Spatio-temporal segmentation with probabilistic sparse models, contributions and applications

Charles-Henri Florin

Thèse

Présentée devant l'Ecole Nationale des Ponts et Chaussées
en vue de l'obtention du grade de Docteur
par
Charles Henri Florin

Spatio-temporal segmentation with probabilistic sparse models, contributions and applications

soutenue le 4 Mai 2007

Président du jury	**Alain Rahmouni**	**CHU Henri Mondor - Paris XII**
Rapporteurs	**Allen Tannenbaum**	**Georgia Tech**
	Jean-Philippe Thiran	**EPFL**
Examinateurs	**Nicholas Ayache**	**INRIA**
	Gilles Fleury	**Supélec**
	Renaud Keriven	**CERTIS, ENPC**
Membre invité	**Gareth Funka-Lea**	**Siemens Corp. Research**
Directeurs de thèse	**Nikos Paragios**	**Ecole Centrale de Paris**
	James Williams	**Siemens Medical Solutions**

Ecole Nationale des Ponts et Chaussées,
77455 Marne la Vallee, France,
http://www.enpc.fr/

To Tom,

Preface

This PhD was co-directed by Nikos Paragios from Ecole Nationale des Ponts et Chaussées, then Ecole Centrale de Paris, and James Williams from Siemens Corporate Research. I would like to warmly thank both my advisers for their time and guidance on both sides of the ocean. Their academic and industrial experience have brought me a wide perspective in the domain of Computer Vision. Nikos Paragios has successfully guided my research remotely from France and James Williams has always offered me the necessary time to share his experience with me.

Renaud Kériven, director of CERTIS lab at Ecole Nationale des Ponts et Chaussées welcomed me at the very beginning of the PhD, where I had the opportunity to meet with talented researchers and benefit from their experience; for that, I am thankful to the CERTIS lab and in particular to Renaud.

Romain Moreau-Gobard guided my first steps at Siemens Corporate Research and supported my first research projects. The exceptional conditions for research created by the people at SCR has had a great impact on my work. I would like to thank in particular Marie-Pierre Jolly for her support; she has brought me both efficient research organization skills and personal support. Mikaël Rousson and Christophe Chefd'hotel for their kind support and friendship, and all the people at SCR with whom I had fruitful and challenging scientific discussions.

Last but not least, I am specially grateful to my fiancée Francesca Lorusso-Caputi for her kind patience and support.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION DÉTAILLÉE

Contribution

Segmenter un objet dans une image numérique consiste à labéliser les pixels qui appartiennent à cet objet et ceux qui appartiennent au fond. La segmentation est une étape préliminaire à l'analyse du contenu de l'image, et est nécessaire dans nombre d'applications telles que le diagnostique en imagerie médicale, ou la reconnaissance d'objet. Le nombre d'images à interpréter ayant augmenté de façon exponentielle ces vingt dernières années, l'analyse automatique est devenue une nécessité à cause de contraintes de temps et d'argent. De plus, dans beaucoup de cas, une évaluation quantitative pour le diagnostique ou la surveillance n'est possible qu'après avoir délimité les contours de l'objet d'intérêt. Les récentes avancées dans ces domaines ont rendu possibles certaines applications telles que: la sécurité et la vidéosurveillance dans les lieux publics (aéroports, hôpitaux et stades sportifs), la reconnaissance automatique de visages pour l'identification ou la lutte contre la criminalité, le controle de trafic depuis l'analyse comportementale des clients dans les centres d'achat à la gestion de trafic routier, l'édition d'image et de vidéo du dessin assisté par ordinateur aux effets spéciaux de l'industrie des jeux et cinématographique, ou du diagnostique et planification de traitement en radiologie.

Les techniques de segmentation se basent sur l'information de l'image (aussi appelée support de l'image dans cette thèse) pour détecter soit les regions qui font partie de l'objet d'intérêt, soit les bords entre l'objet et le fond. Dans le contexte des images de scènes naturelles (e.g. les images médicales), la faible résolution des images et des données corrompues ou pathologiques rendent souvent nécessaires l'utilisation d'une connaissance à priori à propos de l'objet à segmenter. Cette connaissance à priori est utilisée pour construire un modèle et contre-balancer les régions de l'image qui ne supportent que faiblement la segmentation ou la conduisent vers une solution fausse. Ces vingt dernières années ont vu beaucoup de techniques se développer autour de ce problème, nottament les modèles de formes et d'apparence (intensité des pixels). Une analyse

de la distribution statistique de la forme et de l'apparence de l'objet réduit la complexité de la segmentation et guide la solution vers les formes les plus probables statistiquement.

Le travail présenté ici introduit comment des modèles statistiques peuvent être construits et exploités à partir d'une quantité limitée d'information. Cette quantité limitée d'information (aussi appelé information clairsemée) doit être sélectionnée avec soin afin qu'elle soit facilement identifiable dans l'image et permette une reconstruction efficace du reste de l'information. Puisque les images naturelles ont des régions qui supportent mieux la segmentation que d'autres, il est raisonable d'identifier ces régions dans l'ensemble d'apprentissage, et d'apprendre à interpoler le reste de la segmentation, au lieu d'essayer d'extraire le contenu de l'image en incluant des régions qui sont notoirement pauvres en information ou qui conduisent au mauvais résultat. Puisque ces modèles statistiques peuvent être statiques ou non, nous avons examiner les deux cas. Les techniques les plus utilisées pour le suivi d'objet dans des séquences d'images consistent à isoler et suivre ses caractéristiques, ou estimer le champ de déplacement appelé flot optique qui lie une image dans la séquence à l'image successive. L'approche proposée dans cette thèse est plus directe ; elle consiste à modéliser le déplacement de l'objet d'intérêt à partir de la connaissance à priori. Dans la plupart des cas, les caracteristiques de l'objet sont déjà connues (localisation, forme, apparence, ...), ainsi que son déplacement approximativement. En prédisant les caractéristiques de l'objet dans des temps futures, sa segmentation devient plus précise et l'on peut même segmenter toute la séquence d'images à la fois si elle est disponible. Ces modèles de suivi peuvent même être utilisés pour la segmentation classique avec une approche progressive. Dans ce cas, un temps fictif est introduit, et l'état de la solution à un moment donné est utilisé pour prédire le prochain état. Enfin, la segmentation progressive qui consiste à déterminer les caractéristiques les plus probables de l'objet à chaque pas de temps ne permet pas de résoudre certains problèmes les plus difficiles. Pour l'un d'entre eux, nous proposons une nouvelle méthode qui consiste à considérer les caractéristiques de l'objet comme une variable aléatoire et échantilloner l'espace de cette variable aléatoire. Ainsi, des caractéristiques faiblement probables à un temps donné sont conservées et peuvent donner lieu plus tard à la solution correcte.

En travaillant à Siemens Corporate Research, à Princeton, NJ , quatorze certificats d'invention et brevets ont été déposés, et les travaux relatifs à cette thèse ont été publiés dans des conférences renommées dans le domaine de la vision par ordinateur et de l'imagerie médicale:

Time-Varying Linear Autoregressive Models for Segmentation, Charles Florin, Nikos Paragios, Gareth Funka-Lea et James Williams, ICIP 2007, San Antonio, Texas, USA

Liver Segmentation Using Sparse 3D Prior Models with Optimal Data Support, Charles Florin, Nikos Paragios, Gareth Funka-Lea et James Williams, IPMI 2007, Kerkrade, Pays-Bas

Locally Adaptive Autoregressive Active Models for Segmentation of 3d Anatomical Structures, Charles Florin, Nikos Paragios, et James Williams, ISBI 2007, Arlington, VA, E-U

Globally Optimal Active Contours, Sequential Monte Carlo and On-line Learning for Vessel Segmentation, Charles Florin, Nikos Paragios, et James Williams, ECCV 2006, Graz, Autriche

Automatic Heart Isolation for CT Coronary Visualization using Graph-Cuts, Gareth Funka-Lea, Yuri Boykov, Charles Florin, Marie-Pierre Jolly, Romain Moreau-Gobard, Rama Ramaraj et Daniel Rinck, ISBI 2006, Arlington, VA, E-U

Registration of 3D angiographic and X-ray images using Sequential Monte Carlo sampling, Charles Florin, James Williams, Ali Khamene et Nikos Paragios, CVBIA 2005, pages 427-436, Beijing, Chine

Particle Filters, a Quasi-Monte Carlo Solution for Segmentation of Coronaries, Charles Florin, Nikos Paragios et James Williams, MICCAI 2005, pages 246-253, Palm Springs, CA, E-U

Automatic heart peripheral vessels segmentation based on a normal MIP ray casting technique, Charles Florin, Romain Moreau-Gobard et James Williams, MICCAI 2004, vol. 1, pages 483-490, Saint-Malo, France

Présentation Générale de la Thèse

Chapitre 1: Revue de l'Etat de l'Art en Segmentation et Suivi d'Objet

Le premier chapitre effectue une revue des techniques les plus connues de segmentation et suivi d'objet. Il est divisé en trois parties ; dans la première sont présentées les méthodes sans modèle ou à modèle de faible niveau. Les modèles à faible niveau ne dépendent que d'observations directes de l'image tels que l'intensité des pixels, les gradients, ... La seconde partie présente les techniques se basant sur des modèles qui représentent l'objet d'intérêt d'une manière plus avancée telles que des patrons, ou des distributions statistiques. Les techniques à modèle contraignent l'information de l'image avec la connaissance à priori afin de produire une solution au problème de segmentation qui

soit plus robuste aux informations manquantes ou corrompues dans l'image. La troisième partie présente les méthodes de suivi et de segmentation séquentielle. Globalement, il est intéressant de remarquer que les techniques actuelles ne hiérarchisent pas l'information selon le degré de support effectif qu'elle apporte à la segmentation. Pour la segmentation, cela signifie apprendre à partir de la connaissance à priori quelles sont les régions de l'image qui supportent au mieux la solution. Pour le suivi d'objet, cela signifie modéliser les changements de charactéristiques de l'objet au travers du temps. Finallement, il est constaté que les techniques actuelles sous-utilisent les modèles d'incertitude de telle sorte que le résultat de la segmentation séquentielle n'est que la succession des solutions locales les plus probables. Ces trois aspects sont examinés en détail dans la suite de la thèse.

Chapitre 2: Modèle à Information Clairsemée pour la Réduction de Dimension et de Complexité

Ce chapitre examine les modèles de segmentation à partir d'information claisemée. L'image est divisée en régions, et une mesure est associée à chaque région pour quantifier la façon dont elle supporte la segmentation. Ensuite, est introduit un modèle qui n'utilise que les régions qui supportent fortement la segmentation. Le reste de la solution est reconstruit à partir d'une interpolation dirigée par un modèle. Après une généralisation théorique, cette méthode est appliquée à trois problèmes differents d'intérêt majeur en vision par ordinateur. Le premier de ces problèmes est la segmentation d'objet dans une image volumétrique, dans lequel le modèle à information clairsemée est comparé à une technique bien connue de décomposition linéaire afin de mesurer les bénéfices de cette nouvelle approche en termes de réduction de dimension. Le second problème est le suivi et la segmentation du ventricule gauche en échocardiographie, et montre l'efficacité de cette approche dans les cas de haute variance statistique. Une analyse statistique est menée sur les résultats de cette méthode et comparée à la variabilité inter-experts. Finallement, le troisième problème est la reconstruction de surface à partir de distances pointées par laser, dans lequel la robustesse de la nouvelle méthode est analysée avec la précense de bruit sel et poivre.

Chapitre 3: Modèles de Déplacement pour des Dynamismes non-Stationaires par Autoregression

Le troisième chapitre introduit les modèles de déplacement appris à partir de connaissance à priori, et présente comment les adapter à la volée afin qu'à la fois la représentation des formes et celle des dynamismes soient mises à jour avec l'information la plus récente dans la séquence.

En conséquence, l'utilisation de modèles de déplacement n'est pas restreinte à des déplacements qui fassent partie d'un ensemble d'apprentissage, telles que toutes les scènes naturelles par exemple. Ce concept est présenté avec des modèles autoregressifs linéaires par souci de simplcité mathématique, et à cause du grand nombre de cas couverts par ces modèles. Cette technique est appliquée à l'étude de cas d'une silhouette d'un homme qui marche puis court. Des occlusions sont ajoutées numériquements afin d'étudier l'accumulation d'erreur. Cette méthode est également appliquée á des problèmes stationaires pour modéliser toute la séquence à la fois. Enfin, elle est étendue à une technique d'interpolation temporelle basée sur la théorie de restauration de signaux audio.

Chapitre 4: Modèles d'Incertitude pour la Segmentation Séquentielle

Le quatrième chapitre présente la dernière contribution de cette thèse. Le problème de segmentation de structures tubulaires dans des images volumétriques est reformulé comme un problème de suivi en introduisant un temps fictif. Les charactéristiques de ce tube à un temps donné sont représentées par une distribution statistique afin que non seulement la solution la plus probable soit déterminée, mais aussi d'autres solutions hypothétiques de plus faible probabilité. Dans le cas d'images de pathologies ou avec des artefacts d'acquisition, il est prouvé que garder uniquement les solutions les plus probables ne méne pas à la solution globale correcte, alors que des solutions localement imparfaites peuvent mener à la bonne solution globale. Cette technique est implémentée en effectuant un échantillonage séquentiel aléatoire, appelé filtrage particulaire. Puis, afin de comparer le filtrage particulaire avec des techniques de descente du gradient couramment utilisées, un problème statique et paramétrique est examiné: le recalage 2D-3D.

EXTENDED INTRODUCTION

Contribution

The segmentation of an object from a digital image consists in labeling the pixels that belong to this object and those that belong to the background, or alternatively in defining a boundary that encloses the object. Segmentation is a fundamental step to image content analysis whose outcome has numerous potential applications such as diagnosis in medical images or object recognition. The number of images to process in these fields have grown exponentially these past two decades, and have made computer-aided image analysis a major field of study because of time and financial constrains. Furthermore, in most cases, quantitative assessment for diagnosis requires the delineation of the object of interest. Recent advances in this field have allowed new applications to appear such as: security and surveillance in crowded areas (airports, hospitals and stadiums security), automatic face recognition for identification and anti-terrorism, traffic control from customer behavior analysis in shopping malls to road traffic management, image and video editing from computer assisted design to movies special effects, cinematography, or diagnosis and therapy planning from radiology.

Segmentation techniques rely on image information (also called image support in this thesis) to detect either regions that are part of the object of interest, or edges between the object and the background. In the context of images that depict natural scenes (e.g. medical images), low resolution images and corrupted or pathological data often require prior knowledge about the object to segment. Prior knowledge is used to build a model and counter-balance image regions where insufficient support could mislead the segmentation. In the past two decades, many techniques have been developed to tackle this problem, which includes model priors based on shape or appearance (pixel intensities). A statistical distribution analysis of the object's shape and appearance is used to reduce the segmentation problem complexity and guide the solution toward likely shapes.

The present work investigates how statistical models are built and exploited from a limited amount of information. This limited amount of information (also referred to as sparse information) has to be carefully chosen so that it is easily identifiable in the input image and could efficiently be used to reconstruct the remaining information. Since natural images have regions that support the segmentation task in a stronger way than others, it makes sense to identify these regions from prior knowledge, and learn how to express the remaining information used to represent the object instead of trying to extract content from image regions that are known to poorly support or mislead the segmentation. These statistical models may be either static or time-variant in cases where the object geometry changes according to some temporal process; we have investigated both. In image sequences, common approaches to object tracking consists in isolating and tracking features or estimating the velocity field called optical flow that wraps an image to the successive image. A more direct approach is proposed in this thesis; it consists in modeling the object of interest's motion using prior knowledge. In most cases, one already knows the characteristics (e.g. location, shape, appearance, ...) of the object to track and roughly how it moves. By predicting the object's characteristics in future frames, one may segment it more accurately and may even segment the whole sequence of images at once if it is available. Models for tracking in images sequences may even be applied to regular segmentation by progressively solving this problem. A virtual timeline is introduced and the solution's state at one time is used to predict the next state. Finally, progressive segmentation that consists in determining the most probable object features at each time step does not allow the solving of some of the most difficult problems. Such cases refer to important deformations of the object that cannot not be represented by simple linear models. In such cases a more complex probabilistic interpretation of the object dynamics is required. To this end,, we propose a novel approach that consists in representing the object's features as a random variable and sampling the feature space. Thus, features that have a low probability to be exact at a given time step are kept and may correspond to the correct solution at a later time step.

While working for Siemens Corporate Research, Princeton NJ, fourteen invention disclosures and patents relevant to the present thesis were filed. Also, relevant publications were presented in renowned computer vision and medical images conferences:

Time-Varying Linear Autoregressive Models for Segmentation, Charles Florin, Nikos Paragios, Gareth Funka-Lea et James Williams, ICIP 2007, San Antonio, Texas, USA

Liver Segmentation Using Sparse 3D Prior Models with Optimal Data Support, Charles Florin, Nikos Paragios, Gareth Funka-Lea et James Williams, IPMI 2007, Kerkrade, Pays-Bas

Locally Adaptive Autoregressive Active Models for Segmentation of 3d Anatomical Structures, Charles Florin, Nikos Paragios, and James Williams, ISBI 2007, Arlington, VA, USA

Globally Optimal Active Contours, Sequential Monte Carlo and On-line Learning for Vessel Segmentation, Charles Florin, James Williams, and Nikos Paragios, ECCV 2006, Graz, Austria

Automatic Heart Isolation for CT Coronary Visualization using Graph-Cuts, Gareth Funka-Lea, Yuri Boykov, Charles Florin, Marie-Pierre Jolly, Romain Moreau-Gobard, Rama Ramaraj and Daniel Rinck, ISBI 2006, Arlington, VA, USA

Registration of 3D angiographic and X-ray images using Sequential Monte Carlo sampling, Charles Florin, James Williams, Ali Khamene and Nikos Paragios, CVBIA 2005, pages 427-436, Beijing, China

Particle Filters, a Quasi-Monte Carlo Solution for Segmentation of Coronaries, Charles Florin, Nikos Paragios and James Williams, MICCAI 2005, pages 246-253, Palm Springs, CA, USA

Automatic heart peripheral vessels segmentation based on a normal MIP ray casting technique, Charles Florin, Romain Moreau-Gobard and James Williams, MICCAI 2004, vol. 1, pages 483-490, Saint-Malo, France

Thesis Overview

Chapter 1: Review of Prior Art in Segmentation and Tracking

The first chapter reviews the most common techniques for segmentation and tracking. It is divided into three parts; in the first part, model-free and low level models are presented. Low-level models only rely on direct observations from the image such as pixel intensities, image gradients, ... The second part introduces model-based techniques that represent the object of interest in a more advanced fashion using templates, or statistical distributions. Model-based techniques integrate the image information with prior knowledge in order to produce a solution to the segmentation problem that is more robust to missing or corrupt image data. The third part is devoted to tracking techniques and sequential segmentation. Overall, it is noted that current techniques do not hierarchize input information with respect to the amount of information they effectively provide for solving the segmentation problem. For segmentation, this means learning from prior knowledge

which image regions best support the solution. For tracking, this means modeling the changes of the object's characteristics across time. Finally, current segmentation techniques underuse uncertainty models so that the solution of a sequential segmentation is the succession of the most likely individual solutions. These three aspects are investigated in the three following chapters.

Chapter 2: Sparse Information Models for Dimensionality and Complexity Reduction

This chapter examines segmentation models based on sparse information. The input image or shape is divided into regions, and a measure is associated to each region to quantify how it supports the segmentation. Then, a model that uses only the best supported regions is introduced. The solution's remaining is reconstructed using a model-driven interpolation. After a theoretical introduction, this method is applied to three different problems of significant interest in computer vision. The first problem is organ segmentation in medical volumetric images, in which the sparse information model is compared to a common linear decomposition technique to test the benefits of the novel approach in terms of dimensionality reduction. The second problem is tracking and segmentation of the left ventricle in ultrasonic sequences, and proves the effectiveness of this approach in case of high statistical variance. A statistical analysis is conducted to compare the sparse information models with inter-expert variability. Finally, the third problem is surface reconstruction from laser pointed distances, in which the robustness of the novel method is analyzed in the context of salt and pepper input noise.

Chapter 3: Motion Models for Non-Stationary Dynamics via Autoregressive Models

The third chapter introduces motion models learned from prior knowledge and presents how to adapt them on-the-fly so that both the representation of the shapes and the dynamics are updated with the newest information available in the sequence. Consequently, the use of motion models is not constrained to behaviors that are part of the learning set, as merely all natural scenes. This concept is presented with linear autoregressive models for the simplicity of the mathematics and the high number of cases covered by such models. This technique is applied to the study case of a walking silhouette that starts to run. Digital occlusions are added to test the accumulation of errors. This technique is also applied to stationary problems to model the motion in all the sequence at once. It is then extended to a temporal interpolation technique based on audio signal restoration theory.

Chapter 4: Uncertainty Models for Progressive Segmentation

The fourth chapter presents the last contribution of this thesis. In the context of tubular structures segmentation, the multi-hypothesis framework is introduced and implemented with a sequential random sampling procedure called particle filters. Segmentation of tubular structures in volumetric images is reformulated as a geometric tracking problem using a virtual timeline. The characteristics of the tube at one point in the timeline are represented by a statistical distribution so that not only the most likely solution is preserved but also hypothetical solutions of lower probabilities. In the case of pathological data or images with acquisition artifacts, one can claim that the approach consisting in keeping only the most probable solution fails, while some of the less probable solutions may give rise to the correct global solution. This technique is implemented using a sequential Monte-Carlo sampling method called particle filters. The outcome of this method is used to segment highly non-linear structures such as the coronary arteries. Then, in order to compare this framework with commonly used gradient descent approaches, it is extended to a static parametric problem: 2D-3D registration.

Chapter 1

Learning to Segment and to Track

Abstract – *This chapter intends to review the scientific literature in segmentation and tracking with and without models and prior learning. First, techniques based solely on low level information (pixels intensity, image gradient, ...) are presented when the image is modeled by a continuous function and when it is represented in a discreet way. Second, model-based methods are introduced starting by rigid models of few parameters, with none or little adaption, to the soft models that represent the object of interest using probability distributions. Then, these soft models are integrated into low level techniques. Finally, in the context of image sequences, common frameworks combining stochastic modelling and time-series are introduced.*

1.1 Introduction

Object segmentation is equivalent to labelling the pixels in the image that belong to the object and those that belong to the background. Segmentation is the prime step of content extraction and content analysis; therefore a tremendous effort is put on developing algorithms that best segment with respect to human experts, as quickly as possible.

One may distinguish different ways to segment an image. The first is to analyze the pixels intensity value directly and to group the pixels according to intensity similarity, texture or higher order statistical properties and/or detect strong intensity gradients. This method regroups the *model free* and *low level model* techniques. The second method is to use a *model* (e.g. appearance or shape) and to detect the image regions that best match the model. The third is to transform the segmentation problem into a dynamic problem in which the solution (the labeling of the pixels belonging to the object of interest) is progressively achieved using *Bayesian processes*. When the input images form a time sequence, the object's segmentation across time is called *tracking*.

1.2 Model-Free and Low-Level Model Based Segmentation

1.2.1 Active Contours Model

Active contours and snakes

Active contours and snakes models were first introduced by Kass, Witkin and Terzopoulos in 1987 [104]. In many cases, an object distinguishes itself from the background by a boundary: a strong edge, a particular color, ... A general way to note that is to call g a function to minimize locally. Typically, if an object is defined by a strong edge, g is inversely proportional to the image gradient. In the following, the object boundary is denoted by a curve $\mathcal{C}$, parameterized by $p \in [0, 1]$. The curve $\mathcal{C}$ that defines the object's boundary theoretically minimizes the following energy E in equation (1.1) that combines a smoothness term and an image term:

$$E(\mathcal{C}) = \int_0^1 \underbrace{\alpha \left\| \frac{\partial \mathcal{C}}{\partial p} \right\|^2 + \beta \left\| \frac{\partial^2 \mathcal{C}}{\partial p^2} \right\|^2}_{\text{smoothing term}} + \underbrace{\lambda g(\mathcal{C}(p))}_{\text{image term}} dp. \tag{1.1}$$

The smoothness term stands for regularity/smoothness along the curve and has two components (resisting to stretching and bending). The image term g guides the active contour toward the desired image properties (strong gradients). Other terms may be added to take into account user-defined points or other external constraints such as prior knowledge [83][204][192]. External constraints were introduced because the need for models was felt very early. The basic approach for active contours is a local optimization, which makes the method quite sensitive to local minima, occlusions and missing/corrupted data.

The general approach for active contours is to consider a discrete variant of the curve using a number of control points, and evolve that curve $\mathcal{C}$ by minimizing the energy in equation (1.1) at each time step until convergence. Due to the fact that this method was a significant breakthrough in image segmentation, a large part of the following work in the domain is mainly influenced by the idea of designing ever-better energies, to minimize them by more and more efficient numerical schemes to get the globally optimum segmentation result, with respect to that particular energy and that particular scheme. However, this expression may lead to poor results and the snake may shrink to a point in certain regions of the image. In [36] et al. introduced a balloon force ν[1] that avoids the shrinking when no edge information is present.

The main advantages of Active Contour models are their low complexity, the fact that it is easy to introduce prior knowledge, and that they could account for open as well as closed structures. It is also a well established technique, and numerous publications are available. Finally, user interactivity is naturally handled by introducing an extra energy term in the equation. However, they suffer from being myopic (only local image information is taken into account), cannot handle topological changes and the solution to the segmentation problem depends on the selected parameterization.

Geodesic Active Contours

Caselles *et al.* [31] and Kichenassamy *et al.* [105][106] reformulated the active contour equation in a more rigorous way starting back from the snakes equation (equation (1.1)) and omitting certain terms:

$$E(\mathcal{C}) = \alpha \underbrace{\int_0^1 \left\| \frac{\partial \mathcal{C}(p)}{\partial p} \right\|^2 dp}_{\text{intrinsic energy}} + \lambda \underbrace{\int_0^1 g\left(\left|\nabla I\left(\mathcal{C}(p)\right)\right|\right) dp}_{\text{extrinsic energy}}. \tag{1.2}$$

[1] ν is generally a small positive constant designed to inflate the curve like a balloon to prevent it from shrinking when no edge information is present.

By reparameterizing the curve using the Euclidean arc-length ds, $\left\| \frac{\partial \mathcal{C}(p)}{\partial p} \right\| dp = ds$, and using the Malpertuis' Principle, Caselles *et al.* proved in [31] that, when $\alpha = 0$, to find the curve $\mathcal{C}$ that minimizes equation (1.2) is equivalent to find the curve of minimal length L_g in a Riemannian space that depends on the function g:

$$L_g = \int_0^{\mathcal{L}(\mathcal{C})} g\left(\left|\nabla I\left(\mathcal{C}(s)\right)\right|\right) ds, \tag{1.3}$$

where $\mathcal{L}\left(\mathcal{C}\right)$ is the Euclidean length of the curve $\mathcal{C}$. The curve of minimum Riemannian length is determined by taking the first derivative of equation (1.3), and using an iterative method with time step ∂t [30][35][125][107]:

$$\frac{\partial \mathcal{C}}{\partial t} = \underbrace{g\left(\left|\nabla I\left(\mathcal{C}\right)\right|\right) \mathcal{K}\mathcal{N}}_{\text{boundary force}} - \underbrace{\left(\nabla g\left(\left|\nabla I\left(\mathcal{C}\right)\right|\right) . \mathcal{N}\right) \mathcal{N}}_{\text{refinement force}}. \tag{1.4}$$

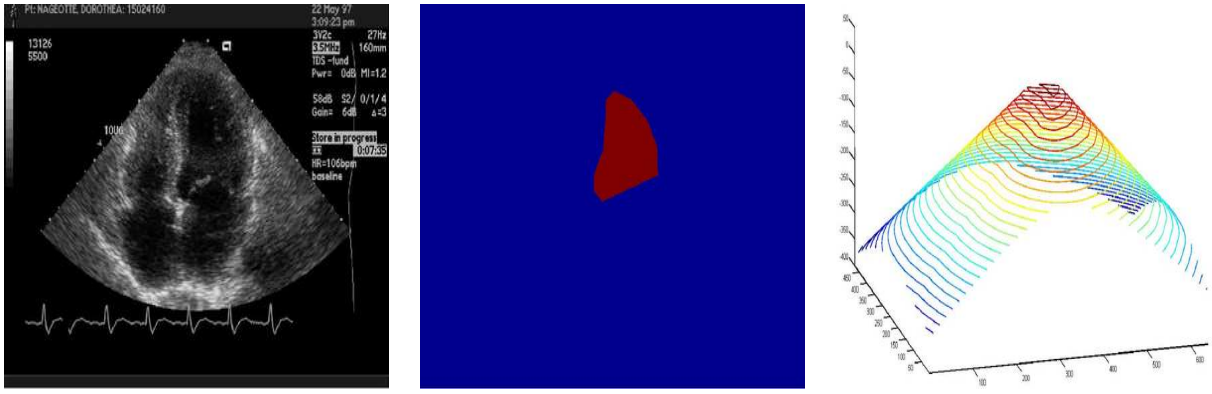
The main drawbacks of active contours are their sensitivity to initialization (the curve always converges to the nearest local minimum), to parameterization and to the chosen topology. If a topology change occurs (e.g. the curve has been initialized as one single piece, but the actual object to segment is composed of two disjoint pieces), a complex heuristic has to break the curve into multiple pieces and reparameterize each piece independently. To overcome this drawback, *level-sets* were introduced by Dervieux and Thomasset [58][57], before been rediscovered independently by Osher and Sethian [137] for flame propagation, then applied to image segmentation by Malladi *et al.* in 1995 [126]. However, this expression may lead to poor results and the curve may shrink to a point in certain regions of the image. In order to cope with that, a force similar to the balloon [36] was also introduced for the case of geometric active contours.

$$\frac{\partial \mathcal{C}}{\partial t} = g\left(\left|\nabla I\left(\mathcal{C}\right)\right|\right)\left(\mathcal{K} + \nu\right)\mathcal{N} - \left(\nabla g\left(\left|\nabla I\left(\mathcal{C}\right)\right|\right) . \mathcal{N}\right)\mathcal{N}. \tag{1.5}$$

Implicit Active Contours: Level Sets

The basic idea of level sets is to represent a given contour $\mathcal{C}$ by the zero level-set of a function[2] ψ (see figure (1.1)) , as written in equation (1.6) where Γ (resp. $\bar{\Gamma}$) is the image region inside (resp.

[2] the Euclidean distance function is used in most of the cases due to certain desired geometric properties



(a) (b) (c)

Fig. 1.1: Distance function that represents a shape (blue curve), positive inside the shape and negative outside. (a) Original image. (b) Binary shape. (c) Distance function isolines.

outside) the object, and $\mathcal{D}(\mathbf{x})$ is the Euclidean distance between point $\mathbf{p}$ and the curve $\mathcal{C}$:

$$\psi(\mathbf{p}) = \begin{cases} 0 & , \mathbf{p} \in \mathcal{C} \\ +\mathcal{D}(\mathbf{p}) > 0 & , \mathbf{p} \in \Gamma \\ -\mathcal{D}(\mathbf{p}) < 0 & , \mathbf{p} \in \bar{\Gamma} \end{cases} \tag{1.6}$$

One can embed in a straightforward fashion [137] the flow described in equation (1.4) in the level-set framework as described in equation (1.6), toward obtaining the following partial derivative equation that guides the propagation of ψ:

$$\frac{\partial \psi}{\partial t} = g\left(|\nabla I\left(\mathcal{C}\right)|\right) \mathcal{K} \left|\nabla \psi\right| + \nabla g\left(|\nabla I\left(\mathcal{C}\right)|\right) . \nabla \psi \tag{1.7}$$

The popularity of level-sets method is largely due to the simplicity with which one may implement them. They do not require any specific parameterization, and implicitly handle topological changes. By a snow-ball effect, an abundant literature [167][141] reviews their mathematical convergence and stability as well as their solution existence and uniqueness, which leads to the level-sets been even more popular.

Region-based active contours started with the work of Zhu and Yuille in [203] where an energy formulation for segmentation of multiple regions was derived from a Bayesian/minimum descriptive length (i.d. posterior probability) expression. Then, Chan and Vese introduced in [33] a

multi-region scheme where each region was defined by the logical combination of several level-set functions' sign. This expression guarantees only $\log(n)$ level-set functions are necessary to represent n regions. This scheme was reformulated for the geodesic active contours in [144] under the name of geodesic active regions. In [198], Yezzi et al. reformulated the multi-object segmentation using a coupled evolution of any arbitrary number of geodesic active contours that is obtained from the definition of binary and trinary flows.

Geodesic active contours aim at representing the segmentation problem in an energetic fashion and optimizing the solution with partial derivative equations. However, the energy terms correspond to a low level model (on curvature, gradient structure, pixels intensity, ...) that, when established once and for all, may poorly fit the image and let the active contour converge to a local solution. To overcome that issue, Juan, Kerivan and Postelnicu [102] propose to add a simulated annealing procedure. Paragios [145] and Rousson [156] use statistical models of shape and appearance that result in softer constrains for the active contours. For more details about level-sets, variational methods and implementations one may refer to [136].

1.2.2 Markov Random Fields and Gibbs Distributions

The probabilistic nature of image acquisition, noise and motion often drive the segmentation toward a low-level statistical model. In this case, instead of representing the solution by a continuous curve and solving partial derivative equations to determine the solution of minimal energy, one describes the probability measure that associates the segmentation to a likelihood given the low-level model's parameters (often based on pixels intensity) and the input image. Different techniques exist to determine the most likely solution. One consists in associating an energy function to the probability function and solving partial derivative equations (see previous sections). Another consists in sampling the random variable space and estimate directly the probability density function (see particle filters in section (1.4) for instance). A third solution consists in using *Gibbs distributions*, *Markov random fields* (MRF for short) equivalence, for segmentation.

The basic idea of MRF is to divide the image into homogeneous regions with respect to certain parameters [78][139][13][54]. To that end, it is assumed a random variable $\mathbf{z}$ describes the observed (corrupted) image, a random variable $\mathbf{x}$ describes the label of the image regions (i.e. object, background, and so on) and a third random variable $\mathbf{y}$ is related to the (a priori unknown) characteristics of each region. The main idea is to determine the labeling $\mathbf{x}$ and characteristics $\mathbf{y}$ that best

fit the image $\mathbf{z}$, or in other words to maximize the *a posteriori* probability:

$$p(\mathbf{x}, \mathbf{y}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x}, \mathbf{y})\, p(\mathbf{x}, \mathbf{y})}{p(\mathbf{z})}. \tag{1.8}$$

The above equation (1.8) is referred to as the *Bayesian rule* equation, and relates the *posterior* probability $p(\mathbf{x}, \mathbf{y}|\mathbf{z})$ to the *conditional* $p(\mathbf{z}|\mathbf{x}, \mathbf{y})$ and *prior* $p(\mathbf{z})$ probabilities.

MRF uses stochastic relaxation and annealing to generate a sequence of random variables that converges toward the maximum a posteriori[3]. An energy is defined by the negative logarithm of the posterior probability function
$E = -\log(p(\mathbf{z}|\mathbf{x}, \mathbf{y})) - \log(p(\mathbf{x}, \mathbf{y}))$; the parameters $\{\mathbf{x}, \mathbf{y}\}$ are locally adapted in the image to maximize this energy. The amplitude allowed for those changes are controlled by the annealing temperature parameter T that slowly decreases with respect to the number of iterations.

This technique has many advantages over hard constrained energies, especially in the presence of noisy images that do not fit any pixels intensity model from prior knowledge. However, simulated annealing avoids certain local minima, but MRF remains a local optimization and the optimum solution cannot be guaranteed in a finite amount of time. In certain problems (denoising for instance), a local solution is sufficient; in others, a global minimum is critical. That is the reason why more efficient numerical schemes that guarantee a global solution are investigated. One example is the max flow/min path principle which has emerged in computer vision using an efficient implementation like the graph-cuts.

Although not exclusively used for computer vision, other discrete optimization techniques are of particular interest. Among them, dynamic programing plays a significant role. Dynamic programing consists in dividing a general problem into a set of overlapping problems, iteratively solving each subproblem and recomposition the general solution. An important example of dynamic programing is provided by Dijkstra [63] for computing shortest paths on a directed graph with non-negative edge weights, such as an image.

Belief propagation networks are a particular instance of belief networks used to optimize one or several probability distribution functions (such as the one in equation (1.8)) using graph theory. A particular example is the expectation-maximization algorithm [53] that is used to iteratively estimate a model's parameters to maximize the likelihood of that model given observations.

[3] This method is often referred to as *Simulated Annealing*.

Last, linear programming [50] aims at solving linear minimization problems subject to constraints. A popular solution to the linear programming problems is given by the simplex algorithm [50]. Another algorithm of the same name, also called the downhill simplex algorithm is introduced by Nelder and Mead [130], and consists in iteratively estimating the solution for different values of the variables and modifying the worst set of variables according to the other possible estimated solutions.

Graph-Cuts

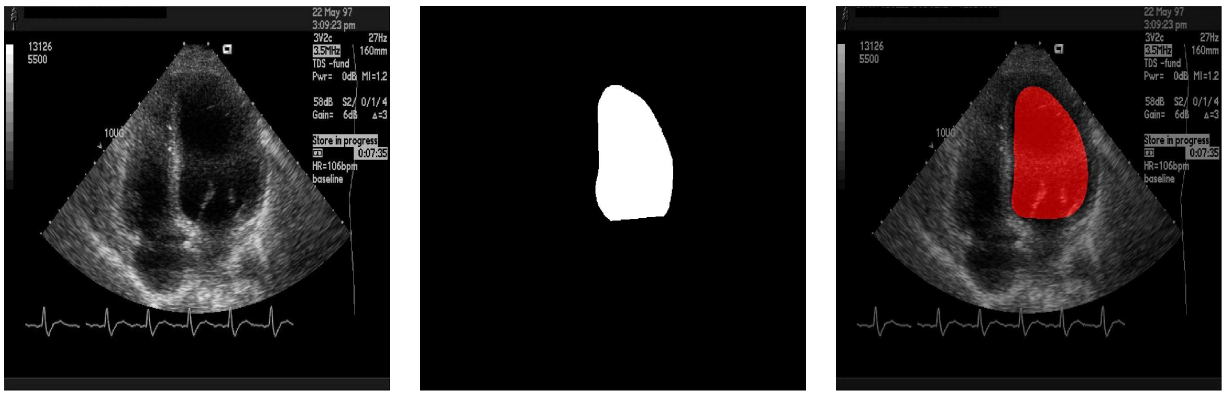
Graph-cuts is a very active field of study that is applied to different problems of computer vision such as segmentation [21][171][189][197][22], image restoration [23], texture synthesis [113] or stereo vision [159][109]. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be an undirected graph whose set of vertices is noted $\mathcal{V}$ and contains two special nodes called terminal, and whose set of edges is noted $\mathcal{E}$ and has non negative weights w_e, $\forall e \in \mathcal{E}$. A *cut*, noted C, is a subset of edges $\mathcal{C} \in \mathcal{E}$ that partitions the original set $\mathcal{E}$ in two such that each partition has one and only one terminal node. The cost of the cut $|\mathcal{C}|$ is the sum of the edges in $\mathcal{C}$:

$$|\mathcal{C}| = \sum_{c \in \mathcal{C}} w_c. \tag{1.9}$$

The graph-cut algorithm consists in finding the cut of minimal cost, and several implementations exist to determine such a cut in a polynomial time (e.g. max-flow and push-relabel [37]). It is of particular interest in the context of segmentation, where each image pixel represents a vertice in the graph, and each edge's weight is related to the relationship between two pixels intensity. For instance, if the edge's weight $w_{e_{ij}}$ between two vertices/image pixels e_i and e_j is defined as

$$w_{e_{ij}} = \frac{1}{1 + |I(e_i) - I(e_j)|}, \tag{1.10}$$

where $I(e_p)$ is the image intensity for the pixel e_p, the minimum cut includes vertices/image pixels that have a low weight $w_{e_{ij}}$, that is neighboring pixels that are very dissimilar. In [22], Boykov and Kolmogorov used graph-cuts to solve a geodesic problem and to include the types of energy that were previously solved using PDEs.



(a) (b) (c)

Fig. 1.2: (a) Original image (b) Template to match using the original image (c) Result of matching

1.2.3 Advantages and Drawbacks of Model Free Segmentation

With model free and low-level model techniques, one does not need to construct a model for the object of interest, which was a considerable advantage when data was not as widely available as it is now. In the particular context of organ segmentation in medical imaging, the inter patient variability of shape and appearance due to various image qualities and pathologies make certain modeling hazardous. The model free techniques are also often simpler to implement and may be quicker to compute. On the other hand, now that imaging data is commonly accessible (in 2005, about 60 millions CT and 1 million PET scans were performed in the US), robust models may be built to direct the segmentation.

1.3 Model-Based Segmentation

Template Matching

The simplest and most straightforward way to use a model for segmentation is *template matching*. Template matching consists in optimizing few parameters (position, rotation angle, scale factors) so that a template best matches the image (see figure (1.2)) according to a certain image measure whose global minimum corresponds to the best template alignment. If the only parameters to be optimized are position, rotation angle, scale factors, the template is called *rigid template*. This optimization is usually performed either using exhaustive search or gradient descent. For further

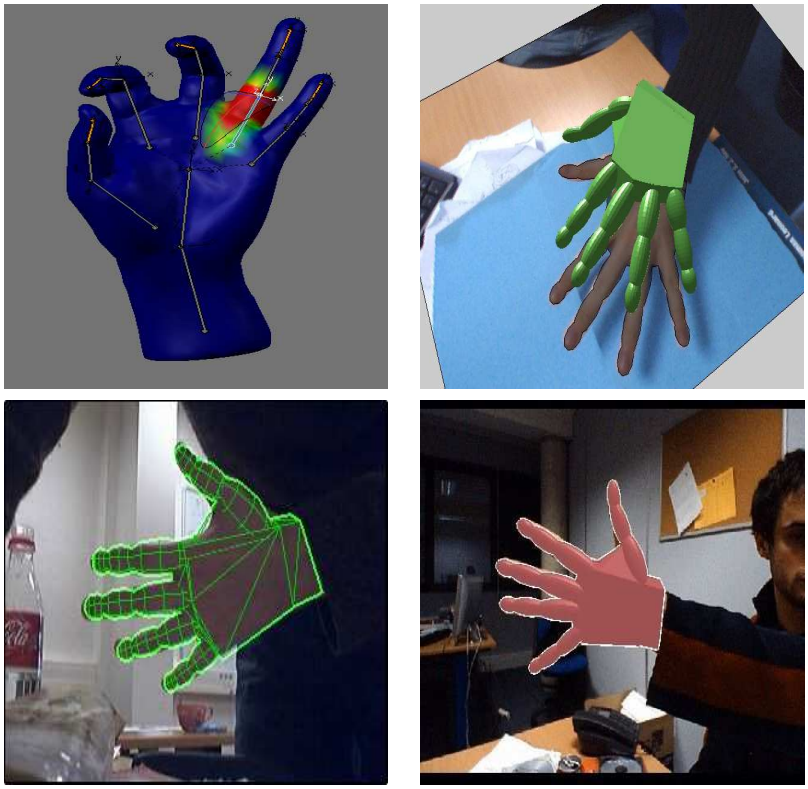


Fig. 1.3: An articulated model of the hand controlled by parameters to control angles and lengths. Courtesy of M. de la Gorce [52].

investigation, one may refer to [26] to study the effectiveness of exhaustive search and template matching in the particular context of watermarking and image security.

Unlike rigid templates, *deformable template* and *articulated templates* are models that can be adapted depending on parameters. An example is given in figure (1.3) where the stick-man shape is controlled by few parameters to adjust the members angle and length to the input image. Articulated models are of particular interest for surveillance and behavior analysis [94]. In the context of medical image analysis, deformable organisms [85] combine sensors and a framework to optimize (deform) a template (organism) with respect to the image.

In many applications, the templates are derived from physical data in a straightforward way. However, in most cases, the templates cannot be defined explicitly, but have to be learned from prior knowledge. A solution consists in establishing a low-dimension basis onto which the object to segment are projected.

1.3.1 Decomposition using Linear Operators

A second approach is to represent the object's model (shape and/or appearance) $\mathbf{Y}$ in a high dimensional space, then project the object's model onto a lower dimension feature space, and optimize the feature vector $\mathbf{X}$ with respect to the image measure:

$$\mathbf{X} = (g_1(\mathbf{Y}), g_2(\mathbf{Y}), ..., g_m(\mathbf{Y})). \tag{1.11}$$

In the following, different choices concerning the linear operators g_is are exposed.

Linear discriminant analysis (or LDA) consists in finding the projection space that best separates two classes/exemplars $p(\mathbf{X}|\mathbf{Y} = \mathbf{Y}_1)$ and $p(\mathbf{X}|\mathbf{Y} = \mathbf{Y}_2)$. LDA commonly relies on Fisher's linear discriminant [72] and assumes the probability density functions (pdfs) are Gaussian with means $\overline{\mathbf{Y}}_1$ and $\overline{\mathbf{Y}}_2$ and the same covariance $\mathbf{\Gamma}$. Fisher's discriminant S is the ratio of the two variances between and within the two classes $\mathbf{Y}_1$ and $\mathbf{Y}_2$ after been projected along the vector $\mathbf{w}$:

$$S = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{\left(\mathbf{w}.\overline{\mathbf{Y}}_1 - \mathbf{w}.\overline{\mathbf{Y}}_2\right)^2}{2\mathbf{w}^T.\mathbf{\Gamma}.\mathbf{w}} \tag{1.12}$$

It can be shown that the vector $\mathbf{w}$ that maximizes the Fischer ratio is

$$\mathbf{w} = \frac{1}{2}\mathbf{\Gamma}^{-1}.\left(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2\right). \tag{1.13}$$

LDA is commonly applied to classification and recognition problems [10]. In a first step, a training set of samples is gathered to estimate the classes distribution properties (mean and variance). Then equation (1.13) is applied to determine the optimal projection.

Recent developments in LDA include a probabilistic LDA [96] that generates a model to extract image features and automatically assigns a weight to each feature according to their discriminative power.

Principal component analysis (PCA) [38] [186] assumes a Gaussian distribution of the training samples. The main idea of PCA is to re-write a high dimensional vector $\mathbf{Y}$ as the sum of a mean vector $\overline{\mathbf{Y}}$ and a linear combination of the principal modes of variation (see figure (1.4). Eigenanalysis of the covariance matrix is used to determine the orthogonal basis formed by the matrix eigenvectors, and the eigenvalues associated to them. This orthogonal basis is composed by

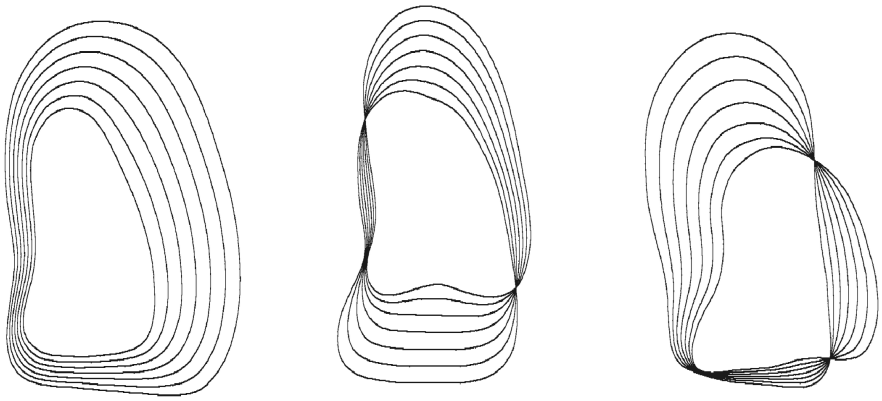


Fig. 1.4: Three principal modes of shape variations for the left ventricle in echocardiography.

the modes of variation, and the variations amplitude is given by the eigenvalues associated to each eigenvector/mode.

Let $\{\mathbf{Y}_i, i \in [|1, N|]\}$ be the set of high dimensional vectors, $\overline{\mathbf{Y}}$ be the mean of these vectors, and $\mathbf{\Gamma} = [(\mathbf{Y}_1 - \overline{\mathbf{Y}})(\mathbf{Y}_2 - \overline{\mathbf{Y}})...(\mathbf{Y}_N - \overline{\mathbf{Y}})][(\mathbf{Y}_1 - \overline{\mathbf{Y}})(\mathbf{Y}_2 - \overline{\mathbf{Y}})...(\mathbf{Y}_N - \overline{\mathbf{Y}})]^T$ be the covariance matrix. Given that $\mathbf{\Gamma}$ is a symmetric real positive definite matrix, its eigenvalues are noted $\lambda_1 > \lambda_2 > ... > \lambda_N > 0$ and the corresponding eigenvectors $\{\mathbf{U}_i, i \in [|1, N|]\}$. Noting $\mathbf{\Lambda}$ the diagonal matrix composed by the eigenvalues, and $\mathbf{U} = [\mathbf{U}_1\mathbf{U}_2...\mathbf{U}_N]$,

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \tag{1.14}$$

Any centered vector $(\mathbf{Y} - \overline{\mathbf{Y}})$ may be projected onto the eigenvectors orthogonal basis using the similarity matrix $\mathbf{U}$ such that

$$\mathbf{Y} - \overline{\mathbf{Y}} = \mathbf{U}\mathbf{X}, \text{ or noting } \mathbf{X} = [x_1...x_N]^T, \mathbf{Y} = \overline{\mathbf{Y}} + \sum_{q=1}^{N} x_q \mathbf{U}_q. \tag{1.15}$$

However, in practical cases, many eigenvectors relates to small eigenvalues (compared to the largest ones); that means these vectors relate to low amplitude variations. Therefore it makes sense to draw a threshold and keep only the eigenvectors whose linear combination approximates up to p percent of the total variation.

However, in some cases, the Gaussian distribution of the data does not stand. PCA extensions have been proposed to solve these cases, the most noticeable of which is *Kernel PCA*. Another decomposition technique that has been developed for non-Gaussian distributions is the *independent component analysis*.

Independent component analysis (ICA) [95] aims at decomposing the input signal into decorrelated non-Gaussian components. Given a signal $\mathbf{Y}$, ICA retrieves the underlying decorrelated non-Gaussian signals $\mathbf{S}_i$ (called *sources*) such that $\mathbf{Y} = \sum_{i=1}^{N} \mathbf{A}_i \mathbf{S}_i$ by optimizing a cost function (neg-entropy, kurtosis, ...) that is minimal when the sources are the farthest away from Gaussian distributions. Among the several methods that exist depending on the cost function, the prominent ones are FastICA [15], InfoMax [11] and JADE [28].

In a practical case, the multiple dimensions of $\mathbf{Y}$ are often correlated which makes the ICA decomposition theoretically impossible. For that reason, the signal is "whitened" by estimating a linear transformation $\mathbf{L}$ such that $\mathbf{Y}' = \mathbf{L}\mathbf{Y}$ and the transformed signal auto-correlation $E[\mathbf{Y}'^T\mathbf{Y}'] = \mathbf{I}$. This property is achieved with $\mathbf{L} = \sqrt{E[\mathbf{Y}^T\mathbf{Y}]}^{-1}$ since $E[\mathbf{Y}'^T\mathbf{Y}'] = E[\mathbf{Y}^T\mathbf{L}^T\mathbf{L}\mathbf{Y}] = \mathbf{I}$.

It is worth noting that, unlike PCA, ICA does not provide any ordering of the different components; therefore, ICA's application to dimensionality reduction is not straightforward. In [187], Uzumcu *et al.* investigate the ordering of the sources for dimensionality purposes. LDA, PCA and ICA retrieve independent components that are intrinsic to the input signal and project the signal onto these components. Another way to proceed is to project the signal onto a canonical orthogonal basis of the functional space.

1.3.2 Decomposition using Non-Linear Operators

Fourier descriptors Let us assume a shape is represented in a parameterized way: $\mathbf{Y}(p)$, $p \in [0, P]$; then, a Fourier analysis may be run over that input signal to retrieve the Fourier coefficients $\mathbf{u_k}$, so that the Fourier transform of the input signal $\widehat{\mathbf{Y}}$ is written as:

$$\widehat{\mathbf{Y}}(\nu) = \int_{t=0}^{\infty} \mathbf{u}(t) e^{2\pi j t \nu} dt, \tag{1.16}$$

or, sampling the shape function on N points, FFT is used to retrieve the discrete FD

$$\mathbf{u_k} = \sum_{k=1}^{N} \mathbf{Y}(p) e^{-2\pi j \frac{kp}{N}}. \tag{1.17}$$

In the context of shape analysis, these Fourier coefficients are referred to as *Fourier descriptors* (FD). FD are used in the context of shape classification [112] and denoising. No need to say that the FD depend on the choice of shape representation; for a comparative study of FD with different shape signatures, one may refer to [201].

Radial basis functions (RBF) [29][64][121][110] is another decomposition basis in the functional space. Just like FD, RBF decomposes the input signal $\mathbf{Y}(p)$ into a linear combination of primitive functions called *kernels* ρ:

$$\mathbf{Y}(p) = \sum_{i=1}^{N} \mathbf{a_k} \rho \left(|p - c_k| \right), \tag{1.18}$$

where the c_k are called *basis function centers*.

Zernike moments have been proposed by Teague [177] to decompose a discrete image. Teague introduced a set of complex polynomials that form an orthogonal basis onto which an image (approximated by a piecewise function) is decomposed. The projection of the image function on a particular Zernike polynomial is called a Zernike moment. The set of moments characterizes an image, therefore Zernike moments are of particular interest for image retrieval [132], denoising [9] and image classification [193]. The main limitation of these methods is that establishing a connection between the prior model and the image domain where the data support is available is not straightforward.

1.3.3 Prior Models for Geodesic Active Contours

Geodesic active contours (see section (1.2.1)) offer two main advantages: they can be embedded in the level-sets framework for an implicit shape representation, and additional energy terms can be introduced to the equation (1.4). These extra energy terms may be intrinsic to the shape (curvature, length, ...) or extrinsic (e.g. referring to the image or to prior knowledge). If a shape constrain is available, it is added as an extrinsic energy (e.g. the L_2 distance between the current solution's

level-sets and the prior constrain's).

Since the level-sets are local optimization techniques, they are sensitive to initialization. Consequently, many model based methods [44][140] use a prior about the initial contour's location, while others [30][105] use geometric models. *Active shape models* [38][186] use the modes of variation to constrain the shape evolution during the energy minimization. These modes of variation are computed from prior knowledge by using PCA built from prior knowledge. In this basis, a shape is characterized by its distance to the mean shape along statistical modes of variation. However, the shapes' distribution on the modes may have very different variances that depend on the Eigenvalues of the covariance matrix (see section (1.14)). Thus, instead of using the Euclidean distance between the solution's level-sets $\mathbf{Y_1} = \{x_i^1\}$ and the prior constrain's $\mathbf{Y_2} = \{x_i^2\}$, the Mahalanobis distance is preferred:

$$D_{\text{mahal}}(\mathbf{Y_1}, \mathbf{Y_2}) = \sqrt{\frac{\sum_i \lambda_i^2 (x_i^1 - x_i^2)^2}{\sum_i \lambda_i^2}}. \tag{1.19}$$

In [114], prior statistical shape analysis is combined with the geodesic active contours to constrain the solution to the most likelihood solutions. When the prior dataset's samples distribution in the PCA space does not fit a Gaussian, a solution consists in using a kernel distance in the PCA space [155], or in using a kernel method for dimensionality reduction [48][44].

1.3.4 Model-Based Segmentation Limitations

Compared to model-free segmentation, model based segmentation uses a priori knowledge mostly in the geometry space about the object to detect. This makes the segmentation process generally more robust but also introduces two main limitations. The first inherent limitation is the constraint one introduces by using a model. By limiting the segmentation to certain degrees of freedom, the role of local variations and image support is diminished. Also, in most cases, models are built from training sets where the available information (geometry and apperance) is treated equally, regardless of the image support. This problem is addressed in chapter 2 where a method is proposed to build a model on carefully chosen sparse information. Furthermore, the models presented in section (1.3) are time-invariant and are insufficient to solve tracking problems such as the one presented in chapter 3. Last, in the context of optimization in time or space, the global solution may not just consist of the collection of locally most likely solutions. Since only the optimal solution

is looked for at each time step, or space scale, and no uncertainty model is used to represent other potential solutions, nothing guarantees the global solution is achieved when the full time or space information is considered. This issue is dealt with in chapter 4, where a segmentation problem is turned into a geometric tracking problem, and an uncertainty model is integrated into the segmentation framework.

1.4 Bayesian Processes for Modeling Temporal Variations

Many segmentation problems contain a time variable either real or virtual. A simple geometric model does not contain time information, which has to be modeled aside. This section introduces time sequence models starting from regression on deterministic variables, then presenting parametric and non-parametric models for random variables with and without linear state transition.

Autoregressive models. Let $\mathbf{X_t}$ be a state variable at time t, and $\mathbf{Y_t}$ be the observation variable. The general form of regression is a function f such that

$$\mathbf{X_t} = f(\mathbf{X_{t-1}}, \mathbf{X_{t-2}}, ...\mathbf{X_{t-p}}, \mathbf{Y_{t-1}}, ...\mathbf{Y_{t-q}}). \tag{1.20}$$

However, many physical time-series are approximated using autoregression laws such as:

$$\mathbf{X_t} = \mathbf{A}[\mathbf{X_{t-1}}^T\mathbf{X_{t-2}}^T...\mathbf{X_{t-p}}^T] + \mathbf{W} + \epsilon, \tag{1.21}$$

such that the random noise ϵ is zero-mean with minimal covariance determinant. This regression law is established from prior knowledge and used to predict and constrain future states. In a general context, the regression law is not updated, therefore this technique is unable to sustain variations in the dynamic system.

The success of autoregressive (AR) models lies in the vast literature dealing with the subject and their easy implementation. However, AR models are mono-modal and poorly suit natural scene problems where the objects to track rarely moves according to one single time-invariant mode. Consequently, either a complex heuristic is developed to mix models, or Markov fields are introduced for multimodality, such as in [1].

In [42], a static autoregressive model was developed to produce a prior for level-set based seg-

mentation using PCA (see section (1.3.1)). This allows the dynamic modeling of shape variations such that future shapes are predicted using the past shapes. The main limitation of such models refers to their *time-invariant* nature. Neither the PCA nor the AR model can sustain changes of dynamism or shapes that have not been learned a priori.

Optimum linear filters: Kalman. As described in [194], the Kalman filter [103] is a set of mathematical equations that estimates the state variables of a process in the least square sense. The filter is very powerful in several aspects: it supports estimations of past, present, and future states, and it can even do so when the precise nature of the modeled system is unknown. In some cases, Kalman filter may track non linear processes [150]; nevertheless, as we shall see in chapter 4, the Kalman filter fails to track tubular structures with inhomogeneities (e.g. branchings, pathologies or corrupt data) such as coronary arteries.

Such a filter assumes that the posterior density is Gaussian at each time step, and that the current state x_t and observation y_t are linearly dependent on the past state x_{t-1}. Such assumptions simplify the Bayesian equations to the following form:

$$\begin{cases} x_t &= F_t x_{t-1} + v_{t-1} \\ y_t &= H_t x_t + n_t, \end{cases} \tag{1.22}$$

where v_{t-1} and n_t refer to zero mean Gaussian noise with covariance matrices Q_{t-1} and R_t that are assumed to be statistically independent. The matrix F_t is considered known and relates the former state x_{t-1} to the current state x_t. The matrix H_t is also known and relates the state x_t to the observation y_t. The pdfs are computed recursively according to the formulas that may be found in Kalman's seminal paper [103].

$$\begin{cases} p(x_{t-1}|y_{1:t-1}) = \mathbf{N}(x_{t-1}; m_{t-1|t-1}, P_{t-1|t-1}) \\ p(x_t|y_{1:t-1}) = \mathbf{N}(x_t; m_{t|t-1}, P_{t|t-1}) \\ p(x_t|y_{1:t}) = \mathbf{N}(x_t; m_{t|t}, P_{t|t}) \end{cases} \tag{1.23}$$

with

$$\begin{cases} m_{t|t-1} &= F_t m_{t-1|t-1} \\ P_{t|t-1} &= Q_{t-1} + F_t P_{t-1|t-1} F_t^T \\ m_{t|t} &= m_{t|t-1} + K_t(y_t - H_t m_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} - K_t H_t P_{t|t-1} \end{cases} \tag{1.24}$$

where

$$K_t = P_{t|t-1} H_t^T \left(H_t P_{t|t-1} H_t^T + R_t\right)^{-1} \tag{1.25}$$

When the Gaussian distribution assumption does not hold, the distribution may be approximated by a sum of Gaussians [99]. When the transition between the states or when the observation function is not linear, a first order Taylor approximation is used; in this context, the Kalman filter is referred to as Extended Kalman filter. When a Taylor development is not sufficient, Unscented Kalman filter [190] consists in a deterministic sampling of states (called sigma-points) that are propagated according to the non-linear transition function to update the states' distribution model.

When the state space is discrete and consists only of a finite number of possibilities, the Grid technique is a convenient alternative to Kalman filter [24]. When the number of possibilities is too high, or when the state space is continuous, a different sampling strategy is developed, such as Gibbs sampling and Particle filtering.

Gibbs sampling was developed by Geman & Geman [78] in 1984 as a sampling strategy for the joint distribution of two or more variables when only the conditional distributions are known. Let the state x_t be composed by n parameters $\{a_i\}_{i=1..n}$. Given an initial value of $\{a_i^{(0)}\}_{i=1..n}$ at iteration 0, the algorithm updates $a_i^{(t)}$ for each i according to $p(a_i^{(t)}|a_1^{(t)}...a_{i-1}^{(t)}, a_{i+1}^{(t)}...a_n^{(t)})$.

This algorithm is of particular interest to sample the posterior distribution of a Bayesian network since it depends only on the conditional distributions. In the context of computer vision, the Gibbs sampler is used to learn image statistics and models of noise or texture for image denoising [181] or segmentation [55] [191] and tracking. An interesting work [154] uses the Gibbs sampler to model the response of a region of interest to different filters.

However, the Gibbs sampler only converges toward the state of local maximum entropy and cannot sustain multi-modes problems. Furthermore, the iteration time does not correspond to a physical time; therefore the Gibbs sampling can only be used for static models. Particle filtering is an alternative to Metropolis-Hastings based algorithms for a dynamic estimation of non-parametric

multi-mode distributions.

Particle filtering [67][122] is a sequential Monte Carlo simulation to recursively estimate the Bayesian posterior probability $p(x_t|y_{1:t})$ with samples $\{x_t^i\}_{i=1..N}$ associated to weights $\{w_t^i\}_{i=1..N}$, such that

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^i \delta(x_t - x_t^i). \tag{1.26}$$

Using the Importance Sampling principle, with Importance Density q, the weights are estimated by

$$w_t^i \propto \frac{p(x_t^i|y_{1:t})}{q(x_t^i|y_{1:t})}, \tag{1.27}$$

and the following sequential estimation is derived with the Bayesian rule:

$$w_t^i \propto w_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i, y_t)}. \tag{1.28}$$

When $p(x_t^i|x_{t-1}^i, y_t)$ is approximated by a Gaussian, the extended Kalman filter provides the optimum estimation of the posterior probability, see above paragraph on Kalman filter. A solution is presented in [99] when $p(x_t^i|x_{t-1}^i, y_t)$ is approximated by a sum of Gaussians. In the most general case, a sub-optimal solution consists in taking the transition probability as importance density ; in this case,

$$w_t^i \propto w_{t-1}^i p(y_t|x_t^i). \tag{1.29}$$

Since particle filtering requires neither a parametric model for the probability density function, nor a linear transition between states, it has become a very popular technique for multi-object [122] [134] and robust tracking [158]. However, sequential Monte Carlo sampling is computationally intensive, especially for large dimension systems, as the number of particle grows exponentially with the dimension [51].

1.5 Limitations of current models

Models are commonly used twofold: either by constraining the solution so that the final result is a trade off between a priori information and actual observation, or by constraining the problem's resolution to follow the statistics of prior knowledge. These statistics are learned over the

whole training set, therefore the models are representative of the training set as a whole and not of each individual exemplar. These approaches are sub-optimal: when the total solution can be divided into subcomponents (in space, time, or feature space), the inter-dependence of these sub-components may be used to intrinsically constrain the solution. The method presented in chapter 2 aims at formulating the relationship between information components and modeling uncertainties to reconstruct the whole information from its sub-components.

Furthermore, in the context of tracking in time sequences, the methods described in section (1.3) model the spatial variation of the structure of interest in a probabilistic fashion. Then, during the inference process a constrain on recovering shapes that belong to the learned family is imposed. The methods of section (1.4) introduce a novel dimension by modeling the system dynamics. However, their application field is limited to time-invariant dynamic systems; for a system whose dynamics change with time - the majority of cases - an adaptive mechanism is to be set up. Adaptive statistical schemes such as Kalman filter updates the statistical distribution of the object's features (mean and covariance matrix), but the transition model needs to be updated externally. In [151], particle filtering, combined with shape model, allows changes of dynamics but such a framework is computationally intensive. Chapter 3 demonstrates how time-information is used to build and adapt on-the-fly a dynamic model for tracking that is computationally lighter than particle filtering under the assumption of mono-modality. This model may even be used for stationary problems to model a whole sequence of images without a sequential procedure.

Last but not least, uncertainties modeling with particle filtering has not been used for volumetric segmentation yet. In chapter 4, we introduce a novel application of particle filtering for a segmentation problem that is not solved by other deterministic methods. An object - coronary vessels in our case - is progressively segmented, such that at a given time, the current segmentation's uncertainties are modeled and used for a robust segmentation of the remaining section of the object. Instead of progressively building the segmentation's solution by considering the most likely local solutions only, lower probability hypothesis are also generated.

Chapter 2

Sparse Information Models for Dimensionality and Complexity Reduction

Abstract – *In this chapter, a novel technique for model-based vision is proposed that performs dimensionality reduction while taking into account explicitly the image support. The central assumption of this approach is that one describes/reconstructs an object of interest using a small number of its elements through interpolation. Therefore this method consists of finding the smallest possible set of robust, most representative, best supported components/features which could provide an optimal reconstruction of the original object through a data-driven interpolation method. Such an objective is met through the use of a variational method that involves the selection of features, and the corresponding interpolation strategy. Toward validation of such a concept, three important applications in computer vision are considered: surface reconstruction from laser range data, object segmentation in image sequences and volumetric segmentation in medical imaging. This method performs better than well known linear models in the case of corrupted data and high statistical variance.*

2.1 Motivation and Overview of the Chapter

Computer vision methods often rely on information given by the whole image to solve problems; however, different image regions may have different qualities. In some regions, the pixel intensity gradients may consistently be low or the object to segment/track/reconstruct may sometimes be occluded, while other regions may have consistently strong gradients and the object easily detaches itself out of the background. We call the first *low support* regions, and the second *high support* regions. Intuitively, instead of using the information provided by the full image, one would like to favor the information provided by the high support regions. One may even go further, in most cases of statistical modeling, the training set consists of examples with varying support/quality. Therefore, one may extend the notion of image support toward being associated with the quality of the samples in general. In this chapter, we propose a straightforward approach: the explicit selection of high support regions to extract elements of the solution, and the reconstruction of the rest of the solution from these elements.

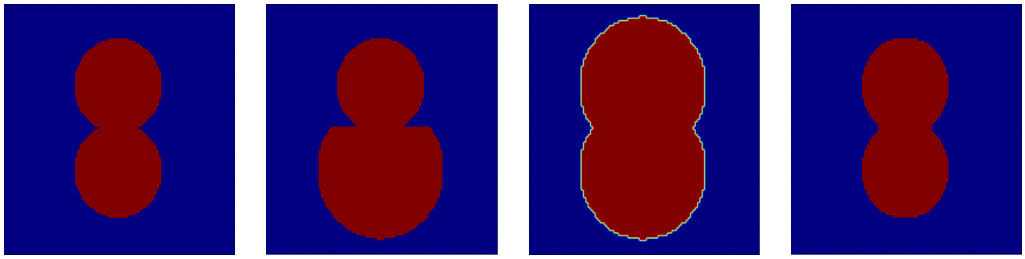
To motivate the learning of data support and its use in prior models, let us consider in section (2.3) an example in segmentation where a shape model that does not account for image support constrains the segmentation toward poorly supported regions. Then, we will introduce the novel technique using sparse information. Section (2.4) generalizes this approach and introduces the notations. Then, this general method is applied to three major problems in computer vision: volumetric segmentation in section (2.5), temporal tracking in section (2.6) and 3D surface reconstruction in section (2.7).

2.2 Prior Art on Image Support Analysis and Interpolation

Confidence measures in the image fall into two classes: measures based on the statistical analysis of the contours and their variability, and analysis of the image support, such as the gradient or the pixels intensity. When the distribution of contours from a learning set is modeled by a Gaussian, the average shape and principal axis are computed with Principal Component Analysis, detailed in chapter 1 section (1.3.1). In [34, 45], only the average shape constrains the segmentation, while in [114] the shape constraint includes the principal modes of variation derived from the PCA. Other distribution models either parametric (Gaussian, multi-Gaussian, kernels, ...) or non-parametric may be used to constrain the segmentation with statistical likelihood from prior knowledge as in

[155, 48]. In [157], the statistical shape variability is compounded in a confidence map so that the local image information weights more in the regions with high statistical shape variability in the training set and less where the shape does not change. The second class of confidence measure is a local likelihood of detecting the contour at a particular location on the image. No exhaustive list of image support measures can be drawn here, but they generally falls into two classes: contour-based or region-based. In the latter, it is usually a function of the gradient (high gradient meaning high probability of contour); in the case of region-based measure, the segmentation favors piecewise smooth regions of similar intensities. Recent works, such as [157], not only take the global shape variability and also use the local variability as a confidence measure in the prior. However, all these methods assume an homogeneous image support. Furthermore, even though PCA is used as a complexity dimension scheme, it still requires the complete reconstruction of the shape during the segmentation process; in other words, the image measure is computed over the whole image, even in regions with notorious low image support. In this chapter, the work focuses on combining both image-based confidence measures and statistical analysis on the contours so that the regions of the image that best support the segmentation and the most stable regions of the contour are selected as a basis to reconstruct the overall surface.

Image interpolation and surface reconstruction have been studied before. The simplest and most common method is to use a spline or piecewise polynomial function [149, 178] that interpolates the contour between explicit points. Other methods use an implicit representation of the contour (a continuous function that takes a zero value on the contour) and interpolating functions such as thin-plate splines [185]. In the context of surface fitting for reconstruction, the main class of fitting method is the *Functional Fitting:* to optimize parameters so that a function best interpolates input points [16]. When the noise is small compared to the data spacing, *Triangulation* methods, such as Delaunay [135], are often preferred. *Volumetric* methods refer to extracting the surface information by computing the volume first. An example of application is the work of Hoppe *et al.* [91] who computed a signed distance function in 3D which is the distance in R^3 to any input point. Then, from the zero levelset of this function is extracted the surface using the Marching Cubes [119]. At last, *Deformable Models* [116] are used to minimize an energy function of the mesh by deforming the mesh so that the mesh is simultaneously attracted to the data by an image term, kept smooth by a tension term and by an optional prior term. More recently, combinations of classic splines and flows of diffeomorphisms have been under study [101], and Younes [200] proposed a numerically more stable solution using affine transformations. For more information on Interpolation for Image Interpolation, Resampling and Registration, the reader may refer to



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

Fig. 2.1: Synthetic example where only the image top half contains relevant information. (a) Correct shape. (b) Corrupted image. (c) PCA result. (d) Sparse information result.

[180][79][123][153].

2.3 Introduction with an Ad Hoc Example

A standard shape model technique consists in using principal component analysis (PCA) [38] (see section (1.3.1)) to project the shape space onto axes that correspond to statistical modes of variation for that class of shapes. The image energy function is optimized in that projected subspace. PCA presents two main advantages: first, the optimization in the subspace depends on few parameters (compared to the original shape space) which makes it faster and less sensitive to local minima, second PCA guarantees the segmentation's result is constrained by statistical modes of variation from the training set. Consequently, the segmentation's result is a trade off between information from the image and statistical coherence. However, missing or corrupted image data may interfere with statistics from the training set and lead to poor results. Let us illustrate this point with a simple example.

Let us consider a synthetic example consisting of sets of two circles with equal but varying radius. The mean shape corresponds to the average radius circles, and the modes of variations account for the circles' radius. Let us further assume the image is only relevant for the top circle; the bottom half of the image may be occluded or corrupted. When the image bottom half section corresponds to a mode of variation (a circle), the prior model leads to false results, results that are incoherent with the image top section (see figure (2.1)). Therefore, data support must be accounted for in choosing which measurements are optimal for the processing of the image. Then, depending on these observations, a learned prior is used to analyze the image and extract the region of interest.

In the present synthetic case, it is clear (see figure (2.1)) that the bottom half section of the image is a low support region that should not be considered; only the partial information contained in the top half is relevant. Therefore, the sparse information technique consists first in dividing the image into m elements (e.g. 2 elements in our case: top and bottom halves) and assigning image support weights $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_m)$ to these elements (e.g. $\mathbf{w}_{\text{top}} = 1$ and $\mathbf{w}_{\text{bottom}} = 0$). Second, one selects the elements, also called *key elements*, best supported by the image (e.g. the top half in our case), and a function ϕ to reconstruct the rest of the solution (e.g. a symmetric reflexion in our case, see figure (2.1)).

Three main criticisms may be objected against this technique: first, time constrains aside, does the selection of measurements provide better results than common dimensionality reduction techniques? Second, since the number of measurements is reduced, what is the sensibility of the solution with respect to input noise? Third, what is the performance of sparse information models in the presence of outliers and salt and pepper noise? These three questions drive us to slightly modify the way we select the key elements: we need an image support measure E_{sup} to select the elements best supported by the input image, we also need a metric E_{int} to measure the quality of reconstruction, and finally a measure E_{var} to avoid selecting key elements that vary a lot with respect to their location. After a general introduction to sparse information models is given in section (2.4), three different experiments are conducted in sections (2.5) to (2.7) to study these questions.

2.4 General Introduction to Sparse Information Models

Let us consider a training set of exemplars $\mathcal{X} = \{\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^P\}$ registered in a reference space Ω_r. Each exemplar is divided into m sub-elements $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_m)$ (see figure (2.2) for an example on volumetric shapes) associated to m measures $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_m)$ which reflect the data support for the observations.

Without loss of generality, we assume the m sub-elements are obtained by a discretization process along one or several axis $\mathbf{v}_0$ using an operator $\rho : [\Omega_r \times \mathbb{R}] \rightarrow \Omega_r/\mathbf{v}_0$:

$$\forall\, i \in [1, m], \quad \mathbf{x}_i = \rho(\mathbf{x}, i) \tag{2.1}$$

In the remaining of the chapter, this continuous parameterization is assumed when not specified.

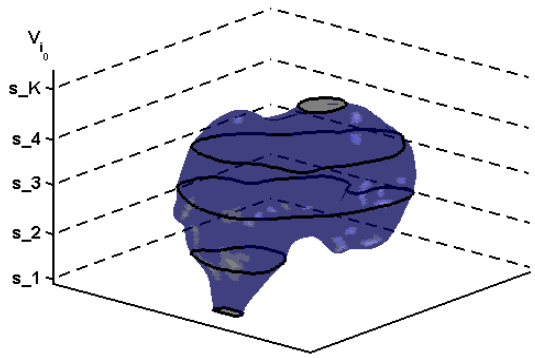


Fig. 2.2: Example of a 3d liver surface $\boldsymbol{x}$ *whose subcomponents* $(\boldsymbol{x}_1, ..., \boldsymbol{x}_m)$ *are obtained by intersecting the 3d shape with the axial plane (dark contours) at specific slice indices* $s_1, s_2, ...s_K$.

The aim of our approach is to recover a minimal description length set of $|\mathcal{B}| = K$ sub-elements $\mathcal{B} = \{\mathbf{x}_{t_k}\}_{k\in[|1,K|]}$ with K small compared to m, and a continuous operator ϕ, from which the whole data $\mathbf{x}$ is deducted:

$$\forall\, k \in [1, m], \quad \phi(\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_K}, k) = \mathbf{x}(k). \tag{2.2}$$

Given such a model, one has to address three important aspects: (i) the number of basis components $|\mathcal{B}| = K$, (ii) the form of these components and (iii) the interpolation strategy ϕ used to reconstruct the training set examples. We have at our disposal the training set $\mathcal{X}$ which consists of many samples associated with certain image support $\mathbf{w}$. Three criteria are now developed to optimize the sparse information model with respect to $\mathcal{X}$, $\mathbf{w}$ and the three important aspects presented above in section (2.3).

Optimal reconstruction of the training set. Toward optimal reconstruction of the training set from the basis $\mathcal{B}$, the distance between the reconstruction and the existing samples is a natural criterion/metric that determines the quality of our model. To this end, let a metric $d : [\Omega \times \Omega] \rightarrow \mathbb{R}^+$ measures the distance between two corresponding sub-elements. Then, assuming the number of components of the training set is fixed, such reconstruction minimizes

$$E_{\text{int}}(\mathcal{B}, \phi) = \sum_{p=1}^{P} \sum_{i=1}^{m} d\left(\mathbf{x}_i^p, \phi_i(\mathbf{x}_{t_1}^p, ..., \mathbf{x}_{t_K}^p)\right). \tag{2.3}$$

Such an approach is purely geometric and does not account for the image support of each sub-element. In the case of shape/surface modeling for example such a distance measures the distance between the reconstructed surface and the one being part of the training set. More advanced measurements like the Hausdorff distance or the metrics comparing the implicit representations of the surfaces can be considered.

Optimal image support. We recall that the sub-elements of a given exemplar have some underlying image support noted $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_m)$. Such simplistic notation is equivalent with saying that the more important the value $\mathbf{w}$ is, the best extraction of the solution in this particular location can be obtained using the data associated with it. The optimum basis $\mathcal{B}$ consists of elements that are confidently extracted from the data; therefore, the basis minimizes

$$E_{\text{sup}}(\mathcal{B}) = \sum_{p=1}^{P} \sum_{k=1}^{K} f\left(\mathbf{w}_k \left(\mathcal{T}_\theta^{-1}(\mathbf{x}_{t_k})\right)\right) \tag{2.4}$$

where f is a monotonically decreasing function, and $\mathcal{T}_\theta^{-1}(\mathbf{x}_{t_k})$ is the inverse mapping between the basis $\mathcal{B}$ and the observation space. This term is evaluated within the entire training set and in other words for a simple surface modeling, we are seeking for the indexes of 2D shapes used to form this surface for which one determines optimally for the corresponding contours for all samples of the training set. The use of such inverse mapping is also to be considered during the application of the model to new data. Therefore, it is critical to have a selection of $\mathcal{B}$ that is relative robust to errors when locating the basis' elements in a new exemplar.

Robustness to parameters variability. Let us consider a slight variation on the selection of the basis, noted $\delta\bar{\mathbf{x}}$. Such a variation can be either on the indexes/positions of the key slices or related with the form of the shapes being present on the key slices. For the interpolation precision of the model not to be significantly affected,

$$\lim_{|\delta\bar{\mathbf{x}}| \to 0} \frac{E_{\text{int}}(\mathcal{B}, \phi) - E_{\text{int}}(\mathcal{B} + \delta\bar{\mathbf{x}}, \phi)}{\delta\bar{\mathbf{x}}} = 0 \tag{2.5}$$

that is reformulated in terms of a cost by defining a smoothness function $\eta()$, like the error-two norm,

$$E_{\text{var}}(\mathcal{B}, \phi) = \eta\left(\nabla_{\mathcal{B}} E_{\text{int}}(\mathcal{B}, \phi)\right). \tag{2.6}$$

Such a penalty term introduces robustness in the basis selection step, as well to the reconstruction process. More precisely, since the aim of our model is to perform knowledge-based segmentation,

one expects that errors will be present in the inference step. Therefore, it is reasonable to assume that when a key element is found in the neighborhood of the optimal position its form slightly deviates from the expected value. The model should be capable to cope with that. Now, one integrates these three constraints into a single cost function:

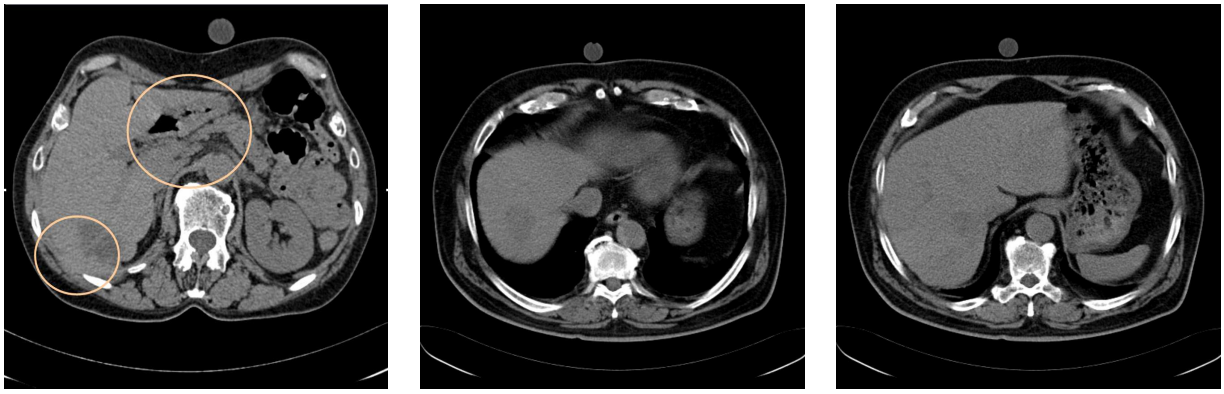
$$E(\mathcal{B}, \phi) = E_{\text{int}}(\mathcal{B}, \phi) + \alpha E_{\text{sup}}(\mathcal{B}) + \beta E_{\text{var}}(\mathcal{B}, \phi) \tag{2.7}$$

The cost function E is minimized with respect to the interpolation function ϕ and the basis $\mathcal{B}$. Such a process cannot be described in a general fashion, but a gradient descent is an excellent choice when considering linear interpolation models, while more advanced non-linear optimization methods like neural networks can be considered for non-linear cases. Last, but not least the residual cost that characterizes the *sparse information model* is used to determine the best number $|\mathcal{B}|$ of key components that optimizes the minimum description length [86]. In order to demonstrate the efficiency of such a model, we consider three different applications: 3D knowledge-based segmentation, multi-frame object segmentation and surface reconstruction from range data.

2.5 Sparse Knowledge-based Segmentation: Automatic Liver Delineation

As mentioned earlier, the example of knowledge-based segmentation is the most direct application of such a dimensionality reduction technique. In this case, the sparse model is built by selecting a minimal set of key element indices $\mathcal{B}$ of 2D contours (represented in an explicit or an implicit fashion) along with an interpolation function ϕ to reconstruct the whole 3D surface in the reference space Ω_r. During the segmentation, one should expect that the pose of the object is different than the one used to learn the model. Therefore, segmentation is equivalent with finding this pose in the new volume, and at the same time determining the geometry of the 2D curves that correspond to the key slices. Therefore, the global transformation $\mathcal{T}_\theta$ that relates the reconstructed model to the volume in the observation space is to be determined, along with the set of 2D key contours that fit the observation. Medical image segmentation is a perfect scenario to demonstrate such a framework. Training sets are available, often manually delineated from experts, and the image support varies according to the physical properties of the organs.

Computerized medical imaging analysis aims at detecting and delineating anatomical structures for diagnosis and therapy. It has gained significant attention in hepatic procedures, specially in oncology. The main aim is to detect tumors and lesions, quantify the ratio of tumors' volume



(a) (b) (c)

Fig. 2.3: Slices of liver dataset with a poor edge regions. (a) Intercostal muscles and neighboring organs have a similar appearance (b) The lowest part of the heart juxtaposes the liver in a way that could correspond to a PCA shape variation (c) Slice with strong image support.

and liver's volume (future liver remnant volume and total liver volume), their localization with respect to the liver's vasculature and the different lobes of the liver [127][152]. Also, in the context of liver transplantation, graft from living donors is increasingly performed due to the shortage of cadaveric donors. This particular procedure requires a pre-operative quantification of the donor's liver volume [89]. However, the segmentation of the liver is an arduous task for two main reasons. First, the liver's appearance and shape has a large inter-patient variability; it is one of the largest organ of the human body, after the skin, and imaged patients may suffer from heavy diseases such as cancer. Second, the neighboring structures have similar appearance in CT and MR, and may juxtapose the liver in a way that corresponds to a statistical shape variation, or without clear edge between the two (see figure (2.3)). In this section, we propose a novel method for liver segmentation that combines a statistical analysis of the data with a reconstruction model from sparse information: only the most reliable information in the volume is used, and the rest of the liver's shape is inferred from the model and the sparse observation.

2.5.1 Prior art on liver segmentation

Given the difficulty of segmenting the liver, a model is commonly used, either as a localization, a shape or an appearance constraint. In [172], a cascading segmentation scheme sequentially detects the different abdominal structures for hepatic surgery planning. In [147][166], the liver is detected by a classification procedure based on the pixels intensity. A semi-automatic procedure is presented in [164] where a live-wire based on dynamic programming assists the user in drawing

the liver's contour slice by slice. In the context of deformable models, the reason for using prior knowledge is that this segmentation method is a local optimization, and therefore is sensitive to local minima. Therefore, the prior work on liver segmentation includes models based on shape variations, and constrains on pixel intensities learned from classification. Intensity based methods are used in [118] with snakes segmenting the liver's contour in a slice-by-slice fashion. However, the vicinity of the liver to neighboring structures of similar appearance makes models attractive for this task [88]. The models used in [88] are based on Cootes et al.'s active shape models [38]. The results obtained in [88] demonstrate active shape models may represent to a certain extent the liver's shape. However, they fall short of accurately segmenting the liver because of the large shape variability of this organ. Furthermore, active shape models are highly dependent on initialization, that is dealt with a multi-resolution approach. Nonlinear models, such as shape-based kernel PCA [49] or Fourier coefficients [56] have been investigated more recently for segmentation. The main limitation of these methods (linear or non-linear) is the explicit assumption of the data distribution that, for example, forms a linear subspace in the case of PCA. These methods process the total amount of data and find the optimum trade-off between an image term and a prior term. Furthermore, the quality of image support is at no point taken into account; it is assumed that should an image region quality be low, another region would compensate. This assumption leads to errors (see section (2.3)) since most of these methods treat segmentation as a statistical estimation problem, where the quality and the support of the training set's exemplars is ignored.

Instead, the approach presented in this section relies on observation at key locations, and on a reconstruction model ; both the key locations and the reconstruction models are learned from a training set that consists of registered manually segmented liver volumes.. This technique provides better results because the segmentation is only supported by the data with the strongest image support, and is also of low complexity because it uses the data in an optimal fashion. We propose a liver model that encodes the shape variations using a small number of carefully chosen key-slices where the organ's contours can be optimally recovered. First, the image or shape to reconstruct is discretized along the longitudinal axis, and all the liver exemplars are registered rigidly so that they fit into the same reference region. Then, a set of slice indices are determined so that it minimizes three different criteria: image support, quality of the reconstruction and sensibility to variations in the projection's subspace. Finally, the reconstruction operator itself is learned over the given liver exemplars. To present this approach, section (2.5.2) is devoted to explicit this model to the particular problem of 3D liver segmentation. Then, to validate the methodology, section (2.5.6) aims at proving the liver's shape is indeed well recovered from few contours at key-slices, and quantifying

the quality of the segmentation obtained in this way. The choice of interpolation function, generalized linear function, is performed empirically by comparing different interpolation functions. The results are reported in appendix A not to interfere with the presentation of the sparse information models.

2.5.2 Model Estimation

The experiment is conducted on segmentation for medical imaging for the case of liver in Computed Tomography (CT). We represent the training set exemplars $\mathbf{x}$ through an implicit model (see section (1.2.1)). A distance function ψ is defined as

$$\forall p \in \Omega, \;\; \psi(p) = \begin{cases} 0 & , p \in \mathcal{C} \\ +\mathcal{D}(p) \geq 0 & , p \in \Gamma \\ -\mathcal{D}(p) < 0 & , p \in \bar{\Gamma} \end{cases} \tag{2.8}$$

and $\mathbf{x}$ is defined as a column vector by aligning the distance value of each pixel in lexicographic order. In the following, the role of ψ and $\mathbf{x}$ may be interchanged by abuse of notation. Such a selection is motivated by the direct ability to determine the properties of such a representation, as well as the ability to introduce surface-based as well as area based criteria in the segmentation process. Furthermore, the liver's topology on the axial plane may change, which is implicitly accounted for by using this representation. Classic explicit parameterizations like triangulated surfaces, or other form of parametric snakes can also be considered.

The acquisition process guides our choice for the definition of the sub-elements: since the image volume is reconstructed slice by slice, with maximum resolution in the slice plane, the axis of projection $\mathbf{v}_0$ is the longitudinal axis. Therefore, a sub-element $\mathbf{x}_i$ corresponds to a particular slice (see figure (2.2)). The geometric transformation T_θ is a translation-scaling that sets $\mathbf{x}$ in a reference space Ω_r with m slices $(\mathbf{x}_1, ...\mathbf{x}_m)$. The choice of the interpolation is rather trivial: we have assumed that any surface slice is determined through a generalized linear interpolation of the key slices.

Optimal Reconstruction. From the experiments of appendix A, it is concluded that generalized linear interpolation for each slice i is a good compromise between complexity and interpolation quality. In other words, the solution (2D contour) at each slice $\mathbf{x}_i$ is reconstructed using a

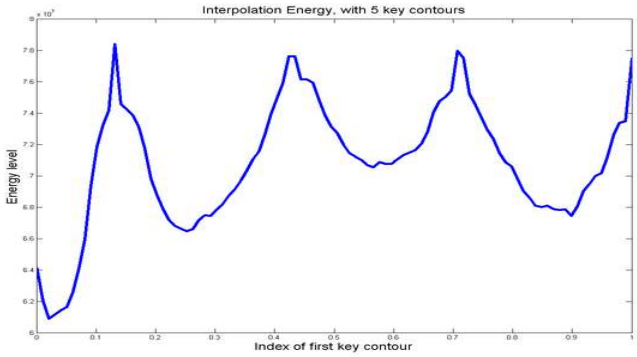


Fig. 2.4: Interpolation quality E_{int} (equation (2.9)) with respect to the first key contour index, with $K = 5$ and the four other contours indices at (0.13, 0.45, 0.73, 1).

particular linear combination $\mathbf{H}_i$ of the key contours $\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}$. This notation is kept in the remaining of the chapter: $\phi = \mathbf{H}$ and $\phi_i = \mathbf{H}_i$. One now defines the objective function component which corresponds to the optimal reconstruction criterion of the training set. The interpolation quality is defined according to the sum of squares difference between the reconstructed distance map and the observed shape's distance map in the reference space Ω_r:

$$E_{\text{int}}(\mathcal{B}, \mathbf{H}) = \sum_{i=1}^{m} \int_{\Omega_r} \left| \mathbf{H}_i \left[\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}\right]^T - \mathbf{x}_i \right|^2 \tag{2.9}$$

This term is often considered to account for prior knowledge within level sets through a direct comparison between the distance function of the model and the one evolving in the image. E_{int} is a quadratic function with global minimum (see figure (2.4)), and since the reference space Ω_r is a continuous space, the minimization of E_{int} benefits from the large literature on quadratic functions minimization. For a particular set of key contours $\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}$, the vector $\mathbf{H}_i$ is obtained through a least squares minimization of equation (2.9) for each individual i. To keep notations simple, let $\mathbf{X}_t = [\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}]$. First, $\mathbf{H}_i\mathbf{X}_t^T = \mathbf{x}_i$ is rewritten as $\mathbf{X}_t\mathbf{H}_i^T = \mathbf{x}_i^T$. Then, a QR decomposition of $\mathbf{X}_t$ is performed from Householder reflections: $\mathbf{X}_t = \mathbf{QR}$ where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper triangular. The optimization of $\mathbf{H}_i$ is then straightforward.

The derivative of E_{int} with respect to the key contours indices is computed directly as

$$\frac{\partial E_{\text{int}}}{\partial k_0} = \sum_{i=1}^{m} \int_{\Omega_r} 2\mathbf{H}_i[0...1...0]^T \left(\mathbf{H}_i\mathbf{X}_t^T - \mathbf{x}_i\right).$$

Quality of Image Support. The next step consists of defining a qualitative metric of the image support for a given slice. In the introduction we have reviewed the state of the art in segmentation. Edge-driven or region-based methods have been considered with the later being less sensitive to noise and initial conditions. To separate the statistical properties of the object from the background has been a common choice to implement region-based terms. Driven from that, the image support $\mathbf{w}_i$ at slice i is defined by the Kullback-Leibler distance between the pixels intensity distributions inside and outside the 2D contour and the a priori learned histograms. Knowing a priori the distributions p_{in} (resp. p_{out}) of the pixels intensity inside (resp. outside) the liver, and computing the normalized pixels intensity histograms h_{in} and h_{out} inside and outside of the reconstructed shape on the key slices,

$$\begin{aligned} E_{\text{sup}}(\mathcal{B}) = &\sum_{k=1}^{K} \int h_{\text{in}}(k,s) \log\left(\frac{h_{\text{in}}(k,s)}{p_{\text{in}}(s)}\right) \, ds \\ &+ \sum_{k=1}^{K} \int h_{\text{out}}(k,s) \log\left(\frac{h_{\text{out}}(k,s)}{p_{\text{out}}(s)}\right) \, ds. \end{aligned} \tag{2.10}$$

This term seeks slices for which the separation of liver/background is the best possible using the observed image intensities. The normalized pixels intensity histograms h_{in} and h_{out} depend on the contour and image index that is to be optimized. Let h'_{in} and h'_{out} be the derivative of these histograms with respect to the contour index. Then, the derivative of $E_{\text{sup}}(\mathcal{B})$ with respect to a particular contour index k_0 is

$$\begin{aligned} \frac{\partial E_{\text{sup}}(\mathcal{B})}{\partial k_0} = &\int h'_{\text{in}}(k_0,s) \log\left(\frac{h_{\text{in}}(k,s)}{p_{\text{in}}(s)}\right)\left(\frac{h_{\text{in}}(k,s)+p_{\text{in}}(s)}{p_{\text{in}}(s)}\right) ds \\ &+ \int h'_{\text{out}}(k_0,s) \log\left(\frac{h_{\text{out}}(k_0,s)}{p_{\text{out}}(s)}\right)\left(\frac{h_{\text{out}}(k_0,s)+p_{\text{out}}(s)}{p_{\text{in}}(s)}\right) ds \end{aligned} \tag{2.11}$$

Figure (2.5) displays the global minimum of E_{sup} with respect to any key contour index t_k.

Robustness to Key-Contours Variability. Finally, the key contours are chosen so as to minimize the impact of little variations in their position, and of little errors in the contours extraction in the key slices. Since a continuous interpolation of the 2D contours is introduced in equation (2.1), the impact of an infinitesimal change ∂k in the slice index may be written as the squared magnitude of the gradient of $\mathbf{x}_{t_k}$ with respect to t_k: $\|\nabla_{t_k}\mathbf{x}_{t_k}\|^2$. In practice, since the contours are represented using distance functions (see equation (2.8)), the derivative of the distance function at

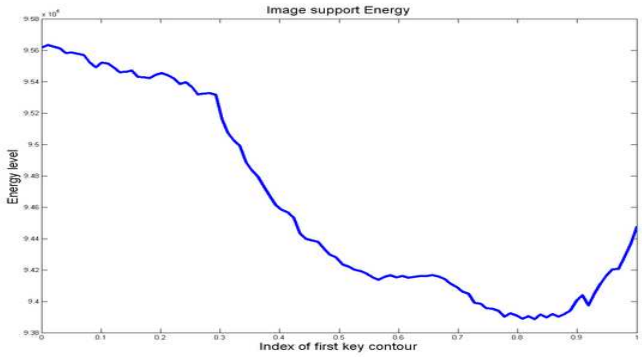


Fig. 2.5: Image support quality E_{sup} (equation (2.10)) with respect to one particular contour index.

index t_k, with respect to the index, is a field of 2d vectors whose squared magnitude is $\|\nabla_{t_k}\mathbf{x}_{t_k}\|^2$.

The second criterion consists in applying a small variation $\partial\mathbf{x}$ to the key contour $\mathbf{x}_{t_k}$ and evaluate how the reconstructed 3D surface changes. For each $i \in [|1, m|]$, let us note the interpolation vector $\mathbf{H}_i = [h_i^1 h_i^2 ... h_i^K]^T$. Then, for a particular $k_0 \in \{1, ...K\}$, the difference between the reconstructed surface before and after adding $\partial\mathbf{x}$ is

$$\begin{aligned}\sum_{i=1}^{m}\int_{\Omega_r}\left|\mathbf{H}_i\left[\mathbf{x}_{t_1},...\mathbf{x}_{t_{k_0}}+\partial\mathbf{x},...\mathbf{x}_{t_K}\right]^T-\mathbf{H}_i\left[\mathbf{x}_{t_1},...\mathbf{x}_{t_K}\right]^T\right|^2 &= \sum_{i=1}^{m}\int_{\Omega_r}\left|\mathbf{H}_i\left[\mathbf{0}_{1:k_0-1}\partial\mathbf{x}\,\mathbf{0}_{k_0+1:K}\right]^T\right|^2\\ &= \sum_{i=1}^{m}\int_{\Omega_r}\left|h_i^{k_0}\partial\mathbf{x}\right|^2\end{aligned}$$

Therefore, the key contours are chosen so as to minimize the integral over the image space of the distance map's gradient at the key locations and the interpolation vector magnitude:

$$E_{\text{var}}(\mathcal{B}) = \sum_{k=1}^{K}\int_{\Omega_r}\|\nabla_{t_k}\mathbf{x}_{t_k}\|^2 + \sum_{i=1}^{m}\mathbf{H}_i^T\mathbf{H}_i. \tag{2.12}$$

Figure (2.6) displays the global minimum of E_{var} with respect to any key contour index $\bar{\mathbf{x}}_i$.

2.5.3 Estimation of the Model Parameters

E_{sup} solely depends on the key contours indices; however, E_{int} and E_{var} depend on both the key contours indices and the interpolation matrix. Therefore, the optimization strategy consists

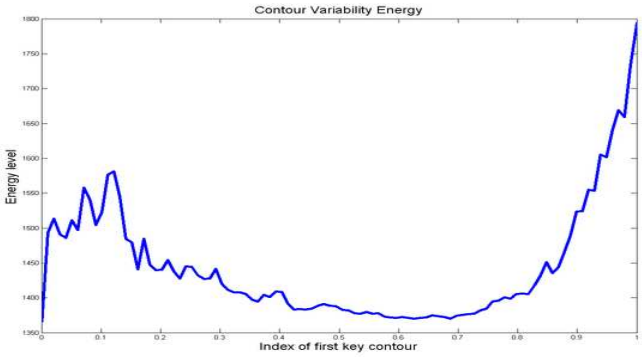


Fig. 2.6: Image support quality E_{var} (equation (2.12)) with respect to one particular contour index.

in alternating a gradient descent step with respect to the indices on $E_{\text{sup}} + E_{\text{var}} + E_{\text{int}}$ with $\mathbf{H}$ constant and the least squares optimization of $\mathbf{H}$, followed by the estimation of the residual error E_{int}. With $\mathbf{H}$ constant, the derivative of E_{sup} with respect to a particular key contour index k_0 is

$$\begin{aligned}\frac{\partial E_{\text{sup}}}{\partial k_0} = \int \frac{\partial h_{\text{in}}}{\partial k}(k_0, s)\log\left(\frac{h_{\text{in}}(k_0, s)}{p_{\text{in}}(s)}\right)\left(\frac{h_{\text{in}}(k_0, s) + p_{\text{in}}(s)}{p_{\text{in}}(s)}\right) \\ + \frac{\partial h_{\text{out}}}{\partial k}(k_0, s)\log\left(\frac{h_{\text{out}}(k_0, s)}{p_{\text{out}}(s)}\right)\left(\frac{h_{\text{out}}(k_0, s) + p_{\text{out}}(s)}{p_{\text{out}}(s)}\right) ds.\end{aligned} \tag{2.13}$$

The derivative of E_{var} with respect to k_0 with $\mathbf{H}$ constant is straightforward:

$$\frac{\partial E_{\text{var}}}{\partial k_0} = \int_{\Omega_r} 2\frac{\partial^2 \mathbf{x}_{t_k}}{\partial k^2}\frac{\partial \mathbf{x}_{t_k}}{\partial k}.$$

As one may see in figures (2.5) and (2.6), $E_{\text{sup}} + E_{\text{var}} + E_{\text{int}}$ is sufficiently smooth to solve this problem with gradient descent. This smoothness is guaranteed to a certain extent by the large size of the training set (31 liver volumes). Should the training set be reduced, or a different type of image be considered, gradient descent may not be suitable anymore. Other techniques such as the simplex optimization [130] should be considered should that situation happen. Furthermore, the factors that depend on the $\{t_1, t_2, ...t_K\}$ in $E_{\text{sup}} + E_{\text{var}} + E_{\text{int}}$ are independent with each other except for the optimization of $\mathbf{H}$. Therefore, it is reasonable to optimize them independently for each key contour index t_k instead of using global optimization schemes such as Gibbs sampling optimization (see chapter 1 section (1.4)). Last but not least, the energy to optimize for $\mathbf{H}$ in equation (2.9) is quadratic. In such a case, a linear least squares optimization scheme is appropriate. The orthogonality constraint on $\sum_{i=1}^{m} \mathbf{H}_i^T \mathbf{H}_i$ is transformed into a spherical constraint:

Initialize $\{t_1, t_2, ...t_K\}$
for $\tau = 1..T_\tau$ **do**
  foreach $i = 1..m$ **do**
    Compute Jacobian $\mathbf{J}$ of $\mathbf{H}_i^{(\tau)}\mathbf{X}^{(\tau)^T} - \mathbf{x}_i$
    $\mathbf{G} = \left(\mathbf{J}^T\mathbf{J}\right)^{-1}\mathbf{J}^T(\mathbf{H}_i^{(\tau)}\mathbf{X}^{(\tau)^T} - \mathbf{x}_i)$
    $\mathbf{H}_i^{(\tau)} = \mathbf{H}_i^{(\tau)} - \mathbf{G} + \mathbf{H}_i^{(\tau)}\mathbf{G}^T\mathbf{H}_i^{(\tau)}$
  end
  foreach t_k **do**
    $t_k = t_k - \frac{\partial E_{\text{sup}}}{\partial k_0} - \frac{\partial E_{\text{var}}}{\partial k_0} - \frac{\partial E_{\text{int}}}{\partial k_0}$
  end
end

Algorithm 1: - General pseudo-code for the estimation of the sparse information models parameters.

$\forall i = [1, m]$ $\mathbf{H}_i^T\mathbf{H}_i = 1$. If a Gauss-Newton strategy is followed for E_{int}, there is no guarantee that the intermediate solution remains on the sphere. Therefore, it is necessary to project the Newton gradient $\mathbf{G}$ on the sphere, so that each step displacement in the Newton optimization is actually $\mathbf{G} - \mathbf{H}_i\mathbf{G}^T\mathbf{H}_i$.

A general pseudo-code for the estimation of the sparse information models parameters is given in the algorithm 1. For simplicity, $[\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}]$ at iteration step τ is noted $\mathbf{X}^{(\tau)}$.

The optimum number K of key contours is estimated using the Schwarz Bayesian Criterion [86] on the residual energy remaining after the minimization is completed. More details on Schwarz Bayesian Criterion are given in the following application, in section (2.6.3).

Using this set of equations one recovers the optimal basis, the optimal interpolation strategy and the optimal number of key elements. In order to validate this process, the first step consists of comparing the sparse information model with standard dimensionality reduction techniques like PCA, ICA, etc.

2.5.4 Comparative Study between PCA and Sparse Information Models for Dimensionality Reduction

To this end, one needs to quantify the error introduced by the sparse models dimension reduction and compare it with common techniques such as PCA. The volumetric data is acquired on Sen-

method	PCA (5 modes)	PCA (30 modes)	Linear interp.	SIM
median symmetric diff.	16.97%	11.70%	10.72%	8.35%
maximum symmetric diff.	29.15%	23.32%	16.13%	13.14%
minimum symmetric diff.	8.61%	6.56%	7.69%	6.28%

Tab. 2.1: Results table showing the median, maximum and minimum symmetric difference between ground truth volumes and reconstructed volumes using PCA (5 and 30 modes), linear interpolation from 5 key slices and sparse information model (SIM) with 5 key slices.

sation 16 CT scanners, with an average resolution of 1 mm in axial plane and 3 mm along the longitudinal axis. 31 volumes (different oncology patients, with or without pathologies such as tumors) are used in our experiments on a leave-one-out basis: 30 volumes are used to build the models (sparse and PCA) and the last one is used for testing.

Table (2.1) summarizes the error introduced by dimensionality reduction for PCA (5 and 30 modes), linear interpolation and sparse information model with 5 slices. This error measure is defined as the symmetric difference [164] between the two volumes V_1 and V_2:

$$\epsilon = 1 - \frac{|V_1 \cap V_2|}{0.5 * (|V_1| + |V_2|)} \tag{2.14}$$

The results clearly demonstrate that the sparse information model with 5 key elements provides a far better reconstruction quality as linear PCA with 5 modes of variation, and slightly better than PCA with 30 modes. The PCA results have a large variance because diseased organs are poorly represented by a Gaussian model in the linear PCA space. A larger study with different pathologies and the use of kernel PCA [114] could improve the reconstruction of the shapes.

Figure (2.7) illustrates different error measures for liver segmentation with linear PCA, liner interpolation and sparse information model. The quality assessment is performed with four error measures: the volumetric error in %, the average surface distance, the root mean square (RMS) distance, and the percentage of surface father than 5mm from the ground truth.

2.5.5 Segmentation Scheme

With sparse model in hand, the volumetric segmentation is boiled down to the segmentation of the shape at key slices; in other words, the whole 3D segmentation problem is reduced to a small set

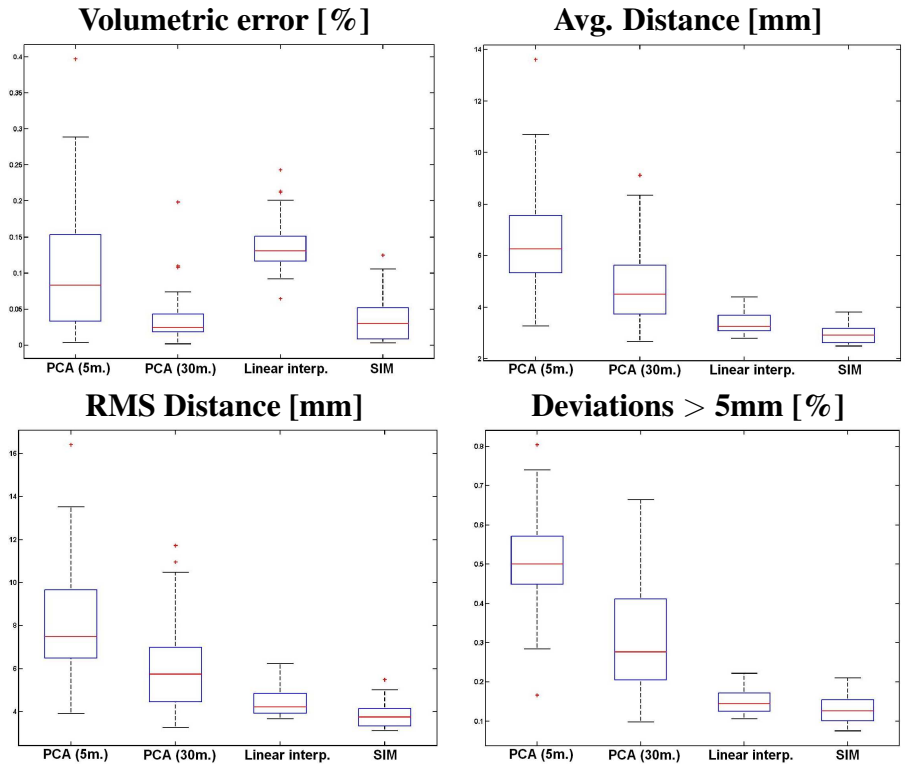


Fig. 2.7: Segmentation result boxplots comparing PCA (5 and 30 modes), linear interpolation and Sparse Information Model. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

of parallel 2D contours to be segmented at specific locations. Therefore, one needs to optimize an image-based cost function with respect to both the set of key contours $\mathcal{B} = \mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}$ in the reference space and the transformation $\mathcal{T}_\theta$ simultaneously. In an iterative optimization scheme, the transformation $\mathcal{T}_\theta$ at a given iteration is used to relate the current set of 2D contours $\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}$ to the image so that both the transformation and the sparse set of contours are optimized concomitantly.

To this end, the cost function consists of the intensity-based likelihood of each pixel, assuming that normalized histograms inside (h_{in}) and outside (h_{out}) the liver are available (if not, one recovers them on-the-fly). Then, the posterior likelihood of the partition with respect to the two

classes is maximized to obtain the key contours $\mathcal{B}$ and the transformation $\mathcal{T}_\theta$:

$$\begin{aligned} E_{\text{seg}}\left(\mathcal{B},\mathcal{T}_\theta\right) = \sum_{k=1}^{K}\int_{\Omega} -\log\left(h_{\text{in}}(I(s))\right)\mathcal{H}(\mathbf{x}_{t_k}(T_\theta(s)))\,ds \\ +\sum_{k=1}^{K}\int_{\Omega} -\log\left(h_{\text{out}}(I(s))\right)\left(1-\mathcal{H}(\mathbf{x}_{t_k}(T_\theta(s)))\right)\,ds, \end{aligned} \tag{2.15}$$

where $\mathcal{H}(\mathbf{x}_{t_k}(s))$ denotes the Heaviside function that is equal to 1 inside the contour $\mathbf{x}_{t_k}$, and 0 outside. During the sparse model's construction the image support has been taken into account in the selection of the key slices. This information has been inherited to the segmentation and, in principle, the slices where one best separates liver from the rest of the background are used (see equation (2.10)). Deriving E_{seg} with respect to each key contour $\mathbf{x}_{t_k}$, one obtains

$$\frac{\partial E_{\text{seg}}}{\partial \mathbf{x}_{t_k}} = \int_{\Omega} -\log\left(\frac{h_{\text{in}}(I(s))}{h_{\text{out}}(I(s))}\right)\delta_{\mathbf{x}_{t_k}}(T_\theta(s))ds, \tag{2.16}$$

where $\delta_{\mathbf{x}_{t_k}}(s) = 1$ if and only if s is on the contour $\mathbf{x}_{t_k}$, and 0 elsewhere. Representing the contour using distance functions (see equation (2.8)) one derives E_{seg} with respect to the geometric transformation parameters θ:

$$\frac{\partial E_{\text{seg}}}{\partial \theta} = \sum_{k=1}^{K}\int_{\Omega} -\log\left(\frac{h_{\text{in}}(I(s))}{h_{\text{out}}(I(s))}\right)\delta_{\mathbf{x}_{t_k}}(T_\theta(s))\nabla\mathbf{x}_{t_k}(T_\theta(s)).\frac{\partial T_\theta}{\partial \theta}(s)ds. \tag{2.17}$$

Using equations (2.16) and (2.17), ones minimizes E_{seg} by gradient descent. When $(\mathcal{B},\mathcal{T}_\theta)$ have reached the energy minimum, the whole volumetric shape $\mathbf{x}$ is reconstructed in Ω_r by applying the linear combination $\mathbf{H}_i$ for each slice i. Finally, the inverse of T_θ is used to transform the reconstructed volume from Ω_r to the image space Ω. Some examples of segmented volumes using such a principle are shown in figure (2.8). In a subsequent step, one may consider refining the results by locally optimizing the solution $\mathbf{x}$ on each slice i, using the sparse model's result as a prior such as [156].

2.5.6 Comparative Study with PCA & kernels for Segmentation

In order to demonstrate that sparse information models can efficiently be used for segmentation a methodology similar to the one considered for the interpolation will be studied.. For that purpose, it is assumed an expert (i.e. either a human expert, or an expert system such as the ones described in the literature) roughly initializes the rigid transformation and the key contours. When no user interaction is available, a preprocessing step, such as exhaustive search or coarse-to-fine search, could be developed. In the case of PCA [155], the segmentation problem is solved by minimizing the cost function resulting from the intensity-based likelihood of each pixel in the volumetric image:

$$\begin{aligned} E_{\text{seg}} = & \int_{\Omega} -\log\left(h_{\text{in}}(I(s))\right) \mathcal{H}(\mathbf{x}(T_\theta(s))) \, d\Omega \\ + & \int_{\Omega} -\log\left(h_{\text{out}}(I(s))\right) (1 - \mathcal{H}(\mathbf{x}(T_\theta(s)))) \; d\Omega, \end{aligned} \tag{2.18}$$

For the PCA segmentation, all the m slices of the volume are used, whereas the sparse information model only segments the K slices determined during the model construction (see equation (2.15)). Let us project the shape $\mathbf{x}$ onto the PCA basis $\{\mathbf{U}_q\}$ (see section (1.3.1)), after subtracting the mean shape $\overline{\mathbf{x}}$

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{q=1}^{N} x_q \mathbf{U}_q. \tag{2.19}$$

As in [155], equation (2.18) is minimized in the PCA's parametric space, where the shapes' distribution is modeled using kernels. The kernels are justified by the poor modeling of the samples distribution by a Gaussian. Using N kernels[1] noted K_n, the likelihood $p(\Lambda)$ of $\Lambda = \{x_q\}$ is

$$p(\Lambda) \propto \sum_{n=1}^{N} K_n(\Lambda).$$

Therefore, the energy to minimize is

$$\begin{aligned} E_{\text{seg}} = & \int_{\Omega} -\log\left(h_{\text{in}}(I(s))\right) \mathcal{H}(\mathbf{x}(T_\theta(s))) \, d\Omega \\ + & \int_{\Omega} -\log\left(h_{\text{out}}(I(s))\right) (1 - \mathcal{H}(\mathbf{x}(T_\theta(s)))) \; d\Omega - \log\left(\sum_{n=1}^{N} K_n(\Lambda)\right), \end{aligned} \tag{2.20}$$

[1] Only Gaussian kernels are considered in [155], but the method stands for any other type of kernels.

method	PCA (30 modes)	SIM	IOV
median symmetric diff.	26.41%	11.49%	5.56%
maximum symmetric diff.	36.84%	17.13%	7.83%
minimum symmetric diff.	16.68%	9.49%	2.96%

Tab. 2.2: Results table showing the average symmetric difference and maximum symmetric between hand-segmented livers and automatic segmentation with PCA and sparse information model (SIM). Also, is also given the Inter-Observer Variability (IOV) statistics.

Then, the derivative of E_{seg} according to a particular x_q is given by

$$\frac{E_{\text{seg}}}{\partial x_q} = \int_{\Omega} -\log\left(\frac{h_{\text{in}}(I(s))}{h_{\text{out}}(I(s))}\right)\delta_{\mathbf{x}}(T_\theta(s))\mathbf{U}_q ds - \frac{\sum_{n=1}^{N}\frac{\partial K_n(\Lambda)}{\partial x_q}}{\sum_{n=1}^{N}K_n(\Lambda)} \tag{2.21}$$

and the derivative with respect to the geometric transformation T_θ is

$$\frac{E_{\text{seg}}}{\partial \theta} = \int_{\Omega} -\log\left(\frac{h_{\text{in}}(I(s))}{h_{\text{out}}(I(s))}\right)\delta_{\mathbf{x}}(T_\theta(s))\nabla\mathbf{x}(T_\theta(s)).\frac{\partial T_\theta}{\partial \theta}(s)ds. \tag{2.22}$$

Table (2.2) summarizes the symmetric difference (see equation (2.14)) between ground truth and the segmented liver obtained using the sparse information model and the PCA-based method [155] (see figure (2.8)). Neighboring structures of similar intensities juxtapose the liver in a way that PCA estimates as a shape variation. On the contrary, the sparse model ignores the regions with low support, and reconstructs the information in these regions based on other visual clues elsewhere in the image. For comparison purposes, the inter-observer symmetric difference in table (2.2) indicates the symmetric difference between livers segmented by different experts using the same semi-automatic tool. Overall, when compared with [88], the results seem to demonstrate sparse information models outperform active shape models. Nevertheless, it must be underlined that the training and evaluation datasets are different. Furthermore, in [88], the shape model is built from smoothed surface meshes, while the training shapes used in this chapter are represented by distance functions (see equation (2.8)) and are not smoothed. However, as one suspects, sparse information models are sensitive to initialization. To quantify this, two different sparse segmentations were performed by segmenting by hand the key slices in the datasets, and comparing the reconstruction results with the ground truth. The difference in quality (symmetric difference with ground truth) between the different reconstructions ranges from 0.02% to 6.73%. Moreover, this variance is not correlated to the IOV (correlation coefficient of 0.47); otherwise stated, a volume with high inter-

observer variability may be segmented by the SIM in a way that is robust to initialization, and reciprocal may be true. Indeed, the IOV depends on the whole organ's structure while the SIM's quality only depends on the key slices. Furthermore, the maximum quality difference of 6.73% is below the maximum IOV symmetric difference (7.83% in table (2.2)).

2.5.7 Discussion on Segmentation from Sparse Information

In this section, we have introduced a novel family of dimensionality reduction techniques based on intelligent selection of key sub-elements with respect to reconstruction quality, image support and variability of these key sub-elements. It is demonstrated that sparse information models can be used to model shape variations, and can efficiently be integrated into a segmentation framework in the context of volumetric organ segmentation. We have applied this technique to the problem of liver segmentation in volumetric images with satisfactory results when compared to standard dimensionality reduction techniques based on linear projections and kernel distributions. On top of interpolation and segmentation quality, this method is also very fast since only the most important and most reliable information is processed for the reconstruction of the whole information. However, as noted in [88], a statistical shape model may not be sufficient to represent the exact shape of the liver ; in a post-processing step, a local optimization - using active contours for instance - may be necessary for better results. This local optimization would not be computed from sparse information. Further work will investigate the use of non-linear models for the interpolation function, as well as a subsequent refinement step that will locally adjust the reconstruction from the model to the actual image information by taking into account the confidence in the reconstruction. More advanced prior models using axial coronal and sagittal sparse information would be an interesting extension of our approach, as it would diminish the quality difference between two differently initialized segmentations. Last, but not least, the use of such methods for feature extraction, classification and content-based image indexing and retrieval is a natural extension on the application side.

Spatial volume segmentation refer to a challenging problem for dimensionality reduction. Similar concept can be created using 2D+time volumes which can be considered to form a surface. In the case of periodic motions, spatio-temporal segmentation consists of recovering the solution for all frames at the same time. In such a scenario one can imagine that the support is varying among the images of the cycle. Cardiac segmentation is an example where such an assumption is valid due to the periodicity of the heart beating.

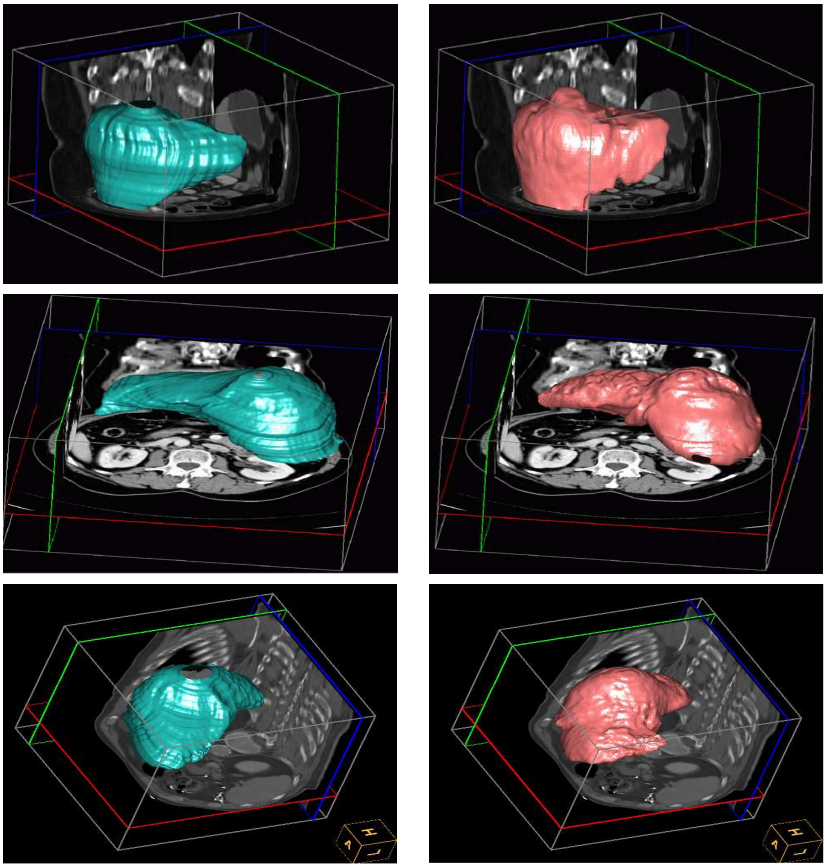


Fig. 2.8: Comparison of liver segmentation obtained by SIM (left column) and expert segmentation (right column).

2.6 Semi-automatic Segmentation of Time Sequence Images using Optimal Information

2.6.1 Medical Motivation

Cardiac ultrasound is a common diagnostic tool for a variety of pathologies. The echocardiographic equipment is cheaper than other imaging devices, and non invasive as the images are acquired by processing the echo of ultrasound waves reflected by the tissues. The most common protocol consists in studying phenotypic and functional abnormalities of the heart's four chambers, valves and aorta to detect and quantify pathologies such as poor function, regurgitation, mitral stenosis, ... An accurate diagnosis demands various measures based on the the difference between the volume of the heart in the systolic and diastolic phase as well as the ventricular wall motion,

which requires the ventricles to be previously segmented [138].

In a broader context, this section presents a possible application of sparse information models to segmentation in time-sequences. The study case consists of 2D left ventricle contours in four chambers view ultrasonic images (see figure (2.9)), but the same concept may be extended to higher dimensions for a 3D tracking such as PET , CT or tagged MR sequences. The only constraint is to work on sequences that are registrable along one particular axis, such as time. The main object of this application is twofold: first, to demonstrate the robustness of sparse information models with respect to noise in the sparse information, and second to compare sparse information model results with interobserver variability.

2.6.2 Prior Art on Segmentation in Echocardiographic Sequences

Shape tracking in time sequences is performed either by taking the object's location and shape at one time step as a prior for the following time step, or by processing the optical flow between consecutive images [93], or by introducing a dynamic model from prior knowledge. A fourth solution consists in modeling the uncertainties of the object's segmentation at one time step, and integrating this model in a sequential Bayesian procedure for a more robust segmentation at the following time step (see section (1.4) in chapter 1), and chapter 4). Geiser and Wilson and their group [170] have developed an automated contour detection scheme in echocardiography by detecting shape segments with arc template filters. Their useful method combines a high-level model of the anatomy with low-level image features; however, the overall method is unstable: the scheme consists of cascading steps, each one of which could fail and cause the whole system to break down. Therefore, a simpler model to represent the contours is often preferred. A common way to represent the shape statistics is to use Cootes' [38] active shape model [18] and active appearance model [17]. A fifth solution consists in representing the time axis as a geometrical axis, and modeling the full sequence of shapes [19].

Nevertheless, these models suffer the same limitation whether they are used for object's tracking or segmentation: since they do not take the local image reliability into account, an image region that consistently contains low information (no edge, neighboring structures, high statistical variance...) may attract the solution toward erroneous results. On the contrary, sparse information models consists in reconstructing the shape at every time steps from the contours at selected key time steps.

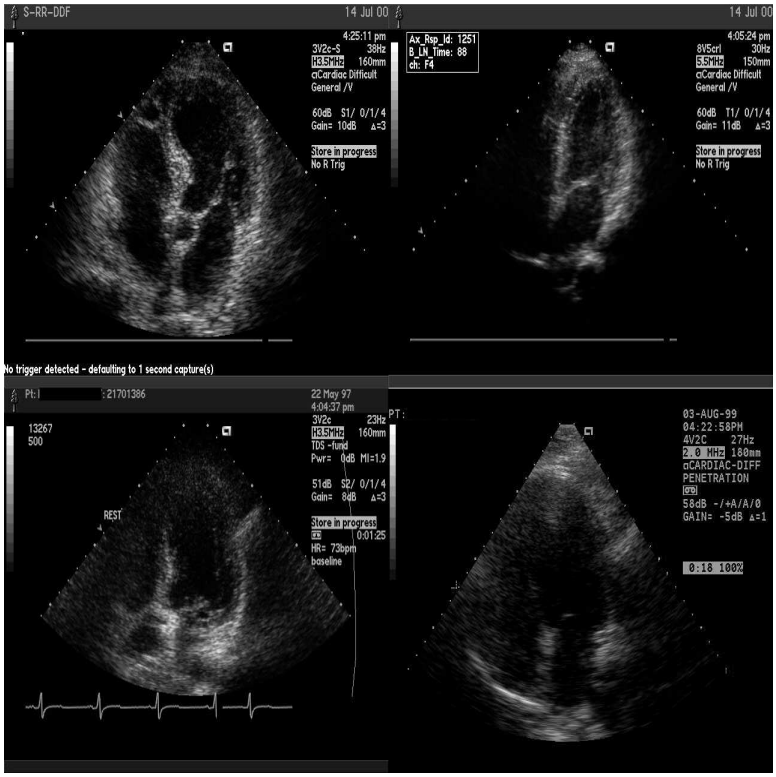


Fig. 2.9: Four chambers views in ultrasound.

2.6.3 Sparse Information Models in Time Sequence Images

For time sequence images, the natural discretization axis $\mathbf{v}_0$ is the time axis. This means the sub-elements of the sequence are defined as the intersections of the shape with the time axis at time t. Let C_t represent the tth contour in the sequence represented though an implicit model ψ_{C_t}, such as in equation (2.8). This implicit function ψ_{C_t} is rewritten as a column vector $\mathbf{Y}$ by aligning the distance value of each pixel in lexicographic order. As in [18], a PCA is performed to represent the contour C_t in a parametric way, after registering the contours with a scaling-translation transformation $\mathcal{T}$; using n modes of variations $\{\mathbf{U}_i\}_{i\in[|1,n|]}$ (n=10 in the experiments below), let the column vector $\mathbf{x}_t = [x_1, x_2, ...x_n]^T$ denote the contour parameters' vector:

$$\mathbf{Y} = \overline{\mathbf{Y}} + \sum_{q=1}^{n} x_q \mathbf{U}_q, \tag{2.23}$$

where $\overline{\mathbf{Y}}$ stands for the mean contour in the reference space.

The error between the ground truth and PCA with 10 modes of variation (99% of total variation) corresponds to the inter-observer variability (0.1mm reported in [100]). The scaling-translation transformation $\mathcal{T}$ registers the first contour of each sequence together and is applied to all successive contours in the sequence. The key time steps $\mathcal{B} = \{t_1, t_2, ...t_K\}$ are determined by minimizing the sum of the three energies (image support E_{sup}, reconstruction quality E_{int} and contour variability E_{var}) after registering the sequences together in time with the retrospective ECG (ElectroCardioGram) signal to a reference interval that is discretized into T time steps. Several interpolating function were tried and generalized linear interpolation has outperformed the others (median distance between ground truth and reconstruction with generalized linear: 0.19mm, linear interpolation: 0.21mm and cubic splines: 0.22mm).

Unlike in section (2.5), the contours are represented implicitly and projected onto the PCA space (equation (2.23)). The reconstruction quality is computed with the Mahalanobis distance $\|.\|$ between reconstructed contours and ground truth in the PCA feature space:

$$E_{\text{int}}(\mathcal{B}, \mathbf{H}) = \sum_{p=1}^{P} \sum_{i=1}^{T} \left\| \mathbf{H}_i \left[\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K} \right]^T - \mathbf{x}_{t_i} \right\|^2 . \tag{2.24}$$

This formula differs from equation (2.9) since the distance is computed in the parametric PCA space. However, the optimization of $\mathbf{H}$, given a particular set of $t_1, ..., t_K$, is performed the same way, using a QR-decomposition on the over-constrained system of equations.

The image support energy is computed from the image gradient norm $|\nabla I_{t_k,p}|$ along the contour, for the p-th sequence in the training set, at time t_k. For a set of P echocardiographic sequences registered in time

$$E_{\text{sup}}(\mathcal{B}) = \sum_{p=1}^{P} \sum_{k=1}^{K} \int_{C_{t_k}} \frac{1}{1 + |\nabla I_{t_k,p}|^2}. \tag{2.25}$$

Such a choice is motivated by the fact that the ventricular tissue is not visible all the time and the LV is neighboring structures which also are filled in with blood, therefore a statistical separation between the intensities of the LV and the background is rather challenging. Noting the parallel with equation (1.2)) for geodesic active contours when $g(x) = \frac{1}{1+x^2}$, and deriving equation (2.25) the same way, one obtains the following derivative:

$$\frac{\partial E_{\text{sup}}(\mathcal{B})}{\partial C_{t_k}} = \sum_{p=1}^{P} \frac{-1}{1 + |\nabla I_{t_k,p}|^2} \text{div} \left(\frac{\nabla \psi_{C_{t_k}}}{|\nabla \psi_{C_{t_k}}|} \right) \tag{2.26}$$

Finally, the contours' sensitivity to small variations of the time index is computed with the magnitude of the derivative of $\mathbf{x}_t$ at $\{t_1, t_2, ...t_k\}$:

$$E_{\text{var}}(\mathcal{B}) = \sum_{p=1}^{P} \sum_{k=1}^{K} \left| \frac{\partial \mathbf{x}_t^p}{\partial t}(t_k) \right|^2 . \tag{2.27}$$

This expression is motivated by two factors. First, the contours selected as basis should not be sensitive to small errors in the interpolation of the $\{t_1, t_2, ...t_k\}$, therefore, their derivative with respect to t should be minimal. Furthermore, in the particular case of diagnostic echocardiography, the images are obtained by interpolating the echo sent back by the structures over time, see equation (2.1). Therefore, when the contours (i.d. the left ventricle endocardium) have minimal variation with respect to time, the integration of the images gives edges with higher contrasts, so better defined contours.

The total energy is defined as the sum of E_{sup}, E_{int} and E_{var}, as in equation (2.7). The derivatives of E_{int} and E_{var} with respect to the key time steps are immediate; however, in a cardiac ultrasound, some images may be of lower quality than others. Thus, a gradient descent following equation (2.26) may lead to local minimums. For this reason, the sum of E_{sup}, E_{int} and E_{var} is minimized using the simplex method [130].

$$SIC = TP\log\left(\frac{RSS}{TP}\right) + K\log(TP). \tag{2.28}$$

The residuals are defined as the integral of the Euclidean distance $\mathcal{D}$ between the reconstructed contours and the ground truth:

$$RSS = \sum_{p=1}^{P} \sum_{k=1}^{K} \int_{C_{t_k}} \mathcal{D}(\mathbf{x}_{t_k}^p, \mathbf{H}_k \left[\mathbf{x}_{t_1}, ...\mathbf{x}_{t_K}\right]^T).$$

It measures how the normalized residual sum of squares increases with respect to the number of key contours, therefore the minimum is preferred.

2.6.4 Reconstruction Scheme & Results

Ultrasound images are difficult to segment individually because of the speckle noise, whereas to consider the whole image sequence allows motion estimation to approximate a 2D contour in re-

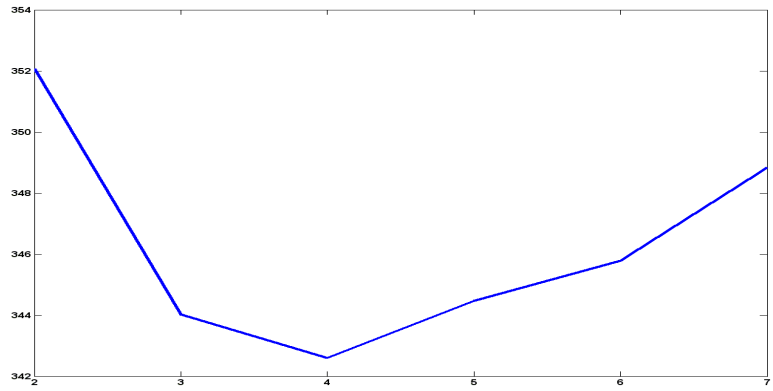


Fig. 2.10: Schwarz information criterion (equation (2.28)) for ultrasound sparse segmentation with respect to the number of key time steps.

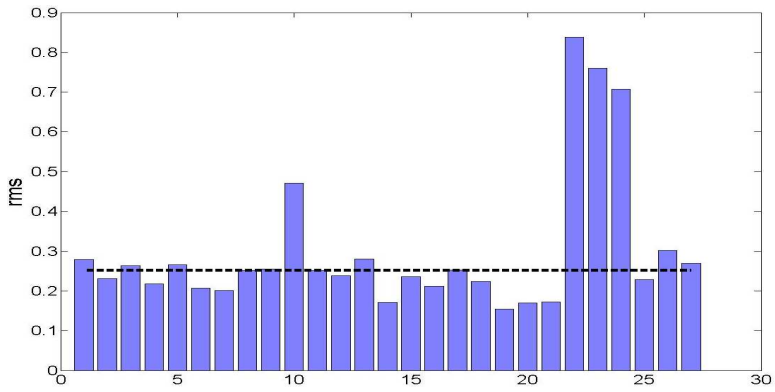


Fig. 2.11: Breakdown of the rms distance for each one of the examplars used in sparse segmentation for ultrasound sequences and median. Cycles 22, 23 and 24 refer to the ungated sequence.

lation with the others. However, either a proper initialization (close to solution) is required for the contours, or a statistical model is to be provided as a regularization constraint for the interpolation of the contours between consecutive frames. The sparse information model provides such a statistical framework so that the segmentation of the full sequence is reduced to the segmentation of few contours. With 4 key contours given using the ground truth, the root mean square (rms) error average distance between the reconstructed contours and the ground truth over 27 different heart cycles in a leave-one-out basis is 5.34 pixels (or 0.25mm average distance) (see figure (2.11)-(2.12)); therefore, it is reasonable to assume that the whole segmentation problem is reduced to the segmentation of the key contours. Segmentation of the left ventricle in ultrasound images in a well investigated field of study (see section (2.6.2)), and is beyond the scope of this section. A semi-automatic method such as [100] could be used to detect the key contours.

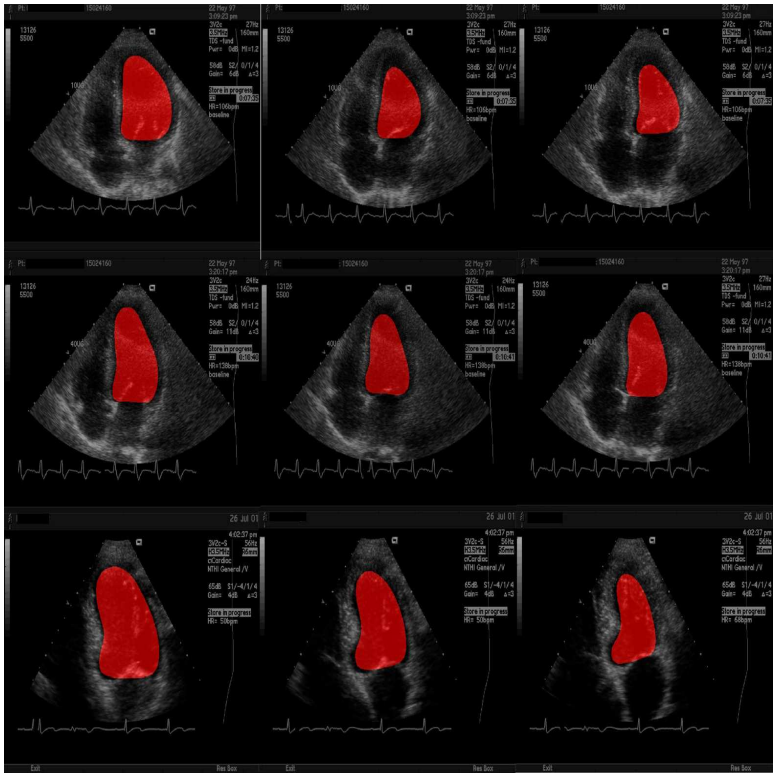


Fig. 2.12: Sparse segmentation results overlayed on ultrasound images sequence.

As segmentation of the whole sequence is reduced to segmentation or initialization on few images, one may suspect a high sensibility to initialization. However, as illustrated in figure (2.13), the mean distance between reconstructed contours and ground truth does not increase exponentially with respect to the mean distance between input contours and ground truth. On the contrary, the correlation stays linear as the amount of error between input contours and ground truth increases. For instance, for input contours with a mean distance of 0.20mm to ground truth, the mean distance over the whole interpolated sequence is 0.24mm to ground truth. Figure (2.14) describes the cumulative distribution of distances between 2 clinicians' contour; the segmentation of 70% of clinicians are consistent up to 0.2mm. The median distance error of 0.15mm (see figure (2.14)) on the key contours produces a mean error of 0.2mm for the whole sequence, which is still lower than 30% of clinicians. This, obviously, does not mean that the automatic method presented in this section "beats" 30% of clinicians, a claim that does not make any sense. It simply means that 30% of contours drawn by clinicians have a larger distance. The last comment one can make about this application is that inhomogeneities in the image support is only taken along the time axis. However, as one sees in figure (2.12), at a given time, the image gradient is not homogeneous

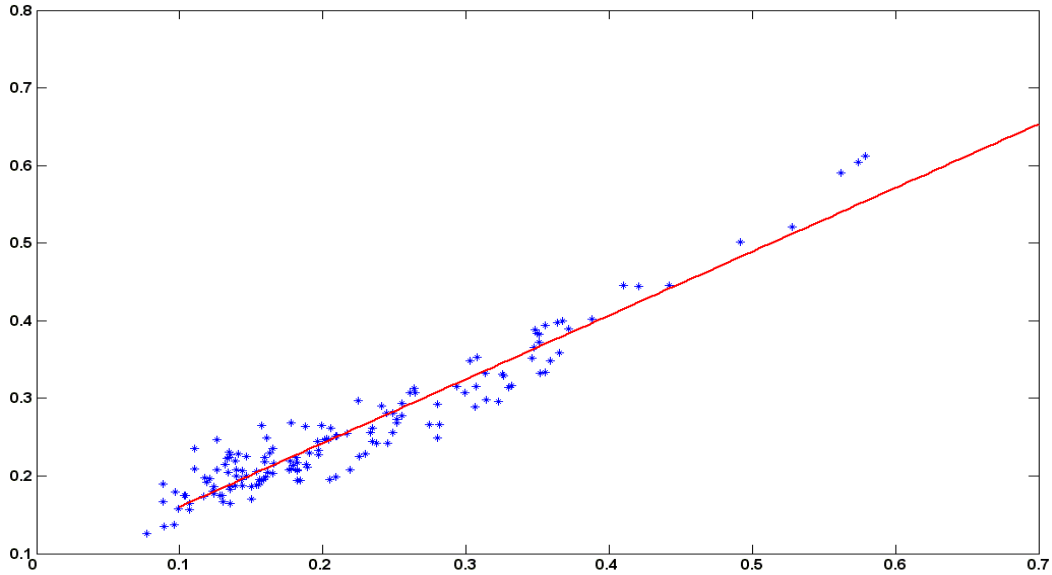


Fig. 2.13: Mean distance between ground truth and interpolated sequence using the sparse information model with respect to the mean distance between ground truth and the initialized contours.

(the pixel intensity gradient is higher on the interventricular septum than on the posterior left ventricular wall, in part because of the papillary muscles). Thus, the method presented above would benefit from a generalization by discretizing the information along the time axis and dividing the 2D images into regions of different support. Both applications considered up to now, refer to well defined problems of recovering smooth surfaces with sufficient data support. The last application considered to evaluate the applicability of the method refers to a problem where the data support is really partial.

2.7 Surface Reconstruction using Sparse Measurement

The aim of this application is to address laser face reconstruction using the 3D RMA dataset benchmark [14] that contains measurement errors and a high data variance. The dataset consists of laser pointed distances between the light source and the subject's face acquired by rows (see figure (2.15)). The image is defined as a volume I whose pixels' value is equal to the distance of these pixels to the nearest laser point. In a common approach to surface reconstruction, the surface to reconstruct is determined by minimizing the sum of the pixels' value on the surface, optionally using PCA-based models [108] or surface meshes to interpolate. On the contrary, Sparse Model is

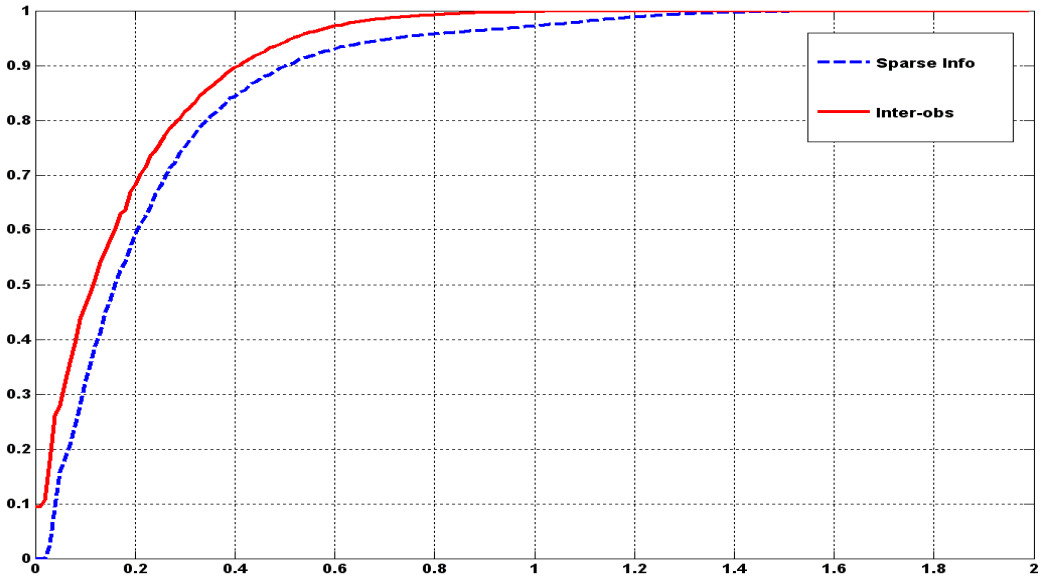


Fig. 2.14: Cumulative distribution of distances between clinicians' contours, and between contours segmented with SIM and ground truth. Curve extracted from Fig. (6) in [100].

used to determine which sub-elements are critical to the reconstruction, and from which the whole surface may be interpolated.

2.7.1 Prior Art on Surface Reconstruction

Surface fitting and surface reconstruction from range data is a well studied field of computer graphics and mainly used for face identification, reverse engineering, CAD, urbanism/architecture or even plastic surgery. Surface extraction from range data is derived from two steps: first the points cloud must be integrated into a common representation, and surfaces are then extracted from that representation. On the one hand, some authors suggest to progressively construct a mesh from the points [162, 184, 169] and then form the surface either by directly connecting the meshes together or by grouping them by regions first [174]. These techniques are sensitive to outliers since any smoothing operations are to be processed on the mesh itself, or on the distance function obtained from the mesh. Others [92, 47, 90] prefer to first integrate the range data into a common representation using a signed distance function, before using the marching cubes algorithm to extract the surface. In [92] Hoppe et al. estimate the local surface tangent plane for each point and then estimate the global surface using a minimum spanning tree of the points. In [47], a signed distance function is estimated from the range data and the surface is defined as the zero-level set of

this function. In [47], this distance function is estimated on a regular 3D grid which facilitates the extraction of the surface. In the following, as far as the estimation of the distance function is concerned, an approach similar to [47] is taken and simplified for the particular case of face reconstruction (e.g. no space carving is needed here). Then, the sparse information model is applied on the distance function to reduce the problem's complexity.

2.7.2 Model Construction

Following the technique presented in [47], we integrate the range data into a common representation using a distance function estimated with a 3D grid. Each laser pointed range defines a 3D point; the points set is projected in a reference space, where the 3D grid is defined, using a geometric transformation T_θ. Then, to each point on the grid is associated the minimal distance to the laser points. To keep the issue induced by the different orientations of the head, the geometric transformation T_θ is defined as a homothety-translation so that all the range volumes $\mathbf{x}$ have the same dimensions in the reference space Ω_r with N rows of M points. The missing points are replaced by the same background range value d_{back}. After projecting the laser points into the reference space, the distance map ψ is estimated on the 3D grid.

Since the acquisition process is sequential (row by row), it is natural to divide the surface $\mathbf{x}$ into sub-elements $\mathbf{x}_i$ by projecting $\mathbf{x}$ onto the rows' axis. A sub-element $\mathbf{x}_i$ is a 2D curve defined as the intersection of the face with the particular row plane i; this 2D curve is represented by a distance map ψ_i (see figure (2.17)).

In the context of surface reconstruction, the interpolation quality measure E_{int} is simply the sum over the N rows of the distance between the reconstructed surface $\phi_n\left(\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_K}\right)$ and the observation ψ_n:

$$E_{\text{int}} = \sum_{n=1}^{N} \psi_n \left(T_\theta^{-1}\left(\phi_n\left(\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_K}\right)\right)\right). \tag{2.29}$$

Since ψ_n is a Euclidean distance function, the distance between the reconstructed surface and the observation is simply the value of ψ_n at the reconstructed surface location.

Given that some rows contain more background points than others, E_{sup} is chosen so as to be

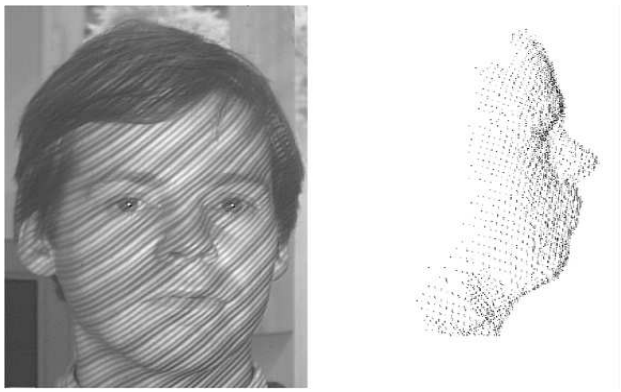


Fig. 2.15: A stripped image of a face and its 3D reconstruction from profile. Courtesy of Charles Beumier, Université Libre de Bruxelles.

proportional to the number of background points in the row:

$$E_{\text{sup}} = \sum_{k=1}^{K} \sum_{j=1}^{M} \delta(\psi_k \left(T_\theta^{-1} \left(\mathbf{x}_{t_k}(j)\right)\right) - d_{\text{back}}). \tag{2.30}$$

Finally, the parameters' variability E_{var} refers to the sensibility of the curves $\bar{\mathbf{x}}_k$ to infinitesimal changes in k, and the sensibility of the reconstructed surface to infinitesimal changes in $\bar{\mathbf{x}}_k$ (see equation (2.12)):

$$E_{\text{var}} = \sum_{k=1}^{K} \sum_{j=1}^{M} \left\| \frac{\partial \mathbf{x}_{t_k}(j)}{\partial k} \right\|^2 + \mathbf{H}_i^T \mathbf{H}_i. \tag{2.31}$$

Several interpolation operators have been tested over the dataset (splines, linear and polynomial interpolation) and, once again, the best results are achieved using generalized linear interpolation: each curve $\mathbf{x}_i$ is a linear combination $\mathbf{H}_i$ of all the key sub-elements $\{\mathbf{x}_{t_k}, k \in [|1, N|]\}$.

Once all these problem-related variables are defined, the sparse problem consists in determining the K row indices and the generalized linear operator ϕ that optimize the three criteria introduced above, equation (2.29)-(2.30)-(2.31). Since all the criteria have a discrete expression, it is reasonable to pre-compute the terms that do not depend on the interpolation function $\mathbf{H}_i$ and perform an exhaustive search. Similar to the segmentation case, an MDL and the calculus of variations with respect to the unknown parameters is used to determine the optimal sparse model. The best description length is achieved with 5% of the rows (see figure (2.17).

2.7.3 Validation

Using the sparse model, the reconstruction problem is simplified: one needs to determine the geometric transformation T_θ between the model space and the range data volume, and the 2D curves $\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_K}$ at key locations. To determine T_θ, one minimizes the distance values taken along the estimated key curves $\mathcal{B} = \mathbf{x}_{t_1}...\mathbf{x}_{t_K}$ in the range volume after geometric transformation T_θ:

$$E_{rec}(\mathcal{B}, T_\theta) = \sum_{k=1}^{K} \int_{\mathbf{x}_{t_k}} \psi\left(T_\theta^{-1}(p)\right) \quad dp. \tag{2.32}$$

To this end, once the range data has been acquired, we apply a fast distance transform [175]. Then, the reconstruction's cost function (equation (2.32)) is defined in this artificial volumetric data and consists of the Chamfer [25] transform of the model. Gradient descent is used to determine the optimal parameters of the transformation. A leave-one-out procedure over the 720 faces provides twice better results than Eigenfaces [186] and linear interpolation, see table (2.3) and figure (2.16). These results demonstrate the robustness of sparse information models to outliers and salt and pepper noise.

method	Eigenfaces	Linear interp.	SIM
average dist.	4.53 %	60.79%	2.09%
maximum dist.	97.73 %	100%	95.11%

Tab. 2.3: Results table showing the average distance difference and maximum distance between acquired and reconstructed faces with Eigenfaces, linear interpolation and Sparse Information Model (SIM) as percentage of the maximum range.

2.8 Conclusion and Discussion

In this chapter, a novel family of dimensionality and complexity reduction techniques are introduced. The solution (segmentation, surface) is discretized along one of its dimension $\mathbf{v}_0$, and key elements along that direction are selected with respect to their image support, their low variability when their index $\mathbf{v}_0$ is slightly changed, and the final reconstruction of the whole solution. Only the most reliable information in the image is used, and the rest of the solution is inferred from the model and the sparse observation. The resulting process is more efficient than standard segmentation/reconstruction since most of the workload is concentrated on the critical points, but also more robust, since the interpolated volume or surface is consistent with the prior knowledge statistics. In the study case of liver segmentation, it is proved that Sparse Information Models have a better

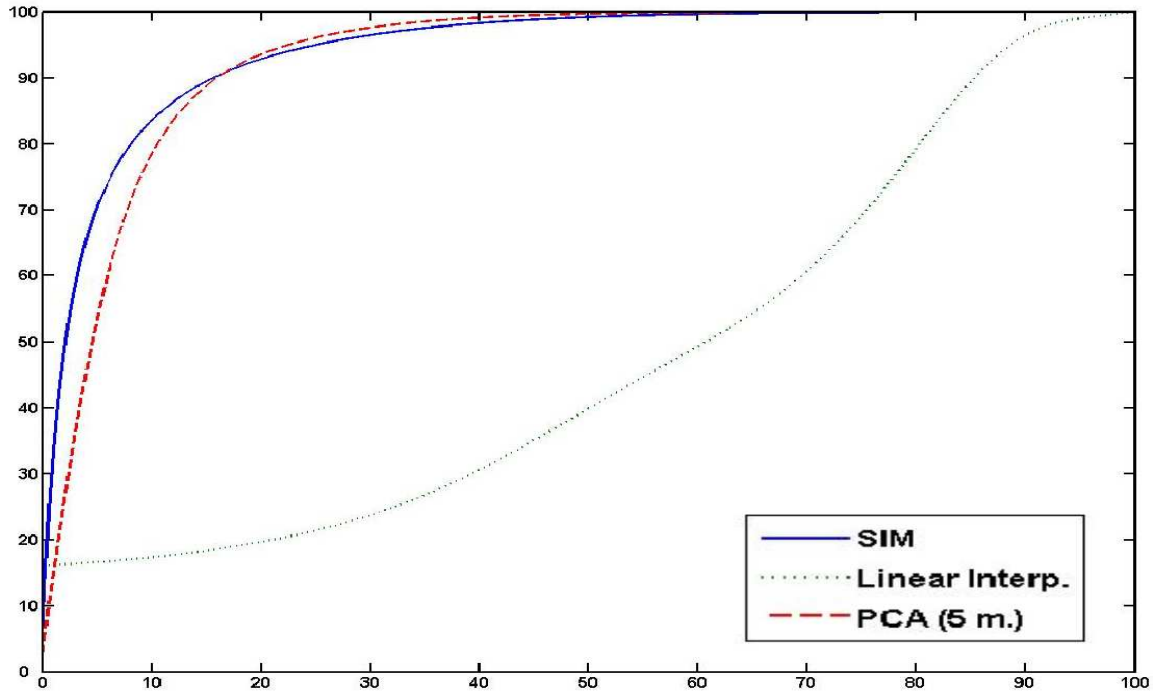
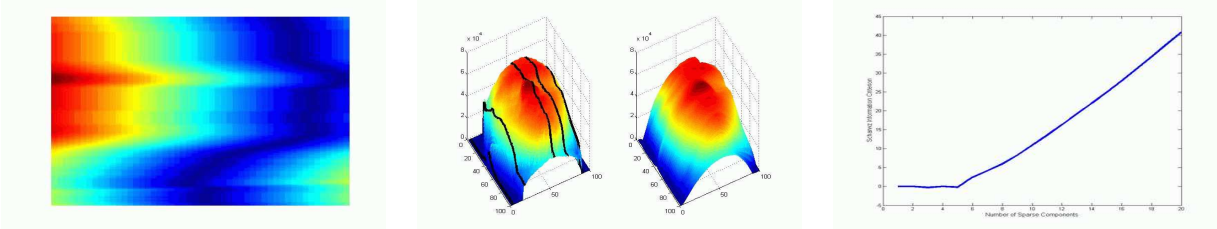


Fig. 2.16: Cumulative histograms of reconstructed surface distances with respect to ground truth. The X axis denotes the distance between reconstructed face and ground truth in percentage of the maximum distance, and the Y axis denotes the percentage of points whose distance is below that value.

dimensionality reduction power than linear decomposition techniques such as PCA. The volumetric surface is efficiently represented by few sparse contours. Moreover, because sparse information models are less sensitive to local minima in low image supported regions, they lead to a better segmentation. The study case of left ventricle segmentation in echocardiography proves that sparse information models perform well even for heavy pathologies and high dynamics variance. Finally, the third study case, surface reconstruction from laser pointed distances, proves the effectiveness of this method for cases with high measurement errors.

However, this method remains static, in the sense that the model is learned once and for all. It is unable to capture new dynamics. The following chapter, chapter 3, introduces a way to model the shape variation statistics using an autoregression scheme, and to adapt this model as new observations are introduced. Thus, the resulting framework is able to capture both new shape variations and changes of dynamics.

Despite the efforts put on generalizing this method as much as possible, certain constrains remain. The major one is in the deterministic procedure of assigning image support and choosing the key elements. Instead of associating a measure w_i to each sub-element $\mathbf{x}_i$, one could assign



(a) (b) (c)

Fig. 2.17: (a) Signed distance function in the 3D grid, colors representing the distance between the points in the 3D grid and the input laser points; (b) the 3D distance function is estimated using the sparse information model, colors representing the distance between the surface and the laser source; (c) Schwarz Bayesian Criterion with respect to the number of key rows.

a probability for $\mathbf{x}_i$ to be selected as measurement. The main advantage of this method is that prior knowledge specific to the particular image of interest could be easily added in the sampling process. The second advantage of sampling information instead of selecting which information to use from training is robustness. Elements of lower interest with respect to the training data may actually be of strong interest for a particular case if this case is an outlier. By randomly sampling the key elements, one randomly allows some of these low interest points to be selected, and makes the sparse information method much more robust. This would mean that, instead of reconstructing the rest of the information from sub-elements using an interpolation function, one would predict the statistical distribution of the total information $\mathbf{x}$ given a statistical analysis of some of its elements. This aspect is studied in more details in chapter 4 in the context of geometric tracking with particle filters.

The second field of study concerns the repartition of the sub-elements. In the three applications above, the sub-elements are chosen along a particular axis, either geometrical or temporal. However, nothing prevents a different approach like selecting partial elements such as surface patches instead of 2D contours in the case of liver segmentation, image region of high gradients in echocardiography, or points of interest instead of rows in the case of surface reconstruction. This would require to develop a more flexible method to register partial elements together. Once again, a statistical analysis of the partial elements distribution in the image would favor the partial elements that best respond to both the three criterias developed in section (2.4) and to measurements of image support in the image of interest.

Chapter 3

Motion Models for Non-Stationary Dynamics via Autoregressive Models

Abstract – *This chapter intends to present the use of autoregressive (or AR for short) models for segmentation. The main contribution is the use of on-the-fly adapting AR models for tracking; however, stationary models are also introduced in the context of a regular (i.e. stationary) cardiac ultrasound sequence, and may even be used to solve static problems such as volumetric segmentation. Tracking highly deforming structures in space and time arises in numerous applications in computer vision. Static models are often referred to as linear combinations of a mean model and modes of variations learned from training examples. In dynamic modeling, the shape is represented as a function of shapes at previous time steps. In this chapter, we introduce a novel technique that uses the spatial and the temporal information on the object deformation. Tracking is reformulated as a high order time series prediction mechanism that adapts itself on-line to the newest results. Samples (toward dimensionality reduction) are represented in an orthogonal basis and are introduced in an auto-regressive model that is determined through an optimization process in appropriate metric spaces. Toward capturing evolving deformations as well as cases that have not been part of the learning stage, a process that updates on-line both the orthogonal basis decomposition as well as the parameters of the autoregressive model is proposed. Promising experimental results in tracking explicit shapes in a video sequence that could be used to impose prior knowledge in segmentation are presented.*

3.1 Introduction

Motion perception is a fundamental task of biological vision, with motion estimation and tracking being the most popular and well-addressed applications. To this end, given a sequence of images, one would like to recover the 2D temporal position of objects of particular interest. These applications often serve as input to high-level vision tasks, like 3D reconstruction, etc. The solutions currently available for this problem consists either in learning a stationary[1] regression law from prior knowledge, or representing the object to track by a random variable and applying a Bayesian framework. However, the Bayesian framework assumes the transition probability density function is known from prior knowledge, and the stationary regression does not handle dynamic changes. Furthermore, in the Bayesian framework, the random variable is defined in a vector space that remains unchanged, which means the model of the object to track is stationary. If the object to track changes its shape into something that has not been learned before, the model cannot represent it. This chapter addresses these questions and proposes a method to dynamically update the regression law and the feature space, and to integrate this dynamic model into a tracking framework based on level-sets. Then, we demonstrate that sequences of shape variations across time and space may be modeled in their totality using stationary autoregressive laws. Two applications are considered: a purely vision-based application that is tracking shape deformations of 2D humans projections, and one that refers to tracking shape deformations of the Left Ventricle in ultrasound sequences.

3.1.1 Prior Art

Tracking non-rigid objects is a task that has gained particular attention in computational vision. Starting from the pioneering formulation of the snake model [104] several attempts to address tracking through the deformation of contours can be found in the literature either model-free [97] or model-based [38]. Level set methods [137] is an alternative technique [136] to track moving interfaces through model-free [143] or model-based [41] methods with the advantage of being implicit, intrinsic and parameter-free. Such methods are able to capture important non-linear deformations.

Introducing prior knowledge within visual perception has been an on-going effort to a number of vision tasks, like segmentation, motion analysis, 3D reconstruction, etc. Tracking was a domain

[1] Stationary means whose law does not change with time.

that has benefited from such an effort, in particular when dealing with objects and structures of limited variation in space and time. To this end, different approaches were considered either based on snakes [46], active shape and appearance models [40, 39], level sets [202] etc. Such approaches model spatial variation of the structure of interest in a probabilistic fashion. Then, during the inference process a constraint on recovering shapes that belong to the learned family is imposed.

Temporal models like Kalman snakes [179], Markov chains, hidden Markov models or multiple hypotheses trackers [97, 182] address tracking in a difference dimension. Constraints/models are imposed in the temporal evolution of the target and prediction mechanisms are used to perform tracking. Shape tracking with autoregressive dynamic models is a step forward in this direction, with different shape spaces being investigated. In [129], a first-order model is used to track cardiac cycles echocardiographic sequences while in [117], Fourier descriptors are used to describe shapes, and a LDM tracks their evolution on time. Tracking articulated structures is problem well suited for autoregressive models and therefore in [2] a method based on a linear dynamic model was proposed. The main limitation of such models refers to their *time-invariant* nature. Temporal models as well as shape representations have been learned from previous sequences, used within tracking and not updated. Consequently, either a complex heuristic is developed to mix models, or Markov fields are introduced for multimodality.

To address this issue, adaptive dynamic models such as adaptive Kalman filter [65] integrate a Kalman filter that estimates additive noise properties (mean and variance) to an AR model. Except for the noise properties, no adaption is made for the signal itself if its own properties change (e.g. new properties cannot be captured by the current feature space) or if the system dynamism (the regression) is modified. On the other hand, in the context of classification and learning, adaptive feature spaces such as adaptive principal component analysis (PCA) [84][115] were developed to take into account the newest results to estimate the feature space. Nevertheless, even if the feature space is adapted to the most recent exemplars, in the context of segmentation and tracking, one also needs an adaptive predictive model to relate segmentation results across time, and to adapt the prediction scheme if changes in the dynamic system occur.

In this chapter, we would like to address these limitations and determine a predictive model that is incrementally adapted to changes both in the system dynamics (and not only the noise properties) and in the feature space. Sparse information models presented in chapter 2 are not sufficient because they need to be trained on whole sequences, and sequences that contain the dynamisms one needs to track. In other words, sparse information models do not sustain unlearned shape

deformations. For that purpose, we propose an on-line technique for tracking based on higher order autoregressive models. Such a technique is based on dimensionality reduction of the parameter space using an orthogonal decomposition of the training set. Then, a linear autoregressive model is built in such space capable of predicting current states from the prior ones. Such a model and its feature space (the orthogonal decomposition of shapes) are updated on-line using new evidence. To this end, a proper geometric distance is used in a robust framework to optimize the parameters of the model.

Section (3.2) introduces the notations and theory of dynamic models, and presents how this model may be update on-the-fly with the study case of a walker silhouette. Section (3.2.4) presents the experiment and results on the silhouette tracking problem. Two other applications are exposed in section (3.3) to demonstrate that stationary AR frameworks may be used to model contours deformation in image sequences or in static volumetric segmentation.

3.2 Autoregressive Models and non-Stationary Dynamic Modeling of Deformations

Time series models are very popular in a number of domains like signal processing. They may be either linear or non-linear; however, for simpler derivations, only the linear case is exposed. Let us assume a set of temporal observations $\mathbf{Y} = \{\mathbf{Y}_t; t \in [0, T]\}$, where each observation $\mathbf{Y}_t \in \Omega$ is a column vector of the N-dimension observation space Ω. In the context of this chapter, these observations are shape representations in the image space. Linear autoregressive models - of order k - consist of expressing the current observation, as a combination of previous samples perturbed by some noise model, as written in equation (3.1)

$$\mathbf{Y}_t = \mathbf{A} \left[\mathbf{Y}_{t-1}^T \; \mathbf{Y}_{t-2}^T \; ... \; \mathbf{Y}_{t-k}^T\right]^T + \mathbf{w} + \eta(\mathbf{0}, \Sigma) \tag{3.1}$$

with N-by-kN matrix $\mathbf{A}$ called the *prediction matrix* and $\eta(\mathbf{0}, \Sigma)$ being the *noise model* vector. A constant vector $\mathbf{w}$ is introduced so that η is assumed to be zero-meaned. For any matrix $\mathbf{M}$, $\mathbf{M}^T$ denotes the matrix transpose. In the most general case, one assumes that the input variable $\mathbf{Y}$ is defined in a high-dimensional space, and therefore some dimensionality reduction is to be performed. Without loss of generality, we assume a set of either linear or non-linear operators $g_i(); i \in [1, m]$, referred to collectively as g, that, when applied to the input variable $\mathbf{Y}$, form a new

basis of observations, as written in equation (3.2).

$$\mathbf{X} = (g_1(\mathbf{Y}), g_2(\mathbf{Y}), ..., g_m(\mathbf{Y})) \tag{3.2}$$

$\mathbf{X}$ is therefore a new random variable. We further assume that such operators are invertible, or otherwise stated, from a feature vector $\mathbf{X}$ one recovers the original observation $\mathbf{Y}$. In that case one restates the autoregressive model in a lower dimensional space, and equation (3.1) is rewritten in equation (3.3).

$$\mathbf{X}_t = \mathbf{A}_g \left[\mathbf{X}_{t-1}^T \mathbf{X}_{t-2}^T ... \mathbf{X}_{t-k}^T\right]^T + \mathbf{w}_g + \eta_g(\mathbf{0}, \Sigma), \tag{3.3}$$

where $.^T$ stands for the vector transpose. $\mathbf{A}_g$ differs from $\mathbf{A}$ since $\mathbf{A}_g$ depends on the decomposition g. The same reasoning holds for η_g; in that case, the noise is defined in the feature space and does not account for the noise introduced by the projection with g. The estimation of $\mathbf{A}_g$ is performed from a set of training examples and robust regression. Let us assume that $T >> k$ observations are available. Once such observations have gone through dimensionality reduction with the operators g, we obtain an over-constrained linear system written in equation (3.4).

$$\begin{aligned} \mathbf{X}_T \leftarrow \hat{\mathbf{X}}_T =& \mathbf{A}_g \left[\mathbf{X}_{T-1}^T \mathbf{X}_{T-2}^T ... \mathbf{X}_{T-k}^T\right]^T + \mathbf{w}_g \\ &\vdots \\ \mathbf{X}_k \leftarrow \hat{\mathbf{X}}_k =& \mathbf{A}_g \left[\mathbf{X}_{k-1}^T \mathbf{X}_{k-2}^T ... \mathbf{X}_0^T\right]^T + \mathbf{w}_g \end{aligned} \tag{3.4}$$

The unknown parameters of such over-constrained system can be determined through a robust least square minimization, equation (3.5),

$$(\mathbf{A}_g, \mathbf{w}_g) = \text{argmin}_{\mathbf{A}_g, \mathbf{w}_g} \left\{ \sum_{t=k}^{T} \rho \left(\mathbf{X}_t, \mathbf{A}_g \left[\mathbf{X}_{t-1}^T \mathbf{X}_{t-2}^T ... \mathbf{X}_{t-k}^T\right]^T + \mathbf{w}_g \right) \right\} \tag{3.5}$$

where ρ is an error metric, in the observation space. Since the reduced space is potentially highly non-uniform (in the case of PCA, this means the variations along one mode are larger than the others), performing the minimization in the observation space greatly reduces the prediction error. In order to simplify the notation, we assume that ρ is the L-2 norm (see equation (3.11)) in the observation space, which transforms equation (3.5) into a linear least square optimization. The number of constraints used in such a procedure is determined off-line using the Schwartz Bayesian criterion (SBC). Following the approach introduced in [131], the least squares estimate of $\mathbf{A}_g$ is estimated for all orders from 1 to $k_{\max}$ and the SBC is computed from the residual errors at each

order. For a particular k, let us note

$$\tilde{\mathbf{A}}_g = [\mathbf{w}_g \mathbf{A}_g],$$

and

$$\tilde{\mathbf{X}}_t = \begin{pmatrix} 1 \\ \mathbf{X}_{t-1} \\ \mathbf{X}_{t-2} \\ \dots \\ \mathbf{X}_{t-k} \end{pmatrix} \tag{3.6}$$

After noting the moment matrices as $\Gamma = \sum_{t=k}^{T} \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^T$ and $\Gamma_x = \sum_{t=k}^{T} \mathbf{X}_t \tilde{\mathbf{X}}_t^T$, the least squares estimate of $\tilde{\mathbf{A}}_g$ is computed as $\tilde{\mathbf{A}}_g = \Gamma_x \Gamma^{-1}$. With this estimate, the residual sum of squared errors (RSS) and the SBC are computed:

$$\text{SBC}(k) = (T-k)m\log\left(\frac{\text{RSS}}{(T-k)m}\right) + (km^2 + m)\log((T-k)m).$$

3.2.1 Online Adaption of Predictive Model

Let us now assume that new observations are present. Two cases need to be distinguished; in the first case, the dynamics are time-invariant and the new observations are directly integrated into equation 3.5. In the second case, the dynamics are time-variant and more importance is given to the latest obtained observations. Once the prediction matrix has been estimated, new observations are introduced in the system toward decreasing the prediction error. To this end, one would like to find the lowest potential of

$$E(\mathbf{A}_g) = \min_{\mathbf{A}_g, \mathbf{w}_g} \left\{ \sum_{t=k}^{T} \rho(\mathbf{X}_t, \mathbf{A}_g \left[\mathbf{X}_{t-1}^T \ \mathbf{X}_{t-2}^T \ \dots \ \mathbf{X}_{t-k}^T\right]^T + \mathbf{w}_g) \right\} \tag{3.7}$$

With ρ defined as the L-2 norm (see equation (3.11)), equation (3.7) is solved with an iterative least squares estimation, with iterative Gauss-Newton method being the most popular technique to address such optimization. In [12], the result is obtained by dividing the sum of squares into blocks, solving the problem for the first block and using this result as initialization once the following block is added to the previous block. Unlike the method presented here, [12] solved the Gauss-Newton iterations using extended Kalman filter for nonlinear measures $E(\mathbf{A}, \mathbf{w}, \Sigma)$. Experiments have shown that few (a couple of dozen) Gauss-Newton iterations are required to achieve far better

results than a simple time-invariant dynamic model.

For non-linear time processes, the local approximation of $(\mathbf{A}_{T+1}, \mathbf{w}_{T+1}, \Sigma_{T+1})$ may not well correspond to the state transition in a very early time step. For that reason, *exponential forgetting* is introduced;

$$E(\mathbf{A}_{g,T+1}, \mathbf{w}_{g,T+1}) = \min_{\mathbf{A}_g, \mathbf{w}_{g,T+1}} \left\{ \sum_{t=k}^{T} w_{T-t} \rho(\mathbf{X}_t, \mathbf{A}_{g,T+1} [\mathbf{X}_{t-1}\ \mathbf{X}_{t-2}\ ...\ \mathbf{X}_{t-k}] + \mathbf{w}_{g,T+1}) \right\} \tag{3.8}$$

with exponential weights $w_t = e^{-t/\tau}$, where τ is the *exponential forgetting* window size. The smaller τ the more reactive but also the more sensitive to noise is the non-stationary autoregressive model.

3.2.2 Feature Space

Before developing the feature space in details, let some notations be introduced straight off. To summarize the notations:

- the shapes are represented implicitly by level-set functions ψ_i (equation (3.9))
- the level-set functions are discretized and represented by observation vectors $\mathbf{Y}_i$
- PCA dimensionality reduction is applied to the observation vectors, after affine transformation, and the resulting parameter vectors (PCA+affine parameters) are noted $\mathbf{X}_i$ (equation (3.10))
- a prediction mechanism $\mathbf{A}_g, \mathbf{w}_g$ is estimated to sequentially predict the future parameter vectors. This prediction mechanism depends on the PCA projection and the transformation since it is optimized in the observation space (equation (3.8)).

When no topology constrains are given, implicit methods are popular shape representations. Let us consider a number of training examples to track $\mathbf{C} = \{\mathcal{C}_i, i \in [1, n]\}$. In [146] a distance

Fig. 3.1: Registered training examples used for initial principal components analysis.

transform representation ψ_i was considered for a given shape $\mathcal{C}_i$, equation (3.9),

$$\forall p \in \Omega, \quad \psi_i(p) = \begin{cases} 0 & , p \in \mathcal{C}_i \\ +\mathcal{D}(p) > 0 & , p \in \Gamma \\ -\mathcal{D}(p) < 0 & , p \in \overline{\Gamma} \end{cases} \tag{3.9}$$

where Ω defines the image domain partitioned in 2 (inside the contour Γ and outside $\overline{\Gamma}$), and $\mathcal{D}(p)$ the Euclidean distance between point $p \in \Omega$ and the exemplar's contour $\mathcal{C}_i$. We call abusively "shape" the distance function ψ. The resulting function is then represented by a column vector $\mathbf{Y}_i$ of dimension n after discreetizing the region of interest in the image domain Ω with n control points. Global registration between shapes is now performed by determining the affine transformation $\mathcal{A}$ that minimizes the integral of squared difference between the alignment shape's distance function and the reference distance function [142].

Since the implicit representation of shapes increases dimensionality, n is usually too large for the computation of $\mathbf{A}$ in equation (3.5); therefore principal component analysis (PCA) is applied for an efficient dimensionality reduction and a more tractable solution (equation (3.3)). PCA has already been introduced in section 1.3.1; a brief outline is given in this chapter to introduce the online updating scheme of section 3.2.3. PCA refers to a linear transformation of variables that retains - for a given number m of operators $g_i(); \ i \in [1, m]$ - the largest amount of variation within the training data. Without loss of generality, a zero mean assumption is considered for the $\{\mathbf{Y}_i\}$ by estimating the mean vector $\overline{\mathbf{Y}}$ and subtracting it from the training samples $\{\mathbf{Y}_i\}$. The n-by-n covariance matrix $\mathbf{\Gamma} = \sum_{i=1}^{m} \mathbf{Y}_i \mathbf{Y}_i^T$ associated to the m training vectors $\mathbf{Y}_i$ is used for an Eigendecomposition. The n Eigenvectors $\mathbf{U}_q$ form an orthonormal basis onto which the vectors $\mathbf{Y}_i$ are projected. Only the m Eigenvectors associated to the highest Eigenvalues are kept, so that the operator g is defined by the affine transformation $\mathcal{A}$ and the projection from the n-dimension space to the m major Eigenvectors, and is invertible if one approximates the $n - m$ smallest Eigenvalues

by 0. The projected vector $g(\mathbf{Y}_i)$ is defined by the coefficients $\{x_q\}_{q=1..m}$ so that:

$$\mathbf{Y}_i = \mathcal{A}(\overline{\mathbf{Y}}) + \sum_{q=1}^{m} x_q \mathbf{U}_q. \tag{3.10}$$

Let us note $\mathbf{X}$ the feature vector $[\mathcal{A}, \Lambda]$ related to the shape $\mathbf{Y}$ in the observation space.

The affine transformation and the PCA decrease the problem dimensionality, but the overall feature space is highly non-uniform since it is composed by PCA parameters and affine transformation factors. In order to overcome such a limitation, we propose to use a metric defined on the original space (i.e. the observation space Ω) to recover the prediction mechanism in the reduced space (i.e. affine transformation from section (3.2.2) and linear factors from equation (3.10)). The simplest metric between two level-sets is the L_2 norm between the two distance functions that correspond to the observation $\mathbf{Y}_t$ and the prediction $\hat{\mathbf{Y}}_t$.

$$\rho(\mathbf{X}_t, \hat{\mathbf{X}}_t) = \int_{\Omega} \left(\psi_{\mathbf{Y}_t}(p) - \psi_{\hat{\mathbf{Y}}_t}(p)\right)^2 dp \tag{3.11}$$

refers to a well behaved distance between observations, and predictions and implicitly accounts for the range of parameters of the autoregressive model. This guarantees that the feature space and autoregressive models are optimum for the L_2 norm for the training set. However, in order to captures changes in shape or varying regression, an explicit on-the-fly update scheme is required.

Once new observations have been introduced to the process, the prediction matrix as well as the orthogonal basis are to be updated. Incremental principal component analysis can be used for the basis, while an exponential forgetting method is more suitable for the prediction matrix. However, in cases where the feature space changes too rapidly, the past basis may not suffice to predict a shape close enough to the solution for convergence. These cases are considered outliers for which either extra prior knowledge has to be integrated in the model, or else a model-free tracking may be more relevant.

3.2.3 Online Adaption of Feature Space

Incremental PCA[84, 115] consists of adding the latest observation to the PCA learning set. Thus, a new feature space is to be used to represent the basis that includes the mean observation and the principal modes of variation $\mathbf{X}$. Using these new variation modes, and the corrected state $\hat{\mathbf{X}}_t$,

the transition model is then updated and is used to predict the following state $\mathbf{X}_{t+1}$. The method presented in [84] can be summarized as follows: given a PCA state at time t-1, that is a mean $\overline{\mathbf{Y}}_{t-1}$, a set of eigenvectors $U_{t-1} = [u_i]$, and their corresponding eigenvalues $\mathbf{D}_{t-1} = \text{diag}\,(\lambda_1, \lambda_2, ...)$, given a new state $\mathbf{Y}_t$, the PCA is updated at time t starting by the mean:

$$\overline{\mathbf{Y}}_t = \frac{(t-1)\overline{\mathbf{Y}}_{t-1} + \mathbf{Y}_t}{t}.$$

One should note that the prediction mechanism provides direct registration between the new example and the basis. The eigenvector matrix is updated by adding the new vector's residual $\mathbf{h}$ and applying a rotation $\mathbf{R}$ on the former eigenbasis:

$$\mathbf{U}_t = \left[\mathbf{U}_{t-1} \frac{\mathbf{h}}{\|\mathbf{h}\|_2}\right] \mathbf{R}. \tag{3.12}$$

For a covariance matrix $\mathbf{\Gamma}_t$, the following relation is assumed $\mathbf{\Gamma}_t \mathbf{U}_t = \mathbf{D}_{t-1}\mathbf{U}_t$, which is solved analytically:

$$\mathbf{\Gamma}_t = \frac{t-1}{t}\mathbf{\Gamma}_{t-1} + \frac{t-1}{t^2}(\mathbf{Y}_t - \overline{\mathbf{Y}}_t)(\mathbf{Y}_t - \overline{\mathbf{Y}}_t)^T,$$

then, one can conclude that $(\mathbf{R}, \mathbf{D}_t)$, (equation (3.12)) is the solution of the eigenproblem

$$\mathbf{GR} = \mathbf{RD}_t,$$

where

$$\mathbf{G} = \frac{t-1}{t}\begin{bmatrix} \mathbf{D}_{t-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{t-1}{t^2}\begin{bmatrix} \mathbf{g}\mathbf{g}^T & \gamma\mathbf{g} \\ \gamma\mathbf{g}^T & \gamma^2 \end{bmatrix},$$

with $\gamma = \mathbf{h}^T\left(\mathbf{Y}_t - \overline{\mathbf{Y}}_t\right)$ and $\mathbf{g} = \mathbf{U}^T\left(\mathbf{Y}_t - \overline{\mathbf{Y}}_t\right)$. Such an iterative procedure updates both the prediction mechanism parameters as well as the PCA basis. These parameters are applied to describe the temporal evolution of the state vector.

The last question to be addressed is the frequency/sampling rate of such a time series. Several constraints are to be satisfied since on one hand, one would expect a model which is capable to express the long-term dynamics. On the other hand, the model should be accurate enough to capture local deformations. One should note that this is a critical component of the model since in the most general case one cannot expect the same frequencies in different sequences.

Linear autoregressive models learned from a training set with a certain temporal sampling rate (i.e. frames per second in video sequences, interslice spacing for volumetric images) are

Fig. 3.2: Vertical occlusion added to the original dataset

Fig. 3.3: Horizontal occlusion added to the original dataset

applicable to any sequence with another sampling rate through a simple conversion. Let $\mathbf{X}[t] = \mathbf{A}_g\mathbf{X}[t-1] + \mathbf{w}_g$ be the autoregressive model learned from prior knowledge (for the sake of simplicity, the example is a first order AR model). Let $\mathbf{X}'$ be a new sequence, and let α be the ratio between the training sequence's sampling rate and the new sequence's, so that $\mathbf{X}'[t] = \mathbf{X}[\alpha t]$. Then, the AR model for $\mathbf{X}'$ is obtained immediately with $\mathbf{X}'[t] = \mathbf{A}_g\mathbf{X}'[t - 1/\alpha] + \mathbf{w}_g$. When $1/\alpha$ is not an integer, the new sequence needs to be interpolated. In section 3.3, a model-guided interpolation procedure will be presented for the case of AR model, but until then, it is assumed that all exemplars have the same sampling rate.

3.2.4 *Multi-frame Segmentation and Tracking with Dynamic Models*

In order to evaluate the performance of such a model, we have considered multi-segmentation and tracking of highly deforming objects (walking pedestrians). Let $\mathbf{Y}_{1:t}$ be the sequence of shapes between time 1 and t that we are looking for, and $I_{1:t}$ the sequence of images. In a Bayesian framework, the general probability that one wants to maximize is $P\left(\mathbf{Y}_{1:t}|I_{1:t}\right)$, which can be expressed

prior method	(1) Stationary AR		(2) Global AR	(3) Adaptive AR	
(a) Original Dataset	correctly seg.	95.20 %	failed	correctly seg.	96.19 %
	under-seg.	2.76 %		under-seg.	2.32 %
	over-seg.	2.03 %		over-seg.	1.49 %
(b) Horizontal Occlusion	failed		failed	correctly seg.	96.66 %
				under-seg.	2.08 %
				over-seg.	1.26 %
(c) Vertical Occlusion	failed		failed	correctly seg.	96.04 %
				under-seg.	1.47 %
				over-seg.	2.48 %

Tab. 3.1: Percentage of correctly segmented, oversegmented and undersegmented pixels for diverse prior, same energy. Method (1) uses the stationary AR, learned from the first frames. Method (2) uses the stationary AR learned from the whole sequence. Method (3) uses the Adaptive AR described in this chapter. Dataset (a) is the original dataset. Dataset(b) is the original dataset with a horizontal occlusion, and dataset (c) presents a vertical occlusion.

in terms of $P\left(I_t|\mathbf{Y}_t\right)$, $P\left(\mathbf{Y}_{1:t-1}|I_{1:t-1}\right)$ and $P\left(\mathbf{Y}_t|\mathbf{Y}_{1:t-1}\right)$:

$$\begin{aligned} P\left(\mathbf{Y}_{1:t}|I_{1:t}\right) &= \frac{P\left(I_t|\left(\mathbf{Y}_{1:t}|I_{1:t-1}\right)\right)P\left(\mathbf{Y}_{1:t}|I_{1:t-1}\right)}{P\left(I_t|I_{1:t-1}\right)} \\ &= \frac{P\left(I_t|\left(\mathbf{Y}_{1:t}|I_{1:t-1}\right)\right)P\left(\mathbf{Y}_t|\mathbf{Y}_{1:t-1}|I_{1:t-1}\right)}{P\left(I_t|I_{1:t-1}\right)} \\ &\times \; P\left(\mathbf{Y}_{1:t-1}|I_{1:t-1}\right) \\ &\propto \; P\left(I_t|\mathbf{Y}_t\right)P\left(\mathbf{Y}_t|\mathbf{Y}_{t-1}\right)P\left(\mathbf{Y}_{1:t-1}|I_{1:t-1}\right) \end{aligned}$$

Noting E_{sup} the image energy to be minimized, $P\left(I_t|\mathbf{Y}_t\right)$ refers to the likelihood of the image given the contour: $\exp(-E_{\text{sup}})$. In the present experiment, E_{sup} refers to the Kullback-Leibler distance between the pixels intensity inside h_{in} (resp. outside h_{out}) the contour $\mathbf{Y}_t$ given the a priori intensity distribution inside p_{in} (resp. outside p_{out}):

$$\begin{aligned} E_{\text{sup}}(\mathbf{Y}_t) = &\int h_{\text{in}}(s,\mathbf{Y}_t)\log\left(\frac{h_{\text{in}}(s,\mathbf{Y}_t)}{p_{\text{in}}(s)}\right)\,ds \\ &+\int h_{\text{out}}(s,\mathbf{Y}_t)\log\left(\frac{h_{\text{out}}(s,\mathbf{Y}_t)}{p_{\text{out}}(s)}\right)\,ds. \end{aligned} \tag{3.13}$$

$P\left(\mathbf{Y}_{1:t-1}|I_{1:t-1}\right)$ is the likelihood of the previous contours, and $P\left(\mathbf{Y}_t|\mathbf{Y}_{t-1}\right)$ is the likelihood of $\mathbf{Y}_t$ given the prediction from the AR model. Since the contours are represented in the PCA

space, $P(\mathbf{Y}_t|\mathbf{Y}_{t-1})$ refers to the mahalanobis distance between the contour $\mathbf{Y}_t$ and the contour predicted using the AR model. With a first-order Markovian assumption, the probability actually maximized is:

$$P(\mathbf{Y}_t|\mathbf{Y}_{1:t-1}, I_t) = \frac{P(I_t|(\mathbf{Y}_t|\mathbf{Y}_{1:t-1}))\, P(\mathbf{Y}_t|\mathbf{Y}_{1:t-1})}{P(I_t)}. \tag{3.14}$$

Equation (3.14) is translated in energy terms by taking the negative logarithm, and the energy to minimized with respect to the PCA parameters is $E(\mathbf{Y}_t) = E_{\text{sup}}(\mathbf{Y}_t) + E_{\text{predict}}(\mathbf{Y}_t)$, where E_{obs} has been previously described and E_{predict} is the Mahalanobis distance between the AR-prediction $\hat{\mathbf{Y}}_t$ and the current contour $\mathbf{Y}_t$:

$$E_{\text{predict}} = \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t\right)^T \mathbf{\Gamma}^{-1} \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t\right). \tag{3.15}$$

One minimizes this cost function using a gradient descend approach (the equations are given in the appendix B). The main innovation of our approach is the dynamic behavior of the model both in the basis and the prediction space. Therefore, it is natural first to compare the performance of such a model with respect to a stationary model as presented in [42].

Comparison with stationary AR models

For comparison purposes we use the same dataset with and without digital occlusions, which shows a man silhouette walking and then running, and test different priors for level-sets [136] evolving according to the same energy. This energy corresponds to the sum of a data-driven term (a histogram-based Chan & Vese functional [32]), and a term associated to the shape prior provided by the dynamic model. The silhouette moves in front of a uniform light colored background so that the data support for the segmentation is optimal which allows to evaluate the performance of the prediction mechanism (one is not interested in the segmentation quality so much as in the dynamic system itself). Furthermore, to properly test the model adaptability, the system dynamic do change; therefore the silhouette walks then runs. The validation is performed by counting the overlap of the ground truth and the tracking result in a region-of-interest centered on the target silhouette. When the predicted contour totally escapes the region-of-interest, it remains in the background for the remaining of the sequence; in this case, the tracking is declared ”failed”. As the results shown in Table (3.1) demonstrate, the adaptive AR model systematically performs better than the stationary AR model, when both are initialized with the first 18 frames (in a 58 frames video sequence). The stationary model learns the dynamic of the walking pace, but is unable to sustain the dynamic

changes when the silhouette starts to run. On the contrary, the adaptive model learns the new dynamics on-line, and is able to make optimal predictions. The learning a stationary AR from the whole sequence (walking and running) drives the model toward an average behavior that does not correspond neither to the walking or the running state. Consequently, it does not stick to the local temporal dynamics. This proves the non-stationary AR model is well adapted to this problem.

In the same way, incremental PCA improves the prediction by 6 % in non-occlusion cases compared to a non-updated PCA (correctly segmented pixels without incremental PCA: 90.23%, correctly segmented pixels with incremental PCA: 95.98 %); but updated PCA alone is not sufficient, and the dynamic model must be updated as well (correctly segmented pixels with incremental PCA but with stationary dynamic model: 91.14%).

Robustness to Occlusion

The main drawback one expects from a locally adaptive method is the potential accumulation of errors. To test that, we introduce occlusions, as shown in figures (3.2-3.3) (one horizontal occlusion that covers one third of the character during 20 frames, and then one vertical of the same width as the character) of the background mean color, and run the tracking scheme with stationary and adaptive priors. Once again, the results demonstrate that the adaptive model sustains these occlusions. Nevertheless, for larger occlusions, errors accumulate and the tracking is lost. Once the performance of the method with respect to highly deforming objects has been proved, the next open issue is the usefulness of such a model to address segmentation. Without loss of generality we consider a stationary case.

3.3 AR Models for Stationary Modeling

Left Ventricle Segmentation in Echocardiography Four Chambers View with Stationary Autoregressive Models

In this section, we propose to use the autoregressive framework to model the dependence of contours in a sequence at one time step with respect to the others. Using the autoregressive model, the segmentation scheme may be developed either sequentially by using the contours at previous frames to predict the following ones, or by using the contours at various frames and interpolating

so as to minimize the variance and bias for the estimation of the remaining contours. The dataset considered in this section is the same as in chapter 2 section (2.6): 27 heart cycles in ultrasound, all ECG-gated but one.

Segmentation with Stationary Autoregressive Models

A sequential segmentation would use the results at previous frames and the AR model to predict the future frames by maximizing the likelihood presented in equation (3.14). The image support energy E_{obs} used in the case of ultrasound images is the weighted sum of a region-based energy and the image gradient energy:

$$\begin{aligned} E_{\text{obs}}(\mathbf{Y}) =& \alpha \int_{\Omega_{\mathbf{Y}}} -\log\left(h_{\text{in}}(I(\mathbf{x}))\right) H_{\mathbf{Y}}(\mathbf{x}) - \log\left(h_{\text{out}}(I(\mathbf{x}))\right)\left(1 - H_{\mathbf{Y}}(\mathbf{x})\right)\, d\mathbf{x} \\ &+ (1-\alpha)\int_{\Omega_{\mathbf{Y}}} \frac{1}{1+|\nabla I(\mathbf{x})|} d\mathbf{x}. \end{aligned} \tag{3.16}$$

To measure the performance of the autoregressive model, one assumes the contours at each step are corrected using the image and equation (3.16), and the error measure is defined as the difference between the ground truth and the predicted (not yet corrected) contours. The regression order is obtained using Schwarz Bayesian criterion (see figure (3.4)). figure (3.5) presents the boxplot[2] results (mean distance between ground truth and predicted contours using the AR model) with respect to the regression order. Several conclusions may be drawn from figure (3.5): first, the order that gives the best results is 2, second, the AR model fails to predict outliers independently of the order, and third non-outliers are consistently well predicted with a 0.2mm mean distance to ground truth. As illustrated in chapter 2 figure (2.14), the mean distance between contours drawn by two different clinicians has 30% chance to be 0.2mm or more, so the results for non-outliers is acceptable for an actual application.

However, the previous results are obtained by optimizing each contour with respect to the image. This method may be subject to accumulation of errors. To measure this, an experiment consists in not optimizing the contours and measuring the quality of prediction with respect to ground truth. In this situation, the errors accumulate without correction and the mean distance be-

[2] The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

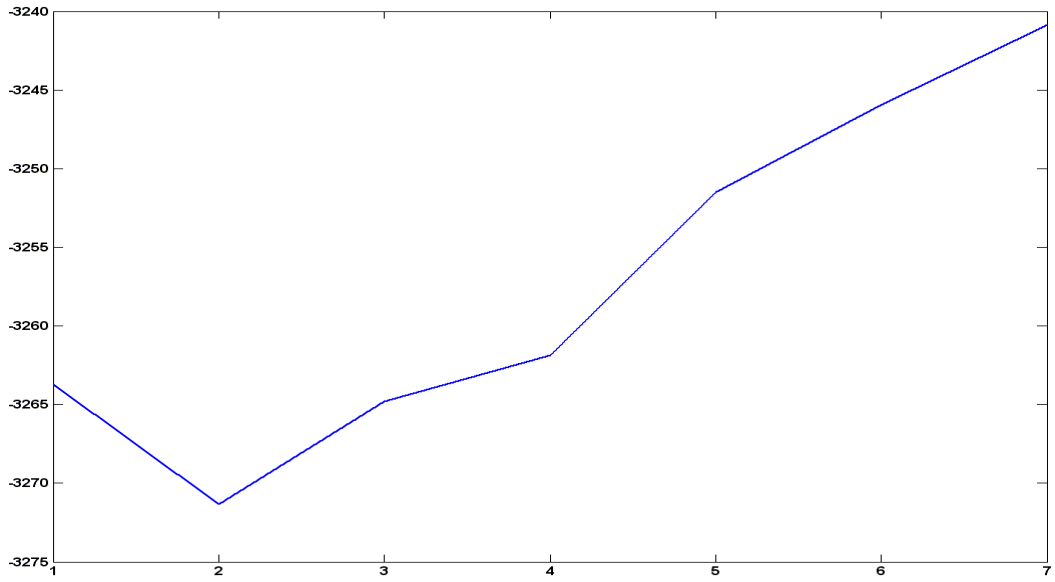


Fig. 3.4: Schwarz Bayesian criterion for autoregressive sequential segmentation of ultrasound sequences.

tween the predicted contours and ground truth is 0.4mm, corresponding to only 10% of clinicians' contours in figure (2.14). The main limitation of this method lies in the assumption of modeling temporal dynamics using a specific order in a sequential fashion. In this particular case, like in many other applications, the full sequence is available at once. Therefore, the autoregression can be used to model the full sequence; that is the object of the following section.

Interpolation with Stationary Autoregressive Models

The accumulation of error described above comes from the fact the initialized contours are all located in the beginning of the sequence. However, the order-k autoregression model indicates that k initial contours suffice to constrain the whole sequence; actually, these k contours do not need to be the first ones in the sequence. Given k contours anywhere in the sequence, the rest of the contours may be interpolated using the autoregression. This technique has been known for long in audio restoration [80], and consists in determining in the least square sense the sequence that best fits the autoregressive model and the input constrains. Let us assume the contours $\mathbf{Y}$s are represented in a parametric space such as PCA (see equation (3.2)) and let us note the regression law for the input signal $\mathbf{X}_t$ as follows:

$$\mathbf{X}_t = \mathbf{A}_1\mathbf{X}_{t-1} + \mathbf{A}_2\mathbf{X}_{t-2} + ...\mathbf{A}_k\mathbf{X}_{t-k} + \epsilon_t, \tag{3.17}$$

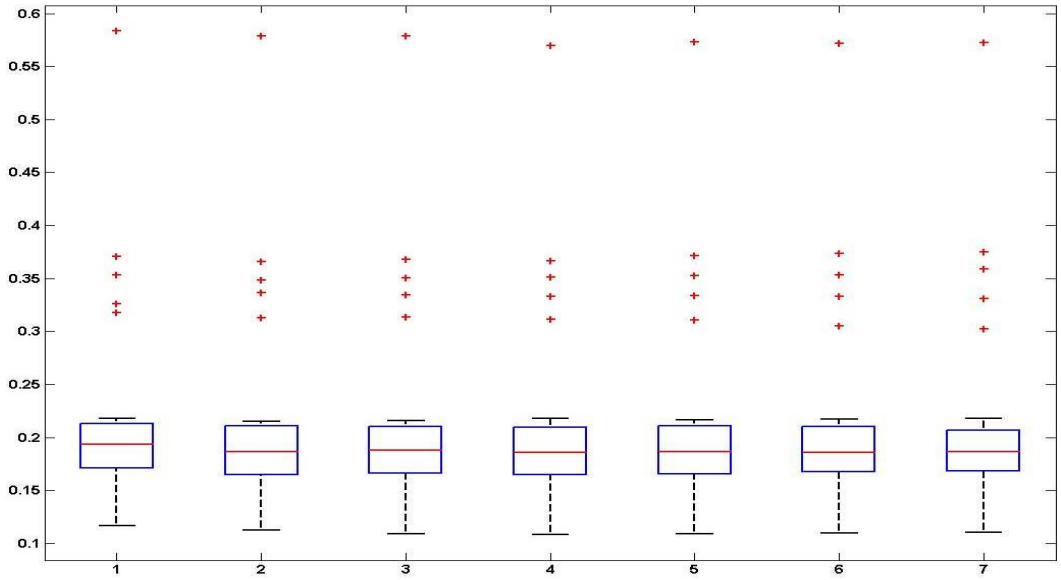


Fig. 3.5: Boxplot results of sequential segmentation of ultrasound images with respect to the regression order (1 to 7).

where ϵ_t denotes the noise between the predicted and actual signal. When each matrix $\mathbf{A}_i$ is diagonal, the parameters of $\mathbf{X}_t$ are totally independent. In the case of left ventricle contours in echocardiography, the hypothesis of independence between the PCA coefficient is justified by the low cross-correlation values compared to the auto-correlations (see figure (3.7)). For the sake of simplicity, $\mathbf{X}_t$ is then assumed to be mono-dimensional, and equation (3.17) may be rewritten as

$$\epsilon = \mathbf{A}\mathbf{X},$$

where ϵ is a column vector whose tth element is $\mathbf{X}_t - (\mathbf{A}_1\mathbf{X}_{t-1} + \mathbf{A}_2\mathbf{X}_{t-2} + ...\mathbf{A}_k\mathbf{X}_{t-k})$, $\mathbf{X} = [\mathbf{X}_T\mathbf{X}_{T-1}...\mathbf{X}_1]$, and $\mathbf{A}$ is composed by the diagonal elements of the $\mathbf{A}_i$s. The contours sequence that best follows the autoregression law minimizes $\epsilon^T\epsilon$. When only k contours are initialized in the sequence, all but these k $\mathbf{X}_i$ are zero vectors; let s be the set of indices of initialized contours. Let $\mathbf{X}^s$ be the input signal composed by the elements of $\mathbf{X}$ that are initialized, and $\mathbf{X}^o$ the vector composed by the elements to be determined. Then, one rewrites the prediction model as follows

$$\epsilon = \mathbf{A}\left(\mathbf{U}\mathbf{X}^s + \mathbf{K}\mathbf{X}^o\right),$$

where $\mathbf{U}$ (resp. $\mathbf{K}$) is a rearrangement matrix of identity matrix $\mathbf{I}$ composed by the k columns of $\mathbf{I}$ whose indices are (resp. are not) in s. The minimum of $\epsilon^T\epsilon$ is obtained by deriving this expression

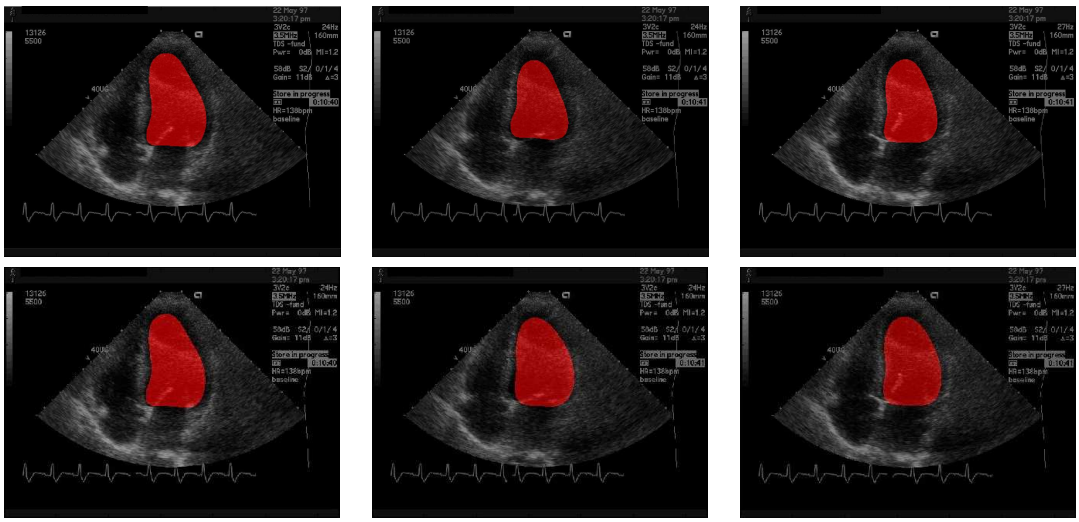


Fig. 3.6: Ultrasound sequence segmented sequentially using an autoregressive model, top row with correction, bottom row without correction.

with respect to X^o:

$$\frac{\partial \epsilon^T \epsilon}{\partial \mathbf{X}^o} = 2\epsilon^T \frac{\partial \epsilon}{\partial \mathbf{X}^o} \tag{3.18}$$

$$= 2\left(\mathbf{AUX}^s + \mathbf{AKX}^o\right)^T \mathbf{AK} \tag{3.19}$$

$$= 0 \Longleftrightarrow \mathbf{X}^o = -\left(\mathbf{AKK}^T\mathbf{A}^T\right)^{-1}\mathbf{K}^T\mathbf{A}^T\mathbf{AUX}^s \tag{3.20}$$

As illustrated by the boxplot results in figure (3.9), a mean distance of 0.2mm between interpolated contours and ground truth is achieved with a second-order AR model and 4 input contours. 4 input contours are certainly too much to ask for a practical application, furthermore, the issue of outliers remain (see figure (3.9)). However, similar results are obtained with a generalized linear interpolation (see section (2.6.3)) using also 4 contours (see figure (3.8)). The same idea can be considered to segment 3D volumes which can be described slices+space sequences. In such a context, we will study the development of an algorithm for liver segmentation.

3.3.1 Sequential Segmentation of volumes using Stationary AR Models

In this application, 3D volumes are discretized along the axial planes; the axial planes are considered as a temporal sequence of images. Volumetric segmentation is considered sequentially on the 2D slices, starting from the ones with strong data support toward the ones of limited support.

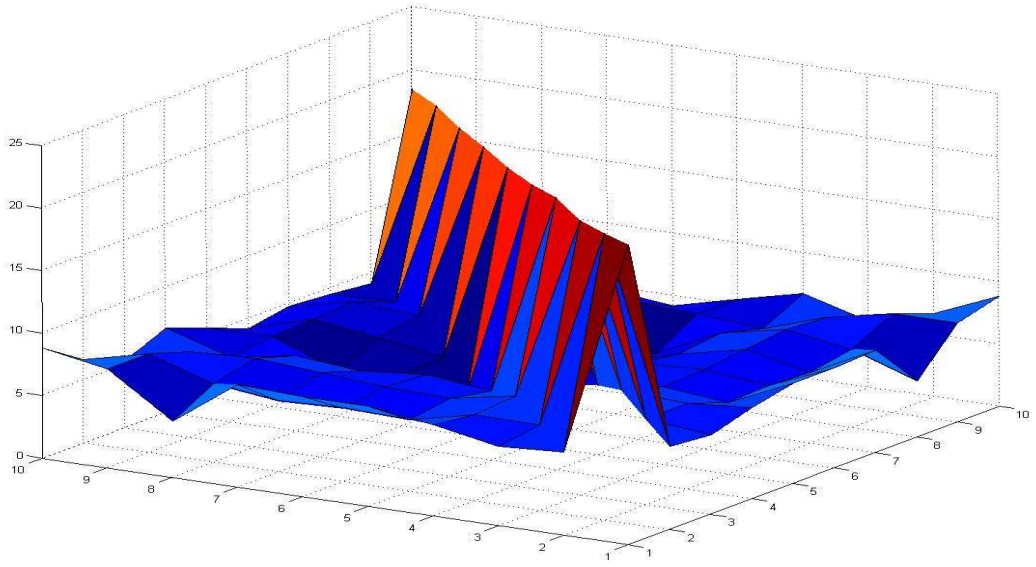


Fig. 3.7: Cross-correlation of PCA coefficients in a logarithmic scale for left ventricle contours in echocardiography. The low correlation between the different PCA coefficients justifies an independent AR model for each component.

Just as in the previous section, section (3.3), the volume may be segmented in a successive fashion by linking the segmentation maps through a locally adaptive autoregressive prediction mechanism - that is learned through training - where confidence in the data from prior slices constrains the results, or else, by interpolating segmentation maps at sparse locations using the AR model.

In this case, the 2D contours $\mathbf{Y}$ are represented in a parametric fashion, using PCA on distance maps for instance, and the contour feature vectors $\mathbf{X}_t$ are predicted using equation (3.3). In order to determine the AR parameters, we have considered 34 manually segmented volumes which have been registered to the same pose. Then, the geometric distance between the prediction and the observation is minimized. The optimal order for the model is 2. In the case of liver segmentation, the image support is measured using a region-based metric E_{obs} using the a priori knowledge of the pixels intensity distribution inside h_{in} and outside h_{out} the liver.

$$E_{\text{obs}}(\mathbf{Y}) = \int_{\Omega_{\mathbf{Y}}} -\log\left(h_{\text{in}}(I(\mathbf{x}))\right) H_{\mathbf{Y}}(\mathbf{x}) - \log\left(h_{\text{out}}(I(\mathbf{x}))\right)(1 - H_{\mathbf{Y}}(\mathbf{x}))\ d\mathbf{x}, \tag{3.21}$$

where $H_{\mathbf{Y}}$ is the Heavyside function equal to 1 inside the contour $\mathbf{Y}$ and 0 outside. The segmentation scheme follows the one presented in section (3.2.4) and equation (3.14). E_{predict} is the Mahalanobis distance between the AR-prediction and the current contour $\mathbf{Y}_t$. The minimization

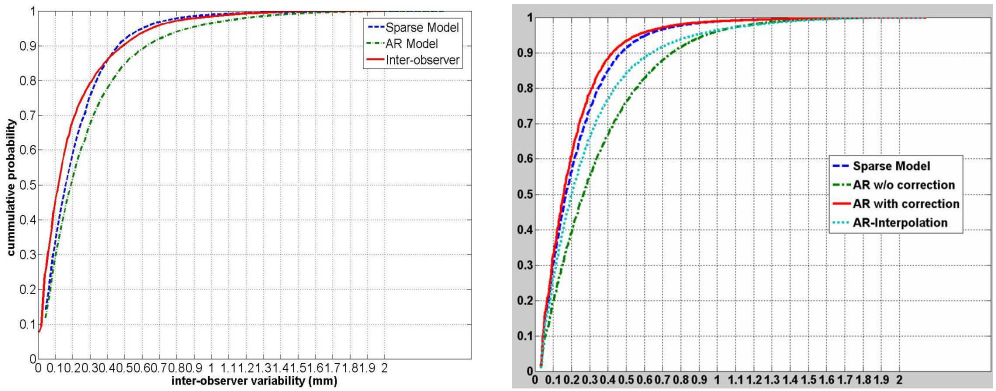


Fig. 3.8: Cumulative distribution of distances between clinicians' contours, SIM and causal AR tracking and AR-Interpolation. X axis in mm, Y axis in percentage of contours. Inter-observer variability curve extracted from Fig. (6) in [100].

of E_{predict} by gradient descent is straightforward, and E_{obs}'s PDE with respect to a particular PCA coefficient x_q may be written as

$$\frac{\partial E_{\text{obs}}}{\partial x_q} = \frac{\partial E_{\text{obs}}}{\partial \mathbf{Y}} \frac{\partial \mathbf{Y}}{\partial x_q} = \iint_{\Omega} \left(-log \left(\frac{h_{\text{in}}(I(\mathbf{x}))}{h_{\text{out}}(I(\mathbf{x}))} \right) \right) \delta_{\mathbf{Y}}(\mathbf{x}) \frac{\partial \mathbf{Y}}{\partial x_q} d\mathbf{x}, \tag{3.22}$$

where the Kronecker operator $\delta_{\mathbf{Y}}(\mathbf{x})$ is equal to one if $\mathbf{x}$ is on the contour, and zero elsewhere.

The validation is based on a 34 liver dataset (Siemens Somatom Sensation 16, 0.6-0.8 mm planar resolution and 1-5 mm interslice spacing), hand-segmented by experts, with all-but-one framework. For each test, 33 livers were used for modeling and one for testing. The percentage of correctly labeled pixels are displayed in table (3.2), along with the mean shape, and active shape model [38] results for comparison (see figure (3.10)). The same observation metric (see equation (3.21)) was used for Autoregressive active models (explicit contours) and 3D active shape model (implicit distance function). For the latter, $P\left(\mathbf{Y}_t | \mathbf{Y}_{1:t-1}\right)$ was estimated using the Mahalanobis distance and the PCA eigenvalues. All datasets were registered rigidly, for a direct comparison with the mean shape (distance function) procedure.

Table (3.2) demonstrates that active models (both ASM and autoregressive active models) improves the segmentation, compared to simply taking the registered mean distance function. Furthermore, since Autoregressive Active Models perform better than ASM (AAM reduces the number of the undersegmented pixels by more than 7%), it demonstrates that, for certain types of 3D organs that undergo large shape variations and with weak image support, a local segmentation, constrained by a 3D statistical model (the 3D autoregression) performs better than a global 3D

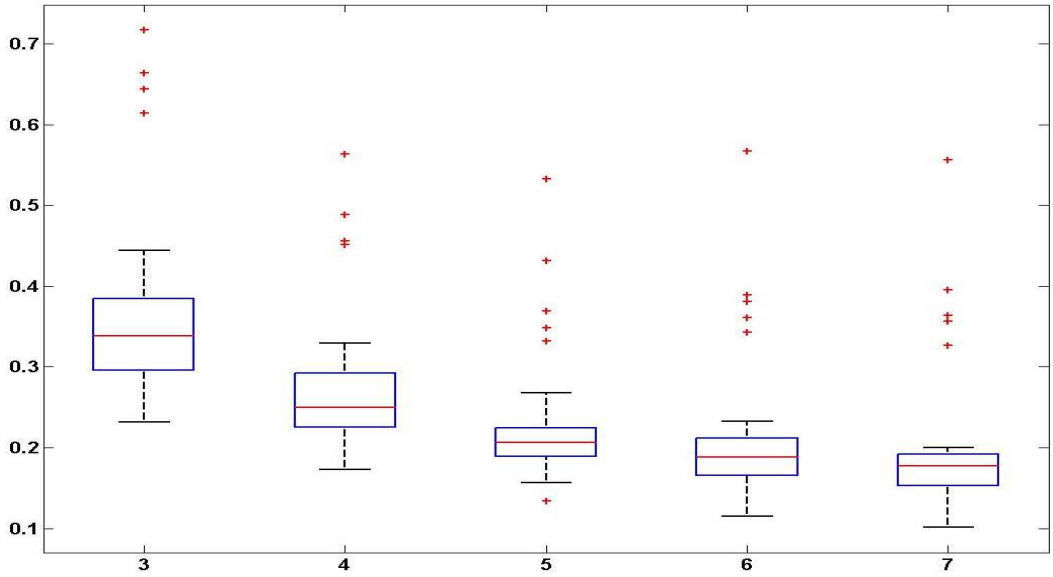


Fig. 3.9: Boxplot results of interpolation using the 2-order autoregressive model with respect to the number of contours inputs (3 to 7).

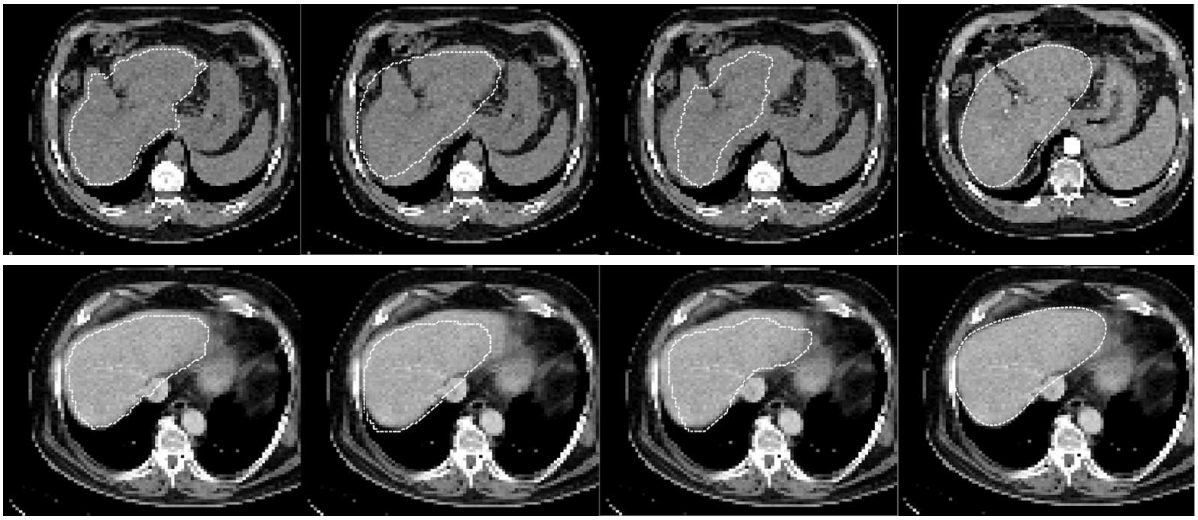
segmentation method	Mean Shape	3D-ASM	AAR
% of overseg. pixels	1.73%	0.67%	0.24%
% of underseg. pixels	30.7%	18.57%	11.47%
% of object true positives	69.33%	81.43%	88.63%
% of full image true positives	95.21%	96.58%	99.63%

Tab. 3.2: Results table showing the percentage of pixels correctly segmented, over a dataset of 34 patients, using autoregressive active model (AAR), 3D mean shape, 3D active shape model (ASM). Oversegmented pixels are background pixels labeled as liver, and undersegmented pixels are liver pixels labeled as background.

approach. The autoregression provides a local update in the sequential segmentation, along with a measure of confidence in the image support, that ASM does not.

3.4 Conclusion and Discussion

In this chapter we have demonstrated that a non-stationary approach introduces novel perspectives within the AR models and shape priors. The present method benefits from the same advantages as stationary models, described in [43], but does not rely on any stationary assumptions. Datasets with time-varying dynamics, that stationary ARs are not able to track, are now successfully processed



(1) (2) (3) (4)

Fig. 3.10: Liver at same slice segmented, from left to right, by expert (1), mean shape (2), ASM (3) and autoregressive active model (4).

by the locally adaptive AR model. Furthermore, to some extent, the non-stationary models handle occlusions and missing data.

In the case of stationary problems, the autoregressive processes are successfully used to model the relationship between consecutive elements of the solution. However, they are subject to accumulation of errors without a correction at each step of the sequence. In order to overcome this limitation, an interpolation procedure is presented so that the difference between the interpolated sequence and the sequence predicted by the autoregression is minimal in the least squares sense. Furthermore, the autoregressive interpolation allows a flexible choice in the indices of the initial constrains which represents an incidental advantage for numerous applications.

The selection of an appropriate representation of the target as well as its dimensionality reduction are open problems in this context. In this chapter we have used an implicit representation for the registration, an explicit representation for the modeling and a PCA for the dimensionality reduction. In terms of registration, being able to account for open structures is an open challenge. In terms of contour representation and parameterization, and distance between observation and prediction more advanced techniques that the Euclidean metric are to be investigated. Last, but not least more appropriate dimensionality reduction techniques like LDA, Fourier or splines bases that could encode a continuous representation of the contour are to be investigated. In order to increase robustness and reactivity, with the same Gaussian noise assumption, the framework pro-

vided by Kalman filter to model the non-stationary characteristics of the noise could be integrated. Furthermore, in the case of occlusions, when the Gaussian assumption does not hold, one might be tempted to use a much heavier nonparametric representation for the distribution, such as particle filtering. A last interesting perspective might also be to incorporate the quality of the segmentation into the on-line learning (in sections (3.2.3)-(3.2.1)) to favor the time steps that gave the best results.

In chapter 2, a method is proposed for solving a class of Computer Vision problems using sparse elements and reconstructing the other solution's elements. In the present chapter, this reconstruction is performed using an adaptive regression law and a segmentation procedure. However, the segmentation uncertainties are not explicitly taken into account, whereas the uncertainty statistics could be modeled from prior learning both in time and space. Chapter 4 presents how a segmentation problem is turned into a geometric tracking problem, and how uncertainties are modeled using a Bayesian framework.

Chapter 4

Uncertainty Models for Progressive Segmentation

Abstract – *This chapter introduces a novel approach in the use of uncertainty models to solve computer vision problems. In the context of tubular structures segmentation, the multi-hypothesis framework is introduced and implemented with a sequential random sampling procedure called particle filters. First, it is shown that the global solution to this problem cannot be achieved by sequentially optimizing the local most likely solution. Then, the structures are modeled by a template of few parameters and segmented in a sequential fashion by randomly sampling the feature space and assigning a probability measure to each sample based on shape and appearance. This probability field is then propagated geometrically along the vessel. The resulting solution performs better than other methods based on front propagation and Hessian analysis. Thanks to the multi-hypothesis framework the pathologies, acquisition artifacts and branchings are successfully segmented. Then, in order to compare this framework with commonly used gradient descent approaches, it is extended to a static parametric problem: 2D-3D registration.*

4.1 Introduction to Uncertainty Models: Static Problem Solved with Condensation

Gaussian assumptions are frequently considered in computer vision and medical imaging. The central idea is that the space manifold of the solution forms a Gaussian density. Even though such an assumption was not made explicitly, it was implicitly encoded in the sparse information models in chapter 2 as well the stationary auto-regressive process in chapter 3. Similar concept was considered in the dynamic ARs, where a succession of such models was considered. Furthermore, these models were mostly presented in terms of mean states with their uncertainties being neglected. In medical imaging and computer vision one expects great benefit from a qualitative interpretation of the results. Therefore, introducing the notion of uncertainties as well as their direct estimation is a quite promising direction. Furthermore, in particularly difficult problems local information may conflict with the global solution; for instance, in the case of coronary arteries segmentation presented below in section (4.2), the global solution is not composed of locally optimum solutions, therefore a multi-hypothesis framework has to be developed.

Since the main focus of our research was single and multi-image segmentation, we consider a particular problem in this area to review the state of the art and introduce the notion of multiple hypotheses testing and uncertainties estimation. In particular we are interested in the segmentation of tubular structures that is a rather tractable problem from parametric perspective which allows easily the study of statistical variations. Furthermore, the outcome of such a process has important clinical value and can be used as a low-level primitive in a number of clinical tools.

4.2 Introduction to Tubular Structures Segmentation using Models of Uncertainty

4.2.1 Medical Motivation

Cardiovascular diseases are the leading cause of death in the western world (more than 30%) and therefore there is a constant demand for improvement of diagnostic tools to detect and measure anomalies in the coronary tree. Such tools could provide early diagnosis of potential vascular anomalies and therefore prevention that can significantly decrease the mortality rate due to cardiac

diseases. Coronary arteries are narrow vessels (between 3 to 5 mms adjacent to the aorta, between 1.5 to 2.5 mms after two branchings) whose role is to feed the heart muscle with oxygenated blood. Their segmentation provides a valuable tool for clinicians to diagnose pathologies such as calcifications and stenoses. Nevertheless, their segmentation is a difficult task because of the low contrast conditions, bifurcations, intensity distortions produced by pathologies and scanner artifacts, and the coronaries' proximity to the heart chambers [160].

Since Computer Tomography (CT) and Magnetic Resonance (MR) imaging of the heart are now widely available, the number of patients scanned has significantly increased in the past few years. Clinicians are now interested in longitudinal studies to measure the development and severity of vascular diseases and their effects on the heart function. Such information is used to determine the time of surgical operation and the effectiveness of treatments. Furthermore, the major diagnosis performed on the coronary arteries involves their diameter (e.g. identification of calcifications and stenoses quantification); consequently CT Angiography's use in quantifying luminal stenosis is growing exponentially [165]. To make actual measurements of the 3D vessel diameter, the clinicians need more than just volume rendering; they need the arteries to be segmented.

4.2.2 Previous work in Tubular Structures Segmentation

Model-free

Tubular structures segmentation techniques consist of model-free and model-based methods. Skeleton-based techniques are the most fundamental among the model-free [173] and aim at detecting the vessel centerlines, from which the whole vessel tree is reconstructed. Region growing methods [199] progressively segment the vessels from a seed point, based on intensity similarity between adjacent pixels. These methods are successful for homogeneous regions, but pathological vessels are more challenging, and may leak into other structures of similar intensity. Morphological operators [71] can be applied to correct a segmentation, smooth its edges or eventually fill holes in the structure of interest, but fail to account for prior knowledge. Tracking approaches [97, 182] are based on the application of local operators to track the vessel. Given a starting condition, such methods recover the vessel centerline through processing vessel cross section information [87]. Various forms of edge-driven techniques, similarity/matching terms between the vessel profile in successive planes, as well as their combination, have been considered to perform tracking. In particular a method that relies on the average outward flux of the gradient vector field of the Euclidean

distance from the vessel boundary to recover skeleton points has been developed in [20]. In [4], a multiscale method to segment thin nets (lines where the gray level is locally extremum). First, the image is filtered by a Gaussian at a certain scale is considered. For each scale, the image maximum curvature is computed and based on differential properties, the points that belong to the vessels centerline are kept.

Deformable models may either be parametric or geometric. Parametric deformable models [161] can be viewed as elastic surfaces (often called snakes), and can handle topological changes with some difficulties. Geometric deformable models [30, 167], on the contrary, can change their topology during the process and may eventually leak into neighboring structures or vasculature. Like snakes, deformable models aim at minimizing the energy computed along the model. Level sets [136] are a way to apply deformable models to non-linear problems, such as vessel segmentation [124]. One may use the fast marching algorithm and its variant for vessel segmentation using the minimal path principle [6, 60, 196] to determine the path of minimal length between two points, backtracking from one point toward the other crossing the isosurfaces perpendicularly. To discourage leaking, a local shape term that constrains the diameter of the vessel was introduced in [128]. One should also mention the method introduced in [120], where the optimization of a co-dimension two active contour was presented to segment brain vessels.

To account for the snake's sensitivity to initialization, in [70] the initial conditions of snakes are determined after a learning process based on a non-parametric estimator. The learning function uses a Parzen window estimator, with a Gaussian kernel. The Parzen window estimator relies on feature value observations, and compare these values with model values. The objective function is then chosen to maximize the probability distribution of these observable quantities.

The maximization of flux was introduced in [188] and was exploited for vessel segmentation in [62] in low contrast conditions using vessel measures introduced in [76]. The vectors normal to the vessel boundaries are collected using the Hessian matrix eigenvectors collected on the points satisfying the vessel measures. The geometric maximizing flux algorithm is then applied to recover the vessel boundaries.

Model-based

Model-based techniques, on the other hand, use prior knowledge and features to match a model with the input image and extract the vessels. The knowledge may concern the whole structure, or

consist of a local region of the vessel. Along this direction, vessel template matching techniques (deformable template matcher) [148] have been investigated. The template model consists of a series of connected nodes that is deformed to best match the input image. Generalized Cylindrical models are modified in Extruded Generalized Cylinders in [133] to recover vessels in angiograms. For highly curved vessels, the local basis used for classical generalized cylinders may be twisted, and a non-orthogonality issue may occur. This problem is solved by keeping the vessel cross section orthogonal to the centerline and the two normal vectors always on the same side of the tangent vector spine as the algorithm moves along the vessel. In [111], the vessel is modeled by a tubular structure, and segmented by filtering the image with a multiscale-structural term derived from the image intensity Hessian matrix [163, 76].

In [27], an anisotropic filtering technique, called *Vesselness Enhancement Diffusion*, is introduced that can be used to filter noisy images preserving vessels boundaries. The diffusivity function relies on the *vesselness* function introduced in [76] to filter along the vessel principal direction and not across. In the resulting image, the background is smoothed, whereas the vessel remains unchanged. Although it is not a segmentation method per se, it dramatically changes the visualization of the vessels and allows filtering based on scale so that only the smallest or biggest structures may be seen. A complete review of tubular structure segmentation in the case of vasculature imaging for MR may be found in [176].

4.2.3 Overview of our method

Existing approaches suffer from certain limitations. Techniques such as local operators, region growing, morphological filters and geometric contours are prone to be sensitive to local minima and fail to take into account prior knowledge on the form of the vessel. Alternatively, cylindrical models, parametric active contours and template matching techniques may not be well suited to account for the eventual distortions of vessel appearance produced by pathologies or scanner artifacts, and require special handling of bifurcations. Tracking methods, on the other hand, may often fail in the presence of missing and corrupted data, or sudden changes. Level sets are time-consuming when they are implemented in the most general way. On the other hand, their efficient implementation using the fast marching method [168, 183] reduces computational burden at the cost of loosing the local implicit function properties.

To improve the segmentation results, one should be able to account for bifurcations, sudden

changes of pixels intensity (pathologies) and missing or corrupt data (motion blur in MR and CT beam hardening). This excludes most of the methods based on parametric models, or linear models, which would require special handling for bifurcations and intensity artifacts.

The segmentation problem is replaced by a tracking problem: the course of the vessel is followed by obtaining the set of 2D planar (tangential to the centerline) segmentation (see figure (4.1)). On a particular plane, the vessel is represented by a model (see section 4.3.1), whose parameters are optimized to fit the image data according to a shape and an appearance measure. However, to follow the centerline by always selecting the maximum likelihood in the parameter space is not sufficient, as demonstrated by the experiment in section 4.3.2. That is the reason why the authors are driven toward a method that would handle multiple hypotheses, and keep only the few most probable following [74, 75]. At each step, a scheme based on particle filtering [98][67] is used to sample the parameters probability density function (pdf). The final segmentation is obtained using the maximum a posteriori of this pdf, and a 2D contour extraction technique called Circular Shortest Path [3].

The remainder of this chapter is organized as follows. In section 4.3, we start by describing the image model in section 4.3.1, then in section 4.3.2, we demonstrate the insufficiency of the Kalman filter and the linear Gaussian model, which leads us to introduce a non parametric method known as particle filtering in section 4.3.3 and define different resampling strategies. In section 4.3.4, we give the details of the image measure used to estimate the pdf. Section 4.3.6 introduces the Circular Shortest Path with which the final segmentation results are obtained. In section 4.4, the experimental method is explained and the results are presented in section 4.4.3. Finally, section 4.6 concludes the chapter with a discussion.

4.3 Segmentation Model & Theoretical Foundations

4.3.1 Description of the Segmentation Model

The vessel is described as a feature space consisting of a series of cross-sections (see figure (4.1) and figure (4.2)). Each cross-section is defined by:

$$\underbrace{\mathbf{x} = (x_1, x_2, x_3)}_{position}, \underbrace{\mathbf{\Theta} = (\theta_1, \theta_2)}_{orientation}, \underbrace{\mathbf{p}_{vessel}}_{appearance} \tag{4.1}$$

Fig. 4.1: The vessels are tracked by sequentially segmenting their cross-sections.

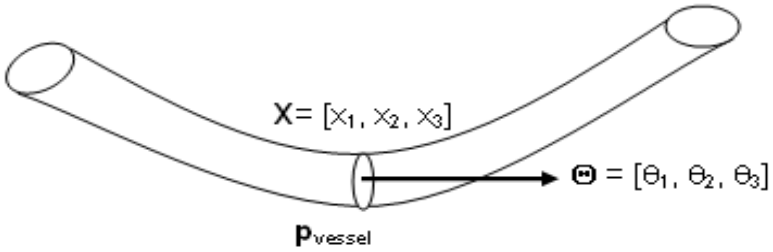


Fig. 4.2: The feature space is defined by the cross-section center position $\mathbf{x} = (x_1, x_2, x_3)$, the cross-section tangential direction $\mathbf{\Theta} = (\theta_1, \theta_2, \theta_3)$ and the lumen pixel intensity distribution $\mathbf{p}_{vessel}$.

where the vessel state vector consists of the 3D location of the vessel $\mathbf{x}$ at the cross-section, the orientation vector $\mathbf{\Theta}$ defined by two angles, and the parameters required for the pdf estimation of the appearance of the vessel, $\mathbf{p}_{vessel}$, as a mixture of two Gaussians:

$$\mathbf{p}_{vessel} = ((P_B, \mu_B, \sigma_B), (P_C, \mu_C, \sigma_C)) \tag{4.2}$$

It is reasonable to assume a rather non uniform structure in the appearance of the vessel because of the presence of calcifications, stents, stenosis and diseased vessel wall (see figure (4.3)). Therefore simple parametric statistical models on the appearance space will fail to account for the statistical properties of the vessel and more complex distributions are to be considered. We consider a Gaussian mixture model that consists of two components to represent the evolving distribution of

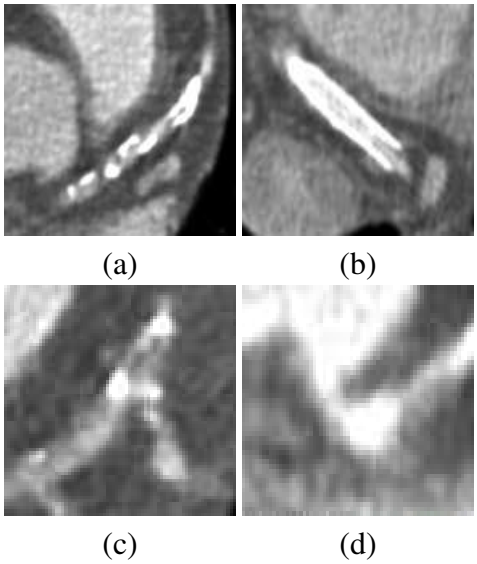


Fig. 4.3: (a) calcification, (b) stent (high intensity prosthesis), (c) branching with obtuse angles, (d) stenosis (sudden reduction of vessel cross section diameter).

the vessel: first, the contrast enhanced blood $(\text{mixture weight} : P_B, \text{mean} : \mu_B, \text{variance} : \sigma_B)$ and next, the high density components, such as calcifications or stent, $(\text{mixture weight} : P_C, \text{mean} : \mu_C, \text{variance} : \sigma_C)$. This mixture is subject to the constraint $[P_C + P_B = 1]$. Then, the vessel may be described at a single cross-section by the following state vector:

$$x_t = (\mathbf{x}, \mathbf{\Theta}, (P_B, \mu_B, \sigma_B), (P_C, \mu_C, \sigma_C)) \tag{4.3}$$

The number of Gaussians in the mixture is justified empirically by the pixel intensity distribution observed in the lumen of different patients. Such a state vector is to be estimated for consecutive planes leading to complete reconstruction of the vessel tree. A simple solution consists in optimizing the state vector in the feature space, and to regularize the centerline with a common tracking procedure (the Kalman filer). However, section 4.3.2 demonstrates why this approach is insufficient, and justifies a process based on multi-hypothesis (see particle filtering in section 4.3.3).

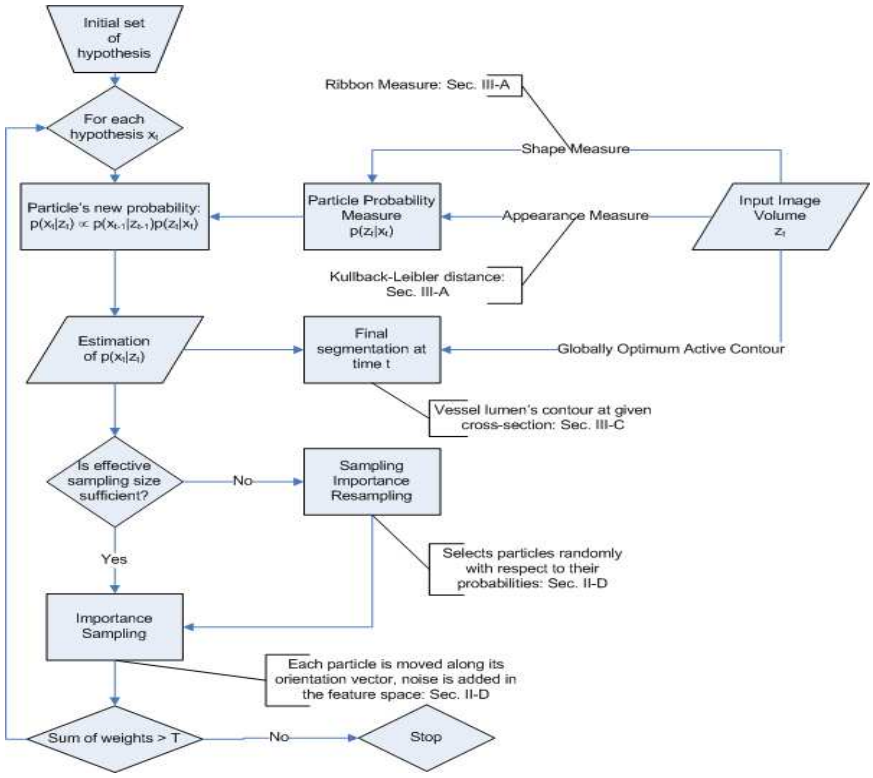


Fig. 4.4: Overview of the different steps involved in the tubular structures sequential segmentation using particle filters.

4.3.2 Introduction to Linear/Gaussian Prediction Models and Limitations

Vessel segmentation is represented by a probability density function in the state vector feature space (see section 4.3.1). A simple representation of a probability field is to model it by a Gaussian, and to assume a linear transition between consecutive states.

As described in [194], the Kalman filter [103] is a set of mathematical equations that estimates the state variables of a process in the least square sense. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. In some cases, Kalman Filter may track non linear processes [150]; nevertheless, as we shall see, the Kalman Filter fails to track inhomogeneous tubular structures (e.g. branchings, pathologies or corrupt data) such as coronary arteries.

Such a filter assumes that the posterior density is Gaussian at each time step, and that the current state x_t and observation z_t are linearly dependent on the past state x_{t-1}. Such assumptions

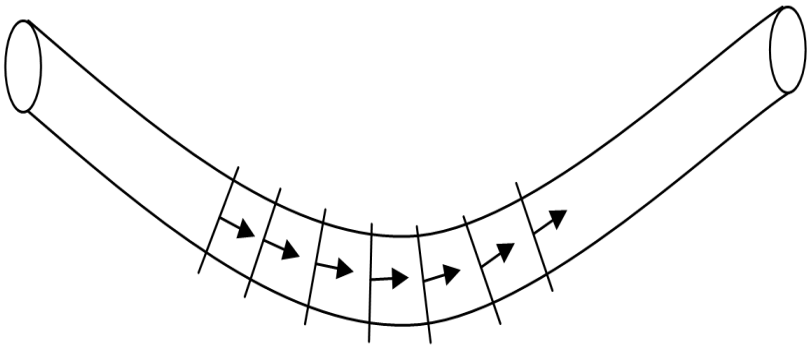


Fig. 4.5: With the Kalman filter, the tubular structure is tracked by optimizing the model's parameters in the feature space, and sequentially determining a set of 2d cross-sectional planes.

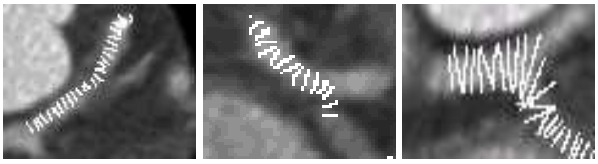


Fig. 4.6: Vessels are segmented by tracking their cross-sections (white bars) using Kalman Filter. The Kalman Filter is unable to track bifurcations and appearance inhomogeneities (artherosclerosis).

simplify the Bayesian equations to the following form:

$$\begin{cases} x_t &= F_t x_{t-1} + v_{t-1} \\ z_t &= H_t x_t + n_t, \end{cases} \tag{4.4}$$

where v_{t-1} and n_t refer to zero mean Gaussian noise with covariance matrices Q_{t-1} and R_t that are assumed to be statistically independent. The matrix F_t is considered known and relates the former vessel state x_{t-1} to the current state x_t. The matrix H_t is also known and relates the vessel state x_t to the observation z_t. In the case of tubular structure segmentation, the vessel's state x_t is described in equation (4.3), the state transition F_t refers to the translation of x_t's location along its orientation vector, all other features remaining unchanged. The observation z_t refers to the image intensities on the cross-section defined by x_t. The pdfs are computed recursively according to the formulas that may be found in Kalman's seminal paper [103].

Kalman filters have been considered to track vessels in retinal images [150] combined with a deterministic step to handle branchings. Such a model requires strong assumptions of Gaussian noise and linearity in the state transitions which make their use in vessel segmentation quite problematic. Some (negative) examples of the application of such a linear model to vessel segmentation are shown in figure (4.6), using the state space earlier introduced and the Kullback-Leibler infor-

mation criterion to measure the distance between prediction and observation. One may claim that neither the observation space (figure (4.3)), nor the structure/geometric space of the vessel (figure (4.6)) are sufficiently homogeneous to allow parametric linear state transitions, and such a method will fail to account for pathological cases where such linearity is absent. These reasons naturally drive us toward a non parametric and non linear approach, such as the particle filters.

4.3.3 Non-linear Systems and Introduction to Particle Filters

Generalities

Particle filters [67, 5, 98] are a sequential Monte-Carlo technique that is used to estimate the Bayesian posterior probability density function (pdf) with a set of samples [81, 195]. In terms of a mathematical formulation, such a method approximates the posterior pdf by M random measures $\{x_t^m, m = 1..M\}$ associated to M weights $\{w_t^m, m = 1..M\}$ such that, given a sequence of observation $\{z_{1:t}\}$

$$p(x_t|z_{1:t}) \approx \sum_{m=1}^{M} w_t^m \delta(x_t - x_t^m). \tag{4.5}$$

where each weight w_t^m reflects the importance of the sample x_t^m in the pdf, given the observations sequence $z_{1:t}$, as shown in figure (4.7). $\delta(x)$ is the Kronecker function equal to 1 when $x = 0$ and 0 elsewhere. Using the Bayes rule, one sequentially estimates $p(x_t|z_{1:t})$ from $p(x_{t-1}|z_{1:t-1})$, knowing $p(x_t|x_{t-1})$ and measuring $p(z_t|x_t)$:

$$\begin{aligned} p(x_t|z_{1:t}) &\propto p(z_t|x_t)p(x_t|z_{1:t-1}) \\ &\propto p(z_t|x_t) \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1} \end{aligned}$$

$p(z_t|x_t)$ is discussed in section 4.3.4, while a novel method to locally estimate $p(x_t|x_{t-1})$ is presented below under the name of reinforced SIR. The samples x_t^m are drawn using the principle of importance density [82] of pdf $q(x_t|x_{1:t}^m, z_t)$. The principle of importance density consists in drawing the samples x_t^m following the distribution q, and assigning their weight according to

$$w_t^m = \frac{p(x_t^m|z_{1:t})}{q(x_t^m|z_{1:t})}.$$

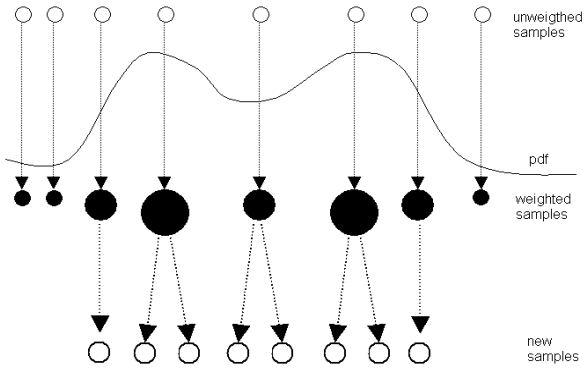


Fig. 4.7: The resampling process: a random selection chooses the samples with the highest weights where a local perturbation is applied.

In the framework of sequential estimation, these weights w_t^m are better expressed in a sequential way by updating them:

$$w_t^m \propto w_{t-1}^m \frac{p(z_t|x_t^m)p(x_t^m|x_{t-1}^m)}{q(x_t^m|x_{t-1}^m, z_t)}, \quad \text{with} \sum_m w_t^m = 1. \tag{4.6}$$

The importance density reflects the way the particles are distributed in the feature space; for instance, a uniform importance density means the samples are uniformly sampled in the feature space regardless of the distribution p. Once a set of samples has been drawn, $p(x_t^m|x_{t-1}^m, z_t)$ is computed out of the observation z_t for each sample x_t^m, and the estimation of the posteriori pdf is sequentially updated using equation (4.6). With such a process many particle weights quickly converge to zero and only the ones that represent the data will present significant weights. Consequently the model will lose its ability to track significant changes on the pdf; therefore a resampling procedure has to be executed on a regular basis. Such a process will preserve as many samples as possible with significant weights. One may find in the literature several resampling techniques [67]. We chose the most prominent one, sampling importance resampling, for its simplicity to implement, and because it allows more hypothesis with low probability to survive, compared to more selective techniques such as stratified resampling [69].

Sampling Importance Resampling

The Sampling Importance Resampling (SIR) algorithm [81] consists of choosing the prior density $p(x_t|x_{t-1})$ as importance density $q(x_t|x_{1:t}^m, z_t)$. This leads to the following condition, from equation

(4.6)

$$w_t^m \propto w_{t-1}^m p(z_t|x_t^m). \tag{4.7}$$

The samples are updated by drawing x_t^m from the distribution $p(x_t|x_{t-1}^m)$, and perturbed according to a random noise vector ϵ, so that $x_t^m \propto p(x_t|x_{t-1}^m)$. The SIR algorithm is the most widely used resampling method because of its simplicity from the implementation point of view. Nevertheless, the SIR uses mostly the prior knowledge $p(x_t|x_{t-1})$, and does not take into account the most recent observations z_t. Such a strategy could lead to an overestimation of outliers. On the other hand, because SIR resampling is performed at each step, fewer samples are required, and thus the computational cost may be reduced compared to other resampling algorithms. Finally, in practice, the estimation of ϵ's law is difficult, and prior knowledge is usually required. A novel method is proposed in the paragraph below to address this issue, by locally estimating $p(x_t|x_{t-1}^m)$.

Reinforced SIR: the State Transition Noise pdf

One overcomes some of the above limitations through the improvement of the perturbation model. Such a step leads to a more effective use of samples and should be driven from the data, or prior knowledge on the form of the vessel. The idea of reinforced SIR is to update the noise model for the random noise vector ϵ added to each sampled during the SIR process. This noise vector refers to $p(x_t|x_{t-1}^m)$.

Let us assume the state transition from x_{t-1} to x_t is known and modeled by

$$x_t = f(x_{t-1}) + \epsilon \tag{4.8}$$

In the case of tubular structures, f represents the location $\mathbf{x}$ displacement along the tangential direction $\boldsymbol{\Theta}$ (see equation (4.3)). From ground truth segmentation, or from previous steps in the current segmentation, the distribution law of ϵ is learned so that the model correctly approximates the noise.

After a particle x_{t-1}^m has been selected by the SIR algorithm, a random noise vector ϵ is added (see previous section paragraph). A straightforward solution consists in using prior knowledge to estimate the law of ϵ (i.d. $p(x_t|x_{t-1}^m)$) once for all. This method presents two difficulties: first, prior knowledge may be limited and/or hard to obtain, second, vessels are linear structures only very locally, therefore the law of ϵ may greatly vary from one part of the vessel to another. That

is the reason why an intermediate step in introduced before the resampling to estimate $p(x_t|x_{t-1}^m)$. This step consists in drawing uniformly distributed samples - called here auxiliary samples - around $f(x_{t-1}^m)$ (i.d. the sample's location **x** at the previous step displaced along its tangential direction $\boldsymbol{\Theta}$), estimating the probability of each auxiliary sample, and updating the model for $p(x_t|x_{t-1}^m)$. Then, in the subsequent step, SIR is performed as above with the updated model for ϵ.

The final paradigm for resampling follows the procedure:

1. particles x_{t-1}^m are selected randomly according to their probability, as in any SIR procedure
2. the selected particles generate N new offspring uniformly distributed
3. these offspring probabilities are estimated, and a pdf $p(x_t|x_{t-1}^m)$ is then drawn for each SIR selected particle
4. this pdf $p(x_t|x_{t-1}^m)$ is used to generate a random noise vector ϵ that perturbs the SIR selected particles

In other words, once the SIR selected a particle x_{t-1}^m to be resampled, $p(x_t|x_{t-1}^m, z_t)$ is estimated in a way similar to equation (4.5):

$$p(x_t|x_{t-1}^m, z_t) \approx \sum_{i=1}^{N} w_t^i \delta(x_t - x_t^i), \tag{4.9}$$

where the x_t^i are generated from $x_{t-1}^m + \epsilon_i$, with the ϵ_i uniformly distributed. The weights w_t^i are estimated from the observation z_t.

This method presents two main advantages. First, as the noise vector ϵ is random, the advantages of SIR over exhaustive search are preserved. Second, the distribution of ϵ is updated at every time step, and for every particle, avoiding the disadvantages of having a noise distribution that would be determined once for all from prior knowledge. Vessels can be straight and suddenly become tortuous, or can have a very homogeneous shape/appearance before encountering a very inhomogeneous region. This *Reinforced SIR* captures the conditions change and adapts the noise vector distribution.

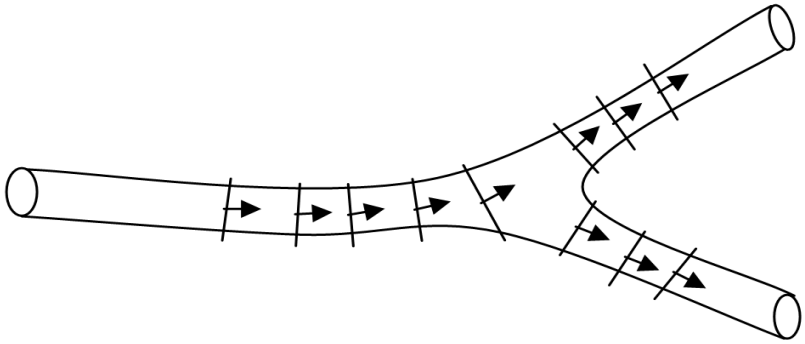


Fig. 4.8: With the particle filter, the model parameters' pdf is estimated at each time step. The maximum a posteriori is then taken to obtain the segmentation

4.3.4 Prediction & Observation: Distance

Particle filtering theory is now applied to estimate the pdf of the feature vector (see equation (4.3)) during the vessel tracking process. Each one of the particles x_t^m represents a hypothetic state of the vessel; a probability measure $p(z_t|x_t^m)$ is used to quantify how the image data z_t fits the vessel model x_t^m. To this end, we are using the image terms, and in particular the intensities that do correspond to the vessel in the current cross-section. The observed distribution of this set is approximated using a Gaussian mixture model according to the expectancy-maximization principle. Each hypothesis is composed by the features given in equation (4.3), therefore, the probability measure is essentially the likelihood of the observation z, given the appearance A model. The following measures (loosely called probabilities) are normalized so that their sum over all particles is equal to one. Assuming statistical independence between shape S and appearance model A, $p(z_t|x_t) = p(z_t|S)p(z_t|A)$.

- Probability measure for shape based on contrast

  Given the vessel model (see equation (4.3)), whose parameters are specified by the particle x_t, a measure of contrast, that we call the *ribbon measure* R, is computed:

$$\begin{cases} R = -\infty \quad , & \mu_{int} \leq \mu_{ext} \\ R = \frac{\mu_{int}-\mu_{ext}}{\mu_{int}+\mu_{ext}}, & otherwise \end{cases} \tag{4.10}$$

  while the correctness of the prediction is given by:

$$p(z|S) = e^{-\frac{|R|}{R_0}} \tag{4.11}$$

where R_0 is a normalizing constant (the average value of R from ground truth), μ_{int} is the mean intensity value for the voxels in the vessel, and μ_{ext} is the intensities mean value for the voxels in a band outside the vessel, such that the band and the vessel's lumen have the same area. This measure is normalized to be equivalent to model a probability measure.

Since the coronary arteries are brighter than the background, the best match maximizes R (see figure (4.11)).

- Probability measure for appearance

  For the vessel lumen pixels distribution $\mathbf{p}_{vessel}$ equation (4.2), the probability is measured as the distance between the hypothesized distribution and the distribution actually observed.

  The distance we use is the symmetrized Kullback-Leibler distance $D(p,q)$ between the model $p(x) = \mathbf{p}_{vessel}$ and the observation $q(x)$:

$$D(p,q) = \int p(x)log(\frac{p(x)}{q(x)}) + q(x)log(\frac{q(x)}{p(x)})dx, \tag{4.12}$$

$$p(z|A) = e^{-\frac{|D(p,q)|}{D_0}}, \tag{4.13}$$

  where D_0 is a normalizing constant, equal to the average value of D from prior knowledge. Once again, this measure $p(z|A)$ is normalized to be equivalent to a probability measure.

The combination of edge-driven and region-based metrics measures the fitness of the observation to the prior knowledge included in the state vector.

4.3.5 Branching Detection

When a branching occurs, the particles naturally split up in the two daughter branches (the case of trifurcation is not studied here), and then track them separately (see figure (4.9)). As branchings are never perfectly balanced, one of them attracts the majority of the particles after few resampling steps. To avoid the collapse of one of the modes, two techniques are available: either to increase the number of particles in the weakest branch, or to treat the two branches separately. The second approach is preferred in this chapter, for the particular context of vessel segmentation. To this end, a simple K-means clustering in the joint space (position+orientation) of the particles is considered at each iteration. When the two clusters are well separated (when the distance between the clusters center is above a certain threshold), the number of particles is doubled and they are equally

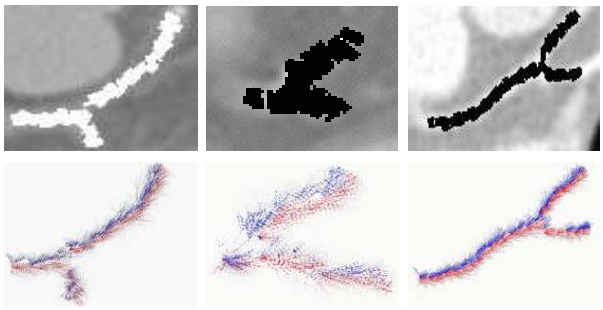


Fig. 4.9: (a) branching points between LCX and LAD for three patients with the particles' mean state overlaid, (b) the particles are clustered at each time step. The branching is detected when the distance between the two clusters center is above a certain threshold.

dispatched in the two branches. The segmentation goes on, according to equation (4.6), with a bi-modal distribution.

The K-means algorithm [68] partitions N points, x_n, into K disjoined clusters, of centers μ_j, minimizing the sum-of-squares

$$J = \sum_{j=0}^{K} \sum_{n=0}^{N} |x_n - \mu_j|^2. \tag{4.14}$$

The K-mean procedure alternates two steps: first each point is associated to the nearest center μ_j, then each center is moved in the barycenter of the cluster.

As illustrated in figure (4.4), the branching detection algorithm is fired at each iteration. Nevertheless, this algorithm is limited in several ways: first, the number of branches at a junction (2 in our case) has to be fixed in advance. In one case (out of 34 tested, see section 4.4.3), one branch is missed because of a 3-branches junction.

4.3.6 Circular Shortest Paths & 2D Vessel Segmentation

The *Circular Shortest Paths by Branch and Bound* (CSP) [3] is a binary search-tree technique to recover the globally optimal active contour, given a point inside the contour and a potential map. First of all, let us note that the problem of finding the globally optimal active contour is equivalent to computing the minimal *weight* path (given a Riemannian metric) that connects a point at angle $\theta = 0$ to its equivalent at $\theta = 2\pi$ across the log-polar transform of the original image. Given a Riemannian metric g (usually inversely proportional to the image gradient), the *weight* W of a path

P is defined as:

$$W(\mathbf{P}) = \int_{\mathbf{P}} g\left(\mathbf{P}(s)\right) ds. \tag{4.15}$$

Given a start point p_0 at $\theta = 0$, the end point $p_{2\pi}$ at $\theta = 2\pi$ of the minimal *non-circular* path is defined as

$$p = argmin_{\mathbf{P}(2\pi)=p} W(\mathbf{P}). \tag{4.16}$$

This end point $p_{2\pi}$ is very quickly found using the well-known Dijkstra [63] algorithm, with the Riemannian metric g (equation (4.15)) playing the role of potential map. To demonstrate the use of a binary search-tree, a property needs to be stated at this point, whose proof is straightforward (see [3]):

> for two sets $S_1 \subseteq S_2$, the minimal path $\mathbf{P}_2$ of S_2 has a lower or equal weight than the minimal path $\mathbf{P}_1$ of S_1, otherwise stated as $W(\mathbf{P}_2) \leq W(\mathbf{P}_1)$.

A corollary is:

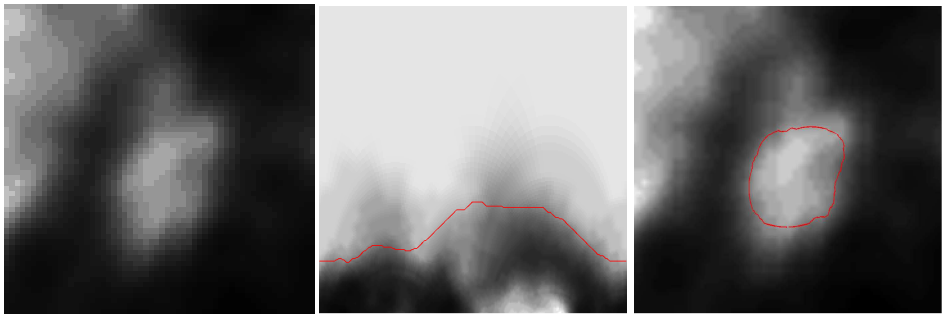
> for any point set S, the weight of the minimal path **P** (circular or not) is a lower bound of the minimal *circular* path weight. Therefore, if $\{S_1, S_2\}$ is a partition of S, and $W(\mathbf{P}_1) \leq W(\mathbf{P}_2)$, the minimal *circular* path of S has its starting point p_0 (and obviously ending point $p_{2\pi}$ as well) in the subset S_1.

Consequently, a binary search-tree is used in the CSP algorithm. First, the tree's root consists of any set of initial points $S = \{p_0\}$ which is divided into two subsets $S_1 = \{p_0\}_1$ and $S_2 = \{p_0\}_2$; second, the minimal *non-circular* paths P_1 and P_2 are computed using equation (4.16) for the two subsets. Each set of vertices is represented by a node in the tree, and has two children nodes (the two subsets). This procedure is then repeated until the subsets are reduced to a single point. At the bottom of the binary search-tree, the subsets are reduced to singletons, and their minimal path are naturally *circular*. The Globally Optimal Circular Shortest Path is summarized in the algorithm 2.

The low cost complexity (for width u and height v, $O(u^{1.6}v)$ average time, or less than a millisecond for 15x15 pixels cross section profile, see figures 4.10 and 4.11) makes this method very attractive for repetitive testings, such as the particle filters presented in section 4.3.3. The metric g in equation (4.15) is equal to the inverse of the image gradient magnitude.

Data: vessel cross-section I
Compute I inverse gradient and its log-polar transform
Initialize $shortest\_part = [\,]$
Initialize $S_1 = \{[1, v/2]\}$ and $S_2 = \{[v/2 + 1, v]\}$
while $shortest\_part$ *is not circular* **do**
  Compute $shortest\_path1$ for S_1 using Dijkstra's algorithm [63]
  Compute $shortest\_path2$ for S_2 using Dijkstra's algorithm [63]
  if $shortest\_path1$ *shorter than* $shortest\_path2$ **then**
    Partition S_1 into S_1 and S_2
  else
    Partition S_2 into S_1 and S_2
  end
end
Result: circular $shortest\_part$ over S

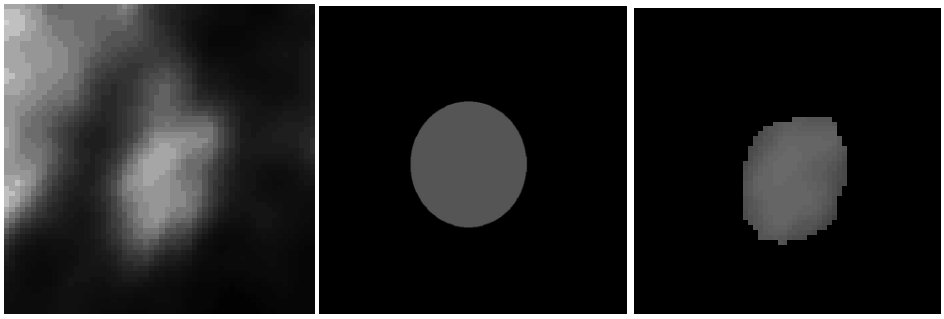
Algorithm 2: - General pseudo-code for the CSP.



(a) (b) (c)

Fig. 4.10: (a) Vessel cross section, (b) cross-section's log-polar transform and minimal weight circular path, (c) CSP in Cartesian view

The CSP algorithm is an efficient technique for image segmentation for closed structures under the assumption that a point is given in the structure interior to compute the log-polar image. One assumes that the maximum a posteriori of the centerline position's pdf provides the start point to the CSP at each step.



(a) (b) (c)

Fig. 4.11: (a) Vessel cross section, (b) elliptic models only roughly represent the cross-section, (c) therefore, CSP's end result is needed for the final segmentation.

4.4 Experimental Validation

4.4.1 Image Modality and Algorithm Specifications

The algorithm was tested on 34 CT images coming from different scanners (SOMATON Emotion, SOMATON Sensation 16 and 64 and SOMATON Definition) and different patients who presented different or no pathologies. A typical voxel resolution is $0.3\text{mm} \times 0.3\text{mm} \times 1\text{mm}$. Contrast agent was used for all images, with different concentration and different products. Table 4.4.1 summarizes the typical intensity range for different tissues, as they are found in a CT angiography volume, with pixels' value coded on 12 bits. No preprocessing is applied before the segmentation procedure described in this article.

	myocardium	vessel lumen / ventricles	calcification	lungs
intensity	900-1100	1100-1300	1400-2000	0-200

Tab. 4.1: Pixel intensity range for different organs coded on 12 bits.

Regarding the initial configuration, the use of approximatively $1,000$ particles gave sufficient results for our experiments. We performed a systematic resampling according to the SIR every time the effective sampling size $N_{eff} = \sum_i 1/w_i^2$ (where w_i is the weight of the ith particle) falls below half the number of particles. As mentioned in Section 4.3.3, the preference for SIR, compared to Stratified Resampling [69], is motivated by the robustness of the segmentation. The *reinforced SIR* strategy exposed in section 4.3.3 gives better results, for a constant number of particles. The tracking stops when the sum of image measures at a given iteration falls below a given threshold.

4.4.2 Comparison with Front Propagation

For comparison purposes, our method is compared with Front Propagation, implemented using the Fast Marching algorithm [59], based on a curvilinear structures detection [163]. The Hessian analysis is used to detect tubular structures; this measure (called "vesselness" in [77]) is integrated into a potential map on which the Fast Marching algorithm is run such as in [61]. Two high and one low amplitude eigenvalues in the Hessian indicate local tubular structure. In few words, Front Propagation computes isosurfaces in a Riemanian space, whose metric is based on the image: the vesselness measure in our case. The front propagates faster along the vessel than in other non-tubular structures. However, in the case of inhomogeneities (either pathologies or bifurcations), this vesselness measure drops and the front propagation either stops or leaks into neighboring structures.

In the synthetic case, the error measure Δ is defined as the symmetric difference between ground truth G and segmentation S:

$$\Delta = 1 - \frac{2\,|G \cap S|}{|G| + |S|}.$$

Since ground truth is not available for the real case studies, an expert visually validates the number of branches correctly segmented and missed.

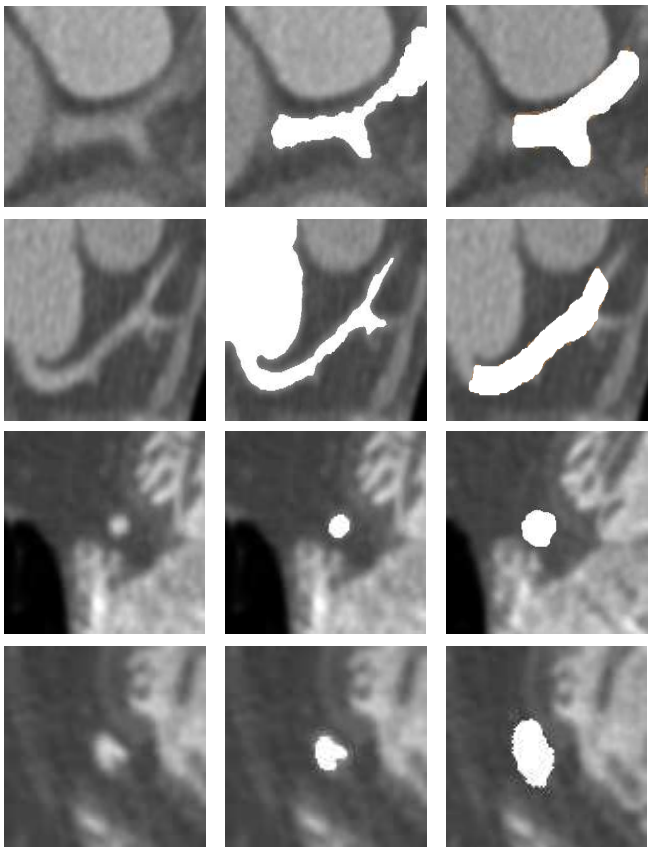


Fig. 4.12: Comparison between hand-labeled ground truth (b) and particle filters results (c) for four short sections. The final segmentation is obtained with CSP, using the centerline provided by the maximum a posteriori.

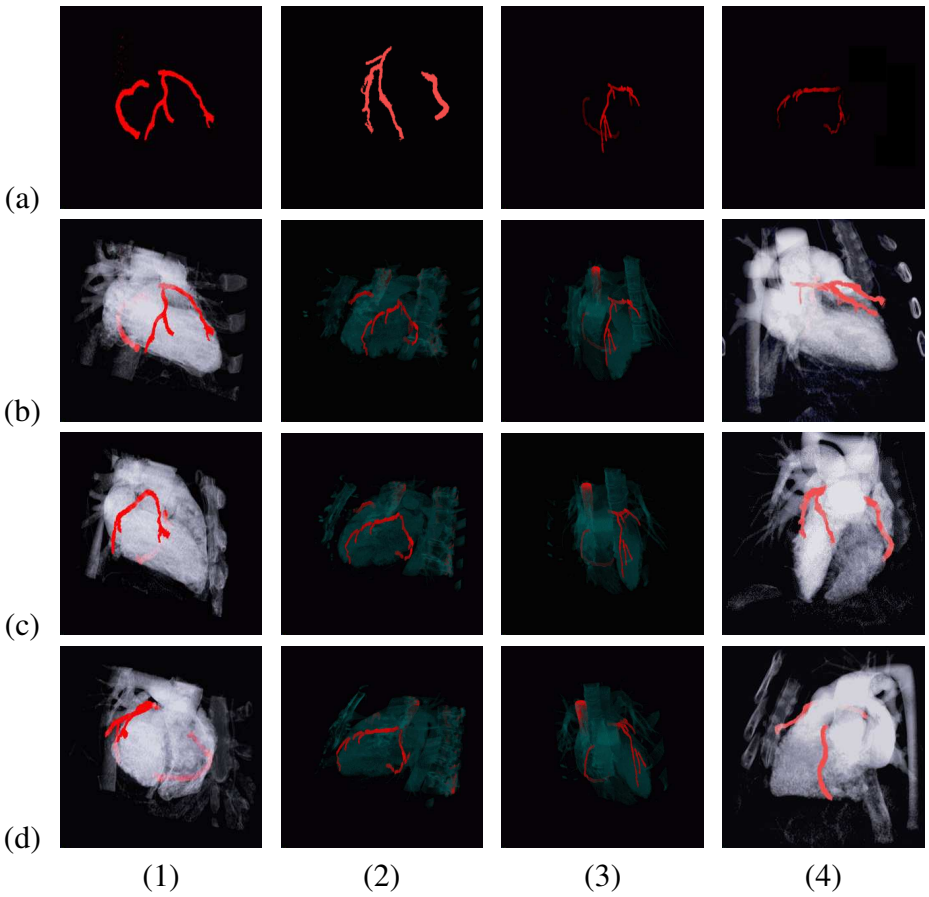


Fig. 4.13: Segmentation of the Left anterior descending coronary artery and Right coronary artery in CTA (in red) for four patients ; (a) coronary tree, (b,c,d) Different $3D$ views super-imposed to the cardiac volume are presented.

vessel name	RCA	Acute Marginal	LAD	First Septal	LCX	Obtuse Marginal
# cases missed, using PF	none	5	none	2	none	2
# cases missed, using FP	12	28	16	23	21	26

Tab. 4.2: Results table showing the number of cases for which branches are incorrectly segmented, over a dataset of 34 patients, using Particle Filters (PF) and Front Propagation (FP), with respect to expert ground truth.

4.4.3 Results

Validation is a difficult part for any coronary segmentation method. The algorithm has been evaluated on 34 patients, and has successfully recovered all the main arteries (RCA, LAD, LCX) for each patient as shown in table 4.2, while a small portion of visual results are also presented in figure (4.13).

The results in table 4.2 corresponds to the number of branches segmented by particle filters and identified by a human expert. For comparison purposes, the same test is performed using Front Propagation based on the image Hessian matrix [163]. These results were achieved with a one-click initialization; the Hessian's eigenvector that corresponds to the smallest eigenvalue gives the approximative initial direction. All patients present some kind of artery pathologies in one, at least, of their coronary vessels (12 cases with calcification, 8 stenosis, 4 stent, 2 bypasses), and many present intensity artifacts (7 stepping, 5 hardening effects). Our approach has successfully segmented both healthy and unhealthy coronaries without leaking into neighboring structures (over-segmentation). The method seems to outperform regarding the detection of the main branchings, while in some cases branching of lower clinical importance at the distal part of the tree have been missed. However, current studies focus on the issue of branchings for narrow vessels in very low contrast conditions. The comparative study demonstrate the proposed framework has the ability to outperform a deterministic Hessian based method in cases with corrupt data (pathologies). Figure (4.12) displays the results for a short section of coronary with corresponding hand-labeled ground truth.

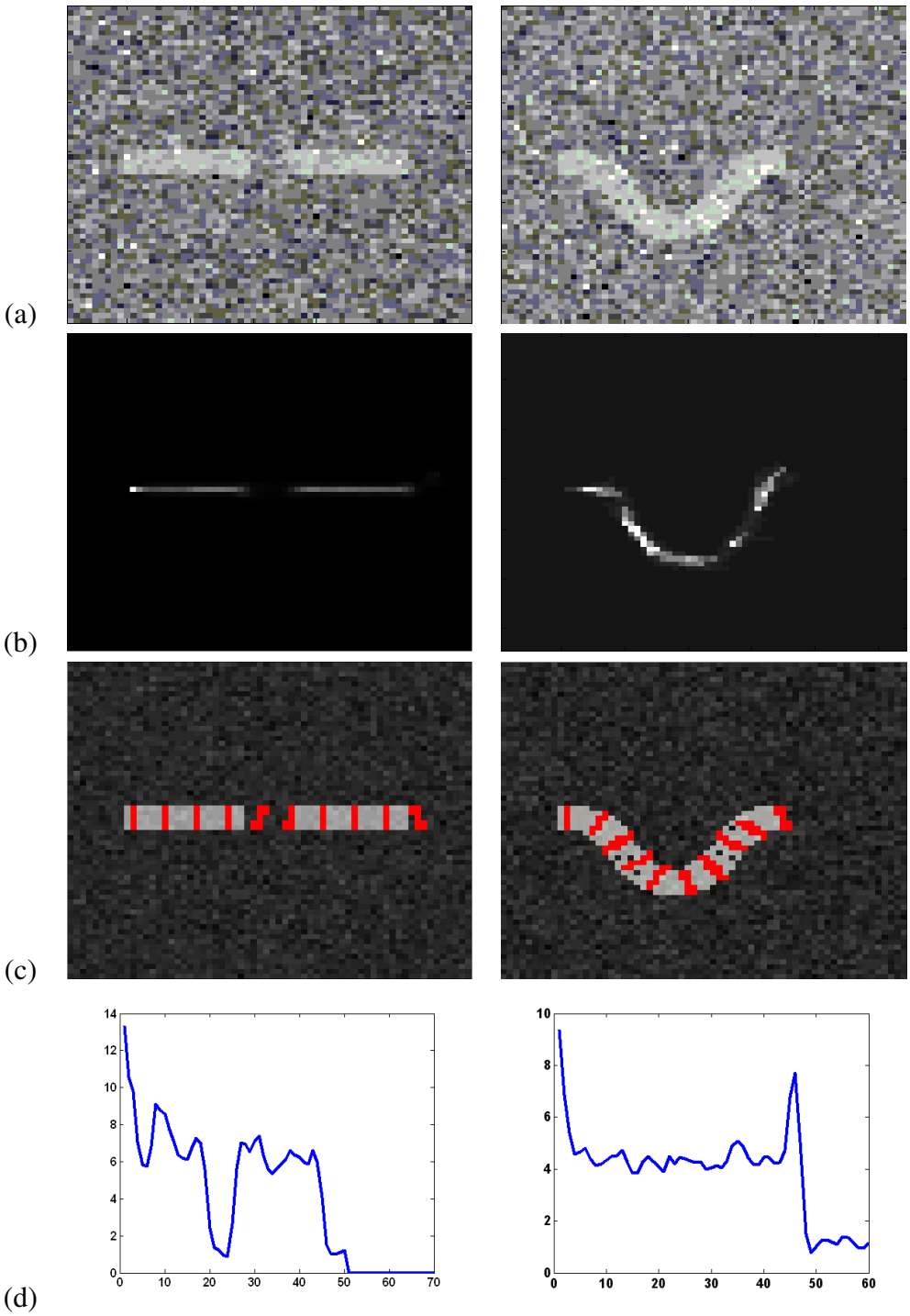


Fig. 4.14: 2 synthetic examples: (left) with total lumen obstruction, (right) with high curvature. (a) Original image. (b) Estimation of the pdf. (c) Cross-sections overlaid on the image. (d) Sum of particles' weight over the steps.

Radius of curvature	3.5	7	10.5	13.3	14
symmetric difference	7.80%	9.29%	9.24%	12.11%	11.07%
Gap width	4	6	8	12	16
symmetric difference	0	2.22%	7.11%	11.56%	24.44%

Tab. 4.3: Results table showing the symmetric difference between maximum likelihood segmentation result (particle with maximum probability at each step), and ground truth on synthetic data.

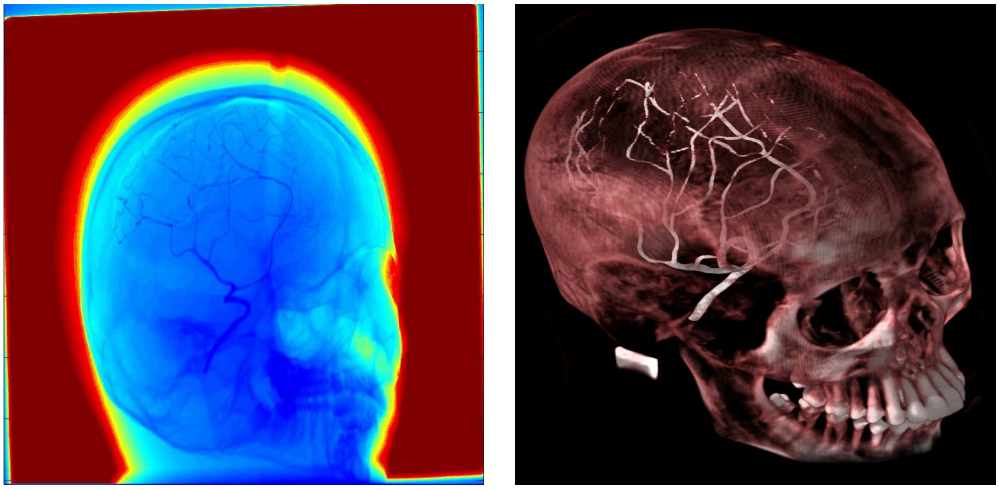
A second test is performed on a 2D synthetic dataset (see figure (4.14)) with Gaussian pixel intensity distributions (vessel: mean = 1200, standard deviation = 50; background: mean = 1000, standard deviation = 100) noisier and with less contrast than real cases. Using real cases radiuses (3-4 pixels lumen diameters), and using the segmentation provided by the particle of maximum weight (the CSP is not used for the CSP's performance not to interfere with the particle filter's), the symmetric difference between segmentation and ground truth is 11.07% for a tubular structure with a 14 pixels radius of curvature, and 2.22% for a 6 pixels wide simulated lumen obstruction. The results in table 4.3 displays the symmetric differences between ground truth and segmentation. In all these cases, the tracking procedure never jumped completely out of the vessel into the background. The algorithm is stopped when the sum of the particles' weight falls below a threshold (see figure (4.14).

4.5 Comparison between Particle Filters and Deterministic Approaches for 2D-3D Registration

Let us consider a problem that is ill-posed in the general form: 2D-3D registration on corrupt images between pre-operative digitally subtraction angiography (DSA) and intra-operative fluoroscopic images.. 2D-3D registration consists in finding the limited number of parameters that define the perspective projection from a given 3D volume to a 2D image (see figure (4.15)). A perspective projection from a 3D homogeneous point P to a 2D homogeneous point p is defined with matrix as

$$p = \mathbf{PCT}P, \tag{4.17}$$

where $\mathbf{T}$ is a 4×4 matrix of the pose relating the pre-operative (3D) coordinate frame to the iso-centric coordinate frame of intra-operative imaging device. $\mathbf{C}$ is a 4×4 matrix defining the transformation between the iso-centric coordinate frame and the coordinate frame centered at the



(X-ray) (CT)

Fig. 4.15: Contrast enhanced X-ray and CT image of right hemisphere vessels.

x-ray imaging source, which also depends on gantry angles. $\mathbf{P}$ is a 3×4 projection matrix defining the projection cone related to the source-detector geometry. We assume $\mathbf{P}$ and $\mathbf{C}$ are known from a calibration step. $\mathbf{T}$ is the unknown pose that encodes the three translation and three rotation parameters. The problem thus consists in optimizing $\mathbf{T}$, with respect to a pose fitness measure.

4.5.1 Pose fitness measure

Intensity based registration approaches require digitally reconstructed radiographs (DRR) generation, which is then compared with the x-ray or fluoro images at each step of the process. The main bottleneck for intensity based approaches is speed. In [205], the authors mentioned several minutes for registering the whole intensity volume, most of the time spent on generating the DRR. For this application, we use the vascular structures as features for registration. Vascular structures are easy to segment in most of cases, sparse enough for the registration to be faster than intensity-based methods, but yet generally well distributed throughout the organs to capture potential misalignments. One obstacle here is to define a fast and robust measure that characterizes the fitness of the pose. The proposed method uses a prior segmentation of the vessels in the 3D volume of interest and the 2D image. From the 2D segmentation result, a distance map [73] is computed (see figure (4.16)). For a given pose matrix $\mathbf{T}$, the measure of fitness is the sum of distances (i.e. D) to the

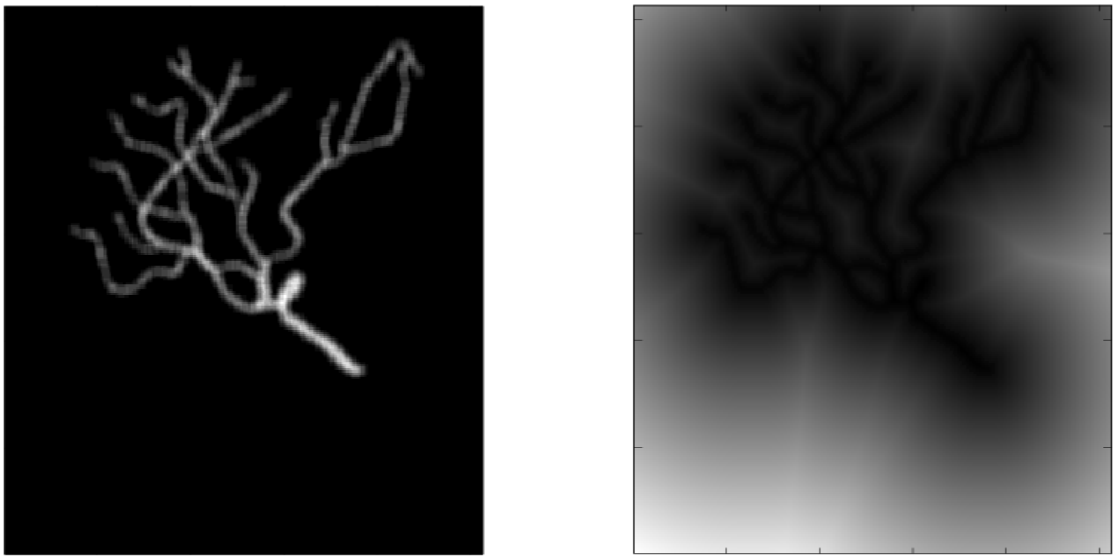


Fig. 4.16: The projected vascular structure and the distance map associated with it.

2D structure for each projected 3D point from the segmented vessel tree:

$$F(\mathbf{T}) = \int_{P \in vessel} D(\mathbf{PCT}P). \tag{4.18}$$

To optimize $\mathbf{T}$, gradient descent is a very common technique [205][73]. However, it is very sensitive to initialization; therefore, in the case of highly corrupted data (see figure (4.17)), gradient descent often fails due to the lack of image support as well as the non-convex cost function. One overcomes these limitations through a multi-hypothesis framework, such as condensation [98][67], presented in section (4.3.3). In this problem, a particle x_t is a vector of pose parameters that define a particular pose matrix $\mathbf{T}$, and the observation z_t is the projection of the vascular tree onto the fluoroscopic image. Each pose matrix (particle) is associated to its fitness measure $F(\mathbf{T})$ defined in equation (4.18); this fitness measure plays the role of conditional probability $p(z_t|x_t)$ once normalized:

$$p(z_t|x_t) \propto e^{-F(\mathbf{T})/F_0}.$$

Then, the SIR algorithm presented in section (4.3.3) is applied: the hypothesis are sampled according to their fitness measure so that hypothesis with high measure are neglected, and those with low measure are selected several times. A random vector is then added to the newly selected hypothesis, and the process is iterated until no better hypothesis can be found, or the maximum number of iteration is reached.

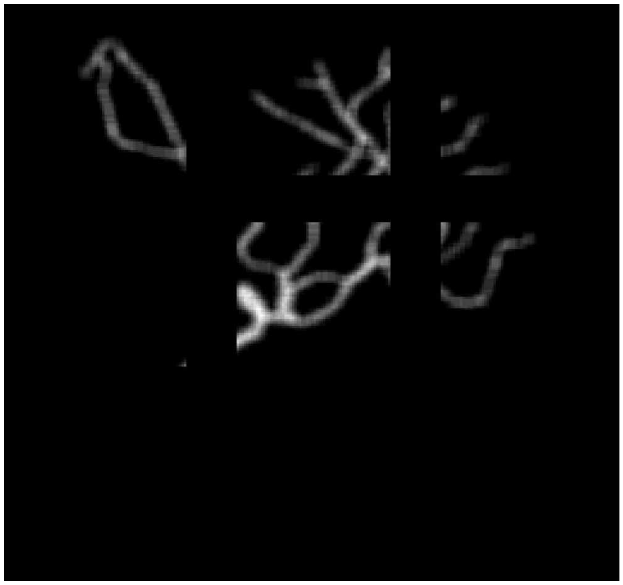


Fig. 4.17: Corrupted X-ray image

First, experiments were conducted to compare the method presented in this chapter with a purely random search, and the results proved that condensation used with x particles during z time steps reach a better registration than $x \times z$ random trials (in the experiments, x=64 and z=100).

4.5.2 Multiple Hypotheses Testing Evaluation in the Context of 2D/3D Registration, Comparison with Levenberg-Marquardt Gradient Descent

Second, experiments were conducted to compare condensation performance (with respect to the measure presented in section (4.5.1)) with classic nonlinear least square methods to find the minimum of a nonlinear function: Levenberg-Marquardt [8] and Gauss-Newton algorithms. Both Levenberg-Marquardt and Gauss-Newton highly depend on initialization, and never led to better results than condensation (over 100 experiments were made with synthetic portal images from real pre-operative volumes, and random patient poses).

The perspective projection may not be perfect, and for optimal parameters, the subtraction of the two registered images may not be exactly null in real cases. For that matter, white noise has been added to the portal image to test the sensibility of the two algorithms, condensation and gradient descent. The results are presented in figures (4.18) and (4.20).

Tests have been performed about the method sensitivity to segmentation error, and the conclu-

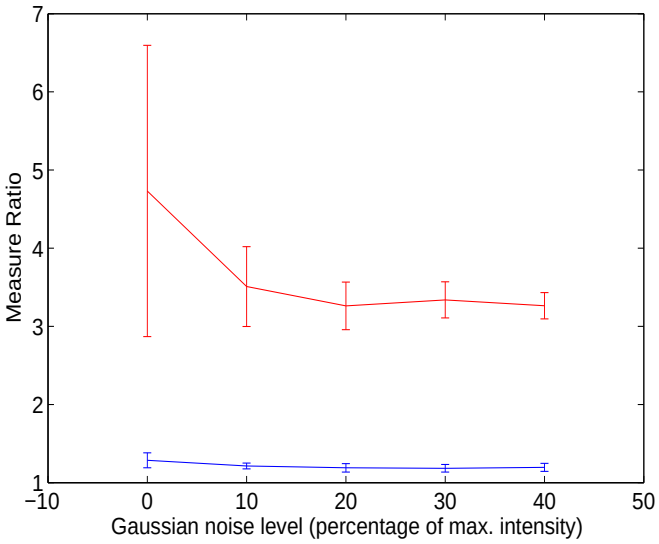


Fig. 4.18: Comparison of condensation (blue) and Levenberg-Marquardt (red) performances when Gaussian noise is introduced in the portal image

sions are presented in figure (4.19). Whereas the gradient descent performances decrease as the segmentation error increases (in Hounsfield units), the condensation results do not vary. As the error level increases, the global minimum's basin of attraction (for gradient descent) diminishes; consequently, the probability for the gradient descent to be correctly initialized diminishes. Since the particles are uniformly initialized in any case, the results are independent of the segmentation error.

For a similar reason, condensation is independent of the capture range, while any Gauss-Newton / gradient descent method inherently depends on the global minimum's basin of attraction width. Furthermore, when features are missing, local minima are created, which are likely to attract the gradient descent method. Particle filters still converge toward the global minimum (see figure (4.17)).

One concludes from this experiment that models of uncertainty may be used to represent a problem's unknown for global optimization. Some hypothesis of lower interest with respect to equation (4.18) survive randomly and are driven toward the global minimum. In this sense, it partially answers the comments made in chapter 2 section (2.8). However, as the problem described here in section (4.1) is essentially static, particle filters resemble simulated annealing. The following section tackles the more difficult problem of segmentation of coronary arteries, using particle

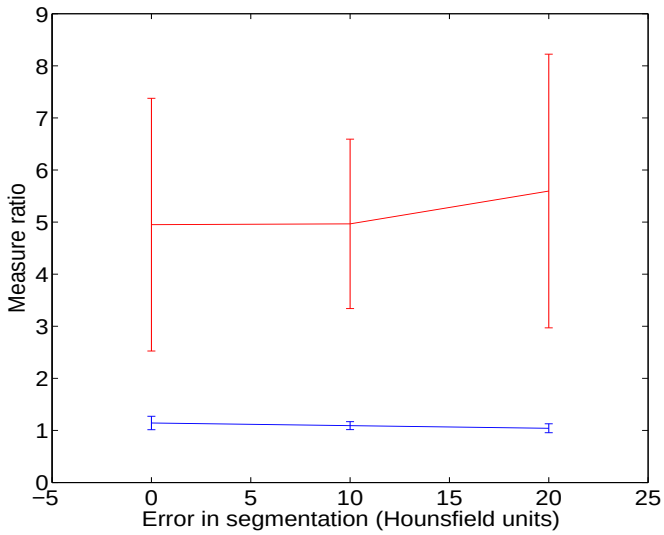


Fig. 4.19: Comparison of condensation (blue) and Levenberg-Marquardt (red) performances when segmentation error is introduced

filters for a progressive segmentation: the model of uncertainty at one point of the segmentation is used to predict the statistical distribution of the artery at a later point.

4.6 Conclusion

In this section, we have shown that Monte-Carlo sampling and multiple hypotheses testing can be used for the segmentation of tubular structures as well as for 2D/3D registration. In the context of vascular segmentation, particle filters sequentially estimate the pdf of segmentations in a particular feature space. The case of coronary arteries was considered to validate such an approach where the ability to handle discontinuities on the structural (branching) as well as appearance space (calcifications, pathological cases, etc.) was demonstrated. The main advantage of such methods lies in their capability to handle intensity inhomogeneities from pathologies and bifurcations. Experiments were conducted on several healthy and diseased patients CTA data sets, segmenting the *Left Main Coronary Artery* and the *Right Coronary Artery* figure (4.13). Another interesting application concerns MR Angiography, as motion artifacts make most deterministic vessel segmentation methods impossible to apply.

As a final remark, it may be underlined that particle filters require heavy computational time in

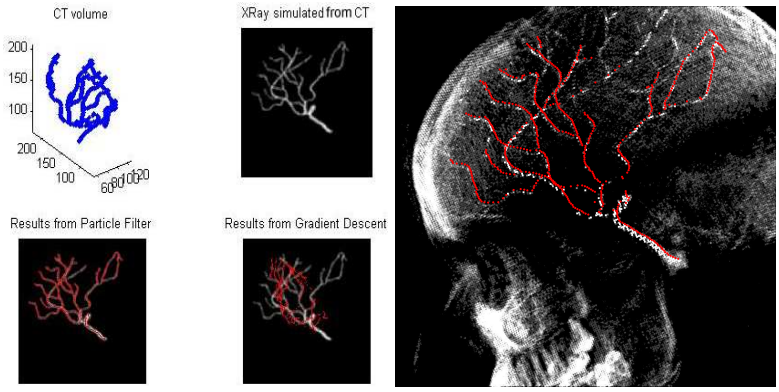


Fig. 4.20: (L) Comparison between Condensation and Levenberg-Marquardt method performance. (R) Registered vascular structure with 2D projection simulated from original CT data

a general context (around an hour for a CT $512\times512\times300$ volumetric image), compared to front propagation. However, in the case of non linear problems, with missing or corrupt data, particle filters provide a better segmentation than deterministic methods. In the particular case of coronary arteries segmentation, due to pathologies, contrast agent heterogeneities and branchings, the use of non deterministic methods have been proved successful. Therefore, when time is a constraint, a compromise is to be found to apply deterministic methods in linear cases and statistical modelization, such as particle filtering, in all other cases.

For the problem of 2D/3D registration, the bootstrap algorithm performs better than Levenberg-Marquardt and Gauss-Newton algorithms, which demonstrates the advantages of using a multi-hypothesis framework. Furthermore, it compares favorably with simulated annealing both in terms of computational time and robustness since the multi-hypothesis framework covers the full feature space in the initialization. Also, if the same problem were to be considered in a temporal sequence, it would naturally provide at convergence an uncertainty model that could be integrated as a prior for the next time step.

Introducing further prior knowledge in the segmentation process is a future direction. One can see such a contribution in two parallel paths. First, building better models that account for the appearance of the vessel seems to be a necessity toward capturing the coronaries at the lowest parts of the vessel tree. The current model is based on the global statistics of the appearance of the vessel and one can claim is a meaningful measure for vessel cross-sections with a certain area. Another extension of this method would include an implicit way to detect branchings during the resampling step, and would avoid using any particle clustering or thresholding the distance

between the clusters. Finally, other non deterministic sampling strategies could reduce the number of particles while keeping the accuracy of the results.

Chapter 5

Conclusion

This thesis intends to motivate the use of models and partial elements of information to extract the desired content of an image. The reconstruction of the global solution may either be performed by the mean of model-guided interpolation (chapter 2), or may be performed in a progressive way (chapter 3 and 4). When the solution is built progressively, the transition step from one state of the solution to the next is either guided by a stationary or a non-stationary model, or the state of the solution is represented by an uncertainty model propagated through time.

Chapter 1

The first chapter has reviewed the literature on segmentation and tracking. It has been shown that model free and low level methods (e.g. based on simple observations such as pixel intensities and image gradients) are of particular interest when the solution cannot be constrained with statistics from prior learning. However, due to the increasing number of images acquired every year, the statistical models built from prior learning are becoming more and more reliable. Thus, the techniques to constrain the solution with shape or appearance statistics have gained popularity in the past decades. Certain optimization methods developed in the context of model-free techniques, such as the level-sets and the graph cuts, have been adapted to include statistical models learned on-the-fly or from prior learning. However, the statistics included in the framework drive the optimization procedure toward a trade-off between prior learning and image information and the resulting solution may poorly fit both the prior and the input image. The uncertainty model proposed in chapter 4 bypasses this problem by estimating the statistics directly from image information in a non parametric fashion. Furthermore, while the reliability of the prior learning

statistics is clearly modeled, the image input is assumed to perfectly support the solution which is not the case in most computer vision problems. In chapter 2, we proposed a model that explicitly accounts for inhomogeneities of the image support reliability. Last but not least, the prior art in dynamic models is presented in chapter 1 and exposes two main limitations: first the dynamic models learned from prior knowledge are static in the sense that they are not adapted to new information provided by the particular case of study. Furthermore, in many sequential problems, the whole sequence of images is available. Therefore, instead of using a dynamic system to model the motion on a frame-to-frame basis, the dynamic system may be used to model and segment the whole sequence at once. This framework is developed in chapter 3.

Chapter 2

The second chapter has introduced a novel method that partitions an input image into different regions and associates a reliability measure to them depending how the image regions support the solution to the segmentation problem. This allows the automatic selection of few image regions that are more reliable than others and from which the whole segmentation solution is recovered using a model driven reconstruction scheme. This procedure is applied to three different problems of computer vision. The first problem -segmentation in volumetric images- is developed in order to prove sparse information models effectively decrease problems' dimensionality compared to common dimensionality reduction techniques such as PCA. Sparse information models are then integrated into a segmentation scheme for the case of three dimensional liver dataset. The second problem -object tracking in time sequences- proves this method is robust to large variances and the reconstruction is roughly similar to human expert's. A comparison between this method and a method based on autoregressive models is presented in the following chapter. The third and last problem -surface reconstruction from laser pointed distances- demonstrates how well this method performs in the case of salt and pepper noise.

Chapter 3

The third chapter has presented the use of autoregressive (or AR for short) models for segmentation. The main contribution was the use of on-the-fly adapting, non-stationary, AR models for tracking; however, stationary models were also introduced in the context of a regular (i.e. stationary) cardiac ultrasound sequence, and may even be used to solve static problems such as volumetric segmentation. Tracking highly deforming structures in space and time arises in numerous applications in computer vision. Static models are often referred to as linear combinations of a mean

model and modes of variations learned from training examples. In dynamic modeling, the shape is represented as a function of shapes at previous time steps. In this chapter, we have introduced a novel technique that uses the spatial and the temporal information on the object deformation. Tracking is reformulated as a high order time series prediction mechanism that adapts itself on-line to the newest results. Samples (toward dimensionality reduction) are represented in an orthogonal basis and are introduced in an auto-regressive model that is determined through an optimization process in appropriate metric spaces. Toward capturing evolving deformations as well as cases that have not been part of the learning stage, a process that updates on-line both the orthogonal basis decomposition as well as the parameters of the autoregressive model is proposed. Promising experimental results in tracking explicit shapes in a video sequence that could be used to impose prior knowledge in segmentation are presented.

Chapter 4

The fourth and last chapter has introduced a novel approach in the use of uncertainty models to solve computer vision problems. In the context of tubular structures segmentation, the multi-hypothesis framework is introduced and implemented with a sequential random sampling procedure called particle filters. First, it is shown that the global solution to this problem cannot be achieved by sequentially optimizing the local most likely solution. Then, the multi-hypothesis framework is introduced and implemented with a sequential random sampling procedure called particle filters. The structures are modeled by a template of few parameters and segmented in a sequential fashion by randomly sampling the feature space and assigning a probability measure to each sample based on shape and appearance. This probability field is then propagated geometrically along the vessel. The resulting solution performs better than other methods based on front propagation and Hessian analysis. Thanks to the multi-hypothesis framework the pathologies, acquisition artifacts and branchings are successfully segmented. Then, in order to compare this framework with commonly used gradient descent approaches, it is extended to a static parametric problem: 2D-3D registration.

Several questions remain open, the first of which is how the sparse information models may be generalized to bypass the need for one or several axis of discretization. Furthermore, since the autoregressive models provide a statistical linkage between the states of the solution at different time, the integration of AR models to the framework of particle filtering seems apropos. The last questions concern practical applications to medical imaging and future perspective of research in that field. There is a trend in the recent studies that tend to demonstrate early diagnostic has a

limited impact on survival rate in a number of pathologies such as certain types of cancer and, noticeably, lung cancer [7]. For these types of pathology the main medical contribution remain therapy. This means that the need for therapy planning software will significantly increase in the near future. Furthermore, expert systems are being developed to assist the clinicians in the decision process, from drug prescription to complementary testings and surgery. These expert systems require statistical models and decision processes that rely only on partial measurements. Measure theory and sparse information models are therefore to be extended into these fields.

Appendix A

Volume Interpolation

A.1 Prior Art on Volume Interpolation

Interpolation models have been studied before. The simplest and most common method is to use a spline or piecewise polynomial function [149, 178] that interpolates the contour between explicit points. Other methods use an implicit representation of the contour (a continuous function that takes a zero value on the contour) and interpolating functions such as thin-plate splines [185]. An example of surface reconstruction is the work of Hoppe *et al.* [91] who computed a signed distance function in 3D which is the distance in $\mathbb{R}^3$ to any input point. Then, from the zero levelset of this function is extracted the surface using the marching cubes [119]. At last, deformable models [116] are used to minimize an energy function of the mesh by deforming the mesh so that the mesh is simultaneously attracted to the data by an image term, kept smooth by a tension term and by an optional prior term.

A.2 Comparative Study for Different Interpolation Models

Table (A.1) resumes the quantitative comparison between the different methods for the same benchmark of hand-segmented livers, and the same key contours regularly sampled between the bottom and top slices. The contours on the axial plane are represented a the zero levelsets of a distance function, and the average distance between the ground truth surface and the reconstructed surface is reported in table (A.1). It clearly motivates a preference for nonlinear interpolation. The

nonlinear interpolation, also called generalized linear model [66], consists in decomposing each contour independently as a linear combination of the key contours. Although it seems like a very straightforward interpolation procedure, it is the one that gave the best experimental results.

modelization method	Piecewise Cubic Spline	Linear Interpolation	Non Linear Model	PCA on features curve
Volume 1	6.53e-2	1.84e-1	2.58e-3	3.38e-3
Volume 2	2.87e-2	2.85e-2	6.24e-2	4.51e-2
Volume 3	1.44e-6	5.65e-2	2.21e-4	5.03e-2
Volume 4	1.95e-4	2.47e-2	2.13e-3	1.63e-1
Volume 5	4.15e-2	2.20e-2	3.35e-2	2.50e-3
Volume 6	1.84e-1	2.60e-2	2.05e-4	9.02e-3
Volume 7	2.10e-3	3.02e-2	6.00e-3	3.23e-2
Volume 9	2.07e-2	1.01e-2	4.77e-5	4.85e-2
Volume 10	4.58e-2	1.52e-1	4.11e-2	2.98e-2
Volume 11	9.75e-2	3.73e-2	4.33e-6	3.73e-2
Volume 12	4.54e-2	8.47e-2	4.24e-3	1.22e-1
Volume 13	3.78e-1	7.31e-2	1.12e-2	1.07e-1
Volume 14	3.36e-2	1.29e-2	2.55e-2	1.21e-5
Volume 15	5.67e-2	1.76e-1	1.52e-2	6.20e-3
Volume 16	4.33e-2	2.41e-1	2.81e-2	8.60e-2
Volume 17	2.10e-3	4.90e-3	3.26e-2	6.27e-3
Volume 19	2.04e-1	6.46e-2	5.19e-4	1.04e-2
Volume 20	1.52e-1	1.56e-2	1.74e-2	2.71e-6
Volume 21	1.52e-1	5.70e-3	2.34e-3	1.24e-2
Volume 22	9.66e-2	1.50e-1	1.06e-2	3.57e-3
Volume 24	1.54e-1	2.99e-3	2.06e-2	3.87e-2
Volume 25	1.39e-1	1.31e-1	3.08e-3	1.43e-2
Volume 26	3.22e-2	1.20e-6	2.10e-4	8.38e-2
Volume 27	9.10e-2	1.75e-3	2.89e-3	4.28e-2
Volume 28	4.73e-2	5.04e-2	2.87e-2	1.29e-1
Volume 29	1.62e-3	2.74e-3	4.08e-4	1.31e-1
Volume 30	1.53e-2	2.26e-4	3.84e-2	2.70e-1
Volume 31	5.37e-2	1.52e-5	3.16e-3	5.90e-3
Volume 32	1.54e-2	7.34e-2	2.67e-6	4.42e-2
Volume 33	3.56e-2	5.90e-7	1.29e-2	4.58e-2
Volume 34	1.88e-4	1.29e-2	3.85e-3	8.36e-2
Mean	7.21e-2	5.40e-2	1.32e-2	5.37e-2
Standard Deviation	8.15e-2	6.55e-2	1.59e-2	6.05e-2
Maximum Deviation	3.78e-1	2.41e-1	6.24e-2	2.70e-1
Minimum Deviation	1.44e-6	5.90e-7	2.67e-6	2.71e-6

Tab. A.1: Results table showing the average distance (in mm) between modeled contours and ground truth contours, for different types of models. Cubic Spline and Linear Interpolation are purely interpolation models, while Non Linear and PCA are trained over the whole dataset.

Appendix B

Partial Derivative Equations, for equations (3.13) and (3.15)

Equation (3.13) states the image support energy with respect to the contour $\mathbf{Y}_t$, by computing the Kullback-Leibler distance between the observed pixels intensity histograms inside (h_{in}) and outside (h_{out}) the contour and the a priori learned histograms p_{in} and p_{out}:

$$\begin{aligned} E_{\text{sup}}(\mathbf{Y}_t) = & \int h_{\text{in}}(s, \mathbf{Y}_t) \log \left(\frac{h_{\text{in}}(s, \mathbf{Y}_t)}{p_{\text{in}}(s)} \right) \, ds \\ & + \int h_{\text{out}}(s, \mathbf{Y}_t) \log \left(\frac{h_{\text{out}}(s, \mathbf{Y}_t)}{p_{\text{out}}(s)} \right) \, ds. \end{aligned} \tag{B.1}$$

The contour $\mathbf{Y}_t$ is represented by an implicit function on which PCA is performed. The expression of this PCA is given here:

$$\mathbf{Y}_i = \mathcal{A}(\overline{\mathbf{Y}}) + \sum_{q=1}^{m} x_q \mathbf{U}_q, \tag{B.2}$$

where the $\mathbf{U}_q$s stand for the modes of variation (Eigenvectors of the covariance matrix), the x_qs are the PCA coefficients, $\overline{\mathbf{Y}}$ is the mean contour and $\mathcal{A}(\overline{\mathbf{Y}})$ is an affine transformation performed on the mean contour. A low-dimension parametric representation of the contour is thus given by the PCA coefficients and the affine transformation parameters.

The partial derivative of equation (3.13) with respect to the PCA coefficient is computed thanks

to the chain rule:

$$\begin{aligned}\frac{\partial E_{\text{sup}}}{\partial x_q} = &\int \frac{\partial h_{\text{in}}}{\partial x_q} \log\left(\frac{h_{\text{in}}(s,\mathbf{Y}_t)}{p_{\text{in}}(s)}\right)\left(\frac{h_{\text{in}}(s,\mathbf{Y}_t)+p_{\text{in}}(s)}{p_{\text{in}}(s)}\right) ds \\ &+ \int \frac{\partial h_{\text{out}}}{\partial x_q} \log\left(\frac{h_{\text{out}}(s,\mathbf{Y}_t)}{p_{\text{out}}(s)}\right)\left(\frac{h_{\text{out}}(s,\mathbf{Y}_t)+p_{\text{out}}(s)}{p_{\text{out}}(s)}\right) ds\end{aligned} \tag{B.3}$$

From an implementation point-of-view, $\frac{\partial h_{\text{in}}}{\partial x_q}$ in equation (B.3) is easily computed by considering it reflects the change of the histogram for an infinitesimal change of the contour along the PCA q-th mode of variation. Mathematically, for a particular pixel intensity s, this is expressed by

$$\frac{\partial h_{\text{in}}}{\partial x_q}(s) = \frac{\int_\Omega \psi_q(p)\delta(\psi(p))\delta\left(I(p)-s\right) dp}{\int\int_\Omega \psi_q(p)\delta(\psi(p))\delta\left(I(p)-s\right) dp\, ds}.$$

where ψ stands for the distance map associated to the contour, and ψ_q stands for the function associated to the q-th mode of variation $\mathbf{U}_q$ in the image domain Ω.

The derivation of equation (3.15) is straightforward:

$$\begin{aligned}E_{\text{predict}} &= \left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t\right)^T \mathbf{\Gamma}^{-1}\left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t\right) \\ \frac{E_{\text{predict}}}{\partial x_q} &= \frac{\partial \mathbf{Y}_t}{\partial x_q}^T \mathbf{\Gamma}^{-1}\left(\mathbf{Y}_t - \hat{\mathbf{Y}}_t\right).\end{aligned}$$

The derivatives with respect to the affine transformation parameters are obtained the same way.

BIBLIOGRAPHY

[1] A. Agarwal and B. Triggs. Tracking articulated motion using a mixture of autoregressive models. In *European Conference on Computer Vision*, volume 3, pages 54–65, 2004.

[2] A. Agarwal and B. Triggs. Tracking articulated motion using a mixture of autoregressive models. In *European Conference on Computer Vision*, pages III 54–65, Prague, May 2004.

[3] B. Appleton and C. Sun. Circular shortest paths by branch and bound. *Pattern Recognition*, 36(11):2513–2520, November 2003.

[4] N. Armande, P. Montesinos, O. Monga, and G. Vaysseix. Thin nets extraction using a multi-scale approach. *Computer Vision and Image Understanding*, 73(2):248–257, 1999.

[5] S. Arulampalam, S. Maskell, N. Gordon, and Tim Clapp. A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 50:174–188, 2002.

[6] B. Avants and J. Williams. An adaptive minimal path generation technique for vessel tracking in CTA/CE-MRA volume images. In *Medical Imaging Computing and Computer-Assisted Intervention*, pages 707–716, 2000.

[7] P. B. Bach, J. R. Jett, U. Pastorino, M. S. Tockman, S. J. Swensen, and C. B. Begg. Computed tomography screening and lung cancer outcomes. *Journal of American Medical Association*, 297:953–961, March 2007.

[8] D. M. Bates and D. G. Watts. *Nonlinear regression and its applications*. New York, Wiley, 1988.

[9] N. Becherer, H. Jödicke, G. Schlosser, J. Hesser, F. Zeilfelder, and R. Männer. On soft clipping of Zernike moments for deblurring and enhancement of optical point spread functions. In C. A. Bouman, E. L. Miller, and I. Pollak, editors, *Computational Imaging IV. Edited by*

Bouman, Charles A.; Miller, Eric L.; Pollak, Ilya. Proceedings of the SPIE, Volume 6065, pp. 73-83 (2006)., pages 73–83, February 2006.

[10] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[11] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[12] D. P. Bertsekas. Incremental least squares methods and the extended kalman filter. *SIAM Journal on Optimization*, 6(3):807–822, 1996.

[13] J. Besag. On the statistical analysis of dirty images. *Journal of Royal Statistics Society*, 48:259–302, 1986.

[14] Charles Beumier and Marc Acheroy. Face verification from 3d and grey level clues. *Pattern Recognition Letters*, 22(12):1321–1329, 2001.

[15] E. Bingham and A. Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal Neural Systems*, 10(1):1–8, 2000.

[16] R. M. Bolle and B. C. Vemuri. On three-dimensional surface reconstruction methods. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 1, pages 1–13, January 1991.

[17] H. Bosch, S. Mitchell, B. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. Reiber. Active appearance - motion models for fully automated endocardial contour detection in time sequences of echocardiograms. In *CARS*, pages 941–947, 2001.

[18] J Bosch, J. Reiber, G. van Burken, J. Gerbrands, W. Gussenhoven, N. Bom, and J. Roelandt. Automated endocardial contour detection in short-axis 2-d echocardiograms; methodology and assessment of variability. In *Proceedings of Computers in Cardiology*, pages 137–140, 25-28 Sep. 1988.

[19] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging*, 21(11):1374–1383, 2002.

[20] S. Bouix, K. Siddiqi, and A. R. Tannenbaum. Flux driven automatic centerline extraction. In *Medical Image Analysis*, 2004.

[21] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *IEEE International Conference in Computer Vision*, pages 105–112, 2001.

[22] Y. Boykov and K. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *IEEE International Conference in Computer Vision*, page 26, Washington, DC, USA, 2003. IEEE Computer Society.

[23] 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.

[24] W. Burgard, D. Fox, and D. Henning. Fast grid-based position tracking for mobile robots. In *KI '97: Proceedings of the 21st Annual German Conference on Artificial Intelligence*, pages 289–300, London, UK, 1997. Springer-Verlag.

[25] M. Butt and P. Maragos. Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10):1477–1484, 1998.

[26] R. Caldelli, A. Piva, M. Barni, and A. Carboni. Effectiveness of st-dm watermarking against intra-video collusion. In *IWDW*, pages 158–170, 2005.

[27] C. Cañero and P. Radeva. Vesselness enhancement diffusion. *Pattern Recognition Letters*, 24(16):3141 – 3151, 2003.

[28] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[29] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 67–76, New York, NY, USA, 2001. ACM Press.

[30] V. Caselles, F. Catté, B. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, 1993.

[31] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. In *IEEE International Conference in Computer Vision*, pages 694–699, 1995.

[32] T. Chan, B. Sandberg, and L. Vese. Active Contours without Edges for Vector-Valued Images. *Journal of Visual Communication and Image Representations*, 2:130–141, 2000.

[33] T. Chan and L. Vese. An Active Contour Model without Edges. In *International Conference on Scale-Space Theories in Computer Vision*, pages 141–151, 1999.

[34] Y. Chen, S. Thiruvenkadam, F. Huang, K. S. Gopinath, and R. W. Brigg. Simultaneous segmentation and registration for functional mr images. In *IARP International Conference on Pattern Recognition*, page 10747, Washington, DC, USA, 2002. IEEE Computer Society.

[35] D. Chopp. *Computing Minimal Surfaces via Level Set Curvature Flow*. PhD thesis, Department of Mathematics, Unibersity of California, Berkeley, 1991.

[36] L. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53:211–218, 1991.

[37] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver. *Combinatorial optimization*. Wiley, 1998. Cook.

[38] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[39] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407:484–498, 1998.

[40] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[41] D. Cremers. A Variational Framework for Image Segmentation Combining Motion Estimation and Shape Regularization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 53–58, 2003.

[42] D. Cremers. Dynamical statistical shape priors for level set-based tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1262–1273, August 2006.

[43] D. Cremers and G. Funka-Lea. Time-variant statistical shape priors for level set based tracking. In *Workshop on VLSM*, 2005.

[44] D. Cremers, T. Kohlberger, and C. Schn *o*rr. Nonlinear Shape Statistics in Mumford-Shah Based Segmentation. In *European Conference on Computer Vision*, volume 2, pages 93–108, 2002.

[45] D. Cremers, N. A. Sochen, and C. Schnörr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *Scale Space Methods in Computer Vision*, pages 388–400, 2003.

[46] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr. Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional. *International Journal of Computer Vision*, 50(3):295–313, 2002.

[47] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, New York, NY, USA, 1996. ACM Press.

[48] S. Dambreville, Y. Rathi, and A. Tannenbaum. Nonlinear shape prior from kernel space for geometric active contours. In *SPIE*, 2006.

[49] S. Dambreville, Y. Rathi, and A. Tannenbaum. Shape-based approach to robust image segmentation using kernel PCA. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 977–984, Washington, DC, USA, 2006.

[50] George B. Dantzig. *Linear Programming and Extensions*. Springer, 1993. DAN g2 93:1 1.Ex.

[51] F. Daum and J. Huang. Curse of dimensionality and particle filters. In *IEEE Proceedings of Aerospace Conference*, volume 4, pages 1979–1993, March 2003.

[52] M. de La Gorce and N. Paragios. Monocular hand pose estimation using variable metric gradient descent. In *British Machine Vision Conference*, volume 3, page 1269, 2006.

[53] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistics Society*, 39(1):1–38, 1977.

[54] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):39–55, 1987.

[55] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):39–55, 1987.

[56] S. Derrode, M.A. Chermi, and F. Ghorbel. Fourier-based invariant shape prior for snakes. In *ICASSP*, 2006.

[57] A. Dervieux and F. Thomasset. A finite element method for the simulation of rayleigh-taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.

[58] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In W. Reynolds and R.W. MacCormack, editors, *Seventh International Conference on Numerical Methods in Fluid Dynamics*, volume 141 of *Lecture Notes in Physics*, pages 158–163, June 1980.

[59] T. Deschamps. *Curve and Shape Extraction with Minimal Path and Level-Sets techniques - Applications to 3D Medical Imaging*. PhD thesis, Université Paris-IX Dauphine, Place du maréchal de Lattre de Tassigny, 75775 Paris Cedex, December 2001.

[60] T. Deschamps and L. D. Cohen. Fast extraction of minimal paths in 3D images and applications to virtual endoscopy. *Medical Image Analysis*, 5(4):281–299, December 2001.

[61] T. Deschamps and L.D. Cohen. Fast extraction of tubular and tree 3d surfaces with front propagation methods. In *IARP International Conference on Pattern Recognition*, Quebec, Canada, August 2002.

[62] M. Descoteaux, L. Collins, and K. Siddiqi. Geometric Flows for Segmenting Vasculature in MRI: Theory and Validation. In *Medical Imaging Computing and Computer-Assisted Intervention*, pages 500–507, 2004.

[63] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[64] H. Q. Dinh, G. Turk, and G. Slabaugh. Reconstructing surfaces by volumetric regularization using radial basis functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1358–1371, 2002.

[65] G. Doblinger. An adaptive kalman filter for the enhancement of noisy AR signals. In *IEEE International Symposium On Circuits and Systems*, pages 1–4, Monterey, CA, USA, May 1998.

[66] A. J. Dobson. *Introduction to Generalized Linear Models, Second Edition*. London: Chapman and Hall/CRC, 2001.

[67] A. Doucet, J. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[68] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[69] P. Fearnhead and P. Clifford. Online inference for well-log data. *Journal of the Royal Statistical Society*, 65:887–899, 2003.

[70] S. D. Fenster, C. G. Kuo, and J. R. Kender. Nonparametric training of snakes to find indistinct boundaries. In *Mathematical Methods in Biomedical Imaging Analysis*, Dec. 2001.

[71] M. Figueiredo and J. Leitao. A nonsmoothing approach to the estimation of of vessel controus in angiograms. *IEEE Transactions on Medical Imaging*, 14:162–172, 1995.

[72] R. A. Fisher. The use of multiple measures in taxonomic problems. *Annals Eugenics*, 7:179–188, 1936.

[73] A. Fitzgibbon. Robust registration of 2d and 3d point sets. In *British Machine Vision Conference*, volume II, pages 411–420, September 2001.

[74] C. Florin, N. Paragios, and J. Williams. Particle filters, a quasi-monte carlo solution for segmentation of coronaries. In *Medical Imaging Computing and Computer-Assisted Intervention*, pages 246–253, 2005.

[75] C. Florin, N. Paragios, and J. Williams. Globally optimal active contours, sequential monte carlo and on-line learning for vessel segmentation. In *European Conference on Computer Vision*, volume 3953, pages 476–489, 2006.

[76] A. Frangi, W. Niessen, P. Nederkoorn, O. Elgersma, and M. Viergever. Three-dimensional model-based stenosis quantification of the carotid arteries from contrast-enhanced MR angiography. In *IEEE Mathematical Methods in Biomedical Image Analysis*, pages 110–118, 2000.

[77] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. *Lecture Notes in Computer Science*, 1496, 1998.

[78] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[79] C. A. Glasbey and K. V. Mardia. A review of image-warping methods. *Journal of Applied Statistics*, 25(2):155–171, April 1998.

[80] S. H. Godsill and P. J. Rayner. *Digital Audio Restoration: A Statistical Model Based Approach*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1998.

[81] N. Gordon. Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEEE Proceedings of Radar and Signal Processing*, 140:107–113, 1993.

[82] N. Gordon. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Journal of Statistics and Computing*, 10:197–208, 2000.

[83] U. Grenander, Y. Chow, and D. M. Keenan. *Hands: a pattern theoretic study of biological shapes*. Springer-Verlag New York, Inc., New York, NY, USA, 1991.

[84] P. Hall and R. Martin. Incremental eigenanalysis for classification. In *British Machine Vision Conference*, volume 1, pages 286–295, 1998.

[85] G. Hamarneh and C. McIntosh. Deformable organisms for medical image analysis. *Parametric and Geometric Deformable Models: An application in Biomaterials and Medical Imagery*, 1:387–443, 2007.

[86] M. H. Hansen and B. Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774, 2001.

[87] M. Hart and L. Holley. A method of Automated Coronary Artery Trackin in Unsubtracted Angiograms. *IEEE Computers in Cardiology*, pages 93–96, 1993.

[88] T. Heimann, I. Wolf, and HP. Meinzer. Active shape models for a fully automated 3D segmentation of the liver - an evaluation on clinical data. In *Medical Imaging Computing and Computer-Assisted Intervention*, volume 2, pages 41–48, 2006.

[89] L. Hermoye, I. Laamari-Azjal, Z. Cao, L. Annet, J. Lerut, B. Dawant, and B. Van Beers. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. *Radiology*, 234(1):171–178, January 2005.

[90] A. Hilton, A.J. Stoddart, J. Illingworth, and T. Windeatt. Reliable surface reconstruction from multiple range images. In *European Conference on Computer Vision*, pages I:117–126, 1996.

[91] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface reconstruction from unorganized points. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 71–78, New York, NY, USA, 1992. ACM Press.

[92] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface reconstruction from unorganized points. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 71–78, New York, NY, USA, 1992. ACM Press.

[93] B. Horn and B. Schunck. Determinating Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[94] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 34(3):334–352, August 2004.

[95] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Network*, 13(4-5):411–430, 2000.

[96] S. Ioffe. Probabilistic linear discriminant analysis. In *European Conference on Computer Vision*, volume 3954 of *LNCS*, pages 531–542. Springer, 2006.

[97] M. Isard and A. Blake. Contour Tracking by Stochastic Propagation of Conditional Density. In *European Conference on Computer Vision*, volume I, pages 343–356, 1996.

[98] M. Isard and A. Blake. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[99] Kotecha J. and Djuric P. Gaussian sum particle filtering. *IEEE Transactions of Signal Processing*, 51(10):2602–2612, 2003.

[100] M.-P. Jolly. Assisted ejection fraction in b-mode and contrast echocardiography. In *IEEE International Symposium on Biomedical Imaging*, pages 97–100, April 2006.

[101] S. Joshi and M. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.

[102] O. Juan, R. Keriven, and G. Postelnicu. Stochastic motion and the level set method in computer vision: *stochastic Active Contours*. *International Journal of Computer Vision*, 69(1):7–25, 2006.

[103] R. Kalman. A New Approach to Linear Filtering and Prediction Problems. *ASME–Journal of Basic Engineering*, 82:35–45, 1960.

[104] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *IEEE International Conference in Computer Vision*, pages 261–268, 1987.

[105] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *IEEE International Conference in Computer Vision*, pages 810–815, 1995.

[106] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Archive for Rational Mechanics and Analysis*, 134:275–301, 1996.

[107] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Geometric active contours for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16:199–209, 1997.

[108] R. Knothe, S. Romdhani, and T. Vetter. Combining pca and lfa for surface reconstruction from a sparse set of control points. In *International Conference on Automatic Face and Gesture Recognition*, pages 637–644, 2006.

[109] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *European Conference on Computer Vision*, volume 3, pages 82–96, 2002.

[110] D. Kovacevic and S. Loncaric. Radial basis function-based image segmentation using a receptive field. *Conference Board of the Mathematical Sciences*, 00:126, 1997.

[111] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset. Model-based multiscale detection of 3D vessels. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 722–727, 1998.

[112] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa. Multiscale fourier descriptor for shape classification. *International Conference on Image Analysis and Processing*, page 536, 2003.

[113] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics, SIGGRAPH 2003*, 22(3):277–286, July 2003.

[114] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Controus. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I:316–322, 2000.

[115] Y. Li. On incremental and robust subspace learning. *Pattern Recognition*, 37(7):1509–1518, 2004.

[116] C.-W. Liao and G. Medioni. Surface approximation of a cloud of 3D points. *Graphical Models Image Processing*, 57(1):67–74, 1995.

[117] C.-B. Liu and N. Ahuja. A model for dynamic shape and its applications. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 129–134, 2004.

[118] F. Liu, B. Zhao, P. Kijewski, L. Wang, and L. Schwartz. Liver segmentation for ct images using gvf snake. *Medical Physics*, 32(12):3699–3706, December 2005.

[119] W. Lorensen and H. Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 163–169, New York, NY, USA, 1987. ACM Press.

[120] L. Lorigo, O. Faugeras, E. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C. Westin. Codimension-Two Geodesic Active Controus for the Segmentation of Tubular Structures. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I:444–451, 2000.

[121] W.-C. Ma, F.-C. Wu, and M. Ouhyoung. Skeleton extraction of 3d objects with radial basis functions. In *SMI '03: Proceedings of the Shape Modeling International 2003*, page 207, Washington, DC, USA, 2003. IEEE Computer Society.

[122] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. *International Journal of Computer Vision*, 39(1):57–71, 2000.

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

[124] R. Malladi and J. Sethian. A Real-Time Algorithm for Medical Shape Recovery. In *IEEE International Conference in Computer Vision*, pages 304–310, 1998.

[125] R. Malladi, J. Sethian, and B. Vemuri. A topology independent shape modeling scheme. In *Geometric Methods in Computer Vision II*, pages 246–256, 1993.

[126] R. Malladi, J. Sethian, and B. Vemuri. Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:158–175, 1995.

[127] HP. Meinzer, M. Thorn, and C. Cardenas. Computerized planning of liver surgery-an overview. *Journal of Computer and Graphics*, 26(4):569–576, August 2002.

[128] D. Nain, A. Yezzi, and G. Turk. Vessel Segmentation Using a Shape Driven Flow. In *Medical Imaging Computing and Computer-Assisted Intervention*, 2004.

[129] J. C. Nascimento, J. S. Marques, and J. M. Sanches. Estimation of cardiac phases in echographic images using multiple models. In *IEEE International Conference on Image Processing*, volume 2, pages 149–152, 2003.

[130] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[131] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transaction of Mathematical Software*, 27(1):27–57, 2001.

[132] Marcin Novotni and Reinhard Klein. Shape retrieval using 3D zernike descriptors. *Computer-Aided Design*, 36(11):1047–1062, 2004.

[133] T. O´ Donnell, T. Boult, X. Fang, and A. Gupta. The Extruded Generalized Cylider: A Deformable Model for Object Recovery. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 174–181, 1994.

[134] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe. A boosted particle filter: Multitarget detection and tracking. In *European Conference on Computer Vision*, volume 1, pages 28–39, 2004.

[135] J. O'Rourke. *Computational geometry in C*. Cambridge University Press, New York, NY, USA, 1994.

[136] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2003.

[137] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[138] C. Otto. *The Practice of Clinical Echocardiography*. Saunders, 2002.

[139] Thrasyvoulos N. Pappas and N. S. Jayant. An adaptive clustering algorithm for image segmentation. In *Second International Conference on Computer Vision*, pages 310–315, Tampa, FL, 1988. IEEE Computer Society.

[140] N. Paragios. Shape-based Segmentation and Tracking in Cardiac Image Analysis. *IEEE Transactions on Medical Imaging*, pages 773– 776, 2003.

[141] N. Paragios, Y. Chen, and O. Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[142] N. Paragios and R. Deriche. Geodesic Active Regions for Motion Estimation and Tracking. In *IEEE International Conference in Computer Vision*, pages 688–674, 1999.

[143] 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*, 22:266–280, 2000.

[144] N. Paragios and R. Deriche. Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.

[145] N. Paragios and R. Deriche. Geodesic Active Regions Level Set Methods for Supervised Texture Segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002.

[146] N. Paragios, M. Rousson, and V. Ramesh. Matching Distance Functions: A Shape-to-Area Variational Approach for Global-to-Local Registration. In *European Conference on Computer Vision*, pages II:775–790, 2002.

[147] H. Park, P. Bland, and C. Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging*, 22(4):483–492, April 2003.

[148] R. Petrocelli, K. Manbeck, and J. Elion. Three Dimentional Structue Recognition in Digital Angiograms using Gauss-Markov Models. *IEEE Computers in Radiology*, pages 101–104, 1993.

[149] B. Pham. Quadratic b-splines for automatic curve and surface fitting. *Computer and Graphics*, 13:471–475, 1989.

[150] P. Quelhas and J. Boyce. Vessel Segmentation and Branching Detection using an Adaptive Profile Kalman Filter in Retinal Blood Vessel Structure. In *Pattern Recognition and Image Analysis: First Iberian Conference*, pages 802–809, 2003.

[151] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. Particle filtering for geometric active contours with application to tracking moving and deforming objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2–9, Washington, DC, USA, 2005. IEEE Computer Society.

[152] B. Reitinger, A. Bornik, R. Beichel, and D. Schmalstieg. Liver surgery planning using virtual reality. *IEEE Computer Graphics Applications*, 26(6):36–47, 2006.

[153] K. Rohr. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[154] S. Roth, L. Sigal, and M. Black. Gibbs likelihoods for bayesian tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 886–893, June 2004.

[155] M. Rousson and D. Cremers. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *Medical Imaging Computing and Computer-Assisted Intervention*, volume 2, pages 757–764, 2005.

[156] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *European Conference on Computer Vision*, pages II:78–93, Copenhangen, Denmark, 2002.

[157] M. Rousson and N. Paragios. Prior knowledge, level set representations and visual grouping. *International Journal of Computer Vision*, 2007. to appear.

[158] D. Rowe, I. Rius, J. Gonzàlez, and J. Villanueva. Robust particle filtering for object tracking. In *International Conference on Image Analysis and Processing*, pages 1158–1165, 2005.

[159] S. Roy. Stereo without epipolar lines: A maximum-flow formulation. *International Journal of Computer Vision*, 34(2-3):147–161, 1999.

[160] F. L. Ruberg. *Computed Tomography of the Coronary Arteries*. London, UK: Taylor & Francis, 2005.

[161] D. Rueckert, P. Burger, S. Forbat, R. Mohiadin, and G. Yang. Automatic Tracking of the Aorta in Cardiovascular MR images using Deformable Models. *IEEE Transactions on Medical Imaging*, 16:581–590, 1997.

[162] M. Rutishauser, M. Stricker, and M. Trobina. Merging range images of arbitrarily shaped objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 573–580, 1994.

[163] Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, and R. Kikinis. 3D Multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In *Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*, pages 213–222, 1997.

[164] A. Schenk, G. Prause, and HO. Peitgen. Efficient semiautomatic segmentation of 3D objects in medical images. In *Medical Imaging Computing and Computer-Assisted Intervention*, pages 186–195, 2000.

[165] P. Schoenhagen, A. E. Stillman, S. S. Halliburton, S. A. Kuzmiak, T. Painter, and White R. D. Non-invasive coronary angiography with multi-detector computed tomography: comparison to conventional x-ray angiography. *International Journal of Cardiovascular Imaging*, 21(1):63–72, Feb. 2005.

[166] KS. Seo, HB. Kim, T. Park, PK. Kim, and JA. Park. Automatic liver segmentation of contrast enhanced ct images based on histogram processing. In *ICNC*, volume 1, pages 1027–1030, 2005.

[167] J. Sethian. A Review of the Theory, Algorithms, and Applications of Level Set Methods for Propagating Interfaces. *Cambridge University Press*, pages 487–499, 1995.

[168] J. Sethian. *Level Set Methods*. Cambridge University Press, 1996.

[169] G.K. Shaffer and R. Pito. Mesh integration based on co-measurements. In *IEEE International Conference on Image Processing*, pages II: 397–400, 1996.

[170] F. Sheehan, D. C. Wilson, D. Shavelle, and E. A. Geiser. Echocardiography. In M. Sonka and J. M. Fitzpatrick, editors, *Handbook of Medical Imaging Volume 2: Medical Image Processing and Analysis*, chapter 11. SPIE press, 2000.

[171] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 2000.

[172] 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. *Computer Aided Surgery (CAS)*, 6(3), August 2001.

[173] E. Sorantin, C. Halmai, B. Erbohelyi, K. Palagyi, K. Nyul, K. Olle, B. Geiger, F. Lindbichler, G. Friedrich, and K. Kiesler. Spiral-CT-based assesment of Tracheal Stenoses using 3D Skeletonization. *IEEE Transactions on Medical Imaging*, 21:263–273, 2002.

[174] M. Soucy and D. Laurendeau. A dynamic integration algorithm to model surfaces from multiple range views. *Machine Vision and Applications*, 8(1):53–62, 1995.

[175] R. Strand and G. Borgefors. Distance transforms for three-dimensional grids with non-cubic voxels. *Computer Vision and Image Understanding*, 100(3):294–311, 2005.

[176] J. Suri, K. Liu, L. Reden, and S. Laxminarayan. A review on mr vascular image processing: skeleton versus nonskeleton approaches: part ii. *IEEE Transactions on Information Technology in Biomedicine*, 6(4):338–350, 2002.

[177] M. R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 8(70):920–930, 1979.

[178] S. Tehrani, T.E. Weymouth, and B. Schunck. Interpolating cubic spline contours by minimizing second derivative discontinuity. *IEEE International Conference in Computer Vision*, pages 713–716, 90.

[179] D. Terzopoulos and R. Szeliski. Tracking with Kalman Snakes. In A. Blake and A. Yuille, editors, *Active Vision*, pages 3–20. MIT Press, 1992.

[180] P. Thévenaz, T. Blu, and M. Unser. Image interpolation and resampling. In I.N. Bankman, editor, *Handbook of Medical Imaging, Processing and Analysis*, chapter 25, pages 393–420. Academic Press, San Diego CA, USA, 2000.

[181] A. Tonazzini and L. Bedini. Monte carlo markov chain techniques for unsupervised mrf-based image denoising. *Pattern Recognition Letters*, 24(1-3):55–64, 2003.

[182] K. Toyama and A. Blake. Probabilistic Tracking in a Metric Space. In *IEEE International Conference in Computer Vision*, pages 50–59, 2001.

[183] J. Tsitsiklis. Efficient Algorithms for Globally Optimal Trajectories. *IEEE Transactions on Automatic Control*, 40:1528–1538, 1995.

[184] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 311–318, New York, NY, USA, 1994. ACM Press.

[185] G. Turk and J. O'Brien. Modeling with implicit surfaces that interpolate. *ACM Transaction on Graphics*, 21(4):855–873, 2002.

[186] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[187] M. Uzumcu, A.F. Frangi, J. Reiber, and B. Lelieveldt. The use of independent component analysis in statistical shape models. In Milan Sonka and J. Michael Fitzpatrick, editors, *SPIE*, volume 5032, pages 375–83, 2003.

[188] A. Vasilevskiy and K. Siddiqi. Flux Maximizing Geometric Flows. In *IEEE International Conference in Computer Vision*, pages I: 149–154, 2001.

[189] O. Veksler. Image segmentation by nested cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1339–, 2000.

[190] E. Wan and R. van der Merwe. The unscented Kalman filter. In S. Haykin, editor, *Kalman Filtering and Neural Networks*, chapter 7. Wiley Publishing, 2001.

[191] X. Wang and H. Wang. Evolutionary gibbs sampler for image segmentation. In *IEEE International Conference on Image Processing*, pages 3479–3482, 2004.

[192] Y. Wang and L. H. Staib. Boundary finding with correspondence using statistical shape models. In *IEEE Conference on Computer Vision and Pattern Recognition*, page 338, Washington, DC, USA, 1998. IEEE Computer Society.

[193] C.-Y. Wee, R. Paramesran, and F. Takeda. New computational methods for full and subset Zernike moments. *Information Sciences—Informatics and Computer Science*, 159(3-4):203–220, 2004.

[194] G. Welch and G. Bishop. An introduction to the kalman filter. Technical Report TR 95-041, University of North Carolina at Chapel Hill, 2004.

[195] W. West. Modelling with mixtures. In J. Bernardo, J. Berger, A. Dawid, and A. Smith, editors, *Bayesian Statistics*. Clarendon Press, 1993.

[196] O. Wink, W. J. Niessen, and M. A. Viergever. Multiscale vessel tracking. *IEEE Transactions on Medical Imaging*, 23(1):130–133, 2004.

[197] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101–1113, 1993.

[198] A. Yezzi, A. Tsai, and A. Willsky. A Statistical Approach to Snakes for Bimodal and Trimodal Imagery. In *IEEE International Conference in Computer Vision*, pages 898–903, 1999.

[199] P. Yim, P. Choyke, and R. Summers. Grayscale Skeletonization of Small Vessels in Magnetic Resonance Angiography. *IEEE Transactions on Medical Imaging*, 19:568–576, 2000.

[200] L. Younes. Combining geodesic interpolating splines and affine transformations. *IEEE Transactions of Image Processing*, 15(5):1111–1119, May 2006.

[201] D. S. Zhang and G. Lu. A comparative study of curvature scale space and fourier descriptors. *Journal of Visual Communication and Image Representation*, 14(1):41–60, 2003.

[202] T. Zhang and D. Freedman. Tracking objects using density matching and shape priors. In *IEEE International Conference in Computer Vision*, pages 1056–1062, 2003.

[203] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

[204] S. C. Zhu and D. Mumford. Prior learning and gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1236–1250, 1997.

[205] L. Zöllei, W. E. L. Grimson, A. Norbash, and W. M. III Wells. 2D-3D rigid registration of x-ray fluoroscopy and ct images using mutual information and sparsely sampled histogram estimators. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 696–703, 2001.