi certainement de détecter des pathologies ou des anormalités telles que des tumeurs.

EXPLOITER LES CHAMPS DE DÉFORMATIONS   Dans notre méthode de segmentation, la transformation optimisée encode la façon dont les modèles se déforment et collent aux contours de l'objet. Cette association entre l'image et l'anatomie de référence est, en soi, une information intéressante. Elle pourrait par exemple être utilisée à des fins éducatives. Par exemple, un utilisateur pourrait cliquer dans l'image et voir la localisation correspondante sur la référence anatomique (qui pourrait être affichée sous forme d'un atlas).

L'information de déformation pourrait aussi être utilisée pour réaliser des analyses statistiques de forme. En effet, si un modèle est utilisé pour réaliser plusieurs segmentations, le résultat de segmentation peut être mis en correspondance avec le modèle de référence. Cela signifie qu'il est possible de projeter les différentes parties anatomiques sur un unique modèle et mettre ces parties en relation chez différents patients. Cette information pourrait probablement permettre d'identifier les parties de l'anatomie avec le plus de variations. De la même façon, les corrections interactives pourraient aussi être projetées sur le modèle de référence. Cela permettrait d'identifier les zones qui requièrent le plus de corrections en générale. Celles-ci pourraient donc être automatiquement indiquées à l'utilisateur en priorité.

VERS LA MULTI-MODALITÉ   De façon similaire aux travaux de Prevost et al. (2013a), notre méthode peut être facilement étendue aux problématiques multimodales, de co-recalage et de co-segmentation. L'idée est de segmenter simultanément les mêmes organes dans deux images (ou plus) de différentes modalités. Chaque organe aurait sa propre pose dans chaque image et se déformerait conjointement (grâce à un unique champ de déformation) par rapport aux différentes forces calculées dans les différentes images. Par exemple, dans le cas du recalage du foie en imagerie IRM et ultrasonore, une telle méthode permettrait de bénéficier de la connaissance de la position du rein droit (quand il est visible).

D'AUTRES APPLICATIONS   Nos travaux sur la localisation d'anévrismes dans des images ultrasonores 3D ont déjà montré que notre approche peut avoir de multiples applications. Par exemple notre méthode pourrait être utilisée pour la segmentation automatique de la prostate et de la vessie. Notre méthode pourrait aussi permettre d'obtenir une modèle virtuel de l'anatomie complète en 3D (par exemple pour enseigner l'anatomie ou pour la chirurgie).

PUBLICATIONS

This work has led to the following publications (in chronological order):

- R. Gauriau, B. Husson, K. Deva, I. Bloch, and C. Adamsbaum. Segmentation du corps calleux : un outil pronostique des pathologies inflammatoires cérébrales ? In *Journées Françaises de Radiologie*, Paris, France, 2013

- R. Gauriau, R. Cuingnet, R. Prevost, B. Mory, R. Ardon, D. Lesage, and I. Bloch. A generic, robust and fully-automatic workflow for 3D CT liver segmentation. In H. Yoshida, S. Warfield, and M. W. Vannier, editors, *Abdominal Imaging. Computation and Clinical Applications*, volume 8198 of *LNCS*, pages 241–250. Springer Berlin Heidelberg, 2013

- R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch. Multi-organ localization combining global-to-local regression and confidence maps. In P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8675 of *LNCS*, pages 337–344. Springer Berlin Heidelberg, 2014

- R. Gauriau, R. Ardon, D. Lesage, and I. Bloch. Multiple template deformation. Application to abdominal organ segmentation. In *International Symposium on Biomedical Imaging (ISBI)*, New York, USA, 2015a

- R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch. Multi-organ localization with cascaded global-to-local regression and shape prior. *Medical Image Analysis*, 23:70–83, 2015

- R. Gauriau, D. Lesage, M. Chiaradia, B. Morel, and I. Bloch. Interactive multi-organ segmentation based on multiple template deformation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS. Springer, 2015b

REMERCIEMENTS

Merci à toutes les personnes qui m'ont embauchée, encadrée, lue, corrigée, encouragée, soutenue, épaulée, réconfortée, stimulée, aiguillée, amusée, distraite et égayée durant cette thèse. Bien plus qu'une expérience de recherche j'ai aussi vécu une très belle aventure humaine. Et parce qu'une image vaut parfois mieux qu'un discours…

CONTENTS

ACRONYMS

1D one-dimensional

2D two-dimensional

3D three-dimensional

4D four-dimensional

AAA abdominal aortic aneurysm

ANN artifical neural networks

ASM active shape model

bagging bootstrap aggregating

CT computed tomography

GPU graphics processing unit

GVF gradient vector flow

ICP iterative closest point

KNN K-nearest neighbors

LDA linear discriminant analysis

MAP maximum a posteriori

MRI magnetic resonance imaging

MSF maximum spanning forests

MSL marginal space learning

PCA principal component analysis

PDF probability density function

PET positron emission tomography

PSNR peak signal-to-noise ratio

RF random forest

RKHS reproducing kernel Hilbert space

ROI   region of interest

RSS   residual sum of squares

SSM   statistical shape model

STAPLE simultaneous truth and performance level estimation

SVM support vector machine

TMDA toolkit for multivariate data analysis

US   ultrasound

CONTEXT AND PRESENTATION

1.1 CONTEXT

Medical imaging is nowadays a key element in medical practice. It provides clinicians with a non-invasive visual representation of the interior of the body which gives valuable information for disease diagnosis, treatment planning and intervention. A varied panel of imaging modalities is now available to meet different needs, each of them having its own characteristics, advantages and drawbacks. We distinguish two types of medical imaging technologies: functional and structural. Functional imaging gives information on physiological aspects (e. g. tissue activity with positron emission tomography (PET)), while structural imaging gives a representation of the tissues and the anatomy (e. g. ultrasound (US), X-ray radiography, computed tomography (CT), magnetic resonance imaging (MRI)). In this work we focus on this second type of imaging.

Understanding and interpreting medical imaging requires a high level of expertise. Indeed, the acquisitions contain complex and varied information being more or less difficult to retrieve considering the following variabilities:

- the intrisic variability of the imaging modalities (contrasts, noise, resolution, two-dimensional (2D)/three-dimensional (3D), bias, fields of view etc.);

- the inter-patient variability (e. g. organ size, shape and appearance, as shown in Figure 1.1);

- the intra-patient variability (e. g. soft organs can move with gravity or breath, or change with time).

For several decades, these challenges have led to the development of computerized methods for better and faster understanding and interpretation of medical images. Medical image analysis and processing is now a specific research field with its own community. Two very active topics are localization and segmentation (sometimes called delineation by clinicians). Localization is the task of locating a target anatomical structure in the image. Segmentation is the task of partitioning an image in different segments, for instance to get the contours of an organ or to extract smaller internal structures such as tumors. This valuable information has various clinical applications such as easiest visualization and understanding, or quantification (e. g. volume). The recent progresses made in this field have opened the way to new possibilities in the
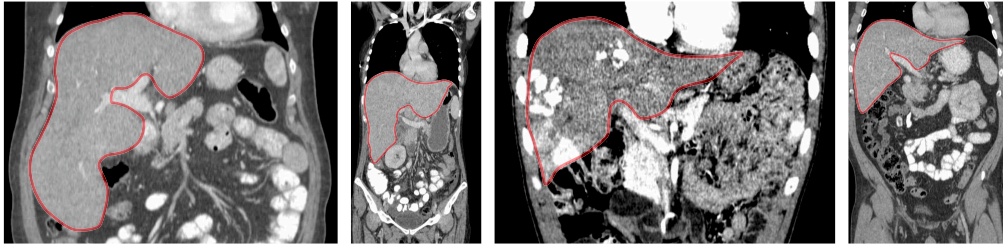
Figure 1.1: Examples of CT images (in coronal plane) for four different patients, showing the anatomical variability of the liver (in red) and variability of fields of view.

medical practice. The development of multi-organ-dedicated localization and segmentation methods make now possible the automatic generation of patient-specific anatomical models. While offering an enhanced understanding of the anatomy in its globality (e. g. organs relationships), it also contributes to the progress and development of medical procedures (e. g. laparoscopic surgery). Despite the great advances made in image processing and analysis, the automation of multi-organ localization and segmentation is still an active field of research trying to overcome the challenging variabilities of medical images mentioned previously.

The main requirements of segmentation tools for the clinical practice are:

- robustness: the algorithm should be usable in various conditions (related to modality and patient);

- accuracy: the algorithm should give precise results in practice;

- speed: the algorithm should process the result shortly;

- ease of use: the algorithm should require as little user intervention as possible.

Our work is an attempt at meeting these clinical needs.

We developed a generic and modular workflow for the localization and the segmentation of multiple organs. Our general methodological framework ensures the robustness and coherence of the solution, while various sub-components (organ models, image features) can be adapted to the application and modality at hand to reach high accuracy results. Finally, particular care was given to keeping our approach as simple as possible while pursuing efficient algorithmic implementations. This results in fast solutions, compatible with current clinical routine.

1.2 OUTLINE

This manuscript is structured around the design of a complete multi-organ segmentation workflow, from localization to user corrections, composed of two main components:

- a localization approach based on a cascade of random forest (RF) regression with shape prior;

- a multipurpose segmentation method based on multiple implicit template deformation that can be used:
  - in an automatic mode;
  - to correct segmentation results in an interactive way.

The proposed approaches can be used independently or chained together to provide a robust automatic segmentation pipeline.

The localization results, provided in the form of organ-dedicated probability maps are accurate and informative enough to be used as is (e.g. for enhanced navigation/visualization experience). The segmentation framework can also be tailored to suit different applications. It can be used in a stand-alone version to get automatic results, or it can integrate user interactions for multiple contours interactive correction. Used together, our components provide a complete toolset for automatic multi-organ segmentation integrating user corrections. The complete pipeline is illustrated in Figure 1.2.

Chapter 2 depicts the context of our contribution. We propose an overview of machine learning techniques and of the different localization and segmentation methods dedicated to multiple structures.

Chapter 3 presents our localization method. This approach gives organ-specific results in 3D in the form of probability maps that we call *confidence maps*. An extended evaluation shows its robustness, accuracy and speed.

Chapter 4 introduces a new variational formulation for the segmentation of multiple organs in 2D or 3D. First experiments show the interest of the formulation with non-overlapping constraints. The definition of a contrast-invariant image-driven force term allows us to reinforce the generic aspect of the approach. Then, we show how to link this work with the previous localization method to get an automatic segmentation pipeline. The segmentation of 5 organs in 3D CT volumes gives good accuracy results within a few seconds only.

Chapter 5 shows how to introduce user feedback in the multiple template deformation. By introducing a new energy formulation allowing local corrections, this provides us with an intuitive and user-friendly tool. A user study performed with medical imaging experts highlights the interest of the approach.

Chapter 6 gives our conclusions and an overview of potential future works.

Lastly, some appendices can be found at the end of the document. They are mainly related to collaborations realized during this thesis. Although they are

not essential for the understanding of this work, they highlight the interest and generic aspect of the approaches presented. Appendix A presents our preliminary work on liver segmentation in 3D CT. This work, realized at the beginning of this thesis opened the way to the extension of the original works we built our thesis on. It can be seen as a mono-organ and simpler version of the work developed in this thesis. Appendix B presents results of liver segmentation in 3D MRI. The method exploits our work on localization. It shows two key interests of this contribution. It also demonstrates that our approach can be used in other modalities than CT. Appendix C shows another application of our localization approach in 3D US images. It shows that our localization approach can be used on other types of anatomical structures than abdominal organs and different modalities. Appendix D gives more details for the computation of the equations used in the segmentation framework and Appendix E presents a method for the construction of mean shape models of organs (with implicit representation).
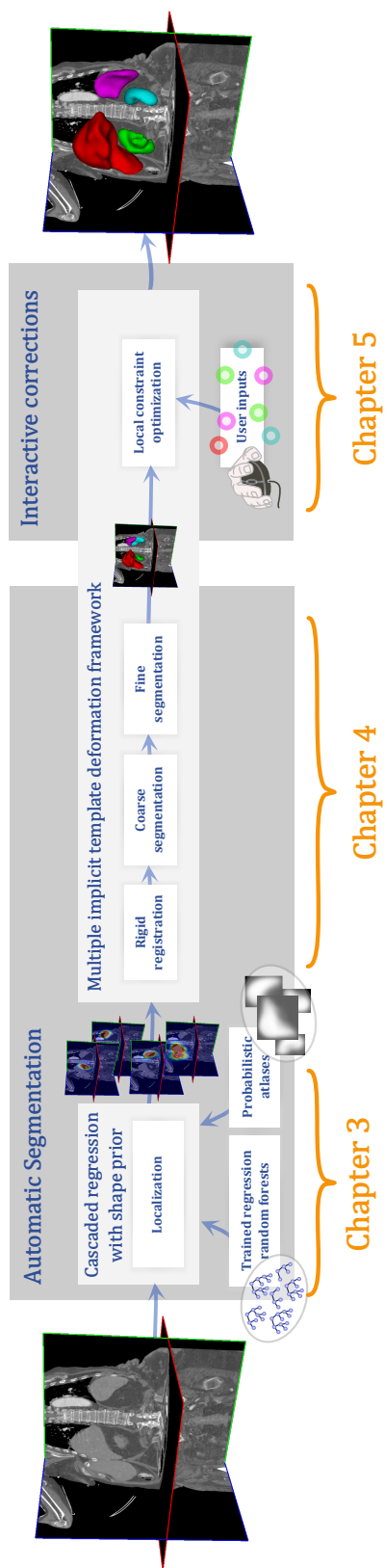
Figure 1.2: Complete pipeline of our automatic segmentation method with corrections.

LITERATURE REVIEW

Abstract

This chapter reviews the main literature keys to understand our work. It starts with a brief overview of the machine learning domain, with a particular focus on supervised methods, decision forests and their variants. Then we review the different methods of localization and segmentation specific to multiple anatomical structures.

Contents

2.1 INTRODUCTION

Localization aims at finding the position of a given structure while segmentation aims at determining the part of the image belonging to the said structure. In the computer vision community the terms localization and detection are often taken as synonyms while they convey different meanings. Object detection is the act of telling whether an object is present or not in the image, while the localization tells its position in the image. Localization and segmentation are intrinsically linked, as the last one often depends on the first one. Segmentation methods may for instance rely on manual localization (e.g. a click from the user). However the automation of these two tasks often leads to different formulations and algorithms.

Localization and segmentation are both vast domains still being actively researched in the computer vision community. An exhaustive review of these fields is out of the scope of this work. We rather focus on the medical imaging sub-domain, which, considering its prolific activity, already gives a good overview of the larger computer vision domain.

Withey and Koles (2008) give an extended review of medical image segmentation methods and list different publicly available pieces of software. Numerous works have been dedicated to single organ segmentation, but their generalization to multiple organs is often not straightforward. Of course any method designed for single organ localization or segmentation may be applied sequentially to several organs (as in the work of Campadelli et al. (2009a)). However the anatomical structures have specific relationships that may be interesting to take into account. For instance abdominal organs such as the kidneys and the liver do not overlap, although they may sometimes be adjacent. Thus, there is some prior knowledge on anatomical structures relationships that could be advantageously expressed to refine the localization and segmentation models. It is more and more frequent to have access to large amount of data, taking benefit of it is a natural inclination. Machine learning methods offer the possibility to model and generalize from sample databases. After a brief overview of the machine learning domain, we focus this bibliographic section on localization and segmentation approaches that are specific to the multi-structure case.

2.2 A SHORT INTRODUCTION TO MACHINE LEARNING

As very well described in the thesis introduction of Pauly (2012), for a human being, the typical process of learning follows three steps: (1) the collection of an amount of observations, (2) the extraction of relevant information, (3) the design of a general model that can explain the most accurately past and future observations. Few decades ago, a new field of artificial intelligence emerged to mimic this learning process from observations: machine learning. The goal of

machine learning is to design algorithms that can learn from past experience to predict new observations. The main interest of this field is to automatically build predictive models from data (usually a huge amount) that would be too complex to design manually. The vocabulary of machine learning is specific and needs to be defined clearly. The term *training data* generally denotes the set of observations from which the model is built. The extraction of the relevant information from the training data is called *features extraction*. The *features* are characteristics that generalize and best describe the original observations (often in a simpler way). The *testing data* are the new observations used to test the model.

Machine learning techniques are classified into three categories: supervised, unsupervised and semi-supervised. The goal of supervised methods is to infer the model (or predictive function) from annotated training data (in general, the collected observations have been annotated by a human). The annotations are usually called *labels* and correspond actually to the output of the predictive function. Unsupervised learning aims at finding hidden structures from training data that have not been labeled. One major class of unsupervised learning methods is clustering whose goal is to discover similar groups in the feature space. Finally, semi-supervised learning corresponds to the case were only a subset of the training data has been labeled. We refer the reader to the excellent book of Hastie (2009) on statistical learning for a complete overview and understanding of learning techniques. Note that in this work we focus on statistical machine learning, but the machine learning domain is larger (e. g. logic-based learning).

As we want to learn from annotations given by an expert, we propose to limit our overview to supervised learning techniques. There are two types of supervised learning approaches: *classification* and *regression*. If the output variable of the predictive function is a discrete variable (we call it a *class*) the problem is called classification. If the output variable is continuous, the problem is called *regression*.

In the following sections, we give a short overview of existing supervised methods and focus on the main challenges they raise. Our goal is not to make an exhaustive review of the field but rather put our work into context by giving the keys to understand our choices.

2.2.1 *Brief overview of supervised learning methods*

Hereafter we list the main approaches of supervised learning and their main principles. We also give appropriate references that would help the reader to find more information on these topics. In Figure 2.1 we show different classification results obtained with some of the main machine learning techniques.

This figure illustrates the different types of decision boundaries they can generate.

2.2.1.1 *Linear methods*

Some of the oldest methods used for supervised learning are the *linear methods* coming from the field of statistics. Although they are mainly used for regression, they can also be used for classification (like the linear discriminant analysis (LDA) method (Hastie, 2009)). The principle is to assume that the output variable can be approximated by a linear combinations of the input variables. The methods differ mainly in the way to estimate the weight of the input variables and the type of optimization procedure to minimize a specific criterion. We refer the reader to the book of Hastie (2009) for more details on the existing approaches.

2.2.1.2 *Support Vector Machines*

With the *support vector machine (SVM)*, introduced by Cortes and Vapnik (1995), the problem of two-classes classification relates to finding the best linear separator between the training observations. This technique is based on the assumption that the best separator is the one that has the largest distance (called *margin*) to the nearest training samples of the two classes (those samples are called *support vectors*). The other idea behind SVM is that non-linear problems can be projected to another high-dimensional space (using a *kernel* function) where the input data can be linearly separated by a hyperplane. One major advantage of this approach is thus to allow for the modeling of non-linear problems. In high dimensions the choice of kernel may however often be problematic and chosen empirically in practice. The SVM framework is originally designed for binary classification problems, but multi-class problems may be reduced to several binary problems (at the cost of building several separated classifiers). Although SVMs have also been extended to regression problems, they are not really adapted to this class of problem, as in the learning process the training samples lying outside of the margin are not taken into account.

2.2.1.3 *Artificial Neural Networks*

The nervous systems of animals inspired the artifical neural networks (ANN) methods. We are still far from mimicking biological nervous networks, but this type of learning approaches gives state-of-the art results in some domains (Caruana and Niculescu-Mizil, 2006). An ANN is presented as a system made up of *neurons* linked with weighted connections called *synapses*. The neurons return different values given the inputs. The learning process aims at finding the weights of the synapses that minimize a given loss function. ANN methods recently raised a new interest when combined with principle of *deep learning*.
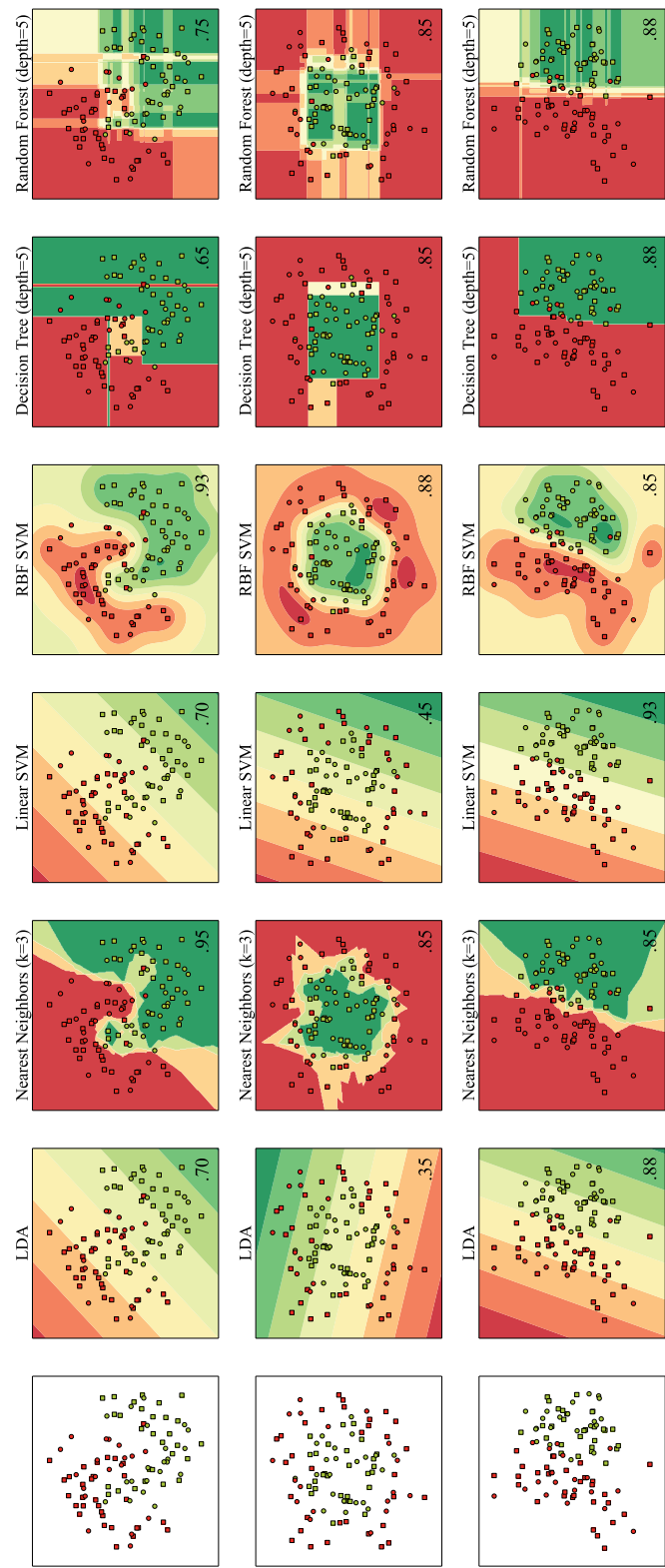
Figure 2.1: Results of binary classification on three different data sets with six different machine learning methods. Squares are for training points, circles for testing points and color gradients represent the prediction confidence. The lower right figure on each image corresponds to the classification accuracy on the testing data. (Illustration generated with the Python source code credited to: <Gael Varoquaux, Andreas Muller, Jaques Grobler> (http://scikit-learn.org/stable/auto_examples/plot_classifier_comparison.html))

The key idea is to chain multiple, sometimes numerous layers of ANN. Each layer aims at higher-level, structured information from the data. Deep learning techniques, combined with convolutional networks, have been particularly popular in visual recognition tasks in the recent years (Russakovsky et al., 2014). However they can be sometimes difficult to explain and understand. The choice of number of units, layers, types of cost functions is still a difficult task and requires a high degree of expertise. We refer the reader to the work of Hinton (2007) and Arel et al. (2010) for the recent advances of this field.

2.2.1.4 *K-Nearest Neighbors*

The *K-nearest neighbors (KNN)* algorithm is a non-parametric method used both for classification and regression. The idea is to find the *k* training samples that have the minimum distance (commonly Euclidean but it can be any other metric) to a new testing sample to infer the value of the output. This kind of approach has the advantage of being simple to implement and understand. It is particularly suited to non-linear problems with many classes (Wu et al., 2008). However it requires generally to have the whole training set available at testing time which may not be feasible for many applications in practice.

2.2.1.5 *Decision trees*

Decision trees are non-parametric methods used for both classification and regression. They are hierarchical learners consisting of an ensemble of simple decisions (binary most of the time). The paternity of *decision trees* is given to Morgan and Sonquist (1963) who first used regression trees for process explanation and prediction. From this time, the principal improvements and methodological developments are due to the work of Breiman et al. (1984) and Quinlan (1993), making them easier to use and understand.

A decision tree is a tree (assumed binary hereafter) composed of *nodes* linked with *branches* and terminated with *leaves*. Each node corresponds to a test on the input data whose answer will guide to a next node or to a leaf containing the possible result.

A decision tree follows a *divide* and *conquer* strategy. First it creates a partition of the input feature space following some decisions. Secondly it estimates the probability of an output given an input for each partition.

We denote a training set by $\{x_n, y_n\}_{n=1}^{N} \in \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} \subset \mathbb{R}^d$ is the feature space and $\mathcal{Y} \subset \mathbb{R}^{d'}$ is the output space. We want to learn a model which gives a prediction y in $\mathcal{Y}$ given an observation x in $\mathcal{X}$. Following a probabilistic framework, we want to learn the posterior $P(y|x)$. We can formulate the problem as a maximum a posteriori problem:

$$\hat{y} = \text{argmax}_{y \in \mathcal{Y}} P(y|x) \tag{2.1}$$

(a) Split node (train)

(b) Split node (test)

(c) Leaf (train)

(d) Leaf (test)
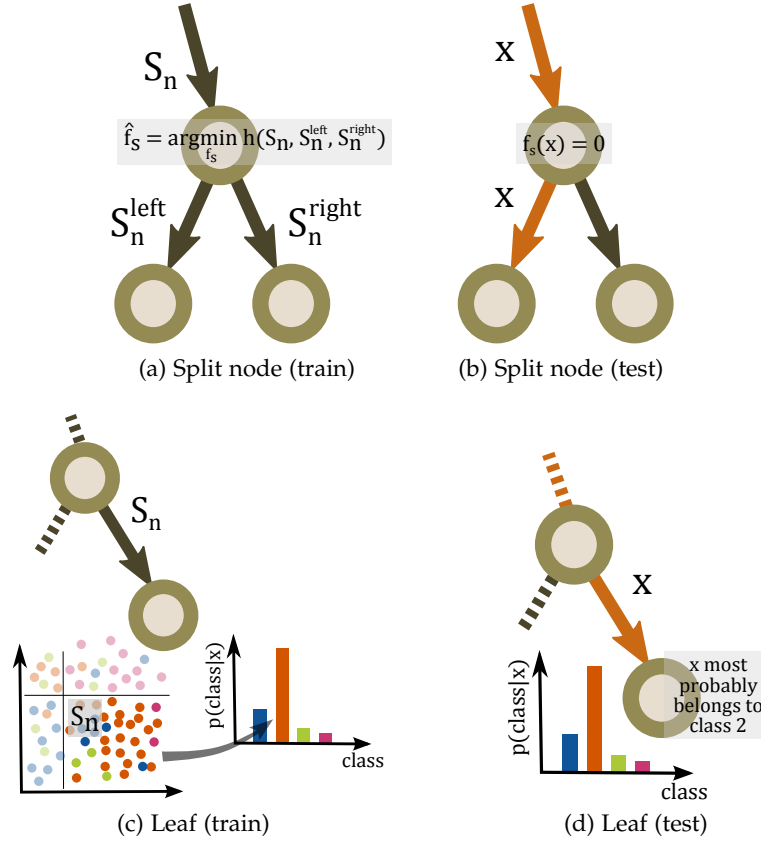
Figure 2.2: Illustration of the different decision tree components: node (b) (a) and leaf (c) (d) (figure inspired by Criminisi and Shotton (2013)).

A decision tree is built recursively from the root node to the leaves, each node performing a partition of the training data. The training samples reaching a leaf are then used to model the posterior distribution locally. During the test phase, this information is used to perform the prediction.

At a node, a binary decision is performed to know if the samples go to the left or to the right nodes. This decision is performed by using a *splitting* function $f_s : \mathcal{X} \rightarrow \{0, 1\}$ defined as:

$$
\begin{cases}
\text{if } f_s(x) = 0, & \text{the observation } x \text{ goes to the left node} \\
\text{if } f_s(x) = 1, & \text{the observation } x \text{ goes to the right node}
\end{cases}
\tag{2.2}
$$

The splitting function often consists of a simple linear projection coupled with a threshold operation. In the case of one-dimensional (1D) features (which will be the case in this work) the splitting function is a simple threshold comparison (also referred to *decision stump* in the boosting literature). Depending on the type of splitting function chosen, there are different parameters to learn during

training. We denote by S_n the training samples at the node n, and by S_n^{left} and S_n^{right} two possible subsets (S_n^{left} and S_n^{right} being a partition of S_n). The parameters of the splitting function are determined by minimizing a given objective function $h(S_n, S_n^{left}, S_n^{right})$ (see Figure 2.2a).

Finally, the node construction can end for different reasons: (i) the objective function gives a result below a chosen threshold, (ii) the tree reaches the maximal tree depth defined by the user, (iii) the number of samples in a node is below a fixed threshold. The current node then becomes a leaf and the posterior distribution of the samples reaching the leaf is stored (see Figure 2.2c).

In the testing phase, a new sample x follows the splitting rule f_s of each node, going to left or right child node (see Figure 2.2b). When it reaches a leaf we get the probability $p(y|x)$ given in the leaf (see Figure 2.2d). In the case of classification, we know the probability of each class y given the observation x. The final prediction can be obtained with the maximum a posteriori ($\hat{y} = $ argmax $p(y|x)$).

Although it is easy to build a tree in practice, it may become computationally expensive. In addition, decision trees often suffer from over-fitting on the training set.

2.2.1.6 *Ensemble methods: the example of Random Forests*

The goal of ensemble methods is to combine several base estimators built with a given algorithm. They are of two types: (i) averaging methods whose principle is to build several estimators independently and then average their result, (ii) boosting methods whose goal is to build several estimators incrementally so as to reduce the bias of the combined estimators on the misclassified training instances (the most famous is *Adaboost* (Freund and Schapire, 1995)).

We will focus on averaging methods by presenting the RF and their derived approaches as it will be useful to understand some theoretical aspects of our contributions.

RANDOM FORESTS   To prevent the problem of over-fitting inherent to decision trees, Amit and Geman (1994) and Ho (1995, 1998) proposed almost at the same time (see the foreword of Amit and Geman in (Criminisi and Shotton, 2013) for the detailed story) to construct an ensemble of weak decision trees, called random forest (RF), instead of constructing one complex decision tree. It follows the principle of ensemble methods: the use of multiple learning algorithms can give better results than a single one. The idea behind RF is to build several trees and add randomness in the node decision tree construction. This decorrelates the trees. Instead of choosing the best split among all possible features, the best split is determined among a subset of randomly chosen features at each node. The authors show that RF can achieve better predictive accuracy and prevents over-fitting. However, learning several *weak* decision trees has the

effect of increasing the bias of each tree. Therefore Breiman (2001) proposes to add even more randomness by learning each tree on a random subset of the training data. This principle is called bootstrap aggregating (bagging). He demonstrates that this gives a very good compromise between bias and variance and makes the algorithm more robust against over-fitting.

Since then, several ways of randomizing even more the tree construction were proposed. For instance, Geurts et al. (2006) proposes to randomize the splitting thresholds selection. This method is called *extremely randomized trees*. When the number of splitting thresholds reduces to 1, the method is called *totally randomized trees*.

We refer the reader to the thesis of Louppe (2014) for a more in-depth discussions on the RF algorithm and its different variants.

WHY USE RANDOM FORESTS?   As explained in Louppe (2014) there are multiple elements making RF attractive in practice:

- they are non-parametric, thus being able to model complex phenomena;

- they perform feature selection intrinsically;

- they scale well to large training sets;

- they are robust to outliers;

- they are relatively simple to understand as they can be visualized;

- several efficient implementations are publicly available.

Several packages and libraries on RF can be found on the web such as the scikit-learn package in Python (Pedregosa et al., 2011) or toolkit for multivariate data analysis (TMDA) in C++ (Hoecker et al., 2007). Implementations on graphics processing unit (GPU) have also been proposed (Sharp, 2008; Grahn et al., 2011). Moreover, the parallelization of tree training and testing is straightforward as each tree can be learned and tested independently.

VARIANTS OF RANDOM FORESTS FOR OBJECT LOCALIZATION   The multiple advantages of RF mentioned below explain the recent popularity of this method in the computer vision community. One competitor for the speed aspect may be the *random ferns* (Ozuysal et al., 2007), a non-hierarchical variant. However it requires very large training sets to approach (sometimes exceed) the performance of classical RF methods. Numerous variants and improvements of RF techniques (often with extra complexity) have been proposed to improve the accuracy and/or to integrate more contextual information. We give hereinafter a brief overview of the methods that have been proposed so far, with a particular focus on object or organ localization. Note that these approaches can often

be applied with other types of classifiers and regressors than RF. Separate decision forests may be trained iteratively to refine successively the predictions as in the work of Shotton et al. (2008). Dollár et al. (2010) leverage this idea of *cascaded* predictions by learning multiple regressors. Each regressor is learned so as to minimize the difference between the true pose of the object to localize and the pose computed by the previous regressors. Tu (2008) proposes to integrate the semantic information from the first classifier into the next classifier (probabilistic boosting tree classifiers (Tu, 2005)). This approach, called *auto-context*, has been successfully applied to brain segmentation (Tu and Bai, 2010). Semantic context may also be introduced directly in the forest construction. The *entangled forests* of Montillo et al. (2011) are built so that tree nodes may depend on sub-results of the same tree. Authors show that it improves accuracy and better captures long-range context. In the *context-sensitive forests* proposed by Kontschieder et al. (2012) the forest entanglement is enhanced by sharing both regression and classification information between nodes. Kontschieder et al. (2013) also generalize the idea of entanglement and auto-context with the *geodesic forest* by using long-range soft connectivity features and a new objective function inspired by conditional random fields. Such entangled methods come with a cost: an increasing complexity and a longer learning process (8 hours to train the entangled forest on 200 CT volumes for the segmentation of 12 organs (Montillo et al., 2011)), which can make parameter optimization delicate. Joint pixel classification and shape regression have also been successfully applied to multi-organ segmentation (Glocker et al., 2012). Computation speed is however not mentioned.

When the objects to localize have a particular structure (which is the case in medical imaging), the RF may be conditioned with global variables. Dantone et al. (2012) propose to tackle the problem of face detection/localization by using the head pose as a prior to learn several RF regressors. Sun et al. (2012) generalize this idea and propose a RF regression model incorporating dependency relationships between output variables through a global latent variable. This has the advantage of improving the accuracy of a classical RF regression at a small computational overhead. The idea of bagging was also extended to *guided bagging* by Lombaert et al. (2014). The idea is to organize the global training set in an *image space* following an image similarity criterion. This allows to build specialized trees. The composition of the forest is then performed at test time. For a new test image, the training set consists of its K nearest neighbors in the image space. Methods based on a similar idea (Konukoglu et al., 2013; Zikic et al., 2013) and more related to atlas-based frameworks are also mentioned in Section 2.4.2.

2.2.2 *Learn more on machine learning*

For a better understanding of the different machine learning techniques we refer the reader to the book of Hastie (2009). The work of Kotsiantis et al. (2007) gives a compact overview of classification techniques, with multiple discussions and comparisons. The work of Fernández-Delgado et al. (2014) evaluates 17 families of classifiers on 121 datasets on 4 different implementations, which gives invaluable insight on the behavior of state-of-the-art methods in practice. Finally the website of Brownlee (2014) provides a huge amount of resources on the topic (videos, courses, implementations, links to books etc.).

Machine learning is being used in more and more domains, in computer science and many other research fields. There exist now several implementations (e.g. Pedregosa et al. (2011)) that can be easily used to test machine learning methods and quickly get reasonable results without extensive prior knowledge of the field. Without being an expert of machine learning it is possible to use these techniques successfully on condition to get much of the *dos and don'ts*. Domingos (2012) sums up brilliantly some key things to know about machine learning.

2.3 MULTI-ORGAN LOCALIZATION IN MEDICAL IMAGES

In the following sections we propose an overview of different localization techniques found in the literature, with a particular focus on multi-organ applications. We classify them into four groups: image-based, atlas-based, graph-based and learning-based. Typical examples are given for each class of method, without claiming for completeness.

2.3.1 *Image-based approaches*

We define as image-based the approaches that take only into account the information given in the image such as grey-level intensities or edges. In some applications, such operations combined with heuristic assumptions may be sufficient. Soler et al. (2001) use a thresholding technique and some morphological operations to initialize the segmentation of the liver. Jang et al. (2012) prefer an adaptive thresholding of the lungs helping to localize the liver thanks to heuristical assumptions. These methods can be viewed as very application- and modality-specific, since a change in intensity range or field of view may lead to a completely wrong localization. The generalized Hough transform (Ballard, 1981) exploits the edge information to detect an arbitrary shape. As shown by Zhang et al. (2010b), it requires a parameterization of the shape and an edge map from the image to be able to extract a consistent shape in the image. This
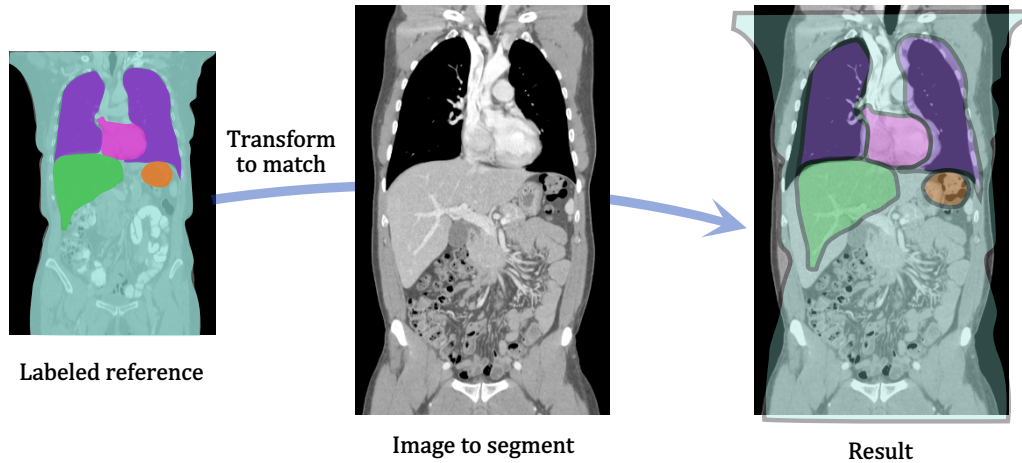
Figure 2.3: Illustration of atlas-based method principle.

method induces a high computational cost and is very shape-dependent. Moreover it is not adapted to organs with high variability.

2.3.2 *Atlas-based approaches*

An atlas is a labeled image used as a reference that can be transformed to match a new image to identify corresponding structures (see Figure 2.3). One interest of atlas-based methods is its ability to provide contextual information by embedding spatial relationships. Atlas-based methods were first dedicated to segmentation problems. We develop this topic in Section 2.4.2. As explained further, atlas-based approaches are computationally expensive. For localization applications, the computational burden can often be lowered as accuracy requirements are generally relaxed compared to segmentation. Atlas registration can for instance be performed at low resolutions. In the work of Zhou and Bai (2005) the atlas approach is used for localization and subsequent segmentation is refined with fuzzy connectedness and level set methods. In the work of Vik et al. (2012) for example, the authors build probabilistic tissue atlases (for air, fat, water and bones) and model the probability densities of landmarks (e. g. the organs centroids) for each of them. A simple global registration of these tissue atlases gives then the spatial positioning probability of the organ landmarks and helps localizing them. Despite low computation times (few seconds), this method remains rather dedicated to large organs or structures, and the accuracy may not be sufficient for small organs localization (e. g. the gallbladder). Moreover, strong pathologies may also induce too large deformations to be captured by such methods.
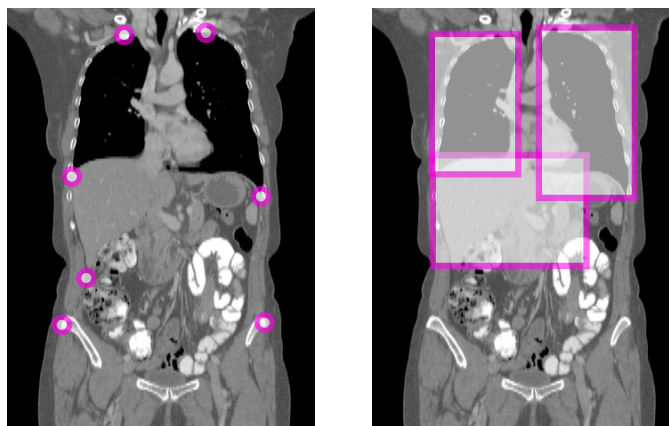
Figure 2.4: Landmarks (right) versus ROI (left) localization.

2.3.3 *Graph-based approaches*

A relatively frequent approach in multi-organ works is the use of graph representations, where organs are modeled as nodes and organ relationships as edges. A common idea to such methods is to capture, through the graph models, the strong prior information on these relationships. Liu et al. (2011) use graph optimization (through dynamic programming) relying mainly on intensity models for each organ. The authors do not give a quantitative evaluation or information about the computation time. Fouquier et al. (2012) propose to localize and segment brain structures in a sequential order, following prior knowledge on the spatial relationships between the different structures. The sequential nature of this process tends to propagate errors and may require a backtracking procedure. To prevent such a problem, Nempont et al. (2013) propose to solve the localization and segmentation problem in a global fashion following a constraint network. Although this approach has the benefit of relying on strong prior knowledge, it is computationally expensive in practice (several hours) and requires very fine parameter tuning which makes it difficult to apply in practice.

2.3.4 *Learning-based approaches*

We distinguish two types of learning-based methods for localization. The first ones are based on landmarks detection, while the second ones aim at finding regions of interest (ROIs) (see Figure 2.4). For both types of methods, regression or classification methods may be used. Classification-based methods generally leverage the local context with a relatively slow speed and are generally computationally expensive, especially for large 3D images. Indeed, in this case every voxel of the image needs to be assigned a label. On the other hand, regression-

based methods can make easier use of the global context and achieve fast speed albeit with lower accuracy. In this case, the full parsing of the image is generally not required. In the following sections we give an overview of the different machine learning methods for multi-organ localization. We refer the reader to Section 2.2.1.6 where we also give an overview of the different RF algorithm variants embedding contextual and spatial information and to (Zhou, 2014), who gives an exhaustive review of learning methods for anatomy localization, with an analysis of challenges and perspectives.

2.3.4.1 *With anatomical landmarks*

The localization of anatomical structures may be done by detecting specific landmarks in the image. The organ models (often in the form of mesh representation) can then be positioned according to these landmarks (Chen and Zheng, 2014) and following prior knowledge. Note that, to our knowledge, the initialization procedure is rarely described (what happens if the landmarks are wrongly detected?). The definition of the landmarks, their number and reproducibility, is also rarely discussed.

In the work of Hugueny and Rousson (2007) and Seifert et al. (2009) the idea is to detect anatomical landmarks with a learned classifier (with a probabilistic boosting tree (Tu, 2005)). An anatomical network is built and used to constrain the organs localization through marginal space learning (MSL) (Zheng et al., 2009b,a) (see Section 2.3.4.2). The organ localization is not evaluated in itself but the landmark detection average accuracy is satisfactory at around 7mm (on 80 volumes), and the computation time is around one second per organ.

In the work of Chen and Zheng (2014) on the other hand, the landmarks are detected independently in the entire image using a regression forest approach. The authors apply their approach successfully to pelvis and femur segmentation on about 200 2D X-ray images (accuracy within a few millimeters).

The landmarks detection may be performed following different strategies and one may notice that some landmarks are easier to detect than others. Their spatial relationships may be interesting to take into account to reinforce the robustness and accuracy of the detection. Such strategies may also advantageously reduce the detection speed. Zhan et al. (2008) propose to schedule the detection using joint landmark spatial relationships (with a multivariate Gaussian distribution model). They show significant improvement.

The idea of Liu et al. (2010) is to reduce the landmarks search range, by finding the best sequential strategy to detect the landmarks. This approach can be used with any classifier and seems to give efficient and accurate results. It is evaluated on about 400 volumes (required around 1400 volumes for training) with a computation time around 300ms per landmark (with GPU implementation). In an extended version, Liu and Zhou (2012) improve this framework by

adding a nearest neighbor matching preprocessing step to estimate roughly the landmark location (about 90 times faster in this version).

Lay et al. (2013) propose to take advantage of both classification and regression-based approaches by combining them to get both good accuracy and low computation cost. They apply their approach to MRI coarse segmentation of 6 thoracic organs with accurate performance and very high speed (a few seconds for the whole pipeline).

Landmark-based approaches seem rather efficient in both computation time and accuracy. However we should mention that they were generally applied on the localization of large structures (e. g. lungs, liver, pelvis) that can tolerate some inaccuracy in the landmark detection. This may not be the case for smaller structures as the gallbladder for example. For such structures, reproducible anatomical landmarks may also be difficult to define.

2.3.4.2 *Estimating several regions of interest*

Instead of detecting landmarks, some researchers proposed to directly localize the ROI of the image containing the organs. These ROI are often represented in the form of bounding boxes.

To find the correct organ position, one possible approach is to parse the entire image and test each organ location according to some features. The best candidate locations are determined using a learned classifier. An exhaustive search on all possible locations becomes however very expensive with the increase of transformation parameters (translation, rotation, scale etc.). Zheng et al. (2009b,a) thus propose to use a cascade of classifiers to refine the transformation parameters estimation in a coarse-to-fine way. A first exhaustive search of the position is then followed by a position-orientation estimation on the best former candidates, which is refined by iteratively re-estimating position and orientation, then position, orientation and scale on reduced sets of candidates. The search space is thus considerably reduced. This method can be used sequentially or independently for multi-organ localization (Seifert et al., 2010; Lu et al., 2012; Wu et al., 2014). Sofka et al. (2010), for instance, perform the localization sequentially in an order determined with a sequential Monte Carlo-based hierarchical detection network. The authors give some results on foetus face detection in 3D echographic US where the accuracy is below 2cm. However when considering multiple structures, this type of approach may lack consistency as it does not incorporate spatial relationships between structures.

Kobatake and Nawano (2006) follow the idea of *eigenfaces* (Turk and Pentland, 1991) coming from face recognition. They build *eigen-organ* spaces with PCA and project each possible volume of interest to this space to find the potential locations. Another global eigen-space is built were all the candidate organ locations are projected and a probabilistic atlas helps to find the best possible locations. This step ensures improved spatial coherence. The projections on every

possible volume of interest of the image are however computationally expensive (about 40 minutes).

Regression approaches have been proposed as an alternative to expensive exhaustive search. Dollár et al. (2010) first propose to regress object position. The idea is to learn directly the estimator giving the answer to the question: *what is the position of the object in the image?* instead of answering to the question: *is the object at this location in the image?*. Their approach, called *cascaded pose regression*, consists in learning sequentially multiple regressors, each regressor depending on the output of the previous regressors. Criminisi et al. (2011, 2013) also propose to use regression for multiple organs localization. The idea is to predict the bounding boxes parameters of several organs according to long-range and contextual features. This type of technique is also successfully applied to localization in MRI (Pauly et al., 2011). It reaches a good trade-off between accuracy and speed (in seconds) and is relatively easy to implement (there exist many random forest open source implementations, see Section 2.2.1.6). Following the work of Dollár et al. (2010) on cascades of regression, Cuingnet et al. (2012) propose to improve the method of Criminisi et al. (2011) with a cascade of regression random forests for kidneys localization. The idea is to first localize the two kidneys and then to refine their position independently. Despite the efficiency and robustness of these methods, one may criticize the rough parameterization of the organs (bounding boxes).

2.3.5   *Conclusion*

This brief literature review shows that each type of method has its own strengths and weaknesses. When facing a new problem, each of them can be potentially considered. The final choice will depend on different aspects of the problem such as the inherent variability of the images, the quality of the images (quantity of noise), the variability of the structures to detect, the reproducibility, etc. While image-based approaches (Section 2.3.1) are sufficient for simple problems, atlas-based methods (Section 2.3.2) can be recommended when looking for a large number of structures (e. g. brain structures), and learning methods (Section 2.3.4) are more dedicated to problems with a high variety of inputs (necessity of learning the variability).

The previous description also clearly highlights the apparent advantages of machine learning methods for the problem of multi-objects localization. They generalize well and allow for generic and accurate solutions. More specifically, regression methods seem to answer to rapidity needs. In this work we propose to explore this type of solution, trying to keep in mind the rapidity aspect. Indeed, even if this type of approach shows better and better performance, the increase in complexity (both in terms of algorithm design and computational speed) may not always be justified.

2.4 MULTI-OBJECT SEGMENTATION IN MEDICAL IMAGES

Although localization and segmentation are two different problems, one can categorize methods from the literature in a similar way. We distinguish the image-based, atlas-based, graph-based, learning-based and deformable model approaches. The latter is given particular attention, as it is closely related to our work (Chapter 4).

2.4.1 *Image-based approaches*

As in Section 2.3.1 we define as image-based the approaches that take only into account the information given in the image such as grey-level intensities or edges for example. One popular approach is to rely on morphological operations, with some prior knowledge needed to determine the size (and/or shape) of the structuring elements. Kobashi and Shapiro (1995) propose to combine the dynamic thresholding to a knowledge-based procedure to segment several organs in CT (the kidneys, the spleen and the liver). Sakashita et al. (2007) model the intensity distributions and perform region refinements using morphological operations in order to extract several organs in multi-phase CT images. Linguraru and Summers (2008) also model the organ intensity distributions but refine the contours with morphological operations and geodesic active contours (Caselles et al., 1997). This type of approach may work well in very specific contexts. They are however very image-specific and thus not adapted to more variable problems (intensity changes, change in field of view, pathologies, etc.).

2.4.2 *Atlas-based approaches*

Atlas-based segmentation was originally proposed by Evans et al. (1991); Gee et al. (1993); Sandor and Leahy (1997) for brain segmentation where the structures are numerous and the spatial relationships are very strong. They proposed to label manually a MRI volume and to register it on new MRI volumes with non-linear deformations. The result is obviously highly dependent on the chosen atlas. Moreover, as the inter-subject variability is not taken into account, such a method may give wrong results for structures with high variability or for patients with pathologies.

Collins et al. (1999) and Park et al. (2003) propose to take into account the shape variability by building probabilistic atlases. This aims at building an atlas from several images giving probabilities of belonging to a given organ at each image location. These atlases are often built organ per organ, as this approach requires a registration step between the training images.

The use of mean atlases was also proposed. The goal is to capture the common parts of several image samples. There are many ways to build a mean atlas.

It often consists in registering all images in a common referential and then average them. For instance, Joshi et al. (2004) propose an unbiased construction (no need of a reference template) in the large deformation diffeomorphic setting.

The mean atlas however smooths out the individual particularities which may then make it difficult to handle high inter-subject variabilities. To tackle this problem, Fenchel et al. (2008) propose to build statistical atlases (with principal component analysis (PCA)). Their approach seems to be fast (10 to 30 seconds on GPU) and it gives reasonable results (around 2cm of mean accuracy on anatomical landmarks). However this kind of atlas remains very specific to one type of field-of-view, which may considerably restrict the possible applications.

Multiple atlases can also be used to capture more information. Heckemann et al. (2006); Blezek and Miller (2007) show that this method gives more robust results. The principle is to register every template to the new image, then to propagate the label information and to fuse the votes with a voting approach such as the method called *simultaneous truth and performance level estimation (STAPLE)* (Warfield et al., 2004). A hierarchical approach recently proposed by Wolz et al. (2012) gives very good results, at a high computational expense however (about one hour). These approaches improve the standard atlas-based methods, albeit at the cost of increased computational time. In that respect, *Atlas Forests* (Zikic et al., 2013) propose an efficient, random forest-based alternative to the registration-based evaluation of textbook multi-atlas methods.

Despite their computational cost, atlas-based methods are under active research and are considered state-of-the-art in many communities (e. g. for brain segmentation). They are indeed particularly suited to the delineation of multiple structures under strong spatial constraints, on condition that pathologies do not deform too severely the structures. However, even if recent improved solutions gave promising results, advanced variants still have difficulties coping with the variability, the pathologies and inherent image-specific challenges. For this reason, atlas-based methods are often followed by an additional contour refinement step (Ding et al., 2005; Duay et al., 2005). The combination with machine learning approaches may also overcome these difficulties. Recent developments such as the works of Zikic et al. (2013) (Atlas forest), Lombaert et al. (2014) (Laplacian Forests) and Konukoglu et al. (2013) (Neighborhood Approximation Forests) exploit similarities between training images to capture more specific, more local variations (in the sense of a similarity metric) than classical global atlases/models. These works relate to multi-atlas approaches in that their key ideas can be used to build more specific, more accurate atlases while still benefiting from (local) clustering of training data. These works focus notably on learning and exploiting the appropriate similarity metrics within the framework of RF (see also Section 2.2.1.6).

We refer the reader to the surveys of Kalinic (2008) and Cabezas et al. (2011) for further information on atlas-based approaches.
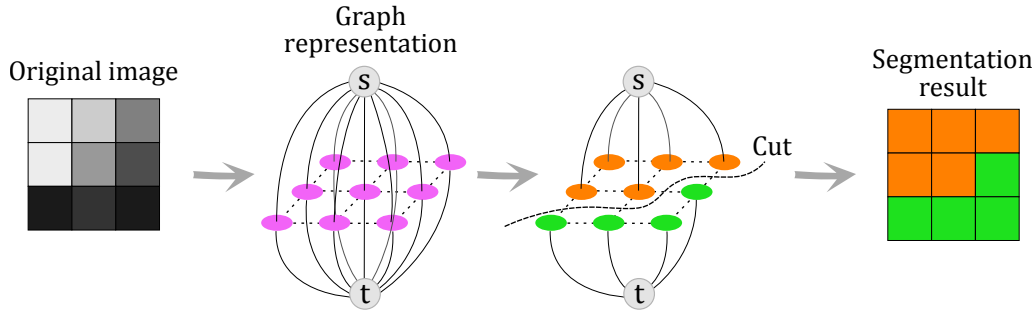
Figure 2.5: Graph-cut segmentation principle.

2.4.3 *Graph-based approaches*

The methods presented in Section 2.3.3 are actually solving localization and segmentation jointly. We refer to this section for more details on knowledge-based graph approaches.

Besides organs and their relationships, a graph can also be used to model an image in its entirety. The most general model represents an image as a connected graph with a node for every pixel (or voxel in 3D) and links between any two pixels. The weights of the edges/links correspond to a similarity criterion based on pixel features. The graph can be considered as a probabilistic model that represents the conditional dependence (the weighted links) between random variables (the nodes). An image is then considered as a realization of random field variables with some model of conditional dependences (Perez et al., 1998). Different models of random fields are used in computer vision, the most well known being the *Markov random field* (Li, 1995). The segmentation problem can then be found by estimating the maximum a posteriori (MAP). Such an optimization is however challenging, especially in the case of large data. Casting it to a *graph-cut* formulation (Boykov and Jolly, 2000) leads to computationally efficient solutions. The problem is then seen as a partition problem where the minimum cut of the graph corresponds to the image segmentation (see Figure 2.5). It is usually solved with a max-flow/min-cut optimization borrowed from graph theory. This approach has the particular interest to be easily used to perform interactive segmentation (Boykov and Jolly, 2001). A specific overview of interactive segmentation approaches is given in Section 5.1.2.

The generalization of graph partitioning approaches to multiple objects is not straightforward. Indeed, the original graph-cut formulation can only find the global optimum of binary labeling problems. Boykov et al. (2001) first showed that it is possible to find approximate solutions for multilabel graph-cut problems. The problem of multi-organ segmentation with graph-cut was also developed by Linguraru et al. (2012b). They use a four-dimensional (4D) graph based on intensity information, for segmenting simultaneously several organs in multi-phase CT. Computation takes about 15 minutes.

Graph-cut methods present several drawbacks: (i) the metrication artifact (or *blockiness*) due to the 4-connectedness in 2D or 8-connectedness in 3D, (ii) the shrinking bias favoring small regions over bigger ones (due to the minimum cut concept), (iii) the memory usage which increases quickly with the image size. A solution to avoid the shrinking bias is to introduce the normalization of the cost of the cut by the size of the resulting object (as in Shi and Malik (2000) for instance). Even if this NP-hard problem may be approximated, the additional computation cost is high.

2.4.4 *Learning-based approaches*

Segmentation can be considered at the pixel/voxel level and formulated as determining whether a voxel belongs to a given structure or not.

Learning-based classification methods can be used to solve multi-structure segmentation problems: one class is associated to each structure. Then given image features, each pixel/voxel of the image is associated to one of the target structures. However, such an approach may lack shape consistency due to the pixel-per-pixel decision. Lee et al. (2003) use a neural network to segment the target structures independently in each slice and then connect each region using fuzzy rules to get coherent regions. As presented in Section 2.2.1.6, machine learning methods may be extended to handle spatial consistency. Entangled decision forests (Montillo et al., 2011) are a good example.

The shape consistency may also be integrated by combining voxel classification with other shape-based approaches. For instance, Glocker et al. (2012) perform a joint classification/regression to segment abdominal organs. They optimize jointly the voxel classification and shape regression. Iglesias et al. (2011) combine the voxel classification with graphical modeling of the spatial relationships and probabilistic atlases.

2.4.5 *Deformable model approaches*

Deformable models have been the subject of very active research for the last three decades, probably due to their cross-disciplinary nature, at the junction of geometry, physics and approximation theory (Moore and Molloy, 2007). In particular, they have given birth to multiple applications in the medical image analysis domain.

As their name suggests, the principle of such approaches is to deform a model so that it meets given criteria (e. g. in segmentation we want that the model fits the contours of a given object) while preserving some properties of the model, such as its regularity and/or topology. Most of the time the model represents a shape or a curve/surface. It can be defined in various ways, for instance by using a parameterization (e. g. a set of points) or by defining the shape contour
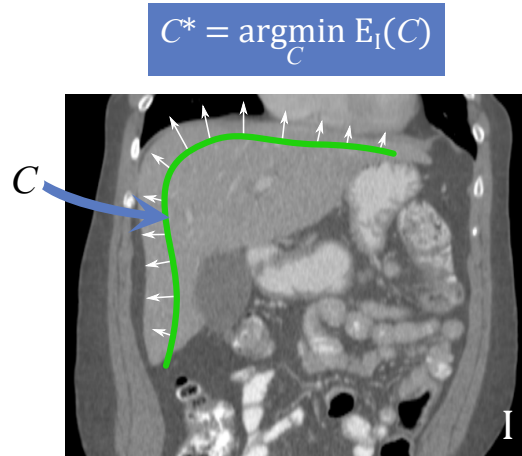
$$C^* = \underset{C}{\operatorname{argmin}} \, \mathrm{E_I}(C)$$

Figure 2.6: Illustration of the snakes approach principle. *C* is the evolving curve and *I* is the image. The colorbar corresponds to the level of the distance function.

as the zero level of a function of superior dimension (implicit representation). Montagnat et al. (2001) give an extensive review of the possible representations and their associated evolution behavior. Depending on this representation, several strategies have been proposed to answer the problem of segmentation. The surveys of McInerney and Terzopoulos (1996); Blake (1998) and more recently Moore and Molloy (2007); Heimann and Meinzer (2009); Jayadevappa et al. (2011) give a good overview of these strategies.

As our work (in particular Chapter 4) fall within this domain, we propose to give a brief overview of its key principles and then we focus on the multi-object extensions.

2.4.5.1 *Deformable model approaches for single object segmentation*

SNAKES   One of the most influential deformable models of computer vision are the original *active contours* (or *snakes* (Kass et al., 1988)). They aim at deforming a hypersurface (a curve in 2D or a surface in 3D) so that it matches the edges of an image (see Figure 2.6). The problem is formulated as an energy minimization. The deformation follows different constraints, such as image forces and internal elastic energy (e.g. the smoothness of the contour), that pull it towards the object contour. The image term basically relies on intensity or edge information (Caselles et al., 1997). The energy minimization is originally performed through gradient descent.

Snakes present however the limitation of converging relatively poorly and of being highly dependent on the model initialization (the energy formuation is not convex). Classical snakes need to be initialized really close to the model to reach the correct local minimum and achieve a proper segmentation. Multiple variants and extensions have been proposed to overcome these problems and
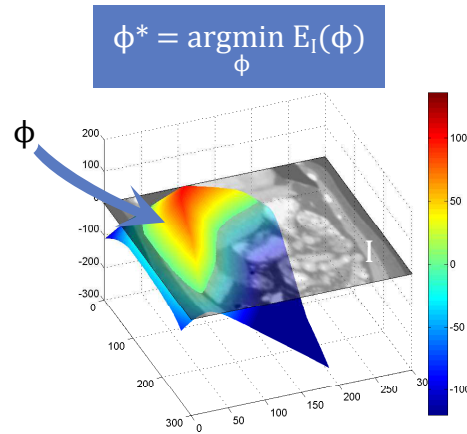
$$\phi^* = \underset{\phi}{\mathrm{argmin}}\, \mathrm{E_I}(\phi)$$

Figure 2.7: Illustration of the level set principle. ϕ is the level set function and *I* is the image.

get more robust algorithms. For instance Cohen (1991) proposes to add a *balloon force* aiming at inflating the contour. The contour initialization thus no longer needs to be close to the solution but in return should be inside the object to segment. Xu and Prince (1997) introduce the gradient vector flow (GVF) snakes where the model evolves on a force field computed from the spatial diffusion of the image edge map gradient. Main interest is to increase the capture range and thus improve the robustness to the initialization.

LEVEL SET   To overcome the problem of numerical issues induced by the explicit contour representation and to handle topology changes, level set approaches have been proposed (first introduced by Osher and Sethian (1988)). The principle, illustrated in Figure 2.7, is to represent a contour as the zero level of an implicit function (in general a distance function). This kind of representation allows the topology to change. On the other hand, level set methods can only be used for closed curves. The evolution equation can be derived from the snakes energy equation. This leads to a partial differential equation requiring careful numerical schemes to avoid potential instabilities. We refer the reader to the survey of Suri et al. (2002) for a detailed overview of the different kinds of level set approaches.

Classical snakes and level set methods do not integrate prior knowledge on the target shape. Rousson and Paragios (2002) propose to introduce a shape prior in the level set formulation, to exploit the strong prior knowledge available in medical image analysis. This shape prior is formulated as a regularization term defined as the L^2 distance of the evolving level set function to a level set of reference. Such a formulation only allows signed distance functions, which should be guaranteed by the optimization framework (with distance function reinitialization at each iteration for instance). Although the topology adaptive-

$$\psi^* = \underset{\psi}{\arg\min} \; E_I(\phi \circ \psi)$$
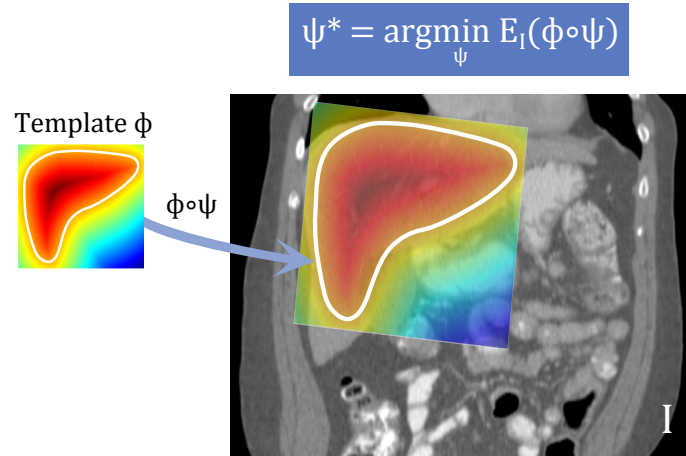
Template ϕ

$\phi \circ \psi$

I

Figure 2.8: Illustration of the template deformation approach. ϕ is the implicit template, ψ is a transformation and I is the image.

ness may be interesting in some computer vision problems, this is however rarely the case in medical imaging where topology preservation is often desirable. Level set methods seem then to have a major drawback in this case. Let us see how to solve this problem with another method based on template deformation.

TEMPLATE DEFORMATION MODELS   In the level set methods, the optimized variable is the shape represented by the level set function. With template deformation approaches, we focus on optimizing the *transformation* of the reference model (also called template) (see Figure 2.8). The shape prior is also intrinsically embedded in the formulation. This representation can be embedded in a region-based variational framework (Mumford and Shah, 1989; Chan and Vese, 2001). The regularization term is applied to the transformation itself, which guarantees nice properties for the contour (more or less smooth for instance). The topology is also guaranteed. Different transformation models were proposed so far. An and Chen (2007) only optimize global similarity transformations while Saddi et al. (2007a) use a non-rigid transformation model regularized with a diffeomorphic fluid model allowing large deformations. Somphone et al. (2008) use a model based on finite element with partition of unity, achieving a compromise between global and local deformations. Huang and Metaxas (2008) use the model of *free form deformations* (Rueckert et al., 1999) in their *Metamorphs* algorithm. Mory et al. (2012) propose another formulation relying on the algebraic definition of shape (Soatto and Yezzi, 2002; Yezzi and Soatto, 2003). The main novelties of this approach are: (i) the invariance of the shape regularization to the pose, (ii) the built-in smoothness of the displacement field using Gaussian convolution, (iii) the possibility of incorporating user interactions easily. These choices contribute to an efficient solution adapted to clinical context.
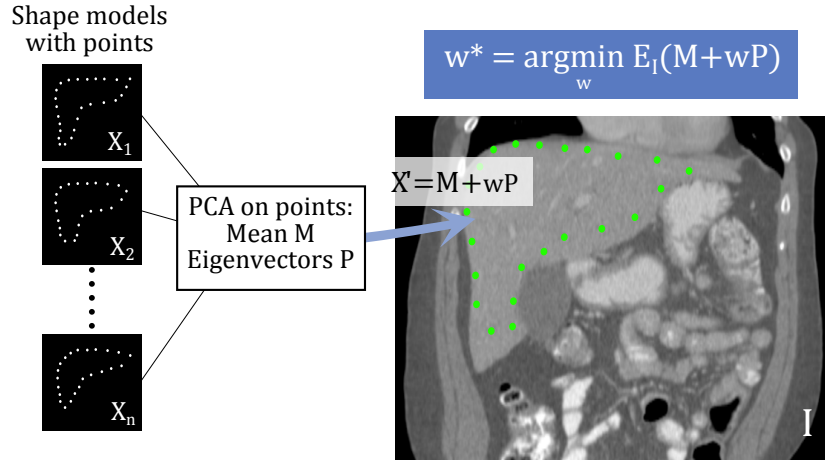
Figure 2.9: Illustration of the SSM approach. $\{X_i\}_{i \in [\![1,n]\!]}$ is the set of point shape models, w is a vector of weights and I is the image.

We give more details on this method in Section 4.2, since it is at the basis of our work. Note that this approach can even be enhanced to take into account shape variations Prevost et al. (2013a).

STATISTICAL MODELS   The principle of statistical models is to embed the statistical shape variations into the deformable model framework. One of the most popular work, called statistical shape model (SSM), was proposed by Cootes et al. (1992, 1995). They use an explicit representation of shape: each shape is represented as a set of reproducible points (e.g. anatomical landmarks or 3D mesh vertices). The idea is to model the distribution of the landmark positions over a set of training shapes (see Figure 2.9). The mean shape is obtained by averaging the points positions over all these training shapes. By assuming that all the shapes of the database are scattered around the mean shape according to a Gaussian distribution, it is then possible to capture the variation information with a PCA. Then, a new shape can be represented as a linear combination of the eigenvectors of the resulting covariance matrix. This shape model is then integrated in an active contour formulation for the segmentation. Heimann and Meinzer (2009) give a comprehensive and detailed review on this type of approaches for medical imaging applications. One can notice the major drawback of the point-to-point correspondence required to align the shapes and perform the statistical computations.

Statistical learning may also be performed with implicit representation of shapes. In the works of Leventon et al. (2000); Tsai et al. (2003) each shape is represented as a signed distance function. They perform the statistical analysis on the set of these functions considered as vectors. Leventon et al. (2000) embeds this information in the geodesic active contours approach, while Tsai et al. (2003) integrates it in the region competition framework. The implicit

representation has the interesting advantage to overcome the problem of point-to-point correspondences. However the approach suffers from a mathematical pitfall: the signed distance functions are not stable by linear combination. Rousson and Paragios (2002) propose a solution to this problem by constraining the function to remain a distance function in the optimization framework.

2.4.5.2 *Extension to multi-object segmentation*

After a brief overview of the main deformable model approaches for single object segmentation, we now give an insight of the extensions proposed for multi-object segmentation. We focus in particular on medical image applications.

IMPLICIT MODEL-BASED METHODS   The extension of the level set approaches to the segmentation of multiple objects mainly takes its principles in the physics domain. Zhao et al. (1996) propose to use multiple level set functions to solve the problem of multiple junctions separating several phases in physics. A constraint of non-overlapping is added to the energy term and solved with the Lagrangian method. The term *multiphase* used in image segmentation comes from the physics domain. In the domain of image segmentation there are actually several ways of representing different objects with level set functions.

A first idea is to use a multiphase representation that directly gives a partition of the image space by construction (no overlap and no vacuums) as in the work of Vese and Chan (2002). The original approach of these authors requires $log(n)$ level set functions for the segmentation of n regions in 2D. They show, according to the *four colors theorem*, that only two level set functions may suffice to represent any partition. However this method has three key limitations: (i) the objects in the same phase should have the same image features, (ii) the image-driven forces are limited to region-based forces, (iii) the penalty on the contour length applies on the level set rather than on the objects. In the work of Fan et al. (2008); Bogovic et al. (2012) only two level set functions and two label functions are needed. Their framework permits to apply different forces on different boundaries. The computation time is however rather high.

Such methods looking for an image partition are not adapted to images where there are for instance several objects and one non-uniform background with various contrasts and textures. In this case the background cannot be considered as another phase and the partition representation may not be the most appropriate. Another idea is thus to associate one level set per component (or phase) as done in the works of Samson et al. (2000); Paragios and Deriche (2000); Yang et al. (2004); Yan et al. (2005); Litvin and Karl (2005); Fan et al. (2008); Kohlberger et al. (2011); Bogovic et al. (2012). The goal is to delineate several structures in the image by defining one level set function for each object with the exception of the background. Nevertheless, this type of solution requires to introduce additional constraints to prevent the problem of overlap-

ping and/or vacuums (if needed). Yang et al. (2004) do not clearly penalize the model with this special type of constraints. They incorporate probabilistic shape information in the contour evolution through a MAP estimation. The image-driven forces are assumed to be sufficient to prevent overlapping. Yan et al. (2005) extend this approach and incorporate repulsive forces preventing objects intersections. Litvin and Karl (2005) enhance the modeling of shape relationships (with PCA) by introducing new features such that the inter-node and inter-object distances or the curvature. Kohlberger et al. (2011) show that this kind of level set approach can be used efficiently for 3D segmentation of abdominal organs (2-3 minutes), at the cost of intricate parameter tuning.

The level set framework is not the only possible choice for the segmentation of multiple objects with implicit representation. For instance, Tsai et al. (2004) give another formulation where the goal is to find the best transformation and linear combination in a shape space (using PCA) according to image features. A gradient descent is used to solve this problem. However in practice, mainly due to the non-convexity of the problem, the solution is highly dependent on initialization.

POINT-BASED METHODS   The extension of point-based approaches to multiple structures may be done in different ways depending on the type of organs or structures to segment. First, the anatomical structures may be represented in a single common model. Zheng et al. (2008) apply this idea for heart segmentation in CT. The heart is composed of different structures (ventricules, epicardium, atrium etc.) that can be represented using a single point distribution model in the form of a mesh. The whole heart is first localized using MSL and the mesh deforms following learned boundary profiles. Chen et al. (2013) apply the same type of principle for the segmentation of abdominal and bone structures. A global mesh model of the different target structures is built, registered and deformed following a segmentation algorithm based on graph-cut and active shape model (ASM). Although such a principle guarantees the non-overlapping of the different target structures, only local deformations can occur as the relative pose of the organs is fixed. The inter-patient variability may then be difficult to handle, especially for soft structures that may have different relative positions (e.g. liver and kidneys). Another type of approach is to segment each organ independently using single organ segmentation method. This is done in the works of Seifert et al. (2010); Lu et al. (2012) for instance. Each object is localized independently (with MSL for instance) and the mesh model of each organ is deformed following image features. The authors assume that the image information will be strong enough to prevent overlapping, which may not be always the case. Wu et al. (2014) thus propose to add a step based on multi-layered graph-cut to remove the overlaps. They obtain accurate and relatively fast (about 90s) results for 3D bone segmentation. In such a framework, we may question the robustness to bad initializations and the final coherence

of the results. Indeed, the spatial relationships of the organs are not taken into account in the core formulation, merely corrected a posteriori. A strong variability in relative positions and organ shapes may be difficult to solve with such approaches. Other works propose to integrate organs relationships constraints directly in the segmentation framework. Gao et al. (2011) extend the active contours to multiple objects while adding a force to prevent overlapping. They segment 11 organs in 8 minutes in CT. Uzunbas et al. (2010) propose another approach based on active contours integrating pose estimation. Although it is very slow (one hour for one CT volume), this approach allows to optimize jointly the poses and the local deformations of the different object contours.

2.4.6 *Conclusion*

Existing methods for multi-object segmentation are generally extensions of single segmentation approaches. However, we saw in this review that the extension is not always straightforward. Modeling organ relationships (non-overlapping, inclusion, adjacency), while maintaining a coherent global result, can be viewed as a key challenge for multi-organ methods. In this light, the literature on advanced multi-organ methods is still relatively scarce but under rapid evolution. We believe that the incorporation of higher level information is of prime importance towards more robust and accurate techniques. This information is becoming more readily available with the advance of large medical databases. Key difficulties include the extraction and modeling of this prior knowledge, and its incorporation into flexible and efficient computational frameworks. Our work is an attempt at providing the latter.

MULTI-ORGAN LOCALIZATION WITH SHAPE PRIOR

Abstract

This chapter presents a fast, accurate and robust method for the localization of multiple organs in 3D medical images. The localization result is given as a confidence map given the chance of each voxel of the image to belong to a said organ. The confidence map also presents the advantage to be organ dedicated and shape consistent.

This method is based on a coarse-to-fine cascade of RF regressions to which is added shape prior by using probabilistic atlases.

This method is evaluated on a large and diverse database of 130 CT volumes for the localization of 6 abdominal organs (the liver, the two kidneys, the spleen, the gallbladder and the stomach). An extensive study of the different learning and testing parameters is first given. It shows both their robustness to medium variations and their influence on the final algorithm accuracy. Secondly, the robustness and the accuracy of the approach is evaluated and demonstrated. By comparing our results with two existing methods, we show significant improvements of the localization performance at a small computation cost (about 5s. to localize 6 organs).

A part of this work has been published in Gauriau et al. (2014) and a more complete and extended version has also been published in Gauriau et al. (2015).

Contents

3.1 INTRODUCTION

3.1.1 *Context*

With the ever growing amount of 3D medical acquisitions, automatic, robust and accurate anatomy localization is of prime interest. First, it directly enables faster data navigation and visualization of target structures which can undoubtedly save the radiologist some time (Andriole et al., 2011). Secondly, the organ localization is a key initialization step for tasks such as segmentation or registration (for liver segmentation for example (Gauriau et al., 2013)). It is, overall, a crucial component to streamline complex workflows such as medical treatment follow-up (in liver radiotherapy for instance, the volume of the liver is required to compute the dose (Murthy et al., 2005)). To be integrated in a clinical context, such a component should be almost transparent. It should help the user or subsequent algorithms without a significant computation overhead. It should of course also be accurate and robust enough to be reliable as an intermediary result. Therefore, the speed and the performance of the anatomy localization are two major aspects to take into account in the design of such an algorithm.

A literature review on the subject is given in Section 2.3. It shows that machine learning methods give today the best results. However the increase in accuracy sometimes goes hand in hand with a complexity raise. This finally results in slower algorithms and a difficulty to tune their parameters. In Section 2.3.4 we have also highlighted the concurrence between landmark- and ROI- based approaches. While the first approach may be faster, the second one seems to be adapted to a larger variability in the type of structures to localize. Moreover, as we mentioned earlier, the identification of reproducible landmarks causes a problem when considering soft tissues with high shape variability or when considering small structures. The choice of ROI-based approaches seems more appropriate to the localization of multiple structures of various sizes and shapes. In Section 2.2.1.6 we reviewed the different strategies specific to ensemble methods, in particular the RF methods, to take into account the spatial relationships (essential aspect in the localization of multiple structures) and the contextual information, which are characteristic of medical images.

In this work we propose a method fulfilling the following properties:

- accuracy,

- speed,

- robustness to various contexts (change in field of view, contrast, morphology).

We show in this chapter that by gathering:

- machine learning regression,
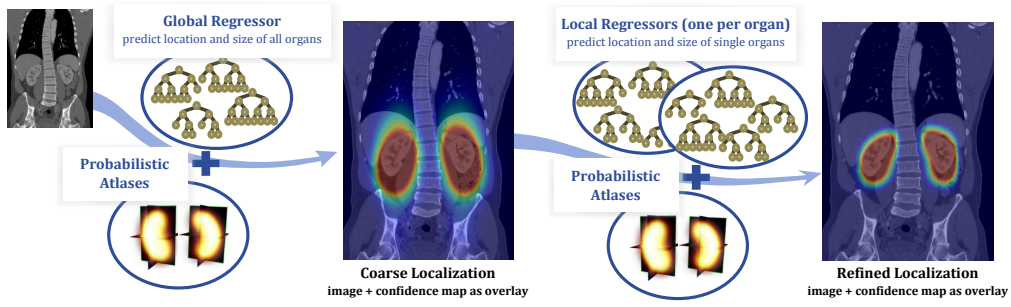
Figure 3.1: Outline of the method applied to the localization of the kidneys in a 3D CT image. A cascade of RF regressors provides a final accurate localization in the form of one confidence map per organ (both displayed as overlay over the image, red indicates a high chance of belonging to the organ while blue indicates a low chance).

- contextual information,

- spatial relationships,

- shape prior,

we can address these requirements. To our knowledge this is the first time that all these aspects are fulfilled together to solve the problem of multiple organs localization.

3.1.2   Outline

To answer the problem of multi-organ localization we propose to perform a cascade of RF regressions conditioned with organ dedicated shape priors. It gives confidence maps for each organ as a localization result. The latter are organ dedicated probability maps given the chance of an image voxel to belong to a said organ. Figure 3.1 outlines the principle of the method and shows some examples of confidence maps. The idea is first to localize roughly all the organs using one RF regression embedding the organ relationships. The resulting vote distributions are combined with organ dedicated probabilistic atlases and used to identify the ROI corresponding to each organ. Afterwards, organ dedicated RF regressors are used to refine independently the localization of each organ.

This relates our method to the works of Criminisi et al. (2011, 2013) on RF regression for anatomy localization, Cuingnet et al. (2012) on cascade of regressors for kidneys localization, and Sun et al. (2012) on conditional regression models. By keeping the method relatively simple to implement, we show that we can get accurate and robust performances at a small computation cost. Indeed our method only requires to learn several classical RF regressors in cascade,

to build one probabilistic atlas per organ and to implement an efficient confidence map computer. The main critical part may be this last point for which we actually propose a fast implementation based on convolution (see Section 3.2.4). The low training computation time also allows us to perform an exhaustive evaluation of the different parameters of our framework (Section 3.4). While giving us a good understanding of the different parameters contributions, this also guarantees to get the best possible accuracy. We show in Section 3.5 that our method provides, in a few seconds, a consistent and accurate localization for multiple organs in 3D medical images.

3.1.3 *Random forest regression for anatomy localization*

In this section we propose an overview of the anatomical localization method as originally introduced by Criminisi et al. (2011, 2013). This will provide the reader with a background which is necessary to understand the method proposed in this chapter.

Detection with regression to leverage the global context was first proposed in (Zhou and Comaniciu, 2007; Zhou, 2010). Criminisi et al. (2011, 2013) proposed to extend it to multiple structures (26 structures such like the left and right femoral necks, the gallbladder, the femurs, the kidneys, the liver, the left and right adrenal glands etc.). Each structure o is represented by its 3D axis-aligned bounding box represented by a 6-dimensional vector $\mathbf{b}_o$ which elements are the face positions (in mm). One single RF regressor is learned which output is the multi-dimensional vector $[\mathbf{b}_1, \mathbf{b}_2..., \mathbf{b}_{N_o}]$.

Section 2.2.1.6 explains what is a RF and how it is trained and used. In the tree nodes, Criminisi et al. (2011, 2013) use similar visual features to those introduced in (Gall and Lempitsky, 2009; Criminisi et al., 2009; Shotton et al., 2009) which are difference of mean intensities over randomly displaced asymmetric 3D patches of the image volume I. We illustrate this type of features in Figure 3.2. Note that a patch can be empty or reduce to a single voxel. These features have the advantage to capture spatial context.

In the work of (Criminisi et al., 2011, 2013), the training set is composed of voxels chosen on a regular grid within $\pm 10\,\mathrm{cm}$ of the center of each axial slice in the training volumes. Let us consider $\mathbf{v}$, a voxel of this training set. For each bounding box structure indexed by o, it is associated with the offset to this bounding box $\mathbf{b}_o$: $\mathbf{d}_o = \hat{\mathbf{v}} - \mathbf{b}_o$ (where $\hat{\mathbf{v}} = [v_x, v_x, v_y, v_y, v_z, v_z]$). The trained RF actually regresses the relative distances $\mathbf{d}_o$ but for convenience we assume that the regressor predicts the bounding boxes $\mathbf{b}_o$.
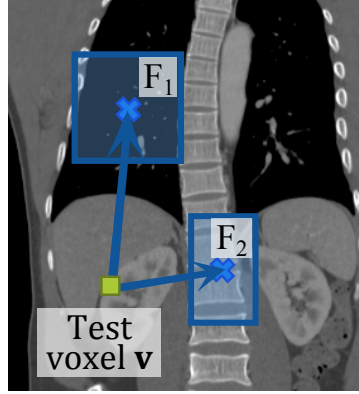
Figure 3.2: Example of 1D features computed at each tree node: difference of mean intensities over two 3D patches F_1 and F_2.

For the training phase, the node optimization is driven by the maximization of the information gain. This information gain IG is the objective function used to find the splitting functions stored in the nodes. At a node n, it is defined as:

$$IG = H(S_n) - \left(\frac{|S_n^{left}|}{|S_n|}.H(S_n^{left}) + \frac{|S_n^{right}|}{|S_n|}.H(S_n^{right}) \right) \qquad (3.1)$$

where H is the entropy and where we keep the notations introduced in Section 2.2.1.6 (S_n represents the training samples at node n and S_n^{left} and S_n^{right} are the candidate subsets going to left or right node). The information gain is the differential entropy between a node and its children. By following a Gaussian distribution model, the minimization of IG implies minimizing the determinants of the 6x6 covariance matrices associated to each structure.

The tree construction stopping criteria are of two kinds: (i) the number of samples reaching the node is below a fixed threshold, (ii) the fixed maximum tree depth has been reached. In the leaves, they store the mean, covariance matrices and class priors $p(o)$ of the samples that reached it.

Once the forest is built, a test can be performed. For a new test image I the test voxels are selected in the same way as in the training phase. Each test voxel is pushed to each tree of the forest. Following the splitting rules defined in the nodes, it finally reaches a leaf l in which are stored the posterior distributions $p(\mathbf{b}_o|l)$ for each structure o. The posterior probability for $\mathbf{b}_o$ given a voxel $\mathbf{v}$ and an image I is defined as:

$$p(\mathbf{b}_o|I, \mathbf{v}) = \sum_{t=0}^{T} \sum_{l \in \hat{\mathcal{L}}_t} p(\mathbf{b}_o|l, \mathbf{v}, I)p(l) \qquad (3.2)$$

where T is the number of trees in the forest and $\hat{\mathcal{L}}_t$ is a subset of leaves of the tree t. The subset $\hat{\mathcal{L}}_t$ is the set of leaves with the smallest uncertainty (by

looking at the trace of the covariance matrices) and containing 75% of the test voxels and $p(l)$ is the proportion of samples reaching the leaf l. The final prediction is finally given by estimating the MAP of $p(\mathbf{b}_o|I,\mathbf{v})$. The authors test this approach with two applications, in 2D and in 3D, for the localization of multiple anatomical structures.

Cuingnet et al. (2012) shows that this method could also be applied by modifying some elements. The objective function used to train the tree nodes can be replaced by the residual sum of squares (RSS) defined as:

$$RSS = \sum_{s \in S_n} (s - \overline{S_n})^2 - \left(\sum_{s_r \in S_n^{right}} (s_r - \overline{S_n^{right}})^2 + \sum_{s_l \in S_n^{left}} (s_l - \overline{S_n^{left}})^2 \right) \qquad (3.3)$$

where $\overline{S_n}$, $\overline{S_n^{right}}$, $\overline{S_n^{left}}$ are the mean over the training samples and right and left training sample candidates respectively. It is faster to compute and gives similar results to the information gain maximization. The combination of the votes for the final prediction can also be replaced by the median over the best votes (this confidence is given by the probabilities $p(\mathbf{b}_o|I,\mathbf{v})$). Cuingnet et al. (2012) claims this is more robust.

In the following work we use the same type of RF regression with this different characteristics:

- the same type of visual features as in Criminisi et al. (2011, 2013),

- the RSS as objective function for the node optimization,

- Gaussian model in the leaves,

- votes combination with the median.

3.2 MULTI-ORGAN LOCALIZATION WITH SHAPE PRIOR

3.2.1 *Notations*

We use the following notations hereafter:

- $\Omega \subset \mathbb{R}^3$ is the image domain,

- $o \in [\![1, N_{\text{org}}]\!]$ are the indices of the N_{org} organs to localize,

- $k \in [\![1, K]\!]$ is the index for one single prediction/regression,

- $G = S \circ T$ is a geometric transformation of $\mathbb{R}^3$ composed of a translation T and anisotropic scaling S, and $G_{o,k}$ denotes the geometric transformation associated to organ o and regression k,

- $\mathbf{b} \in \mathbb{R}^6$ denotes a 3D bounding box parameterization either as its center location and size $(\mathbf{c}, \mathbf{s})$ $(\mathbf{c}, \mathbf{s} \in \mathbb{R}^3)$ or as its extremal vertices $(\mathbf{b}_{min}, \mathbf{b}_{max})$ $(\mathbf{b}_{min}, \mathbf{b}_{max} \in \mathbb{R}^3)$, $\mathbf{b}_{o,k}$ then denotes the bounding box parameterization of organ o and regression k,

- $\mathbf{x} \in \Omega$ denotes any voxel of an image I,

- $\mathbf{v_k}$ denotes a test voxel for regression k.

3.2.2 *Presentation of the method*

The method presented here is designed for the localization of multiple organs. In Figure 3.1 we present an outline of the method for the localization of the two kidneys. The method works in two steps in a global-to-local fashion. It makes use of three main components: (i) bounding boxes predictors (in our case: RF regression as described in Section 3.1.3), (ii) probabilistic atlases for each organ, (iii) a confidence map computer. Confidence maps combine votes aggregation with a probabilistic atlas, resulting in organ-dedicated probabilistic maps. The first step of our method aims at finding the global organ locations with a single regressor while the second step aims at refining individually the organ positions with organ dedicated regressors. In Section 3.2.4 we explain the principle of confidence maps and in Section 3.2.5 we detail the cascade approach.

In the main works on organ localization with regression, spatial vote distributions are not fully exploited. However, as shown with the Hough forests (Gall and Lempitsky, 2009), vote aggregation can give more information than a single measure of the distribution (e.g. mean or median). For this purpose we introduce the notion of *confidence map*, which encodes the confidence in finding a target organ at a given location. It is built through an aggregation process, making use of both the spatial distribution of regression votes and of organ shape priors through probabilistic atlases.

3.2.3 *Probabilistic atlas construction*

To compute a probabilistic atlas of an organ we first register the binary masks of several samples of this organ. For this purpose let $\{M_i\}_{i \in [\![1, N_\mathrm{m}]\!]}$ be the set of N_m different cropped binary masks of the organ such that $\forall i, 1 \leq i \leq N_m$ $M_i : \Omega \to [0, 1]$ and where M_r ($r \in [\![1, N_\mathrm{m}]\!]$) is a mask of reference chosen arbitrarily. We transform each mask with a transformation Ψ_i (rigid and anisotropic scaling) in order to scale up the masks M_i to the same size as M_r (considering the bounding
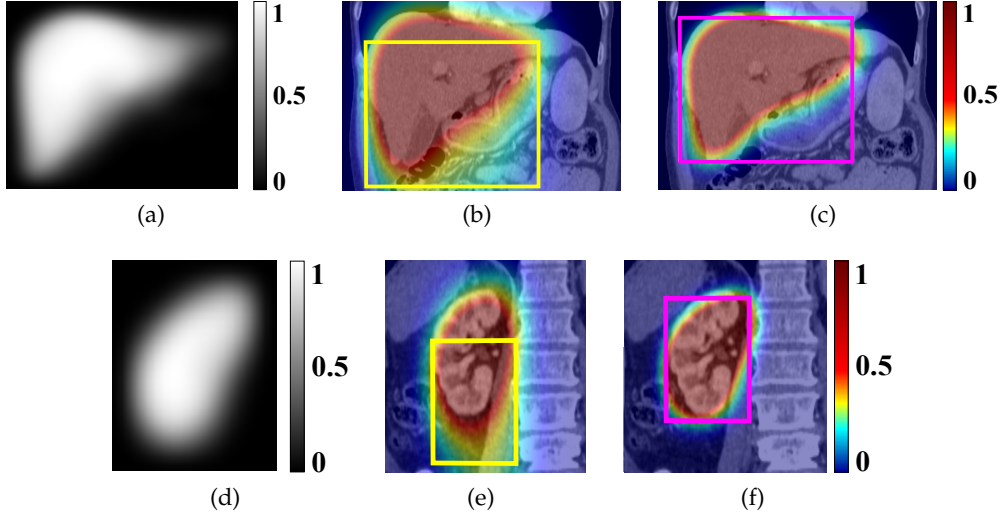
Figure 3.3: Atlases of the liver (a) and the kidney (d), localization of the liver and the right kidney: images with confidence map as overlay (images have been cropped) and predicted median box after global step (b)(e) and local step (c)(f).

boxes of the binary shapes). The probabilistic atlas A is then computed as an average of these masks:

$$A(\mathbf{x}) = \frac{1}{N_{\mathrm{m}}} \sum_{i=1}^{N_{\mathrm{m}}} M_i \circ \Psi_i(\mathbf{x})$$

Each value $A(\mathbf{x})$ evaluates the probability of a voxel $\mathbf{x}$ to belong to the organ. Figure 3.3a and Figure 3.3d show probabilistic atlases for the liver and the right kidney respectively. We normalize the atlas in location and scale in the range $[-1, 1]^3$. Thus if $\mathbf{x} \notin [-1, 1]^3$ we have $A(\mathbf{x}) = 0$.

3.2.4 *Confidence map construction*

Here we explain the construction of the confidence maps and propose an implementation to compute them efficiently. The principle is illustrated for the kidney localization in Figure 3.4. Numerous voxels of the image vote for the kidney bounding box location and each of these votes are combined with the right kidney probabilistic atlas (following an aggregation process).

ORGANS BOUNDING BOXES PREDICTION   Let us consider a regressor $\mathbf{R}$. For a test voxel $\mathbf{v}_k$ and its associated image features from an image I, it gives a prediction of the bounding box $\mathbf{b}_{o,k}$ of an organ o. Notice that, as in the work of Criminisi et al. (2013), the label regressed is actually the relative distance

of the voting voxel to the extremal vertices of the bounding box. For the sake of comprehension we simplify the notation and assume that the regressor predicts the bounding box parameters. If the regressor is a RF, one can store the distributions of the training labels in the leaves of the trees. During the testing phase this allows us to compute the posterior probability $p(\mathbf{b}_{o,k}|I, \mathbf{v}_k)$ that we denote $\alpha_{o,k}$. It gives a confidence score about the vote of v_k for the location of the bounding box $\mathbf{b}_{o,k}$ of the organ o. We can thus write $[\mathbf{b}_{o,k}, \alpha_{o,k}] = R(\mathbf{v}_k, I)$.

CONFIDENCE MAP CONSTRUCTION   If we want to consider the localization at a voxel level in the image, either the voxel $\mathbf{x}$ belongs to the box or it does not. The probability $p(\mathbf{x} \in o|\mathbf{b}_{o,k})$ of a voxel to belong to the organ o given the predicted bounding box $\mathbf{b}_{o,k}$ then follows a uniform distribution. This approximation is rough, especially for organs with non convex shapes as the liver for example. Thus we propose to enhance this hypothesis by using a shape prior through the use of probabilistic atlases A_o of organ o. We assume that $p(\mathbf{x} \in o|\mathbf{b}_{o,k}) = A_o \circ G_{o,k}(\mathbf{x})$ where $G_{o,k}$ is the transformation which translates and resizes (anisotropically) the bounding box $\mathbf{b}_{o,k}$ to fit into $[-1, 1]^3$.

We want to know the joint probability of localizing an organ and that a voxel of the image belongs to this organ, which means that we want to compute the probability of having both a predicted bounding box $\mathbf{b}_{o,k}$ and $\mathbf{x} \in o$, given an image I and a test voxel $\mathbf{v}_k$:

$$p(\mathbf{x} \in o, \mathbf{b}_{o,k}|I, \mathbf{v}_k) = p(\mathbf{x} \in o|\mathbf{b}_{o,k}, I, \mathbf{v}_k).p(\mathbf{b}_{o,k}|I, \mathbf{v}_k) \quad (3.4)$$

As to deduce the probability of $(\mathbf{x} \in o)$, knowing the bounding box $\mathbf{b}_{o,k}$ or knowing the image I and the voting voxel v_k are equivalent ($p(\mathbf{x} \in o|\mathbf{b}_{o,k}, I, \mathbf{v}_k) = p(\mathbf{x} \in o|\mathbf{b}_{o,k})$), we can consider that the random variables $(\mathbf{x} \in o|\mathbf{b}_{o,k})$ and $(I, \mathbf{v}_k|\mathbf{b}_{o,k})$ are independent conditionally to $\mathbf{b}_{o,k}$. We can write:

$$\begin{aligned} p(\mathbf{x} \in o, \mathbf{b}_{o,k}|I, \mathbf{v}_k) &= p(\mathbf{x} \in o|\mathbf{b}_{o,k}).p(\mathbf{b}_{o,k}|I, \mathbf{v}_k) \\ &= \alpha_{o,k} \, . \, A_o \circ G_{o,k}(\mathbf{x}) \end{aligned} \quad (3.5)$$

In the work of Criminisi et al. (2013) and Cuingnet et al. (2012), the authors assume that the localization results from a measure over the predictions given by multiple test voxels of the image (maximum a posteriori, average or median). Here we consider the entire set of K predictions and model the final confidence score F in each voxel of the image I, defining the confidence maps of organ o as follows:

$$F(\mathbf{x} \in o, \{\mathbf{b}_{o,k}\}_{k \in [\![1,K]\!]}|I) = C_o(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^{K} \alpha_{o,k}.A_o \circ G_{o,k}(\mathbf{x}) \quad (3.6)$$
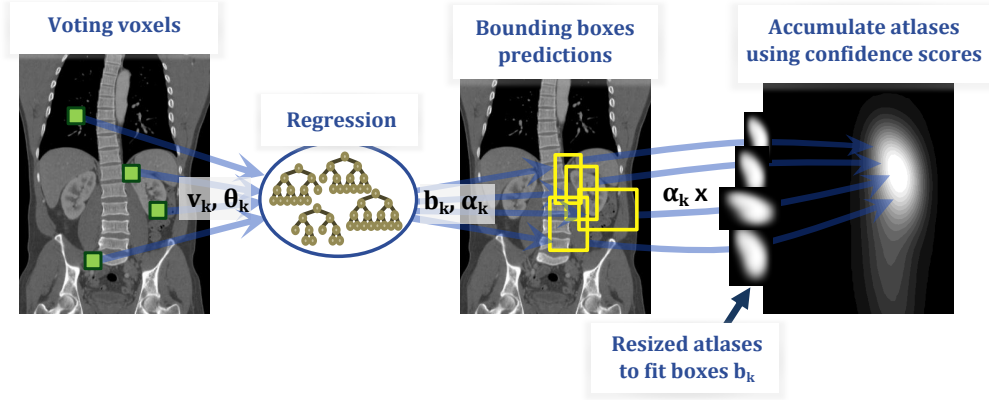
Figure 3.4: Outline of the confidence map computation, here illustrated on the right kidney localization only (for the sake of understanding).

where C_o is the confidence map of organ o. We then normalize the map C_o by its maximum to get values in the range $[0, 1]$. For a voxel $\mathbf{x}$ of the image I, $C_o(\mathbf{x})$ can be seen as the chance of $\mathbf{x}$ to belong to the organ o.

Some examples of confidence maps are given in Figure 3.3b, Figure 3.3c, Figure 3.3e and Figure 3.3f. Figure 3.3b and Figure 3.3e show that the maps capture the ambiguity of the vote distribution, as we observe that some voxels were correctly voting for the box upper wall position, an information that the median was not able to capture.

FAST CONFIDENCE MAP IMPLEMENTATION   The confidence map C for a given organ is actually built by translating and scaling the organ probabilistic atlas A according to each vote k and accumulating the result in C with weight α_k. The map C gives a confidence score about the presence of the organ at a given location in the image. The pseudo-code is given in Algorithm 1.

The computation of these maps may be expensive. Therefore we propose a fast implementation which considerably reduces the computation time while not degrading significantly the accuracy (see Section 3.5). The idea is to uniformly discretize the space of predicted bounding boxes dimensions (over each dimension x, y and z). With K votes, each corresponding to a bounding box location and size $\mathbf{d}_k$, it is easy to compute the corresponding centers $\mathbf{c}_k$. We denote $\mathbf{d}_{min}$ and $\mathbf{d}_{max}$ the minimum and maximum box sizes over the K predictions. Assume we want to discretize the space of these box sizes ensemble. We do it the simplest way by defining the jth (where $j \in [\![1, J]\!]$) discretized size $\hat{\mathbf{d}}_j \in \mathbb{R}^3$ as:

$$\hat{\mathbf{d}}_j = \frac{j}{J}(\mathbf{d}_{max} - \mathbf{d}_{min})$$

$C \leftarrow 0;$
for $k \leftarrow 1$ **to** K **do**
$\quad [\boldsymbol{b}_{min}, \boldsymbol{b}_{max}, \alpha] \leftarrow \boldsymbol{R}_\theta[v_k];$
$\quad X_b \leftarrow \texttt{getVoxelsInBox}(\boldsymbol{b}_{min}, \boldsymbol{b}_{max});$
$\quad \boldsymbol{c} \leftarrow \frac{1}{2}(\boldsymbol{b}_{min} + \boldsymbol{b}_{max});$
$\quad \boldsymbol{d} \leftarrow \boldsymbol{b}_{max} - \boldsymbol{b}_{min} ;$
$\quad A_t \leftarrow \texttt{rescaleAtlas}(A, \boldsymbol{d}) ;$
$\quad$ **foreach** $\boldsymbol{x} \in X_b$ **do**
$\quad\quad C[\boldsymbol{x}] \leftarrow C[\boldsymbol{x}] + \frac{\alpha}{K} A_t(\boldsymbol{x} - \boldsymbol{c});$

Algorithm 1: Confidence map computation. The variables $\mathbf{b}_{min}$, $\mathbf{b}_{max}$, $\mathbf{c}$ (the center), $\mathbf{d}$ (the diagonal) and $\mathbf{x}$ are in $\mathbb{R}^3$. The function $\texttt{getVoxelsInBox}$ retrieves the set of voxels X_b included in the box defined with parameters $\mathbf{b}_{min}$ and $\mathbf{b}_{max}$. The function $\texttt{rescaleAtlas}$ rescales atlas A anisotropically to fit dimensions d.

Each size dimension is discretized the same way. Then for each possible discretized size of box $\hat{\mathbf{d}}_j$, we create a volume C_j of the same size as the image (it is however possible to work at a lower resolution) initialized to 0. For each prediction we set the box centers to their confidence score value in the volume C_j. We define the probabilistic atlas A_j which has been anisotropically rescaled to the jth sample size. This atlas is then convolved with the volume C_j. The final map C is computed as the sum of each C_j and normalized by the maximum to get the final probability map. The pseudo-code is given in Algorithm 2. With a uniform discretization on 27 boxes sizes (3 per spatial dimension, which is good trade-off between speed and precision), the computation is about 30 times faster for a limited loss of accuracy (see Section 3.5).

3.2.5 *Cascaded regression for multi-organ localization*

Our approach consists of two steps in a coarse-to-fine fashion. A first regressor, aiming at capturing the spatial relationships between organs, is learned using global information: voxels of the whole image vote for the positions of all the organs simultaneously (all bounding boxes are regressed with the same forest, as in Criminisi et al. (2013)). Then, new regressors, dedicated to each organ, are learned using more local information. We may wonder why not to localize directly each organ independently like in the second step. Our experiments showed that the global step helps increasing the robustness of the approach as it encodes the global anatomy prior. The benefit of the cascade approach has also already been shown in (Cuingnet et al., 2012). The authors localize the two kidneys by using a cascade of three RF regressions. The first regression predicts the two bounding boxes simultaneously. In a second step, two regressors refine

```
C ← 0, d_max ← 0, d_min ← max ;
for k ← 1 to K do
    [b_min, b_max, α_k] ← R_θ[v_k];
    c_k ← ½(b_min + b_max);
    d_k ← b_max − b_min ;
    d_max ← max(d_max, d_k);
    d_min ← min(d_min, d_k);
d̂ ← quantize(d_max, d_min, J) ;
foreach d̂_j ∈ d̂ do
    indices ← getIndicesOfVotes(d̂_j, d̂);
    C_j ← 0;
    foreach ind ∈ indices do
        C_j[c_ind] ← C_j[c_ind] + α_ind ;
    A_j ← rescaleAtlas(A, d̂_j) ;
    C_j ← convolve(C_j, A_j) ;
    C ← C + C_j ;
```

Algorithm 2: Fast confidence map computation. The variables $\mathbf{b}_{min}$, $\mathbf{b}_{max}$, $\mathbf{d}_{min}$, $\mathbf{d}_{max}$, $\mathbf{c}_k$ (the center), $\mathbf{d}_k$ and $\mathbf{x}$ are in $\mathbb{R}^3$. The variable $\hat{\mathbf{d}}$ denotes the set of discretized size samples. The function getBoxDimensions computes the size of the box defined with parameters $\mathbf{b}_{min}$ and $\mathbf{b}_{max}$. The function quantize computes the different sample sizes from the range of boxes sizes. The function getIndicesOfVotes retrieves the indices of votes corresponding to the sample size $\hat{d}_j$. The function rescaleAtlas rescales atlas A anisotropically to fit to dimensions d.

the left and right kidneys bounding box centers independently. For these last two regressors, the voting voxels (at training and testing time) are selected in a ROI around the bounding box centers predicted by the first regressor.

Here we propose to introduce the use of confidence maps for refining the votes in the cascade of regressors. Moreover, instead of only refining the bounding box centers in the second step, we refine the prediction of all the bounding box parameters. The selection of voxels which vote in the local step may benefit of the information given by the confidence maps, that is to say the vote distributions and the shape prior. This method can be applied with any multi-variate regressor. In Section 3.3 we detail how to implement this method with the RF as a regressor.

3.2.5.1 *Global step*

In the first step, a random subset of K_g voxels $\{\mathbf{v}_k\}_{k \in [\![1, K_g]\!]}$ of the image I vote for the bounding boxes parameters $\{\mathbf{b}_{k,o}\}_{k \in [\![1, K_g]\!]}$ of all organs o. By regressing all the parameters jointly note that the relationships between the organs are

implicitly embedded during learning. Votes are performed according to long-range features computed from the image (see Section 3.3.1 for details on the features). These features are chosen to encapsulate global information from the image. Then the confidence map C_o for each organ o is computed using Algorithm 1 or 2, given the probabilistic atlas A_o. Figure 3.3b and Figure 3.3e show some examples of results after this global step for the liver and the right kidney localization. In these figures we compare the localization with the confidence maps and the one given by the bounding boxes (computed from the median over the predictions). We see that capturing the distribution of the votes can be more informative than just a median.

3.2.5.2   *Local step*

The second step aims at improving the previous localization. Each organ o is re-localized individually using a region of interest (ROI) selected thanks to the previous global localization and the resulting confidence maps. To find the ROI we compute the binary mask B_o from the map C_o thresholded at a value t_g (see Section 3.4 for numerical values). Then we select a random subset of K_l voxels $\{\mathbf{v}_k\}_{k\in[\![1,K_l]\!]}$ such that each voxel is in the ROI ($B_o(\mathbf{v}_k) = 1$). Each voxel $\mathbf{v}_k$ votes for the bounding box of organ o using a regressor specifically trained for this organ. Contrarily to the previous step this regressor is now learned using shorter-range features (see Section 3.4) and computed in the vicinity of the organ o thanks to the ROI given by the confidence map. This gives more importance to local information. We then use the votes to compute new and more accurate confidence maps C_o' for each organ. Figure 3.3c and Figure 3.3f show the benefit of adding this local step after the global step: the confidence maps are more focused on each organ and less spread around it.

3.3   AN APPLICATION: THE LOCALIZATION OF ABDOMINAL ORGANS

In the following sections we show how to apply the method for the localization of abdominal organs in CT images. We propose to localize the liver, the two kidneys, the spleen, the stomach and the gallbladder. In this section we detail the method pipeline, the material and the implementation. In Section 3.4 we show a study and optimization of the different testing and training parameters. Then in Section 3.5 we present an evaluation of the localization results.

3.3.1   *Using the regression forest as a regressor*

To regress the parameters $\mathbf{b}_o$ of each organ o (see notations in Section 3.2.1) we choose to use the RF as in (Criminisi et al., 2013). The authors showed that the RF regression is well suited to rough localization especially in terms of accuracy
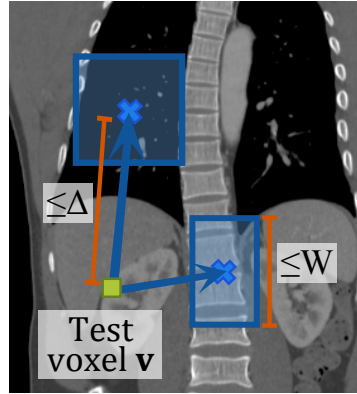
Figure 3.5: Example of 1D features computed at each tree node: difference of mean intensities in two patches of random size and location in a certain range.

and speed. We are able to reinforce its robustness and accuracy using our global-to-local approach with confidence maps.

We use the same kind of features as in Criminisi et al. (2013), similar to those introduced in Gall and Lempitsky (2009); Shotton et al. (2009). These contextual features are defined as the difference of mean intensities in two 3D patches of random size and random location in a given range. We define Δ as the maximum offset in each dimension of the patches to the test voxel and W as the maximum width size of the 3D patches. Figure 3.5 shows an example of this type of features. Notice that the patches may be reduced to a single voxel or an empty box, thus giving unary features. For the global step of the algorithm, long-range features (high values of Δ and W) are used (they are computed from the image after Gaussian smoothing, with kernel width of 3 mm). For the local step, more local features (shorter values of Δ and W) are used. In each tree leaf the multivariate Gaussian distributions of the parameters to regress. This statistical information is used to compute the confidence scores of each vote are stored.

The first global forest is learned from a random subset of voxels from images of the training set. After learning this forest we proceed to the localization on the same training images and we compute the confidence maps associated with this result. Notice that in practice the votes with low confidence scores are discarded. To compute the confidence maps we only keep the best votes according to the confidence score (if N is the number of votes, $K \leq N$ is the number of votes that we keep). Then each local regression forest is learned from training voxels in the vicinity of the corresponding organ thanks to the confidence maps (hypothesis demonstrated right hereafter). After the forest of the global step is trained, we use it to test the training images. The resulting confidence maps are then thresholded to select the training voxels that are used to train the local forests. The testing phase proceeds exactly the same way.

3.3.2 *Material and implementation*

The database is composed of 130 3D CT images coming from 112 patients with diverse medical conditions (healthy and pathological subjects, no organ missing). It includes volumes with varied fields of view, body shapes, resolution and use or not of contrast agents. Slices and inter-slices resolutions range from 0.5 to 1 mm and from 0.5 to 3 mm, respectively. For both training and testing of each of the following experiments we work at a resolution of 3 mm. All the organs have been manually segmented in these 130 volumes. The dataset is split randomly into 50 and 80 volumes for training and testing, respectively. Our method is implemented in C++ and running times are given for a machine with two 2.3 GHz cores and 8 Go RAM[1].

3.4 PARAMETERS STUDY

To reach the best performance and analyze each aspect of the algorithm, we perform an extensive greedy optimization of the algorithm parameters. Random forests can become very complex to analyze. There are many parameters to take into account and each of them may be more or less linked to the others. Moreover considering the computation time (learning one tree with maximum depth 12 takes about 2 minutes), we choose to use a greedy and sequential parameters optimization.

3.4.1 *Protocol*

For each parameter we perform a 5-fold cross-validation on the training set (40 for training and 10 for testing). For computation time and memory reasons, the accuracy of the algorithm is measured as the mean over each organ of the mean distances of the predicted boxes to the ground truth bounding boxes. We assume this would give nearly equivalent results if we looked at the accuracy of the confidence maps. The predicted boxes are simply computed by taking the median of the votes over the best votes K_g and K_l. For each test image we run the algorithm 5 times to be able to measure and reduce the variability induced by the randomness of the method. For learning the forests we perform subbagging (Andonova et al., 2002) with uniform random draw with replacement. This means that we learn each tree on a random subset of the training set the size of which is inferior to the training set size (in the original bagging method, the bootstrap has the same size as the whole training set). If subbagging is interesting to reduce the training computation time, we will also show that it can also make the algorithm more robust. The node feature selection is

---

1 Due to industrial confidentiality concerns we are not able to release our own code. However several packages and libraries on RF can be found on the web (see Section 2.2.1.6).

also randomized (we compute only 30 random features at each node). For each experiment we use 7 trees, as this gives a good trade-off between low variance and computation time (also shown in later experiments, see Figure 3.13). Before each learning we de-correlate the data with a whitening transform. This is a traditional statistical method which forces the variance of the training data along each dimension to become equal to one, thus preventing to give more importance to one or another dimension during learning.

We first initialize every parameter arbitrarily with values already giving reasonable results. These values were quite easy to find, taking into account the literature on the subject and a few experiments. Then we optimize each parameter one-by-one by grid-search and we replace its value by the optimized one.

PARAMETERS TO BE OPTIMIZED   We use the index g for the global forest and l for the local forests.

For the global forest we study the following parameters (in this order):

- T_g,S_g: the maximum tree depth and the number of voxels per image used for training respectively (studied jointly),

- B_g: the bootstrap size (percentage of training samples used to learn one tree),

- Δ_g,W_g: feature patches maximum offset and maximum width (studied jointly),

- N_g,K_g: the number of test voxels per image and the percentage of test voxels kept for the final prediction computation respectively (studied jointly).

For the local forests we study the same parameters: T_l, S_l, Q_l, Δ_l, W_l, N_l and K_l.

We also study the parameter t_{cmap} which is the threshold of the confidence map computed after global step for selecting the voxels used for learning or testing the local forest.

3.4.2 *Off-line training results*

All the final results of the parameters greedy optimization are given in Table 3.1.

The graphics of Figure 3.6 and Figure 3.7 show the optimization of the parameters of the global step. Each of them gives the testing cross validation accuracy computed for the different parameters values in the same order as they were optimized. Figure 3.6a shows that the accuracy is increasing with the maximum tree depth until $T_g = 14$, and for this value the optimal number of training voxels per image is $S_g = 25000$. The maximum tree depth and the number of training voxels per image being fixed, we can look at the optimal bootstrap

Table 3.1: Regression forest parameters.

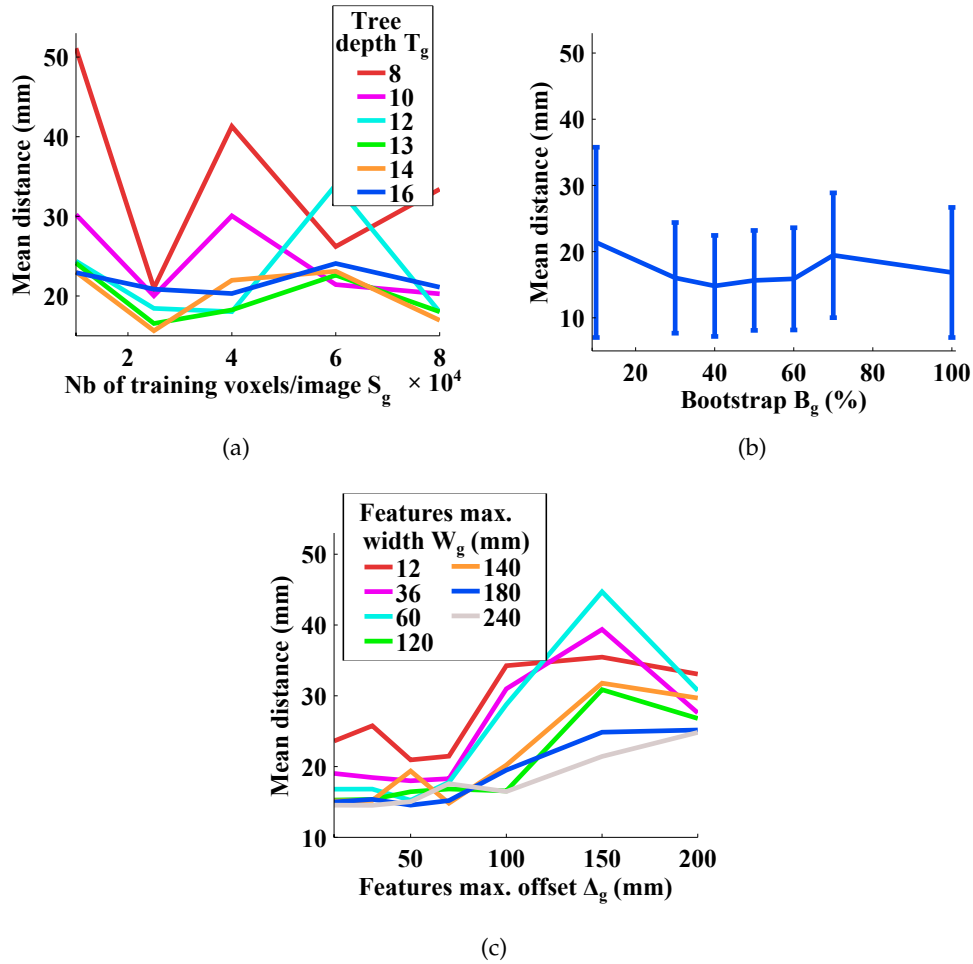| Parameters	Maximum tree depth	Nb training voxels/image	Bootstrap size	Range of features (mm)	Confidence map threshold	Nb of votes	Nb of best votes
Global step	$T_g = 14$	$S_g = 25000$	$B_g = 40\%$	$\Delta_g = 50, W_g = 180$		$N_g = 60000$	$K_g = 10\%$
Local step	$T_l = 14$	$S_l = 10000$	$B_l = 50\%$	$\Delta_l = 40, W_l = 40$	$t_{cmap} = 40\%$	$N_l = 10000$	$K_l = 10\%$

Figure 3.6: Optimization results of parameters corresponding to the global RF. Performances are given for (a) the number of training voxels per image S_g, each curve corresponding to a given maximum tree depth, (b) the bootstrap B_g with error bars corresponding to the standard deviation of the localization accuracy, (c) the features parameter Δ_g and W_g.
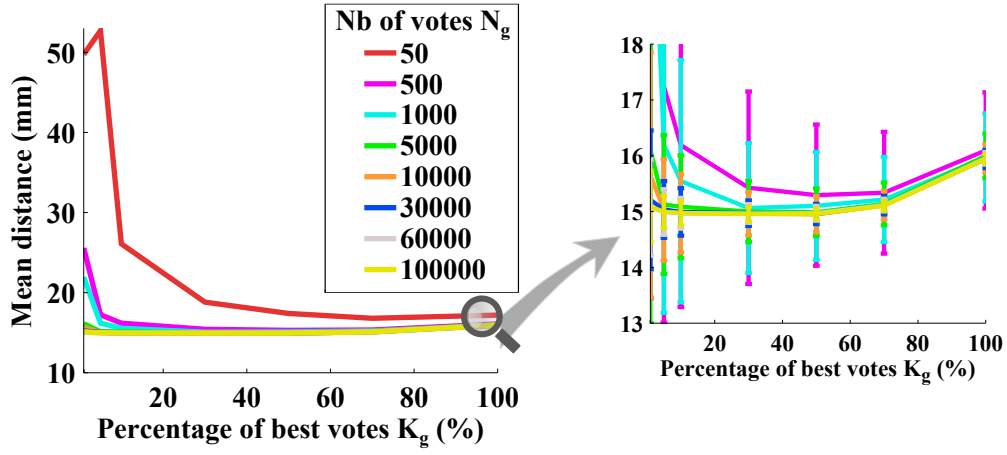
Figure 3.7: The left plot corresponds to the performance in function of the percentage of selected test voxels K_g, each curve corresponding to different values of the number of test voxels N_g, and the right plot is a zoom with error bars related to the standard deviation of the results over the multiple tests.

value. In Figure 3.6b we see that an optimal value is around $B_g = 40\%$ although close values give similar accuracy. We notice that a bootstrap of 100% (equivalent to original bagging) gives the same result but with higher variance (mean variance over all the organs). Here the subbagging helps to achieve more robustness. In Figure 3.6c we can see the influence of the feature parameters. It shows the algorithm accuracy in function of the maximum offset value Δ_g of the 3D patches. Each curve is given for different values of the patches maximum width W_g. We see that big patches lead to better accuracy, reaching a plateau at about $W_g = 60$mm of maximum width. We also notice that increasing values of Δ_g degrades the accuracy. Some experiments show that this can be explained by an increasing number of patches that fall outside of the image domain, the features being then undetermined. In our implementation we do not discard labels with undefined features. At a node, labels with undefined features just go right and left in the tree, as these labels may have defined features at other nodes of the tree and then be useful. While former results were related to the training part, the next figure shows results on the testing part. For this test we keep the best forest given after the last parameter tuning. In Figure 3.7 we show the accuracy compared to the percentage of best test voxels K_g. Each curve corresponds to different numbers of test voxels per image N_g. We notice that from $N_g = 1000$ we reach almost the same accuracy. However, looking at the right plot which is a zoomed version showing the standard deviation over the multiple tests, we see that the standard deviation over the multiple tests decreases when the number of test voxels increases. With $N_g = 60000$ and $K_g = 10\%$ the results reach

a standard deviation below 0.2mm. We deemed such performance reasonable enough, as increasing this value would also increase the computation time.

The graphics of Figure 3.8 and Figure 3.9 show the optimization of parameters for the local RF (second step in the cascade). In particular Figure 3.8a shows that from a maximum tree depth of $T_l = 14$ the convergence is reached, and that a number of $S_l = 10000$ training voxels per image gives optimal results. Figure 3.8b shows a very good robustness with respect to the bootstrap value, and we set it to $B_l = 50\%$. Figure 3.8c shows the accuracy in function of the feature parameters. Notice that more local features than in the global forest are preferred. This confirms the interest of using a coarse-to-fine approach. As highlighted in (Zhou et al., 2010), such a framework mimics the human visual system in a certain way: global information is first used to define the context and then, while restricting the ROI, more local information is then required to refine the understanding analysis of this area. For the best performance, these features should have displacements up to $\Delta_l = 40$mm and patches widths up to $W_l = 40$mm. Finally in Figure 3.9 we show the results in function of the testing voxels for the local part. From $N_l = 1000$ we get similar accuracies, but looking at the right plot we set $N_l = 10000$ and $K_l = 10\%$, which corresponds to a standard deviation in test below 0.2mm. Finally, Figure 3.8d shows the performance result in function of the confidence map threshold t_{cmap}. This parameter actually controls the size of the region of interest in which the local step is performed. Setting the value to $t_{cmap} = 40\%$ gives the best results, but we notice that this parameter is very stable and values between 20% and 80% give similar results (this is also explained by the relatively small spread of the confidence maps, see Figure 3.3c and Figure 3.3f. This result demonstrates the interest of restraining the ROI in the cascade approach.

3.4.3 *Conclusion on parameters optimization*

Although very costly in terms of computation time (more than two full weeks in cumulative time), this study on parameters is very informative. First of all, it shows that the choice of parameters may be crucial for having a good algorithm accuracy. Fortunately, our results also show that there is no need to tune every parameter very precisely, since the results remain quite stable in certain range of parameter values. We prove then that our implementation is robust. This will be also demonstrated by the results presented next, showing that the same parameter optimization can be used for a varied and large set of images.

The experiments also showed the interest of using subbagging. While reducing the training computation time, it also reinforces the robustness of the algorithm. Finally this part also confirms the intuition that the range of features directly relates to the level of information we want to retrieve. Indeed the first global step for coarse localization requires features covering large parts of the
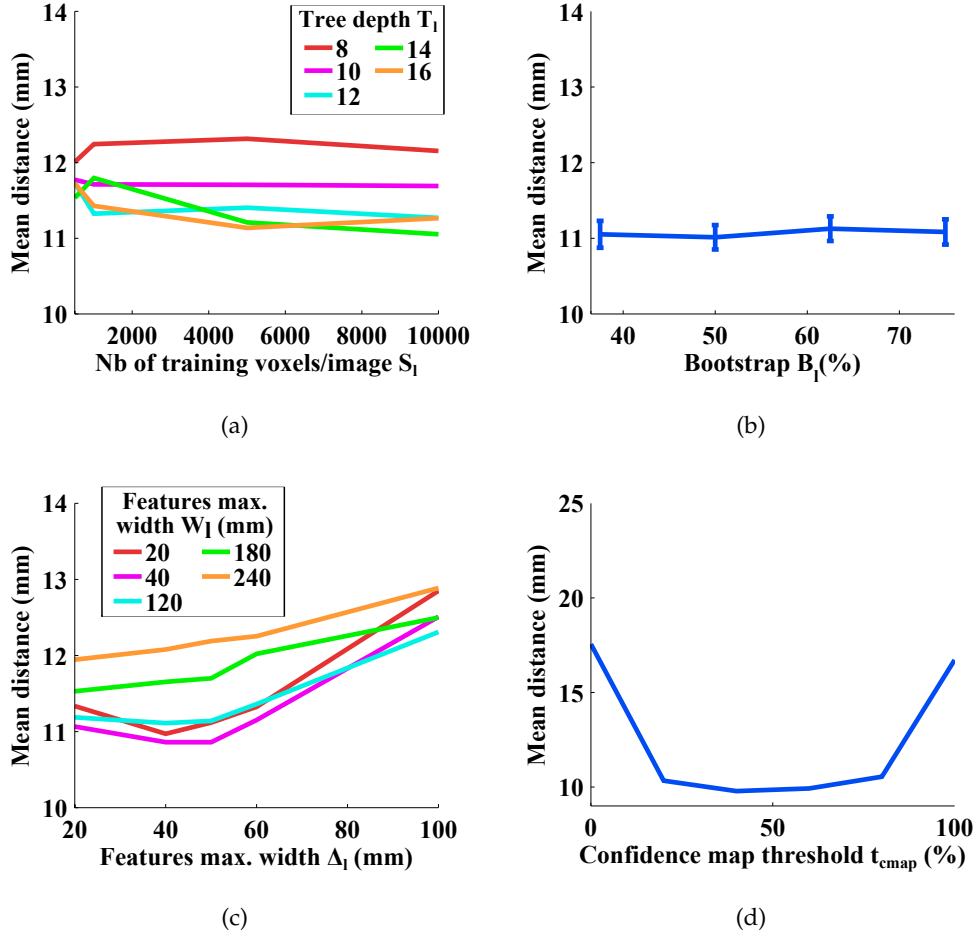
Figure 3.8: Optimization results of parameters corresponding to local RF. Performances are given for (a) the number of training voxels per image S_l, each curve corresponding to a given maximum tree depth, (b) the bootstrap B_l with error bars corresponding to the standard deviation over the localization accuracy, (c) the features parameter Δ_l and W_l, (d) the threshold of the confidence map for local forest voxel selection.
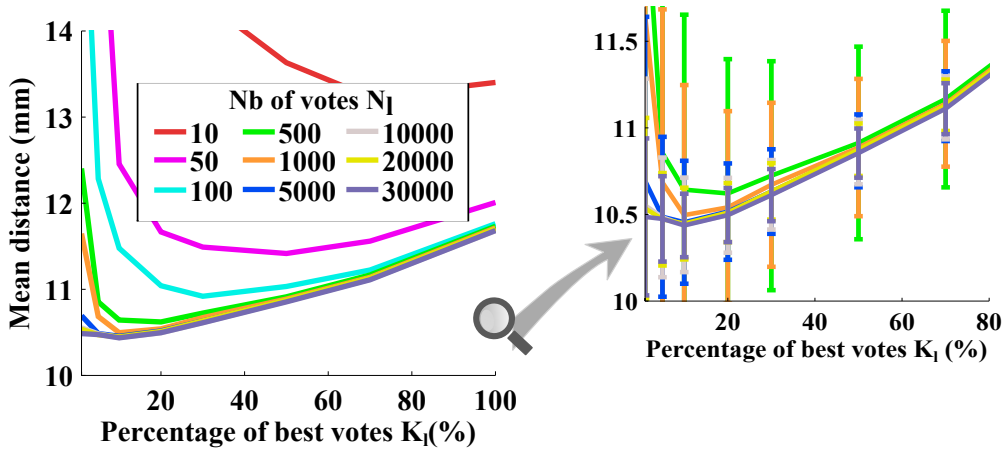
Figure 3.9: The left plot corresponds to the accuracy in function of the percentage of selected test voxels K_l, each curve corresponding to different values of the number of test voxels N_l, and the right plot is a zoom with error bars related to the standard deviation of the results over the multiple tests.

image, as if we had to look at the image from far away. On the other hand, more local features are needed in the local step in order to refine the previous global localization: the ROI is narrower and we look at the image more closely.

3.5 EVALUATION AND VALIDATION

In this section we evaluate our approach, taking into account the parameters optimization realized in the previous section. We perform here two types of experiments. The first one aims to show that the confidence maps can be used as a result in itself, while the second experiment demonstrates several aspects of our work: the interest of the cascade approach with a comparison to other state-of-the-art methods, the benefit of a parameter optimization, and finally the interest of using the confidence maps in the cascade for the voting voxels selection.

3.5.1 *Protocol of the experiments*

For the two following sets of experiments, we use the same database as the one described in Section 3.3.2. We learn the forests on the training set, following the conclusions of the parameters optimization of Section 3.4, and we perform our experiments on the testing set (composed of 80 volumes). For the first ex-

periment we consider the confidence maps as a result, which corresponds to our own original approach. The confidence maps are computed by using Algorithm 2. Our experiments showed a very small difference in accuracy compared to the naive version of Algorithm 1 (below 3%). This confirms the interest of our fast implementation. In the second experiment we want to see the influence of the use of the confidence maps in the cascade and compare the obtained results with those of existing methods. Therefore we use the median bounding boxes as a final result as in (Cuingnet et al., 2012).

3.5.2 *Processing times*

Implementation details are given in Section 3.3.2. Training the seven forests (one for the global step and six for the local step) and building the probabilistic atlases takes approximately 20 minutes. Testing a classical abdominal image (approximate size: 512 x 512 x 230 voxels) takes around 5 seconds. The confidence maps computation takes about half that time (one map computation takes around 200 ms).

3.5.3 *First experiment: evaluating the interest of confidence maps as a result*

The first objective of our evaluation is to show that the confidence maps can be used as a localization result itself, giving more consistent information than the bounding boxes alone. Confidence maps are computed at a 5 mm isotropic spacing. In Figure 3.10 are shown the best and the worst results for each organ according to the mean distance to ground truth. Moreover an exhaustive visualization of the results can be seen online[1].

On the first line of Table 3.2 are given the results of the mean distance (in millimeters) of the thresholded confidence maps contours (with threshold set to 50%) to the ground truth contours. The results statistics (median and standard deviation) confirm that our method is robust to the variety of test images. Figure 3.11 details these results by giving the distribution of mean distances to ground truth for each organ. Notice the low spread of the results and the very few number of outliers (no result below 40mm). These results confirm that our approach can be very useful in various contexts such as segmentation initialization or fast anatomy detection, added to the fact that it runs in about 5 seconds. This makes our approach adapted to clinical applications, especially as the code may still be optimized.

The goal of this first experiment is also to show that the confidence maps give much more information than a simple binary mask or contour. For that purpose we propose adapted evaluation measures taking into account the fuzziness of the maps. If C denotes a confidence map and B a binary mask of the organ

---

1 http://perso.telecom-paristech.fr/~gauriau/MOLoc_SupMat.html
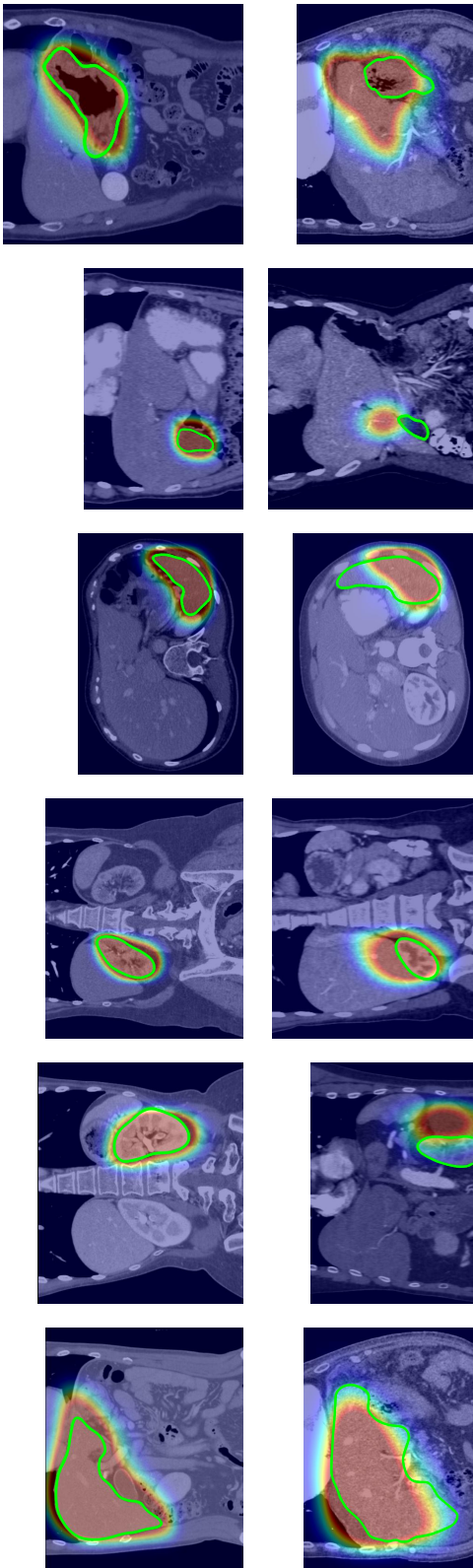
Figure 3.10: Best and worst results per organ (first and second line respectively) according to the mean distance to ground truth. Confidence maps are overlaid on the images and the green contour corresponds to the organ manual segmentation. From left to right, organs are considered in this order: liver, left kidney, right kidney, spleen, gallbladder and stomach.

59
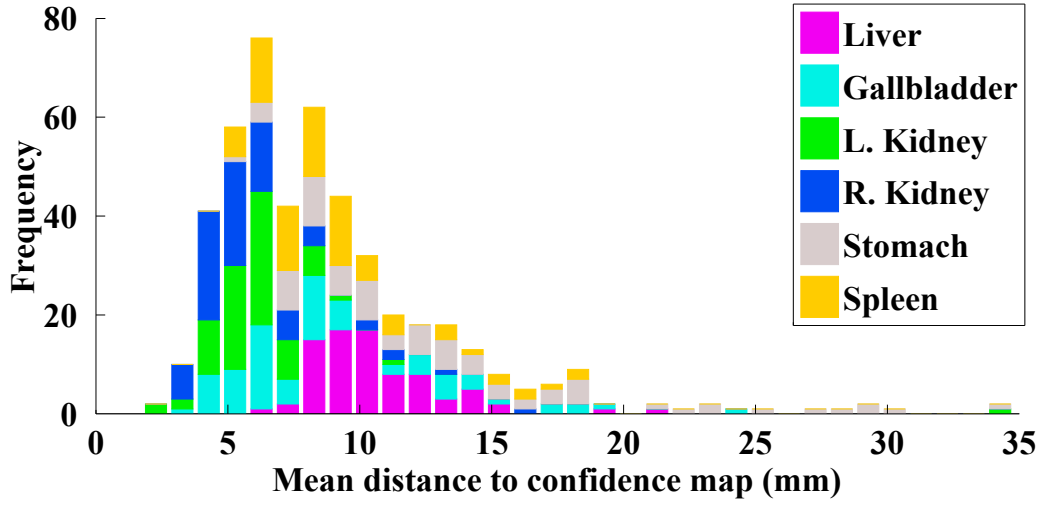
Figure 3.11: Histogram of results on confidence maps, frequencies are stacked over each other. Mean distances of ground truthes to thresholded confidence maps (50%) for each organ.
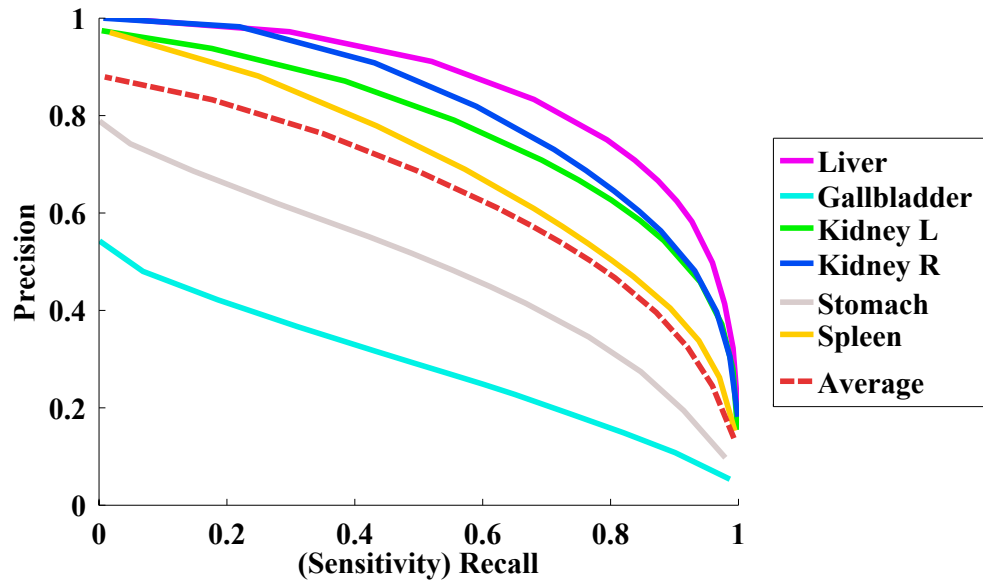
Figure 3.12: Receiver operating characteristic (ROC) curves showing how the precision and sensitivity change depending on the confidence map threshold.

ground truth, then the true positive values TP, the false negative values FN and the false positive values FP are defined as:

$$TP = \sum_{x \in \Omega} B(x)C(x)$$
$$FN = \sum_{x \in \Omega} B(x)(1 - C(x))$$
$$FP = \sum_{x \in \Omega} (1 - B(x))C(x)$$

Then following the definitions of the sensitivity (or recall) S, the precision P and the dice coefficient D:

$$S = \frac{TP}{TP + FN}, \ P = \frac{TP}{TP + FP}, \ D = \frac{2TP}{2TP + FP + FN} \quad (3.7)$$

we are able to propose weighted versions of these measures. The corresponding figures are reported in Table 3.2. In Figure 3.12 we also show the receiver operating characteristic (ROC) curves for varying confidence map thresholds. We note a good compromise between sensitivity and precision. In terms of accuracy, results are deemed as satisfying for localization purposes. Performances are lower for the stomach and the gallbladder which are challenging organs due to their shape and location variability. However the performance results given in Table 3.2, Figure 3.12 and Figure 3.11 show that they are still correctly detected, meaning the localization may not be perfect but still reasonably overlapping the ground truth.

3.5.4 *Second experiment: evaluating the cascade approach*

In this second experiment, we show the interest of using the confidence maps in the cascade approach and compare our results with existing methods. We evaluate the methods of Cuingnet et al. (2012) and Criminisi et al. (2013) on our database. We gathered all the implementation details that we could find in their articles. If we could not find any parameter detail we used the same as ours. For the three methods we work at the same resolution (3mm). The main differences between our implementation and the one of Criminisi et al. (2013) are: (i) they store histograms in the leaves, (ii) the final prediction is computed from the leaves with less uncertainty and by finding histogram maxima, (iii) the tree depth is 12, (iv) the minimum node size is 25. The differences between our implementation and the one of Cuingnet et al. (2012) are: (i) the voting voxels selection in the second step (ROI around the bounding box centers), (ii) the tree depth is 15, (iii) the minimum node size is 100, (iv) the bounding box centers only are re-estimated in the second regression step. In comparison our implementation has the following characteristics: (i) Gaussian distributions are stored in the leaves, (ii) the final prediction is the median over the votes with

Table 3.2: Results with confidence maps (5 mm isotropic spacing) with Algorithm 2 with mean distance (mm), dice (%), weighted dice (%), weighted sensitivity (%) and weighted precision (%) (mean±standard deviation (median)). Mean distance and weighted dice are computed given the confidence map thresholded to 50%.

|	Liver	L. Kidney	R. Kidney	Spleen	Gallbladder	Stomach	All organs	Time
Mean distance (mm)	10.6 ± 2(10)	6.3 ± 4(6)	5.8 ± 2(5)	9.1 ± 3(8)	8.8 ± 4(8)	13.9 ± 7(12)	9.0 ± 2(8)	
Dice (%)	74.9 ± 5(76)	69.6 ± 11(72)	70.8 ± 9(73)	60.3 ± 12(62)	33.1 ± 18(34)	48.3 ± 14(50)	59.5 ± 7(62)	
W. dice (%)	74.9 ± 10(79)	75.5 ± 8(78)	77.4 ± 6(79)	72.7 ± 8(74)	51.7 ± 16(56)	58.1 ± 10(60)	68.4 ± 14(73)	5s
W. sensitivity (%)	78.3 ± 6(80)	69.5 ± 8(71)	70.9 ± 8(72)	69.3 ± 11(71)	51.7 ± 17(54)	52.7 ± 10(54)	65.4 ± 14(69)	
W. precision (%)	67.2 ± 19.8(77)	80.3 ± 11(83)	83.2 ± 8(87)	74.6 ± 14(76)	54.6 ± 18(58)	62.0 ± 17(62)	70.3 ± 18(75)	
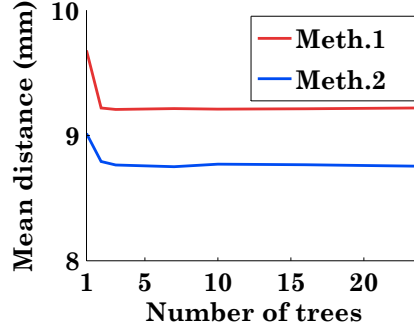
Figure 3.13: Results on bounding boxes for methods 1 and 2 and for various numbers of trees (same number of trees in the global and local step).

less uncertainty, (iii) the tree depth is 14, (iv) the voting voxels of the second step are selected thanks to the confidence maps, (v) all the bounding box parameters are re-estimated in the cascade, (vi) the minimum node size is 50.

The results are computed considering the bounding box wall distances (our results are computed by taking the median bounding box over the K_l best predictions). The figures are given in Table 3.3.

First of all, the performances after the global step (third line) show the benefit of the parameter optimization, as these results reach better accuracy than the results presented by Criminisi et al. (2013). Then we propose to compare the way of selecting the test voxels for the second local step:

- method 1: from the predicted boxes after the global step,

- method 2: from the thresholded confidence map (using Algorithm 2).

The corresponding results are given in the last two lines of Table 3.3. The results after the method 1 first show the benefit of the cascade approach (additional iterations did not show significant improvements of the results), compared to the single step one. Figure 3.14 confirms visually the benefit of the cascade approach. The results after the method 2 show the difference between our method and the original cascade approach. Significant improvements are especially noticed for the liver (gain of 1mm, about 10% improvement) and the spleen (gain of 0.8mm, about 10% improvement) whose shapes are very far from the bounding box approximation. This confirms the benefit of the shape prior for such organs. Moreover, a comparison of our results to those of Cuingnet et al. (2012) for the kidneys shows that the cascade approach is scalable and the increasing number of organs does not degrade the performances. Figure 3.13 shows the results of method 1 and method 2 for a varying number of trees in the forest. We observe a very fast convergence of the accuracy, forests with 3 trees already giving very good results.

Table 3.3: Box walls mean distances per organ (mean distance (mm) ± standard deviation (median)), per method and per experiment. In purple are the best figures. The results are given for the same database (see Section 3.3.2).

Method	Liver	L. Kidney	R. Kidney	Spleen	Gallbladder	Stomach	All organs	Time(~)
Criminisi et al. (2013)	14.0 ± 5(14)	12.3 ± 7(11)	13.2 ± 6(12)	14.2 ± 6(13)	15.5 ± 8(14)	16.8 ± 16(16)	14.4 ± 7(13)	0.5s
Cuingnet et al. (2012)	12.2 ± 4(12)	6.8 ± 6(6)	6.4 ± 4(5)	9.0 ± 5(8)	11.4 ± 8(10)	14.2 ± 7(13)	10.0 ± 7(9)	3s
Our method								
After global step	13.5 ± 4(13)	11.3 ± 6(10)	12.8 ± 5(12)	13.0 ± 5(13)	15.1 ± 6(14)	15.1 ± 6(14)	13.5 ± 6(13)	0.5s
After local step, method 1	11.7 ± 4(11)	5.8 ± 4(5)	6.2 ± 3(5)	8.7 ± 4(8)	9.5 ± 4(9)	13.4 ± 6(13)	9.2 ± 5(8)	2s
After local step, method 2	**10.7 ± 4(10)**	**5.5 ± 4(5)**	**5.6 ± 3(5)**	**7.9 ± 4(7)**	**9.5 ± 4(8)**	**13.2 ± 5(13)**	**8.8 ± 5(8)**	3.2s
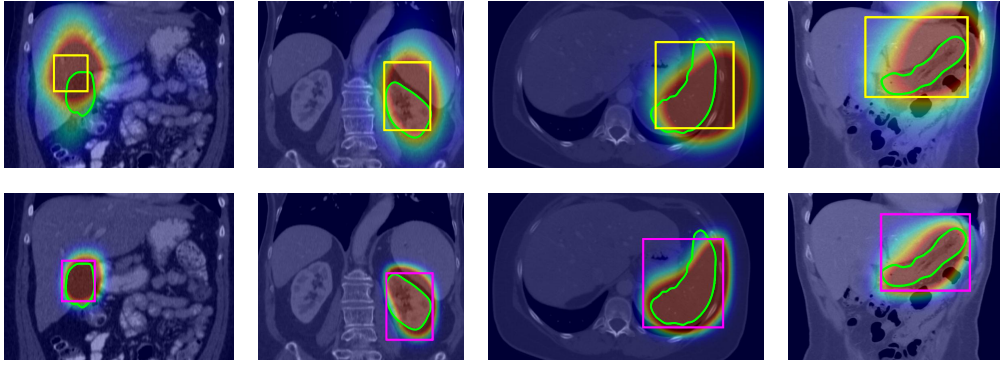
Figure 3.14: Results after the global step (first row) and the local step (second row). The confidence maps and the median boxes are overlaid on the images, the green contour corresponds to the organ manual segmentation. From left to right, organs are in this order: gallbladder, left kidney, spleen and stomach.

3.5.5 *Conclusion*

This set of experiments highlights the interest of approach on different aspects:

- the quantitative evaluation, supported by the comparison to two existing recent state-of-the-art approaches, revealed both the accuracy and robustness of the method,

- the estimated running times show its speed,

- the relative simplicity of the approach allows us to perform a parameter optimization enabling us to take the best of the approach,

- the method does not require a large number of training images to be efficient.

3.6 CONCLUSION

3.6.1 *Contributions*

In this chapter we proposed an accurate, robust and fast method for the localization of multiple organs. We extended the idea of cascade of RF regressors while introducing the concept of confidence map, which models the vote distributions with global shape prior conditioning. We showed that the confidence map, with a proposed fast implementation, can enhance the consistency and accuracy of multi-organ localization for a limited computational overhead. It is a generic tool with promising potential, which can be used with any type of regressor and which is adaptable to different applications and modalities (e.g.

CT, MRI as shown in Appendix B, US as shown in Appendix C). Moreover its fuzziness property may be useful in many types of clinical applications, such as segmentation (for initialization or image-driven forces computation as we show in Section 4.5 and in Appendix B) or visualization (to target the structures of interest for 3D rendering) for instance.

This work also focused on parameters optimization and features understanding. The use of RF is here relevant as there is no need of a fine-tuning of the parameters. On the other hand, a coarse optimization of the parameters helps improving the localization results. Our parameter study also highlights the link between the image information level and the range of features.

3.6.2 *Limits and perspectives*

First, our method does not handle the missing organ case. Even if an organ is missing, the algorithm will give a localization prediction. In Figure 3.15 we present an example of localization in a CT image with a missing right kidney. The result is probable (the kidney is located below the liver) even if the organ is not present. As mentioned in (Criminisi et al., 2013), the organ could be declared present if the confidence of the prediction is above a given threshold. Although they mention it, they do not present an evaluation or validation. Indeed, missing organs remain relatively rare. Therefore, the training and testing databases should be very large to be statistically representative of this characteristic. This may be difficult to achieve in practice. Another way of solving the missing organ problem could be to add a checking procedure after the final localization. This one could be based on prior knowledge of the organs appearances. This would require to model the organ's appearance and to find a similarity measure to be able to compare and take a decision. Methods used in anomaly detection could be investigated (Chandola et al., 2009).

Another limit of our approach relates to the assumption that the images are correctly oriented. In practice, for a MRI or CT acquisition, the patient lays on the table always in the same orientation. For US acquisitions however, the probe orientation depends on the clinician. As a result, the organ of interest can have varying orientation in the image. This problem can be easily solved by adding rotation parameters to the organ ROI representation model, as in the work of Zhou and Comaniciu (2007); Dollár et al. (2010). However, this should be done cautiously, as this greatly increases the parameterization complexity and thereby the complexity of the regression problem. This can be compensated with a larger training database able to represent the existing variability.

Our framework also does not guarantee the shape consistency of the confidence maps (i. e. not too far from the original organ shape). In practice, as the local step gives more confident results and the votes are done close to the organ, the final appearance of the map is not too far from the original organ
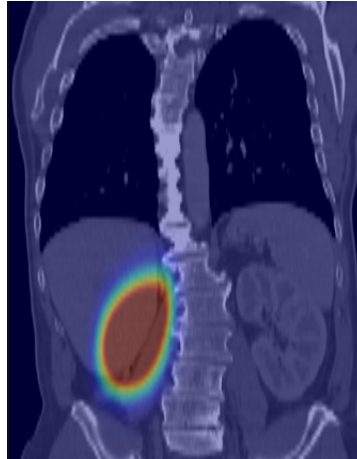
Figure 3.15: Example of localization in the case of a missing right kidney.

shapes. Nevertheless this could be ensured by modifying the best vote selection. For instance, we could perform a spatial mean shift procedure (Fukunaga and Hostetler, 1975) on the vote ensemble and keep the bigger cluster.

Other variants of our method could also be thought of. The shape consistency of the confidence maps may probably be improved with the use of multiple probabilistic atlases per organ.

However, all the improvements mentioned above have a cost, whether it be related to computation, memory, database enlargement or algorithm complexity increase. We believe that keeping this method simple and with a low complexity gives already reasonable results and makes it a fast and powerful component, that can be easily integrated into more advanced workflows. We demonstrate its full potential in the following chapters by integrating it in an automatic segmentation framework.

4

MULTI-ORGAN SEGMENTATION WITH TEMPLATE DEFORMATION

Abstract

This chapter presents a new variational formulation for the segmentation of multiple organs in 2D or 3D. The approach extends the original template deformation framework of Mory et al. (2012) dedicated to single organ segmentation. First we formulate a new energy functional under constraints guaranteeing non-overlapping of the different organs. Qualitative examples in 2D show the interest of our formulation under constraint. Secondly we propose and analyze a generic image-driven term taking into account both intensities and edge information. Thirdly, we propose to link our previous work on localization with this new segmentation method to get a complete automatic segmentation workflow. Experiments on the segmentation of abdominal organs in CT volumes show a good tradeoff between accuracy and speed (about 30s) and demonstrate the interest of our approach for clinical applications.

A condensed version of this work has been published in Gauriau et al. (2015a).

Contents

4.1 INTRODUCTION

In Chapter 1 we mentioned the different challenges and requirements of segmentation methods in medical images. Robustness, accuracy and speed are the keys of a good segmentation solution. Designing an algorithm fulfilling all these constraints is however not straightforward. The search for robustness and accuracy often leads to more complex and less efficient solutions.

In this work we propose a solution for multiple organs segmentation that finds a good trade-off between robustness, accuracy and speed. The method extends the work of Mory (2011); Mory et al. (2012) on template deformation to multiple objects. We give an overview of the original approach in Section 4.2 and present the principle of our multiple template deformation in Section 4.3. The idea is to deform shape models according to image information and such that the segmentation contours of the different organs do not overlap. The shape prior ensures shape-consistent and topology-preserved segmentations. We also suggest specific implementation details to make the numerical optimization efficient. The algorithm can accomodate various kinds of image information to drive the segmentation, which makes the approach generic and applicable to different types of modalities. In Section 4.4 we give an example of a generic edge- and intensity-based region term that can be used with our framework. Finally in Section 4.5 we propose a complete automatic segmentation workflow composed of our localization method (presented in Chapter 3) and the segmentation method presented in this chapter. We apply this framework to the segmentation of abdominal CT data and propose a qualitative and quantitative evaluation.

NOTATIONS We use the following notations hereafter:

- d is the image dimension, $d = 2$ for 2D images and $d = 3$ for 3D volumes,

- I is an image defined as $I : \Omega \rightarrow \mathbb{R}$,

- $\Omega \subset \mathbb{R}^d$ is the image domain,

- $\Omega_n \subset \mathbb{R}^d$ is the domain of template n,

- $\mathbf{x} \in \mathbb{R}^d$ is a point,

- ϕ denotes an implicit function, whose zero-level corresponds to the contour and which is positive inside the contour and negative outside,

- $\mathbf{u} : \Omega \rightarrow \mathbb{R}^d$ is a vector field, usually defined in the image domain,

- ∇ is the gradient operator in $\mathbb{R}^d$,

- div is the divergence operator in $\mathbb{R}^d$,

- Δ is the Laplacian operator in $\mathbb{R}^d$, such that $\Delta = div(\nabla)$,

- $H : \mathbb{R} \rightarrow [0,1]$ is the Heaviside step function, if $x \leq 0$ then $H(x) = 1$ and if $x > 0$ then $H(x) = 0$,

- δ is the dirac function,

- Id is the identity transformation,

- $\circ$ is the functional composition,

- $*$ denotes the convolution.

4.2 PRESENTATION OF THE IMPLICIT TEMPLATE DEFORMATION FRAMEWORK

In Section 2.4.5.1 we gave a literature overview of the different deformable models for single object segmentation. In this section we present the implicit template deformation framework introduced by Mory (2011); Mory et al. (2012), on which our approach is based. We also show how Prevost (2013) modifies the original framework to solve theoretical pitfalls.

This presentation is largely inspired by the Ph.D. work of Mory (2011), the Ph.D. work of Prevost (2013) and the conference paper of Mory et al. (2012).

4.2.1 *The framework*

The implicit template deformation framework is based on a region-based variational formulation for segmentation. It formulates an energy term which depends on image information computed outside and inside the evolving contour. Instead of finding directly the implicit function that minimizes the energy, as it is done in level set approaches, the idea is to find a regularized transformation that deforms an implicit template model while minimizing the energy functional. Among others, this has the advantage to avoid the re-initialization of the implicit function as a signed distance function.

We denote a shape template by ϕ_0, fixed and defined in its own referential Ω_0, and a regular transformation $\psi : \Omega \rightarrow \Omega_0$ (we explain later how to define the transformation ψ to ensure its regularity). Figure 4.1 shows how the transformation deforms the model template ϕ_0 with the transformation ψ.

In the implicit template framework, the transformation ψ is optimized and the equation of the problem is written as follows:

$$\min_{\psi} \left\{ E[\psi] = \int_{\Omega} H(\phi_0 \circ \psi(\mathbf{x})) f(\mathbf{x}) d\mathbf{x} + \lambda \mathcal{R}(\psi) \right\}$$

$$\text{with } f(\mathbf{x}) = f_{int}(\mathbf{x}) - f_{ext}(\mathbf{x})$$

$$(4.1)$$

(a) Star-shaped prior   (b) Implicit template ϕ_0 (c) Deformed template $\phi = \phi_0 \circ \psi$
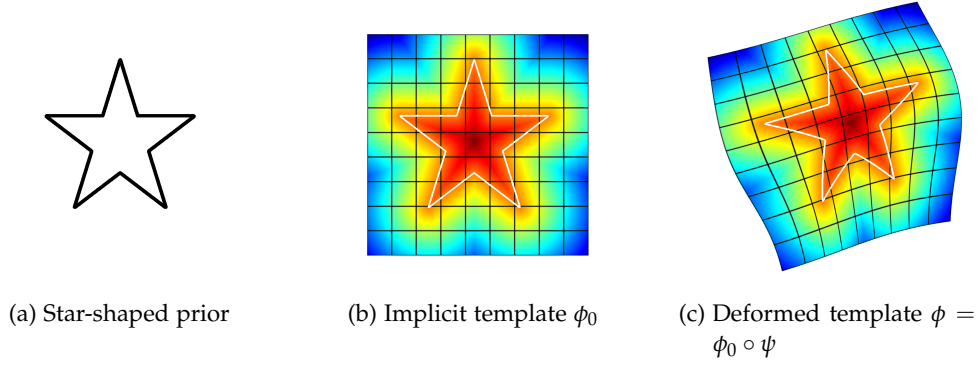
Figure 4.1: Illustration of the principle of template deformation (ϕ_0 is the implicit template model, ψ is a transformation, ϕ corresponds to the warped template). (Source: Mory (2011)).

where $\mathcal{R}$ is a regularization term penalizing the transformation ψ and $f(\mathbf{x})$ is the region term related to the image information (we detail this term later).

A SHAPE-BASED MODEL   The transformation model exploited in the template deformation algorithm of Mory et al. (2012) relies on an algebraic definition of shape as in (Soatto and Yezzi, 2002; Yezzi and Soatto, 2003). The idea is to follow the definition of a shape as a geometrical representation of a portion of space occupied by an object without any notion of location, orientation or size. Thus, the shape prior previously defined by ϕ_0 is defined in an abstract referential $\Omega_0 \subset \mathbb{R}^d$ that is normalized and centered. A local deformation $\mathcal{L} : \Omega_0 \rightarrow \Omega_0$ deforms the original shape in the domain Ω_0 thus giving a new shape. Then a global transformation $\mathcal{G} : \Omega \rightarrow \Omega_0$ maps this shape from the abstract domain Ω_0 to the domain Ω of the image, thus giving the object position, orientation and scale. We call this transformation the *pose* of the model. The transformation ψ can thus be defined by the composition:

$$\psi = \mathcal{L} \circ \mathcal{G} \tag{4.2}$$

Such a decomposition make it possible to define the regularization term intrinsically to the shape such that $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L})$. The shape regularization term is thus independent of the pose. A possible definition of $\mathcal{R}$ is to penalize the magnitude of the induced deformation via the L_2-norm as follows:

$$\mathcal{R}(\mathcal{L}) = \frac{1}{2} \|\mathcal{L} - Id\|_2^2 \tag{4.3}$$

In the case of ambiguous image information, the regularization term will dominate and keep $\mathcal{L}$ close to the identity such that the segmentation contour remains close to the shape prior. Now let us see how to parameterize the transformation $\mathcal{G}$ and define the deformation $\mathcal{L}$ such that it remains regular.

4.2.2   *Definition of the transformation ψ*

PARAMETRIZATION OF THE POSE $\mathcal{G}$ The pose $\mathcal{G}$ is a parametric transformation that can be a rigid transformation or a similarity for instance. In the case of a similarity, we represent it by a matrix in homogeneous coordinates defined with 7 parameters $\mathbf{p} = \{p_l\}_{l=1,..7}$.

A REGULAR DEFORMATION $\mathcal{L}$ Most organs have a fixed topology. With a proper definition of the transformation ψ, the template deformation model is able to ensure this property. Prevost (2013) highlights that specific properties of the deformation $\mathcal{L}$ are required to ensure the topology preservation: (i) no folds should be generated (which means that $\mathcal{L}$ should be injective), (ii) no holes should be created ($\mathcal{L}$ should be onto). Adding the constraint of regularity, diffeomorphisms are thus suitable candidates.

As in the work of Rueckert et al. (1999), Mory et al. (2012) defines the deformation $\mathcal{L}$ as:

$$\mathcal{L} = Id + \mathbf{u} \tag{4.4}$$

where $\mathbf{u} \in L^2(\Omega_0, \mathbb{R})$ is a displacement field. To keep the deformation $\mathcal{L}$ regular, Mory et al. (2012) propose to define $\mathbf{u}$ as a filtered version of an integrable displacement field $\mathbf{v} \in L^2(\Omega_0, \mathbb{R})$:

$$\mathbf{u}(\mathbf{x}) = [K_\sigma * \mathbf{v}](\mathbf{x}) = \int_{\Omega_0} K_\sigma(\mathbf{x} - \mathbf{y}) \mathbf{v}(\mathbf{y}) d\mathbf{y} \tag{4.5}$$

where K_σ is a Gaussian kernel with scale σ. This approach, similar to the *Demons* algorithm Thirion (1998), enforces smoothness in a range related to σ. However, Prevost (2013) shows that this formulation encounters theoretical issues (non guaranteed smoothness and ill-posed optimization problem). To solve them, he proposes a new formulation by defining the displacement field $\mathbf{u}$ lying in a reproducing kernel Hilbert space (RKHS) U defined by the Gaussian kernel K_σ on $L^2(\Omega, \mathbb{R}^d)$. Equation 4.1 can then be written as:

$$\min_{\mathbf{p}, \mathbf{u}} \left\{ E[\mathbf{p}, \mathbf{u}] = \int_\Omega H(\phi_0 \circ \psi_{\mathbf{p}, \mathbf{u}}(\mathbf{x})) f(\mathbf{x}) d\mathbf{x} + \frac{\lambda}{2} \|\mathbf{u}\|_U^2 \right\}$$
$$\text{with } f(\mathbf{x}) = f_{int}(\mathbf{x}) - f_{ext}(\mathbf{x}), \tag{4.6}$$
$$\psi_{\mathbf{p}, \mathbf{u}} = \mathcal{L} \circ \mathcal{G},$$
$$\mathcal{L} = Id + \mathbf{u}, \|\mathbf{u}\|_U^2 = \langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\Omega, \mathbb{R}^d)} = \langle \mathbf{u}, K_\sigma^{-1} \mathbf{u} \rangle_{L^2(\Omega, \mathbb{R}^d)}$$

Although this formulation ensures a regular deformation $\mathcal{L}$, Prevost (2013) shows that a proper optimization scheme should be used to ensure that $\mathcal{L}$ is a diffeomorphism. We detail this in Section 4.2.4.

Table 4.1: Catalog of global/local region statistics and flux maximization. These models can be used for the foreground ($i = 1$) or the background regions ($i = 2$). (Source: Mory (2011)).

	Region Term function	Parameters
Piecewise-constant	$f_i(x) = (I(\mathbf{x}) - c_i)^2$	c_i = mean
Gaussian	$f_i(\mathbf{x}) = \frac{1}{2\sigma_i^2}(I(\mathbf{x}) - c_i)^2 + \log(\sigma_i)$	(c_i, σ_i^2) = (mean,var)
Non-parametric	$f_i(\mathbf{x}) = \log(p_i(I(\mathbf{x})))$	p_i = Parzen PDF
Piecewise-smooth	$f_i(\mathbf{x}) = \int_\Omega W(\mathbf{x} - \mathbf{y})(I(\mathbf{x}) - s_i(\mathbf{y}))^2 d\mathbf{y}$	s_i = normalized convolution
Local non-parametric	$f_i(\mathbf{x}) = -\int_\Omega W(\mathbf{x} - \mathbf{y}) \log p_i(I(\mathbf{x}), (\mathbf{y})) d\mathbf{y}$	p_i = local PDF
Gradient flux	$f_1(\mathbf{x}) = \pm\Delta I(\mathbf{x})$ and $f_2(\mathbf{x}) = 0$	-

4.2.3 *The region term*

The template deformation framework is based on a region-based variational formulation. The region term (or image-driven forces) is defined in Equation 4.6 by the function $f(\mathbf{x}) = f_{int}(\mathbf{x}) - f_{ext}(\mathbf{x})$ where $f_{int}(\mathbf{x})$ and $f_{ext}(\mathbf{x})$ correspond to image information related to the interior and the exterior of the contour, respectively. Multiple expressions have been proposed so far for these functions in the literature as shown in Table 4.1, where some examples are given.

Most of these terms rely on intensity information. The gradient flux, based on the Laplacian of the image, exploits the edge information. However it is dependent on the contrast of the object to segment with regards to the background. Although it may give good results in some cases (e. g. for kidney segmentation in CT and US (Mory et al., 2012; Cuingnet et al., 2012)), its application is however limited. If an object to segment is surrounded with objects of higher and lower intensities, the use of the gradient flux may fail. We see in Section 4.4 how to tackle this problem by comparing other contrast-invariant edge-driven terms.

4.2.4 *Numerical optimization*

Mory et al. (2012) propose to solve the problem of Equation 4.6 by gradient descent. This simple optimization procedure has the advantage of limiting the computations to the gradients of the energy function. As noticed by Prevost (2013), the fact that Equation 4.6 is highly non-convex is not that problematical. The design of an energy providing the global optimum corresponding exactly to the organ of interest is indeed often more challenging than ensuring a good localization to find the correct local minimum.

Following the work of Saddi et al. (2007b), Prevost (2013) shows that to ensure a diffeomorphic local deformation $\mathcal{L}$, the gradient descent update should be done with composition updates. Starting from an initialization, the variables $\mathbf{p} = \{p_l\}_{l=1,..7}$ and $\mathbf{u}$ are updated iteratively:

$$
\begin{aligned}
p_l &\leftarrow p_l - \Delta t \frac{\partial E}{\partial p_l} \\
\mathbf{u} &\leftarrow \mathbf{u} \circ (Id - \Delta t \nabla_{\mathbf{u}} E) - \Delta t \nabla_{\mathbf{u}} E
\end{aligned}
\tag{4.7}
$$

where $\Delta t \in \mathbb{R}^+$ is the time step constant. We refer to Prevost (2013) for the computations and expressions of the gradients $\frac{\partial E}{\partial p_l}$ and $\nabla_{\mathbf{u}} E$.

4.3   A FRAMEWORK FOR MULTIPLE IMPLICIT TEMPLATE DEFORMATION

In this section we present the core of our multi-object segmentation framework. We want to segment N objects indexed by $n \in [\![1, N]\!]$ and for which we have shape models. The idea is to find the transformations (rigid and/or non-rigid) of these shape models that best map to the object boundaries in an image I. By extending the work of Mory (2011); Mory et al. (2012), we propose a variational framework to solve this problem. In Section 4.3.1 we describe the base of the framework. In Section 4.3.2 we propose to add non-overlapping constraints to the framework. Then in Section 4.3.3 we present the numerical optimization scheme chosen to solve our variational approach and we give the associated implementation details in Section 4.3.4. Finally we conclude on this approach in Section 4.3.6. The notations used are described in Section 4.1.

4.3.1   *The core of the framework*

Let us denote N objects to segment indexed by $n \in [\![1, N]\!]$. We define their associated shape templates (built for instance from several examples, see Appendix E) with implicit functions $\phi_n : \Omega_n \to \mathbb{R}$ (for $\mathbf{x} \in \Omega_n$, $\phi_n(\mathbf{x}) > 0$ if $\mathbf{x}$ is inside the shape contour, $\phi_n(\mathbf{x}) < 0$ if $\mathbf{x}$ is outside the shape contour and $\phi_n(\mathbf{x}) = 0$ if $\mathbf{x}$ is on the contour). We define $\psi_n : \Omega \to \Omega_n$ the transformation acting on object n that maps the image domain to the template domain. Each object n is associated with image-driven forces denoted by $f_n : \Omega \to \mathbb{R}$.

We want to find the transformations ψ_n such that the data-driven forces are minimal inside each object contour. The energy to minimize can be written as:

$$
E[\psi_1, ..., \psi_N] = \sum_{n=1}^{N} \left(\int_{\Omega} H(\phi_n \circ \psi_n(\mathbf{x})).f_n(\mathbf{x}) d\mathbf{x} \right)
\tag{4.8}
$$

where H is the Heaviside step function.
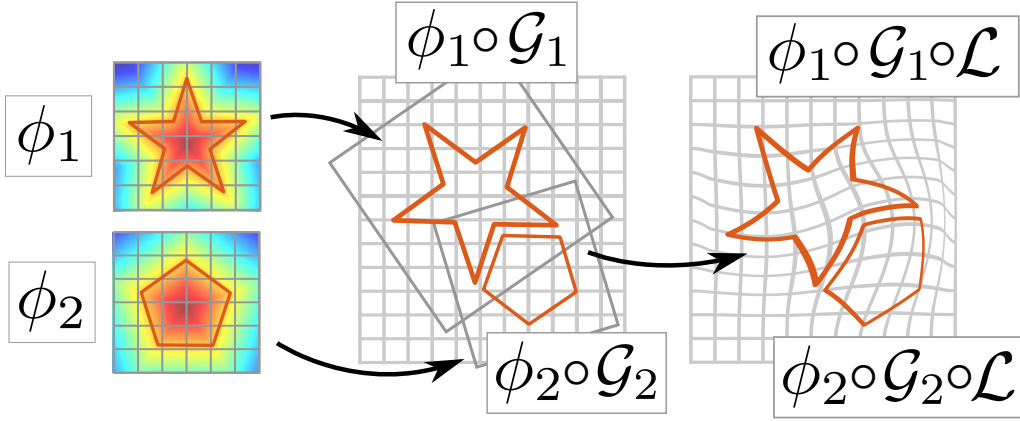
Figure 4.2: Illustration of the deformation framework with two objects.

The transformations can be defined as the composition of a global transformation (or pose) and a non-rigid deformation. The non-rigid deformation, denoted by $\mathcal{L} : \Omega \to \Omega$, deforms locally the image domain. The global transformation of object n, denoted by $\mathcal{G}_n : \Omega \to \Omega_n$, transports the image domain to the template domain. We write $\psi_n = \mathcal{G}_n \circ \mathcal{L}$ for each object n. We later show that this assumption is not restrictive and greatly simplifies the formulation and associated optimization. We assume that the non-rigid transformation $\mathcal{L}$, acting in the image domain, is common to all the objects. We see later what allows us to make this assumption and its interest. Figure 4.2 illustrates the deformation principle. The equation of the energy to minimize can now be written as follows:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) \tag{4.9}$$

This energy definition however does not guarantee that the local deformation induced by $\mathcal{L}$ maintains regularity. Thus we add a regularization term to the energy definition to ensure this property:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_2^2 \tag{4.10}$$

where λ is a fixed constant controlling the importance of the regularization.

In the following paragraphs we detail the model of the local deformation and the regularization term.

REGULARIZATION TERM   The decomposition $\psi_n = \mathcal{G}_n \circ \mathcal{L}$ allows us to define the regularization term independently of the pose (second term in Equation 4.10) which is of high interest. Indeed, the templates of the objects remain fixed while the transformations G_n are restricted to rigid deformations such as scaling, rotation, etc. Controlled with the constant λ, the regularization term

penalizes high deformations while preserving the shape prior. For high values of λ the local deformations should remain very small and very close to the original shape, while for small values the local deformations may be larger and get the contours away from the shape model.

MODEL OF THE LOCAL DEFORMATION   We follow the same definition of the local deformation presented in Section 4.2.2 and introduced by Mory (2011); Mory et al. (2012); Prevost (2013).

The local deformation is defined as $\mathcal{L} = Id + \mathbf{u}$ where $\mathbf{u} : \Omega \rightarrow \mathbb{R}^d$ is a displacement field $\mathbf{u}$ lying in a RKHS U defined by the Gaussian kernel K_σ on $L^2(\Omega, \mathbb{R}^d)$. Equation 4.10 can now be written as:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2$$

$$\text{with } \mathcal{L} = \mathbf{Id} + \mathbf{u}, \|\mathbf{u}\|_U^2 = \langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\Omega, \mathbb{R}^d)} = \langle \mathbf{u}, K_\sigma^{-1}\mathbf{u} \rangle_{L^2(\Omega, \mathbb{R}^d)}$$

(4.11)

To ensure that $\mathcal{L}$ is a diffeomorphism, a proper optimization is needed as detailed in Section 4.3.3.

IMAGE-DRIVEN FORCES   The image-driven forces can follow any formulation coming from the region-based variational framework. A list of possible image forces formulations is given in Section 4.2.3. According to the structures of interest they can rely on global or local information and on intensities or on edges. This choice will depend on the type of application. In Section 4.4 we propose a type of image-driven forces taking into account both intensities and edge information. Note that these forces could also be derived from classification maps as done in the work of Cuingnet et al. (2012); Prevost (2013).

4.3.2   *Non-overlapping constraints*

The basis formulation given in Equation 4.11 does not guarantee non-overlapping of the objects. To prevent such a problem we propose to introduce new constraints on the global pose transformations $\mathcal{G}_n$. The problem to minimize becomes:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2$$

$$\text{subject to } \forall (i,j) \in [\![1, N]\!]^2, i < j,$$

$$C(\mathcal{G}_i, \mathcal{G}_j) = \int_\Omega H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x}))d\mathbf{x} = 0$$

(4.12)

The constraints are designed such that intersections of each pair of objects should reduce to the empty set. Note that the constraints are applied on the objects after their pose transformation $\mathcal{G}_n$. This means that when they are correctly rigidly placed in the image domain, they can then be deformed with the transformation $\mathcal{L}$. As the local deformation $\mathcal{L}$ is a diffeomorphism, non-overlapping is guaranteed (as illustrated in Figure 4.2).

4.3.3 *Numerical optimization scheme*

In this section, we propose to solve the minimization of the energy given in Equation 4.12. Depending on the nature of the function to optimize there exist different ways of solving a constraint problem. In the case of a nonlinear programming problem with equality constraints, the most well-known methods are the quadratic penalty method and the augmented Lagrangian method. These methods are used to turn a constrained problem into a series of unconstrained ones.

Unconstrained minimization methods can then be applied to solve these unconstrained problems. Here we propose to use the gradient descent method as in Mory et al. (2012). As noticed in Prevost (2013), the choice of a gradient descent with a fixed time-step is more efficient in practice than using methods to decrease the number of iterations needed for convergence (e. g. *line search*, (Nocedal and Wright, 2006)). Indeed, such methods require to compute the energy which would involve the computation of volume integrals, while the gradients only require surface integral computations (see Section 4.3.4).

We detail hereafter two possible constrained optimizations with the quadratic penalty and the augmented Lagrangian methods. We refer the reader to the book of Nocedal and Wright (2006) for more details on numerical optimization under constraint.

4.3.3.1 *More details about the variables of the problem*

The parametric transformations $\mathcal{G}_n$ are defined by N_g parameters $\mathbf{p_n} = \{p_{n,l}\}_{l=1...N_g}$ (e.g. $N_g = 6$ if $\mathcal{G}_n$ is rigid in $\mathbb{R}^3$).

We assume that the forces f_n are fixed during the optimization.

4.3.3.2 *Formulation with the quadratic penalty method*

The quadratic penalty method was introduced by (Courant, 1943). It aims at solving a series of unconstrained minimization problems by adding a multiple of the square of the violation of each constraint to the energy function. As a result, the additional terms are positive when the constraints are violated and zero otherwise.

Following this principle, Equation 4.12 turns into a series of unconstrained minimizations expressed as:

$$
\min_{\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}} \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}]
$$

$$
\text{with } \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] = E[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] + \frac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i,\mathcal{G}_j)^2 \tag{4.13}
$$

where μ_k is the penalty coefficient. At each optimization k, μ_k is increased, the unconstrained problem is solved by gradient descent and used as the initialization for the next iteration. In this way, the constraints are penalized more and more severely at each unconstrained minimization.

To simplify the notations we define $V_Q = \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i,\mathcal{G}_j)^2$. The problem becomes:

$$
\min_{\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}} \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}]
$$

$$
\text{with } \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] = E[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] + \mu_k V_Q \tag{4.14}
$$

For each optimization k, $p_{n,l}$ and $\mathbf{u}$ are updated jointly and iteratively:

$$
p_{n,l(t+1)} \leftarrow p_{n,l(t)} - \Delta t_p \frac{\partial \hat{E}_k}{\partial p_{n,l}}
$$

$$
\mathbf{u}_{(t+1)} \leftarrow \mathbf{u}_{(t)} \circ (\mathbf{Id} - \Delta t_u \nabla_u \hat{E}_k) - \Delta t_u \nabla_u \hat{E}_k \tag{4.15}
$$

with Δt_p, Δt_u fixed time steps. The initialization of the optimization $k + 1$ is the result of optimization k after convergence. The update of $\mathbf{u}$ is done by composition to ensure that the local deformation is a diffeomorphism, as explained by Prevost (2013).

The details of the gradient calculation are given in Section D.1.1. The resulting evolution equations are:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_\Omega \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x})\rangle.f_n(\mathbf{x})d\mathbf{x} + \nabla_{p_{n,k}} V_Q$$

$$\nabla_u \hat{E}_k(\mathbf{x}) = K_\sigma * \left\{ \sum_{n=1}^N \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}) \right\} + \lambda \mathbf{u}(\mathbf{x})$$

with

$$\nabla_{p_{n,k}} V_Q = \sum_{\substack{1 \le i \le N \\ i \ne n}} \left(C(\mathcal{G}_i, \mathcal{G}_n) \int_\Omega \delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,k}}\rangle dx \right)$$

(4.16)

REMARK   Due to its simplicity, the method of the quadratic penalty is often used in practice. However, as demonstrated by Nocedal and Wright (2006), this approach suffers from disadvantages. In particular, the ill-conditioning of the Hessian causes numerical difficulties when μ_k becomes large, and the global minimizer cannot be attained in general. To construct a penalty problem that is equivalent to the original problem and that reduces the possibilities of ill-conditioning, Lagrange multipliers can be explicitly integrated in the function. This method is called the augmented Lagrangian. We show how to implement it in the next section.

4.3.3.3   *Formulation with the augmented Lagrangian method*

The augmented Lagrangian method was proposed by Hestenes (1969) and Powell (1969). It aims at modifying the quadratic penalty formulation by introducing a summation term involving Lagrange multipliers. Other variants are proposed and compared in (Birgin et al., 2004). They show that this formulation is the most efficient in practice.

Following this principle, Equation 4.12 turns into a series of unconstrained minimizations as follows:

$$\min_{\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}} \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}]$$

$$\text{with } \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] = E[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] - \sum_{\substack{1 \le i \le N \\ i < j \le N}} \alpha_{i,j}^k.C(\mathcal{G}_i,\mathcal{G}_j) + \frac{\mu_k}{2} \sum_{\substack{1 \le i \le N \\ i < j \le N}} C(\mathcal{G}_i,\mathcal{G}_j)^2$$

(4.17)

where μ_k is the penalty coefficient. At each optimization k, the unconstrained problem of Equation 4.17 is solved by gradient descent, μ_k is increased, the $\alpha_{i,j}^k$ are updated and the solution is used as the initialization for the next iteration.

The poses $\mathcal{G}_i^*$ and $\mathcal{G}_j^*$ are the optimal solutions found after optimization k. In the case of equality-constraints such as ours, the Lagrangian multipliers can be updated as follows (Nocedal and Wright, 2006):

$$\alpha_{i,j}^{k+1} = \alpha_{i,j}^k - \mu_k C(\mathcal{G}_i^*, \mathcal{G}_j^*), \quad \forall(i,j), 1 \leq i \leq N, i < j \leq N \tag{4.18}$$

To simplify the notations we define

$$V_L = - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \alpha_{i,j}^k . C(\mathcal{G}_i, \mathcal{G}_j) + \frac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i, \mathcal{G}_j)^2$$

and the problem notation simplifies to:

$$\min_{\mathcal{G}_1, .., \mathcal{G}_N, \mathcal{L}} \hat{E}_k[\mathcal{G}_1, .., \mathcal{G}_N, \mathcal{L}]$$
$$\text{with } \hat{E}_k[\mathcal{G}_1, .., \mathcal{G}_N, \mathcal{L}] = E[\mathcal{G}_1, .., \mathcal{G}_N, \mathcal{L}] + V_L \tag{4.19}$$

As in Section 4.3.3.2, for each optimization k, $p_{n,l}$ and $\mathbf{u}$ are updated jointly and iteratively following Equation 4.15, the same update equation as in the previous section.

The details of the gradient calculation are given in Appendix D. The resulting evolution equations are:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_\Omega \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})) . \langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x}) \rangle . f_n(\mathbf{x}) d\mathbf{x} + \nabla_{p_{n,k}} V_L \tag{4.20}$$

$$\nabla_u \hat{E}_k(\mathbf{x}) = K_\sigma * \left\{ \sum_{n=1}^N \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})) . f_n(\mathbf{x}) . (d\mathcal{G}_n)^t \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}) \right\} + \lambda \mathbf{u}(\mathbf{x}) \tag{4.21}$$

with

$$\nabla_{p_{n,l}} V_L = \sum_{\substack{1 \leq i \leq N \\ i \neq n}} \left(\mu_k C(\mathcal{G}_i, \mathcal{G}_n) - \alpha_{i,n}^k \right) . \int_\Omega \delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})) . H(\phi_i \circ \mathcal{G}_i(\mathbf{x})) . \langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x} \tag{4.22}$$

4.3.3.4 *Comparison between the quadratic penalty and the augmented Lagrangian methods*

As mentioned earlier, the augmented Lagrangian method is more stable than the quadratic penalty method, which suffers ill-conditioning. The augmented Lagrangian approach also does not need the parameter μ to take infinite values so that the optimization converges to the optimum. In the following experiment, we evaluate the difference of convergence speed between the two methods.
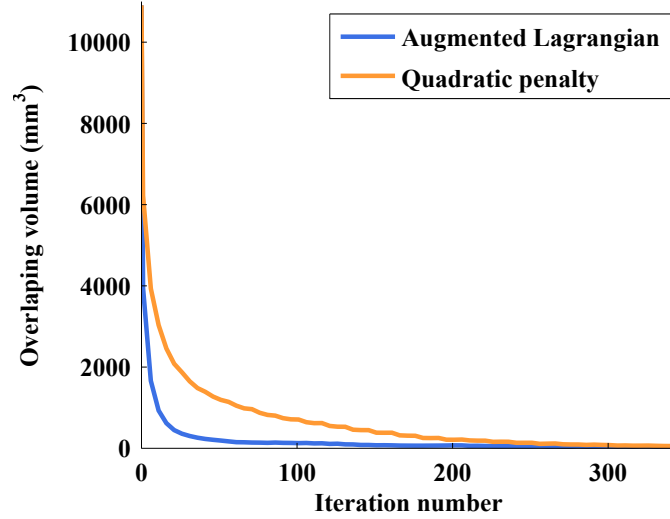
Figure 4.3: Comparison of the convergence of the augmented Lagrangian and squared penalty methods. The total overlapping volume is given in function of the iterations.

EXPERIMENT   We selected arbitrarily 10 abdominal CT volumes on which we performed our optimization framework with non-overlapping constraints on 6 organs (the liver, the two kidneys, the heart, the spleen and the gallbladder). In this experiment we fix the local deformation $\mathcal{L} = Id$ as we are interested in the impact of the overlapping constraint. We performed 350 iterations on each volume and computed the volume overlap of each couple of organs after each 5 iterations. The Lagrangian multipliers are updated after each 15 iterations and the penalty term is multiplied by 1.2 after each 30 iterations (for both methods). These parameters were fixed experimentally and for fair comparison the penalty parameter is the same in both experiments. The results are then averaged over the 10 images. This experiment is performed twice, with the quadratic penalty method and with the augmented Lagrangian method. We show the results in Figure 4.3.

CONCLUSION   We see in Figure 4.3 that the overlap volume decreases significantly faster with the augmented Lagrangian than with the quadratic penalty method. Given these results, we use the augmented Lagrangian approach in the following experiments.

4.3.4 *Implementation details for fast execution*

We see in Equations 4.20, 4.21 and 4.22 that there are multiple terms to compute. We see hereafter how to reduce the computation cost in order to get a fast implementation.

FROM INTEGRAL OVER THE VOLUME TO INTEGRAL OVER THE SURFACE
In Equations 4.20 and 4.22 we notice two integrals over the image domain Ω.
The first expression to integrate includes the term $\delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}))$ which is
nothing more than the contour of the warped template in the image domain.
Thus, the gradient equations correspond to integrals over a surface term. As
explained by Prevost (2013), in most existing works (e. g. Chan and Vese (2001))
this kind of term is numerically computed by using an approximation $\delta_\epsilon \in C^\infty$
of the Dirac distribution such that

$$\forall a \in \mathbb{R}, \delta_\epsilon(a) = \frac{\epsilon}{\pi(a^2 + \epsilon^2)}$$

Prevost (2013) notices that this approximation has a non-compact support which
implies to compute the integral at least in a narrow band of the surface. More-
over the selection of ϵ clearly influences the accuracy of the final result. He
shows that another strategy based on the *generalized scaling property*[1] can be
used. This property shows that the integrals given in Equations 4.20 and 4.22
by:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_\Omega \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x})\rangle.f_n(\mathbf{x})d\mathbf{x} + \mu_k.\nabla_{p_{n,k}}V_L$$

$$\nabla_{p_{n,l}}V_L = \sum_{\substack{1 \le i \le N \\ i \ne n}} \left(C(\mathcal{G}_i, \mathcal{G}_n) - \alpha_{i,n}^k \right).\int_\Omega \delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\langle \nabla\phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}}\rangle d\mathbf{x}$$

can be properly turned into surface integrals, as follows:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_{\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})=0} \langle \frac{\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})}{\|\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})\|}, \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x})\rangle.f_n(\mathbf{x})d\mathbf{x} + \mu_k.\nabla_{p_{n,k}}V_L \quad (4.23)$$

$$\nabla_{p_{n,l}}V_L = \sum_{\substack{1 \le i \le N \\ i \ne n}} \left(C(\mathcal{G}_i, \mathcal{G}_n) - \alpha_{i,n}^k \right).\int_{\phi_n \circ \mathcal{G}_n(\mathbf{x})=0} \frac{H(\phi_i \circ \mathcal{G}_i(\mathbf{x}))}{\|\phi_n \circ \mathcal{G}_n(\mathbf{x})\|}.\langle \nabla\phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}}\rangle d\mathbf{x}$$

$$(4.24)$$

In practice, the marching cube algorithm (Lorensen and Cline, 1987) is run
on the warped template zero level to obtain a set of points sampling the surface
integral.

---

[1] Let $\Omega \in \mathbb{R}^d$, $f : \Omega \to \mathbb{R}$ 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: $\int_\Omega \delta(\phi(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \int_{\phi(\mathbf{x})=0} \frac{f(\mathbf{x})}{\|\nabla\phi(\mathbf{x})\|d\mathbf{x}}$
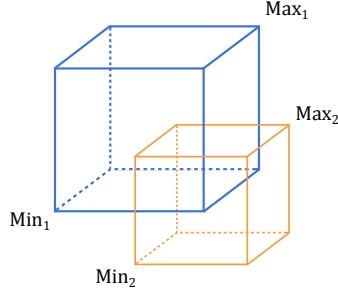
Figure 4.4: Two overlapping axis-aligned bounding boxes and extremal vertices.

input : $Min_1[3], Max_1[3],$
$\qquad Min_2[3], Max_2[3]$
output: Boolean
for $i \leftarrow 1$ **to** 3 **do**
$\quad$ **if** $Max_1[i] < Min_2[i]$ **then**
$\quad\quad \llcorner$ return False
$\quad$ **if** $Max_2[i] < Min_1[i]$ **then**
$\quad\quad \llcorner$ return False
return True;

Algorithm 3: Algorithm for axis-aligned bounding boxes collision detection.

COMPUTATIONS NEAR THE ZERO LEVEL   As in Mory et al. (2012), the equations allow us to compute the terms $\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}$ and $\phi_n \circ \mathcal{G}_n \circ \mathcal{L}$ only near their zero level. By defining ϕ_n as distance functions, they can be computed in an efficient coarse-to-fine approach using octrees (refer to Mory et al. (2012) for details).

COLLISION DETECTION   The term of Equation 4.22 corresponds to the sum of non-overlapping constraints per couple of objects. If two objects are far from each other, there is then no need to compute the corresponding constraint term. Thus we propose to add a simple collision detection on pairs of object bounding boxes to reduce the computations.

With axis-aligned boxes this is very simple: it only requires to compare the minimum and maximum vertices dimension per dimension. We give an illustration of the principle in Figure 4.4 and Algorithm 3.

In Table 4.2 we give computation times per iteration without non-overlapping constraints, with and without collision detection step. We observe a computation acceleration of about 60%.

For more information on collision detection we refer the reader to the thesis work of Lin (1993) which gives an extensive review and study on the subject and for different types of objects.

WORK AT LOWER RESOLUTION   Another way to reduce the computation time is to reduce the working resolution. The most critical parameter is the resolution of the implicit function $\phi_n \circ \psi_n$. Indeed, this term needs to be computed at each iteration. We showed just above that it is only required near its zero level and can be efficiently computed using an octree-based method. It is possible to reduce even more the computation time by computing $\phi_n \circ \psi_n$ at a lower resolution. As seen in Figure 4.5, the time per iteration varies ex-

Table 4.2: Computation times with and without collision detection. Experiment realized on one image on 200 iterations. Times are given per iteration.

	2 organs (liver, left kidney)	6 organs (liver, kidneys, spleen, heart, gallbladder)
Without non-overlapping constraint	2.5 ms	85 ms
With non-overlapping constraints without collision detection	25 ms	165 ms
With non-overlapping constraints with collision detection	15 ms (-60%)	100 ms (-60%)
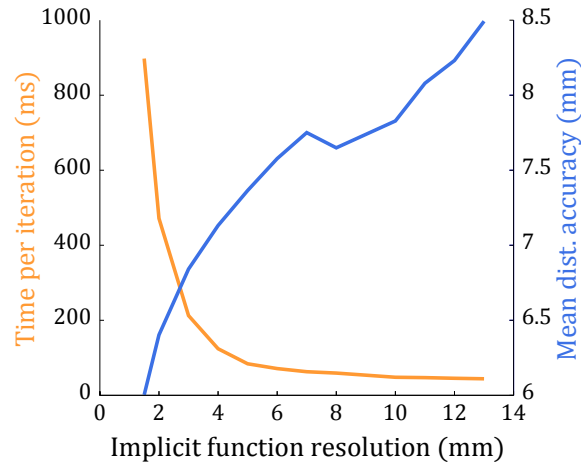
Figure 4.5: Computation time and mean distance accuracy in function of the implicit function $\phi_n \circ \psi_n$ resolution. Experiment realized on 400 iterations for the segmentation of 4 organs in 3D.

ponentially with the resolution. This has however a significant impact on the segmentation accuracy. A trade-off should then be found between the expected accuracy and the computation time. In practice we generally use a resolution of 3 to 5mm.

4.3.5 *Illustrative experiments in* 2D

In this section we present some results on toy examples in 2D. Note that the framework is further evaluated in Section 4.5 on real 3D medical images.

In the following experiments we propose to segment three objects simultaneously: one star, one pentagon and one disk.

(a) Initialization

(b) Result without non-overlapping constraints

(c) Result with non-overlapping constraints
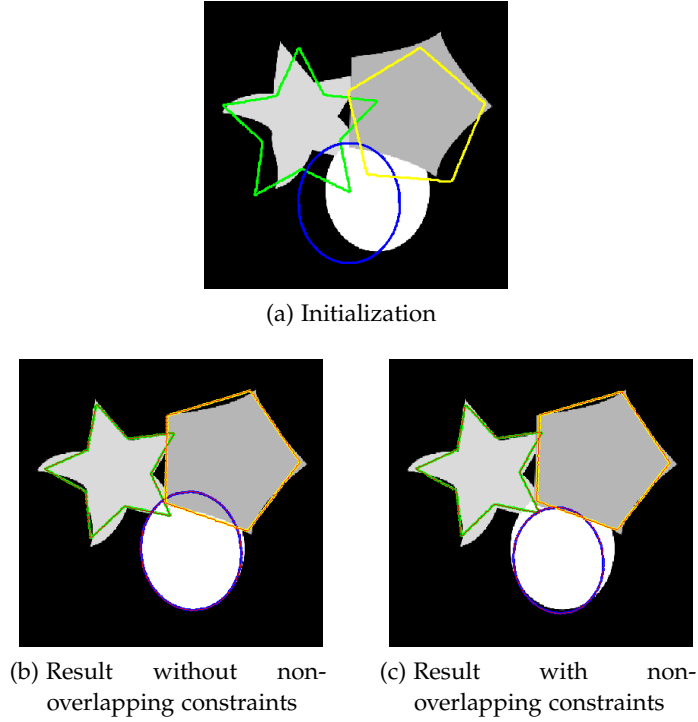
Figure 4.6: Experiment 1. Initialization and segmentation results with three objects where $\psi_n = \mathcal{G}_n$.

EXPERIMENT 1: NON-OVERLAPPING CONSTRAINTS   In this experiment we show the impact of adding the non-overlapping constraints. We launch the optimization on the poses $\mathcal{G}_n$ only (the local deformation $\mathcal{L} = Id$).

Figure 4.6 shows the initialization and the segmentation results with and without non-overlapping constraints. The non-overlapping constraints have the expected impact on the objects: they do not overlap.

EXPERIMENT 2: CONTOUR EVOLUTION WITH $\psi_n = \mathcal{G}_n \circ \mathcal{L}$ In this experiment we show how the local deformation $\mathcal{L}$ is optimized after a given number of iterations. Figure 4.7 shows some examples where both the evolving contours and the poses of each object are displayed. We see that once the objects are optimally placed in the image, the deformation $\mathcal{L}$ modifies the evolving contours to best match the edges of the image.

EXPERIMENT 3: ROBUSTNESS TO THE INITIALIZATION   In this experiment we test the robustness of the algorithm to different initializations. In Figure 4.8 we present the results after 800 iterations and for four different initializations. In the three first examples, the algorithm seems to converge to the same local minimum. In the last example, however, the poses of objects seem to be stuck in a local minimum that the local deformation compensates with difficulty. Indeed,

(a) Initialization

(b) After 100 iterations

(c) After 200 iterations

(d) After 500 iterations

(e) After 1000 iterations
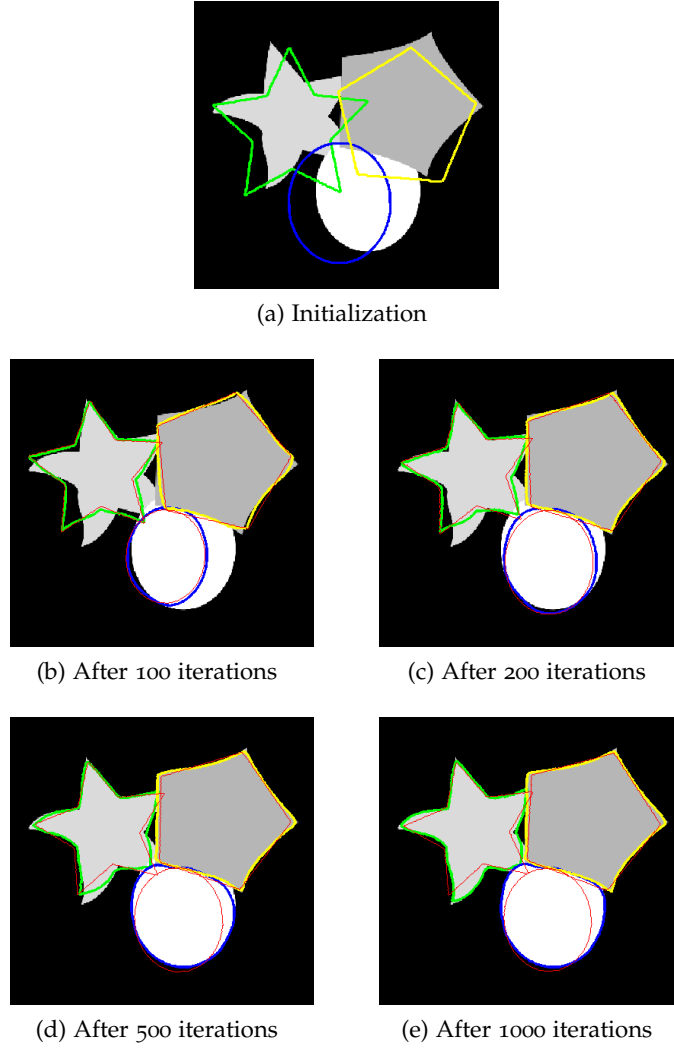
Figure 4.7: Experiment 2. Initialization and segmentation results with three objects where $\psi_n = \mathcal{G}_n \circ \mathcal{L}$. In red we show the optimized poses of each object.

the sharp borders of the pentagon template are hardly mapped to the sides while the sides of the pentagon are hardly mapped to vertices. This is due to the intrinsic regular definition of the deformation $\mathcal{L}$. Further iterations seem to give an approximate correct segmentation. Despite this limitation, this experiment still shows a good robustness to the initialization.

4.3.6 *Conclusion*

In this section we presented a new approach for the segmentation of multiple organs extending the work of Mory (2011) on template deformation. By formulating the segmentation as a variational problem optimizing the deformation of

(a) Initialization

(b) Result (after 800 it.)

(c) Initialization

(d) Result (after 800 it.)

(e) Initialization

(f) Result (after 800 it.)

(g) Initialization

(h) Result (after 800 it.)

(i) Result (after 2400 it.)

Figure 4.8: Experiment 3. Results (left column) after different initializations (right colum). In red we show the optimized poses of each object.

shape models, this ensures shape prior and shape consistency (e. g. sharp edges remain sharp). Decomposing the optimized transformation into pose and local deformations allows us: (i) to formulate a shape regularizer penalizing large deformations from the initial model, (ii) to use the algorithm alternatively for different purposes (e. g. registration or fine segmentation). Moreover, adding specific non-overlapping constraints guarantees the non-overlap of the different object segmentations. We also showed that the formulation into a constrained problem guarantees non-overlapped solutions which is not the case in existing approaches embedding soft constraints (e. g. an additional term in the energy).

Finally, we gave implementation details providing for an efficient solution. We see further in Section 4.5 that we can segment 6 abdominal organs in a typical CT volume in less than one minute.

In the next section we detail how to define a robust region term.

4.4   A ROBUST REGION TERM f_n: EXPLOITING INTENSITIES AND EDGE

As said earlier, the template deformation framework is a region-based variational approach. Region-based approaches rely on the image features lying both inside and outside the regions defined by the evolving curve/surface. Original region-based formulations often rely on intensity differences between foreground and background areas (see Section 4.2.3). To achieve the best accuracy, it is desirable to exploit both types of information: intensity- and edge-based. Intuitively, the intensity information should have a global impact on the curve/-surface evolution while the edge information should have more influence when the curve/surface gets close to the image boundaries.
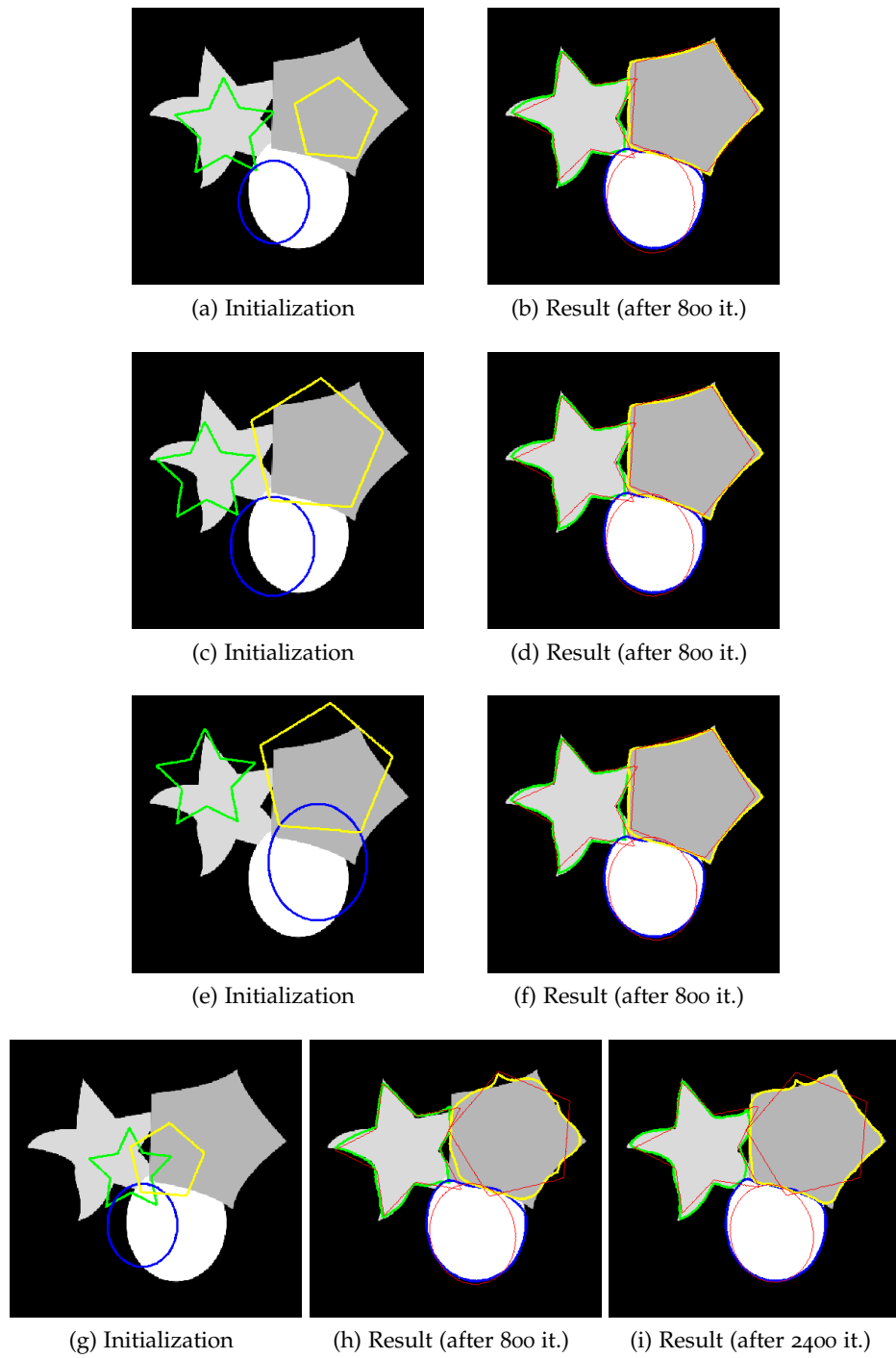
By nature, the edge information, related to the image intensities gradient ∇I, is a very local feature. It thus appears difficult to benefit from it unless the evolving contour is close to the image boundaries. This problem has actually drawn the attention of many researchers since the apparition of active contours at the end the 90's. One solution consists in diffusing the edge information in the image. Xu and Prince (1997) propose for instance a method called GVF which generates a field of forces from the gradient vectors of the image. The resulting field has a large capture range to attract the contours including towards concavities. Another approach, called *flux maximization*, proposed by Vasilevskiy and Siddiqi (2002), consists in using the Laplacian of the image ΔI on a filtered version of the image, this having the effect to diffuse the Laplacian information. The Laplacian of the image has been introduced earlier by Marr and Hildreth (1980) for edge detection. Indeed, the zero-crossings of the Laplacian correspond to local maxima of the image gradient. While GVF methods are time consuming, the flux maximization method proposed by Vasilevskiy and Siddiqi (2002) presents the disadvantage of being contrast dependent. Indeed the Laplacian of the image is negative or positive depending on the direction of

the image gradient. In Section 4.4.1 we compare the original flux maximization with two other approaches: one derived from Kimmel and Bruckstein (2003) and another one that we propose. Then in Section 4.4.2 we show how to define a robust region based on an optimal signed Laplacian function and edge information.

4.4.1 *Contrast-invariant forces*

Let us define $\vec{n}$, the unit normal to the evolving curve/surface $\mathcal{C}$. When the curve reaches the image boundaries, the gradient of the image ∇I and the normal to the contour $\vec{n}$ should coincide. The scalar product $\langle \nabla I, \vec{n} \rangle$ is then maximal in each point of the curve. If the curve matches perfectly the image boundaries, then the integral over the curve

$$\oint_{\mathcal{C}} \langle \nabla I, \vec{n} \rangle ds$$

should be maximal. This term is sometimes called *alignement term*. Such a term is used in active contours formulations. In the template deformation framework, however, a volumetric term is needed. Note that if the evolving contour $\mathcal{C}$ is the zero level of an implicit function ϕ, then the unit normal to the contour can be expressed as

$$\vec{n} = \frac{\nabla \phi(x)}{\|\nabla \phi(x)\|}$$

The divergence theorem turns the alignment term into an integral over the volume as follows:

$$\oint_{\mathcal{C}} \langle \nabla I, \vec{n} \rangle ds = \int_{\Omega_c} \Delta I(x) dx \tag{4.25}$$

where Ω_c is the domain of the image included in the closed curve $\mathcal{C}$. Note that the Laplacian term does not depend on the curve, which may be a significant advantage in terms of computations. Indeed, when segmenting an image, this term only needs to be computed once. Equation 4.25 shows the link between the contrast direction and the sign of the Laplacian, also illustrated in Figure 4.9.

In Figure 4.11 we show a toy image and its Laplacian (the toy image is used again in the following illustrations of this section). The image is composed of a disk and a background with different regions of higher and lower intensities. By looking at the corresponding Laplacian image, one can notice the contrast dependency of the Laplacian (∇I and $\vec{n}$ may be in the same direction or in the opposite direction). This example shows that the Laplacian is not suited for the segmentation of objects with contrast-varying backgrounds.

(a) $\langle \nabla I, \vec{n} \rangle < 0$ (b) $\langle \nabla I, \vec{n} \rangle > 0$
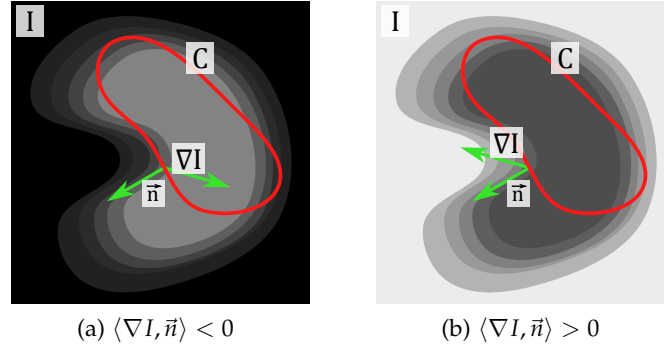
Figure 4.9: Illustration of the contrast dependency of the Laplacian. The evolving curve is in red.

SIGNED LAPLACIAN   Kimmel (2003) proposes to solve this issue with a robust version of the alignment term:

$$\oint_{\mathcal{C}} |\langle \nabla I, \vec{n} \rangle| ds = \iint_{\Omega_c} \text{sign}(\langle \nabla I, \vec{n} \rangle) \Delta I dv = \iint_{\Omega_c} f_{SL}(I, \mathcal{C}) dv \qquad (4.26)$$

The idea is to look at the projection of the image gradient on the surface normal without sign distinction. The corresponding volumetric term corresponds to the Laplacian weighted with the sign of the projection. However in return, this means that the region term depends on the position of the object in the image.

In Figures 4.11c and 4.11d we give two examples of this volumetric term for two different initializations of a contour (red circle).

OPTIMAL SIGNED LAPLACIAN   Inspired by the works of Haralick (1984); Canny (1986) on robust edge detectors, Kimmel and Bruckstein (2003) propose another term relying on the Laplacian of the image. The Haralick-Canny edge detector is based on the zero-crossing of the Laplacian. The idea is to express the second derivatives of the Laplacian along the gradient direction (corresponding to the orthogonal direction to the image level sets) instead of along the x and y directions. Although this is similar analytically, this seems to provide more accurate results numerically.

The tangent and orthogonal unit vectors to the image level sets are defined as follows. In 2D the gradient vector field $\vec{\xi} : \mathbb{R}^2 \to \mathbb{R}$ is

$$\vec{\xi}(x, y) = \frac{\nabla I(x, y)}{\|\nabla I(x, y)\|} = \frac{1}{\|\nabla I(x, y)\|} \begin{pmatrix} I_x \\ I_y \end{pmatrix} \qquad (4.27)$$

where I_x and I_y are the derivatives of I along x and y respectively. The orthogonal vector field $\vec{\eta} : \mathbb{R}^2 \to \mathbb{R}$, such that $\langle \vec{\eta}, \vec{\xi} \rangle = 0$, is

$$\vec{\eta}(x, y) = \frac{\bar{\nabla} I(x, y)}{\|\nabla I(x, y)\|} = \frac{1}{\|\nabla I(x, y)\|} \begin{pmatrix} -I_y \\ I_x \end{pmatrix} \qquad (4.28)$$
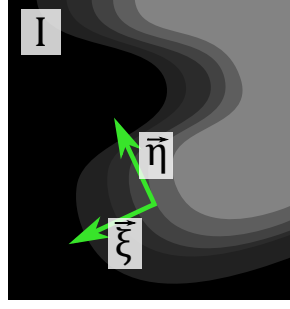
Figure 4.10: Gradient and orthogonal directions to the level-sets at a point of an image I.

They are illustrated in Figure 4.10.

By using the rotation invariance property of the Laplacian it is possible to write

$$
\begin{aligned}
I_{\xi,\xi} &= I_{\xi,\xi} + I_{\eta,\eta} - I_{\eta,\eta} \\
&= \Delta I - I_{\eta,\eta}
\end{aligned}
\tag{4.29}
$$

where $I_{\xi,\xi}$ and $I_{\eta,\eta}$ are the second derivatives of I along the tangent and orthogonal directions to the image level sets. The operator $I_{\xi,\xi}$ is equal to 0 when the gradient of the image is maximal along the orthogonal direction to the image level set. The idea is then to replace the use of the Laplacian operator with $I_{\xi,\xi}$.

Kimmel and Bruckstein (2003) show that $I_{\eta,\eta}$ corresponds to the curvature of the image level sets weighted by the image gradient amplitude, which gives

$$
I_{\xi,\xi} = \Delta I - \|\nabla I\| \mathrm{div}\left(\frac{\nabla I}{\|\nabla I\|}\right)
$$

We see that $I_{\xi,\xi}$ is expressed as the Laplacian minus a curvature term. We refer to (Kimmel and Bruckstein, 2003) for a proof of this result. Then, as for the signed Laplacian operator presented above, it is also possible to solve the contrast dependency as follows:

$$
f_{\xi s}(x) = \mathrm{sign}(\langle \nabla I, \vec{n}\rangle) I_{\xi,\xi} = \mathrm{sign}(\langle \nabla I, \vec{n}\rangle)\left(\Delta I - \|\nabla I\|\mathrm{div}\left(\frac{\nabla I}{\|\nabla I\|}\right)\right)
\tag{4.30}
$$

In Figures 4.11e and 4.11f we give two examples of this signed optimal Laplacian term with two different initialization of the model template.

SQUARED ALIGNEMENT TERM   We also propose another operator by using the square function:

$$
\oint_{\mathcal{C}}(\langle \vec{\nabla} I, \vec{n}\rangle)^2 ds = \iint_{\Omega_c}\mathrm{div}(\langle \nabla I, \vec{n}\rangle \nabla I)dv = \iint_{\Omega_c} f_{Square}(I, \mathcal{C})dv
\tag{4.31}
$$

Such a formulation tends to favor the edges with higher contrasts and those being well aligned with the evolving curve. We can see some examples in Figures 4.11g and 4.11h with two different initializations of the curve.

(a) Image I

(b) Laplacian ΔI

(c) Signed Laplacian $\text{sign}(\langle \nabla I, \vec{n} \rangle) \Delta I$

(d) Signed Laplacian $\text{sign}(\langle \nabla I, \vec{n} \rangle) \Delta I$

(e) Optimal Signed Laplacian $\text{sign}(\langle \vec{\nabla} I, \vec{n} \rangle) I_{\xi,\xi}$

(f) Optimal Signed Laplacian $\text{sign}(\langle \vec{\nabla} I, \vec{n} \rangle) I_{\xi,\xi}$

(g) Square term $\text{div}(\langle \vec{\nabla} I(x), \vec{n}(x) \rangle \vec{\nabla} I(x))$

(h) Square term $\text{div}(\langle \vec{\nabla} I(x), \vec{n}(x) \rangle \vec{\nabla} I(x))$

Figure 4.11: A toy image example (a) and the corresponding edge-based terms: signed Laplacian (c), (d), signed optimal Laplacian (e),(e) and square term (g), (h), with the zero-level of the initial template (in red).

COMPARISON OF THE CONTRAST INDEPENDENT TERMS   In Figure 4.11 we show a toy image example and the corresponding force terms presented before. Examples are given with two different template initializations. We observe that the optimal signed Laplacian highlights more the aligned edges than the signed Laplacian term. The zero-crossings seem more distinct. In comparison, the square term has a stronger effect and tends to give more importance to

edges that are well aligned with the template normals. We see hereafter the effect of this terms in case of noisy images.

4.4.1.1 *Comparison on noisy and real image*

In this section we compare the different edge-driven terms presented earlier

$$f_{SL}(\mathbf{x}) = \text{sign}(\langle \vec{\nabla} I(\mathbf{x}), \vec{n}(\mathbf{x}) \rangle) \Delta I(\mathbf{x})$$

$$f_{\zeta_S}(\mathbf{x}) = \text{sign}(\langle \vec{\nabla} I(\mathbf{x}), \vec{n}(\mathbf{x}) \rangle) \left(\Delta I(\mathbf{x}) - \|\vec{\nabla} I(\mathbf{x})\| \text{div} \left(\frac{\vec{\nabla} I(\mathbf{x})}{\|\vec{\nabla} I(\mathbf{x})\|} \right) \right) \quad (4.32)$$

$$f_{Square}(\mathbf{x}) = \text{div}(\langle \vec{\nabla} I(\mathbf{x}), \vec{n}(\mathbf{x}) \rangle \vec{\nabla} I(\mathbf{x}))$$

In a first experiment we use the toy image from Figure 4.11a to which we add noise (Gaussian noise + Gaussian filtering, peak signal-to-noise ratio (PSNR) equal to 33dB). The associated terms are given in Figure 4.12. The optimal signed Laplacian seems to be more robust to noise compared to the signed Laplacian. Indeed, the zero-crossing better aligns with the image boundaries. The square term is also more robust than the signed Laplacian but it tends to attenuate the influence of less contrasted contours. In Figure 4.13 we present a comparison on a 2D abdominal CT image (without additional noise). As seen in the previous toy example, the optimal signed Laplacian gives a more robust map than the signed Laplacian. The zero-crossing is smoother and seems to follow the image contours. The square term is also more robust but attenuates more significantly the boundaries with soft contrasts. Moreover the zero-crossing does not follow as well the image level sets.

4.4.1.2 *Conclusion*

Our experiments to compare different contrast-invariant edge terms highlight that the optimal signed Laplacian proposed by Kimmel and Bruckstein (2003) and the square term that we suggested are good candidates compared to the signed Laplacian. They both seem more robust to noise. Compared to the square alignment term, the optimal signed Laplacian may be more advantageous in some applications. Indeed, it follows more smoothly the image boundaries and attenuates less severely poorly contrasted regions. It is important to note that these contrast-invariant terms are strongly dependent on the evolving contour position. If the initialization is done outside the target object, this may have an irreversible impact on the segmentation result. Finally, compared to the Laplacian which only needs to be computed once, these terms need to be computed each time the evolving contour changes (in practice, if we assume that the template is correctly positioned at the beginning, the term does not need to be updated after each iteration).

(a) Noisy image I

(b) Signed Laplacian $f_{SL} = \text{sign}(\langle \nabla I, \vec{n} \rangle) \Delta I$

(c) Optimal Signed Laplacian $f_{\xi s} = \text{sign}(\langle \vec{\nabla} I, \vec{n} \rangle) I_{\xi,\xi}$

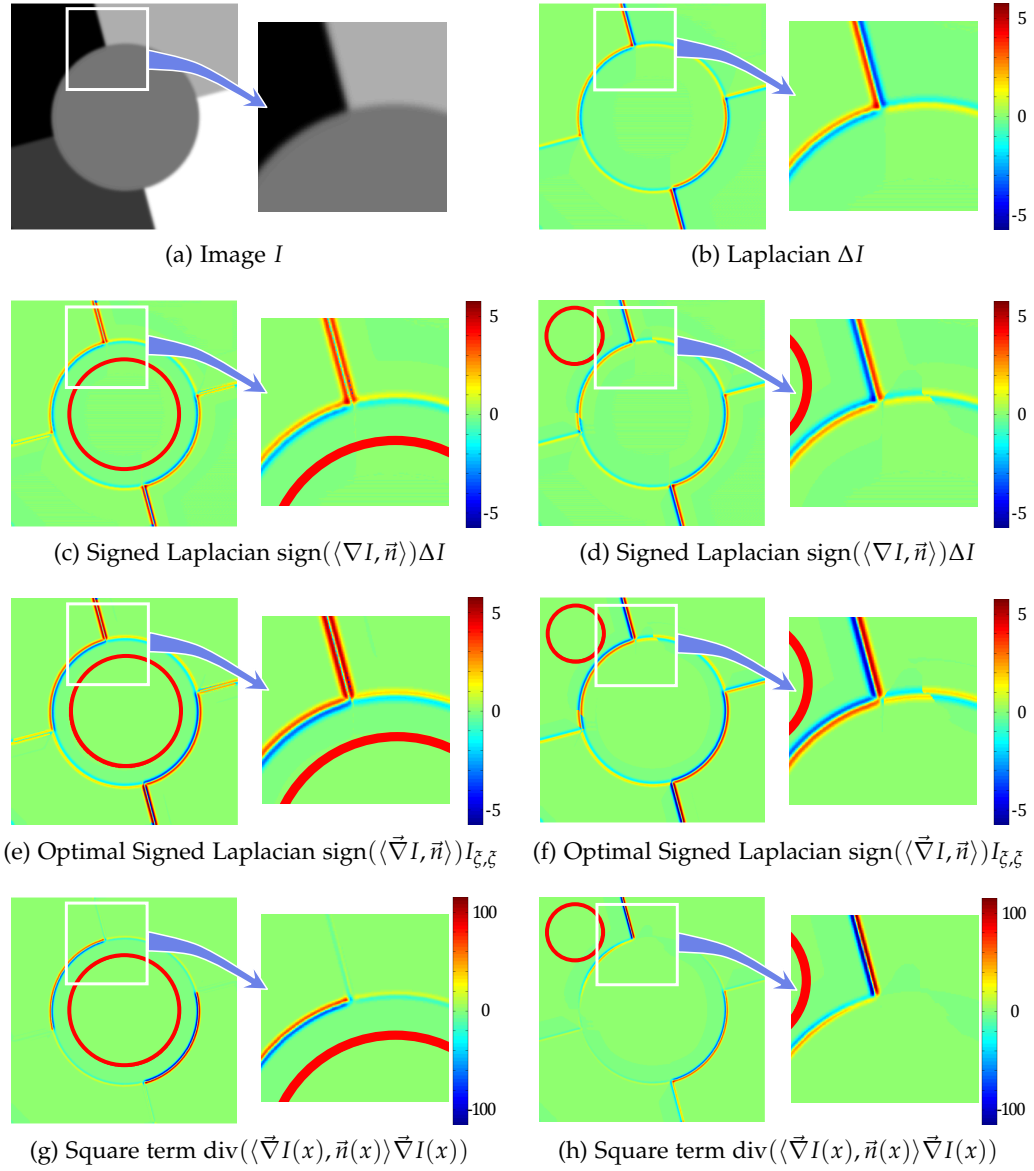(d) Square term $f_{Square} = \text{div}(\langle \vec{\nabla} I, \vec{n} \rangle \vec{\nabla} I)$

Figure 4.12: A toy noisy image example (a) and the corresponding edge-based terms: signed Laplacian (b), signed optimal Laplacian (c) and square term (d). (the zero level of the implicit template (in red))

(a) Noisy image I

(b) Signed Laplacian $f_{SL} = \text{sign}(\langle \nabla I, \vec{n} \rangle) \Delta I$

(c) Optimal Signed Laplacian $f_{\xi s} = \text{sign}(\langle \vec{\nabla} I, \vec{n} \rangle) I_{\xi,\xi}$

(d) Square term $f_{Square} = \text{div}(\langle \vec{\nabla} I, \vec{n} \rangle \vec{\nabla} I)$
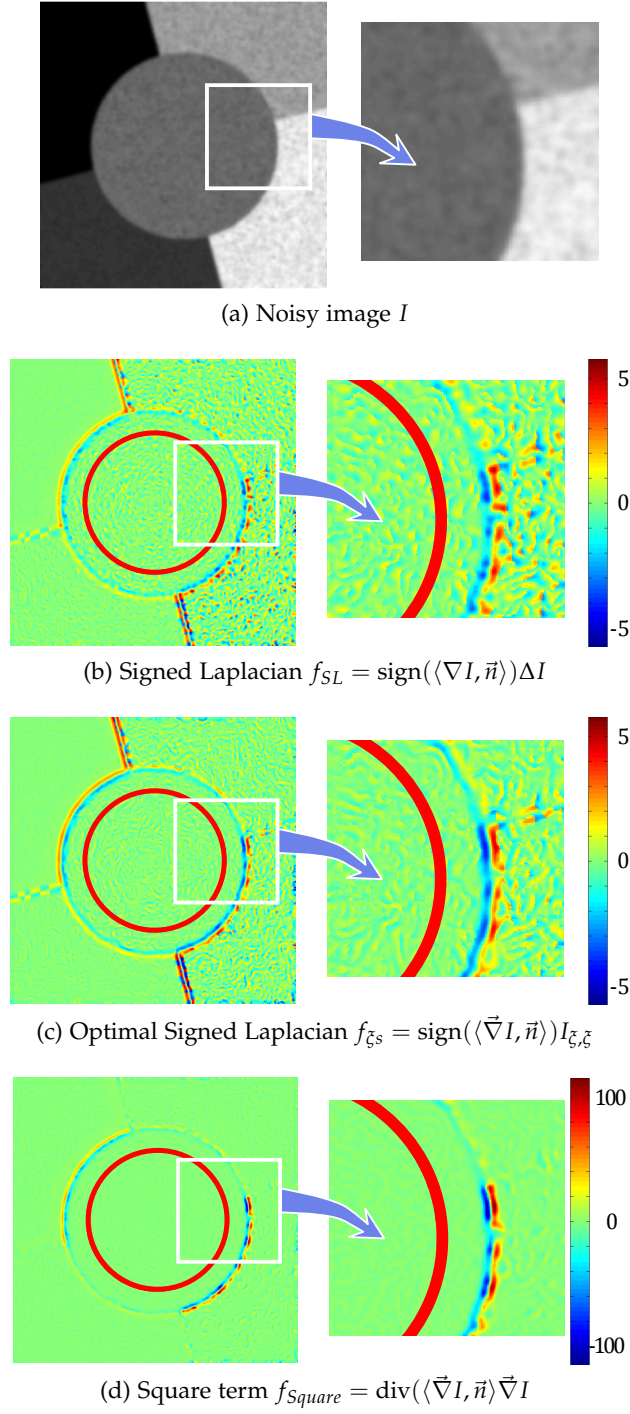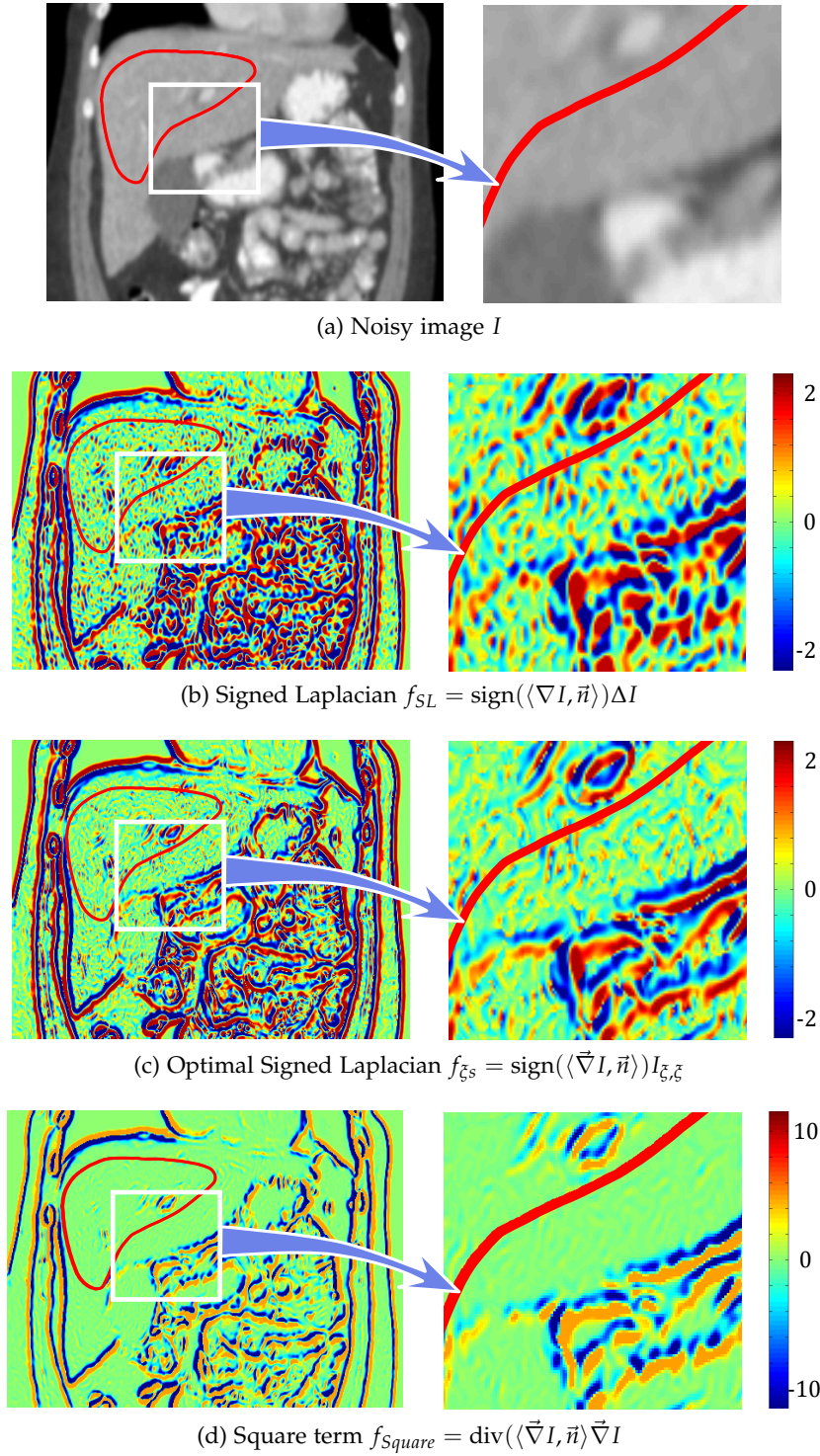
Figure 4.13: A 2D CT image (a) and the corresponding edge-based terms: signed Laplacian (b), signed optimal Laplacian (c) and square term (d), with the zero level of the implicit template (in red). Values are saturated for clearer visualization.

4.4.2 *Definition of a robust region term*

The region term $f : \Omega \to \mathbb{R}$ that we propose and that will be used in the following experiments takes into account both intensities and edges. We use a non-parametric model for the modeling of the intensities. Our region term is formulated as follows:

$$f(\mathbf{x}) = (1 - \alpha) f_{intensities}(\mathbf{x}) + \alpha f_{edge}(\mathbf{x}) \tag{4.33}$$

where $\alpha \in [0, 1]$ is a constant, and

$$f_{intensities}(\mathbf{x}) = \frac{p_{in}(I(\mathbf{x})) - p_{out}(I(\mathbf{x}))}{p_{in}(I(\mathbf{x})) + p_{out}(I(\mathbf{x}))}$$

$$f_{edge}(\mathbf{x}) = \text{sign}(\langle \nabla I, \frac{\nabla \phi(x)}{\|\nabla \phi(x)\|} \rangle) \left(\Delta I(\mathbf{x}) - \|\nabla I(\mathbf{x})\| \text{div} \left(\frac{\nabla I(\mathbf{x})}{\|\nabla I(\mathbf{x})\|} \right) \right) \tag{4.34}$$

where $p_{in}(I(\mathbf{x}))$ and $p_{out}(I(\mathbf{x}))$ are the probabilities of the intensity $I(\mathbf{x})$ inside and outside the evolving curve/surface, respectively. These probabilities can be expressed using the kernel density estimation method as

$$p_{in}(I(\mathbf{x})) = \frac{\int_{\Omega} H(\phi(\mathbf{y})) K((I(\mathbf{x}) - I(\mathbf{y})) d\mathbf{y}}{\int_{\Omega} H(\phi(\mathbf{y})) d\mathbf{y}}$$

$$p_{out}(I(\mathbf{x})) = \frac{\int_{\Omega} (1 - H(\phi(\mathbf{y}))) K((I(\mathbf{x}) - I(\mathbf{y})) d\mathbf{y}}{\int_{\Omega} (1 - H(\phi(\mathbf{y}))) d\mathbf{y}} \tag{4.35}$$

where K is a Gaussian kernel.

Thanks to the constant α, it is possible to adjust the influence of intensities or edge in the segmentation process. In Figure 4.14 we give an example of force map with different values of α. In particular, this enables to follow a global to local strategy of segmentation as done in Gauriau et al. (2013). The segmentation of the liver was performed by using the template deformation algorithm in several steps where α was increasing from about 0 to 1. The idea was to rely first on intensities to guide coarsely the segmentation and then to rely on edge so that the evolving surface reaches the liver boundaries.
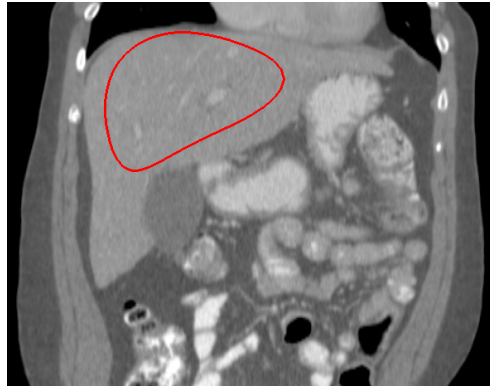
(a) CT image and template
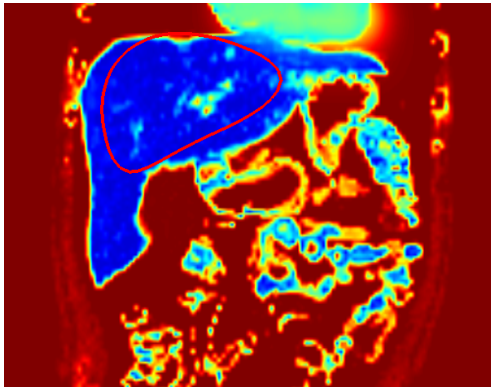
(b) Force map with $\alpha = 0$
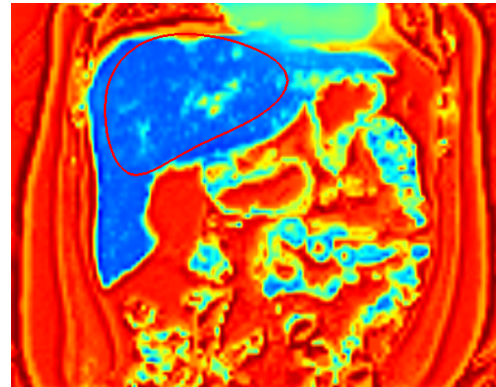
(c) Force map with $\alpha = 0.3$
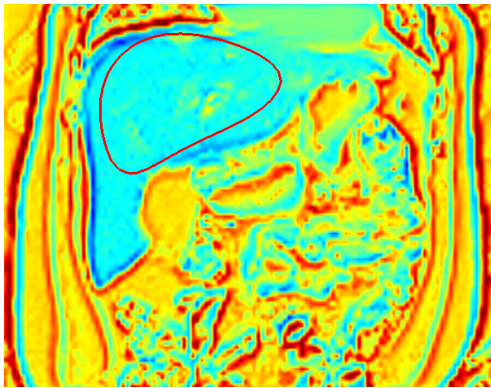
(d) Force map with $\alpha = 0.7$
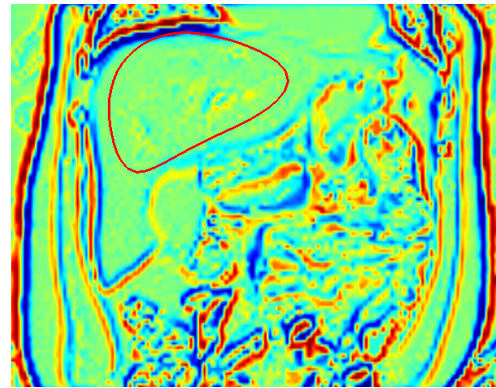
(e) Force map with $\alpha = 1$

Figure 4.14: A CT slice with liver template (in red) (a) and associated force maps with different α values (b),(c),(d),(e). (same type of colormap as in Figure 4.13)
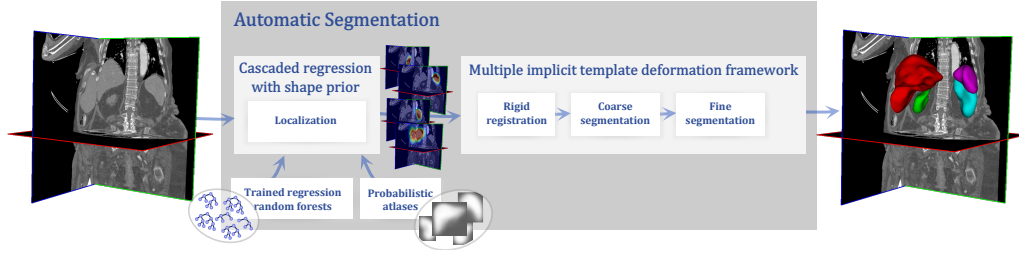
Figure 4.15: Automatic segmentation workflow.

4.5 APPLICATION TO THE AUTOMATIC SEGMENTATION OF ABDOMINAL ORGANS IN 3D CT

In this section we propose to chain the localization method presented in Chapter 3 and the segmentation framework presented just before to design a complete pipeline for the automatic segmentation of multiple organs in CT volumes. While showing the potential of the proposed multi-organ segmentation formulation, we also demonstrate that our contributions on both localization and segmentation can be linked together to get a full and fast automatic solution. Moreover, the validation presented hereafter on 106 volumes proves the accuracy of the method and its potential for a clinical application.

This section is an extension of the conference article (Gauriau et al., 2015a). Note that compared to this previous publication, we update our quantitative results on a larger database (106 versus 50 volumes).

4.5.1 *Workflow*

The method aims at segmenting multiple abdominal organs from an image. Our workflow, illustrated in Figure 4.15, is composed of two main steps:

1. organ localization with regression RF with confidence maps (method presented in Section 3.2),

2. organ segmentation with multiple template deformation (method presented in Section 4.3).

The first step returns 3D confidence maps for each organ. These maps indicate the chance of each voxel of the image to belong to the target organs. We decide to use these maps directly to initialize the segmentation algorithm. They present the advantage of being both organ- and patient-specific. Although the shape-consistency of the confidence maps is not guaranteed, we have seen that it is the case in practice (Chapter 3). By defining $T_l \in [0, 1]$ a high-confidence level

(a) CT slice with liver and heart ground truths (in red)

(b) Forces map for the liver

(c) Forces map for the liver with heart confidence map attenuation
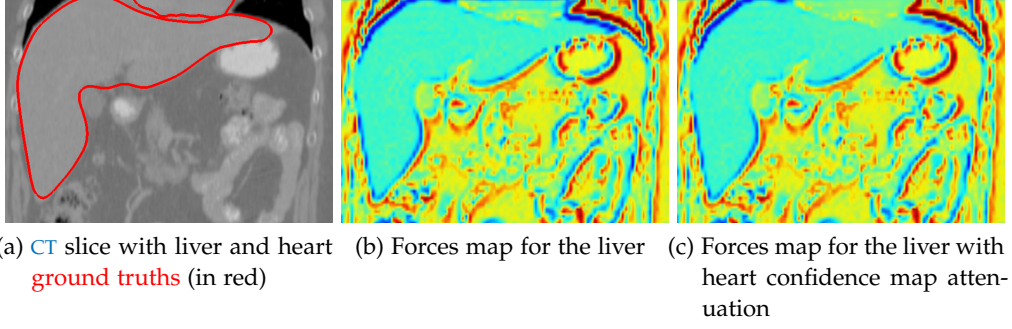
Figure 4.16: Example of forces with attenuation using confidence maps.

(T_l being close to 1), the implicit template ϕ_n of an organ n can be built from its confidence map $C_n : \Omega \rightarrow [0, 1]$ as:

$$\forall \mathbf{x} \in \Omega_n, \phi_n(\mathbf{x}) = C_n \circ \Psi(\mathbf{x}) - T_l$$

where $\Psi : \Omega_n \rightarrow \Omega$ is a similarity transformation mapping Ω_n to Ω. Then the zero level of ϕ_n corresponds to the shape model of organ n.

The segmentation formulation presented in Equation 4.12 can thus be applied using these implicit templates from confidence maps, and the associated numerical optimization under constraint with augmented Lagrangian can be processed for segmentation. In addition, the segmentation is performed in several steps, following a coarse-to-fine strategy detailed thereafter.

ENHANCED IMAGE FORCES WITH CONFIDENCE MAPS   The confidence maps may also be used advantageously in the design of the image forces f_n that are specific to each organ. Indeed, the intrinsic organs relationships information embedded in the confidence maps can contribute to more meaningful image forces.

We extend the image-driven forces presented in Section 4.4.2. If $\{C_k\}_{k \in [\![1,K]\!]}$ is a set of confidence maps of $K \leq N$ neighboring organs of the organ n, the new force term can be defined as:

$$\hat{f}_n(\mathbf{x}) = \left(\prod_{\substack{k=1 \\ k \neq n}}^{K} (1 - C_k(\mathbf{x})) \right) . f_n(x) \tag{4.36}$$

where f_n can be the force term defined in Section 4.4.2 for example.

For instance, the liver and the heart have, in general, a common interface as illustrated in Figure 4.16a. In some cases (depending on the contrast agent), they may have the same contrast and the boundary between them may be difficult to distinguish. Adding the heart confidence map information to the liver forces map may thus be a solution to prevent the liver contour from leaking into the heart region (see Figures 4.16b and 4.16c).
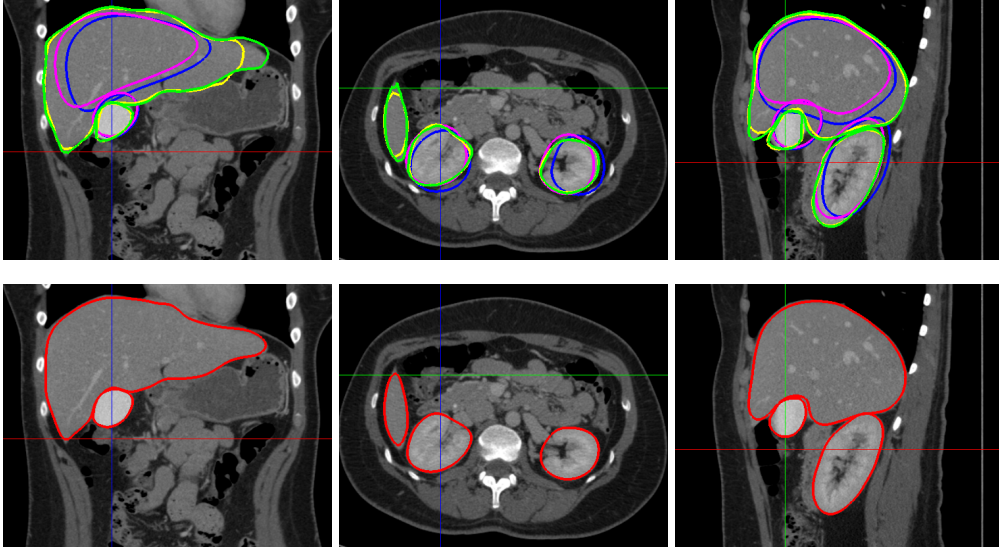
Figure 4.17: Three orthogonal views of the global-to-local 3D segmentation in CT after each step on first row (blue: localization, pink: step 1,yellow: step 2, green: step 3) and ground truth on second row. Five organs are segmented (liver, kidneys, spleen, gallbladder).

In Appendix B we show another application for liver segmentation in MRI where these enhanced image forces are used advantageously.

GLOBAL-TO-LOCAL SEGMENTATION STRATEGY   After localization, the organ templates are generated from the confidence maps and the segmentation is processed in three steps in a coarse-to-fine fashion. Figure 4.17 gives an example of segmentation results for 5 organs after each step. This method of segmenting in several steps with template deformation has been used successfully in (Gauriau et al., 2013) for the segmentation of the liver.

The first step corresponds to a rigid registration step to adjust the template positions and reduce the overlapping. The transformations ψ_n are such that $\psi_n = \mathcal{G}_n^r \circ Id$ with $\mathcal{G}_n^r$ being a rigid transformation (6 parameters: 3 translations, 3 rotations). In this step, the algorithm relies on intensity-based forces only ($\alpha = 0$). Intensity forces have two advantages at this stage. First, they are less local, smoother than edge-based ones and thus well adapted to drive the estimation of the global rigid transformation. Second, they tend to slightly under-segment the organs, resulting in contours lying mostly within the target structures. Due to the heterogeneity of the target object surroundings compared to the object himself, the expansion behavior is more likely to reach the correct local minimum than shrinking. We see in Figure 4.17 the difference between the result after localization (in blue) and after this first step (in pink).

The second step is a coarse segmentation phase optimizing both pose and local deformations: $\psi_n = \mathcal{G}_n^s \circ \mathcal{L}$. The transformation $\mathcal{G}_n^s$ is here a similarity transformation (7 parameters: 3 translations, 3 rotations, 1 scale) to adapt the template to the organ shape and size. It is initialized with the parameter values from the first step. The non-overlapping constraint avoids contours intersections. The image-driven forces are now relying on both intensities and edges ($\alpha = 0.5$), intensity-based forces aim at expanding the contours while the edge-based forces drive the contours close to the object boundaries. After this step, we get a coarse segmentation of the organs. In Figure 4.17 we see in yellow the contours after this coarse step. Although for the kidneys this step may be already satisfactory, for the liver the result remains rough and requires some refinement.

The third step aims at obtaining a finer segmentation. The poses of the objects are assumed to have converged in the previous steps and are thus fixed in this step. Only local deformations are involved in this refinement process. Moreover, the forces now rely mainly on edge ($\alpha = 0.8$). The purpose of this step is to better fit the organ contours to the target boundaries. We see in Figure 4.17 the corresponding green contours. The result for the liver is more accurate. The tip of the liver is better segmented for instance.

Note that the image forces are fixed during optimization but they are re-computed after each step given the previous segmentation results.

4.5.2 *Experiments*

To evaluate our algorithm we propose to segment simultaneously several organs (the liver, the heart, the two kidneys, the spleen and the gallbladder) in 3D CT images. Note that we do not evaluate the segmentation of the heart as it is only partially visible in most of the images. Its segmentation acts more as a help to segment the liver and prevent its contours to leak into the heart region.

4.5.2.1 *Material and Implementation*

The database includes 156 CT volumes from 130 different patients, with varied fields of view, body shapes, resolution and use or not of contrast agents. Slice and inter-slice resolutions range from 0.5 to 1 mm and from 0.5 to 3 mm, respectively. The database is split into 50 volumes for training (localization part and parameters tuning) and 106 for testing. The organs (liver, kidneys, spleen, gallbladder and heart) were segmented manually by an expert for training and evaluation purposes.

The training of the regression RF for the localization part is done using the parameters obtained in Section 3.4, as the problem is quasi identical (the only difference is the localization of the heart instead of the stomach) and the parameters optimization has been performed globally (parameters are not specific to

each organ). We use image forces enhanced with confidence maps. For the liver forces we use the heart confidence map and for the gallbladder we use the liver confidence map. These features have been fixed experimentally and according to anatomical knowledge.

The method is implemented in C++ and running times are given for a machine with four 2.3 GHz cores and 8 Gb RAM.

4.5.2.2 *Results and interpretation*

In Figures 4.18 and 4.19 we give the histograms of the Dice coefficients and mean distances compared to ground truth results for each organ and after each step of the pipeline. The histograms of each organ are represented stacked on each other. They show the progression in accuracy after each step. After localization the accuracy is widespread around 60% for the Dice coefficient. Then, after each step, the histograms shift, corresponding to better accuracy results. The Dice results do not seem to change significantly from coarse to fine segmentation, but the mean distance histograms better highlight the interest of the coarse-to-fine approach. In Figure 4.19d we observe that the last step is especially meaningful for the liver. This can be explained by its stronger inter-patient shape variability compared to the kidneys and the spleen. These results are complementary and give different types of information about the results. For the gallbladder for instance, the mean distance final results seem improper in Figure 4.19d, however we can observe in Figure 4.18d that the Dice coefficients are small in general for the gallbladder. The gallbladder is a particularly challenging organ to segment, due to its smallness and variability in shape and appearance. In our experiments, we observe that its localization is already problematic (Figure 4.18a). Subsequent steps are often not able to compensate for such large localization errors.

In Table 4.3 we give different measures and precise figures comparing the final segmentation results and the ground truth: mean distance, Dice coefficient and Hausdorff distance (i.e. maximum distance). The median results show good accuracies for the liver, the kidneys and the spleen (between 1 and 2 millimeters). Given the resolutions of our images, this corresponds to errors within a couple of voxels, which is reasonable in many clinical applications (liver resection for instance). We obtain results of the order of the best reported multi-organ segmentation methods (e.g., 2.9 and 1.35mm in average for the liver and kidneys in Kohlberger et al. (2011)). Our worst results are obtained for the gallbladder, an organ with high shape and appearance variability. High localization errors largely impede subsequent optimization.

In our experiments we evaluated the results with original force maps and with confidence maps enhancement. We did not observe significant difference on the average results. However we noticed that in some cases, the enhancement could improve the segmentation of the gallbladder.
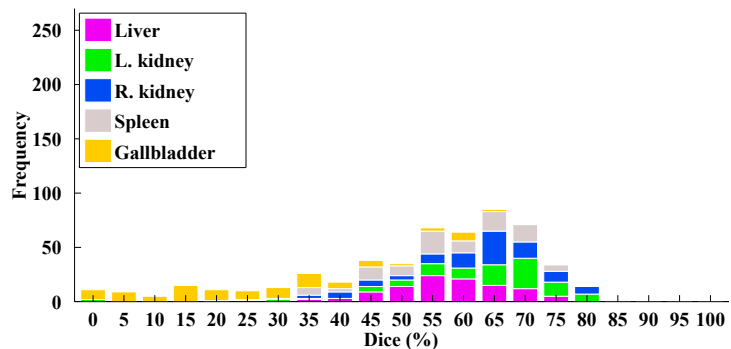
(a) Localization
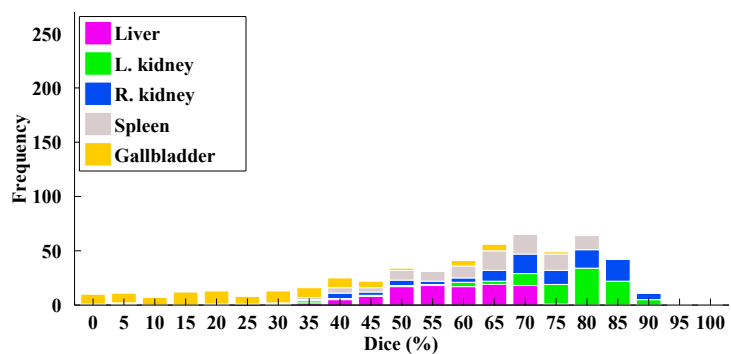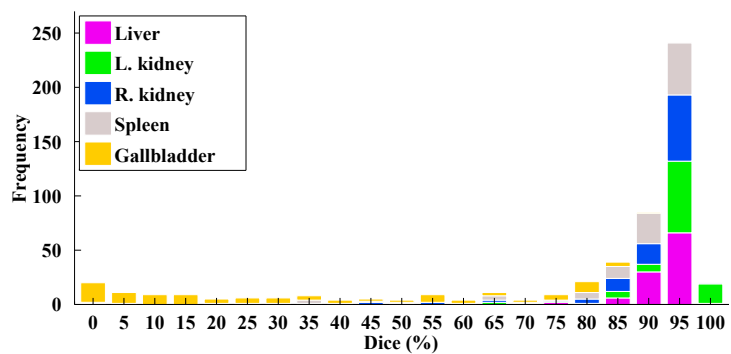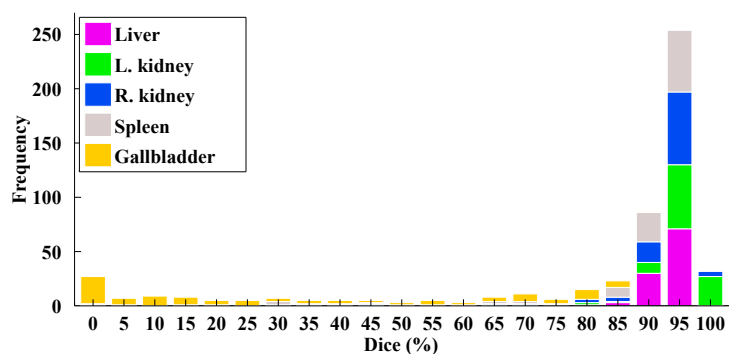
(b) Step 1

(c) Step 2

(d) Step 3

Figure 4.18: Histograms of Dice (%) results compared to ground truth after each step. Organ histograms are stacked upon each other.
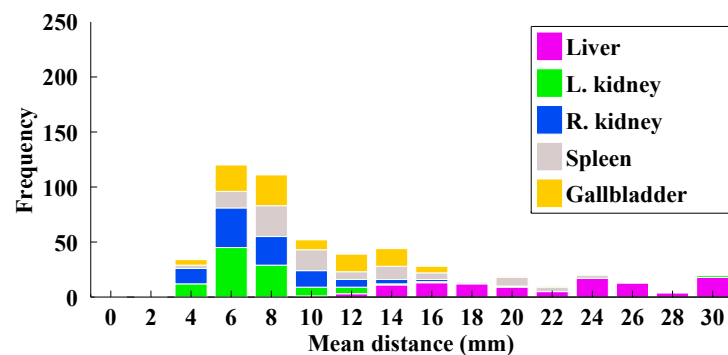
(a) Localization
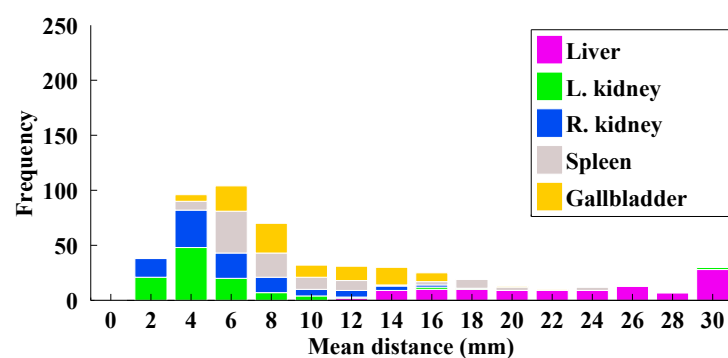
(b) Step 1
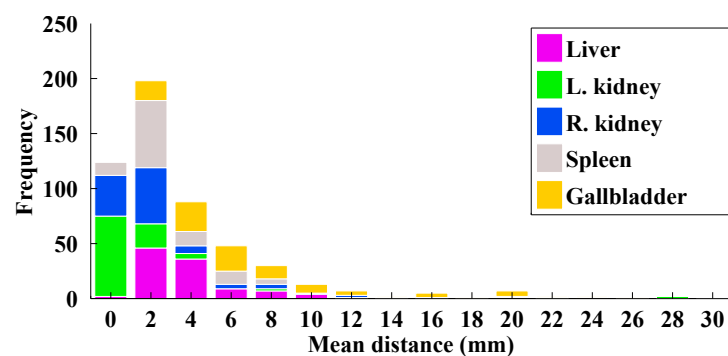
(c) Step 2
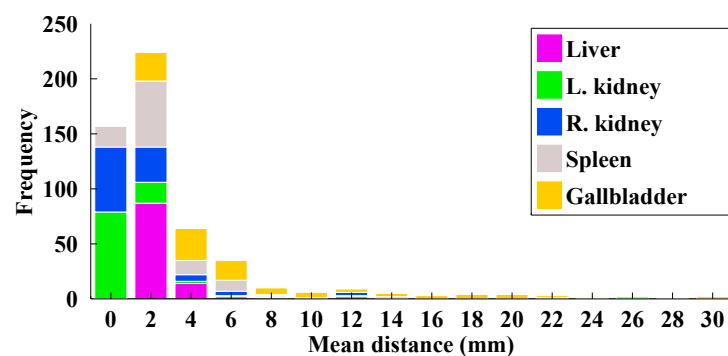
(d) Step 3

Figure 4.19: Histograms of mean distances (in mm) to ground truth after each step. Organ histograms are stacked upon each other.

Table 4.3: Final results: mean distance (mm), Dice coefficient (%) and Hausdorff distance to ground truth. (average ± standard deviation (median))

	Liver	L. Kidney	R. Kidney	Spleen	Gallbladder	**All organs**
Mean Dist.	$2.5 \pm 2(2)$	$1.7 \pm 4(1)$	$1.8 \pm 2(1)$	$2.5 \pm 3(1)$	$7.3 \pm 7(5)$	$3.2 \pm 4(2)$
Dice	$93.0 \pm 3(94)$	$91.8 \pm 16(97)$	$90.0 \pm 13(94)$	$86.4 \pm 18(93)$	$35.0 \pm 29(30)$	$80.6 \pm 26(92)$
Hausdorff	$17.7 \pm 7(17)$	$7.4 \pm 6(4)$	$9.3 \pm 5(8)$	$10.2 \pm 6(8)$	$10.2 \pm 5(10)$	$11.5 \pm 7(10)$

Table 4.4: Computation time per step (in seconds). (average ± standard deviation)

	Localization	Step 1	Step 2	Step 3	**All**
Time (s)	5 ± 0	5 ± 2	11 ± 5	11 ± 2	32 ± 9

In Table 4.4 we give the average computation times for each step. The whole process runs in about 30 seconds which makes it adapted to clinical context.

In Figure 4.20 we show some examples of results processed with and without the non-overlapping constraints. This confirms the interest of this additional constraint to the framework. In Figure 4.22 we present the best, median and worst results per organ (according to the Dice coefficient). These examples illustrate well the interpretation of the quantitative results presented before. In Figure 4.21 we present two examples of results in 3D. We notice how easier it is to visualize the organs and their relative spatial locations.

4.5.3 *Conclusion*

In this section we presented a complete automatic workflow for the segmentation of multiple organs in CT volumes. We first showed how to link our works on localization and segmentation to get a consistent pipeline. The confidence maps returned after the localization step can be advantageously used to get more consistent force maps and thus more accurate results. We also proposed a segmentation in several steps in a coarse-to-fine fashion to reach the correct local minimum more reliably. The results show the interest of such a strategy and demonstrate the accuracy of the method for all the organs, except the gallbladder. The poor results for the gallbladder are mainly explained by the difficulty to get accurate localization and by its varying appearance (no homogeneous intensities in some cases).

We believe that these results are very promising, given that we kept our method as generic as possible. In particular, we used the same type of forces for each organ and did not perform advanced parameter tuning. The entire pipeline runs in about 30s on average. Such run times are compatible with most clinical workflows, such as liver surgery planning for instance.
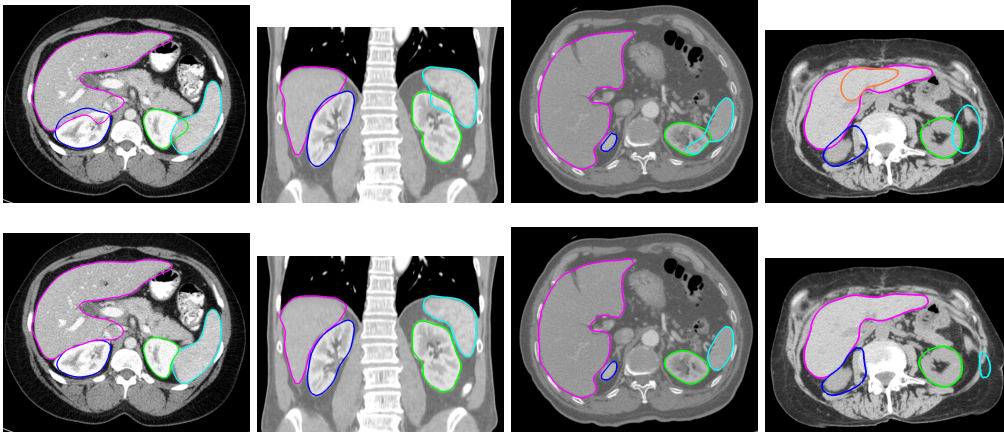
Figure 4.20: Examples of results without and with non-overlapping constraint (first and second rows, respectively). Liver (in pink), right kidney (in blue), left kidney (in green), spleen (in cyan) and gallbladder (in orange).
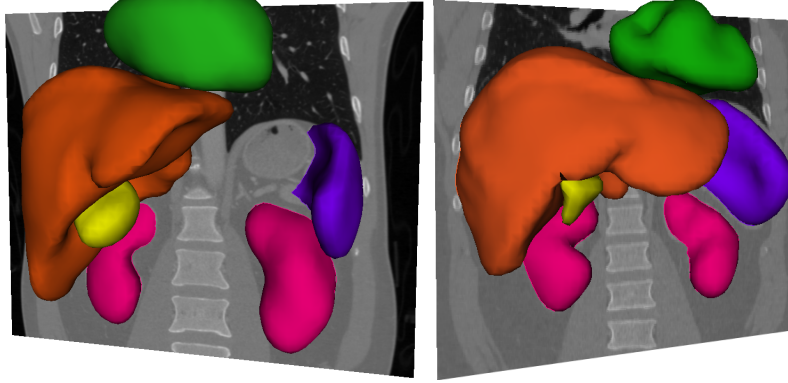
Figure 4.21: Examples of segmentation results in 3D.

4.6 CONCLUSION AND DISCUSSION

4.6.1 *Contributions*

In this chapter we presented a new variational framework for the segmentation of multiple organs. This approach extends the work of Mory et al. (2012) on implicit template deformation to multiple organs. By estimating organ-specific global transformations and one common diffeomorphic transformation for local deformations we ensure consistent deformations. The introduction of non-overlapping constraints acting on the organ-specific global transformations guarantees that the organs do not intersect. We also proposed optimizations leading to a fast implementation (around 30s for 6 organs on 3D CT data). We also presented a contrast-invariant force term based on edge information (using the work of Kimmel and Bruckstein (2003) on regularized zero-crossing) and on
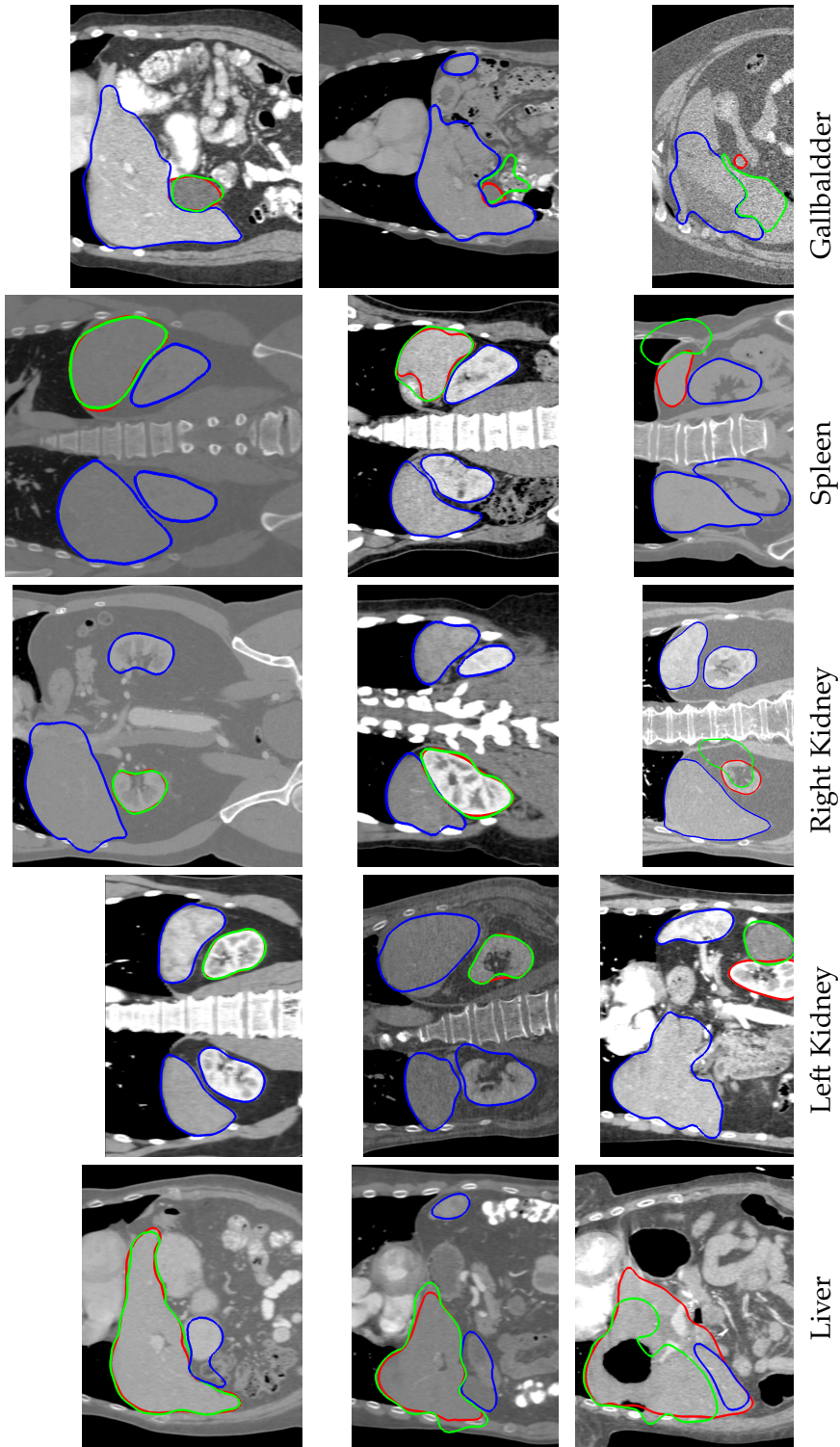
Figure 4.22: Best, median and worst results for each organ according to Dice coefficient (respectively first, second and third lines). Ground truth (red), segmentation result of the organ (green), segmentation results of other organs (blue).

region intensities information. This generic force term can be used in our multiple template deformation framework and reinforce its generic aspect. Indeed, in our experiments, the same parameters were used for all the organs. Experiments in 2D and 3D showed the interest of the approach, whether in terms of accuracy, robustness and speed. We showed that the proposed segmentation method can be chained with our localization method (Chapter 3) in a straightforward manner, resulting in a complete pipeline for automatic, multi-organ segmentation. The quantitative and qualitative evaluation on a database of 106 CT volumes show state-of-the-art results. We believe our framework achieves an interesting balance between accuracy, robustness and computational speed. Furthermore, we show in the following chapter (Chapter 5) how to introduce user feedback in the framework to allow for simple result corrections.

4.6.2   *Limits and perspectives*

Our approach is dedicated to structures with fixed-topology, not interlocked and for which we can define a template. The segmentation of tissues such as fat or muscles would be challenging to segment with our method., due to the difficulty of defining a sensible template. The segmentation of the liver and its inner tumors would also not be feasible, as this would contradict our non-overlapping constraints. This would necessitate to define other constraints than the non-overlapping constraints that we proposed. The addition of other types of constraints (e. g. inclusion and adjacency) would be an interesting extension of our framework.

The definition of a template is sometimes problematic, e. g. for elongated structures or highly deformable tissues. For example, the intestine has a shape that varies too much from patient to patient to be able to define a common mean shape. However, note that our segmentation framework can be used with any initialization template. A coarse segmentation obtained with another method could be used as a template for our algorithm. The shape models may also be enhanced by incorporating shape variabilities as in the work of Prevost et al. (2013a).

Finally, the success of the segmentation depends highly on the initial position of the templates. Using a robust localization method as we did greatly alleviates the issue. Another approach could be to define new image-driven forces, independent (or less dependent) of the template positions. Introducing new information in the forces with voxel-based classification may be a first step in that direction.

<div style="text-align: right; font-size: 3em;">5</div>

INTERACTIVE SEGMENTATION WITH MULTIPLE TEMPLATE DEFORMATION

Abstract

In this chapter we introduce user constraints in the multiple template deformation formulation (Chapter 4). The goal is to allow the user to correct automatic segmentation results with simple interactions. By taking advantage of the original formulation, we are able to get coherent deformations. After presenting the constrained variational formulation, we propose a new energy formulation with an intuitive, local behavior. By this mean, the impact of user interactions is limited to a local range.

A user study shows the potential of the approach for the coherent correction of multiple segmentations. The proposed approach provides a user-friendly tool to correct automatic segmentation results in a few minutes.

A part of this work has been published in Gauriau et al. (2015b).

Contents

5.1   INTRODUCTION

5.1.1   *Context*

Despite the recent improvements of fully automatic segmentation approaches, perfect accuracy remains unreachable in many image processing scenarios. It is especially noteworthy when inter- and intra-patient variabilities are important. In a clinical context, the possibility of adding user feedback thus appears a necessity. As explained by Olabarriaga and Smeulders (2001), the introduction of user interactions is not straightforward. From the user point of view, the requirements are the following:

- the interactions should be easy to perform (only few keyboards or mouse button used, a limited number of user interactions to get the final results);

- the impact of the interaction should be displayed quickly;

- the algorithm should be *clever* enough to meet the user's needs.

Such conditions necessitate technological solutions that are fast (almost real-time) and that find an intuitive trade-off between the user interactions and the image information. In the case of organ segmentation, the algorithm should find the solution that both fulfills the user constraints and is consistent with the image information such as the organ shape and the contrasts. Moreover, in the case of multiple segmentation corrections, new issues appear. How to preserve non-overlapping between the objects? How to get consistent deformations when objects are close to each other? The formulation proposed in this chapter aims at answering these problems.

In Section 5.1.2, we give an overview of the literature on interactive segmentation methods, with a specific focus on medical imaging context. In Section 5.1.3 we present the work of Mory et al. (2012) on real-time segmentation with implicit template deformation. Following this work, we propose in Section 4.3 an extension of our multi-object variational framework presented in Section 4.3 to take into account user interactions. Finally, we present a user study in Section 5.3 to evaluate our approach in a clinical-like context.

NOTATIONS   We remind the notations used hereafter:

- d is the image dimension, $d = 2$ for 2D images and $d = 3$ for 3D volumes,

- I is an image defined as $I : \Omega \rightarrow \mathbb{R}$,

- $\Omega \subset \mathbb{R}^d$ is the image domain,

- $\Omega_n \subset \mathbb{R}^d$ is the domain of template n,

- $\mathbf{x} \in \mathbb{R}^d$ is a point,

- ϕ denotes an implicit function, whose zero-level corresponds to the contour and which is positive inside the contour and negative outside,

- $\mathbf{u} : \Omega \rightarrow \mathbb{R}^d$ is a vector field, usually defined in the image domain,

- ∇ is the gradient operator in $\mathbb{R}^d$,

- Δ is the Laplacian operator in $\mathbb{R}^d$, such that $\Delta = div(\nabla)$,

- $H : \mathbb{R} \rightarrow [0,1]$ is the Heaviside step function, if $x \leq 0$ then $H(x) = 1$ and if $x > 0$ then $H(x) = 0$,

- δ is the dirac function,

- Id is the identity transformation,

- $\circ$ is the functional composition,

- $*$ denotes the convolution.

5.1.2 *Background on interactive segmentation*

Segmentation methods are often split into three categories according to the required user interactions. Automatic methods do not require any user intervention, while semi-automatic methods require limited user interaction (a click or a stroke often used for the algorithm initialization) and interactive methods are more demanding (several user interactions). In this section we focus on interactive segmentation approaches, although the frontier with semi-automatic approaches can be sometimes fuzzy. Most of the methods that we can find in the computer vision literature are dedicated to a single object and 2D as we will show hereafter.

GUIDED MANUAL SEGMENTATION   There are two possible ways of segmenting manually an object. The user may color the pixels (or voxels) belonging to the target object, or he can draw the contour separating the object from the background. Such operations can be computer-assisted. The *live wire* technique (Mortensen et al., 1992), for instance, aims at assisting the user with the delineation of the boundaries of the object of interest, in 2D. Between pairs of user-clicked contour points, the algorithm extrapolates the boundary using a minimal path algorithm. In the case of clear, well-contrasted contours, such a technique gives good results. However if the contrasts are poor, the algorithm may produce false paths.

The extension to 3D is not straightforward. A possibility, introduced by Falcao and Udupa (1997), is to use the original 2D live wire technique in several slices and then to approximate the final 3D volume. Improvements have also been proposed to accelerate the algorithm and make it usable in real-time (Falcao

Table 5.1: Graph-based algorithms generalization (Source: Couprie et al. (2009)).

q \ p	0	finite	∞
1	reduction to seeds	Graph cuts	Maximum Spanning Forest (Watershed)
2	l_2-norm Voronoi	Random walker	Power watershed $q = 2$
∞	l_1-norm Voronoi	l_1-norm Voronoi	Shortest Path Forest

et al., 2000). Even if some efforts have been done to reduce the user interactions (Hamarneh et al., 2005), such a method remains tedious and time-consuming.

GLOBAL OPTIMIZATION FROM USER INITIALIZATION   Another strategy is to ask the user to give some initialization and then to launch the segmentation algorithm.

In general, the user draws some strokes or points in the background and in the foreground and then, according to this information and to the image information, the algorithm finds the best partition of the image. Multiple methods solve this problem through an energy minimization via discrete optimization. In particular, a general formulation was proposed by Couprie et al. (2009) to describe a broad family of such methods:

$$\min_x E_{p,q}(x) = \min_x \underbrace{\sum_{e_{i,j} \in \mathcal{E}} w_{i,j}^p |x_i - x_j|^q}_{\text{smoothness term}} + \underbrace{\sum_{v_i} w_{F_i}^p |x_i|^q + \sum_{v_i} w_{B_i}^p |x_i - 1|^q}_{\text{data fidelity term}}$$

$$\text{subject to} \begin{cases} x_i = 1, & \text{if } v_i \in F \\ x_i = 0, & \text{if } v_i \in B \end{cases}$$

(5.1)

where x is the target configuration (the final segmentation being given by $s = \{s_i | s_i = 1 \text{ if } x_i \leq 0.5 \text{ or } 0 \text{ otherwise}\}$, $(p,q) \in \mathbb{N}^2$, $G = (V, \mathcal{E})$ is a graph with vertices $v \in V$ and edges $e \in \mathcal{E}$, $w_{i,j}$ is the weight of the edge $e_{i,j}$ between the vertices v_i and v_j, w_{F_i} and w_{B_i} are the unary weights penalizing foreground and background affinity at node v_i. This formulation unifies major graph-based image segmentation approaches including maximum spanning forests (MSF) (whose a well-known special case are *watersheds*), graph-cuts (see short bibliographic review in Section 2.4.3), random walker (Grady, 2006) and shortest paths algorithms. Table 5.1 gives an overview of the methods according to particular values of the parameters p and q.

There are multiple ways for defining a watershed (Beucher and Meyer, 1992). Among other formulations, Cousty et al. (2009) show that the watershed algorithm can be formulated as a MSF. A MSF is a generalization of the minimum spanning tree for unconnected graphs: each component of the graph corresponds to a minimum spanning tree. In the case of segmentation, each tree

is connected to one seed component (user input) and the weight of the set of trees should be maximum. The separations (or cuts) of these trees correspond to the segmentation result (and equivalent to watersheds result). Although watersheds are commonly used in image segmentation, they may sometimes suffer from over-segmentation.

Graph-cuts, introduced by Boykov and Jolly (2000), are very popular for interactive image segmentation. As we already gave literature background on graph-cuts in Section 2.4.3, we propose here to focus on improvements of this method that are specific to interactive segmentation. Kéchichian et al. (2011) propose to accelerate the computations by performing a voxel clustering (with centroidal Voronoi diagram clustering algorithm) and then to apply the graph-cut on these super-pixels. The computation takes several minutes. Jacinto et al. (2012) use this approach in a 3D web interface. The graph-cut algorithm is however sensible to the seed quantity and it tends to favor the smaller cuts, which may be unfavorable for elongated structures. The solution of normalized cut to reduce the shrinking bias is computationally expensive (Shi and Malik, 2000) and thus not applicable for interactive purposes. Price et al. (2010) propose another solution by adding geodesic confidence to the edge information in the graph-cut framework. This adds a computational cost and this method has only be tested in 2D color images. Rother et al. (2004) enhance the graph-cut algorithm for segmentation with simplified interactions. Their method, called *GrabCut*, only requires from the user to draw a bounding box around the object of interest. This method combines a more powerful and iterative optimization and the use of *border matting* algorithm to improve the boundaries detection.

On the other hand, the random walk algorithm, first introduced by (Grady, 2006), tends to favor average cuts. Zhang et al. (2010a) show that this algorithm can actually be seen as a special case of a more generic segmentation framework based on anisotropic diffusion. Yang et al. (2010) propose to enhance the random walk framework by integrating soft and hard constraints. For instance, instead of drawing strokes inside and outside the target object, the user can also add strokes on the target boundary.

The principle of shortest paths algorithms (or geodesics) is to assign foreground label to the pixels that can be connected to the foreground seed with shortest path than to any background seed. Paths are weighted using the image content in the same manner as with graph-cuts and random walker methods. This approach, popularized by Bai and Sapiro (2007), gives better results for elongated structures and requires simpler scribbles than with graph-cut and random walker but they can be more sensible to noise and seeds position. Variants of this idea have also appeared in the works of Falcão et al. (2004); Alvino et al. (2007); Criminisi et al. (2008). For instance, Criminisi et al. (2008) use the geodesic transform for the spatial regularization. However the geodesic filter acts directly on the energy unaries and not on the user strokes. Criminisi et al. (2008) show that a careful implementation, including parallelization and

memory optimization, can considerably accelerate processing times, making geodesic methods suited for highly interactive tasks.

Graph-cuts, random walks and geodesic approaches have a common drawback: they are not very robust to initialization changes. If the user scribbles differently from another user, the segmentation result may be different. Nguyen et al. (2012) propose a more robust solution. They use the result of a geodesic or random walk algorithm as an initialization step for a convex active contour algorithm (Bresson et al., 2007). They demonstrate more robust results to user input.

The image partitioning from user inputs can be also formulated as a continuous variational problem. Zhao et al. (2010) use a continuous convex formulation of the image partitioning. By modeling the foreground and background regions with discriminative learning they can exploit richer image features and get better accuracy. Compared to graph-based optimization algorithms generally limited to local, pairwise information, this type of framework can more easily exploit more global and contextual information.

The type of methods listed in this section can be sometimes frustrating for the user. These algorithm can take new user inputs but do not take into account previous segmentation results. They proceed by relaunching the optimization with a different initialization. For difficult images, the user may be required to add so many corrections that a fully manual segmentation would be more desirable. This type of problem is rather frequent in medical imaging, as the target boundaries may often be intricate. More sequential methods seem thus more appropriate.

SEQUENTIAL SEGMENTATION WITH USER INPUTS   Gulshan et al. (2010) propose a solution designed to reduce the user effort. They use the star-convexity model and geodesic paths to represent the set of possible shapes. This shape constraint is embedded in the same graph-cut energy formulation as in the work of Boykov and Jolly (2001). The set of possible shape contours is then reduced to the subset of star-convex shapes. The authors also add two properties allowing for more intuitive interactions: (i) the current segmentation should remain partially valid after a user edit, (ii) the shape constraint should only change locally, close to the new user input. They give an extensive evaluation and comparison with other approaches with regards to the user effort (number of strokes). About 10 strokes are required in average to get a correct segmentation result. Computation times are not given.

Another way of considering the problem is proposed by Cremers et al. (2007). They extend the level set segmentation framework (see Section 2.4.5.1) to take into account external user constraints. The method consists in adding a new term in the energy formulation that penalizes non verified constraints. The probabilistic image-driven forces are also modified according to the user inputs. In practice, the user clicks on a location to include or exclude from the contour

and then waits a few seconds so that the contour evolves and converges to a new minimum. Such a formulation implies to specify two additional parameters controlling the strength and the spatial range of the user constraints. Ben-Zadok et al. (2009) also propose to modify the level set formulation by adding a new term to the functional. The results presented only show corrections very close to the initial contour.

In such sequential approaches, the computation time is crucial, as the user is waiting after each interaction to examine the new result and proceed to further corrections.

Mory et al. (2009) propose a technique relying on the minimization of an energy over a low-dimensional subset of possible implicit functions spanned by non-Euclidean *Radial Basis Functions*. These functions are built upon image features such as intensities or edge information. The authors show that they can take advantage of the non-convexity of their energy minimization formulation. Indeed, the user interactions will lead to a specific local minimum of the objective function. The responsiveness of their implementation also allows the user to see the impact of his interactions in real-time, to make the appropriate corrections rapidly and thus drive the algorithm to the correct local minimum.

The optimization of non-convex problems is also exploited in the work of Mory et al. (2012) (detailed in Section 5.1.3 and from which we derive the method presented in this chapter). The user interactions are integrated as hard constraints into the segmentation framework with template deformation. The proposed solution gives real-time feedback to the user which allows him to correct a wrong segmentation result rapidly.

This type of methods works sequentially in such a way that each new user interaction is supposed to improve the current segmentation result. However, one underlying problem is that the optimization is still solved globally. This means that a new user input at a location x does not guarantee that the contour at a location y far from x will not be affected. This may result in counterintuitive behaviors, as the user generally expects his corrections to have only a local impact.

LOCAL USER CORRECTIONS   To the best of our knowledge, the scientific literature on purely local corrections is scarce, especially for 3D segmentation. Kang et al. (2004) introduce different tools for 3D contour editing: hole-filling, point-bringing and surface-dragging. These tools can be cumbersome to use, mainly because of their complex parameterization. Grady and Funka-Lea (2006) exploit the graph-cut algorithm locally and take into account both the image information and the initial segmentation. Reported computation times remain relatively long for interactive setups (several seconds). Harrison et al. (2013) use the random walker algorithm in a sub-volume around the user corrections. The user draws a stroke along the correct contour in a slice of the 3D volume. Then the algorithm learns the image features, classifies the voxels in the sub-volume and

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

(b) $\phi = \phi_0 \circ \psi$ satisfying the user constraints
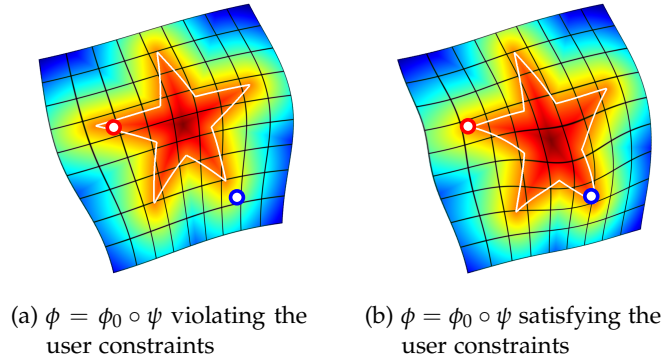
Figure 5.1: Illustration of the user constraints principle. Blue and red dots indicate user constraints that should lie inside and outside the final segmentation. (Source: Mory (2011)).

the random walk algorithm is run. With a GPU implementation, results are obtained rapidly (in 3 seconds) but fall short of real-time update rates. The qualitative comparison only shows cases where the contrasts of the ground truth boundaries are strong. A closer approach to our work, proposed recently by Barbosa et al. (2014), extends the B-spline active surface segmentation framework (Barbosa et al., 2012) with user interactions. The method is designed for the interactive correction of automatic segmentation results. It consists in adding a new term in the original energy formulation presented in (Barbosa et al., 2012). The surface deforms locally according to the user inputs. The range of action is controlled by the size of a regularization kernel. Benefiting from a real-time implementation, this method seems to give intuitive interactions and allows the user to perform the corrections rapidly.

We identified various types of approaches for interactive segmentation, with different workflows and mathematical formulations. We found however very few methods dedicated to multi-object corrections, even less in 3D. Even if automatic methods become more and more accurate, the possibility of editing the result appears necessary, especially in a medical context. The introduction of user constraints into our segmentation framework (Chapter 4) aims at providing an efficient, user-friendly solution to this problem.

5.1.3 *Implicit template deformation framework with user interactions*

Mory et al. (2012) show that the template deformation framework (presented in Section 4.2) can be enhanced to handle user interactions. The idea is to add constraint points that should lie inside or outside the object curve/surface.

The user constraint points are denoted by $\{\mathbf{x}_k\}_{k\in[\![1,K]\!]} \subset \Omega^K$. They can be mapped to the template domain to know their position with regards to the evolving contour. Each point is associated with a variable γ_k that takes value 1 or -1 whether the point should lie inside or outside the contour respectively. As ϕ_0 is positive inside and negative outside its zero-level, the constraints can be formulated as follows:

$$\forall k \in [\![1, K]\!], \gamma_k \phi_0 \circ \psi(\mathbf{x}_k) \geq 0 \tag{5.2}$$

Figure 5.1 illustrates the principle of violating and satisfying constraints.

The segmentation problem of Equation 4.6 then becomes a constrained optimization problem that can be written as:

$$\min_{\mathbf{p},\mathbf{u}} \left\{ E[\mathbf{p}, \mathbf{u}] = \int_{\Omega} H(\phi_0 \circ \psi_{\mathbf{p},\mathbf{u}}) f(\mathbf{x}) d\mathbf{x} + \frac{\lambda}{2} \|\mathbf{u}\|_U^2 \right\}$$
$$\text{subject to } \gamma_k \phi_0 \circ \psi_{\mathbf{p},\mathbf{u}}(\mathbf{x}_k) \geq 0, \forall k \in [\![1, K]\!], \tag{5.3}$$
$$\text{with } \psi_{\mathbf{p},\mathbf{u}} = \mathcal{L} \circ \mathcal{G}, \mathcal{L} = Id + \mathbf{u}, \|\mathbf{u}\|_U^2 = \langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\Omega,\mathbb{R}^d)} = \langle \mathbf{u}, K_\sigma^{-1}\mathbf{u} \rangle_{L^2(\Omega,\mathbb{R}^d)}$$

This new energy formulation is optimized following a general augmented Lagrangian methodology (Nocedal and Wright, 2006) consisting in turning the problem into a series of unconstrained problems. We refer the reader to Mory et al. (2012) to know more about the optimization scheme. Note that we use this type of optimization approach in Section 5.2.

REMARK Mory et al. (2012) limit the optimization to one unconstrained energy minimization by taking a large enough penalty term. This is actually equivalent to turn the hard constraints into soft ones. Although this may work well in practice, it should be done carefully as it may induce instable results, especially when the user constraints are far from the evolving curve/surface.

5.2 ADDING USER INTERACTIONS TO THE MULTIPLE TEMPLATE DEFORMATION FRAMEWORK

Following the work of Mory et al. (2012) on interactive segmentation (presented in Section 5.1.3), we propose to extend the segmentation framework presented in Section 4.3 to take into account user interactions. To do that, we extend the method given by Mory et al. (2012) to our multi-object framework. User-defined constraints take the form of inside or outside labeled control points. An additional requirement of our multi-object setup is the need to identify on which organ a constraint should apply. Similarly to the overlapping constraints of Section 4.3.2 and to the work of Mory et al. (2012), the problem can be written

(a) Inside constraint is added to the star

(b) The objects deform so that the constraint is verified
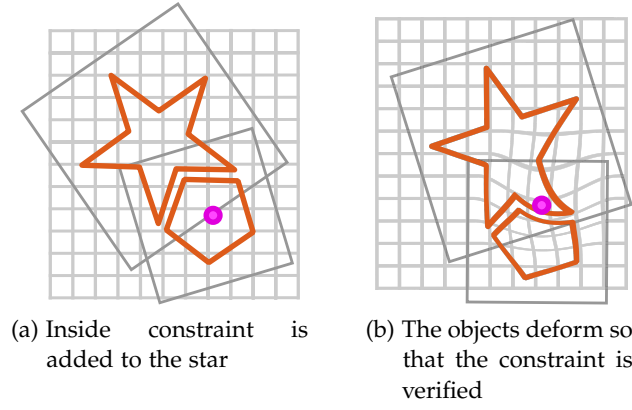
Figure 5.2: A constraint is added to the star shaped object (a). The objects are deformed so that the constraint gets verified (b).

as an optimization under constraint. We extend Equation 4.12 and write the minimization problem as follows:

$$\min_{\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}} E[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}]$$

$$\text{where } E[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}] = \sum_{n=1}^{N} \left(\int_{\Omega} H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2$$

subject to

$$\forall(i,j) \in [\![1,N]\!]^2, i < j, C(\mathcal{G}_i,\mathcal{G}_j) = \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x}))d\mathbf{x} = 0,$$

$$\forall n \in [\![1,N]\!], \forall q_n \in [\![1,K_n]\!],\ \gamma_{q_n}\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n}) \geq 0,\ \gamma_{q_n} \in \{-1,1\}$$

$$(5.4)$$

where :

- K_n is the number of user constraints for the organ n,

- $\{\mathbf{x}_{q_n}|\forall n \in [\![1,N]\!], \forall q_n \in [\![1,K_n]\!], \mathbf{x}_{q_n} \in \Omega\}$ is the set of user constraints,

- γ_{q_n} takes value -1 if the control point should be outside the contour of organ n and value 1 if it should be inside.

Suppose that the user adds a point at a location $\mathbf{x}_{q_n}$ outside the organ contour (defined by the zero level of the function $\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})$) and that he wants this point to be inside the organ ($\gamma_{q_n} = 1$). Before any optimization, $\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n})$ is strictly negative as the point $\mathbf{x}_{q_n}$ is outside the contour and the constraint $\gamma_{q_n}\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n}) \geq 0$ is not verified. The optimization framework must find the transformations $\mathcal{G}_n$ and $\mathcal{L}$ that deform all the object contours so that the constraint point $\mathbf{x}_{q_n}$ ends up inside the warped contours of object n. Figure 5.2 gives an illustration of this example.

5.2.1 *Numerical optimization*

Contrarily to the constrained problem that we have seen earlier for the non-overlapping constraint (Equation 4.12), the problem of Equation 5.4 includes inequality constraints. This type of problem can also be turned into an unconstrained formulation with the augmented Lagrangian method. However the numerical scheme cannot be the same, as the one given in Equation 4.18 is specific to equality constraints. Nocedal and Wright (2006) give a numerical optimization scheme to solve non-linear problems under inequality constraints. To be able to use this method, we propose to transform the equality constraints corresponding to non-overlapping terms into inequality constraints. This can be easily done by considering the positivity of the Heaviside function. Equation 5.4 becomes:

$$
\min_{\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}} E[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}]
$$

$$
\text{where } E[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}] = \sum_{n=1}^{N}\left(\int_{\Omega} H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x}\right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2
$$

$$
\text{subject to}
$$

$$
\forall(i,j) \in [\![1,N]\!]^2, i < j, -C(\mathcal{G}_i,\mathcal{G}_j) \geq 0,
$$

$$
\forall n \in [\![1,N]\!], \forall q_n \in [\![1,K_n]\!], \ F_{q_n}(\mathbf{x}_{q_n}) \geq 0
$$

$$
\text{where}
$$

$$
C(\mathcal{G}_i,\mathcal{G}_j) = \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x}))d\mathbf{x},
$$

$$
F_{q_n}(\mathbf{x}_{q_n}) = \gamma_{q_n} \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n}), \ \gamma_{q_n} \in \{-1,1\}
$$

$$(5.5)$$

Following the indications of Nocedal and Wright (2006), we can transform this constrained optimization problem into a series of unconstrained minimization problems of this form:

$$
\min_{\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}} \hat{E}_k[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}]
$$

$$
\text{where}
$$

$$
\hat{E}_k[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}] = \max_{\substack{\lambda \geq 0 \\ \lambda' \geq 0}}\left\{ E[\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}] - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}}\alpha_{i,j}C(\mathcal{G}_i,\mathcal{G}_j) - \frac{1}{2\mu_k}\sum_{\substack{1 \leq i \leq N \\ i < j \leq N}}(\alpha_{i,j} - \alpha_{i,j}^k)^2 \right.
$$

$$
\left. - \sum_{\substack{1 \leq n \leq N \\ 1 \leq q_n \leq K_n}}\alpha'_{q_n}F_{q_n}(\mathbf{x}_{q_n}) - \frac{1}{2\mu'_k}\sum_{\substack{1 \leq n \leq N \\ 1 \leq q_n \leq K_n}}(\alpha'_{q_n} - \alpha'^k_{q_n})^2 \right\}
$$

$$(5.6)$$

We denote:

- α the vector composed of the Lagrange multipliers $\alpha_{i,j}$ for the non-overlapping constraints,

- α' the vector composed of the of Lagrange multipliers α'_{q_n} for the user constraints,

- μ_k and μ'_k the penalty parameters for the non-overlapping and user constraints respectively.

Nocedal and Wright (2006) show that the maximization of the energy of Equation 5.6 can be performed explicitly and that the Lagrange multipliers update can be done as follows:

$$\alpha_{i,j}^{k+1} = \begin{cases} 0 & \text{if } -C(\mathcal{G}_i, \mathcal{G}_j) + \frac{\alpha_i^k}{\mu_k} \geq 0 \\ \alpha_i^k - \mu_k C(\mathcal{G}_i, \mathcal{G}_j) & \text{otherwise.} \end{cases} \tag{5.7}$$

The update of the Lagrange multipliers for the user constraints follows the same function. By replacing in Equation 5.6, the energy equation becomes:

$$\hat{E}_k[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] + \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} h(C(\mathcal{G}_i, \mathcal{G}_j), \alpha_{i,j}^k, \mu_k)$$
$$+ \sum_{\substack{1 \leq n \leq N \\ 1 \leq q_n \leq K_n}} h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k) \tag{5.8}$$

where the function h of three scalar arguments is defined as follows:

$$h(c, \alpha; \mu) = \begin{cases} -\alpha c + \frac{\mu}{2} c^2 & \text{if } c - \frac{\alpha}{\mu} \geq 0, \\ -\frac{\alpha^2}{2\mu} & \text{otherwise.} \end{cases} \tag{5.9}$$

Hence, at each iteration k, $\hat{E}_k[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}]$ is minimized by gradient descent for fixed Lagrange multipliers and then Equation 5.7 is used to obtain the Lagrange multipliers estimates for optimization $k + 1$. For each optimization k, the update of $p_{n,l}$ and $\mathbf{u}$ is done as in Section 4.3.3.3 following Equation 4.15.

The gradient of $\hat{E}_k$ is easily computed by using the gradient equations given in Equation 4.22 and by following the gradient of h with regards to a variable x:

$$\frac{\partial h(c(x), \alpha; \mu)}{\partial x} = \begin{cases} (\mu c(x) - \alpha) \frac{\partial c(x)}{\partial x} & \text{if } c(x) - \frac{\alpha}{\mu} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$

The gradient equations are defined as follows:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_\Omega \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x})\rangle.f_n(\mathbf{x})d\mathbf{x}$$

$$+ \sum_{\substack{1 \leq i \leq N \\ i \neq n}} \frac{\partial h(C(\mathcal{G}_i, \mathcal{G}_n), \alpha_{i,n}^k; \mu_k)}{\partial p_{n,l}} + \sum_{1 \leq q_n \leq K_n} \frac{\partial h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k)}{\partial p_{n,l}} \quad (5.11)$$

$$\nabla_u \hat{E}_k(\mathbf{x}) = K_\sigma * \left\{ \sum_{n=1}^N \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}) \right\} +$$

$$+ \lambda \mathbf{u}(\mathbf{x}) + \sum_{1 \leq q_n \leq K_n} \nabla_u h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k) \quad (5.12)$$

Details of the gradient computations are given in Appendix D.2.

5.2.2 *Implementation*

The same kind of implementation optimizations used in Section 4.3.4 can be used again (integral over the surface instead of integral over the volume, computations near the zero level, collision detection, work at lower resolutions).

SIMPLE INTERACTIONS   In a practical context, the user first needs to select the organ on which he wants to apply the corrections before choosing the location of the correction. The type of correction (outside or inside point, $\gamma_{n,k} = -1$ or 1 respectively) does not need to be specified by the user. It can be easily guessed as we have access to the function $\phi_n \circ \mathcal{G}_n \circ \mathcal{L}$. For a point $\mathbf{x}$, the sign of $\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})$ directly indicates if the point is located inside or outside the contour of the object n. This also provides the distance to the contour, which may be an interesting information indicating how far from the contour the user asked for a correction.

5.2.3 *A new energy formulation for local constraints optimization*

When correcting the contours of pre-segmented objects, a user may expect the impact of his inputs to remain spatially local. A proper algorithm behavior would take into account the user inputs while relying on the image information in a region of interest around the user input location. In such a case we suppose that the objects are already correctly positioned in the image and that only local deformations occur, which means that the energy only depends on the deformation $\mathcal{L}$. Therefore we propose a new formulation of the energy E of Equation 5.5:
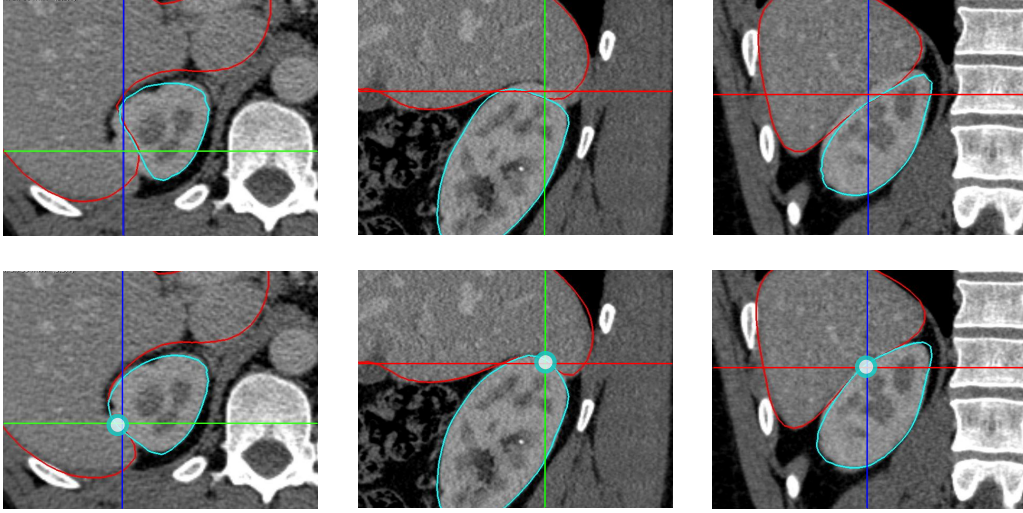
Figure 5.3: Example of 3D result after a user constraint in the left kidney (cyan dot). Top row: initial state of the contours. Bottom row: result after the introduction of the user constraint. We display the results in three orthogonal views (coronal, sagittal and axial).

$$E[\mathcal{L}] = \sum_{n=1}^{N} \left\{ \int_{\Omega} \left(K_s * \sum_{q_n=1}^{K_n} \delta_{q_n}(\mathbf{x}) \right) H(\phi_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right\} + \frac{\lambda}{2} \|\mathcal{L} - Id\|_U^2 \quad (5.13)$$

where K_s is a Gaussian kernel with fixed width s and $\delta_{q_n}(x) = \delta(x - x_{q_n})$ (δ is the Dirac distribution). With this new energy, the image-driven forces act only in the neighborhood of the user inputs. Note that this formulation guarantees shape-consistent solutions.

The numerical optimization with this new energy equation is similar to the one presented in Section 5.2.1, except that the pose transformations are not optimized and the non-overlapping constraint term is reduced to the empty set.

5.2.4 *Examples*

In this section we give some examples of typical results. In Figure 5.3 we show the result before and after the introduction of one user constraint in the kidney. The algorithm is performed in 3D. This example shows how the kidney contours move with the liver contours to take into account the user constraint.

In the next example, Figure 5.4, we show the interest of relying on image information when optimizing the user constraints. This experiment uses the local constraints formulation presented in Section 5.2.3. It is performed in 3D but we show only the coronal view. After adding two user constraints, the

(a) Initial state
(b) Result after local user constraints without image forces
(c) Result after local user constraints with image forces
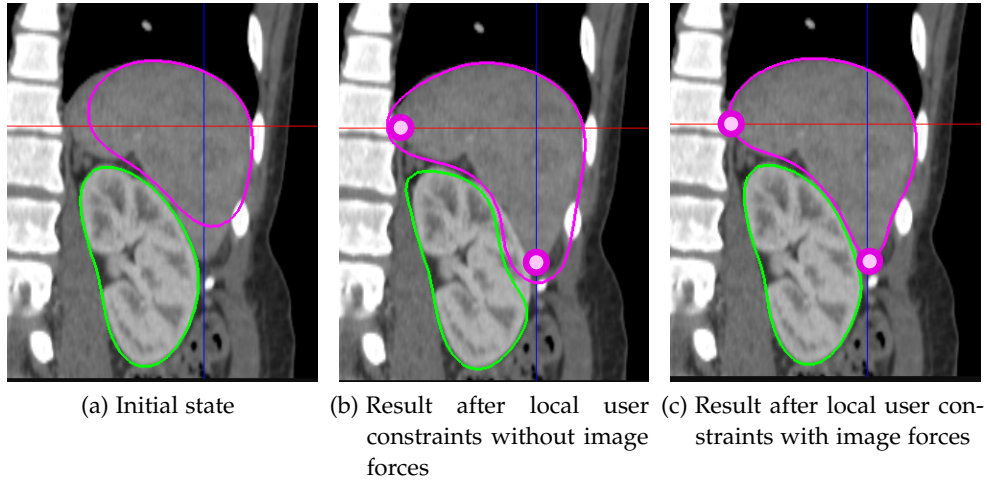
Figure 5.4: Results after user constraints (spleen and left kidney). Comparison between energy optimization with (c) and without (b) image-driven forces.
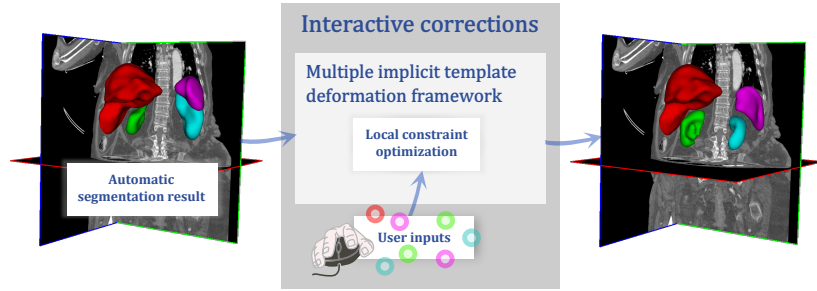
Figure 5.5: Interactive user correction workflow.

segmentation result with image-driven forces is more accurate than without forces, as expected. Indeed, if not taking into account the image information, the optimization will provide a local minimum that is not necessarily consistent with the organ boundaries.

5.3 APPLICATION TO THE CORRECTION OF ABDOMINAL ORGAN SEGMENTATIONS IN 3D CT

In this section, we evaluate our interactive correction method through experiments reproducing a clinical context. The idea behind our experiment is to reproduce a clinical context. The clinician could use automatic segmentation results (given from the automatic segmentation framework presented in Chapter 4, possibly run off-line) and correct them with our method with local corrections (presented in Section 5.2.3) if needed. The principle is illustrated in Figure 5.5. To evaluate fairly the user-friendliness and acceptance of our method,
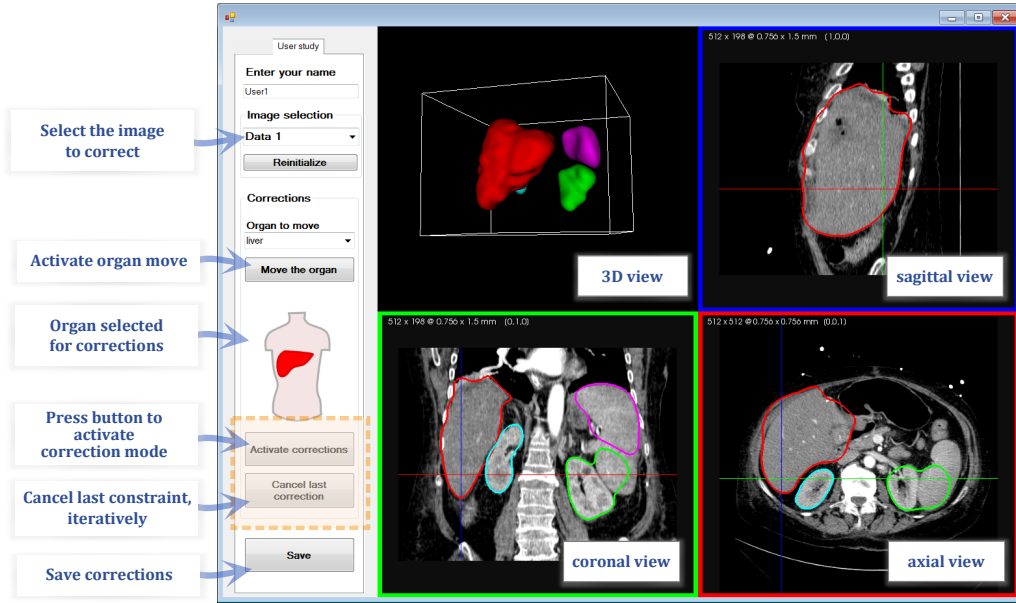
Figure 5.6: User interface for the user study on automatic segmentation corrections. In the orange box are highlighted the only buttons active during active corrections mode.

we organized a user study with a specific protocol. First, we present the interface that we designed for this study. Secondly, we detail the user study material and protocol and we present the results.

5.3.1   *User interface*

The interface used is presented in Figure 5.6. It is kept as simple as possible.

At the beginning of the study the user enters his name and then he corrects each data in the given order (*Image selection > Data 1*). For each image, the corrections can be of two types:

- global corrections: organ move, rotate and/or scale, in the 3D view

- local corrections: with clicks in any slice view.

We recommended the users to perform the global corrections (if needed) at the beginning of each image correction. In practice this tool was used in only one difficult case. To perform these global corrections the user proceeds as follows:

- he selects the organ he wants to modify,

- he presses the button *Move the organ*,

- he performs the transformations in the 3D view with mouse clicks (middle click allows for translation in the visualization plane, left click allows for rotations, right click allows for scaling),

- when he has finished, he presses again the button *Move the organ*

The local corrections are allowed after pressing the button *Activate corrections*. Once activated, the algorithm runs continuously until this button is pressed again. The user can add any correction at any time and they are taken into account by the algorithm *on-the-fly*. The interactions can be performed in any orthogonal view (coronal, sagittal or axial). The button clicks correspond to the following interactions:

- left click: select the organ to correct (the selected organ name appears near the mouse pointer and a corresponding icon appears in the left toolbar),

- right click: add a constraint point (if the click location is outside the selected organ then the constraint is attractive and if the click is inside then the constraint is repulsive),

- middle click down: navigate in the 2D views.

The user can cancel the last constraint he added by clicking on *Cancel last correction*. However, it does not put back the segmentations in the former state. This feature could be added relatively easily in a further version, with careful implementation and memory management.

5.3.2 *Material and implementation*

For the sake of continuity we use the same database as in Section 4.5. We recall that it is composed of 156 CT volumes from different 130 different patients with varied fields of view, body and organ shapes, resolution and use or not of contrast agent. The database has been split in 50 volumes for training and 106 for testing. We also recall that the method is implemented in C++ (for the algorithm and visualization) and C# (for the interface). In this experiment we use the automatic segmentation results obtained in Section 4.5.

For the correction steps, the image-driven forces presented in Section 4.4.2 are used with $\alpha = 0.5$ to rely both on image intensities and edge. The energy for local constraints (Section 5.2.3) is used with an isotropic Gaussian kernel K_s with width $s = 20$mm. These parameters have been set experimentally. They provided the expected behavior.

5.3.3 *Protocol*

We assume that the automatic segmentation method presented in Chapter 4 can be used off-line in clinical applications. Clinicians can then choose to correct the

segmentation results if they are not satisfactory. Starting from the automatic segmentation results of the evaluation performed in Section 4.5 on 106 images, we propose to constitute a small database of 10 volumes. To reflect the practical variability of automatic results, we sort the volumes with regards to their Dice coefficient to the ground truth and we select 10 volumes uniformly spread on this basis. These datasets are sorted in an arbitrary order for the user study to prevent any bias related to the difficulty of the corrections. The study is limited to the correction of four organs: the liver, the two kidneys and the spleen.

We asked 10 experts of the medical imaging domain (among them two radiologists) to correct the automatic segmentation results in these 10 images. They only had few minutes to familiarize themselves with the tool before starting the experiment. Users were asked to perform corrections until they reached what they considered satisfactory, clinically usable results. During the experiments, each click was recorded and intermediate segmentation results were saved (after each 100 iterations). At the end, each user was asked to fill a form with different questions about the usability of the interface.

5.3.4 *Results*

ACCURACY   On average the users spent 345 seconds per volume (median: 228s). Figure 5.7 gives, for each organ, the average mean distance and Dice coefficient per image according to the ground truth, in function of time. The accuracy measures converge rather rapidly. After about 300 seconds, an average distance of 1.5 mm is reached. Considering that there are four organs to look at and correct in 3D, this remains reasonable in terms of time. Note that the liver reaches a mean distance of 2 mm, higher than for other organs. We explain this mainly by the user variability and tolerance (e. g. sometimes the user included or excluded part of the aorta and the inferior vena cava).

Figure 5.8 presents the mean distance accuracy per user for each image of the database in function of time. Note that sometimes the user decided that the automatic segmentations were good enough and did not require corrections. Each color corresponds to one user. The graphs are displayed in ascending order starting from the best mean distance accuracy to the worst. The results of the radiologists correspond to the dashed plots. We notice some outliers as in Figure 5.8c in cyan. They are due to wrong clicks (the user forgets to select the organ he wants to correct) and the impossibility to undo a modification. The time spent on the corrections seems correlated with the automatic segmentation accuracy. In general, wrong segmentations are corrected, with more or less accuracy depending on the user. We also observe that in some cases (e. g. Figure 5.8d), good segmentation results may be difficult to improve.

Figure 5.9 presents some examples in two different images and for each orthogonal view. In the first example (two first rows) the user added 7 constraint
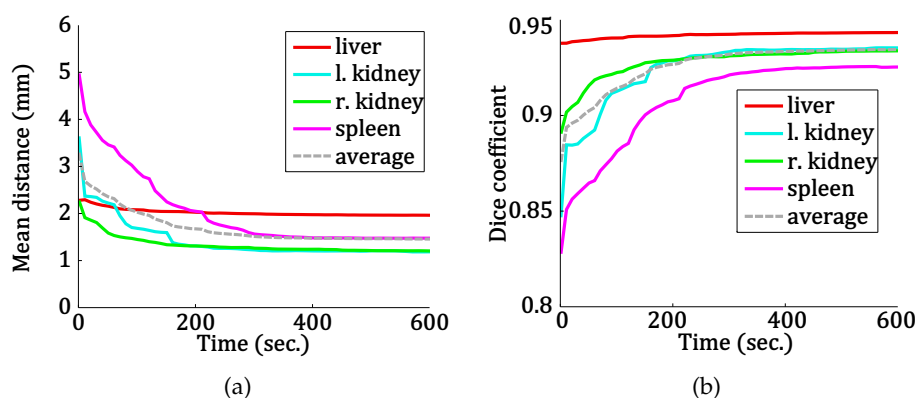
Figure 5.7: Average mean distance and Dice coefficient per image in function of the correction time.

points (2 on the right kidney and 5 on the spleen). The second column shows these points and the result after convergence. We see in the coronal view (last column), that at first the spleen and the kidney contours were adjacent at a location they should not. The corrections introduced allowed correcting this problem. In the second example (two last rows), the user added 16 constraint points (14 on the spleen and 2 on the liver). At first the spleen was wrongly segmented, leaking into the liver. The corrections enabled to get correct segmentations while being consistent with the interface between the tip of the liver and the top of the spleen.

USER FEEDBACK   In Figure 5.10 we give the results of the user form feedback. Considering the small number of participants we chose to ask for grades between 1 and 5. The grades of the segmentation quality before and after corrections seems to show that the users are rather satisfied with the segmentation results (note that there could be a bias as the users graded their own results, however we found they were rather critical in their judgments and comments). Moreover it also highlights that the automatic segmentation quality appears to be good (the two radiologists said they were rather impressed by these automatic results). The user-friendliness of the interface seems also rather good, between easy and very easy except for one user Moreover, we observed in practice that the global corrections were difficult to use. They could be easily replaced by a simpler operation (e. g. one click in the center of the target organ). Most users found the interface simple and easy to learn, requiring only minimal training (a few minutes). One negative point is the duration of the corrections. Although the user can visualize the optimization of its constraints in real-time, reaching convergence can take up to a few seconds. In practice, we observed that many users were watching the optimization until convergence instead of di-

(a)

(b) No corrections for 2 users

(c) No corrections for 2 users

(d)

(e) No corrections for 1 user

(f)

(g)

(h)

(i)

(j)
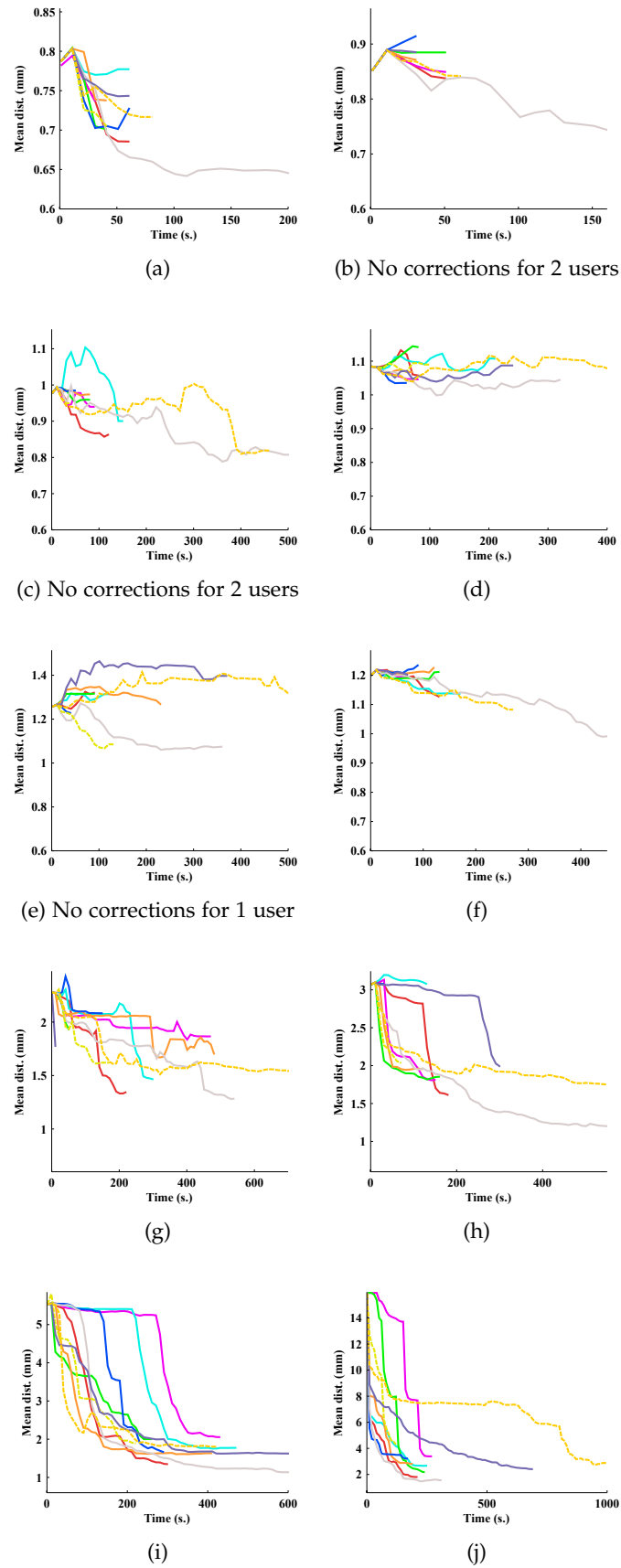
Figure 5.8: Mean distance accuracy (mm) for each test image of the user study and per user in function of time (s). One color corresponds to one user.

Figure 5.9: Examples of results in two different volumes. For each case, the first row corresponds to initial automatic segmentation results and the second row corresponds to results after correction.
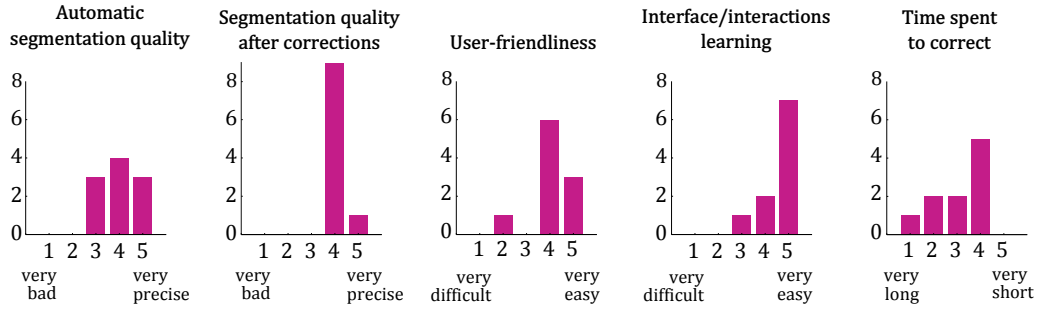
Figure 5.10: Feedback form results from the 10 experts.

rectly focusing on other corrections. This can be a bit frustrating for some users. Further optimization is needed to improve the responsiveness of the interface. In fact, since the user study, we significantly improved the responsiveness of the algorithm. In particular, we accelerated the convergence of the constraint optimization, leading to a much improved user experience. In addition, the algorithm is not adapted to large and irregular deformations. For instance if the liver contour leaked into the aorta it may be difficult to rectify the segmentation as the deformation field $\mathcal{L}$ can hardly remain a diffeomorphism in this case. This behavior is inherent to our formulation.

5.4 CONCLUSION

5.4.1 *Contributions*

In this chapter we presented how to extend the original multiple template deformation to take into account user interactions. The constraints are represented as control points that should lie or not inside a specified organ. They are formulated as additional constraints plugged into the template deformation framework. An augmented Lagrangian scheme allows turning the problem into an unconstrained one and to optimize it by gradient descent. This hard constraint formulation ensures the stability of the optimization process, contrary to most state-of-the-art approaches where the user constraints are formulated as soft constraints. In order to solve the problem of global optimization for corrections, we proposed a new energy formulation adapted to local corrections. This avoids unexpected deformations (for instance a deformation far from the user interaction). A user study with 10 medical imaging experts highlighted the usability of this framework for automatic segmentation results corrections in a practical, clinical-like setup. By keeping the interactions as simple as possible we showed that our interface can be easily used without advanced training. The correction time remains reasonable, depending on the accuracy required. Target accuracy

depends on the clinical application, though in a large number of cases, the automatic segmentation already gave satisfactory results.

5.4.2 *Limits and perspectives*

First, we observed in practice that the intrinsic diffeomorphic formulation may not be adapted to all types of correction. We plan on incorporating complementary tools (local smoothing, cropping or brush tools) to solve such issues. The addition of tools should however thought of carefully. Indeed, more complexity leads to more training time. Such a tool should then be easy to use.

Secondly, we think that our user interface would highly benefit from an acceleration of the algorithm. Indeed, instantaneous feedback should facilitate the usability of the tool. We think there is still room for implementation improvements.

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

We presented a complete and coherent pipeline for the localization, segmentation and user-guided correction of multiple organs in medical images. By performing extensive evaluations on large databases, we showed the robustness, accuracy and speed of each component of our pipeline, contributing to a globally reliable and efficient framework. Radiologists and medical imaging experts gave enthusiastic feedback on the algorithm results, thus proving its clinical potential.

From a methodological point of view, our work brings contributions in both localization and segmentation fields.

We proposed a new machine-learning-based localization approach for multiple anatomical structures, giving organ shape-dedicated results. This method extends the work of Criminisi et al. (2013) on anatomical localization and the work of Cuingnet et al. (2012) on cascades of RF regressions. It exploits shape prior knowledge to provide localization results in the form of organ-specific probability maps instead of organ bounding boxes. These confidence maps can have different purposes. In this work they are used to initialize and guide subsequent segmentation steps.

We also proposed a novel segmentation approach dedicated to multiple structures. This method extends the work of Mory et al. (2012) on implicit template deformation to multiple templates. The originality of our solution is to take advantage of the transformation decomposition into pose and local deformations. Each object has its own pose transformation to place it correctly in the image. Local deformations are handled by a non-rigid deformation field common to all objects. By introducing non-overlapping hard constraints, we ensure consistent segmentation results between the different target structures. We detail specific optimizations leading to a fast implementation. We also defined robust generic forces based both on region intensities and image gradients. The forces can be easily enhanced by introducing the information contained in the confidence maps. By using the same set of parameters for every organ, we get robust and accurate results in diverse conditions. The definition of application-tailored forces can also be considered and easily integrated in the framework.

We showed that this segmentation framework can be extended to integrate user-driven corrections. We took special care in the energetic formulation to ensure an intuitive behavior (especially, the locality of the corrections). User-

provided constraints are incorporated in a natural and consistent fashion into our framework. In particular, the formulation allows for coherent deformations between a corrected organ and its neighbors: organs move together according to the different constraints.

In conclusion, we developed a powerful and generic tool answering to the main clinical needs: robustness, accuracy, speed and ease-of-use. The clinical applications presented in this work demonstrate the potential of our approach:

- application to different modalities;

- usable from one to multiple structures;

- usable in automatic or interactive modes.

Lastly, we showed the soundness of injecting prior knowledge, whether through machine learning or shape priors, for better robustness and accuracy.

QUICK SUMMARY OF THE CONTRIBUTIONS

A complete pipeline for the automatic segmentation and correction of multiple organs:

- both generic and customizable;

- robust, accurate and computationally efficient;

- user-friendly.

A new multi-organ localization approach:

- robust thanks to the use of spatial relationships, contextual information and shape prior;

- giving organ-dedicated results thanks to confidence maps;

- providing probabilistic information that can have multiple applications.

A new multi-organ segmentation approach:

- robust thanks to the template deformation formulation, the shape prior and contrast-invariant forces;

- coherent results ensured by non-overlapping constraints;

- intuitive local user corrections.

6.2 PERSPECTIVES

At the end of each chapter we already gave possible directions for the improvement of each different component. We propose here to give more general comments. Given the versatility of this work, the perspectives for both improvements and applications are numerous.

IMPROVING THE LOCALIZATION OF SMALL STRUCTURES   We have seen that small structures with high shape variations (e. g. the gallbladder) are not localized very accurately with our approach. This seems highly related with the size of the structure, as we observed that the mean distance of the result to the ground truth is generally small. Starting from the region obtained by regression, fine-grained classification-methods could be integrated to refine the localization result. Indeed, local information would probably give valuable additional clues.

IMPROVING THE PATIENT-SPECIFIC SHAPE CONSISTENCY   In this work, the shape prior embedded in our models relies on mean organ shapes. Shape variations may also be exploited. The works of Zhou and Comaniciu (2007); Cootes et al. (2012); Chen and Zheng (2014) on direct shape regression may be explored to get confidence maps better highlighting the inter-patient variability. The integration of shape variations in the segmentation framework may also be explored as in the work of Prevost et al. (2013a). The construction of joint organs variations is a particularly interesting research lead (e. g. a large liver is likely to influence the position and size of the neighboring right kidney).

FURTHER EXPLOITING THE CONFIDENCE MAPS   Confidence maps hold valuable information that could be exploited further. The incertitude they contain may be used more deeply in our segmentation algorithm, for instance to balance the influence of the image forces.

In terms of applications, they could also be used to enhance the user experience: automatic organ focus, software customization (with adapted tools when the user moves the mouse in different parts of the image), improved 3D rendering etc.

FURTHER EXPLOITING THE IMAGE-DRIVEN FORCES   It would be interesting to investigate the use of the image-driven forces towards measures of the incertitude on the final segmentation. Indeed, these maps embed valuable information about the organ appearance and their location relationships (thanks to the confidence maps). This information could be used for instance to indicate to the user the locations that are uncertain and likely to need corrections.

Prior knowledge on organ appearance could also be introduced in the image-driven forces, for instance using machine learning techniques. This would prob-

ably improve the robustness of the algorithm. It could also be a way to detect pathologies and abnormalities such as tumors.

FURTHER EXPLOITING THE TEMPLATE DEFORMATIONS   The transformations optimized in the template deformation framework encode how templates deform to fit the target object boundaries. This *mapping* between the image and an anatomical reference is, in itself, a valuable source of information. For instance, it could be used for educational purpose. A user could click in the image and see the corresponding location in the anatomical reference (that could be displayed in the form of an atlas). It could also be used to perform statistical shape analysis. Indeed, if one model is used to perform multiple segmentations, the segmentation results can be mapped back to the model. This means that we can project the different anatomical parts on one single model and relate these parts between different subjects. Such an information can be used to identify the most varying anatomical parts and to model shape variations. In the same way, the different user corrections can be mapped to the original model. This could be used to identify the parts of the organs that require the most of corrections in general. These areas could be indicated to the user as regions that would need to be checked as a priority.

TOWARDS MULTI-MODALITY   In a way similar to the work of Prevost et al. (2013b), our framework could be easily extended to multi-modal, multi-organ co-registration and co-segmentation. The idea is to segment simultaneously the same organs in two (or more) images of different modalities. Each organ would be assigned different poses according to the images and they would deform jointly (one single deformation field) according to different forces computed from each image (adapted to the different modalities). For instance, in the case of liver MRI-US registration and segmentation, such a framework would probably benefit from adding the right kidney in the model (when it is visible).

OTHER APPLICATIONS   While the work on abdominal aortic aneurysm (AAA) (see Appendix C) should be investigated more in depth, we believe that our methods have a great potential for other applications. For instance, our work could be applied to the automatic segmentation of the prostate and the bladder. Our method could also be turned into a powerful tool to get a complete virtual patient 3D model (e. g. useful for teaching anatomy or surgery).

A

AUTOMATIC SEGMENTATION OF THE LIVER WITH TEMPLATE DEFORMATION IN CT IMAGES

In this part we present our preliminary work on fully automatic organ segmentation in CT images. It exploits the two main components on which we built our work in this thesis: (i) anatomy localization with RF regression Criminisi et al. (2013), (ii) segmentation with implicit template deformation (Mory et al., 2012). The method is dedicated to liver, although it could be adapted to other organs. The contributions of this work is twofold:

- a complete automatic segmentation pipeline composed of fast and robust methods ;

- a global-to-local segmentation strategy with the implicit template deformation framework.

While providing a robust and fast method for liver segmentation, this also paved the way to our work on multi-organ approaches.

We give an overview of the approach and the results hereafter (Section A.1 and Section A.2). We refer the reader to the publication (Gauriau et al., 2013) for more details.

A.1 THE METHOD

The liver is a vital organ of the human body with about 500 functions (Zakim and Boyer, 2003). It is the only organ capable of natural regeneration of lost tissues, making possible the resection of parts of the liver. The liver is also the second most transplanted organ (about 21000 liver transplants each year according to the World Health Organization (2007)). Liver segmentation from medical images is then required in many clinical contexts, providing anatomical measures that are crucial for clinical diagnosis, surgery planning (Hermoye et al., 2005) and radiation dose calculation (Murthy et al., 2005) for instance. The use of CT acquisitions is rather widespread in most of these cases as they provide 3D images with good quality (MRI is also alternatively used in certain applications). The automation of the liver segmentation is however challenging, due to its high variability of shape, pathologies and contrasts (different CT phases).

The literature on liver segmentation includes a vast variety of interactive, semi-automatic and automatic methods. We refer the reader to the recent and extensive reviews of Campadelli et al. (2009b) and Mharib et al. (2011) and to the *SLIVER*07 segmentation challenge (Heimann et al., 2007; Heimann and et
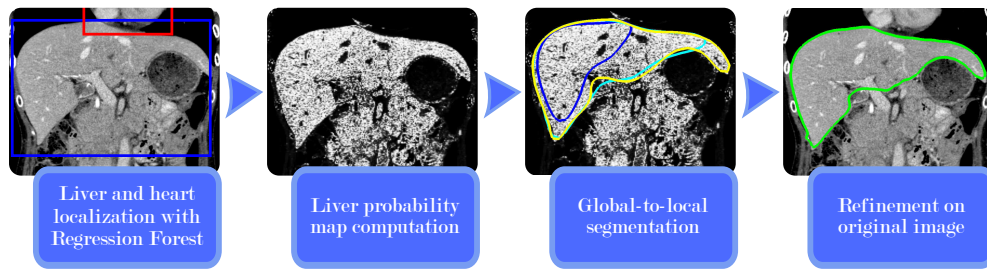
Figure A.1: Workflow of our fully-automatic liver segmentation method.

al., 2009) for more detailed bibliographic overviews. Those reviews highlight different groups of methods such as intensity-based, active contours, statistical shape models, graph-cuts and atlas-based registration. They also show that most segmentation methods focus on CT images with specific fields of view and particular CT phases (often contrast-enhanced) which may restrict their clinical use. Quantitative evaluation is another key point when comparing different methods. The challenge SLIVER07 (Heimann et al., 2007) has become a reference for liver segmentation, enabling fast and easy comparisons (Heimann and et al., 2009). Unfortunately this database is limited to 20 training and 10 testing datasets. In the literature, only few methods are validated on extensive databases. One can cite the works of (Ling et al., 2008; Zhang et al., 2010b; Linguraru et al., 2010), where the authors used proprietary databases composed of 277, 75 and 48 images respectively. However the differences in evaluation criteria and database composition make comparisons of these methods rather difficult.

A.1.1 *The pipeline*

We propose a fully-automatic and robust method for the segmentation of the liver in CT images. The workflow is inspired by the work of Cuingnet et al. (2012), which proved its efficiency for CT kidney segmentation. Our approach consists of four steps built on state-of-the-art algorithmic components (see Figure A.1).

The liver and the heart are first localized using RF regression (Criminisi et al., 2013). A liver probability map (or image-driven forces) is computed based on the localization results. Then the template deformation framework is used to perform the liver segmentation using a global-to-local strategy. A final refinement step is then applied using the original image.
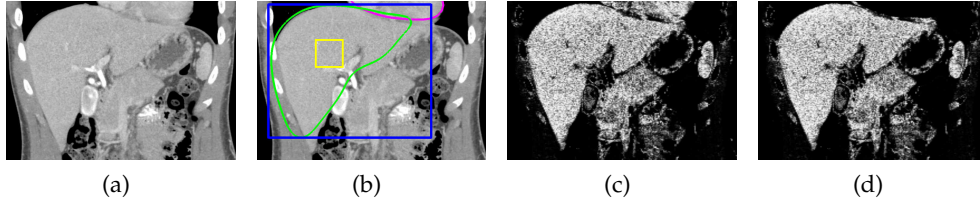
Figure A.2: Probability map computation steps: (a) original image, (b) liver predicted box (blue), fitted mean liver (green), heart segmentation (pink), voxels patch (yellow), probability map (c) before and (d) after heart masking.

A.1.2 *Liver and heart localization with RF regression*

Following the work of Criminisi et al. (2013) on anatomy localization (see Section 3.1.3), we use the same method to localize the liver and the heart. It provides bounding boxes for both organs (see Figure A.1), each of them parameterized with a vector of $\mathbb{R}^6$ composed of the coordinates of the two extremal vertices.

The training phase is performed using random subsets of voxels of the training images. The features used are the same as in (Criminisi et al., 2013): for each voxel we compute the mean intensities in two randomly displaced boxes. This exploits the fact that the intensities in CT images have a real physical meaning. In the testing phase a random selection of voxels votes for the predicted labels. We refer to Section 3.1.3 for more details on the method.

This approach provides robust estimates of the positions and sizes of the liver and heart which are used to derive a liver probability map described hereafter.

A.1.3 *Liver probability map computation*

Segmenting the liver directly in the image may provide insufficient results, in particular in images with poor contrast and fuzzy liver contours. Consequently, we propose to also take advantage of intensity distribution to pre-process the image and enhance liver voxels as shown in Figure A.2.

A.1.3.1 *Fitting a mean liver model in the predicted bounding box*

MEAN LIVER COMPUTATION   A mean liver model is built using a set of manually segmented liver shapes, represented as meshes. We first register the shape meshes using the fast and robust registration method of Fitzgibbon (2003). The main interest of this method over the classical Iterative Closest Point approach (Besl and McKay, 1992) is to overcome the problem of point correspondence and to use a robust norm to obtain a consistent registration of the shapes. The

mean liver model is obtained by averaging the implicit functions of the registered shapes.

MEAN LIVER FITTING   The mean liver shape is scaled anisotropically so as to best fit the predicted bounding box of the liver.

A.1.3.2   *Estimation of the intensity histogram*

From the previously fitted mean liver model barycenter we select a cuboidal patch ($\sim 60^3$ mm^3) of voxels, as illustrated in Figure A.2b. The intensity histogram is computed in this patch. Its normalization provides a function $h : \mathbb{R} \rightarrow [0,1]$ which gives an estimation of the probability of each intensity value to belong to the liver.

A.1.3.3   *Coarse segmentation of the heart*

In addition to the estimation of liver intensities, we roughly segment the heart in the image. This segmentation is performed on the original image with the template deformation framework described in Section 4.2. We initialize the algorithm with a mean heart model, built similarly to the mean liver shape and fitted in the predicted heart bounding box. Deformation parameters (see Table A.1) are set so as to prevent the heart contour from leaking in the liver. Conversely the rough binary mask M_h we obtain will prevent the liver contour from leaking in the heart in the subsequent steps.

A.1.3.4   *Probability map computation*

The liver probability map $M_l : \Omega \rightarrow [0,1]$ is defined as:

$$\forall \mathbf{x} \in \Omega, M_l(\mathbf{x}) = (1 - M_h(\mathbf{x}))\ h(I(\mathbf{x}))$$

where Ω is the image domain and I is the image. A subsampling of the image can be previously applied to increase the computational efficiency.

This probability map (see Figure A.2d is used in subsequent segmentation steps.

A.1.4   *Global-to-Local implicit template deformation*

We use the template deformation framework introduced in Mory et al. (2012) to extract the liver contours from the probability map and the original image. The framework is described in Section 4.2.

A.1.4.1 *Template deformation framework*

The model-based approach of Mory et al. (2012) is specially suited when target objects have partially unclear edges. Thanks to the intrinsic shape prior and regularization term, the algorithm fairly extrapolates the contours in difficult areas. Key advantages of the method are its speed, robustness and its ability to use both contour and region information. We refer the reader to Section 4.2 for more details on the variational formulation and optimization of the method. The approach is used with a region term relying on flux maximization and non-parametric region statistics. For easier understanding we recall the equation of the approach:

$$
\min_{\mathbf{p},\mathbf{u}} \left\{ E[\mathbf{p},\mathbf{u}] = \int_{\Omega} H(\phi_0 \circ \psi_{\mathbf{p},\mathbf{u}}) f(\mathbf{x}) d\mathbf{x} + \frac{\lambda}{2} \|\mathbf{u}\|_U^2 \right\}
$$

$$
\text{with } f(\mathbf{x}) = \alpha \Delta J(x) + (1-\alpha) log\left(\frac{p_{int}(J(x))}{p_{ext}(J(x))}\right), \tag{A.1}
$$

$$
\psi_{\mathbf{p},\mathbf{u}} = \mathcal{L} \circ \mathcal{G},
$$

$$
\mathcal{L} = Id + \mathbf{u}, \|\mathbf{u}\|_U^2 = \langle \mathbf{u},\mathbf{v} \rangle_{L^2(\Omega,\mathbb{R}^d)} = \langle \mathbf{u}, K_\sigma^{-1}\mathbf{u} \rangle_{L^2(\Omega,\mathbb{R}^d)}
$$

where $J : \Omega \rightarrow \mathbb{R}$ is the working image, $\mathbf{p}$ is the vector of parameters for the global transformation $\mathcal{G}$, $\mathbf{u}$ is a deformation field, ϕ_0 is the implicit shape template (for the liver in this case), $\lambda \in [0,1]$ is a constant parameter balancing the shape constraint, K_σ is a Gaussian kernel defining the RKHS U, p_{int} and p_{ext} are the intensities distributions inside and outside the deformed object regularly estimated on the working image J.

A.1.4.2 *Global-to-Local Segmentation Workflow*

Liver segmentation is performed according to an original global-to-local strategy on the probability map using the above-mentioned algorithm. Four steps help refining the contour progressively: 3 steps are performed on the probability map M_l and the final step on the original image I as shown on Figure A.3. Parameters have been set experimentally and are kept identical in all processed examples (see Table A.1). We observed a relatively low sensitivity to parameter variations in practice.

Table A.1: Parameters used for global-to-local segmentation.

	J	ψ	α	λ
Heart	I	$\mathcal{L} \circ Id$	0.8	0.1
Liver				
step 1	M_l	$Id \circ \mathcal{G}_r$	0.4	–
step 2	M_l	$\mathcal{L} \circ \mathcal{G}_s$	0.8	0.03
step 3	M_l	$\mathcal{L} \circ Id$	1	0.01
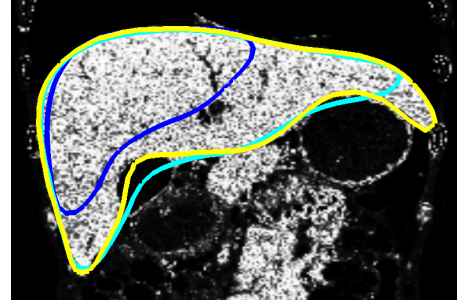
Figure A.3: The three segmentation steps on the probability map: step 1 (blue), step 2 (cyan), step 3 (yellow).

STEP 1: INITIALIZATION   The liver shape model is fitted to the predicted liver bounding box. As the template deformation method tends to favor expansion displacements, we scale down the model by a factor 0.7. Then a first step aims at globally registering the shape model without any local deformation. Equation A.1 is minimized using the parameters reported in Table A.1. The high weight on the region term constrains the shape inside the liver, thus facilitating expansion in subsequent deformation steps.

STEP 2: COARSE SEGMENTATION   We deform the previous result, still minimizing Equation A.1 but with different parameters (see Table A.1) allowing some local deformations. The flux term is now more important than the region term so that the model contours match liver edges more accurately.

STEP 3: LOCAL DEFORMATION   The third step helps refining the segmentation. Now we only optimize local deformations (see Table A.1). The flux term is used alone in order to reach the contours. Releasing the shape constraint finally helps reaching stretched parts such as liver tips.

STEP 4: REFINEMENT ON THE ORIGINAL IMAGE   To improve the accuracy of the final segmentation, we apply guided filtering (He et al., 2010) on the binary mask of the previous segmentation result, using the original image as guide. As detailed in the work of He et al. (2010), the guided filter acts as a fast, local matting/feathering refinement step enabling the final segmentation to better match the edges of the original image.

A.2 EXPERIMENTS AND RESULTS

In this part we present the experiments performed on two databases. The first database is SLIVER07 (Heimann et al., 2007), which has become a reference for liver segmentation evaluation. The second database is composed of 268 diverse CT volumes, further demonstrating the accuracy and robustness of the method. In those two experiments both the regression forest and the mean liver model are learned solely on the 20 training datasets of SLIVER07 (publicly available), for the sake of result reproducibility. Computational times are given for a C++ implementation on a machine with four 2.3GHz cores (Core i7-2820QM) and 8Gb RAM.

A.2.1 *Training part*

For mean liver model building and RF training, we use the 20 training images of SLIVER07. There is no pre-processing of the images before training. The forest of 7 trees and 12 decision levels is learned after randomly selecting a subset of 40.000 voxels per image. The minimum node size is 50 and the training computational time is about 10 minutes.

A.2.2 *Evaluation on the SLIVER07 benchmark*

We tested our method on the SLIVER07 challenge database (Heimann et al., 2007) which is composed of 20 training and 10 testing 3D CT volumes (average slice and inter-slice resolution are 0.7mm±0.1 and 1.5mm±0.9, respectively) rather focused on the liver. The RF and mean liver shape were trained on the 20 training samples while we tested the algorithm on the 10 testing volumes. We compare our results with the best reported 3D methods (Kainmüller et al., 2007; Wimmer et al., 2009; Linguraru et al., 2012a; Huang et al., 2012) of the challenge, pointing out that the first three did not obtained those results in the challenge conditions (training on the 20 samples). Among these automatic methods ours comes in fifth position. In Table A.2 we report the same validation measures (OVE: overlap error, VOD: volume difference, AVD: average distance, RMSD: root mean squared distance, MAXD: maximum distance) and inter-observer scores (see (Heimann and et al., 2009)) as used in the challenge. Those results have also been published online (Heimann et al., 2007).

On this challenge database we get comparable results to best scored methods (see last column in Table A.2). Our method also presents the advantage of being faster and requiring few samples for training. Moreover we thoroughly evaluate the robustness of liver localization in various conditions hereafter.

Table A.2: The five best automatic methods on SLIVER07 database. We report the computational time (per image), the number of training volumes and the SLIVER07 measures. (n/a: non available)

	Time	Training volumes	OVE [%]	OVE Score	VOD [%]	VOD Score	AVD [mm]	AVD Score	RMSD [mm]	RMSD Score	MAXD [mm]	MAXD Score	Total Score
Kainmüller et al. (2007)	15min	122	**6.1**	**76.2**	−2.9	84.7	**0.95**	**76.3**	**1.9**	**74.0**	18.7	75.4	**77.3 ± 9.4**
Wimmer et al. (2009)	3min	112	6.5	74.7	**1.0**	**86.4**	1.0	74.5	2.0	72.3	**18.3**	**75.9**	76.8 ± 3.8
Linguraru et al. (2012a)	n/a	n/a	6.4	75.1	2.3	85.0	1.0	74.9	1.9	73.4	20.8	72.7	76.2 ± 5.9
Huang et al. (2012)	n/a	-	7.6	70.4	−1.3	85.4	1.3	68.0	2.4	67.4	22.1	70.9	72.4 ± 8.6
Our method	**46s**	**20**	7.2	71.7	2.6	85.0	1.3	67.0	2.6	64.2	23.1	69.6	71.5 ± 10.0

Table A.3: Results and computational time after each step of the algorithm reported as Mean ± Standard-Deviation.

	Time [sec.]	OVE [%]	VOD [%]	AVD [mm]	RMSD [mm]	MAXD [mm]
RF localization	1.5	-	-	10.8 ± 6.8	-	-
Proba. map	6	-	-	-	-	-
Step 1	1	57.7 ± 10.4	57.3 ± 11.1	23.8 ± 9.9	31.6 ± 12.3	84.9 ± 27.4
Step 2	10	10.4 ± 6.7	-1.0 ± 6.0	2.3 ± 3.2	4.0 ± 4.6	23.3 ± 14.9
Step 3	23	9.6 ± 6.7	-2.4 ± 5.7	1.9 ± 2.9	3.3 ± 4.1	21.5 ± 13.7
Step 4	5	8.4 ± 6.7	-1.5 ± 5.6	1.7 ± 2.4	3.0 ± 3.6	21.7 ± 14.1

A.2.3 *Evaluation on a large and varied database*

A.2.3.1 *Database description*

The database we use in this experiment is composed of 268 3D CT images coming from 127 patients with diverse medical conditions (41% of the patients present a significant alteration of their liver shape and/or appearance). The database includes volumes with varied body shapes, fields of view (for 28% of the database the images include a large part of the trunk), resolution and use or no use of contrast agents (19% of delayed or non contrasted scans, 31% of hepatic arterial phases and 50% of portal venous phases) as shown in Figure A.5. Slices and inter-slices resolution ranges from 0.5 to 1 mm and from 0.5 to 3 mm, respectively. The 268 images have been segmented manually by an expert. The regression forest we use in this experiment is the same as the one used previously.

A.2.3.2 *Results*

In Table A.3 we report the results after each step thus showing the relevance of the global-to-local strategy. For the sake of consistency we use the same evaluation measures as in the first experiment. The localization with regression forest is fast (1.5s) and robust as the average distance (mean distance of box faces) is of 10.8mm for a maximum of 46.8mm. In comparison, Criminisi et al. (2013) obtain an average distance of 15.7mm for liver localization. We again emphasize that the regression forest was trained on only 20 datasets, which further highlights the robustness of the method.

After the last refinement we obtain mean and median distances of 1.7mm and 1.3mm, respectively. The median value highlights the presence of a limited number of outliers. Indeed for more than 90% of the database the overlap error is below 15.8% and the average distance below 3mm. Outliers can be principally
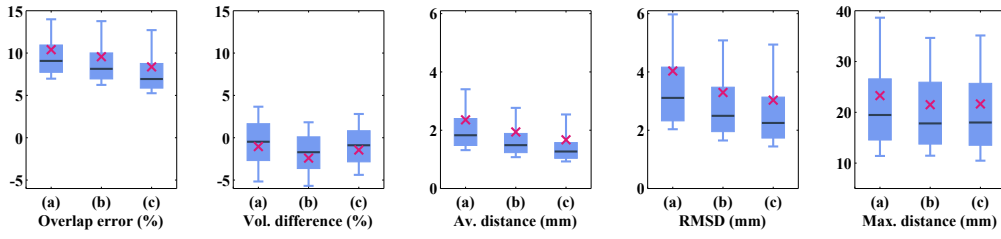
Figure A.4: Boxplots for (a) step 2 (b) step 3 and (c) step 4. They represent 1st and 9th decile, 1st and 3rd quartile, median (dark-blue dash) and mean (pink cross).
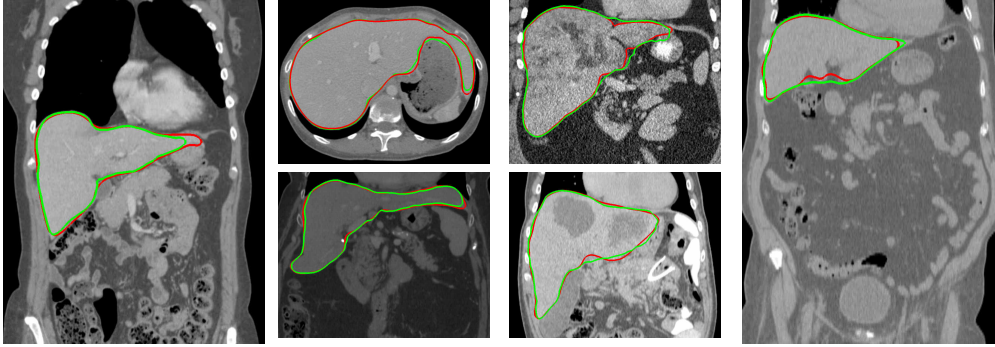
Figure A.5: Examples of segmentation results (red: ground truth, green: our result) in varied situations: different fields of view, liver contrast, shape and pathologies.

explained by a wrong initial position of the shape model (imprecise bounding box) and by diseases giving a very atypical appearance to the liver. In Figure A.4 we display boxplots of the different measures, showing their relatively compact dispersion. These results confirm those obtained on the SLIVER07 database and are good, despite the much larger variability of the database. Figure A.5 shows the segmentation accuracy in various situations.

A.3 CONCLUSION

In this appendix we presented a fully-automatic pipeline for CT liver segmentation. We showed that the approach is robust to a large variety of imaging conditions: different fields of view, different CT phases, healthy and diseased livers. The method relies on a localization step performed with RF regression and several template-based deformation steps sequentially done in a global-to-local fashion.

The results on the SLIVER07 similar to the best state-of-the-art performances, are confirmed by an additional extensive evaluation on a large and heterogeneous database (268 volumes). This validation demonstrates that our framework reaches an attractive balance between robustness, accuracy (mean distance to ground truth of 1.7mm) and speed (46s). We emphasize the genericity and relative simplicity of our framework, which required very limited liver-specific tuning. It is reproducible and could be improved in a number of ways. For instance, failed segmentations could be detected and easily corrected by the clinician as the template deformation framework we employ can handle user interactions (see Chapter 5).

REMARK   This preliminary work on automatic segmentation gave us a first practical insight of the potential of the random forest and template deformation approaches. Both methods are robust and fast which make them adapted to clinical applications. From these conclusions, the extension to multiple organ applications appeared natural.

AUTOMATIC LIVER SEGMENTATION IN MRI

In this appendix, we present a method to segment the liver in 3D in MRI. This pipeline, mainly developed by Roberto Ardon (Senior Scientist, Philips Research), works in a similar way as the method presented in Appendix A for the segmentation of the liver in CT. It is composed of two components:

- a localization step provided by the method presented in Chapter 3: the confidence maps of the liver and neighbor organs (the heart and the right kidney) are computed;

- a coarse-to-fine segmentation pipeline (similar to the one presented in Appendix A) based on the implicit deformation framework (Prevost et al., 2013a).

This appendix does not cover all the technical details of the method, but rather demonstrates:

- the generic aspect, with regards to modality, of our localization method with cascade of RF and shape prior;

- the interest of the confidence maps when used as attenuation maps to compute the image-driven forces (as presented in Section 4.5).

MATERIAL   The database used for the experiment is composed of 175 T1-weighted MRI volumes from different patients. The images come from the same clinical site. All the volumes were segmented manually by an expert. Slice and inter-slice resolutions range from 0.5 to 2 mm and from 2 to 10 mm, respectively.

EXPERIMENT AND RESULTS   To show the interest of using confidence maps in the image-driven forces used for segmentation, we propose to launch the automatic segmentation algorithm with and without the confidence maps. The quantitative results are given in Table B.1. They show a significant improvement when using the confidence maps. Mean and median Dice coefficients are improved by about 1%. The mean distance improves by about 10% (decreasing from 4.5mm to 4.0mm in average). The Hausdorf distance measure confirms this trend with a gain of almost 2mm.

In Figure B.1 we show qualitative examples of the improvement. We see the influence of the confidence maps, avoiding the segmentation to leak into the heart and into the right kidney.

Table B.1: Final results: Dice coefficient (%), mean distance (mm) and Hausdorff distance to ground truth. (average $\pm$ standard deviation (median))

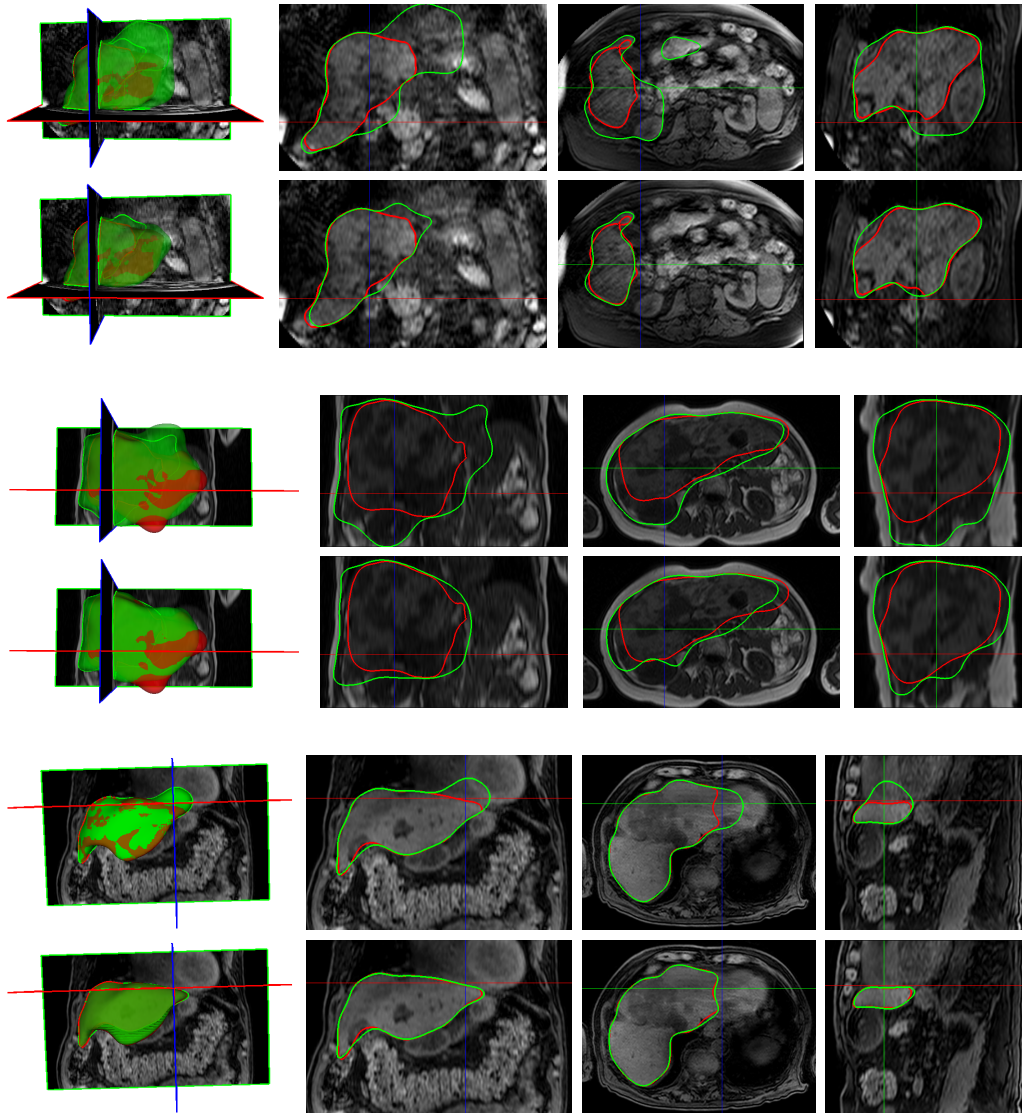|	Dice (%)	Mean dist. (mm)	Hausdorf dist. (mm)
Without confidence maps attenutation	$90.3 \pm 14.0(94.8)$	$4.5 \pm 9.2(2.0)$	$22.7 \pm 12.5(20.0)$
With confidence maps attenuation	$91.2 \pm 13.3(95.6)$	$4.0 \pm 8.5(1.7)$	$21.0 \pm 12.0(17.7)$

Figure B.1: Each couple of rows corresponds to one case where the first and second rows show the segmentation result without confidence maps and with confidence maps, respectively. In red: ground truth, in green: segmentation result.

C

ABDOMINAL AORTIC ANEURYSM (AAA) LOCALIZATION IN 3D ULTRASOUND

In this appendix, we present preliminary results on the application of our localization method (Chapter 3) for AAA in 3D US images. This experiment, performed in collaboration with Laurence Rouet (Senior Scientist, Philips Research), shows that our method can be used for different applications and imaging modalities. This work has been submitted to *IEEE International Ultrasonics Symposium, 2015*.

C.1 CONTEXT

The aorta is a major blood vessel of the human body. As shown in Figure C.1, it runs from the heart through the chest and to the abdomen where it divides at the aorto-illiac bifurcation.

AAA is a local blood-filled enlargement of the abdominal aorta such that its diameter is 50% larger compared to the normal aortic diameter (Solomon and Kent, 2014). Over time this bulge can become weak and the blood pressure may cause its rupture. In such a case, mortality rate is of about 85% (Solomon and Kent, 2014). It affects between 2 and 8% of the men over the age of 65 and four times less women. Causes are not exactly known. There are however well-defined risk factors such as smoking, hypertension or genetics (Wittels, 2011).

There are two main treatment options for AAA: open surgical repair (replacement of the diseased area with a graft) and endovascular repair (minimally invasive surgery with insertion of a stent graft in the aorta). US imaging is today the reference imaging technique for diagnosis and monitoring for surveillance of small AAA. Follow-up after endovascular surgery is mostly performed with CT. The repair decision is generally based on AAA diameter measurements from 2D US images (a diameter above 50mm is an indication for intervention). The reproducibility and reliability of the measurement is a clinical challenge. The use of volume measurement in 3D US has been recently explored by Rouet et al. (2010); Long et al. (2013); Bredhal et al. (2013). Although it is not routinely used in clinical practice, a follow-up with volume measurements has already been proven more predictive of the AAA evolution than a follow-up with diameter measures for the CT modality Prinssen et al. (2003).
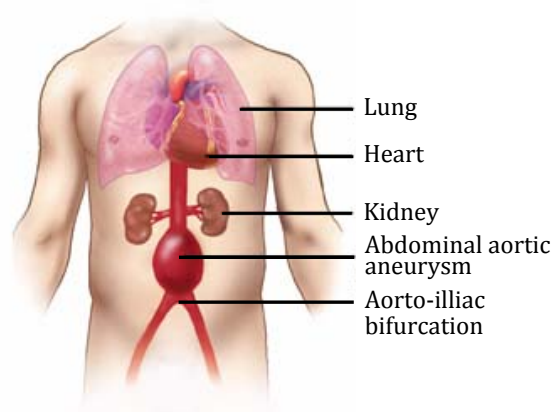
Figure C.1: AAA in the human body. (Source: Society of Interventional Radiology (SIR) (2015))

C.2   SEGMENTATION OF AAA IN 3D US

Long et al. (2013); Bredhal et al. (2013) propose a workflow to segment AAA in 3D and quantify different parameters (e. g. maximum diameter, area, volume). By comparing with 2D US measurements and CT quantification, they show that this approach is very promising.

One of the novelty of the approach is to use a semi-automatic segmentation tool allowing user corrections. Further improvements have also be done for these publications and automation of the tool has been implemented. This tool already gives good segmentation results. However, in an attempt to get more robustness and accuracy (especially in post-endovascular surgery), we started preliminary experiments to apply our localization method (Chapter 3) for post-endovascular AAA in 3D US.

C.3   AAA LOCALIZATION WITH CONFIDENCE MAPS

We use the same workflow as the one presented in Chapter 3: a cascade of two RF regressors and shape prior. The only difference is the type of features. Indeed, difference of intensities does not make sense in US images. We thus propose to use two types of features. The first feature type is binary and is based on the polarity of mean intensity differences (same approach as in Section 3.3.1). The second feature type is similar to the first one but using the Laplacian of the image instead of the image itself.

Similar parameters as in Section 3.5 are used. We trained two sequential RF on 30 post-endovascular 3D US and performed qualitative evaluation on other datasets.

(a)   (b)
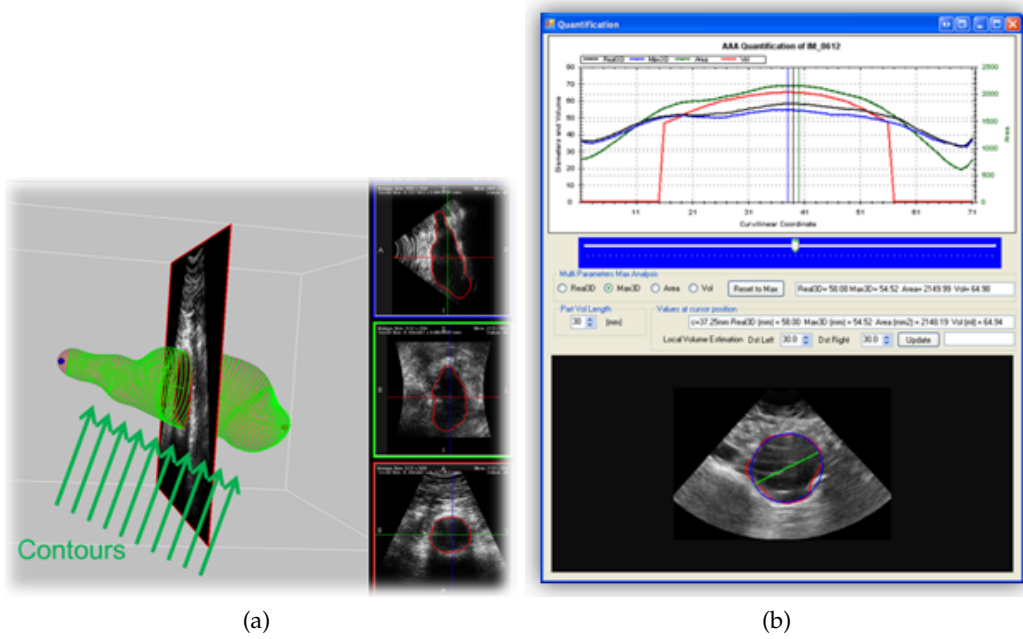
Figure C.2: Example of AAA 3D segmentation with visualization (a) and quantification tool (b). (Source: Laurence Rouet)

In Figures C.3 and C.4 we give some examples of localization of AAA. We observe on the first examples that the algorithm seems to be robust to various aneurysm shapes and sizes and to different fields of view. On the examples of Figure C.4 we note that the algorithm is less accurate in the case of aneurysms with uncommon shapes and size (very small or very big).

The results are promising and show the interest of our localization approach. We believe they may be improved by: (i) performing a dedicated parameters optimization as in Section 3.4, (ii) studying other features more adapted to US, (iii) increasing and varying the training database. Quantitative experiments and evaluation of the impact of this localization on the segmentation will be done in a near future.
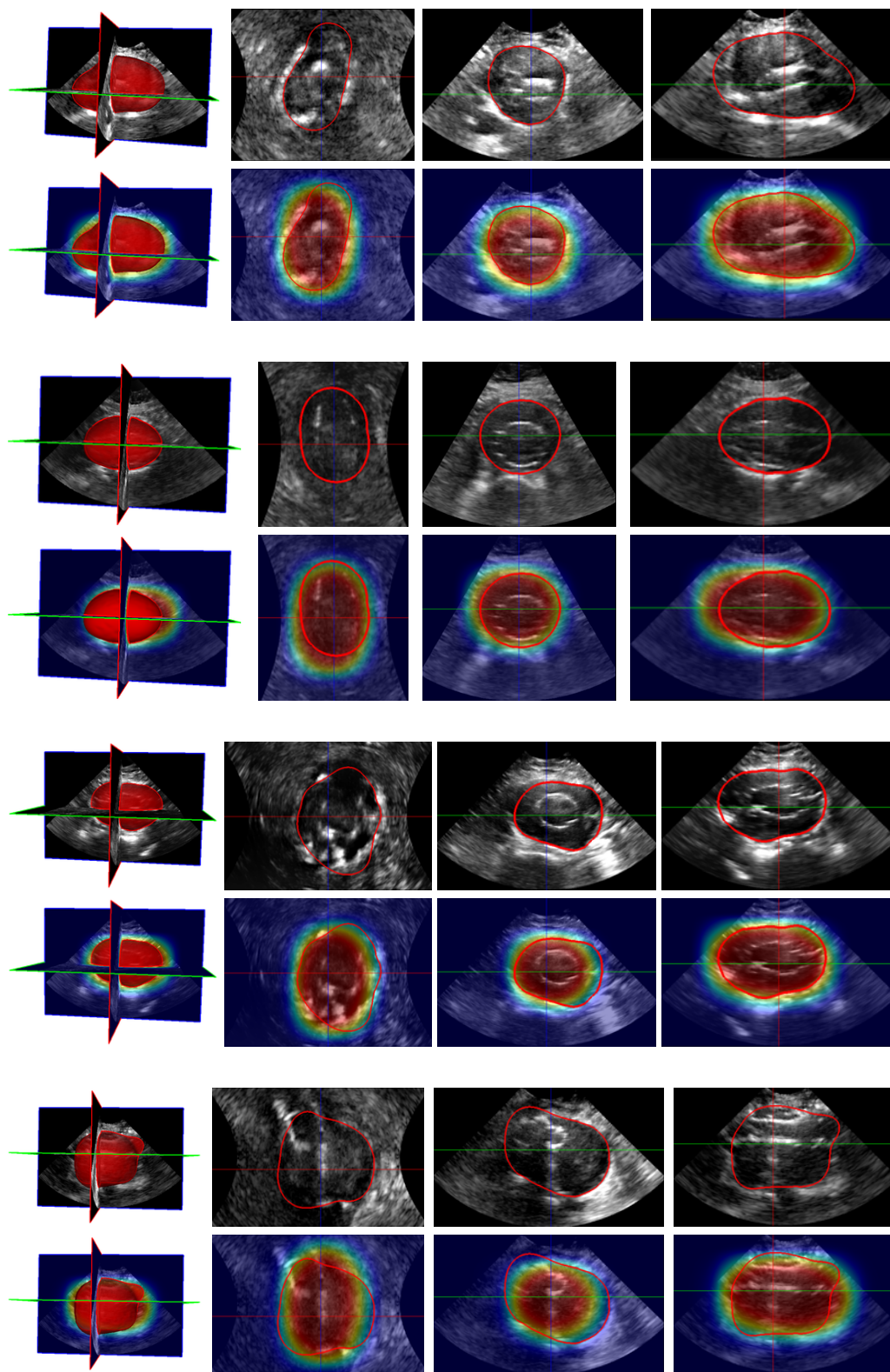
Figure C.3: Examples of accurate localization of AAA. Each pair of rows corresponds to a different patient. In red is the ground truth and the confidence map is set as an overlay on the image in the second row. First is given the 3D view and then three orthogonal views.
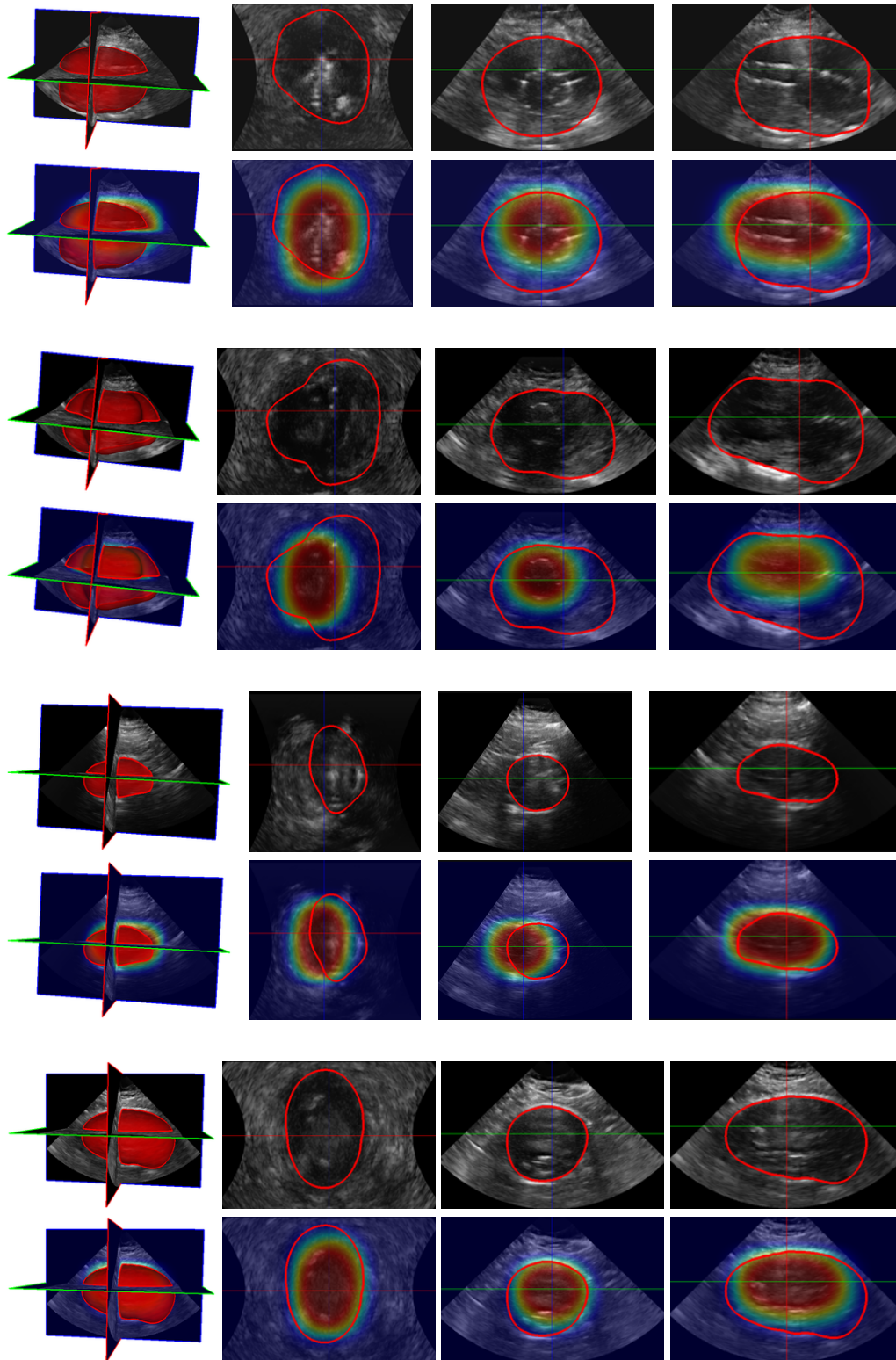
Figure C.4: Examples of less accurate localization of AAA. Each pair of rows corresponds to a different patient. In red is the ground truth and the confidence map is set as an overlay on the image in the second row. First is given the 3D view and then three orthogonal views.

GRADIENT EQUATIONS FOR THE MULTIPLE IMPLICIT TEMPLATE DEFORMATION FUNCTIONAL

D.1 DERIVATION DETAILS FOR THE FUNCTIONAL UNDER NON-OVERLAPPING CONSTRAINTS

In this section we detail the calculus of variation for the gradient equations given in Section 4.3.3 for the quadratic penalty method and the augmented Lagrangian method. We recall the energy equation to minimize given in Equation 4.12:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_{\Omega} H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(x)d\mathbf{x} \right) + \frac{\lambda}{2} \|\mathcal{L}(\cdot) - Id\|_U^2$$

$$\text{with } \mathcal{L} = \mathbf{Id} + \mathbf{u}, \|\mathbf{u}\|_U^2 = \langle \mathbf{u}, K_\sigma^{-1}\mathbf{u} \rangle_{L^2(\Omega, \mathbb{R}^d)}$$

$$\text{subject to } \forall (i, j) \in [\![1, N]\!]^2, i < j,$$

$$C(\mathcal{G}_i, \mathcal{G}_j) = \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x}))d\mathbf{x} = 0$$

D.1.1 *With the penalty method*

Following the penalty method the problem is written as (Equation 4.13):

$$\min_{\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}} \hat{E}_k[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}]$$

$$\text{with } \hat{E}_k[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] + \frac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i, \mathcal{G}_j)^2$$

We detail further the computation of the different terms:

- $\nabla_{p_{n,k}} E$,

- $\nabla_{p_{n,k}} V_Q$ with $V_Q = \frac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i, \mathcal{G}_j)^2$,

- $\nabla_u E_k$.

D.1.1.1 *Gradient with respect to the pose*

Here we give the gradient of the functional with respect to the pose. We split this computation in two terms: $\nabla_{p_{n,k}} E$ and $\nabla_{p_{n,k}} V$. We recall that the parametric transformations $\mathcal{G}_n$ are defined by N_g parameters $\mathbf{p_n} = \{p_{n,l}\}_{l=1...N_g}$ (e.g. $N_g = 6$ if $\mathcal{G}_n$ is rigid in $\mathbb{R}^3$).

$$
\begin{aligned}
\nabla_{p_{n,l}} E &= \sum_{i=1}^{N} \left(\int_{\Omega} \frac{\partial}{\partial p_{n,l}} (H(\phi_i \circ \mathcal{G}_i \circ \mathcal{L}(\mathbf{x}))).f_i(\mathbf{x}) d\mathbf{x} \right) \\
&= \int_{\Omega} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\frac{\partial}{\partial p_{n,l}} (\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x}) d\mathbf{x} \\
&= \int_{\Omega} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial}{\partial p_{n,l}} (\mathcal{G}_n \circ \mathcal{L}(\mathbf{x})) \rangle .f_n(\mathbf{x}) d\mathbf{x} \\
&= \int_{\Omega} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x}) \rangle .f_n(\mathbf{x}) d\mathbf{x}
\end{aligned}
\tag{D.1}
$$

Now we detail the computation of the gradient of the penalty term.

$$
\begin{aligned}
\nabla_{p_{n,l}} V_Q &= \frac{1}{2} \frac{\partial}{\partial p_{n,l}} \sum_{\substack{1 \le i \le N \\ i < j \le N}} C(\mathcal{G}_i, \mathcal{G}_j)^2 \\
&= \sum_{\substack{1 \le i \le N \\ i < j \le N}} C(\mathcal{G}_i, \mathcal{G}_j).\frac{\partial}{\partial p_{n,l}} \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_j \circ \mathcal{G}_j(\mathbf{x})) d\mathbf{x} \\
&= \sum_{\substack{1 \le i \le N \\ i \ne n}} C(\mathcal{G}_i, \mathcal{G}_n).\frac{\partial}{\partial p_{n,l}} \int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).H(\phi_n \circ \mathcal{G}_n(\mathbf{x})) d\mathbf{x} \\
&= \sum_{\substack{1 \le i \le N \\ i \ne n}} C(\mathcal{G}_i, \mathcal{G}_n).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\frac{\partial}{\partial p_{n,l}} H(\phi_n \circ \mathcal{G}_n(\mathbf{x})) d\mathbf{x} \\
&= \sum_{\substack{1 \le i \le N \\ i \ne n}} C(\mathcal{G}_i, \mathcal{G}_n).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\frac{\partial}{\partial p_{n,l}} (\phi_n \circ \mathcal{G}_n(\mathbf{x})) d\mathbf{x} \\
&= \sum_{\substack{1 \le i \le N \\ i \ne n}} C(\mathcal{G}_i, \mathcal{G}_n).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x}
\end{aligned}
\tag{D.2}
$$

D.1.1.2 *Gradient with respect to the deformation field*

The deformation $\mathcal{L} : \Omega \to \Omega$ is parameterized by the displacement field $\mathbf{u} \in L^2(\Omega, \mathbb{R}^d)$ such that $\mathcal{L} = Id + \mathbf{u}$. If we denote $\mathbf{u_t} = \mathbf{u} + t\eta$ a perturbation of $\mathbf{u}$ in the direction η and $\mathcal{L}_t = Id + \mathbf{u_t}$, then

$$
\frac{\partial \mathcal{L}}{\partial t} = \lim_{t \to 0} \frac{\mathcal{L}_t - \mathcal{L}}{t} = \eta
$$

Following the same kind of approach given in Prevost (2013), we compute $D_\eta E$ the Gâteaux derivative of E with respect to $\mathbf{u}$ in the direction η. This helps expressing E as a scalar product to exhibit the gradient we are looking for.

We recall the expression of E:

$$E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}] = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(x)).f_n(x)dx \right) + \frac{\lambda}{2}\|\mathcal{L}(\cdot) - Id\|_U^2$$

For simplicity of notation, we write E instead of $E[\mathcal{G}_1, ..., \mathcal{G}_N, \mathcal{L}]$ in the following computations:

$$D_\eta E = D_\eta \left(\sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).f_n(\mathbf{x})d\mathbf{x} \right) + \frac{\lambda}{2}\|\mathcal{L}_t(\cdot) - Id\|_U^2 \right)$$

$$= \int_\Omega \sum_{n=1}^{N} \left(\frac{\partial}{\partial t}(H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).f_n(\mathbf{x}) \right) d\mathbf{x} + \frac{\lambda}{2}\frac{\partial}{\partial t}\|\mathcal{L}_t(\cdot) - Id\|_U^2$$

$$= \int_\Omega \sum_{n=1}^{N} \left(\delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).\frac{\partial}{\partial t}(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).f_n(\mathbf{x}) \right)$$
$$+ \frac{\lambda}{2}\frac{\partial}{\partial t}\langle \mathcal{L}_t(\mathbf{x}) - \mathbf{x}, \mathcal{L}_t(\mathbf{x}) - \mathbf{x}\rangle_U \, d\mathbf{x}$$

$$= \int_\Omega \sum_{n=1}^{N} \left(\delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).\langle\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x}), \frac{\partial}{\partial t}(\mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x}))\rangle_{L^2(\Omega,\mathbb{R}^d)}.f_n(\mathbf{x}) \right)$$
$$+ \lambda\langle \mathcal{L}_t(\mathbf{x}) - \mathbf{x}, \eta(\mathbf{x})\rangle_U d\mathbf{x}$$

$$= \int_\Omega \sum_{n=1}^{N} \left(\delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).\langle\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x}), d\mathcal{G}_{n_{(\mathcal{L}_t(\mathbf{x}))}}.\eta(\mathbf{x})\rangle_{L^2(\Omega,\mathbb{R}^d)}.f_n(\mathbf{x}) \right)$$
$$+ \lambda\langle \mathcal{L}_t(\mathbf{x}) - \mathbf{x}, \eta(\mathbf{x})\rangle_U d\mathbf{x}$$

$$= \int_\Omega \langle \sum_{n=1}^{N} \left(\delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t.\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x}) \right) + K_\sigma^{-1}*(\lambda(\mathcal{L}_t(\mathbf{x}) - \mathbf{x})), \eta(\mathbf{x})\rangle_{\mathcal{L}^2(\Omega,\mathbb{R}^d)}d\mathbf{x}$$

$$\tag{D.3}$$

As it is preferable to convolve than to deconvolve we prefer to write:

$$D_\eta E = \int_\Omega \langle K_\sigma*\left(\sum_{n=1}^{N} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t.\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}_t(\mathbf{x}) \right) + \lambda\mathbf{u}(\mathbf{x}), \eta(\mathbf{x})\rangle_{\mathcal{L}^2(\Omega,\mathbb{R}^d)}d\mathbf{x}$$

$$\tag{D.4}$$

Equation D.4 shows that the gradient of E with respect to $\mathbf{u}$ that maximizes the energy variation represented by the Gâteaux derivative $D_\eta E$ is then:

$$\nabla_\mathbf{u}E(\mathbf{x}) = K_\sigma*\left(\sum_{n=1}^{N} \delta(\phi_n \circ G_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t.\nabla\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}) \right) + \lambda\mathbf{u}(\mathbf{x})$$

$$\tag{D.5}$$

D.1.2   *With the augmented Lagrangian method*

We recall the problem to derive (Equation 4.17):

$$\min_{\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}} \hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}]$$

with $\hat{E}_k[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] = E[\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}] - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \alpha_{i,j}^k.C(\mathcal{G}_i,\mathcal{G}_j) + \dfrac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i,\mathcal{G}_j)^2$

We define

$$V_L = - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \alpha_{i,j}^k.C(\mathcal{G}_i,\mathcal{G}_j) + \frac{\mu_k}{2} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i,\mathcal{G}_j)^2$$

The derivatives $\nabla_{p_{n,k}} E$ and $\nabla_u E_k$ are identical than in the quadratic penalty formulation (see Section D.1.1), while the derivation of V_L varies. We detail it below.

$$\nabla_{p_{n,l}} V_L = - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \frac{\partial}{\partial p_{n,l}} \left(\alpha_{i,j}^k.C(\mathcal{G}_i,\mathcal{G}_j) \right) + \frac{\mu_k}{2} \frac{\partial}{\partial p_{n,l}} \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} C(\mathcal{G}_i,\mathcal{G}_j)^2$$

$$= - \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \alpha_{i,j}^k.\frac{\partial}{\partial p_{n,l}} C(\mathcal{G}_i,\mathcal{G}_j)$$

$$+ \mu_k \sum_{\substack{1 \leq i \leq N \\ i \neq n}} C(\mathcal{G}_i,\mathcal{G}_n).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x}$$

$$= - \sum_{\substack{1 \leq i \leq N \\ i \neq n}} \alpha_{i,n}^k.\frac{\partial}{\partial p_{n,l}} C(\mathcal{G}_i,\mathcal{G}_n)$$

$$+ \mu_k \sum_{\substack{1 \leq i \leq N \\ i \neq n}} C(\mathcal{G}_i,\mathcal{G}_n).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x}$$

$$= \sum_{\substack{1 \leq i \leq N \\ i \neq n}} \left(\mu_k C(\mathcal{G}_i,\mathcal{G}_n) - \alpha_{i,n}^k \right).\int_{\Omega} H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x}$$

$$\tag{D.6}$$

D.2   DERIVATION DETAILS FOR THE FUNCTIONAL UNDER USER CONSTRAINTS

In this section we detail the computations of the gradient equations given in Equation 5.11:

$$\frac{\partial \hat{E}_k}{\partial p_{n,l}} = \int_{\Omega} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x}) \rangle.f_n(\mathbf{x}) d\mathbf{x}$$

$$+ \sum_{\substack{1 \leq i \leq N \\ i \neq n}} \frac{\partial h(C(\mathcal{G}_i,\mathcal{G}_n), \alpha_{i,n}^k; \mu_k)}{\partial p_{n,l}} + \sum_{1 \leq q_n \leq K_n} \frac{\partial h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k)}{\partial p_{n,l}}$$

and Equation 5.12:

$$\nabla_u \hat{E}_k(\mathbf{x}) = K_\sigma * \left\{ \sum_{n=1}^{N} \delta(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x})).f_n(\mathbf{x}).(d\mathcal{G}_n)^t \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}) \right\} + \\ \lambda \mathbf{u}(\mathbf{x}) + \sum_{1 \leq q_n \leq K_n} \nabla_u h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k)$$

The function h is detailed in Equation 5.9 and its derivative is detailed in Equation 5.10.

D.2.1 *Details for the computation of Equation 5.11*

Let us detail the computation for the last two terms of the first equation (Equation 5.11). The first term, related to the non-overlapping constraints, is computed as follows:

$$\frac{\partial h(C(\mathcal{G}_i, \mathcal{G}_n), \alpha^k_{i,n}; \mu_k)}{\partial p_{n,l}} = \begin{cases} (\mu_k C(\mathcal{G}_i, \mathcal{G}_n) - \alpha^k_{i,n}) \frac{\partial C(\mathcal{G}_i, \mathcal{G}_n)}{\partial p_{n,l}} & \text{if } C(\mathcal{G}_i, \mathcal{G}_n) - \frac{\alpha^k_{i,n}}{\mu_k} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

(D.7)

with

$$\frac{\partial C(\mathcal{G}_i, \mathcal{G}_n)}{\partial p_{n,l}} = \int_\Omega H(\phi_i \circ \mathcal{G}_i(\mathbf{x})).\delta(\phi_n \circ \mathcal{G}_n(\mathbf{x})).\langle \nabla \phi_n \circ \mathcal{G}_n(\mathbf{x}), \frac{\partial \mathcal{G}_n(\mathbf{x})}{\partial p_{n,l}} \rangle d\mathbf{x} \quad (D.8)$$

The second term, related to the user constraints, is computed as follows:

$$\frac{\partial h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k)}{\partial p_{n,l}} = \begin{cases} (\mu'_k F_{q_n}(\mathbf{x}_{q_n}) - \alpha'_{q_n}) \frac{\partial F_{q_n}(\mathbf{x}_{q_n})}{\partial p_{n,l}} & \text{if } F_{q_n}(\mathbf{x}_{q_n}) - \frac{\alpha'_{q_n}}{\mu'_k} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

(D.9)

with

$$\frac{\partial F_{q_n}(\mathbf{x}_{q_n})}{\partial p_{n,l}} = \gamma_{q_n} \langle \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}), \frac{\partial \mathcal{G}_n}{\partial p_{n,l}} \circ \mathcal{L}(\mathbf{x}) \rangle \quad (D.10)$$

D.2.2 *Details for the computation of Equation 5.12*

Let us detail the computation for the last term of the second equation (Equation 5.12).

$$\nabla_u h(F_{q_n}(\mathbf{x}_{q_n}), \alpha'_{q_n}, \mu'_k) = \begin{cases} (\mu'_k F_{q_n}(\mathbf{x}_{q_n}) - \alpha'_{q_n}) \nabla_u F_{q_n}(\mathbf{x}_{q_n}) & \text{if } F_{q_n}(\mathbf{x}_{q_n}) - \frac{\alpha'_{q_n}}{\mu'_k} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

And following the same type of derivation method as in with Gâteaux derivatives we get:

$$\nabla_u F_{q_n}(\mathbf{x}_{q_n}) = K_\sigma * \left(\gamma_{q_n} (d\mathcal{G}_n)^t \nabla \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n}) \right) \tag{D.11}$$

A METHOD TO BUILD MEAN SHAPE MODELS

In this appendix, we propose a method to construct mean shape models from implicit representations of shape samples. Given a set of shape samples represented with implicit functions, the idea is to register these samples on a reference using a robust version of the iterative closest point (ICP) (Fitzgibbon, 2003). Then the registered samples are averaged and the corresponding zero-level corresponds to the mean shape. We detail the method for 3D shapes, the same procedure can be also used in 2D.

A ROBUST ICP   The ICP is a method introduced by Besl and McKay (1992) to register a point cloud on another reference point cloud. The idea is to find the global transformation between the two sets of points that minimizes the distance between them. Multiple variants of the algorithm exist and can be found in (Rusinkiewicz and Levoy, 2001; Wild, 2010; Oswald, 2010).

A robust version, introduced by Fitzgibbon (2003), gives good and fast results thanks to the use of distance function representation. We detail this method hereafter. Note that the *marching cube* algorithm (Lorensen and Cline, 1987) enables to turn the implicit representation into a point-based representation of the object's surface.

Let use define a reference shape $\mathcal{R}$ represented by the zero-level of a distance function $\phi_r : \mathbb{R}^3 \to \mathbb{R}$ and a shape model to register $\mathcal{M}$ represented by a set of N_d points of $\mathbb{R}^3$ denoted by $\{\mathbf{m}_i\}_{i=1}^{N_d}$. For the sake of simplicity, we will call *reference* the reference shape and *model* the shape model to register. We want to find the global transformation $T(\mathbf{a}) : \mathbb{R}^3 \to \mathbb{R}^3$, where $\mathbf{a} = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, s]$ are the translation, rotation and scale parameters, that minimizes the sum of the distances between the two shapes.

The Euclidean distance d_E between a point $\mathbf{m}_i$ of the model transformed by $T(\mathbf{a})$ and the reference $\mathcal{R}$ is defined as:

$$d_E(T(\mathbf{m}_i; \mathbf{a}), \mathcal{R}) = \phi_m(T(\mathbf{m}_i; \mathbf{a})) \tag{E.1}$$

We introduce the error function $\varepsilon : \mathbb{R}^+ \to \mathbb{R}^+$ such that $||x|| \to \varepsilon^2(|x|)$ (where $x \in \mathbb{R}$) and ε is a monotonic increasing function. The ε-distance d_ε between a transformed point $T(\mathbf{m}_i; \mathbf{a})$ and the reference $\mathcal{R}$ can be defined as:

$$d_\varepsilon(T(\mathbf{m}_i; \mathbf{a}), \mathcal{R}) = \varepsilon^2(\phi_m(T(\mathbf{m}_i; \mathbf{a}))) \tag{E.2}$$
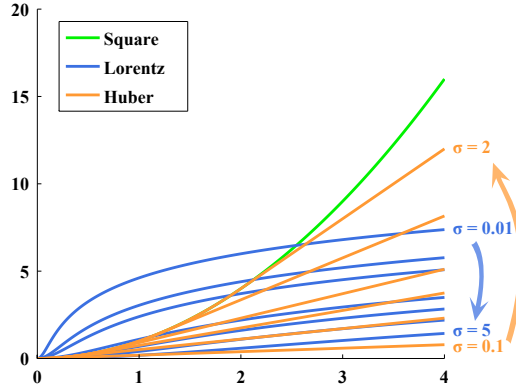
Figure E.1: Lorentz, Huber and square error functions.

Finding the best transformation $T(\mathbf{a})$ registering $\mathcal{M}$ on $\mathcal{R}$ can then be formulated as finding the transformation minimizing the following energy:

$$E(\mathbf{a}) = \sum_{i=1}^{N_d} \varepsilon^2(T(\mathbf{m}_i, \mathbf{a})) \tag{E.3}$$

Fitzgibbon (2003) uses two classical error functions: the Lorentz function and the Huber function, respectively defined as:

$$\varepsilon_l(r) = \log(1 + \frac{r^2}{\sigma}) \quad \text{and} \quad \varepsilon_h(r) = \begin{cases} r^2 & \text{if } r < \sigma \\ 2\sigma|r| - \sigma^2 & \text{otherwise} \end{cases} \tag{E.4}$$

The Huber function is a function equal to the square function below a certain threshold σ and equal to the identity function above this threshold. Figure E.1 illustrates these error functions for different values of σ and compared to the square function. We observe that they increase less rapidly thus limiting the negative effect of outliers.

Fitzgibbon (2003) optimizes the energy function Equation E.3 by using a non-linear approach with the Levenberg-Marquardt algorithm (Marquardt, 1963). This algorithm interpolates between the Gauss-Newton and the gradient descent methods. It has the advantage of having a larger capture range, meaning that the initialization can be farther from the minimum to reach.

EXAMPLE   To show the interest of introducing robust error functions in the ICP algorithm, we propose an experiment in which we try to register a manual segmentation mesh of a liver and a modified version of this mesh. To modify the initial mesh we added a sphere, supposed to simulate a strong local deformation (e.g. a tumor). In Figure E.2 we show the reference mesh and the modified mesh and the different results obtained with the Lorentz, Huber and square error functions.

(a) Reference   (b) Model to register

(c) Lorentz   (d) Huber   (e) Square
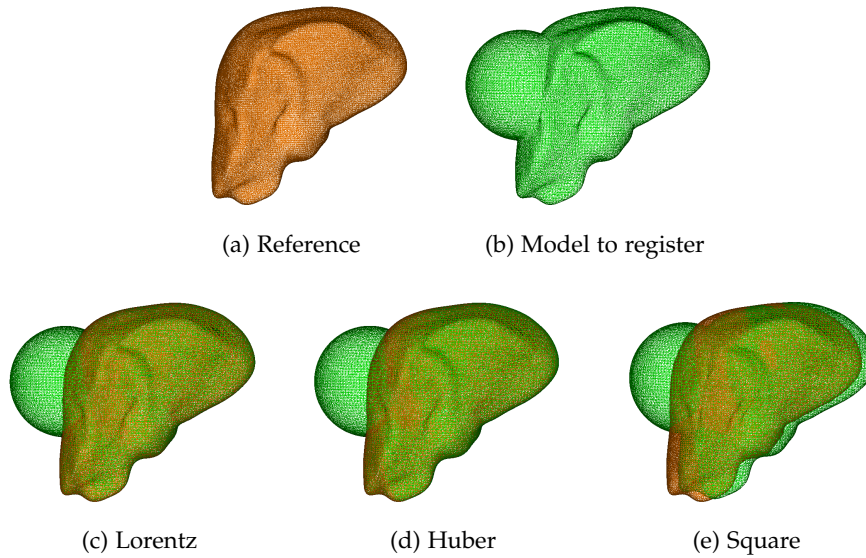
Figure E.2: Registration examples of a mesh model (b) to a reference (a) mesh using Lorentz (c), Huber (d) and square (e) error functions.

We observe the averaging effect of the outliers with the square function, which does not happen with the Huber and Lorentz functions. Note that the value of the parameter σ controls this averaging effect.

BUILD A MEAN LIVER SHAPE   The goal of our experiments was to build a mean shape of the liver from multiple manual segmentation samples. The liver is a soft organ with high shape variabilities. Determining a common referential for the different liver shape samples is not straightforward. One common observation is that the liver and the lungs present a smooth and regular interface at the top of the liver. We chose to take this part of the liver as a reference to register all the samples in the same referential. The robust registration presented before is then used to perform this task. In practice we observed that the Lorentz function gave more satisfying results. The procedure is simple, given a database of liver shapes, we choose arbitrarily a reference having a *standard* shape. Then the other liver shapes are registered on this reference using the robust ICP of Fitzgibbon (2003). In practice we observed that the choice of the reference had no significant influence on the final mean shape result.

After the registration step, the shape samples distance functions are averaged. Note that the space of distance functions is not linear. To solve this issue, a distance function with the same zero-level as the resulting function can be approximated using the method of Sussman et al. (1994).

In another experiment we tried to see the influence of the number of shape samples on the resulting mean shape for the liver. We used a database of 90
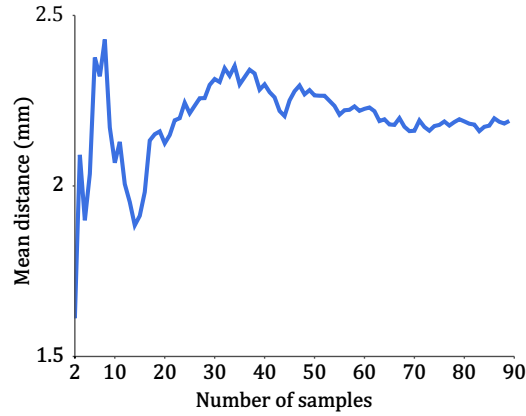
Figure E.3: Average distance of the mean shape to the first sample in function of the number of samples used to build the mean shape.

different liver shapes obtained from manual segmentation in CT volumes (from different patients). We computed the mean liver shape starting from 2 samples and adding each time one new sample, until reaching 90 samples. For each mean shape we compute its average distance to the first sample. This is a way to observe the changes between different mean shapes. The results are given in Figure E.3. From about 20 samples, the resulting mean shape seems to be stable. This is confirmed in Figure E.4 where we show mean liver models built from various number of samples.

CONCLUSION   We presented a simple method to build mean shapes from shape samples in 3D. It consists in registering the shape samples on a reference shape using a robust ICP approach. Then the distance functions of the shape samples are averaged and the result is turned into a distance function (which can be used directly in our implicite template deformation framework, Chapter 4). We also showed that starting from 20 samples, the mean shape seems to be stable despite the high variability of the liver. The intuition that there is generally no need for particularly large databases to build stable models was confirmed on other organs (e. g. kidneys, spleen).

(a) 10 samples   (b) 20 samples   (c) 30 samples

(d) 40 samples   (e) 50 samples   (f) 60 samples

(g) 70 samples   (h) 80 samples   (i) 90 samples

Figure E.4: Mean liver shapes built from various numbers of samples.

BIBLIOGRAPHY

C. Alvino, G. Unal, G. Slabaugh, B. Peny, and T. Fang. Efficient segmentation based on eikonal and diffusion equations. *International Journal of Computer Mathematics*, 84(9):1309–1324, 2007. (Cited on page 115.)

Y. Amit and D. Geman. Randomized inquiries about shape; an application to handwritten digit recognition. Technical report 401, Dept of Statistics, University of Chicago, 1994. (Cited on page 14.)

J.-H. An and Y. Chen. Region based image segmentation using a modified Mumford-Shah algorithm. In *Scale Space and Variational Methods in Computer Vision*, pages 733–742. Springer, 2007. (Cited on page 29.)

S. Andonova, A. Elisseeff, T. Evgeniou, and M. Pontil. A simple algorithm for learning stable machines. In *European Conference on Artificial Intelligence (ECAI)*, pages 513–517, 2002. (Cited on page 50.)
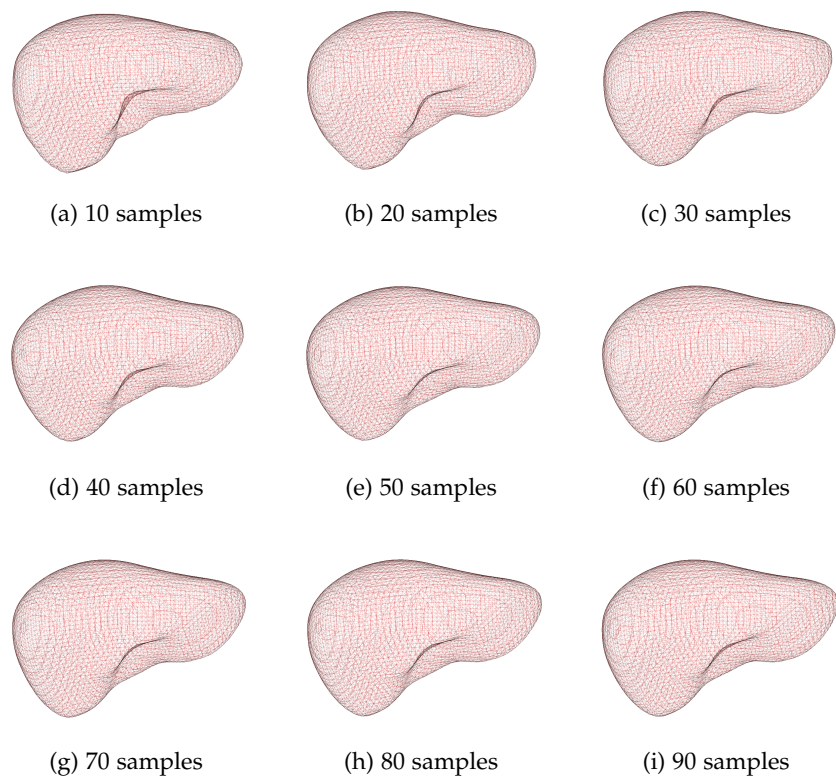
K. P. Andriole, J. M. Wolfe, R. Khorasani, S. T. Treves, D. J. Getty, F. L. Jacobson, M. L. Steigner, J. J. Pan, A. Sitek, and S. E. Seltzer. Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section CT scan can ruin your whole day. *Radiology*, 259(2):346–362, 2011. (Cited on page 37.)

I. Arel, D. C. Rose, and T. P. Karnowski. Deep machine learning - a new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine*, 5 (4):13–18, 2010. (Cited on page 12.)

X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. (Cited on page 115.)

D. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981. (Cited on page 17.)

D. Barbosa, T. Dietenbeck, J. Schaerer, J. D'hooge, D. Friboulet, and O. Bernard. B-spline explicit active surfaces: An efficient framework for real-time 3-D region-based segmentation. *IEEE Transactions on Image Processing*, 21(1):241–251, 2012. (Cited on page 118.)

D. Barbosa, B. Heyde, M. Cikes, T. Dietenbeck, P. Claus, D. Friboulet, O. Bernard, and J. D'hooge. Real-time 3D interactive segmentation of echocardiographic data through user-based deformation of B-spline explicit active surfaces. *Computerized Medical Imaging and Graphics*, 38(1):57–67, 2014. (Cited on page 118.)

N. Ben-Zadok, T. Riklin-Raviv, and N. Kiryati. Interactive level set segmentation for image-guided therapy. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1079–1082, 2009. (Cited on page 117.)

P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analaysis and Machine Intelligence*, 14(2):239–256, 1992. (Cited on pages 141 and 165.)

S. Beucher and F. Meyer. The morphological approach to segmentation: the watershed transformation. *Optical Engineering*, 34:433–481, 1992. (Cited on page 114.)

E. G. Birgin, R. A. Castillo, and J. M. Martinez. Numerical comparison of augmented lagrangian algorithms for nonconvex problems. *Computational Optimization and Applications*, 31:31–56, 2004. (Cited on page 81.)

A. Blake. *Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer, London ; New York, 1998. (Cited on page 27.)

D. Blezek and J. Miller. Atlas stratification. *Medical Image Analysis*, 11:443–457, 2007. (Cited on page 24.)

J. A. Bogovic, J. L. Prince, and P.-L. Bazin. A multiple object geometric deformable model for image segmentation. *Computer Vision and Image Understanding*, 117(2):145–157, 2012. (Cited on page 31.)

Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. In S. L. Delp, A. M. DiGoia, and B. Jaramaz, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 1985, pages 276–286. Springer Berlin Heidelberg, 2000. (Cited on pages 25 and 115.)

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (11):1222–1239, 2001. (Cited on page 25.)

Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 105–112, 2001. (Cited on pages 25 and 116.)

K. Bredhal, A. Long, M. Taudorf, L. Lönn, L. Rouet, R. Ardon, H. Sillesen, and J. Eiberg. Volume estimation of aortic sac after EVAR using 3-D ultrasound – a novel, accurate and promising technique. *European Journal of Vascular and Endovascular Surgery*, 45(5):456, 2013. (Cited on pages 153 and 154.)

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. (Cited on page 15.)

L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984. (Cited on page 12.)

X. Bresson, S. Esedoḡlu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007. (Cited on page 116.)

J. Brownlee. Best machine learning resources for getting started - machine learning mastery. http://machinelearningmastery.com/best-machine-learning-resources-for-getting-started/, 2014. Accessed: 2014-12-08. (Cited on page 17.)

M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, 104(3):158–177, 2011. (Cited on page 24.)

P. Campadelli, E. Casiraghi, S. Pratissoli, and G. Lombardi. Automatic abdominal organ segmentation from CT images. *Electronic Letters on Computer Vision and Image Analysis*, 8(1):1–14, 2009a. (Cited on page 8.)

P. Campadelli, E. Casiraghi, and A. Esposito. Liver segmentation from computed tomography scans: A survey and a new algorithm. *Artificial Intelligence in Medicine*, 45(2-3):185–196, 2009b. (Cited on page 139.)

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. (Cited on page 92.)

R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168. ACM, 2006. (Cited on page 10.)

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997. (Cited on pages 23 and 27.)

T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. (Cited on pages 29 and 84.)

V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009. (Cited on page 66.)

C. Chen and G. Zheng. Fully automatic segmentation of AP pelvis X-rays via random forest regression with efficient feature selection and hierarchical sparse shape composition. *Computer Vision and Image Understanding*, 126:1–10, 2014. (Cited on pages xx, 20, and 137.)

X. Chen, J. K. Udupa, A. Alavi, and D. A. Torigian. GC-ASM: Synergistic integration of graph-cut and active shape model strategies for medical image

segmentation. *Computer Vision and Image Understanding*, 117(5):513–524, 2013. (Cited on page 32.)

L. D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, 1991. (Cited on page 28.)

D. Collins, A. Zijdenbos, W. Baaré, and A. Evans. ANIMAL+INSECT: improved cortical structure segmentation. In *Information Processing in Medical Imaging (IPMI)*, pages 210–223, 1999. (Cited on page 23.)

T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proceedings of British Machine Vision Conference*, volume 9, page 18, 1992. (Cited on page 30.)

T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995. (Cited on page 30.)

T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer. Robust and accurate shape model fitting using random forest regression voting. In A. Fitzgibbon, S. Lazebnik, P. Perona, S. Yoichi, and C. Schmid, editors, *European Conference on Computer Vision (ECCV)*, pages 278–291. Springer, 2012. (Cited on pages xx and 137.)

C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. (Cited on page 10.)

C. Couprie, L. Grady, L. Najman, and H. Talbot. Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In *IEEE International Conference on Computer Vision*, pages 731–738, 2009. (Cited on page 114.)

R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1):1–23, 1943. (Cited on page 79.)

J. Cousty, G. Bertrand, L. Najman, and M. Couprie. Watershed cuts: Minimum spanning forests and the drop of water principle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1362–1374, 2009. (Cited on page 114.)

D. Cremers, O. Fluck, M. Rousson, and S. Aharon. A probabilistic level set formulation for interactive organ segmentation. In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Medical Imaging*, volume 6512, 2007. (Cited on page 116.)

A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis*, 17(8):1293–1303,

2013. (Cited on pages xiii, xiv, xviii, 22, 38, 39, 41, 43, 44, 46, 48, 49, 61, 63, 64, 66, 135, 139, 140, 141, and 147.)

A. Criminisi and J. Shotton. *Decision forests for computer vision and medical image analysis*. Springer, London; New York, 2013. (Cited on pages 13 and 14.)

A. Criminisi, T. Sharp, and A. Blake. Geos: Geodesic image segmentation. In D. Forsyth, P. Torr, and A. Zisserman, editors, *European Conference on Computer Vision (ECCV)*, pages 99–112. Springer, 2008. (Cited on page 115.)

A. Criminisi, J. Shotton, and S. Bucciarelli. Decision forests with long-range spatial context for organ localization in CT volumes. In *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, pages 69–80, 2009. (Cited on page 39.)

A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu. Regression forests for efficient anatomy detection and localization in CT studies. In B. Menze, G. Langs, Z. Tu, and A. Criminisi, editors, *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, volume 6533 of *LNCS*, pages 106–117. Springer Berlin Heidelberg, 2011. (Cited on pages 22, 38, 39, and 41.)

R. Cuingnet, R. Prevost, D. Lesage, L. Cohen, B. Mory, and R. Ardon. Automatic detection and segmentation of kidneys in 3D CT images using random forests. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7512 of *LNCS*, pages 66–74. Springer Berlin Heidelberg, 2012. (Cited on pages v, vii, xiii, xiv, xviii, 22, 38, 41, 44, 46, 58, 61, 63, 64, 75, 78, 135, and 140.)

M. Dantone, J. Gall, G. Fanelli, and L. Van Gool. Real-time facial feature detection using conditional regression forests. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2578–2585, 2012. (Cited on page 16.)

F. Ding, W. Leow, and S. C. Wang. Segmentation of 3D CT volume images using a single 2D atlas. *Computer Vision for Biomedical Image Applications*, pages 459–468, 2005. (Cited on page 24.)

P. Dollár, P. Welinder, and P. Perona. Cascaded pose regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1078–1085, 2010. (Cited on pages 16, 22, and 66.)

P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012. (Cited on page 17.)

V. Duay, N. Houhou, and J. P. Thiran. Atlas-based segmentation of medical images locally constrained by level sets. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages II–1286, 2005. (Cited on page 24.)

A. C. Evans, W. Dai, D. L. Collins, P. Neelin, and S. Marrett. Warping of a computerized 3-D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. In *Medical Imaging V: Image Processing*, pages 236–246. International Society for Optics and Photonics, 1991. (Cited on page 23.)

A. X. Falcao and J. K. Udupa. Segmentation of 3D objects using live wire. In K. M. Hanson, editor, *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, pages 228–235, 1997. (Cited on page 113.)

A. X. Falcão, J. Stolfi, and R. de Alencar Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29, 2004. (Cited on page 115.)

A. Falcao, J. Udupa, and F. Miyazawa. An ultra-fast user-steered image segmentation paradigm: live wire on the fly. *IEEE Transactions on Medical Imaging*, 19 (1):55–62, 2000. (Cited on page 113.)

X. Fan, P. L. Bazin, J. Bogovic, Y. Bai, and J. L. Prince. A multiple geometric deformable model framework for homeomorphic 3D medical image segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–7, 2008. (Cited on page 31.)

M. Fenchel, S. Thesen, and A. Schilling. Automatic labeling of anatomical structures in MR FastView images using a statistical atlas. In D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5241 of *LNCS*, pages 576–584. Springer Berlin Heidelberg, 2008. (Cited on page 24.)

M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15:3133–3181, 2014. (Cited on page 17.)

A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(13-14):1145–1153, 2003. (Cited on pages 141, 165, 166, and 167.)

G. Fouquier, J. Atif, and I. Bloch. Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. *Computer Vision and Image Understanding*, 116(1):146–165, 2012. (Cited on page 19.)

Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995. (Cited on page 14.)

K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975. (Cited on page 67.)

J. Gall and V. Lempitsky. Class-specific hough forest for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1022–1029, 2009. (Cited on pages 39, 42, and 49.)

Y. Gao, A. Tannenbaum, and R. Kikinis. Simultaneous multi-object segmentation using local robust statistics and contour interaction. *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pages 195–203, 2011. (Cited on page 33.)

R. Gauriau, R. Cuingnet, R. Prevost, B. Mory, R. Ardon, D. Lesage, and I. Bloch. A generic, robust and fully-automatic workflow for 3D CT liver segmentation. In H. Yoshida, S. Warfield, and M. W. Vannier, editors, *Abdominal Imaging. Computation and Clinical Applications*, volume 8198 of *LNCS*, pages 241–250. Springer Berlin Heidelberg, 2013. (Cited on pages 37, 98, 102, and 139.)

R. Gauriau, B. Husson, K. Deva, I. Bloch, and C. Adamsbaum. Segmentation du corps calleux : un outil pronostique des pathologies inflammatoires cérébrales ? In *Journées Françaises de Radiologie*, Paris, France, 2013.

R. Gauriau, R. Ardon, D. Lesage, and I. Bloch. Multiple template deformation. Application to abdominal organ segmentation. In *International Symposium on Biomedical Imaging (ISBI)*, New York, USA, 2015a. (Cited on pages xvi, 69, and 100.)

R. Gauriau, D. Lesage, M. Chiaradia, B. Morel, and I. Bloch. Interactive multi-organ segmentation based on multiple template deformation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS. Springer, 2015b. (Cited on page 111.)

R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch. Multi-organ localization combining global-to-local regression and confidence maps. In P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8675 of *LNCS*, pages 337–344. Springer Berlin Heidelberg, 2014. (Cited on pages xiv and 35.)

R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch. Multi-organ localization with cascaded global-to-local regression and shape prior. *Medical Image Analysis*, 23:70–83, 2015. (Cited on pages xiv and 35.)

J. C. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225–236, 1993. (Cited on page 23.)

P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006. (Cited on page 15.)

B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi. Joint classification-regression forests for spatially structured multi-object segmentation. In *European Conference on Computer Vision (ECCV)*, volume IV, pages 870–881. Springer-Verlag, 2012. (Cited on pages 16 and 26.)

L. Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006. (Cited on pages 114 and 115.)

L. Grady and G. Funka-Lea. An energy minimization approach to the data driven editing of presegmented images/volumes. In R. Larsen, M. Nielsen, and J. Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4191 of *LNCS*, pages 888–895. Springer, 2006. (Cited on page 117.)

H. a. Grahn, N. Lavesson, M. H. Lapajne, and D. Slat. Cudarf: A cuda-based implementation of random forests. In *9th ACS International Conference on Computer Systems and Applications (AICCSA)*, pages 95–101. IEEE, 2011. (Cited on page 15.)

V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3129–3136, 2010. (Cited on page 116.)

G. Hamarneh, J. Yang, C. McIntosh, and M. Langille. 3D live-wire-based semi-automatic segmentation of medical images. In J. M. Fitzpatrick and J. M. Reinhardt, editors, *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, pages 1597–1603, 2005. (Cited on page 114.)

R. M. Haralick. Digital step edges from zero crossing od second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1): 58–68, 1984. (Cited on page 92.)

A. P. Harrison, N. Birkbeck, and M. Sofka. Intelledits: Intelligent learning-based editor of segmentations. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8151 of *LNCS*, pages 235–242. Springer Berlin Heidelberg, 2013. (Cited on page 117.)

T. Hastie. *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer, New York, NY, 2nd edition, 2009. (Cited on pages 9, 10, and 17.)

K. He, J. Sun, and X. Tang. Guided image filtering. *European Conference on Computer Vision (ECCV)*, pages 1–14, 2010. (Cited on page 144.)

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006. (Cited on page 24.)

T. Heimann and et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009. (Cited on pages 139, 140, and 145.)

T. Heimann, M. Styner, and B. van Ginneken. Sliver07. http://www.sliver07.org, 2007. Accessed: 2013-05-12. (Cited on pages 139, 140, and 145.)

T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, 2009. (Cited on pages 27 and 30.)

L. Hermoye, I. Laamari-Azjal, Z. Cao, L. Annet, J. Lerut, B. M. Dawant, and B. E. Van Beers. Liver segmentation in living liver transplant donors: Comparison of semiautomatic and manual methods. *Radiology*, 234(1):171–178, 2005. (Cited on page 139.)

M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969. (Cited on page 81.)

G. E. Hinton. Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10):428–434, 2007. (Cited on page 12.)

T. K. Ho. Random decision forests. In *IEEE Conference on Document Analysis and Recognition*, volume 1, pages 278–282, 1995. (Cited on page 14.)

T. K. Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998. (Cited on page 14.)

A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA: Toolkit for Multivariate Data Analysis. *ArXiv Physics e-prints*, 0703039, 2007. (Cited on page 15.)

C. Huang, F. Jia, Y. Li, X. Zhang, H. Luo, C. Fang, and Y. Fan. Fully automatic liver segmentation using probability atlas registration. In *International Conference on Electronics, Communications and Control 2012*, pages 126–129, 2012. (Cited on pages 145 and 146.)

X. Huang and D. N. Metaxas. Metamorphs: deformable shape and appearance models. *Pattern Analysis and Machine Intelligence*, 30(8):1444–1459, 2008. (Cited on page 29.)

S. Hugueny and M. Rousson. Hierarchical detection of multiple organs using boosted features. In *Computer Analysis of Images and Patterns*, pages 317–325, 2007. (Cited on page 20.)

J. Iglesias, E. Konukoglu, A. Montillo, Z. Tu, and A. Criminisi. Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In *Information Processing in Medical Imaging (IPMI)*, pages 25–36, 2011. (Cited on page 26.)

H. Jacinto, R. Kéchichian, M. Desvignes, R. Prost, and S. Valette. A web interface for 3D visualization and interactive segmentation of medical images. In *Proceedings of the 17th International Conference on 3D Web Technology*, pages 51–58. ACM, 2012. (Cited on page 115.)

Y. Jang, H. Hong, J. W. Chung, and Y. H. Yoon. Automatic segmentation of the liver using multi-planar anatomy and deformable surface model in abdominal contrast-enhanced CT images. In D. R. Haynor and S. Ourselin, editors, *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, volume 8314, page 831436, 2012. (Cited on page 17.)

D. Jayadevappa, S. S. Kumar, and D. S. Murty. Medical image segmentation algorithms using deformable models: A review. *IETE Technical Review*, 28(3): 248–255, 2011. (Cited on page 27.)

S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004. (Cited on page 24.)

D. Kainmüller, T. Lange, and H. Lamecker. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In *3D Segmentation in the Clinic - A Grand Challenge*, pages 109–116, 2007. (Cited on pages 145 and 146.)

H. Kalinic. Atlas-based image segmentation: A survey. Technical report, Department of Electronic Systems and Information Processing, Universiy of Zagreb, 2008. (Cited on page 24.)

Y. Kang, K. Engelke, and W. A. Kalender. Interactive 3D editing tools for image segmentation. *Medical Image Analysis*, 8(1):35–46, 2004. (Cited on page 117.)

M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, pages 321–331, 1988. (Cited on page 27.)

R. Kéchichian, S. Valette, M. Desvignes, and R. Prost. Efficient multi-object segmentation of 3D medical images using clustering and graph cuts. In *18th IEEE International Conference on Image Processing (ICIP)*, pages 2149–2152, 2011. (Cited on page 115.)

R. Kimmel. Fast edge integration. In *Geometric Level Set Methods in Imaging, Vision, and Graphics*, pages 59–77. Springer New York, 2003. (Cited on page 91.)

R. Kimmel and A. M. Bruckstein. Regularized Laplacian zero crossings as optimal edge integrators. *International Journal of Computer Vision*, 53(3):225–243, 2003. (Cited on pages xv, 91, 92, 93, 95, and 108.)

M. Kobashi and L. G. Shapiro. Knowledge-based organ identification from CT images. *Pattern Recognition*, 28(4):475–491, 1995. (Cited on page 23.)

H. Kobatake and S. Nawano. Probabilistic atlas-guided eigen-organ method for simultaneous bounding box estimation of multiple organs in volumetric CT images. *Medical Imaging Technology*, 24(3):191–200, 2006. (Cited on page 21.)

T. Kohlberger, M. Sofka, J. Zhang, N. Birkbeck, J. Wetzl, J. Kaftan, J. Declerck, and S. Zhou. Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6893 of *LNCS*, pages 338–345. Springer Berlin Heidelberg, 2011. (Cited on pages 31, 32, and 104.)

P. Kontschieder, S. R. Bulò, A. Criminisi, P. Kohli, M. Pelillo, and H. Bischof. Context-sensitive decision forests for object detection. In *Advances in neural information processing systems*, pages 431–439, 2012. (Cited on page 16.)

P. Kontschieder, P. Kohli, J. Shotton, and A. Criminisi. Geof: Geodesic forests for learning coupled predictors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 65–72, 2013. (Cited on page 16.)

E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi. Neighbourhood approximation using randomized forests. *Medical image analysis*, 17(7):790–804, 2013. (Cited on pages 16 and 24.)

S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268, 2007. (Cited on page 17.)

N. Lay, N. Birkbeck, J. Zhang, and S. Zhou. Rapid multi-organ segmentation using context integration and discriminative models. In J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells, and L. Zöllei, editors, *Information Processing in Medical Imaging (IPMI)*, volume 7917 of *LNCS*, pages 450–462. Springer Berlin Heidelberg, 2013. (Cited on page 21.)

C. C. Lee, P. C. Chung, and H. M. Tsai. Identifying multiple abdominal organs from CT image series using a multimodule contextual neural network and spatial fuzzy rules. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):208–217, 2003. (Cited on page 26.)

M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, 2000. (Cited on page 30.)

S. Z. Li. *Markov random field modeling in Image Analysis*. Springer, first edition edition, 1995. (Cited on page 25.)

M. C. Lin. *Efficient collision detection for animation and robotics*. PhD thesis, University of California, Berkeley, CA, 1993. (Cited on page 85.)

H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu. Hierarchical, learning-based automatic liver segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. (Cited on page 140.)

M. G. Linguraru and R. M. Summers. Multi-organ automatic segmentation in 4D contrast-enhanced abdominal CT. In *International Symposium on Biomedical Imaging*, pages 45–48, 2008. (Cited on page 23.)

M. G. Linguraru, J. A. Pura, V. Pamulapati, and R. M. Summers. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Medical Image Analysis*, 16(4):904–914, 2012a. (Cited on pages 145 and 146.)

M. G. Linguraru, J. A. Pura, V. Pamulapati, and R. M. Summers. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase 4D. *Medical Image Analysis*, 2012b. (Cited on page 25.)

M. G. Linguraru, J. K. Sandberg, Z. Li, F. Shah, and R. M. Summers. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. *Medical Physics*, 37(2):771, 2010. (Cited on page 140.)

A. Litvin and W. Karl. Coupled shape distribution-based segmentation of multiple objects. In *Information Processing in Medical Imaging (IPMI)*, pages 171–233, 2005. (Cited on pages 31 and 32.)

D. Liu and S. K. Zhou. Anatomical landmark detection using nearest neighbor matching and submodular optimization. In *Medical Image Computing and Computer-Assisted Intervention*, pages 393–401. Springer, 2012. (Cited on page 20.)

D. Liu, K. S. Zhou, D. Bernhardt, and D. Comaniciu. Search strategies for multiple landmark detection by submodular maximization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2831–2838, 2010. (Cited on page 20.)

X. Liu, Q. Song, P. Mendonca, X. Tao, and R. Bhotika. Organ labeling using anatomical model-driven global optimization. In *First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, pages 338–345, 2011. (Cited on page 19.)

H. Lombaert, D. Zikic, A. Criminisi, N. Ayache, A. Criminisi, and D. Zikic. Laplacian forests: Semantic image segmentation by guided bagging. In P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8674 of *LNCS*, pages 496–504. Springer Berlin Heidelberg, 2014. (Cited on pages 16 and 24.)

A. Long, L. Rouet, A. Debreuve, R. Ardon, C. Barbe, J. Becquemin, and E. Allaire. Abdominal aortic aneurysm imaging with 3-D ultrasound: 3-D-based maximum diameter measurement and volume quantification. *Ultrasound in Medicine & Biology*, 39(8):1325–1336, 2013. (Cited on pages 153 and 154.)

W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *ACM Siggraph Computer Graphics*, volume 21, pages 163–169, 1987. (Cited on pages 84 and 165.)

G. Louppe. *Understanding Random Forests: From Theory to Practice*. PhD thesis, Université de Lièges, 2014. (Cited on page 15.)

C. Lu, Y. Zheng, N. Birkbeck, J. Zhang, T. Kohlberger, C. Tietjen, T. Boettger, J. Duncan, and S. Zhou. Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7511 of *LNCS*, pages 462–469. Springer Berlin Heidelberg, 2012. (Cited on pages 21 and 32.)

D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Society of Industrial and Applied Mathematics (SIAM) Journal*, 11(2):431–441, 1963. 00000. (Cited on page 166.)

D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217, 1980. (Cited on page 90.)

T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996. (Cited on page 27.)

A. M. Mharib, A. R. Ramli, S. Mashohor, and R. B. Mahmood. Survey on liver CT image segmentation methods. *Artificial Intelligence Review*, 37(2):83–95, 2011. (Cited on page 139.)

J. Montagnat, H. Delingette, and N. Ayache. A review of deformable surfaces: topology, geometry and deformation. *Image and Vision Computing*, 19(14):1023–1040, 2001. (Cited on page 27.)

A. Montillo, J. Shotton, J. Winn, J. Iglesias, D. Metaxas, and A. Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In *Information Processing in Medical Imaging (IPMI)*, pages 184–196, 2011. (Cited on pages 16 and 26.)

P. Moore and D. Molloy. A survey of computer-based deformable models. In *IEEE Conference on International Machine Vision and Image Processing (IMVIP)*, pages 55–66, Los Alamitos, Calif, 2007. (Cited on pages 26 and 27.)

J. N. Morgan and J. A. Sonquist. Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302):415–434, 1963. (Cited on page 12.)

E. Mortensen, B. Morse, W. Barrett, and J. Udupa. Adaptive boundary detection using live-wire two-dimensional dynamic programming. In *IEEE Proceedings of Computers in Cardiology*, pages 635–638, 1992. (Cited on page 113.)

B. Mory, O. Somphone, R. Prevost, and R. Ardon. Real-time 3D image segmentation by user-constrained template deformation. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7510 of *LNCS*, pages 561–568. Springer Berlin Heidelberg, 2012. (Cited on pages v, vii, xiv, xvii, xviii, 29, 69, 71, 72, 73, 74, 75, 76, 78, 79, 85, 108, 112, 117, 118, 119, 135, 139, 142, and 143.)

B. Mory. *Interactive Segmentation of 3D Medical Images with Implicit Surfaces*. PhD thesis, EPFL, 2011. (Cited on pages 71, 72, 73, 75, 76, 78, 88, and 118.)

B. Mory, R. Ardon, A. J. Yezzi, and J. Thiran. Non-euclidean image-adaptive radial basis functions for 3D interactive segmentation. In *IEEE International Conference on Computer Vision*, pages 787–794. IEEE, 2009. (Cited on page 117.)

D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989. (Cited on page 29.)

R. Murthy, R. Nunez, J. Szklaruk, W. Erwin, D. C. Madoff, S. Gupta, K. Ahrar, M. J. Wallace, A. Cohen, D. M. Coldwell, A. S. Kennedy, and M. E. Hicks. Yttrium-90 microsphere therapy for hepatic malignancy: Devices, indications, technical considerations, and potential complications. *Radiographics*, 25(suppl_1):S41–S55, 2005. (Cited on pages 37 and 139.)

O. Nempont, J. Atif, and I. Bloch. A constraint propagation approach to structural model based image segmentation and recognition. *Information Sciences*, 246:1–27, 2013. (Cited on page 19.)

T. Nguyen, J. Cai, J. Zhang, and J. Zheng. Robust interactive image segmentation using convex active contours. *IEEE Transactions on Image Processing*, 21 (8):3734–3743, 2012. (Cited on page 116.)

J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, New York, 2006. (Cited on pages 79, 81, 82, 119, 121, and 122.)

S. D. Olabarriaga and A. W. Smeulders. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 5(2):127–142, 2001. (Cited on page 112.)

S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988. (Cited on page 28.)

L. Oswald. Recent development of the iterative closest point algorithm from 2008 to 2010. Technical report, Eidgenössische Technische Hochschule Zürich, 2010. (Cited on page 165.)

M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. (Cited on page 15.)

N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):266–280, 2000. (Cited on page 31.)

H. Park, P. H. Bland, and C. R. Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging*, 22(4):483–492, 2003. (Cited on page 23.)

O. Pauly. *Random Forests for Medical Applications*. PhD thesis, Technischen Universität München, 2012. (Cited on page 8.)

O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. Möller, S. Nekolla, and N. Navab. Fast multiple organ detection and localization in whole-body MR dixon sequences. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6893 of *LNCS*, pages 239–247. Springer Berlin Heidelberg, 2011. (Cited on page 22.)

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. (Cited on pages 15 and 17.)

P. Perez et al. *Markov random fields and images*. Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), 1998. (Cited on page 25.)

M. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic press, new york edition, 1969. (Cited on page 81.)

R. Prevost. *Variational methods for model-based image segmentation - application in medical imaging*. PhD thesis, Université Paris-Dauphine, CEREMADE - Centre de Recherche en Mathématiques de la Décision, 2013. (Cited on pages 72, 74, 75, 76, 78, 79, 80, 84, and 161.)

R. Prevost, R. Cuingnet, B. Mory, L. D. Cohen, and R. Ardon. Incorporating shape variability in image segmentation via implicit template deformation. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, pages 82–89. Springer, 2013a. (Cited on pages xxi, 30, 110, 137, and 151.)

R. Prevost, R. Cuingnet, B. Mory, J.-M. Correas, L. D. Cohen, and R. Ardon. Joint co-segmentation and registration of 3d ultrasound images. In *Information Processing in Medical Imaging (IPMI)*, volume 7917, pages 268–279. Springer, 2013b. (Cited on pages xx and 138.)

B. L. Price, B. Morse, and S. Cohen. Geodesic graph cut for interactive image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3161–3168, 2010. (Cited on page 115.)

M. Prinssen, E. Verhoeven, H. Verhagen, and J. Blankensteijn. Decision-making in follow-up after endovascular aneurysm repair based on diameter and volume measurements: A blinded comparison. *European Journal of Vascular and Endovascular Surgery*, 26(2):184–187, 2003. (Cited on page 153.)

J. R. Quinlan. *C4.5: programs for machine learning*, volume 1. Morgan kaufmann, 1993. (Cited on page 12.)

C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3): 309–314, 2004. (Cited on page 115.)

L. Rouet, R. Ardon, J.-M. Rouet, B. Mory, C. Dufour, and A. Long. Semi-automatic abdominal aortic aneurysms geometry assessment based on 3D ultrasound. In *IEEE Ultrasonics Symposium (IUS)*, pages 201–204. IEEE, 2010. (Cited on page 153.)

M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision (ECCV)*, volume 2351 of *LNCS*, pages 78–92. Springer Berlin Heidelberg, 2002. (Cited on pages 28 and 31.)

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999. (Cited on pages 29 and 74.)

S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Third International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001. (Cited on page 165.)

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and others. Imagenet large scale visual recognition challenge. *arXiv preprint arXiv:1409.0575*, 2014. (Cited on page 12.)

K. A. Saddi, M. Rousson, and F. Cheriet. Global-to-local shape matching for liver segmentation in CT imaging. In *Proceedings of MICCAI Workshop 3-D Segmentation Clinic: A Grand Challenge*, pages 207–214, 2007a. (Cited on page 29.)

K. A. Saddi, C. Chefd'Hotel, M. Rousson, and F. Cheriet. Region-based segmentation via non-rigid template matching. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–7, 2007b. (Cited on page 76.)

M. Sakashita, T. Kitasaka, K. Mori, Y. Suenaga, and S. Nawana. A method for extracting multi-organ from four-phase contrasted CT images based on CT value distribution estimation using EM-algorithm. In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Medical Imaging*, volume 6514, pages 1–12, 2007. (Cited on page 23.)

C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A level set model for image classification. *International Journal of Computer Vision*, 40(3):187–197, 2000. (Cited on page 31.)

S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging*, 16(1):41–54, 1997. (Cited on page 23.)

S. Seifert, A. Barbu, S. K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, and D. Comaniciu. Hierarchical parsing and semantic navigation of full body CT data. *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Medical Imaging*, 7259:725902, 2009. (Cited on page 20.)

S. Seifert, M. Kelm, M. Moeller, S. Mukherjee, A. Cavallaro, M. Huber, and D. Comaniciu. Semantic annotation of medical images. In B. J. Liu and W. W. Boonn, editors, *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Medical Imaging*, pages 762808–762808–8, 2010. (Cited on pages 21 and 32.)

T. Sharp. Implementing decision trees and forests on a GPU. In D. Forsyth, P. Torr, and A. Zisserman, editors, *European Conference on Computer Vision (ECCV)*, volume 5305 of *LNCS*, pages 595–608. Springer Berlin Heidelberg, 2008. (Cited on page 15.)

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. (Cited on pages 26 and 115.)

J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. (Cited on page 16.)

J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1): 2–23, 2009. (Cited on pages 39 and 49.)

S. Soatto and A. J. Yezzi. DEFORMOTION deforming motion, shape average and the joint registration and segmentation of images. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision (ECCV)*, pages 32–47. Springer, 2002. (Cited on pages 29 and 73.)

Society of Interventional Radiology (SIR). Abdominal aortic aneurysm diagnosis and treatment options - SIR. http://www.sirweb.org/patients/abdominal-aortic-aneurysms/, 2015. Accessed: 2015-03-19. (Cited on page 154.)

M. Sofka, J. Zhang, S. K. Zhou, and D. Comaniciu. Multiple object detection by sequential Monte Carlo and hierarchical detection network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1735–1742, 2010. (Cited on page 21.)

L. Soler, H. Delingette, G. Malandain, J. Montagnat, N. Ayache, C. Koehl, O. Dourthe, B. Malassagne, M. Smith, D. Mutter, and J. Marescaux. Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery. *Official journal of the International Society for Computer Aided Surgery*, 6(3):131–142, 2001. (Cited on page 17.)

C. G. Solomon and K. C. Kent. Abdominal aortic aneurysms. *New England Journal of Medicine*, 371(22):2101–2108, 2014. (Cited on page 153.)

O. Somphone, B. Mory, S. Makram-Ebeid, and L. Cohen. Prior-based piecewise-smooth segmentation by template competitive deformation using partitions of unity. In *European Conference on Computer Vision (ECCV)*, pages 628–641, Berlin, Heidelberg, 2008. Springer-Verlag. (Cited on page 29.)

M. Sun, P. Kohli, and J. Shotton. Conditional regression forests for human pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3394–3401, 2012. (Cited on pages 16 and 38.)

J. Suri, Kecheng Liu, S. Singh, S. Laxminarayan, Xiaolan Zeng, and L. Reden. Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, 6(1):8–28, 2002. (Cited on page 28.)

M. Sussman, P. Smereka, and S. J. Osher. A level set approach for computing solutions to incompressible two-phase flow. *Journal of Computational Physics*, 114:146–159, 1994. (Cited on page 167.)

J. P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998. (Cited on page 74.)

A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154, 2003. (Cited on page 30.)

A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky. Mutual information in coupled multi-shape model for medical image segmentation. *Medical Image Analysis*, 8(4):429–445, 2004. (Cited on page 32.)

Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1589–1596, 2005. (Cited on pages 16 and 20.)

Z. Tu. Auto-context and its application to high-level vision tasks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. (Cited on page 16.)

Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1744–1757, 2010. (Cited on page 16.)

M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–591. IEEE, 1991. (Cited on page 21.)

M. G. Uzunbas, O. Soldea, D. Unay, M. Cetin, G. Unal, A. Ercil, and A. Ekin. Coupled nonparametric shape and moment-based intershape pose priors for multiple basal ganglia structure segmentation. *IEEE Transactions on Medical Imaging*, 29(12):1959–1978, 2010. (Cited on page 33.)

A. Vasilevskiy and K. Siddiqi. Flux maximizing geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1565–1578, 2002. (Cited on page 90.)

L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002. (Cited on page 31.)

T. Vik, D. Bystrov, N. Schadewaldt, H. Schulz, and J. Peters. A new method for robust organ positioning in CT images. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 338–341, 2012. (Cited on page 18.)

S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004. (Cited on page 24.)

M. Wild. Recent development of the iterative closest point (ICP) algorithm from 2002 to 2007. Technical report, Eidgenössische Technische Hochschule Zürich, 2010. (Cited on page 165.)

A. Wimmer, G. Soza, and J. Hornegger. A generic probabilistic active shape model for organ segmentation. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5762 of *LNCS*, pages 26–33. Springer Berlin Heidelberg, 2009. (Cited on pages 145 and 146.)

D. J. Withey and Z. J. Koles. Medical image segmentation: A review. *International Journal of Bioelectromagnetism*, 10(3):125–148, 2008. (Cited on page 8.)

K. Wittels. Aortic emergencies. *Emergency Medicine Clinics of North America*, 29 (4):789–800, 2011. (Cited on page 153.)

R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert. Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7510 of *LNCS*, pages 10–17. Springer, 2012. (Cited on page 24.)

World Health Organization. WHO | WHO proposes global agenda on transplantation. http://www.who.int/mediacentre/news/releases/2007/pr12/en/, 2007. Accessed: 2015-03-11. (Cited on page 139.)

D. Wu, M. Sofka, N. Birkbeck, and S. K. Zhou. Segmentation of multiple knee bones from CT for orthopedic knee surgery planning. In P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8673 of *LNCS*, pages 372–380. Springer International Publishing, 2014. (Cited on pages 21 and 32.)

X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008. (Cited on page 12.)

C. Xu and J. L. Prince. Gradient vector flow: A new external force for snakes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 66–71, 1997. (Cited on pages 28 and 90.)

P. Yan, W. Shen, A. Kassim, and M. Shah. Segmentation of neighboring organs in medical image with model competition. In J. S. Duncan and G. Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3749 of *LNCS*, pages 270–277. Springer Berlin Heidelberg, 2005. (Cited on pages 31 and 32.)

J. Yang, L. H. Staib, and J. S. Duncan. Neighbor-constrained segmentation with level set based 3-D deformable models. *IEEE Transactions on Medical Imaging*, 23(8):940–948, 2004. (Cited on pages 31 and 32.)

W. Yang, J. Cai, J. Zheng, and J. Luo. User-friendly interactive image segmentation through unified combinatorial user inputs. *IEEE Transactions on Image Processing*, 19(9):2470–2479, 2010. (Cited on page 115.)

A. Yezzi and S. Soatto. Stereoscopic segmentation. *International Journal of Computer Vision*, 53(1):31–43, 2003. (Cited on pages 29 and 73.)

D. Zakim and T. D. Boyer, editors. *Hepatology: a textbook of liver disease*. Saunders, Philadelphia, 4th ed edition, 2003. (Cited on page 139.)

Y. Zhan, X. Zhou, Z. Peng, and A. Krishnan. Active scheduling of organ detection and segmentation in whole-body medical images. In D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5241 of *LNCS*, pages 313–321. Springer Berlin Heidelberg, 2008. (Cited on page 20.)

J. Zhang, J. Zheng, and J. Cai. A diffusion approach to seeded image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2125–2132, 2010a. (Cited on page 115.)

X. Zhang, J. Tian, K. Deng, Y. Wu, and X. Li. Automatic liver segmentation using a statistical shape model with optimal surface detection. *IEEE Transactions on Biomedical Engineering*, 57(10):2622–2626, 2010b. (Cited on pages 17 and 140.)

H.-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1):179–195, 1996. (Cited on page 31.)

Y. Zhao, S.-C. Zhu, and S. Luo. Co3 for ultra-fast and accurate interactive segmentation. In *Proceedings of the international conference on Multimedia*, pages 93–102. ACM, 2010. (Cited on page 116.)

Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging*, 27(11):1668–1681, 2008. (Cited on page 32.)

Y. Zheng, B. Georgescu, H. Ling, S. K. Zhou, M. Scheuering, and D. Comaniciu. Constrained marginal space learning for efficient 3D anatomical structure detection in medical images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 194–201, 2009a. (Cited on pages 20 and 21.)

Y. Zheng, B. Georgescu, and D. Comaniciu. Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images. In *Information Processing in Medical Imaging (IPMI)*, pages 411–422, 2009b. (Cited on pages 20 and 21.)

K. S. Zhou. Discriminative anatomy detection: Classification vs regression. *Pattern Recognition Letters*, 43:25–38, 2014. (Cited on page 20.)

S. K. Zhou. Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. *Medical Image Analysis*, 14(4):563–581, 2010. (Cited on page 39.)

S. K. Zhou and D. Comaniciu. Shape regression machine. In N. Karssemeijer and B. P. F. Lelieveldt, editors, *Information Processing in Medical Imaging (IPMI)*, volume 4584 of *LNCS*, pages 13–25. Springer-Verlag Berlin Heidelberg, 2007. (Cited on pages xx, 39, 66, and 137.)

X. S. Zhou, Z. Peng, Y. Zhan, M. Dewan, B. Jian, A. Krishnan, Y. Tao, M. Harder, S. Grosskopf, and U. Feuerlein. Redundancy, redundancy, redundancy: the three keys to highly robust anatomical parsing in medical images. In *Proceedings of the international conference on Multimedia information retrieval*, pages 175–184. ACM, 2010. (Cited on page 55.)

Y. Zhou and J. Bai. Atlas based automatic identification of abdominal organs. In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Medical Imaging*, volume 5747, pages 1804–1812, 2005. (Cited on page 18.)

D. Zikic, B. Glocker, and A. Criminisi. Atlas encoding by randomized forests for efficient label propagation. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8151 of *LNCS*, pages 66–73. Springer Berlin Heidelberg, 2013. (Cited on pages 16 and 24.)

Méthodes multi-organes rapides avec a priori de forme pour la localisation et la segmentation en imagerie médicale 3D

Romane GAURIAU

RESUME : Avec l'utilisation de plus en plus répandue de l'imagerie dans la pratique médicale (diagnostic, suivi, planification d'intervention, etc.), le développement d'algorithmes d'analyse d'images est devenu primordial. Ces algorithmes permettent aux cliniciens d'analyser et d'interpréter plus facilement et plus rapidement des données de plus en plus complexes. Dans ce contexte, la localisation et la segmentation de structures anatomiques sont devenues des composants critiques dans les processus cliniques modernes. La conception de tels outils pour répondre aux exigences de robustesse, précision et rapidité demeure cependant un réel défi technique.

Ce travail propose une méthode complète pour la segmentation de plusieurs organes dans des images médicales. Cette méthode, générique et pouvant être appliquée à un nombre varié de structures et dans différentes modalités d'imagerie, est constituée de trois composants : (i) un algorithme de localisation automatique, (ii) un algorithme de segmentation, (iii) un outil de correction interactive. Ces différentes parties peuvent s'enchaîner aisément pour former un outil complet et cohérent, mais peuvent aussi bien être utilisées indépendemment.

L'accent a été mis sur des méthodes robustes et efficaces afin de répondre aux exigences cliniques. Notre méthode de localisation s'appuie sur une cascade de régression par forêts aléatoires (Cuingnet et al., 2012). Elle introduit l'utilisation d'informations a priori de forme, spécifiques à chaque organe (grâce à des atlas probabilistes) pour des résultats plus cohérents avec la réalité anatomique. Notre méthode de segmentation étend la méthode de segmentation par modèle implicite (Mory et al., 2012) à plusieurs modèles. La formulation proposée permet d'obtenir des déformations cohérentes, notamment en introduisant des contraintes de non recouvrement entre les modèles déformés. En s'appuyant sur des forces images polyvalentes, l'approche proposée se montre robuste et performante pour la segmentation de multiples structures. Toute méthode automatique n'est cependant jamais parfaite. Afin que le clinicien garde la main sur le résultat final, nous proposons d'enrichir la formulation précédente avec des contraintes fournies par l'utilisateur. Une optimisation localisée permet d'obtenir un outil facile à utiliser et au comportement intuitif. Ce dernier composant est crucial pour que notre outil soit réellement utilisable en pratique.

Chacun de ces trois composants a été évalué sur plusieurs grandes bases de données cliniques (en tomodensitométrie, imagerie par résonance magnétique et ultrasons). Une étude avec des utilisateurs nous a aussi permis de recueillir des retours positifs de plusieurs experts en imagerie médicale. Les différents résultats présentés dans ce manuscrit montrent l'intérêt de notre méthode et son potentiel pour une utilisation clinique.

MOTS-CLEFS : segmentation, imagerie médicale, 3D, localisation, déformation de modèles, forêts aléatoires

ABSTRACT : With the ubiquity of imaging in medical applications (diagnostic, treatment follow-up, surgery planning. . .), image processing algorithms have become of primary importance. Algorithms help clinicians extract critical information more quickly and more reliably from increasingly large and complex acquisitions. In this context, anatomy localization and segmentation is a crucial component in modern clinical workflows. Due to particularly high requirements in terms of robustness, accuracy and speed, designing such tools remains a challenging task.

In this work, we propose a complete pipeline for the segmentation of multiple organs in medical images. The method is generic, it can be applied to varying numbers of organs, on different imaging modalities. Our approach consists of three components : (i) an automatic localization algorithm, (ii) an automatic segmentation algorithm, (iii) a framework for interactive corrections. We present these components as a coherent processing chain, although each block could easily be used independently of the others.

To fulfill clinical requirements, we focus on robust and efficient solutions. Our anatomy localization method is based on a cascade of Random Regression Forests (Cuingnet et al., 2012). One key originality of our work is the use of shape priors for each organ (thanks to probabilistic atlases). Combined with the evaluation of the trained regression forests, they result in shape-consistent confidence maps for each organ instead of simple bounding boxes. Our segmentation method extends the implicit template deformation framework of (Mory et al., 2012) to multiple organs. The proposed formulation builds on the versatility of the original approach and introduces new non-overlapping constraints and contrast-invariant forces. This makes our approach a fully automatic, robust and efficient method for the coherent segmentation of multiple structures. In the case of imperfect segmentation results, it is crucial to enable clinicians to correct them easily. We show that our automatic segmentation framework can be extended with simple user-driven constraints to allow for intuitive interactive corrections. We believe that this final component is key towards the applicability of our pipeline in actual clinical routine.

Each of our algorithmic components has been evaluated on large clinical databases. We illustrate their use on CT, MRI and US data and present a user study gathering the feedback of medical imaging experts. The results demonstrate the interest in our method and its potential for clinical use.

KEY-WORDS : segmentation, 3D, localization, medical imaging, deformable model, random forest, machine learning, anatomical intelligence