Cue Integration and Front Evolution in Image Segmentation

Mikaël Rousson

THÈSE

présentée et soutenue à

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS

le 6 décembre 2004

pour obtenir le titre de

DOCTEUR EN SCIENCES

École Doctorale Sciences et Technologies de l'Information et de la Communication

Spécialité

Informatique et Traitement du Signal

Mikaël Rousson

Cue Integration and Front Evolution in Image Segmentation

Intégration d'attributs et évolutions de fronts en segmentation d'images

Advisor: **Rachid Deriche**

Président	Olivier Faugeras	INRIA Sophia Antipolis, France
Rapporteurs	Luis Alvarez	University of Las Palmas, Spain
	Guillermo Sapiro	Univesity of Minesota, Minneapolis, USA
	Kaleem Siddiqi	McGill University, Montreal, Canada
Examinateurs	Michel Barlaud	Universtity of Nice-Sophia Antipolis, France
	Rachid Deriche	INRIA Sophia Antipolis, France

Abstract

Automatic detection and selection of regions of interest inside an image is a key step in image understanding. Many studies have been dedicated to this issue during the past decades. Efficient and robust algorithms have been developed for many applications. However, most of them make use of heuristics inherent to a particular class of images. The limiting factor to obtain a general algorithm is the large variety of cues available to characterize a region of interest. Examples include gray-level, color, texture and shape.

In this thesis, we propose a general formulation able to deal with each one of these characteristics. Image intensity, color, texture, motion and prior shape knowledge are considered. For this purpose, a probabilistic inference is obtained from a Bayesian formulation of the segmentation problem. Then, reformulated as an energy minimization problem, the most probable image partition is obtained using front evolution techniques. Level-set functions are introduced to represent the evolving fronts while region statistics are optimized in parallel. This framework can naturally handle scalar and vector-valued smooth images but more complex cues are also integrated. Texture and motion features, as well as prior shape knowledge are successively introduced. Complex medical images are considered in the last part of the thesis, with a focus on diffusion magnetic resonance images and their associated 3D probability density fields.

Résumé

La détection et l'extraction automatique de régions d'intéret à l'intérieur d'une image est une étape primordiale pour la compréhension des images. Une multitude d'études dédiées à ce problème ont été proposées durant les dix dernières années. Des algorithmes efficaces et robustes ont été développés pour diverses applications. Cependant, la plupart d'entre eux introduisent des heuristiques propres au type d'image considéré. La variété des caractéristiques possibles définissant une région d'intéret est le principal facteur limitant leur généralisation. Ces critères région peuvent être le niveau de gris, la couleur, la texture, la forme des objets, etc...

Dans cette thèse, nous proposons une formulation générale qui permet d'introduire chacune de ces caractéristiques. Plus précisément, nous considérons l'intensité de l'image, la couleur, la texture, le mouvement et enfin, la connaissance à priori sur la forme des objets à extraire. Dans cette optique, nous obtenons un critère probabiliste à partir d'une formulation Bayésienne du problème de la segmentation d'images. Ensuite, une formulation variationnelle équivalente est introduite et la segmentation la plus probable est finalement obtenue par des techniques d'évolutions de fronts. La représentation par ensembles de niveaux est naturellement introduite pour décrire ces évolutions, tandis que les statistiques régions sont estimées en parallèle. Ce cadre de travail permet de traîter naturellement des images scalaires et vectorielles mais des caractéristiques plus complexes sont considérées par la suite. La texture, le mouvement ainsi que l'à priori sur la forme sont traîtés successivement. Finalement, nous présentons une extention de notre approche aux images de diffusion à résonnance magnétique où des champs de densité de probabilité 3D doivent être considérés.

Acknowledgments

This thesis was accomplished in the project **Odysée** at **INRIA, Sophia-Antipolis** under the supervision of **Rachid Deriche**. It has been funded by the European project **Cogvisys**.

First, I would like to thank **Olivier Faugeras**, director of the project Odyssée, for welcoming me in his research team where I had the opportunity to meet and work with many talented people.

I would like to thank **Rachid Deriche**, my PhD supervisor, who guided and supported me during these three years. I greatly acknowledge him for his wise advice and his involvement in our research.

All this could not have happened without the influence of **Nikos Paragios** who made me interested in research and in image analysis during my internship at Siemens Corporate Research. Also, I would like to thank him for his help and encouragement before and during my PhD studies.

Several people have widely contributed to the work presented in this manuscript. Next to Rachid Deriche and Nikos Paragios, **Thomas Brox** and **Christophe Lenglet** were active collaborators. Throughout this thesis, I will emphasize their important contributions in my work.

I would also like to thank my reviewers **Luis Alvarez**, **Kaleem Siddiqi** and **Guillermo Sapiro** for the time they spent reading my manuscript and for their fruitful comments. I am also very grateful to the ones who had a long trip to come to my defense in Sophia-Antipolis. Finally, I thank **Michel Barlaud** for having accepted to be in my jury.

Contents

Introduction **13**
A Bayesian formulation for image segmentation 14
Organization of this thesis 15

Résumé détaillé (version française) **19**

1 Image Segmentation by Front Evolution **41**
1.1 Boundary-based curve evolutions: from *Snakes* to geodesic active contours 42
1.1.1 The *Snake* model 42
1.1.2 The level-set representation for front evolutions 43
1.1.3 Geometric active contours 45
1.1.4 Geodesic active contours 46
1.2 Region-based contour evolutions 48
1.2.1 The level-set representation as an optimization framework . . 49
1.2.2 The Mumford-Shah functional 51
1.2.3 Bayesian formulation 53
1.3 Conclusion 57

2 Bayesian Region-Based Front Evolution **59**
2.1 Maximum a posteriori frame partition 60
2.2 Partition constraint 60
2.3 Region-based term 60
2.4 Energy formulation 61
2.5 Adaptive region-based segmentation 62
2.5.1 Parametric region densities 62
2.5.2 Nonparametric region densities 69
2.6 Implementation 76
2.6.1 The general case 76
2.6.2 Multi-resolution implementation 76
2.7 Connections with classical clustering methods 77
2.7.1 K-means 77

2.7.2 Maximum-likelihood and the expectation-maximization (EM) algorithm . . . 79
2.8 Contributions and conclusions . . . 81

3 Cue Extraction & Integration 83
3.1 Static cue: intensity/color and structure tensor . . . 84
3.1.1 Dealing with texture . . . 84
3.1.2 Nonlinear structure tensor . . . 85
3.1.3 Combination of intensity and texture cues . . . 88
3.1.4 Unsupervised segmentation of textured images . . . 89
3.1.5 Limitations and possible extensions . . . 101
3.1.6 Conclusion . . . 101
3.2 Motion cue: optical flow . . . 102
3.2.1 Optical flow extraction . . . 102
3.2.2 Combining spatial and temporal cues . . . 103
3.2.3 Motion segmentation . . . 104
3.2.4 Tracking . . . 106
3.3 Contributions and conclusions . . . 108

4 Implicit Representation for Prior Shape Knowledge 113
4.1 Implicit shape registration/alignment . . . 116
4.1.1 The level-set representation for global registration . . . 117
4.1.2 Global registration . . . 118
4.2 Implicit shape modeling . . . 125
4.2.1 Capturing local variability : a stochastic shape model . . . 127
4.2.2 Capturing principal variations through principal component analysis . . . 133
4.2.3 When should the mean representation be maintained to a distance function? . . . 135
4.3 Shape constraints in object extraction . . . 136
4.3.1 Shape constrained surface evolution . . . 137
4.3.2 Constrained object extraction . . . 142
4.4 Contributions and conclusion . . . 147

5 Segmentation of 3D Probability Density Fields: Application to Diffusion MRI 151
5.1 Introduction . . . 151
5.2 Data acquisition, DTI . . . 153
5.3 Diffusion tensor images segmentation . . . 154
5.3.1 Linear approximation . . . 156
5.3.2 Probability density fields segmentation . . . 157

5.3.3 A Riemmanian approach to DTI segmentation 159
5.4 Experimental Results and Comparisons 161
5.4.1 Synthetic data 162
5.4.2 Real diffusion tensor data 164
5.4.3 Experiments for Method 3 166
5.5 Contributions and conclusions 166

Conclusion **169**

Conclusion (version française) **171**

Introduction

Analyzing image content has become vital in many applications. Personal as well as national security make use of image or video monitoring which require automatic extraction of meaningful information. Medical imaging has also become an important application in image analysis by helping doctors in the detection and diagnosis of diseases and assisting them during the operating phases. Other applications are coming to light with the fast development of communications and entertainment which need more and more tools to compress and manipulate multimedia content. The large number of images and their very high resolutions make manual analysis obsolete and automatic algorithms need to be developed. This domain of computer-aided image processing is still very young, since it became practical only in the early eighties. Seminal works were mostly heuristic but recent studies are using mathematical techniques from geometry, statistics and learning.

Among all image processing tools, segmentation is probably one of the most important since it is a necessary step in numerous algorithms. Loosely speaking, it consists in partitioning an image in regions of interest. There is no general definition of a region of interest since it may depend on the type of image and on the considered application. It can be defined as homogeneous regions with respect to a given measure or as regions that fit the best with some prior knowledge.

For example, in natural images, one will try to mimic the human visual system to obtain results equivalent to the ones that would be given by a human. Given the numerous varieties of natural images [Figure 1], resolving this problem is in general quite difficult. Image intensity, color and texture are probably the most prominent features but, when analyzing a scene, in addition to the image content, a human makes use of prior knowledge learned from its own experience or inherited from the species. Even if more and more neurobiological studies are starting to propose models of the visual cortex, its high complexity is far from being completely understood. Therefore, rather than considering arbitrary images, most studies concentrate only on a subset of real images with particular characteristics. Another problem is the segmentation of medical images. In this case, the objective is not to copy the visual system but to extract real structures/organs from a given modality of acquisition. Hence, understanding and modeling the acquisition process is a key step to extract information from acquired images. Despite all advances in medical image acquisition, many important structures

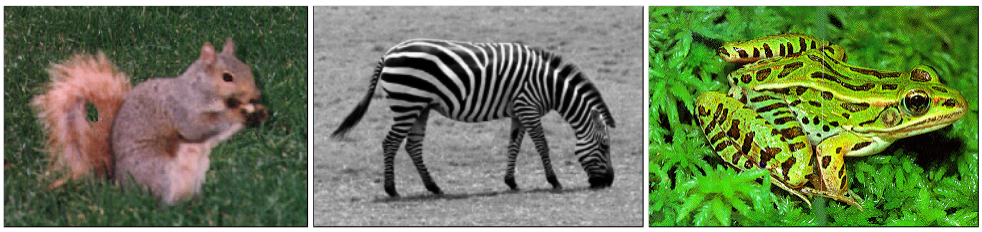


Figure 1: Examples of natural images.

remain hardly visible. All information that can be obtained from prior knowledge is of great interest and it has to be used to help the segmentation process.

A wide range of methods has been proposed to tackle the problems posed by image segmentation. They include direct thresholding based on image gray-level, grouping from edge detector results, classical clusterings on pixel values (k-means, Gaussian mixture models), graph-cuts, Markov random fields and others. In this thesis, we consider a geometric approach based on front evolution techniques. The geometric flow is obtained from a general Bayesian formulation of the segmentation problem which allows for various cues to be considered and therefore, a large set of images.

The choice of a geometric approach has been made for several reasons. The segmentation problem is by definition a geometric one since geometric structures have to be extracted from the 2D image domain. Optimizing with respect to contours or shapes, we consider an optimization space that permits a straightforward formulation of partitioning problems. As we will see along this thesis, geometric constraints can then be naturally expressed to integrate properties of object borders. Simple spatial regularizations can be introduced but more complex geometric knowledge will also be presented. The second characteristic of our framework is the definition of the geometric flow from a Bayesian formulation. Allowing us to integrate multiple cues, the simplification of the general partitioning probability density function can be done by making explicit assumptions which gives us a very clear view of the limits of our algorithms. Aiming at removing parts of these assumptions, extensions are then possible. In the following, we introduce briefly this Bayesian formulation, the base of the work presented in this thesis.

A Bayesian formulation for image segmentation

We consider a general framework for image segmentation integrating different cues: gray-value, color, texture, motion and prior shape knowledge. The main idea is to evolve a front subject to internal and external forces such that it maps the boundaries of the object(s) to be extracted at convergence. Using a Bayesian formulation of the problem, different cues can be integrated in this front evolution. Let I be the image

to be segmented and $\mathcal{P}(\Omega)$, a partition of the image domain Ω. Following [107], we consider $p_{\mathcal{S}}(\mathcal{P}(\Omega)|I)$, the *a posteriori* frame partition probability, given the observed image I. The optimal partition of the image is obtained by maximizing this probability according to the associated hypothesis. The Bayes rule permits one to express this probability as:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) = \frac{p(I|\mathcal{P}(\Omega))}{p(I)} p(\mathcal{P}(\Omega)), \tag{1}$$

where $p(\mathcal{P}(\Omega))$ and $p(I)$ are respectively the probability of a partition $\mathcal{P}(\Omega)$ and an image I. The third term, $p(I|\mathcal{P}(\Omega))$ represents the *a posteriori* segmentation probability for the image I, given the partition $\mathcal{P}(\Omega)$. When segmenting a given image, the term $p(I)$ is constant and (1) simplifies to:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) \propto p(I|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)). \tag{2}$$

Both terms $p(I|\mathcal{P}(\Omega))$ and $p(\mathcal{P}(\Omega))$ are studied in this thesis.

An approximation of $p(I|\mathcal{P}(\Omega))$ for piecewise smooth images will be first proposed in Chapter 2 assuming pixel values to be independent realizations of a random process. This approximation will be applied to scalar and vector-valued images, considering either parametric or non-parametric probability density approximations for each region. Then, an extension to textured images will be presented in Chapter 3 by considering a two-step approach where texture features are first extracted.

The probability of a partition $\mathcal{P}(\Omega)$ will be also studied in detail. It will first be expressed as a simple regularization constraint in Chapter 2 but more complex models will be proposed in Chapter 4. In a given application, the objects to be extracted often have a similar shape. Assuming a set of training samples to be available, it is possible to model a family of shapes by learning the distribution of the training shapes. Then, this distribution can be used to express the probability of a new partition $\mathcal{P}(\Omega)$. Two different models will be presented in Chapter 4.

Organization of this thesis

This thesis is organized in five chapters. The first one is a detailed review of state of the art techniques in front evolution methods for image segmentation. Considering such techniques, the second part introduces a general region-based unsupervised segmentation approach from which all the other parts are based. Hence, complex cues are integrated in this framework: texture and motion are considered in Chapter 3 while Chapter 4 deals with prior shape knowledge. The fifth and last chapter shows an extension to a relatively new modality in medical imaging by considering diffusion magnetic resonance images. A more detailed summary of each one of these parts is presented below.

CHAPTER 1
In the first part, we review recent progress in level-set based surface evolutions which represent efficient techniques for object extraction. Boundary and region-based methods are presented and we discuss the advantages of each variant.

CHAPTER 2
Based on a Bayesian formulation, the second chapter proposes a general formulation for region-based image partitioning of scalar as well as vector-valued images. A direct application of this approach is presented on gray and color images. Parametric and nonparametric probability density functions are considered to approximate region properties. Links with more "classical" clustering techniques are also discussed.

CHAPTER 3
Textured images do not fit the assumptions made to simplify the Bayesian formulation. Contrary to smooth images, spatial links between pixel values must be considered. To this end, we apply the idea of the nonlinear structure tensor which turns out to have very good discrimination properties while inducing only three feature channels. Motion information is also introduced through the estimation of the optical flow, extending a segmentation technique of static images to the detection of moving objects in image sequences.

CHAPTER 4
The fourth part focuses on the introduction of prior shape knowledge within the segmentation process. In many applications like medical imaging, hand-segmented data are often available and structures usually exhibit similar shapes. Analyzing and incorporating this information can be of great interest to improve the robustness of the algorithm and to deal with critical images. Two different models are proposed, both are based on implicit shape representations. We first deal with voxel-wise shape variability through a stochastic shape modeling. We then extend Active Shape Models [38] by analyzing the principal variations of the implicit shapes.

CHAPTER 5
Finally, we propose a novel approach to extract structures from diffusion magnetic resonance images. Our goal is to extract structures of interest in the white matter. These types of images differ from classical modalities by describing the stochastic motion of water molecules at each voxel. We extend our region-based formulation to segment 3D probability density fields, by recasting this problem into the recent framework of information geometry. Promising results are obtained on synthetic as well as on real datasets.

At the end of each chapter, we recall its main contributions and mention possible

extensions. To summarize, this thesis contributes to the problem of image segmentation by:

1. presenting a detailed review of state of the art methods in front evolution for image segmentation (chapter 1),
2. extending region-based approaches to vector-valued images with parametric or non-parametric statistical models (chapter 2),
3. estimating region statistics, jointly with the partitioning process (chapter 2),
4. extracting and integrating texture and motion cues (chapter 3),
5. modeling and integrating prior shape knowledge (chapter 4),
6. extending all this framework to the use of the more complex diffusion tensor images (chapter 5).

Résumé détaillé (version française)

Introduction

L'analyse d'images est devenue un élément vital de nombreuses applications. Elle est par exemple omniprésente dans les systèmes de sécurité, que ce soit pour défendre des intérêts collectifs ou plus personnels. Ainsi, de plus en plus de systèmes de surveillance ont recours à des techniques de traitement d'images afin d'extraire les automatiquement les informations pertinentes. Dans un tout autre domaine, l'imagerie médicale a pris une place importante dans le monde médical en offrant aux docteurs de nouveaux moyens pour la détection et le suivi de maladies ainsi qu'une assistance en phases opératoires. Un troisième secteur important est celui des contenus multimédias qui sont récemment devenus entièrement digitaux. De nombreux challenges sont alors apparu pour la compression, la manipulation et l'indexation de telles données. Le nombre important d'images et leur haute résolution rend l'analyse de toutes ces données impossible par un observateur humain et l'élaboration d'algorithmes automatiques est nécessaire. L'analyse d'images assistée par ordinateur est un domaine encore très récent. Les premiers travaux étaient pour la plupart heuristiques mais de nombreuses études plus récentes reposent sur des bases mathématiques telles que la géométrie, les statistiques ou les théories de l'apprentissage.

Dans toues ces applications, la segmentation est une étape clé dans la plupart des algorithmes d'analyse d'images utilises. De manière générale, segmenter une image consiste à extraire les différentes régions d'intérêt. La caractérisation de ces zones dépend à la fois du type d'image considéré et de l'application visée et il n'existe donc pas de définition unique. Par exemples, elles peuvent être caractérisées par certaines mesures d'homogénéité ou à partir d'informations à priori.

La segmentation d'images naturelles est un problème particulier puisque le but d'obtenir des résultats comparables a ceux que donnerait un observateur humain. Étant donné la grande diversité des images naturelles, résoudre ce problème est en pratique très compliqué. L'intensité de l'image, la couleur et la texture sont certainement les caractéristiques les plus importantes mais lorsque nous analysons une scène, nous employons des sources d'information autres que son contenu propre. Nous utilisons notamment les informations acquises par expérience. Bien que de plus en plus d'études

neuro-biologiques aient pour objectif de comprendre notre système visuel, elles sont encore loin de pouvoir modéliser sa grande complexité. Par conséquent, au lieu de considérer des images quelconques, la plupart des approches se concentrent sur un sous-ensemble d'images naturelles ayant des propriétés particulières.

La segmentation d'images médicales est un problème légèrement différent. L'objectif n'est plus de copier un système visuel mais d'extraire des structures/organes réels à partir d'une modalité d'acquisition donnée. Ainsi, la compréhension et la modélisation du système d'acquisition est une étape importante afin d'extraire l'information des images obtenues. Malgré tous les progrès effectués en acquisition d'images médicales, de nombreuses structures sont à peine visibles et l'utilisation d'informations à priori est alors primordiale.

Différents types d'approches ont été proposées pour résoudre les nombreux problèmes posés par la segmentation d'images. Parmi les méthodes les plus populaires, on trouve de simples seuillages basés sur l'intensité de l'image, des approches de regroupement suivant la valeur des pixels (k-moyennes, modèles de mixtures de Gaussiennes), des techniques utilisant le parcellement de graphes (graph-cut), d'autres introduisent des champs aléatoires de Markov, etc... Dans cette thèse, nous considérons une approche géométrique basée sur les techniques d'évolution de fronts. Un critère géométrique est obtenu à partir d'une formulation Bayésienne du problème de segmentation qui permet d'intégrer différents attributs et par conséquent, de traiter une grande diversité d'images.

Une formulation Bayésienne pour la segmentation d'images

Nous considérons un cadre général permettant d'intégrer les attributs suivants: niveau de gris, couleur, mouvement et connaissance à priori sur la forme. L'idée principale est de propager un front soumis à des forces internes et externes afin qu'une fois à l'équilibre, il délimite les différentes régions d'intérêt. Grâce à une formulation Bayésienne du problème, nous montrons que différents attributs peuvent être intégrés dans une telle évolution du front. Soit I l'image à segmenter et $\mathcal{P}(\Omega)$ une partition du domaine Ω de l'image. En se basant sur les idées développées dans [107], nous considérons $p_{\mathcal{S}}(\mathcal{P}(\Omega)|I)$, la probabilité *à posteriori* d'un partitionnement d'une image donnée I. La partition optimale de l'image est obtenue en cherchant à maximiser cette probabilité par rapport aux hypothèses associées. La règle de Bayes permet d'exprimer cette probabilité comme:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) = \frac{p(I|\mathcal{P}(\Omega))}{p(I)} p(\mathcal{P}(\Omega)) \tag{3}$$

où $p(\mathcal{P}(\Omega))$ et $p(I)$ sont respectivement les probabilités d'une partition $\mathcal{P}(\Omega)$ et d'une image I. Le troisième terme $p(I|\mathcal{P}(\Omega))$, représente la probabilité *à postériori* d'une image I pour un partitionnement $\mathcal{P}(\Omega)$. Lorsque l'on segmente une image donnée, le

terme $p(I)$ est constant et l'équation (3) se simplifie pour donner:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) \propto p(I|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)) \tag{4}$$

Dans cette thèse, nous étudions les deux termes de cette relation: $p(I|\mathcal{P}(\Omega))$ et $p(\mathcal{P}(\Omega))$.

Une approximation du terme $p(I|\mathcal{P}(\Omega))$ pour des images lisses par morceaux est tout d'abord proposée dans le chapitre 2 en introduisant l'hypothèse que les pixels d'une même région sont des réalisations indépendantes d'un processus aléatoire. Cette approximation est appliquée à des images scalaires et vectorielles avec des approximations paramétriques et non-paramétriques des densités de probabilité régions. Dans le chapitre 3, nous présentons une extension de cette approche aux images texturées. La stratégie en deux étapes considérée consiste à extraire dans un premier temps les composantes texture et de segmenter ensuite l'image vectorielle formée par l'ensemble des attributs souhaités.

La probabilité d'une partition $\mathcal{P}(\Omega)$ est aussi étudiée en détail. Elle est tout d'abord exprimée comme une simple contrainte de régularisation dans le chapitre 2 mais des modèles plus complexes sont introduits par la suite dans le chapitre 4. Dans une application donnée, les formes extraites d'images différentes sont souvent similaires. En supposant qu'un jeu de formes d'apprentissage soit disponible, il est alors possible de modéliser la famille de forme correspondant au problème en apprenant leur distribution. Cette modélisation permet par la suite d'exprimer la probabilité d'une nouvelle partition de l'image. Deux modélisations différentes de cette information sur la forme sont proposées dans le chapitre 4.

Organisation de la thèse

Cette thèse est organisée en cinq chapitres. Le premier présente un bref état de l'art des techniques de segmentation par évolution de fronts. Dans la seconde partie, nous utilisons ces techniques pour définir une approche générale de segmentation d'images non-supervisée qui sera la base des autres parties. Ainsi, des attributs complexes sont par la suite introduits dans ce cadre de travail: la texture et le mouvement sont considérés au cours du chapitre 3 et le chapitre 4 est dédié à l'introduction d'information à priori sur la forme. La cinquième et dernière partie présente une extension de ces travaux à une modalité d'imagerie médicale relativement récente, les images de diffusion à résonance magnétique. Un résumé plus détaillé de chacune de ces parties est présenté ci-dessous.

CHAPITRE 1

Dans ce chapitre, nous présentons les avancées récentes en techniques d'évolution de surfaces implicites pour l'extraction de structures dans une image. La représentation par ensembles de niveaux est décrite et des méthodes dites frontières et régions sont

introduites. Nous parlons en particulier des limitations de chacune d'entre elles.

CHAPITRE 2
En ce basant sur le cadre de travail présenté dans le chapitre 1, nous proposons une formulation région pour la segmentation d'images scalaires et multivaluées. Une application directe de cette méthode est présentée sur des images en niveaux de gris et sur des images couleur. Des densités de probabilités paramétriques et non-paramétriques sont introduites pour approximer les propriétés de chaque région. Les relations liant notre approche et des méthodes générales d'estimation de paramètres statistiques sont finalement présentées.

CHAPITRE 3
Les images texturées ne satisfont pas aux hypothèses introduites au cours du chapitre 2. Contrairement aux images "lisses", des relations spatiales entre pixels doivent être considérées. Afin de capturer cette information, nous utilisons le tenseur de structure non-linéaire qui a de très bonnes propriétés de discrimination, tout en introduisant un nombre réduit de composantes. Une approche similaire est proposée pour intégrer le mouvement grâce à une estimation du flot optique. Cela permet d'étendre notre approche de segmentation d'images statiques à la détection et au suivi d'objets en mouvement dans des séquences vidéos.

CHAPITRE 4
Dans le quatrième chapitre, nous introduisons un à priori sur la forme dans ce processus de segmentation. Dans de nombreuses applications, telles que l'analyse d'images médicales, des données segmentées à la main sont disponibles et les structures d'intérêt ont souvent des formes similaires d'une image à l'autre. L'introduction de cette information sur la forme peut grandement améliorer la robustesse d'un algorithme d'extraction automatique. Deux modèles différents sont proposés, tous deux basés sur une représentation implicite des formes. Le premier modèle incorpore une variabilité voxelique sur la forme à priori alors que le deuxième capture les variations principales d'un jeu de représentations implicites, étendant les modèles actifs de formes aux ensembles de niveaux.

CHAPITRE 5
Finalement, nous proposons une nouvelle approche pour la segmentation d'images de diffusion à résonance magnétique, notre objectif étant d'extraire des structures d'intérêt à l'intérieur de la matière blanche du cerveau. Ce type d'images est différent des images classiques car elles décrivent le mouvement stochastique des molécules d'eau en chaque voxel. Nous étendons notre formulation région à la segmentation de champs de densité de probabilité 3D, utilisant des travaux récents effectués dans le domaine de la géométrie de l'information. Des résultats prometteurs sont obtenus sur

des images synthétiques et réelles.

Chapitre 1
Segmentation d'images par évolutions de fronts

Les techniques d'évolution de fronts ont été introduites en segmentation d'images par Kaas et al. dans [74]. Leurs travaux proposent de détecter les frontières entre les objets en faisant se déplacer une courbe (souvent appelée "snake") dans l'image. L'évolution de courbe proposée dans [74] est introduite pour minimiser une énergie définie le long de la courbe. Cette énergie incorpore des contraintes géométriques ainsi qu'un terme de recouvrement avec les forts gradients de l'image. Cette approche fut une avancée importante dans le développement des techniques de segmentation et tout une partie du domaine s'en est fortement inspirée. Diverses extensions du modèle "Snake" ont rapidement été proposées afin d'éliminer plusieurs de ses limitations initiales. Le principal problème était la nécessité d'avoir une très bonne initialisation du contour pour obtenir le résultat escompté. Parmi ces contibutions, on trouve notamment les travaux de Blake [15], Cohen [35], Berger [9], Fua et Leclerc [58]. Cependant, ces améliorations n'éliminent pas entièrement l'importance de l'initialisation et la topologie des objets à extraire doit toujours être connue au préalable. Une autre étape a été franchie lorsque la représentation par ensembles de niveaux a été utilisée pour la propagation de fronts. Introduite par Dervieux et Thomasset dans [47, 48], cette représentation implicite fut redécouverte et plus largement diffusée quelques années plus tard par Osher et Sethian [106]. Cette représentation implicite et intrinsèque du contour fournit naturellement une solution aux principales limites des contours actifs (le changement de topologie et la paramétrisation du contour). De plus, les propriétés géométriques de la courbe telles que la courbure, le vecteur normal ou tangent, peuvent être facilement exprimées à l'aide de cette représentation. De nombreux flots bénéficiant de cette formulation ont donc été proposés [27, 34, 91, 76, 28]. Des schémas d'implémentation rapides dont la convergence et la stabilité sont prouvées ont aussi été proposés [132, 1]. Le modèle dit des "contours actifs géodésiques" est certainement l'un des plus élégants en reformulant le problème de la segmentation par celui de la recherche d'un chemin de distance minimale dans un espace Riemmanien dont la métrique est induite par des propriétés de l'image.

Plus récemment, des approches ont proposé d'utiliser des termes régions pour contrôler l'évolution du contour [171, 30, 107, 122, 68]. Alors que les contours actifs ou les évolutions de fronts classiques incluent uniquement l'information gradient pour définir un critère d'arrêt, les approches région considèrent une formulation plus globale où des quantités sur le domaine complet de l'image sont intégrées. Contrairement à l'approche gradient qui repose sur des mesures locales, elles sont plus robustes aux conditions initiales et il n'est plus nécessaire de régler de manière précise un

critère d'arrêt. Ces méthodes régions peuvent être classées en deux catégories: celles basées sur la fonctionnelle de Mumford-Shah ou issues de minimisations des "Variations Totales" [102, 15, 30, 31, 144] et celles obtenue par formulations Bayésiennes [171, 107, 122].

Dans le prochain chapitre, nous revisitons ces méthodes régions. Nous étendons celle présentée dans [107] en ajoutant l'estimation des statistiques régions en parallèle à l'évolution de courbe. Nous considérons des images scalaires et multivaluées dans le cas d'approximations paramétriques et non-paramétriques des statistiques régions.

Chapitre 2
Évolutions de fronts basés sur l'information région

Dans ce chapitre, nous développons des évolutions de fronts basées région et issues de la formulation Bayésienne du problème de segmentation. Ce type d'approche a été choisi pour plusieurs raisons. Tout d'abord, le choix d'une approche région semble le meilleur pour définir un algorithme robuste à l'initialisation et pouvant s'appliquer à une large variété d'images. De plus, l'objectif principal de cette thèse étant l'incorporation de différents attributs, l'approche considérée doit intégrer un processus de décision afin de sélectionner dynamiquement les attributs importants et la formulation Bayésienne est tout à fait adaptée.

La fonctionnelle que nous considérons dans ce chapitre est proche de celle présentée en section 1.2.3. Plusieurs travaux [81, 171, 107] ont proposé des critères similaires obtenus à partir de formulations différentes. Dans [81, 171], cette énergie est obtenue à partir d'un critère dit "Descripteur de Longueur Minimale" alors que dans [107], elle est prouvée être équivalente à approximation du maximum à posteriori de la probabilité de partitionnement de l'image. Pour cela, plusieurs hypothèses explicitées par les auteurs sont nécessaires.

Soit I l'image à segmenter et $\mathcal{P}(\Omega)$ une partition du domaine Ω de l'image. En se basant sur les idées développées dans [107], nous considérons $p_{\mathcal{S}}(\mathcal{P}(\Omega)|I)$, la probabilité *à posteriori* d'un partitionnement d'une image I. La partition optimale de l'image est obtenue en cherchant à maximiser cette probabilité par rapport aux hypothèses associées. La règle de Bayes permet d'exprimer cette probabilité comme

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) = \frac{p(I|\mathcal{P}(\Omega))}{p(I)} p(\mathcal{P}(\Omega)), \tag{5}$$

où $p(\mathcal{P}(\Omega))$ et $p(I)$ sont respectivement les probabilités d'une partition $\mathcal{P}(\Omega)$ et d'une image I. Le troisième terme $p(I|\mathcal{P}(\Omega))$, représente la probabilité *à posteriori* d'une image I pour un partitionnement $\mathcal{P}(\Omega)$. Lorsque l'on segmente une image donnée, le terme $p(I)$ est constant et l'équation (5) se simplifie pour donner:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) \propto p(I|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)) : \tag{6}$$

Contrainte de partitionnement

Le terme $p(\mathcal{P}(\Omega))$ permet d'intégrer de l'information à priori sur le partitionnement de l'image. La modélisation et l'introduction de contraintes de forme sont étudiées dans chapitre 4 mais pour l'instant, nous considérons une contrainte plus générale en utilisant ce terme pour exprimer des propriétés de régularité géométrique sur l'interface entre les régions. Soit $\mathcal{C}$ cette frontière, une contrainte de régularisation peut être introduite en favorisant les partitions avec une interface de longueur minimale:

$$p(\mathcal{P}(\Omega)) = \frac{\nu}{2}\, e^{-\nu\,|\mathcal{C}|},$$

où $|\mathcal{C}|$ est la longueur de l'interface et ν est une constante permettant de contrôler l'importance donnée à ce terme de régularisation.

Terme région

Le second terme $p(I|\mathcal{P}(\Omega))$ ne peut pas être estimé dans le cas général et plusieurs simplifications doivent être introduites. Tout d'abord, nous supposons que le nombre de régions est connu et qu'aucune corrélation entre labels n'existe. Le premier facteur de l'expression (6) s'écrit alors:

$$p(I|\mathcal{P}(\Omega)) = p(I|\{\Omega_1, \ldots, \Omega_N\}) = p(I|\Omega_1)p(I|\Omega_2)\ldots p(I|\Omega_N),$$

où $p(I|\Omega_X)$ est la probabilité de l'image I sachant que Ω_X est une région d'intérêt.

Une autre simplification est introduite en supposant que les pixels d'une région sont des réalisations indépendantes et identiquement distribuées d'un même processus aléatoire. Cela permet de remplacer les distributions régions par les probabilités jointes des pixels de chaque région:

$$p(I|\mathcal{P}(\Omega)) = \prod_{\mathbf{x}\in\Omega_1} p(I(\mathbf{x})|\Omega_1) \prod_{\mathbf{x}\in\Omega_2} p(I(\mathbf{x})|\Omega_2)\cdots \prod_{\mathbf{x}\in\Omega_N} p(I(\mathbf{x})|\Omega_N).$$

Cette approximation n'est généralement pas valide puisque les régions de l'image sont souvent caractérisées par des relations spatiales à un niveau plus ou moins local. Dans le cas d'images texturées par exemple, ces relations spatiales sont les propriétés principales des régions d'intérêt. Par conséquent, ce modèle n'est valide que pour des images lisses par morceaux et éventuellement bruitées. Néanmoins, comme nous le verrons dans le chapitre 3, des filtres peuvent être utilisés afin d'extraire ces propriétés spatiales pour ensuite être intégrées dans cette approche.

Formulation variationelle

L'estimation du maximum à posteriori est équivalente à minimiser l'énergie obtenue en appliquant la fonction $[-\mathbf{log}()]$. En négligeant les termes constants, nous obtenons

l'énergie suivante:

$$E(\{\Omega 1, \ldots, \Omega_N\}) = -\sum_i \int_{\Omega_i} \log p(I(\mathbf{x})|\Omega_i)\, d\mathbf{x} + \nu\,|\mathcal{C}|. \tag{7}$$

Par la suite, nous considérons différents types de densités de probabilité régions. L'estimation de ces densités est effectuée pendant le processus de segmentation. Cela est possible en optimisant l'énergie (7) à la fois par rapport aux statistiques régions et par rapport au partitionnement. Des approximations paramétriques et non-paramétriques de ces densités sont proposées.

Densités régions paramétriques

Soit $p(I|\theta_i)$ une représentation paramétrique de la densité de probabilité dans la région Ω_i, paramétrisée par θ_i. La segmentation est obtenue en minimisant

$$E(\{\Omega 1, \ldots, \Omega_N\}, \{\theta 1, \ldots, \theta_N\}) = -\sum_i \int_{\Omega_i} \log p(I(\mathbf{x})|\theta_i)\, d\mathbf{x} + \nu\,|\mathcal{C}|. \tag{8}$$

Cette énergie dépend de deux types de paramètres: les régions Ω_i, sous-ensembles ouverts et bornés de Ω, et les paramètres statistiques appartenant à un espace de paramètres Θ donné. Une combinaison de fonctions d'ensembles de niveaux permet de reformuler cette énergie. L'optimisation par rapport à des sous-domaines de l'image est alors remplacée par une minimisation d'énergie qui dépend de fonctions Lipschitziennes définissant implicitement le partitionnement de l'image. Par souci de simplicité, nous considérons le cas élémentaire d'un bi-partitionnement ($N = 2$)[1]. Soit ϕ la fonction distance signée à la frontière entre Ω_1 et Ω_2, positive dans Ω_1 et négative dans Ω_2. L'énergie (8) s'écrit

$$\begin{aligned} E(\phi, \{\theta_1, \theta_2\}) = &-\int_\Omega \Big(H(\phi) \log p(I(\mathbf{x})|\theta_1) + (1 - H(\phi)) \log p(I(\mathbf{x})|\theta_2) \Big)\, d\mathbf{x} \\ &+ \nu \int_\Omega |\nabla H(\phi(\mathbf{x}))|\, d\mathbf{x}. \end{aligned} \tag{9}$$

Pour θ_1 et θ_2 fixés, cette fonctionnelle peut être minimisée par rapport à ϕ par descente de gradient. L'équation d'Euler-Lagrange pour ϕ donne l'équation d'évolution suivante:

$$\frac{\partial \phi}{\partial t}(\mathbf{x}) = \delta(\phi) \left(\nu \,\mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \log \frac{p(I(\mathbf{x})|\theta_1)}{p(I(\mathbf{x})|\theta_2)} \right), \tag{10}$$

[1] Plusieurs travaux proposent des formulations par ensembles de niveaux pour plus de deux régions [167, 107, 155, 20] mais la complexité devient plus importante et ces approches sont alors plus sensibles aux conditions initiales. Le but de notre travail étant l'estimation de paramètres statistiques durant le partionnement, nous nous limitons au cas simple de deux régions. Cependant, ce cas simplifié est pertinent dans la plupart des extractions d'objets où un seul objet doit être séparé du reste de l'image.

tandis que les paramètres statistiques doivent être régulièrement mis à jour afin de vérifier la contrainte

$$\theta_i = \underset{\theta \in \Theta}{\text{Arg min}} \int_{\Omega_i} \log p(I(\mathbf{x})|\theta)\, d\mathbf{x} \quad \text{for} \quad i = 1, 2. \tag{11}$$

Dans le cas d'approximations par des densités Gaussiennes, les formes analytiques de ces paramètres statistiques sont disponibles et ils peuvent être mis à jour après chaque itération de l'évolution du front.

Densités régions non-paramétriques

Choisir une famille de densités paramétriques peut être vu comme une limitation de la modélisation des statistiques régions. Contrairement aux modèles paramétriques, l'estimation de densités non-paramétriques peuvent approximer n'importe quel type de distribution pour un jeu de données suffisamment large. Parmi les méthodes non-paramétriques, la méthode dite des *fenêtres de Parzen* et couramment utilisée. Appliquée à notre problème, l'estimation de Parzen de la distribution de l'intensité dans une région Ω_X s'écrit:

$$p_X(z) = \frac{1}{|\Omega_X|} \int_{\Omega_X} \frac{1}{h} K\left(\frac{z - I(\mathbf{x})}{h}\right) d\mathbf{x}. \tag{12}$$

Notre énergie devient alors:

$$E(\{\Omega_1, \ldots, \Omega_N\}) = -\sum_i \int_{\Omega_i} \log\left(\frac{1}{|\Omega_i|} \int_{\Omega_i} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}}\right) d\mathbf{x} + \nu|\mathcal{C}|. \tag{13}$$

Contrairement au cas paramétrique, cette fonctionnelle ne dépend que des sous-domaines Ω_i mais des intégrales emboîtées apparaissent. Ce type de fonctionnelle peut être minimisé en utilisant le formalisme de *Gradient de Forme* introduit dans [4]. Appliqué à nore problème, nous obtenons la proposition suivante:

Proposition 0.1 *La dérivée de Gâteaux du terme région défini par une estimation de Parzen avec écart type fixé*

$$F(\Omega_X) = \int_{\Omega_X} \log\left(\frac{1}{|\Omega_X|} \int_{\Omega_X} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}}\right) d\mathbf{x}, \tag{14}$$

dans la direction d'un champs de vecteurs $\mathbf{V}$ *est:*

$$< F'(\Omega_X), V >= -\int_{\partial\Omega_X} \Bigg(\log p_X(I(\mathbf{x})) + \frac{1}{|\Omega_X|} \Bigg(\int_{\Omega_X} \frac{\frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right)}{p_X(I(\hat{\mathbf{x}}))}\, d\hat{\mathbf{x}} - 1 \Bigg) \Bigg) (V(\mathbf{x}) \cdot N(\mathbf{x})) d\mathbf{a}(\mathbf{x}). \tag{15}$$

Dans le cas d'un noyau Gaussien, certains termes peuvent être négligés et l'équation d'évolution complète est alors proche de celle du cas paramétrique:

$$\begin{cases} p_i(I(\mathbf{x})) = \dfrac{1}{|\Omega_i|}\displaystyle\int_{\Omega_i} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}}, \quad i = 1, 2 \\ \dfrac{\partial \phi}{\partial t}(\mathbf{x}) \approx \delta(\phi)\left(\nu \operatorname{div}\left(\dfrac{\nabla \phi}{|\nabla \phi|}\right) + \log \dfrac{p_1(I(\mathbf{x}))}{p_2(I(\mathbf{x}))}\right) \end{cases} \tag{16}$$

Implémentation

L'algorithme général est similaire dans les deux approches. Deux étapes alternent jusqu'à convergence: (i) l'estimation des statistiques régions et (ii), l'évolution de front.

Les détails techniques pour mettre à jour les statistiques régions ont été abordées mais l'évolution de front nécessite quelques clarifications. Tout d'abord, les équations d'évolution de fronts (10) et (16) ne doivent pas nécessairement être appliquées au domaine de l'image complet. La fonction Dirac étant en facteur dans chacune de ces équations, seuls les pixels proches du niveaux zéro doivent être mis à jour. Il est ainsi possible de travailler seulement dans une bande étroite autour du passage par zéro de ϕ. En pratique, une liste de points correspondant à cette bande doit être maintenue et modifiée à chaque itération. Cette bande est en fait obtenue facilement à partir de la ré-initialisation de l'ensemble de niveaux à la fonction distance [142, 132].

Chapitre 3
Extraction et intégration d'attributs

L'approche présenté jusqu'ici est limitée aux images lisses par morceaux, i.e. sans structures. Ceci est dû à l'hypothèse restrictive faite dans le chapitre précédent où es pixels d'une même région sont supposés être des réalisations indépendantes et identiquement distribuées d'un processus aléatoire:

$$p(I|\Omega_i) \approx \prod_{\mathbf{x} \in \Omega_i} p(I(\mathbf{x})|\Omega_i).$$

Par conséquent, seuls les histogrammes des régions sont utilisés pour modéliser les statistiques régions. Cette approximation est valide pour des images lisses mais dès que de la texture est présente, cette approche ne peur capturer l'information pertinente. Par définition, une texture est liée à une distribution spatiale des pixels dans un certain voisinage. Comme notre approche n'intègre pas de tels modèles spatiaux, l'information texture n'a pas d'influence sur la segmentation.

Afin d'étendre notre approche aux images non-lisses/texturées, deux possibilités s'offrent à nous. La première est de construire des modèles statistiques de textures d'ordres élevés à partir de distributions de probabilités sur des champs aléatoires en

utilisant la théorie des champs de Markov [64, 43]. La deuxième approche utilise la théorie du filtrage qui s'inspirent d'études neurobiologiques du cortex visuel suggérant la présence d'une multitude de filtres orientés [94, 135]. Les méthode les plus utilisés sont basées sur les filtres de Gabor [59, 151] ou sur des transformées en ondelettes [93, 136].

Afin d'obtenir un compromis entre simplicité et généricité, nous proposons une approche en deux étapes dont la première peux s'inscrire dans les méthodes de filtrage. Des caractéristiques textures sont tout d'abord extraites en considérant une version non-linéaire du tenseur de structure. Ensuite, elles sont combinées à l'intensité pour former un vecteur caractéristique permettant de partitionner l'image. Les techniques d'évolutions de courbes basées régions présentées dans le chapitre précédent sont utilisées pour segmenter l'image à partir du vecteur caractéristique qui peut être vu comme une image vectorielle. De nombreux tests expérimentaux sur des images très variées illustrent le potentiel de la méthode, nous fournissant ainsi un algorithme non-supervisé de segmentation d'images texturées.

Une approche similaire en deux étapes est proposée pour intégrer l'information mouvement. Dans un premier temps, cette information est extraite dans la séquence d'images grâce à une estimation robuste du flot optique. Ensuite, les composantes mouvement sont combinées avec l'intensité (ou la couleur) et la texture pour définir un nouveau vecteur caractéristique. Ce vecteur peut être utilisé pour segmenter/détecter des objets en mouvement et ensuite pour les suivre au cours du temps. Une nouvelle formulation couplant plusieurs ensembles de niveaux est introduite afin de suivre plusieurs objets, tout en gérant les cas de recouvrements.

Caractéristiques statiques: intensité/couleur et tenseur de structure

L'utilisation de filtres de Gabor ou de modèles basés sur les champs de Markov a le grand inconvénient d'introduire de nombreux paramètres difficiles à estimer. Une solution très intéressante à ce problème de discrimination de textures a été proposée par Bigün et al [11] avec le tenseur de structure. Comparé aux autres méthodes, le tenseur de structure introduit seulement trois composantes à une échelle donnée. Dans le paragraphe suivant, nous démontrons que, combinées à l'intensité, ces trois composantes sont suffisantes pour la discrimination de nombreux types de textures. Le seul problème dans le travail original [11] est l'utilisation d'un lissage Gaussien qui déplace les frontières et conduit ainsi à des segmentations peu précises. Pour résoudre ce problème, Weickert et al. ont développé dans [162] une version non-linéaire du tenseur de structure basée sur la diffusion non-linéaire d'images à valeurs matricielles.

Le tenseur de structure non-linéaire

Nous proposons d'extraire les caractéristiques textures à partir du *tenseur de structure* [55, 11, 118, 87, 63]:

$$J_\rho = K_\rho * (\nabla I \nabla I^\top) = \begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix},$$

où K_ρ est un noyau Gaussien ayant comme déviation standard ρ. Ainsi le tenseur de structure fourni trois composantes à un niveau d'échelle donné. Nous traitons ici le cas simplifié où seul un niveau d'échelle est considéré. Il est possible d'incorporer plusieurs niveaux d'échelle en ajoutant les composantes correspondantes. En comparaison aux méthodes basées sur les filtres de Gabor, l'orientation de la texture est ici obtenue à partir de versions régularisées des dérivées de l'image. Cette est totalement contenue dans les dérivées de l'image. Ainsi, le tenseur de structure peut être aussi efficace qu'un ensemble de filtres de Gabor pour discriminer les textures. Il est en outre invariant par rotation.

Le problème principal du tenseur de structure classique et le déplacement des contours causé dû au lissage Gaussien. Cela conduit à une segmentation peu précise près des contours. Comme mentionné plus tôt, l'idée introduite dans [162] pour régler ce problème est de remplacer le lissage Gaussien par une diffusion non-linéaire. Suivant cette idée, nous proposons une version améliorée pour la segmentation de textures. La diffusion non-linéaire est basée sur les travaux précurseurs de Peronal et Malik [115]. L'idée principale est de réduire le lissage sur les contours. L'équation de diffusion résultante est $\partial_t u = \text{div}\,(g(|\nabla u|)\nabla u)$ où $u(t=0)$ est l'image I et g est une fonction décroissante. Cette équation peut seulement être utilisée sur des valeurs scalaires mais Gerig et al. [61] ont proposé une version pour les données vectorielles:

$$\partial_t u_i = \text{div}\left(g\left(\sum_{k=1}^{N} |\nabla u_k|^2\right)\nabla u_i\right) \qquad \forall i = 1, \ldots, N,$$

où u_i est une composante du vecteur caractéristique et N est le nombre de composantes. Il faut noter que les composantes sont couplées par l'intermédiaire d'une diffusion jointe, un contour dans une composante réduisant ainsi le lissage dans les autres.

Si les composantes d'une matrice sont vues comme les composantes d'un vecteur (ce qui est raisonnable puisque la norme de Frobenius d'une matrice est égale à la norme Euclidienne du vecteur résultant), il est possible de diffuser une matrice telle que le tenseur de structure. Cela nous conduit à un tenseur de structure non-linéaire légèrement différent de celui utilisé dans [162]. La diffusion correspondante sur la matrice est en fait équivalente à celle proposée dans [145, 146]. Il est à noter, d'après [162], que le couplage entre les composantes permet de conserver la semi-positivité.

Ensuite, un point important est le choix de la fonction de diffusivité g. Pour le cas de caractéristiques texture, le flot à *variations totales* (TV) [127, 22] semble être

très bien adapté puisqu'il permet d'éliminer les oscillations et conduit à un résultat constant par morceaux. Cela est très utile étant donné que le tenseur de structure, qui contient des dérivées premières de l'image, a des réponses très locales. Le lissage a pour objectif principal de fermer les zones entre ces phénomènes locaux, tout en préservant les contours importants. Cependant, lorsque le gradient est proche de zéro, le flot TV conduit à des problèmes numériques. Ce cas est traité en ajoutant une constante ϵ au module du gradient: $g(|\nabla u|) = \frac{1}{|\nabla u|+\epsilon}$.

Pour l'implémentation, nous appliquons la technique AOS [164] qui permet un calcul efficace du flot TV, même pour des valeurs très faibles de ϵ (proche de 0.001), introduisant ainsi peu de flou. Dans ce cas précis, la technique AOS est environ 4 fois plus rapide qu'une simple implémentation explicite.

Combinaison de l'intensité et des composantes textures

L'étape suivante est la combinaison de l'information texture et de l'intensité dans le but de simplifier les données et de fermer les structures. Pour cela, nous utilisons la diffusion d'images multivaluée présentée dans le paragraphe précédent car elle couple chaque composante avec la même diffusivité. De cette manière, l'information de toutes les composantes influe sur la probabilité d'avoir une frontière à un point donné. Cependant, pour un couplage équilibré, l'équation de diffusion suppose que toutes les composantes aient des valeurs comparables. Malheureusement, une simple normalisation sur toutes les composantes n'est pas une approche adéquate puisqu'elle risque d'amplifier le bruit dans les composantes avec peu d'information. Ce problème doit donc être résolu d'une manière différente.

Pour les composantes textures, cela peut être fait en remplaçant le tenseur de structure pas sa racine carrée. Notons $J_0 = T(\lambda_i)T^\top$, une décomposition en valeurs propres du tenseur de structure, sa racine carrée peut être calculée de la manière suivante:

$$\tilde{J_0} := \sqrt{J_0} = T(\sqrt{\lambda_i})T^\top.$$

Dans le cas d'images en niveaux de gris, cela revient à:

$$\tilde{J_0} = \begin{pmatrix} \frac{I_\mathbf{x}}{|\nabla I|} & -\frac{I_\mathbf{y}}{|\nabla I|} \\ \frac{I_\mathbf{y}}{|\nabla I|} & \frac{I_\mathbf{x}}{|\nabla I|} \end{pmatrix} \begin{pmatrix} |\nabla I| & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{I_\mathbf{x}}{|\nabla I|} & \frac{I_\mathbf{y}}{|\nabla I|} \\ -\frac{I_\mathbf{y}}{|\nabla I|} & \frac{I_\mathbf{x}}{|\nabla I|} \end{pmatrix} = \frac{J_0}{|\nabla I|}.$$

En couplant ainsi l'intensité et le tenseur de structure, la diffusion non-linéaire est appliquée au vecteur caractéristique u composé de l'intensité et des composantes de $\tilde{J_0}$.

Segmentation non supervisée d'images texturées

Comme nous l'avons déjà mentionné, l'approche de segmentation développée dans le chapitre précédent est limitée au cas d'images lisses car les pixels de chaque région

sont supposés indépendants. Cependant, les composantes textures que nous venons d'extraire incluent une information de cohérence spatiale des pixels. Par conséquent, il devient possible de segmenter des images plus compliquées incluant de la texture en appliquant l'approche développée dans le chapitre précédent sur ce vecteur caractéristique:

$$p(I|\Omega_i) \approx p(u|\Omega_i) \approx \prod_{\mathbf{x}\in\Omega_i} p(u(\mathbf{x})|\Omega_i),$$

où u inclut l'intensité et les composantes du tenseur de structure après diffusion couplée.

Suivant le cadre de travail développé précédemment, le problème de segmentation peut être reformulé comme une minimisation d'énergie. De plus, au lieu de modéliser la distribution jointe de chaque composante, nous les supposons non-corrélées. Soit $p_{ij}(\mathbf{x})$, la densité de probabilité conditionnelle d'une valeur $u_j(x)$ d'apparaître dans la région Ω_i, la fonctionnelle de bi-partitionnement correspondante est:

$$E(\phi) = \int_\Omega \left(H(\phi) \sum_{j=1}^{4} \log p_{1j}(u(\mathbf{x})) + (1 - H(\phi)) \sum_{j=1}^{4} \log p_{2j}(u(\mathbf{x})) \right) d\mathbf{x} + \int_\Omega \nu |\nabla H(\phi)| \, d\mathbf{x}.$$

Une approximation Gaussienne appairait raisonnable pour représenter chaque densité région et une optimisation alternée permet alors de minimiser cette énergie par rapport à ϕ, et par rapport aux paramètres statistiques. Le calcul des variations de ϕ conduit à l'évolution d'ensembles de niveaux suivante:

$$\frac{\partial \phi}{\partial t}(\mathbf{x}) = \delta(\phi) \left(\nu \, \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \sum_{j=1}^{4} \log \frac{p_{1j}(I(\mathbf{x}))}{p_{2j}(I(\mathbf{x}))} \right) \quad \forall \mathbf{x} \in \Omega,$$

tandis que les paramètres statistiques optimaux peuvent être estimés en parallèle directement à partir de leurs estimations empiriques.

Caractérisation du mouvement: le flot optique

Si l'on considère le mouvement, l'estimation du flot optique est la méthode principale pour intégrer cette information. L'estimation du flot optique est en soit un domaine de recherche entier et de nombreuses méthodes existent [7, 99, 141]. Il est intéressant de constater que le tenseur de structure non-linéaire peut aussi être utilisé dans le cadre d'estimation du mouvement [19]. Dans ce cas, un tenseur de structure non-linéaire spatio-temporal est nécessaire. Il est obtenu de la même manière que dans la section précédente mais en changeant la condition initiale: $u(t = 0) = (I_{\mathbf{x}}^2, I_{\mathbf{y}}^2, I_{\mathbf{z}}^2, 2I_{\mathbf{x}}I_{\mathbf{y}}, 2I_{\mathbf{x}}I_{\mathbf{z}}, 2I_{\mathbf{y}}I_{\mathbf{z}})$ où z est la coordonnée temporelle.

Segmentation d'objets en mouvement

Lorsque l'on considère une séquence d'images, l'information mouvement devient disponible et les composantes du flot optique peuvent alors être ajoutées au vecteur caractéristique. Pour une séquence d'images en niveaux de gris, le vecteur caractéristique contient alors 6 éléments: l'intensité, les 3 composantes du tenseur de structure et les deux composantes du flot optique. Pour la détection d'objets en mouvement, il est souvent utile d'augmenter le poids des composantes du flot, sans quoi des objets non-mobiles mais avec une intensité ou une texture différente du fond risquent aussi d'être détectés.

Suivi d'objets en mouvement

Le problème du suivi d'objets est similaire. La principale différence est que la position ainsi que le nombre d'objets dans l'image précédente sont supposés connus. Cela permet d'éliminer l'hypothèse de bi-partitionnement et il devient possible de suivre chaque objet séparément. Mais il est alors nécessaire d'introduire un couplage entre plusieurs ensembles de niveaux décrivant chaque objet mobile. Ce couplage est important afin d'éviter des recouvrements entre objets.

Un ensemble de niveaux est assigné à chaque objet mobile Ω_i tel que $\phi_i(\mathbf{x}) > 0$ si $\mathbf{x} \in \Omega_i$. En ajoutant la contrainte de non-recouvrement, une fonction caractéristique χ_i peut être associée a chaque objet et une autre fonction χ_b caractérise le fond de l'image. Soit e_i la log-vraisemblance d'un pixel pour la région Ω_i et e_b, celle du fond, le suivi de M objets est alors obtenu en minimisant l'énergie suivante:

$$E(\{\phi_k, k = 1..M\}) = -\int_\Omega \chi_b(\mathbf{x}) e_b(\mathbf{x})\, d\mathbf{x} - \sum_{i=1}^{M} \int_\Omega \chi_i(\mathbf{x}) e_i(\mathbf{x})\, d\mathbf{x} + \nu \sum_{i=1}^{M} \int_\Omega |\nabla H(\phi_i)|\, d\mathbf{x}$$

La minimisation de cette énergie fournit un système d'équations correspondant aux évolutions de chaque ensemble de niveaux. De la même manière que précédemment, les statistiques région peuvent être mises à jour régulièrement à partir de leurs estimations empiriques.

Chapitre 4
Représentation implicite pour la connaissance à priori sur la forme

Afin d'introduire un à priori sur la forme dans les méthode basées sur les ensembles de niveaux, la première étape consiste à construire un modèle de forme. La représentation

par ensemble de niveaux apparaît adaptée pour la construction de tels modèles, comme le prouvent de nombreux travaux publiés récemment [86, 33, 143, 125, 42]. L'objectif est alors de retrouver une représentation compacte d'un jeu d'ensembles de niveaux $[\phi_1, \phi_2, \ldots, \phi_N]$ correspondant aux contours ou surfaces d'apprentissage. Une simple moyenne de ces représentation peut être estimée et utilisée comme modèle [33] mais cette technique ne permet de capturer aucune variabilité autour de cette forme moyenne.

Lorsqu'un nombre relativement important d'exemples est disponible, il est possible de capturer les variations principales des formes et une analyse en composantes principales (ACP) peut être appliquée afin de capturer ces variations [86, 143]. L'ACP est une transformation linéaire des variables considérées qui permet d'extraire les modes de variation principaux.

Nous présentons comment extraire ce modèle de formes et nous proposons une nouvelle méthode pour l'introduire dans des évolutions de fronts classiques.

Extraction des variations principales par ACP

Dans une application donnée, les formes d'apprentissage sont supposées appartenir a une même famille de formes. En analysant ce jeu d'apprentissage, le but est de modéliser cette famille. Pour cela, nous considérons l'approche introduite dans [86, 143]. Ce modèle fait l'hypothèse que cette famille peut être générée à partir de combinaisons linéaires d'ensembles de niveaux. Dans un premier temps, les exemples ϕ_i sont centrées par rapport à une représentation moyenne ϕ_s, $[\psi_i = \phi_i - \phi_s]$. Ensuite, les principaux modes de variation peuvent être obtenus en effectuant une ACP sur la matrice de covariance:

$$\Sigma_{i,j} = \frac{1}{|\Omega|} \int_{\Omega} \psi_i(\mathbf{x}) \psi_j(\mathbf{x})\, d\mathbf{x}, \quad 1 \leq i \leq N, 1 \leq j \leq N.$$

Gardant seulement les m modes de variation principaux, une nouvelle forme appartenant à la même famille est obtenue par combinaison linéaire:

$$\phi = \phi_s + \sum_{j=1}^{m} \lambda_j\, U_j, \tag{17}$$

où les U_j sont les principaux modes de variation sélectionnés et λ_j sont les poids associés a chacun de ces modes. La valeur de ces poids étant limitée dans un certain intervalle défini par les valeurs propres correspondantes.

Contraintes de forme et extraction d'objets

L'étape suivante est l'introduction de ce modèle de forme à priori dans une évolution de front en définissant la probabilité d'un partitionnement de l'image $p(\mathcal{P}(\Omega))$ lorsqu'un modèle de forme est disponible. Nous nous limitons à l'extraction d'un seul objet (pas

nécessairement connexe) et nous représentons le front par un ensemble de niveaux ϕ. La probabilité d'un partitionnement est alors obtenue à partir de la distribution à priori $p(\phi|$"modèle de forme").

La même représentation de forme étant utilisée lors de la modélisation et de l'extraction, la définition de cette probabilité à priori est des plus naturelle. A partir d'une formulation Bayésienne, nous proposons un critère permettant de contraindre l'évolution de fronts en incluant une estimation dynamique de la pose de l'objet. Ensuite, nous développons deux modules de contrainte de forme: (i) l'à priori sur la forme est constitué d'un seul ensemble de niveaux et (ii), le modèle ACP de la section précédente est intégré.

Évolution de fronts sous contrainte de forme

Nous considérons maintenant l'évolution d'un front représenté par une fonction implicite ϕ. Le but est de contraindre l'évolution de ce front par certaines propriétés de forme $\phi_{\mathcal{M}}$, modulo une transformation donnée $\mathcal{A}$. Il est possible de modéliser l'espace joint du front et de la transformation vers le modèle à priori en utilisant une densité de probabilité. Soit $[p(\phi, \mathcal{A}|\phi_{\mathcal{M}})]$, la distribution à priori du couple $(\phi, \mathcal{A})$ pour un modèle $\phi_{\mathcal{M}}$ donné. Cette distribution n'est pas connue, elle varie entre différents objets et elle ne peut pas être estimée dans le cas général. Cependant, les méthodes d'échantillonnage dites de Monte-Carlo ou d'autres techniques peuvent être utilisées pour retrouver ce type de distribution lorsque des informations empiriques sont disponibles. En l'absence d'une telle connaissance, nous pouvons considérer une formulation Bayésienne de cette densité:

$$p(\phi, \mathcal{A}|\phi_{\mathcal{M}}) = \frac{p(\phi_{\mathcal{M}}|\phi, \mathcal{A})}{p(\phi_{\mathcal{M}})} p(\phi, \mathcal{A}) = \frac{p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)}{p(\phi_{\mathcal{M}})} p(\phi, \mathcal{A}).$$

Le terme constant $p(\phi_{\mathcal{M}})$ peut être ignore et l'espace joint des interfaces et des transformations est supposé uniforme. Cette hypothèse est souvent considérée en l'absence de connaissance sur les propriétés de l'objet à extraire. Ensuite, retrouver l'interface optimale et la transformation est équivalent à trouver le maximum à posteriori $p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)$ obtenu en maximisant

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|\phi(\mathbf{x})).$$

où les pixels sont supposés indépendants. Il faut aussi tenir compte de la transformation du front ϕ dans une pose similaire de celle du modèle $\phi_{\mathcal{M}}$. Si l'on considère une similitude (transformation rigide et facteur d'échelle), les variations d'échelle entraîne des changements "prédictifs" de la fonction distance [111] et il doit être introduit comme facteur multiplicatif de ϕ. Si on note s le facteur d'échelle, on obtient:

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})).$$

La distribution à priori $[p_{\mathbf{x}}()]$ est obtenue à partir de la phase de modélisation et trouver le maximum à posteriori est équivalent à minimiser le potentiel obtenu en appliquant la fonction $-\log$:

$$E(\phi, \mathcal{A}) = -\log\left[\prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x}))\right] = -\int_{\Omega} \log(p_{\mathbf{x}}(s\,\phi(\mathbf{x})))\,d\mathbf{x}$$

Modèle de forme statique: Si l'on considère un modèle de forme le plus simple possible, i.e. une seule forme à priori $\phi_{\mathcal{M}}$, nous devons définir la probabilité du front ϕ, modulo la transformation $\mathcal{A}$. Une distribution Gaussienne en chaque voxel est considérée:

$$p_{\mathbf{x}}(s\,\phi(\mathbf{x})) = p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(s\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2}{2\sigma^2}\right),$$

où σ représente la confiance en la forme à priori $\phi_{\mathcal{M}}$. En supposant que cette confiance soit la même en tout point de l'image, l'évolution de front et les paramètres de la transformation doit doivent minimiser l'énergie suivante:

$$E_{shape}(\phi, \mathcal{A}) = \frac{1}{2\sigma^2}\int_{\Omega}(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2\,d\mathbf{x}.$$

Le facteur constant σ peut être omis ou remplacé par un poids lorsque ce terme est utilisé comme module dans une énergie incluant un terme image. En limitant l'intégrale au voisinage du niveau zéro de ϕ, nous obtenons l'expression analytique suivante:

$$E_{shape}(\phi, \mathcal{A}) = \int_{\Omega}\delta_{\epsilon}(\phi)(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2\,d\mathbf{x}, \tag{18}$$

où δ_{ϵ} est la version lissée de la fonction Dirac usuelle.

Modèle actif de forme: Lors de la phase de construction du modèle, nous avons analysé les modes de variation principaux dans le jeu d'apprentissage. Incluant cette information, la transformation idéale doit mettre en correspondance le front courant avec la "meilleure" représentation par ensemble de niveaux appartenant à la famille générée par notre modèle. Comme déjà mentionné, une forme $\phi_{\mathcal{M}}$ appartenant a cette classe peut être obtenue à partir des m modes de variation principaux:

$$\phi_{\mathcal{M}} = \phi_s + \sum_{j=1}^{m}\lambda_j\,U_j.$$

Le modèle de forme est alors compose de la représentation moyenne ϕ_s et des modes de variation principaux $\{U_j, j = 1..m\}$. Nous devons estimer $[p(\phi, \mathcal{A}, \lambda|\mathcal{M})]$, la

distribution jointe à priori de l'interface, de la transformation et des poids des modes pour un modèle de forme donné. Une formulation Bayésienne de cette densité donne:

$$p(\phi, \mathcal{A}, \lambda|\mathcal{M}) = \frac{p(\mathcal{M}|\phi, \mathcal{A}, \lambda)}{p(\mathcal{M})} p(\phi, \mathcal{A}, \lambda) = \frac{p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)}{p(\mathcal{M})} p(\phi, \mathcal{A}, \lambda).$$

Le terme constant $p(\mathcal{M})$ peut être ignore et la distribution jointe $p(\phi, \mathcal{A}, \lambda)$ est supposée uniforme. L'interface optimale, la transformation et les poids de modes sont obtenus en maximisant:

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})).$$

Un modèle plus complexe pourrait facilement être déduit à partir de la phase de modélisation en apprenant la distribution marginale $p(\lambda)$. En l'absence d'une telle information, nous gardons l'hypothèse d'une distribution uniforme. De la même manière que pour le modèle statique, $p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x}))$ est approximée par une Gaussienne centrée en $\phi_{\mathcal{M}}$ pour chaque pixel. Nous appliquons le fonction $-\log$ à cette densité et nous limitons l'intégrale autour du niveau zéro de ϕ. Les paramètres optimaux doivent alors minimiser l'énergie

$$E_{shape}(\phi, \mathcal{A}, \lambda) = \int_{\Omega} \delta_{\epsilon}(\phi) \left(s\,\phi(\mathbf{x}) - \left(\phi_s(\mathcal{A}(\mathbf{x})) + \sum_{j=1}^{m} \lambda_j\, U_j(\mathcal{A}(\mathbf{x})) \right) \right)^2 d\mathbf{x}.$$

Extraction d'objets sous contrainte de forme

Dans ce paragraphe nous intégrons les deux différents modèles de forme que nous venons de définir dans la méthode de segmentation présentée dans le chapitre 2.

Comme nous l'avons démontré précédemment, il est possible d'intégrer une information de forme à priori dans l'évolution de front représenté par un ensemble de niveaux. Cette information peut donc être intégrée avec les termes images proposés dans les chapitre 2 et 3. Le terme d'attache aux données conduira l'interface vers l'objets d'intérêt alors que l'à priori sur la forme garantira que la forme reste en accord avec le modèle de la forme de l'objet. Pour cela, nous pouvons intégrer les contraintes de forme dans la formulation Bayésienne du chapitre 2. Dans le cas de l'extraction d'un seul objet dont le contour est donné par un ensemble de niveaux ϕ, la formulation Bayésienne du chapitre 2 s'écrit

$$p(\phi|I) \propto p(I|\phi)p(\phi).$$

Le premier terme de cette expression a été intensivement étudié dans les chapitre 2 et 3 mais le second terme n'a jusqu'ici été utilisé que pour introduire une contrainte de régularité sur l'interface. Nous pouvons maintenant définir un terme plus ciblé en intégrant des modèles de forme. La formulation variationnelle correspondante sera

composée de deux termes: un module image ou d'attache aux données et un module de contrainte sur la forme.

Dans le cas d'un à priori statique sur la forme, la fonctionnelle suivante doit être minimisée:

$$E(\phi, \mathcal{A}) = b\, E_{shape}(\phi, \mathcal{A}) + (1 - b)\, E_{data}(\phi),$$

où E_{shape} est le terme d'attraction vers la forme à priori, E_{data} est le module image et le poids b permet de contrôler l'influence de chacun des termes.

Comme nous l'avons démontré dans le paragraphe précédent, le modèle actif de forme peut aussi être incorporé de manière dynamique dans un tel processus de segmentation en optimisant par rapport aux poids des modes de variation retenus dans le modèle:

$$E(\phi, \mathcal{A}, \lambda) = b\, E_{shape}(\phi, \mathcal{A}, \lambda) + (1 - b)\, E_{data}(\phi).$$

La minimisation de ces deux fonctionnelles par rapport à ϕ et $\mathcal{A}$ est effectuée par descente de gradient. Quand aux poids des modes de variation présents dans la deuxième énergie, leurs valeurs optimales peuvent être facilement estimées en résolvant un système linéaire après chaque évolution du front.

Chapitre 5
Segmentation d'un champ de densité de probabilité 3D: application à l'IRM de diffusion

Dans ce chapitre, nous proposons trois approches originales pour la segmentation d'un champ de densités de probabilités 3D. Ce travail a de nombreuses applications en imagerie médicale, en particulier pour les images de diffusion à résonance magnétique où une fonction représentant le mouvement moyen des molécules d'eau est assignée à chaque voxel. L'extraction automatique de structures importantes dans la matière blanche, telles que le *corpus callosum*, pourrait permettre d'améliorer grandement notre connaissance des connectivités cérébrales. Dans la première approche, nous utilisons une loi Gaussienne multivariée afin d'approximer la distribution des composantes du tenseur de diffusion à l'intérieur de chaque sous région du volume DTI. La seconde technique est basée sur la version symétrisée de la distance de Kullback-Leibler et de la modélisation de sa distribution dans chaque région d'intérêt. Dans la troisième approche, nous considérons la variété statistique de dimension 6 définie par les paramètres du tenseur de diffusion et nous proposons un nouvel algorithme en définissant distance géodésique et moyenne intrinsèque dans cette variété Riemannienne. Les formulations variationnelles du problème conduisent à trois évolutions différentes sur des ensembles de niveaux qui convergent vers la segmentation optimale correspondante. Nous validons ces approches sur des données synthétiques et

nous montrons des résultats prometteurs pour l'extraction du *corpus callosum* et des ventricules latéraux.

Conclusion

Un cadre de travail général pour la segmentation non-supervisée d'images a été présenté dans cette thèse. Une évolution de front basée région a été obtenue à partir d'une formulation Bayésienne du problème. La méthode proposée améliore celles existantes et propose différentes extensions. Cela nous a permis d'extraire des structures avec une grande précision dans une grande variété d'images naturelles et médicales. La principale contribution de cette thèse est la définition d'un cadre général permettant d'incorporer différents attributs pour résoudre le problème de la segmentation dans des cas pratiques. Cet objectif a été atteint en combinant et en améliorant plusieurs techniques de l'état de l'art:

1. Nous avons introduit une évolution de front basée région issue d'une formulation Bayésienne qui permet d'intégrer des modèles statistiques dans une approche géométrique.

2. La représentation par ensembles de niveaux a été introduite pour définir la fonctionnelle de partitionnement, conduisant naturellement à une évolution de front implicite.

3. Des approximation paramétriques et non-paramétriques des densités de probabilité régions ont été estimées de manière dynamique en parallèle à l'évolution de courbe.

4. Des caractéristiques texture et mouvement ont été extraites et introduites dans ce cadre de travail afin de traiter une plus grande diversité d'images.

5. Une approche implicite a été présentée pour modéliser et intégrer une connaissance à priori sur la forme des objets à extraire.

6. Cette approche a finalement été étendue à des données plus complexes à valeurs matricielles, en particulier au cas des images de diffusion a résonance magnétique.

Tout au long de cette thèse, nous étions concentré sur l'aspect non-supervisé des approches présentées. Afin d'éviter des modèles statistiques trop compliqués dont l'estimation des paramètres aurait pu être un frein important, nous avons essayé d'identifier et d'extraire les informations importantes qui définissent les régions d'intérêt dans un type d'image donné. Nous y sommes arrivé avec un certain succès puisque nous avons démontré l'efficacité de nos approches en fournissant de meilleurs

résultats que de nombreux travaux récents et avec une complexité moindre. Cependant, en vu d'une robustesse accrue, l'apprentissage des quelques paramètres introduits et l'utilisation d'estimateurs plus robustes pourraient être considérés. Nous pensons que ce cadre de travail est une bonne base pour élaborer des systèmes efficaces d'analyse et de reconnaissance d'images.

Nous concluons sur une vision plus générale du problème de segmentation pour les images naturelles. Sur ce type d'images, ce que nous essayons d'accomplir en intégrant différents attributs et de la connaissance à priori, c'est d'imiter le complexe et toujours mystérieux système visuel humain. Les progrès effectués ces dernières années en imagerie cérébrale apportent des moyens inespérés pour étudier ce système complexe et pour comprendre comment nous, humains, effectuons ces taches d'extractions d'objets. Afin de franchir les limites des méthodes d'analyse d'images actuelles et de les rapprocher de nos propres performances, il pourrait être bénéfique de réunir informaticiens et neurologistes afin de créer des modèles motivés biologiquement de plus en plus proche de la réalité [80].

Chapter 1

Image Segmentation by Front Evolution

Curve evolution was first proposed for image segmentation in the seminal work of Kaas et al. [74]. This technique tries to find edges by evolving a curve in the image domain, often referred to as a snake. The curve evolution was initially based on the minimization of an energy functional defined along the curve, reflecting curve regularity and overlap with image edges. In [74], the initial energy is composed of three terms: the first two ones stand for intrinsic properties of the curve like smoothness and elasticity while the last one is an external term whose role is to attract the curve toward image edges. This work was a breakthrough in the very early development of image segmentation and a whole part of the domain has been influenced by this view of the problem. Various direct extensions of the *Snake* model were rapidly proposed to overcome some of its important limitations. The principal problem was the need for a very good initialization to obtain the expected result. The works of Blake and Zisserman [15], Cohen [35], Berger [9], Fua and Leclerc [58] are some examples of such contributions. However, these approaches did not remove the dependency on the parameterization and the object topology had still to be known in advance. Then, a step was taken when the *Level-Set* representation was introduced for front propagation. First discovered by Dervieux and Thomasset in [47, 48], it was later rediscovered and more largely popularized by Osher and Sethian [106]. This implicit and intrinsic representation of the contour gives a solution to the two main limitations of the active contours mentioned above, namely the changes of topology and the parameterization of the contour. Geometric properties of the curve can be easily approximated from the *Level-Set* representation, like the curvature, the normal and tangent vectors. Therefore, various geometric flows have been proposed, some based on an energy minimization [76, 77, 28] and others [27, 34, 91]. Efficient schemes have also been proposed for their

implementation, a detailed description can be found in [1]. Moreover, convergence and stability as well as mathematical proof of existence and uniqueness of the solution have been widely studied [132]. The *Geodesic active contour* model [76, 77, 28] is certainly the most elegant one, where the segmentation problem is reformulated as the one of finding a minimal geodesic path in a Riemannian space induced by image properties.

A more recent approach is the introduction of region-based terms to control the evolution of the curve [171, 167, 30, 107, 122, 68]. While classical active contours or front propagation consider only gradient information to define a stopping criterion, region-based approaches use more global partitioning formulation which integrates quantities over the whole image domain. Contrary to gradient-based approaches which rely on local measures, they are more robust to the initial conditions and there are fewer critical stopping functions to be tuned. Region-based active contours can be classified in two classes: the first one is based on the Mumford-Shah functional or on Total Variation minimization while the second one comes from a Bayesian formulation.

In the following, we give a detailed insight into the evolution of these techniques. In the first part, we introduce classical contour evolution where the evolving curve aims at finding image discontinuities/high gradients. The second part focuses on recent contributions to region-based curve evolution techniques.

1.1 Boundary-based curve evolutions: from *Snakes* to geodesic active contours

1.1.1 The *Snake* model

The classical *Snake* model aims at finding image boundaries by moving a contour/snake whose evolution is guided by internal as well as external forces. It was initially formulated by the means of an energy minimization by Kaas et al. in [74]. Let $I : \Omega \in \mathbb{R}^2 \rightarrow \mathbb{R}^+$ be a scalar image, the *Snake* functional is defined along a parameterized contour $\mathcal{C} : [0, 1] \rightarrow \Omega$:

$$E(\mathcal{C}) = \int_0^1 \alpha |\mathcal{C}'(p)|^2 \, dp + \int_0^1 \beta |\mathcal{C}''(p)|^2 \, dp - \lambda \int_0^1 |\nabla I(\mathcal{C}(p))| \, dp. \tag{1.1}$$

The first two terms stand for internal properties of the curve, rigidity and elasticity respectively, while the last term is the external energy integrating image data. It is intended to be minimum when the contour overlaps image boundaries (high gradients). This seminal model has important limitations and numerous extensions and generalizations have been proposed in the literature. The data term is for example quite restrictive: image intensities may have quite a different dynamic in different images and thus, a more general term should be considered.

An other important issue to be dealt with is the minimization of this energy. Simple gradient descent-based methods were initially proposed to update a parametric repre-

sentation of the contour until convergence. In this direct approach, the functional has the disadvantages that it depends on the parameterization of the curve and topological changes are not allowed. To resolve the parameterization issue, different types of basis functions have been proposed [174, 117, 6, 44]. The most promising are B-splines [44] for evolving contours in 2D images and finite elements [36, 95] for evolving surfaces in 3D images or in higher dimension. Topological changes have also been introduced in several works [95]. The level-set framework of Osher and Sethian [106] provides a new way of representing evolving fronts which permits both issues to be resolved. The major negative aspect of this new representation, compared to classical parametric contours, is the additional computational cost. However, efficient and stable algorithms are now available.

In summary, parametric active contours are still the most time-efficient and they are often favored to level-sets in real-time applications. However, the level-set framework permits new constraints to be defined, without worrying about the type of parameterization. Moreover, it is often possible to restrict the level-set evolution to a small narrow-band around its zero level, decreasing dramatically the complexity. In the next paragraph, we introduce very briefly this representation for evolving contours or surfaces.

1.1.2 The level-set representation for front evolutions

The level-set representation was introduced for flame propagation by Osher and Sethian in [106] and later for image segmentation by Malladi et al. in [92] (let us note that the same idea was already present, almost ten years earlier, in the work of Dervieux and Thomasset [47, 48]). This framework is well-suited for front evolution and it has naturally become the standard way of expressing contour evolution techniques [105]. Level-sets are implicit and intrinsic representations of interfaces of any dimension, topological changes are naturally possible, efficient and stable numerical schemes have been proposed and theoretical correctness can be provided by the theory of viscosity solutions [39, 132].

The basic idea is to see the interface between two regions as the zero level of a level-set function ϕ which is defined on the whole image domain $\Omega \in \mathbb{R}^p$. An evolving front $\mathcal{C} : [0,1] \times \mathbb{R}^+ \rightarrow \mathbb{R}^p$ is then implicitly represented by a Lipschitz function $\phi : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$, i.e. verifying:

$$\exists C \in \mathbb{R}^+, \forall \mathbf{x}, \mathbf{y} \in \Omega, \quad |\phi(\mathbf{x}) - \phi(y)| \leq C\|\mathbf{x} - \mathbf{y}\|.$$

A common choice is the signed distance function which has other nice properties to express geometric properties of the curve. It is defined as follows:

$$\phi(\mathbf{x};t) = \begin{cases} 0, & \mathbf{x} \in \mathcal{C}(t) \\ +\mathcal{D}(\mathbf{x}, \mathcal{C}(t)) > 0, & \mathbf{x} \in \mathcal{C}_{in}(t) \\ -\mathcal{D}(\mathbf{x}, \mathcal{C}(t)) < 0, & \mathbf{x} \in \mathcal{C}_{out}(t) = [\Omega - \mathcal{C}_{in}(t)] \end{cases}$$

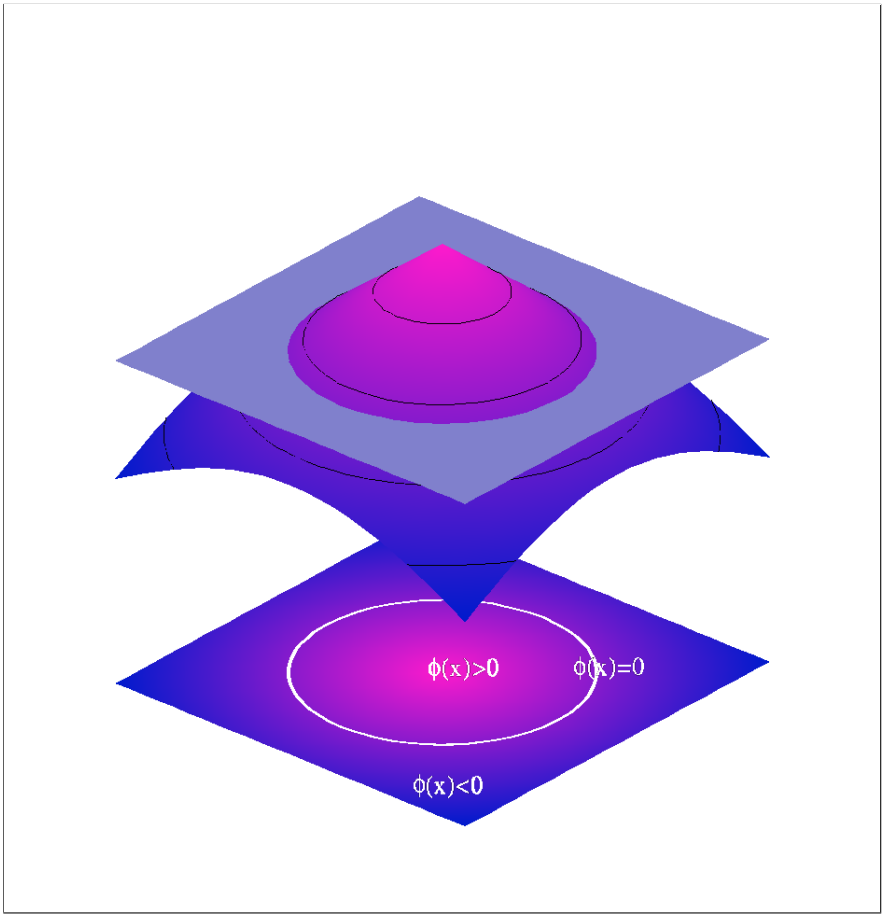


Figure 1.1: Illustration of a level-set function obtained for a circle.

where $\mathcal{D}(\mathbf{x}, \mathcal{C}(t))$ is the minimum Euclidean distance between the pixel $\mathbf{x}$ and $\mathcal{C}(t)$. The signed distance function obtained for a circle is shown in [Figure 1.1]. Geometric properties of the interface can be easily expressed with its implicit representation:

- The inward unit normal vector to $\mathcal{C}$ is $\mathcal{N} = -\frac{\nabla\phi}{|\nabla\phi|}$,
- The mean curvature at any point of $\mathcal{C}$ can be directly estimated by: $\kappa = \nabla \cdot \frac{\nabla\phi}{|\nabla\phi|}$.

These relations permit to rewrite easily classical contour evolutions. For example, let us consider a simple front evolution without tangential component: $C_t = F\mathcal{N}$. For each point $\mathbf{x}(t) \in \mathcal{C}(t)$, we have:

$$\frac{\partial \mathbf{x}(t)}{\partial t} = F(\mathbf{x}(t)) \frac{\nabla\phi(\mathbf{x})}{|\nabla\phi(\mathbf{x})|} \quad \text{and} \quad \phi(\mathbf{x}(t), t) = 0.$$

Then, applying the chain rule to the second relation, we obtain the following equality:

$$\frac{\partial \phi(\mathbf{x}(t), t)}{\partial t} + \nabla\phi(\mathbf{x}(t)) \cdot \frac{\partial \mathbf{x}(t)}{\partial t} = 0.$$

Now, we can deduce the evolution equation for the level-set function:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0, \tag{1.2}$$

while the initial level-set $\phi(\mathbf{x}, t = 0)$ is given by the initial curve $\mathcal{C}(t = 0) = \mathcal{C}_0$. For certain forms of speed functions F, the equation 1.2 is a standard Hamilton-Jacobi equation and the corresponding theoretical analysis applies [132].
General curve evolutions with possible tangential speed ($\mathcal{C}_t = \mathbf{v}$) can be also embedded in a level-set evolution. The corresponding evolution equation is a convection of ϕ by the velocity field v:

$$\frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi = 0. \tag{1.3}$$

Before ending this section, we review an important point of the level-set formulation, namely the "re-distancing" step. This implicit way of propagating fronts remains valid only if the level-set ϕ remains Lipschitz function. This has been an important issue since there is no guarantee from the evolution equations that the level-set embedding maintains such properties. The easier way to deal with this limitation is to reinitialize the level-set to the distance function when is needed. Several methods have been proposed for this task, one of the most used aims at solving the following PDE [142]:

$$\frac{\partial \phi}{\partial t} = (1 - \text{sign}\,(\phi_0))\,(1 - |\nabla \phi|)\,,$$

where ϕ_0 is the level-set we wish to reinitialize. Another efficient re-distancing algorithm is based on the Fast Marching technique [132] and has a very low complexity in time but an interpolation step around the zero level makes it less accurate. From our experience, a mixed approach appears to be the most accurate and efficient: the PDE is used for reinitialization in a small neighborhood of the zero level while the Fast Marching permits to extend the distance function to a larger band. Let us also mention the elegant work of Gomes et al. [62] where the level-set evolution equation was rewritten so that the level-set remains a distance function. However, this modification may decrease the convergence speed of the evolution and the re-distancing step is often preferred.

In this section, we have presented a general formulation which applies to evolving interfaces of any dimension. The next section will focus on the influence of this framework on boundary-based curve evolution techniques. Then, the second section will present region-based evolutions which also make use of the level-set formulation. In particular, we will see how the level-set representation can be also useful as an optimization framework to define boundary as well as regional partitioning functionals.

1.1.3 Geometric active contours

The level-set representation [106] was initially used for geometric active contours by Caselles et al. [27], Chopp et al. [34] and Malladi et al. [91]. They are based on

the *geometric heat equation* $C_t = \kappa\mathcal{N}$, which permits to express internal properties of the curve as in the *Snake* model by smoothing the curve. This flow can be directly expressed with the level-set function ϕ:

$$\frac{\partial\phi}{\partial t} = |\nabla\phi| \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right). \tag{1.4}$$

Following the idea of the *Snake* model, this equation must be modified in two ways to be able to detect edges: (1) the evolving interface should be attracted toward the edges and, (2) it should stop on the edges. For this purpose, a function of the image gradient is generally defined such that:

$$\begin{cases} g\big(|\nabla\hat{I}|\big) \to 0 & \text{, if} \quad |\nabla\hat{I}| \to \infty \\ g\big(|\nabla\hat{I}|\big) \to 1 & \text{, if} \quad |\nabla\hat{I}| \approx 0 \end{cases}$$

where $\hat{I}$ is a smoothed version of the image, obtained from a convolution with a Gaussian kernel. This function can be seen as an edge detector and a common choice is:

$$g\big(|\nabla\hat{I}|\big) = \frac{1}{1+|\nabla\hat{I}|^n} \quad , n = 1 \text{ or } 2.$$

The heat flow is then modified as follows:

$$\frac{\partial\phi}{\partial t} = g\big(|\nabla\hat{I}|\big)|\nabla\phi| \left(\operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \nu\right). \tag{1.5}$$

where ν is a balloon force [35] whose role is to push the curve toward a given direction. This geometric flow has some nice properties: it will shrink or expand the curve (depending on the sign of $\nu + \kappa$) toward the image edges and thanks to the function g, its speed will fall to zero on the edges. Moreover, the method benefits from the level-set formulation and the contour can split or merge, making it able to detect objects with arbitrary topology. However, the contour remains to be initialized completely inside or outside the objects, the balloon force ν being set respectively positive or negative. An other important practical limitation is the definition of the stopping function g. This can be an arduous task because image gradients of edges may vary a lot for different images and even inside one image. A badly defined stopping function will make the contour go through the edge without any possible backward propagation.

1.1.4 Geodesic active contours

Geodesic active contours were proposed by Caselles et al. in [28] and by Kichenassamy et al. in [76, 77]. Being a natural extension of geometric active contours, the geometric flow is obtained from the minimization of a given functional and an additional term that improves the evolution appears. In both papers, [28] and [76, 77], similar flows are obtained but following slightly different formulations.

In [28], the authors start from the classical snakes approach [74]. Neglecting the rigidity term, the snakes model is modified to:

$$E(\mathcal{C}) = \alpha \int_0^1 |\mathcal{C}'(p)|^2 \, dp + \lambda \int_0^1 g(|\nabla I(\mathcal{C}(p))|)^2 \, dp, \tag{1.6}$$

where $g()$ is function of the image gradient as the one defined in the last paragraph. While only the ratio $\frac{\alpha}{\lambda}$ has to be fixed, this energy still depends on the parameterization of the curve. Thanks to the Maupertuis'Principle, Caselles et al. proposed in [28] to reformulate the problem of minimizing (1.6) by the one of finding the curve of minimal geodesic length in a Riemannian space whose metric is induced by image gradients[1]. The reformulation introduces a single free parameter corresponding to the difference between the internal and external parts of the energy (1.6), i.e. the trade-off between α and λ. The choice of this parameter is discussed [28]. Setting it to 0, minimizing (1.6) is shown to be equivalent to the minimization of the intrinsic energy:

$$E(\mathcal{C}) = \int_0^1 g(|\nabla I(\mathcal{C}(p))|)|\mathcal{C}'(p)| \, dp = \int_0^{\mathcal{L}(\mathcal{C})} g(|\nabla I(\mathcal{C}(p))|) \, ds, \tag{1.7}$$

where $ds = |\mathcal{C}'(p)| \, dp$.

Then, the problem of fitting the curve to the image edges is translated to the one of finding geodesics of minimal distance curves in a Riemannian space by minimizing (1.7). Taking the first variations of (1.7) (we refer to [76, 77, 28] for details), the direction that minimizes the length as fast as possible is given by:

$$\frac{\partial \mathcal{C}}{\partial t} = g()\kappa\mathcal{N} - (\nabla g() \cdot \mathcal{N})\mathcal{N}.$$

This curve evolution can be expressed with a level-set function and the corresponding flow is then

$$\frac{\partial \phi}{\partial t} = g() \, |\nabla \phi| \, \mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \nabla g() \cdot \nabla \phi.$$

This evolution equation shares almost the same first term with the geometric evolution equation (1.5). The additional weight ν is generally also introduced in this new evolution to improve convergence speed. Then, we obtain the final geodesic active contour model:

$$\frac{\partial \phi}{\partial t} = g() \, |\nabla \phi| \, \left(\mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \nu\right) + \nabla g() \cdot \nabla \phi. \tag{1.8}$$

Now, the only difference with geometric active contours is the new term $\nabla g() \cdot \nabla \phi$. This term improves the geometric flow in several aspects: the attraction toward the edges is increased when the contour gets close to them, the curve evolution is no more "one way" and the cases where the contour did not stop on low contrast edges are

[1] The equivalence between geodesic active contours and a modified version of the classical *Snake* model was further analyzed in [5].

more easily handled without setting a complex stopping function, the weight ν is only an optional parameter that is introduced to increase the convergence speed.

Despite the improved flexibility of this last evolution becomes in terms of initialization and choice of boundary function, this approach remains very local. The contour will be attracted to the closest edges and in homogeneous parts, the speed of the evolution will be only controlled by the artificial force ν. Several extensions have been proposed to overcome this last limitation. Siddiqi et al. proposed in [133] a shape and area minimizing flow whose level-set form is:

$$\frac{\partial \phi}{\partial t}(\mathbf{x}) = \alpha \left(g() \, |\nabla \phi| \, \mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \nabla g() \cdot \nabla \phi \right) + \frac{1}{2} \, \mathrm{div}\left(\mathbf{x} g() \right) |\nabla \phi|, \qquad (1.9)$$

where the second term minimizes the area enclosed by the contour and provides the inflatory term ν, artificially introduced before, as well as a new attraction force in the vicinity of edges.

Gradient Vector Flows [166] lead to important improvements. They permit to attract the contour to an edge for any initialization by normalizing and extending edge attraction terms to the whole image domain. This extension permits also to increase the robustness with respect to holes along boundaries, hence reducing leak problems. However, the contour will still converge toward the closest edges and important object boundaries may be missed. In several applications, a good guess of the object boundaries may be available and this will not be an issue. However, without such knowledge, these techniques require manual initializations. Relying on more global information (not only along the curve) may help in improving the robustness with respect to the initial curve position. As we will see in the next section, the minimization of a region-based functional drives to a new curve evolution where the region-based part can be seen as an adaptive balloon force which makes the approach much more independent to the initialization.

Before starting with region-based front evolutions, we want to emphasize that geodesic active contours and area minimizing flow have been deeply studied and mathematical proofs of correctness and existence of the solution are established.

1.2 Region-based contour evolutions

Region-based contour evolutions are defined from a global functional and they are intended to give a complete partitioning of the image domain. This vision has been shown to be quite promising to improve classical active contours and many recent works are based on this approach [102, 81, 171, 167, 30, 31, 144, 107, 122, 120, 68]. Before describing some of these works, we first show how the level-set formulation can be an efficient framework to define regional criteria. Then, we present two classes of region-based contour evolution techniques: the first one is based on the Mumford-Shah functional while the second one comes from a Bayesian formulation.

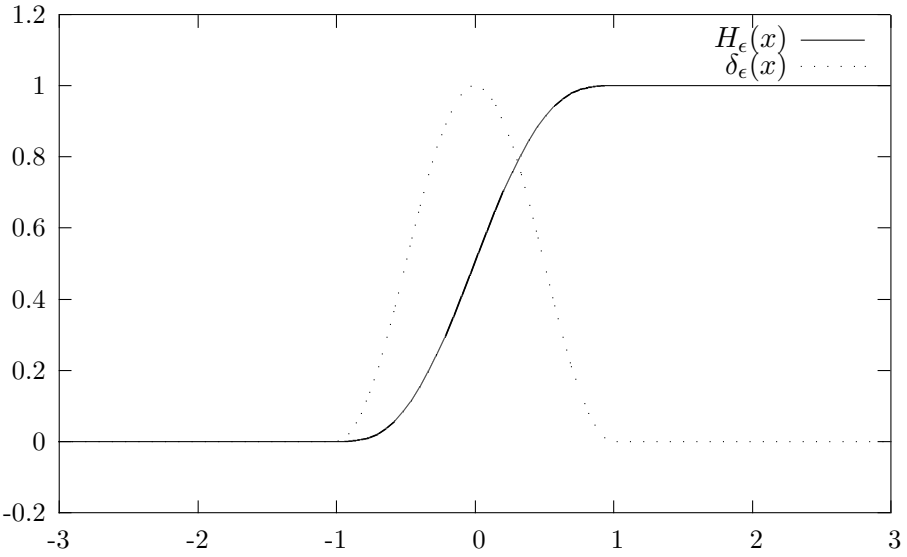


Figure 1.2: Heaviside and Dirac functions with $\epsilon = 1$

1.2.1 The level-set representation as an optimization framework

The level-set representation can be also usefull as an optimization framework [170]. Let δ and H denote the usual Dirac and Heaviside functions. These functions permit to define quantities along the curve or over the inside/outside region:

$$E_{curve} = \int_\Omega |\nabla H(\phi)|\, f(\mathbf{x}, \mathcal{C})\, d\mathbf{x} = \int_\Omega \delta(\phi)|\nabla\phi|\, f(\mathbf{x}, \mathcal{C})\, d\mathbf{x},$$

$$E_{in} = \int_\Omega H_\alpha(\phi(\mathbf{x}))\, f_{in}(\mathbf{x}, \mathcal{C})\, d\mathbf{x}, \qquad E_{out} = \int_\Omega (1 - H_\alpha(\phi(\mathbf{x})))\, f_{out}(\mathbf{x}, \mathcal{C})\, d\mathbf{x}.$$

For example, we can obtain the length of the interface from the first energy by setting f to 1 and the other ones will furnish the area of the respective region for a similar choice. Regularized versions of the Heaviside and Dirac functions must be used for implementation, a possible choice is the following one:

$$H_\epsilon(\phi) = \begin{cases} 1, & \phi > \epsilon \\ 0, & \phi < -\epsilon \\ \dfrac{1}{2}\left(1 + \dfrac{\phi}{\epsilon} + \dfrac{1}{\pi}\sin\left(\dfrac{\pi\phi}{\epsilon}\right)\right), & |\phi| < \epsilon. \end{cases}$$

$$\delta_\epsilon(\phi) = \begin{cases} 0, & |\phi| > \epsilon \\ \dfrac{1}{2\epsilon}\left(1 + \cos\left(\dfrac{\pi\phi}{\epsilon}\right)\right), & |\phi| < \epsilon. \end{cases}$$

A general region-based partitioning objective functional can be written by associating a penalty function e_i to each pixel of the region Ω_i. Assuming the number

of regions to be known and equal to N, a continuous formulation can be written by integrating these "error" functions on the corresponding regions:

$$E(\{\Omega_i, i = 1..N\}) = \sum_i \int_{\Omega_i} e_i(\mathbf{x})\, d\mathbf{x}. \tag{1.10}$$

The simple binary case of bi-partitioning can be easily expressed within the level-set framework. Using the same notations as in the previous section, one level-set ϕ stands for the implicit representation of the interface between Ω_1 and Ω_2. The expression of (1.10) for $N = 2$ is then:

$$E(\phi) = \int_\Omega \Big(H_\epsilon(\phi(\mathbf{x}))e_1(\mathbf{x}) + (1 - H_\epsilon(\phi(\mathbf{x})))e_2(\mathbf{x}) \Big)\, d\mathbf{x}. \tag{1.11}$$

The generalization to N regions is more challenging. The idea is to couple a set of level-set functions $\Phi = \{\phi_1, \ldots, \phi_n\}$ by defining a characteristic function $\chi_i(\Phi)$ for each Ω_i such that:

$$\begin{cases} \chi_i(\mathbf{x}) > 0, & \text{if } \mathbf{x} \in \Omega_i \\ \chi_i(\mathbf{x}) = 0, & \text{otherwise} \end{cases} \tag{1.12}$$

Two other constraints must be added to respect the partitioning definitions $\cup_{i=1}^{N} \Omega_i = \Omega$ and for all $i \neq j$, $\Omega_i \cap \Omega_j = \varnothing$.

A straightforward choice to verify (1.12) is to associate one level-set ϕ_i to each region Ω_i and then the corresponding characteristic function is defined as $H_\epsilon(\phi_i)$. A coupling term must be generally added if the energy minimization does not maintain the partitioning constraint valid as in [170]. Yezzi et al. also proposed in [167] to couple several contour evolutions to segment an arbitrary number of regions. Using N coupled contour evolutions, they are able to segment the image in $N+1$ regions. Also, the coupling between the curves is not an "artificial" additive term but it is obtained from the definition of binary and ternary flows.

In [30], Chan and Vese proposed an other way of defining these characteristic functions that has two nice advantages: (i) only $\log(N)$ level-sets are necessary to represent N regions and (ii), the partitioning constraint is naturally maintained by construction. Their idea is to define a characteristic function from each logical combination of the level-sets' signs. For example, if two level-sets ϕ_1 and ϕ_2 are considered, 4 regions can be represented by the combinations $\{(\phi_1 > 0, \phi_2 > 0), (\phi_1 < 0, \phi_2 > 0), (\phi_1 > 0, \phi_2 < 0), (\phi_1 < 0, \phi_2 < 0)\}$. The re-definition of the energy with a combination of Heaviside functions of the level-sets is direct.

The next two parts present two different classes of region-based contour evolutions for image segmentation. The first class starts from the Mumford-Shah functional and is purely based on geometric properties of images. Two cases are presented, both trying to solve the segmentation problem within the level-set framework. The first one is

restricted to a simplification of the general Mumford-Shah functional by assuming the image to be well-approximated by a piece-wise constant model. The second one proposes an optimization framework for the general case of a piece-wise smooth model. The second class of methods is based on a Bayesian formulation where geometric and statistical concepts are combined. Geometric properties are intrinsic to the contour evolution technique while statistical properties of each region can be learnt from the intensity histogram or from filter responses. Moreover, several links between these two different approaches for particular/asymptotic cases will be presented.

1.2.2 The Mumford-Shah functional

The *week membrane model* [102, 15] assumes that the image can be approximated by a piece-wise smooth image. Mumford and Shah proposed in [102] to obtain this "ideal" representation u of the image I by minimizing the following functional:

$$E_{MS}(u,\mathcal{C}) = \int_{\Omega} |u(\mathbf{x}) - I(\mathbf{x})|^2 \, d\mathbf{x} + \mu \int_{\Omega\setminus\mathcal{C}} |\nabla u(\mathbf{x})|^2 \, d\mathbf{x} + \nu|\mathcal{C}|. \tag{1.13}$$

The interpretation of the three terms of this energy is straightforward: the first one is the usual mean-square data term, the second one makes the cartoon model smooth everywhere but on the image discontinuities and the last one enforces the boundaries $\mathcal{C}$ to be smooth too. This functional is closely related to the general total variation minimization functional [127] but here, image discontinuities are introduced explicitly with $\mathcal{C}$. However, since the level-set representation permits to express implicitly these discontinuities, both functionals may be equivalent by restricting u to a certain family of functions. Such examples can be found in the first chapter of [105].

Active Contours Without Edges - Chan and Vese [30]

Rather than minimizing the Mumford-Shah functional (1.13) for a general function u, Chan and Vese considered in [30] an easier one, "the minimal partition problem" [103], assuming u piecewise constant. If $\{\Omega_i, i = 1..N\}$ is a partitioning of the image and $u(\mathbf{x}) = c_i$ for $x \in \Omega_i$, the Mumford-Shah functional (1.13) can be simplified as:

$$E^0_{MS}(u,\mathcal{C}) = \sum_{i=1}^{N} \int_{\Omega_i} |c_i - I(\mathbf{x})|^2 \, d\mathbf{x} + \nu|\mathcal{C}|. \tag{1.14}$$

In [30], the authors proposed to represent implicitly the boundaries $\mathcal{C}$, using a combination of level-set functions. In the simplest case of bi-partitioning, one level-set is sufficient and the above energy can be re-written the following way:

$$\begin{aligned} E^0_{CV}(c^+,c^-,\phi) = &\int_{\Omega} |c^+ - I(\mathbf{x})|^2 H_\epsilon(\phi(\mathbf{x})) \, d\mathbf{x} + \int_{\Omega} |c^- - I(\mathbf{x})|^2 (1 - H_\epsilon(\phi(\mathbf{x}))) \, d\mathbf{x} \\ &+ \nu \int_{\Omega} |\nabla H_\epsilon(\phi(\mathbf{x}))| \, d\mathbf{x}. \end{aligned} \tag{1.15}$$

This energy must be minimized with respect to two types of parameters, the region constants c^+ and c^- and the level-set function ϕ. A two-step approach is employed [30]: at fixed ϕ, the optimal constant c^+ and c^- are nothing but the empirical intensity means in the corresponding regions: $c^+ = \int_\Omega I\, H_\epsilon(\phi) / \int_\Omega H_\epsilon(\phi)$ and $c^- = \int_\Omega I\,(1 - H_\epsilon(\phi)) / \int_\Omega (1 - H_\epsilon(\phi))$, and the level-set is updated according to the following evolution equation:

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \left(\nu \,\mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|} \right) - |c^+ - I|^2 + |c^- - I|^2 \right). \tag{1.16}$$

This method is quite efficient to segment scalar images that can be modeled by a piecewise constant function. In such cases, the regularization term permits to deal with noisy data by imposing spatial regularity of the contour. However, no noise model is integrated and thus, the regularization weight ν must be increased for highly noisy images. This can bring two important drawbacks: (i) sharp edges of the objects to be detected will be smoothed and (ii), the convergence[2] speed decreases when ν increases because the schema is not stable. In Section 1.2.3, a similar contour evolution is obtained from a Bayesian formulation which can be seen as a direct extension of this model but integrating intensity variances over each region.

The case of N regions is also considered with the coupling of $\log(N)$ level-sets as presented before. In [29], the authors proposed a similar method to deal with vector-valued images. Next to the straightforward use of vector means as region approximations, logical operators can be defined to design channel-selective segmentation strategies.

Level-set implementation of the Mumford-Shah functional

The extension of the Active Contours Without Edges [30] approach to the more general piecewise smooth model was proposed in [31] and [144]. The idea is still to use a combination of level-set functions to represent the boundaries but a general smooth function u_i is considered instead of a constant to model image intensity on each region Ω_i. The corresponding bi-partitioning functional is:

$$\begin{aligned} E_{CV}(u^+, u^-, \phi) = & \int_\Omega |u^+(\mathbf{x}) - I(\mathbf{x})|^2 H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x} \\ & + \int_\Omega |u^-(\mathbf{x}) - I(\mathbf{x})|^2 (1 - H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x} \\ & + \mu \int_\Omega |\nabla u^+(\mathbf{x})| H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x} + \mu \int_\Omega |\nabla u^-(\mathbf{x})| (1 - H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x} \\ & + \nu \int_\Omega |\nabla H_\epsilon(\phi(\mathbf{x}))|\, d\mathbf{x}. \end{aligned} \tag{1.17}$$

[2] The existence of a minimizer for this functional and more generally for the Mumford-Shah functional has been proved in [103].

This energy must be minimized with respect to the three different functions u^+, u^- and ϕ. In both works [31, 144], the authors obtained the same system of coupled equations:

$$\begin{aligned} u^+ &= I + \mu\Delta u^+ \text{ on } \{\phi > 0\}, \frac{\partial u^+}{\partial \mathcal{N}} = 0 \text{ on } \{\phi = 0\} \cup \mathcal{C}, \\ u^- &= I + \mu\Delta u^- \text{ on } \{\phi < 0\}, \frac{\partial u^-}{\partial \mathcal{N}} = 0 \text{ on } \{\phi = 0\} \cup \mathcal{C}, \\ \frac{\partial \phi}{\partial t} &= \delta_\epsilon(\phi)\left(\nu \,\text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - |u^+ - I|^2 + |u^- - I|^2 - \mu|\nabla u^+|^2 + \mu|\nabla u^-|^2\right). \end{aligned} \tag{1.18}$$

This approach permits to segment and restore an image at the same time. The segmentation takes benefit from the smoothed image while the smoothing integrates explicitly the contour position as discontinuities. Contrary to the piecewise constant model, unconnected components may have different intensity properties and thus, only two coupled level-sets permit to extend the method to an arbitrary number of regions, thanks to the Four Color Theorem [31].

However, several issues remain open, as the balancing strategy between the two parallel optimization processes. As mentioned in the seminal work of Mumford and Shah, it is still not clear how far the two Poisson equations in (1.18) should be conducted at each iteration step of the contour evolution. The time complexity may also be critical since after each displacement of the curve, a new smoothing process has to be done on each part of the image.

An other important limitation is the piece-wise smooth model which is limited to a very small set of images. As soon as regions are textured, the method cannot be applied. In the next section, we present a parallel way of tackling non-smoothed images which leaves out the Mumford-Shah model but introduces statistical analysis over regions within a variational framework.

1.2.3 Bayesian formulation

The Bayesian formulation for image segmentation is based on the assumption that pixel intensities within each region are the realization of a "random" process with a given density function. Let p_i be the probability distribution in the region Ω_i, Paragios and Deriche showed in [107] that assuming pixel intensities to be independent realizations of the corresponding random process, the *maximum a posteriori* partitioning of the image can be obtained by minimizing the energy:

$$E(\{\Omega_i, i = 1..N\}) = -\sum_i \int_{\Omega_i} \log p_i(I(\mathbf{x}))\, d\mathbf{x} \tag{1.19}$$

The spatial structure of the image is only visible through the unknown regions Ω_i and if no additional spatial constraint is imposed, this functional is a simple clustering

criteria of a set of random variables. This equation is spatially constrained by adding a regularization term on the interface $\mathcal{C}$ between the regions Ω_i:

$$E(\{\Omega_i, i = 1..N\}) = -\sum_i \int_{\Omega_i} \log p_i(I(\mathbf{x}))\, d\mathbf{x} + \nu|\mathcal{C}| \tag{1.20}$$

Several works are based on this functional. For example, in [171], Zhu et al. considered this functional for gray, color and textured images and they showed the equivalence with a minimum descriptor length (MDL) based criteria [81] as well as the Mumford-Shah functional by choosing Gaussian probability densities and setting $\mu \to \infty$ in (1.13).

Geodesic active regions

The geodesic active regions (GAR) framework was presented by Paragios et al. in [107]. Their approach combines the MAP region-based term of equation (1.20) with the geodesic active contour model (1.7). This model permits to obtain a speed function that is controlled by global region properties while accuracy on boundaries is guaranteed by the boundary term. The objective functional of the GAR model is:

$$E(\{\Omega_i, i = 1..N\}) = -\sum_i \int_{\Omega_i} \log p_i(I(\mathbf{x}))\, d\mathbf{x} + \nu \int_0^1 \left|\frac{\partial \mathcal{C}}{\partial p}\right| g()\, dp \tag{1.21}$$

Another contribution in [107] is the generalization of the boundary function $g()$, which was initially a direct function of image gradients, to a function of a predefined probability density function of edge points. This probability density function is usually learnt from a selected set of edge detectors. For the minimization of the energy (1.21), the authors introduced one level-set to represent each region. A partitioning in N regions of the image is then obtained through the following system of evolution equations:

$$\begin{cases} \forall i \in [1, N] \\ \dfrac{\partial \phi_i}{\partial t} = |\nabla \phi_i| \left(\nu \left(g()\kappa_i + \nabla g() \cdot \dfrac{\nabla \phi_i}{|\nabla \phi_i|} \right) + \log \dfrac{p_i(I)}{p_{O_i}(I)} \right) \end{cases} \tag{1.22}$$

where O_i is the competing region at the current pixel. A coupling force was also proposed to avoid vacuum and overlaps between regions, driving to a more complex system.

Until now, region probability densities were not specified. A straightforward approach is to apply a clustering technique on the image intensity histogram. In [107], Gaussian densities or mixtures of Gaussian densities (for textured images) were learnt a priori from image intensity or filter responses (for textured images). In the early *Region Competition* approach proposed by Zhu et al. [171], single Gaussians where used for each region but their parameters were estimated dynamically during the curve evolution process, following the *Expectation-Minimization* algorithm [65]. In the following section, we revisit this approach using the level-set framework and we show the

similitude with the *Active Contours Without Edges* model of section 1.2.2 and other recent works.

Expectation-minimization approach

Following [122, 120], we introduce the statistical parameters as unknowns. For the sake of simplicity, we start with the simplest case of bi-partitioning and we consider only the Euclidean curve length as regularization. The image is assumed to be composed by two regions Ω_1 and Ω_2, and each region intensity distribution can be approximated by a single Gaussian:

$$p_i(I|\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(I-\mu_i)^2}{2\sigma_i^2}}$$

Using one level-set ϕ as implicit representation of the interface between Ω_1 and Ω_2, the corresponding MAP energy (1.21) can be written as follow:

$$\begin{aligned} E(\phi,\{\mu_{1,2},\sigma_{1,2}^2\}) = &\int_\Omega H_\epsilon(\phi)\left(log(2\pi\sigma_1^2) + \frac{(I(\mathbf{x})-\mu_1)^2}{2\sigma_1^2}\right) d\mathbf{x} \\ &+ \int_\Omega (1-H_\epsilon(\phi))\left(log(2\pi\sigma_2^2) + \frac{(I(\mathbf{x})-\mu_2)^2}{2\sigma_2^2}\right) d\mathbf{x} + \nu|\mathcal{C}| \end{aligned} \tag{1.23}$$

This energy is minimized by following an *Expectation-Minimization* algorithm: at fixed ϕ, the optimal statistical parameters are estimated, then the energy is minimized with respect to ϕ and both steps are iterated until convergence. For a fixed level-set function, close form of the optimal statistical parameters can be easily obtained from the derivation:

$$\begin{cases} \mu_1 = \dfrac{\int_\Omega I(\mathbf{x}) H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x}}{\int_\Omega H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x}}, & \sigma_1^2 = \dfrac{\int_\Omega (I(\mathbf{x})-\mu_1)^2 H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x}}{\int_\Omega H_\epsilon(\phi(\mathbf{x}))\, d\mathbf{x}}, \\ \mu_2 = \dfrac{\int_\Omega I(\mathbf{x})(1-H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x}}{\int_\Omega (1-H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x}}, & \sigma_2^2 = \dfrac{\int_\Omega (I(\mathbf{x})-\mu_2)^2 (1-H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x}}{\int_\Omega (1-H_\epsilon(\phi(\mathbf{x})))\, d\mathbf{x}}. \end{cases} \tag{1.24}$$

Then, the minimization step is done through the following gradient decent:

$$\frac{\partial\phi}{\partial t} = \delta_\epsilon(\phi)\left(\nu\,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \frac{(I-\mu_1)^2}{2\sigma_1^2} + \frac{(I-\mu_2)^2}{2\sigma_2^2} - \log\frac{\sigma_1^2}{\sigma_2^2}\right) \tag{1.25}$$

Remark 1: Setting the variances to one gives exactly the same evolution equation as the one obtained for the *Active Contours Without Edges* model (1.16). Here, the introduction of the variance permits to capture naturally regions with different variances and noisy images are more easily handled.

Remark 2: In [68], Jehan et al. obtained the same evolution equation but considering a slightly different formulation of the problem. Region statistics are also assumed to be Gaussian and their parameters (means and variances) are set to their

empirical estimation while in the Bayesian formulation, these estimates are obtained from the minimization of the energy (1.23). Then, the authors propose to segment the image by minimizing region descriptors depending on region variances. The entropy of Gaussian densities is one of these descriptors and the following functional has to be minimized:

$$E(\mathcal{C}) = \int_{\mathcal{C}_{in}} \log(\sigma_1^2(\mathcal{C})) + \int_{\mathcal{C}_{out}} \log(\sigma_2^2(\mathcal{C})) + \nu|\mathcal{C}| \tag{1.26}$$

The level-set formulation is not introduced in the energy but an Eulerian minimization approach is employed. Once the gradient descent is estimated for $\mathcal{C}$, the authors introduce the level-set formulation to represent implicitly the evolving contour and the evolution equation (1.25) is obtained.

Kernel density estimation

Rather than considering Gaussian densities, which is quite restrictive on region intensity distributions, a kernel-based estimation of these distributions can be introduced within a variational framework. Actually, a Gaussian model assumes the image to have piecewise smooth intensities but once textured objects are present, this assumption is no longer valid. In [78], Kim et al. proposed to consider the Parzen window estimator for textured gray images. The intensity distribution in a region Ω_i is estimated by:

$$p_i(I) = \frac{1}{|\Omega_i|} \int_{\Omega_i} g_\sigma(I - \hat{I}(\mathbf{x}))\, d\mathbf{x} \tag{1.27}$$

where $g_\sigma(z) = (2\pi\sigma^2)^{-1/2} exp(-z^2/2\sigma^2)$ is a Gaussian kernel with standard deviation σ. The objective function for bi-partitioning is then:

$$\begin{aligned} E(\mathcal{C}) = &- \int_{\mathcal{C}_{in}} \log\left(\frac{1}{|\mathcal{C}_{in}|}\int_{\mathcal{C}_{in}} g_\sigma(I(\mathbf{x}) - \hat{I}(\hat{\mathbf{x}}))\, d\hat{\mathbf{x}}\right) d\mathbf{x} \\ &- \int_{\mathcal{C}_{out}} \log\left(\frac{1}{|\mathcal{C}_{out}|}\int_{\mathcal{C}_{out}} g_\sigma(I(\mathbf{x}) - \hat{I}(\hat{\mathbf{x}}))\, d\hat{\mathbf{x}}\right) d\mathbf{x} + \nu|C| \end{aligned} \tag{1.28}$$

This functional is composed by nested region integrals; the calculus of variations of such types of functionals was studied in [4]. Using these *Shape Derivatives* tools and applying the chain rule, the following gradient descent is obtained for the implicit representation of $\mathcal{C}$:

$$\begin{aligned} \frac{\partial\phi}{\partial t}(\mathbf{x}) = \delta(\phi)\Bigg(&\underbrace{\nu \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right)}_{\text{Regularization}} + \underbrace{\log\frac{p_{\mathcal{C}_{in}}(I(\mathbf{x}))}{p_{\mathcal{C}_{out}}(I(\mathbf{x}))}}_{\text{Region Competition}} \\ &+ \underbrace{\sum_{\mathcal{C}_X=\{\mathcal{C}_{in},\mathcal{C}_{out}\}} \frac{1}{|\mathcal{C}_X|}\int_{\mathcal{C}_X} \frac{g_\sigma(I(\mathbf{x}) - I(\hat{\mathbf{x}}))}{p_{\mathcal{C}_X}(I(\hat{\mathbf{x}}))} d\hat{\mathbf{x}}}_{\text{Additional Term}}\Bigg). \end{aligned} \tag{1.29}$$

If we compare this evolution equation to the static case given in equation (1.22), a new term appears which accounts for region density variations. Even if this density estimation does not make use of a parametric representation of region intensity distributions, this approach is not parameter-free since the kernel size σ has to be chosen. Actually, this parameter is very important to remove noise and to capture density modes, but a heuristic choice is not possible in general. However, a multi-scale strategy can be employed, similarly to simulated annealing, by minimizing the functional (1.28) successively for decreasing standard deviation σ, initializing the contour from its last final position. This strategy permits to obtain the good segmentation for the ideal σ and the contour should keep this position for lower σ, thanks to the regularization constraint on the contour. A remaining issue is the assumption of similar noise properties for each region.

1.3 Conclusion

We have presented a short summary of contour evolution in image segmentation. Two different types of contour evolution have been considered: boundary-based and region-based. Boundary-based methods were the natural formulation of early evolution techniques but strong limitations were present, in particular concerning the initial contour position. On the other hand, region-based methods are more global by minimizing a functional defined on the whole image domain. The contour evolution is guided by region information and its initial position is no longer so important. While most region-based techniques only incorporate a length regularization as boundary term, the geodesic active contour model can be integrated within a region-based formulation [107], taking benefits from both methods. Two classes of region-based techniques have been presented. The first one is purely based on geometric properties while the second one incorporates statistical models of each region. The geometric view of the problem is elegant and may perform well. However, in most cases, images are subject to noise or unknown perturbing phenomena and the introduction of statistical models is necessary. Motivated by a Bayesian formulation, the second technique models region statistics and so, it can handle a wider range of natural and real images.

In the next chapter, we revisit the region-based method presented in [107]. Scalar and vector-valued images are considered with parametric and non-parametric approximations of pixel values distribution over each region. The main contribution is the estimation of region statistics during the partitioning process itself.

Chapter 2

Bayesian Region-Based Front Evolution

In this chapter, a region-based front evolution is obtained from a Bayesian formulation of the image partitioning problem. This approach has been favored in this thesis for several reasons. The choice of a region-based approach seems to be the best one to define robust algorithms which can deal with a large set of images and with arbitrary initializations. Besides, since the main objective of this thesis is the incorporation of different cues within front evolution, the chosen framework need to incorporate a decision process to dynamically select relevant cues and the Bayesian formulation is well adapted for this task.

Our objective functionals are issued from the one presented in Section 1.2.3. Several works [81, 171, 107] have proposed to use very similar optimization criteria from different formulations. In [81, 171], this energy was obtained from a Minimum Descriptor Length [50] criteria while in [107], it was proven to be equivalent to an approximation of the maximum *a posteriori* (MAP) frame partition by making several explicit assumptions/simplifications.

In the next part, we recall this MAP formulation and we discuss the validity of the assumptions and the possibilities of removing them. Then, we revisit this segmentation criteria with the level-set formulation and we introduce region statistics as unknown parameters. Applications to scalar and vector-valued images are considered for parametric and nonparametric distributions. This framework is validated on numerous synthetic and real images. In the last section, we present several similarities with classical clustering techniques and we discuss the advantages of geometric methods when dealing with images.

2.1 Maximum a posteriori frame partition

Following [107], we consider $p_{\mathcal{S}}(\mathcal{P}(\Omega)|I)$, the *a posteriori* frame partition probability, given the observed image I. The optimal partition of the image is obtained by maximizing this probability according to the associated hypothesis. The Bayes rule permits one to express this probability as:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) = \frac{p(I|\mathcal{P}(\Omega))}{p(I)} p(\mathcal{P}(\Omega)), \tag{2.1}$$

where $p(\mathcal{P}(\Omega))$ and $p(I)$ are respectively the probability of a partition $\mathcal{P}(\Omega)$ and the probability of an image I. The third term, $p(I|\mathcal{P}(\Omega))$ represents the *a posteriori* segmentation probability of the image I, given the partition $\mathcal{P}(\Omega)$. When segmenting a given image, the term $p(I)$ is constant and (2.1) simplifies to:

$$p_{\mathcal{S}}(\mathcal{P}(\Omega)|I) \propto p(I|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)).$$

2.2 Partition constraint

The term $p(\mathcal{P}(\Omega))$ allows for the integration of *a priori* knowledge about the image partition. When, in a particular application, shapes keep similar properties from one image to another, learning prior shape distributions from training samples may be of great interest to design this term. Chapter 4 is dedicated to the modeling and introduction of such information but, right now, we consider a more general constraint, using this term to express geometric regularity on the border between image regions. Let $\mathcal{C}$ be this interface, a simple regularization constraint can be imposed by favoring image partitions with the shortest interface:

$$p(\mathcal{P}(\Omega)) \propto e^{-\nu |\mathcal{C}|}$$

where $|\mathcal{C}|$ is the length of the interface and ν, a weighting constant controlling the regularization.

2.3 Region-based term

Now, we concentrate on the term $p(I|\mathcal{P}(\Omega))$, the *a posteriori* segmentation probability for the image I, given the partition $\mathcal{P}(\Omega)$. This probability cannot be recovered in the general case and several simplifications have to be done. First, we assume the image domain Ω to be composed of N regions: $\mathcal{P}(\Omega) = \{\Omega_1, \ldots, \Omega_N\}$ with no correlation between region labeling. Then, we can write:

$$p(I|\mathcal{P}(\Omega)) = p(I|\{\Omega_1, \ldots, \Omega_N\}) = p(I|\Omega_1)p(I|\Omega_2)\ldots p(I|\Omega_N)$$

where $p(I|\Omega_X)$ denotes the probability of having the image I when Ω_X corresponds to a region of interest.

Another important approximation is generally made by assuming pixels inside a region to be independent and identically distributed realizations of a same random process. This permits us to replace region distributions by the joint probability of the pixels belonging to the region:

$$p(I|\mathcal{P}(\Omega)) = \prod_{\mathbf{x}\in\Omega_1} p(I(\mathbf{x})|\Omega_1) \prod_{\mathbf{x}\in\Omega_2} p(I(\mathbf{x})|\Omega_2) \cdots \prod_{\mathbf{x}\in\Omega_N} p(I(\mathbf{x})|\Omega_N)$$

This approximation is generally not valid since image intensities are often characterized by local spatial relations as modeled in the Markov Random Field theory. In the case of textured images, these spatial relations are the main characterization of the regions of interest. Therefore, this model is only valid for piecewise smooth images. Nevertheless, as we will see in the chapter 3, spatial filters can be applied to the image to extract pertinent features that can be used for the segmentation.

2.4 Energy formulation

The maximization of the *a posteriori* segmentation probability is equivalent to the minimization of the obtained energy after applying the negative logarithm:

$$E(\{\Omega 1, \ldots, \Omega_N\}) = -\sum_i \int_{\Omega_i} \log p(I(\mathbf{x})|\Omega_i)\, d\mathbf{x} + \nu\,|\mathcal{C}| \tag{2.2}$$

This energy was presented in Section 1.2.3 and it is the basis of several works [81, 171, 129, 107]. The MAP formulation permits all the approximations that are needed to obtain this criteria to be made explicit. Then, it becomes easier to understand which conditions should be verified by an image so that it can be segmented by this approach. Let us recall the principal simplifications and their consequences on the possible set of images:

1. The partition constraint is approximated by a regularization on the interface: this has no limitation on the type of image, but in a given application, the distribution of possible partitions can be learnt *a priori* and it may be useful to constraint the optimization process. This issue will be studied in Chapter 4 of this thesis.

2. The image domain Ω is composed of N regions: the number of regions need to be known *a priori*, limiting the unsupervised aspect of the approach. However, heuristics can be used to decide whether two regions can be merged or not like in [20].

3. Pixels within a region are independent and identically distributed realizations of a random process: this is actually the strongest assumption of the approach. It restricts the set of possible images to piecewise smooth images. A two-step approach will be considered in Chapter 3 to deal with a wider range of images.

As explained in Section 1.2.3, a regularization term on the interface is necessary to include spatial constraints. While a simple length constraint as the one considered so far is generally employed, more complex boundary-based terms can be used to integrate image features on the interface. In [107], the *geodesic active contour* functional stands as the boundary term.

Comparing the works presented in [81, 171, 129, 107], several important differences arise. In [171], a region competition approach is employed to segment gray, color and textured images. Region densities are approximated by Gaussian distributions whose parameters are estimated during the optimization process. In [129], the use of the energy is limited to the task of classification while region densities are learnt *a priori*. Finally, in [107], region densities are still known in advance but the region-based classification is fused with the *geodesic active contours*, driving to a new partitioning energy referred as *geodesic active regions*.

2.5 Adaptive region-based segmentation

In this section, we consider the energy (2.2) with different types of region probability densities. The estimation of these densities is done online, i.e. during the optimization process itself. This can be achieved by optimizing the objective functional with respect to region statistics in parallel to the partitioning. Parametric as well as nonparametric densities are considered. As far as the parametric case is concerned, we consider single Gaussians for scalar images and multivariate Gaussians for the more general case of vector-valued images. The Parzen window estimate [112] is used for the nonparametric case.

In the case of unsupervised methods, pixel value distributions inside each region are not known *a priori* and they must be estimated during the partitioning process. Two different approaches are presented; in the parametric case, region densities are estimated explicitly while in the nonparametric formulation, region densities are introduced as functions of the partition.

2.5.1 Parametric region densities

The definition of a general criteria without making any assumption on region intensity distributions is not possible. However, a straightforward approach is to assume each intensity distribution to belong to the same parametric family but with different parameters. Then, the segmentation criteria can be optimized with respect to two set of parameters: the image partition and the statistical parameters. The energy (2.2) can be used to define such an optimization criteria by simply introducing the statistical parameters as unknown. If $p(I|\theta_i)$ is a parametric representation of the *a posteriori* intensity distribution in the region Ω_i parameterized by θ_i, the segmentation of the

image is obtained from the minimization of:

$$E(\{\Omega 1,\ldots,\Omega_N\},\{\theta 1,\ldots,\theta_N\}) = -\sum_i \int_{\Omega_i} \log p(I(\mathbf{x})|\theta_i)\, d\mathbf{x} + \nu\,|\mathcal{C}|. \qquad (2.3)$$

This energy has clearly two different types of parameters: the image sub-domains Ω_i, open and bounded subsets of Ω and the statistical parameters θ_i which belongs to a parameter space Θ. As explained in the previous chapter, a combination of level-set functions permits to reformulate this energy. The optimization with respect to image sub-domains is replaced by a new minimization with unknown Lipschitz functions defined on the image domain. For the sake of simplicity, we consider the case of bi-partitioning ($N = 2$) [1]. Let ϕ be the signed distance function to the border between Ω_1 and Ω_2, the energy (2.3) becomes:

$$\begin{aligned} E(\phi,\{\theta_1,\theta_2\}) = &-\int_\Omega \Big(H(\phi)\log p(I(\mathbf{x})|\theta_1) + (1-H(\phi))\log p(I(\mathbf{x})|\theta_2) \Big)\, d\mathbf{x} \\ &+ \nu \int_\Omega |\nabla H(\phi(\mathbf{x}))|\, d\mathbf{x}. \end{aligned} \qquad (2.4)$$

For θ_1 and θ_2 fixed, this functional can be minimized with respect to the function ϕ through a gradient descent. The Euler-Lagrange equation for ϕ gives the following evolution equation:

$$\frac{\partial \phi}{\partial t}(\mathbf{x}) = \delta(\phi)\left(\nu\,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \log\frac{p(I(\mathbf{x})|\theta_1)}{p(I(\mathbf{x})|\theta_2)}\right) \qquad (2.5)$$

while the statistical parameters are regularly updated according to:

$$\theta_i = \underset{\theta\in\Theta}{\mathrm{Arg\,min}} \int_{\Omega_i} \log p(I(\mathbf{x})|\theta)\, d\mathbf{x} \quad \text{for} \quad i = 1,2 \qquad (2.6)$$

As will be the case for Gaussian approximations, closed form solutions may exist for optimal values of the statistical parameters and therefore, they can be continuously updated during the level-set evolution.

Single Gaussian approximations for scalar images

A Gaussian approximation of region densities is well-adapted to smoothed gray-valued images since it models each region by a constant intensity with possible noise. Let us first recall the well-known Normal or Gaussian law:

$$p(z|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \qquad (2.7)$$

[1] Several works propose level-set formulations for more than two regions [167, 107, 155, 20] but the increased complexity makes them more sensitive to the initial conditions. Since the purpose of our approach lies in the estimation of region statistics jointly with the image partionning, we stick to the simplest case of bi-partitioning. However, the binary case is still pertinent for most object extractions where a single object of interest has to be separated from the rest of the image. For the purpose of demonstration, we show some promising results obtained with our approach and the coupling method developed in [20] in [Figure 2.9].

According to the last paragraph, the objective energy can be minimized by iteratively estimating the optimal statistical parameters (μ_i, σ_i^2) for a fixed level-set and evolving the level-set with these parameters. Here, the optimal mean and variances of the Gaussian law are simply their empirical values in the corresponding region. Then, the bi-partionning energy can be minimized using the following system of coupled equations:

$$\begin{cases} \mu_i = \dfrac{1}{|\Omega_i|}\displaystyle\int_{\Omega_i} I(\mathbf{x})\,d\mathbf{x}, \quad \sigma_i^2 = \dfrac{1}{|\Omega_i|}\displaystyle\int_{\Omega_i} (I(\mathbf{x})-\mu_i)^2 d\mathbf{x} \quad \text{for} \quad i=1,2 \\ \dfrac{\partial\phi}{\partial t}(\mathbf{x}) = \delta(\phi)\left(\nu\,\text{div}\left(\dfrac{\nabla\phi}{|\nabla\phi|}\right) - \dfrac{(I(\mathbf{x})-\mu_1)^2}{2\sigma_1^2} + \dfrac{(I(\mathbf{x})-\mu_2)^2}{2\sigma_2^2} - \log\dfrac{\sigma_1^2}{\sigma_2^2}\right) \end{cases} \tag{2.8}$$

Let us note that this system simplifies to the *Active Contours Without Edges* model introduced by Chan and Vese in [30] if the variances of both regions are set to a same constant (i.e. if the region model does not incorporate variance).

At each iteration, the statistical parameters have to be updated and if they had to be estimated directly, the whole image domain would have to be considered after each iteration of the curve evolution, leading to a very high complexity for the complete algorithm. However, this can be done in a more "clever" way by watching the pixels that are changing domain at each iteration of the level-set evolution. Then, the new statistical parameters can be expressed with respect to their previous update. Let n^+ and n^- be the number of pixels going respectively from Ω_1 to Ω_2 and inversely. The associated domains are named Ω_+ and Ω_-. We assume the statistical parameters at time t, $\mu_i^{(t)}$ and $\sigma_i^{(t)}$, to be known. First, we can express the domains at time $t+1$: $\Omega_1^{(t+1)} = \Omega_1^{(t)} + \Omega_+ - \Omega_-$ and $\Omega_2^{(t+1)} = \Omega_2^{(t)} - \Omega_+ + \Omega_-$ and consequently: $|\Omega_1^{(t+1)}| = |\Omega_1^{(t)}| + n^+ - n^-$ and $|\Omega_2^{(t+1)}| = |\Omega_2^{(t)}| - n^+ + n^-$. Then, the new values of the means are:

$$\begin{cases} \mu_1^{(t+1)} = \dfrac{1}{|\Omega_1^{(t+1)}|}\displaystyle\int_{\Omega_1^{(t+1)}} I(\mathbf{x})\,d\mathbf{x} \\ \qquad\quad = \dfrac{1}{|\Omega_1^{(t+1)}|}\left(|\Omega_1^{(t)}|\mu_1^{(t)} + \displaystyle\int_{\Omega_+} I(\mathbf{x})\,d\mathbf{x} - \int_{\Omega_-} I(\mathbf{x})\,d\mathbf{x}\right) \\ \mu_2^{(t+1)} = \dfrac{1}{|\Omega_2^{(t+1)}|}\displaystyle\int_{\Omega_2^{(t+1)}} I(\mathbf{x})\,d\mathbf{x} \\ \qquad\quad = \dfrac{1}{|\Omega_2^{(t+1)}|}\left(|\Omega_2^{(t)}|\mu_2^{(t)} - \displaystyle\int_{\Omega_+} I(\mathbf{x})\,d\mathbf{x} + \int_{\Omega_-} I(\mathbf{x})\,d\mathbf{x}\right) \end{cases} \tag{2.9}$$

The same trick can be used to update region variances. For this purpose, we recall the relation between the first and second order moments of a random variable X:

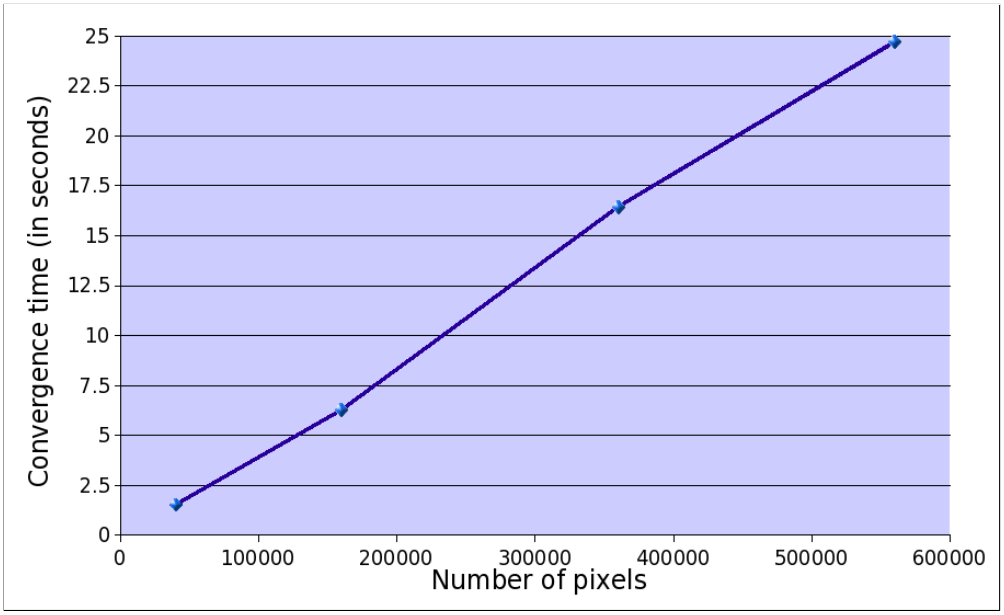


Figure 2.1: Convergence time for the entire process on different scales of the same image.

$V(X) = E(X^2) - E(X)^2$. The following update rules follow:

$$\begin{cases} \sigma_1^{(t+1)^2} = \dfrac{1}{|\Omega_1^{(t+1)}|} \displaystyle\int_{\Omega_1^{(t+1)}} I^2(\mathbf{x})\, d\mathbf{x} - \mu_1^{(t+1)^2} \\ \sigma_2^{(t+1)^2} = \dfrac{1}{|\Omega_2^{(t+1)}|} \displaystyle\int_{\Omega_2^{(t+1)}} I^2(\mathbf{x})\, d\mathbf{x} - \mu_2^{(t+1)^2} \end{cases} \tag{2.10}$$

Of course the same decomposition in $\Omega_i^{(t)}$, Ω_+ and Ω_- can be used to estimate the first terms. To summarize, we need to estimate the expectation of the intensity and the squared intensity in each domain at the initialization. Then, only integrals over the "changing" domains Ω_+ and Ω_- are necessary. The complexity of each iteration of the front evolution is thus decreased to the length of the contour multiplied by the number of pixels changing regions. Assuming the evolving contour to visit each pixel only once, the total complexity is linear in the image size as shown in [Figure 2.1][2].

In [Figure 2.2], we show the curve evolution as well as a plot of the corresponding energy for a simple synthetic image composed of two regions with different mean gray values and with additional Gaussian noise. An initialization with small circles is considered to detect holes more easily. The evolution of the statistical parameters are presented in [Figure 2.3]. We also show an example on a real hand image in [Figure 2.4]. Both examples show the benefits of using the level-set representation as the

[2]The convergence times of this experiment, as well as the other ones for other results were obtained on a computer with a 1 Gz CPU and 1 Gb of RAM.

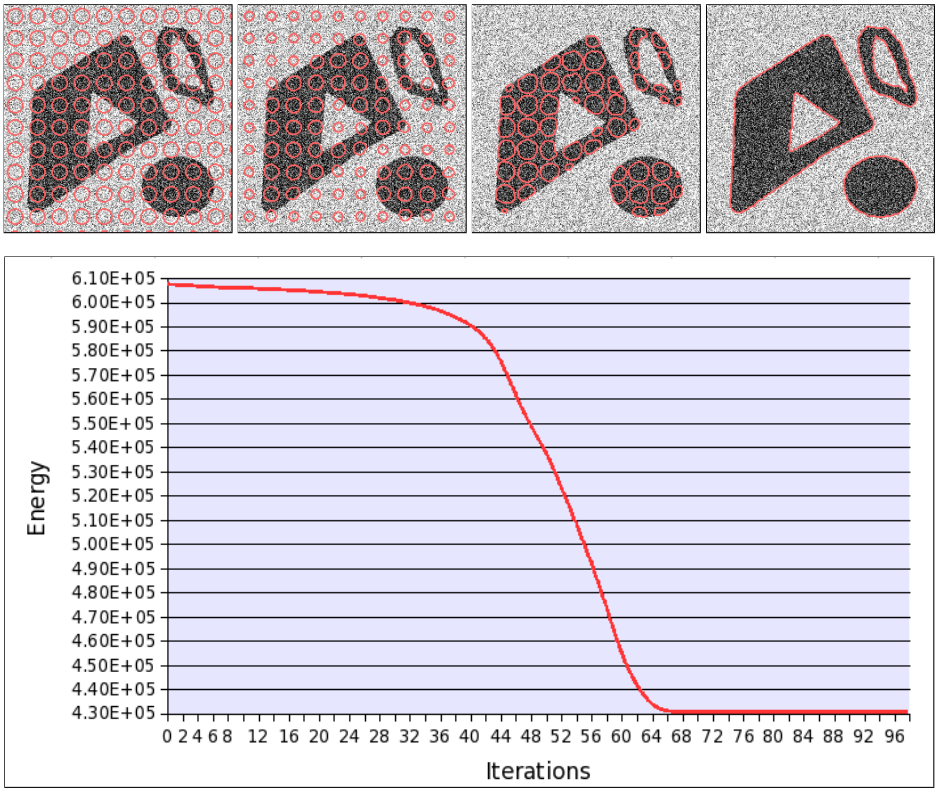


Figure 2.2: Synthetic example composed of two regions with different mean gray values (256x256)- TOP: Curve evolution ($\nu = 5$, convergence time 12 seconds), BOTTOM: Corresponding energy evolution.

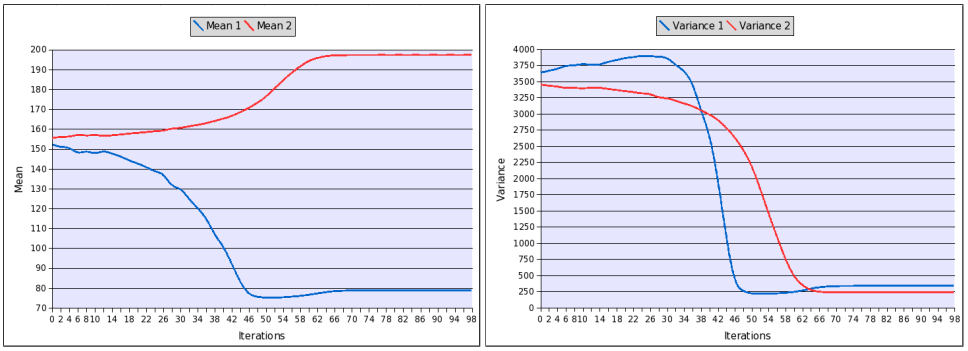


Figure 2.3: Synthetic example of [Figure 2.2] - LEFT: Mean evolutions, RIGHT: Variance evolutions.

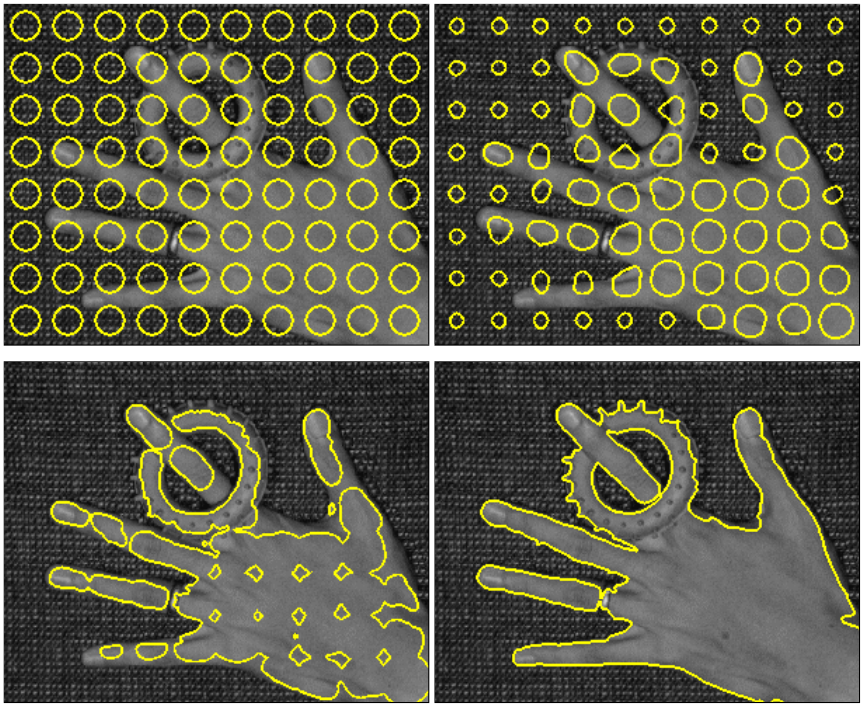


Figure 2.4: Real gray-valued hand image (303x243) - Curve evolution ($\nu = 3$, convergence time: 11 seconds).

contour changes topology many times before convergence. Other initializations may be considered but the use of tiny circles is quite efficient in terms of convergence speed and for the detection of small parts and holes.

Multivariate Gaussian approximations for vector-valued images

A direct extension to vector-valued images and more particularly to color images is to use a multivariate Gaussian density as approximation of region statistics:

$$p(z|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|} e^{-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)}. \tag{2.11}$$

The derivation of the corresponding energy is quite similar to the scalar case. For a fixed contour, the optimal statistical parameters are also their empirical estimations in the corresponding regions. Therefore, for the bi-partitioning case, the image partition is obtained by updating the level-set function according to the following system of

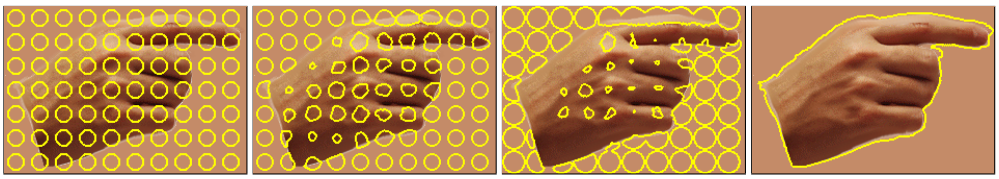


Figure 2.5: Synthetic hand image with constant background (234x161) - Curve evolution ($\nu = 5$, convergence time: 6 seconds).

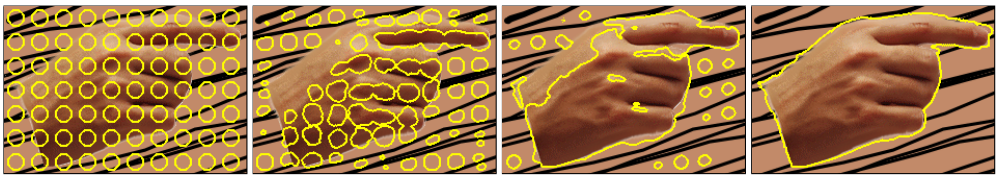


Figure 2.6: Synthetic hand image with complex background (234x161) - Curve evolution ($\nu = 5$, convergence time: 10 seconds).

coupled equations:

$$\begin{cases} \mu_i = \dfrac{1}{|\Omega_i|}\displaystyle\int_{\Omega_i} I(\mathbf{x})\,d\mathbf{x}, \\ \Sigma_i = \dfrac{1}{|\Omega_i|}\displaystyle\int_{\Omega_i} (I(\mathbf{x})-\mu_i)^T(I(\mathbf{x})-\mu_i)\,d\mathbf{x} \quad \text{for} \quad i=1,2 \\ \dfrac{\partial\phi}{\partial t}(\mathbf{x}) = \delta(\phi)\left(\nu\,\text{div}\left(\dfrac{\nabla\phi}{|\nabla\phi|}\right)\right. \\ \qquad \left. -(I(\mathbf{x})-\mu_1)^T\Sigma_1^{-1}(I(\mathbf{x})-\mu_1)+(I(\mathbf{x})-\mu_2)^T\Sigma_2^{-1}(I(\mathbf{x})-\mu_2)-\log\dfrac{|\Sigma_1|}{|\Sigma_2|}\right). \end{cases} \tag{2.12}$$

Similarly to the scalar case, the estimation of the statistical parameters can be optimized to avoid full computation over the whole image domain at each iteration. Here, it becomes a bit more technical since cross-components products appear in the covariance matrices but the final complexity is identical to the one obtained in the scalar case (modulo the number of parameters).

Several segmentation examples are presented in [Figures 2.5, 2.6, 2.7 and 2.8]. First we built two synthetic images composed of a real hand with different backgrounds. In the first one shown in [Figure 2.5], the background color was chosen similar to the mean color of the hand and a successful segmentation is obtained due to the estimation of the covariance matrix inside each region. A more complex background with black stripes is considered in the second example [Figure 2.6]. Here, the regularization term plays an important role to make our algorithm able to extract the hand.

Experiments on real images are more challenging. When considering color im-

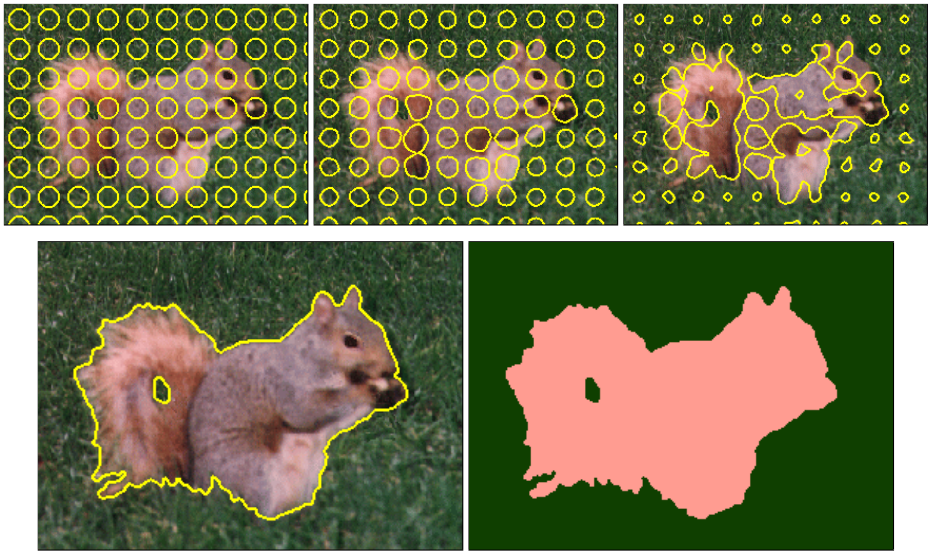


Figure 2.7: Segmentation of a *Squirrel* color image (288x209) - Curve evolution ($\nu = 5$, convergence time: 10 seconds) and final segmentation.

ages, the choice of a pertinent color space is important. If we want our region statistics to make sense, distances between pixel values should reflect perceptual difference between the associated colors. Hopefully, the CIE-Lab color space was designed to fulfill this requirement. Several promising results are shown in [Figure 2.7 and 2.8]. The *squirrel* image is rather simple to segment since the colors are quite different but the grass is still noisy and a strong regularity has to be imposed to obtain the smooth result presented in [Figure 2.7]. The other example presented in [Figure 2.8] is more difficult since the frog color is really close to the color of the background and some texture is present in the image.

As mentioned before, several level-sets can be coupled to segment the image in an arbitrary number of regions. The extension of our framework with such coupling methods like the ones presented in [107, 155, 20] is straightforward and we show some nice results in [Figure 2.9] which were obtained using the algorithm proposed in [20], jointly with our region-based evolution.

2.5.2 Nonparametric region densities

The choice of a parametric family to approximate intensity distributions may be seen as a limitation to model region statistics. Contrary to parametric models, nonparametric density estimations can approximate any type of distribution for a sufficiently large data set. Within nonparametric methods, *Kernel* or equivalently *Parzen density estimates* are the most widely used. Going back to our segmentation problem, the Parzen

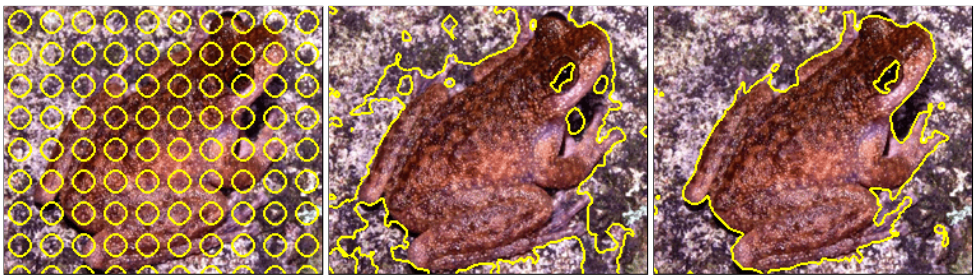


Figure 2.8: Segmentation of a *Frog* color image (250x210) - Curve evolution ($\nu = 3$, convergence time: 16 seconds).

density estimate of the intensity in a region Ω_X can be written as follows:

$$p_X(z) = \frac{1}{|\Omega_X|} \int_{\Omega_X} \frac{1}{h} K\left(\frac{z - I(\mathbf{x})}{h}\right) d\mathbf{x}, \tag{2.13}$$

where K is a *kernel* function and h is the scale parameter or *bandwidth*. In general, density estimates are expressed directly from the histogram. If h_X is the histogram in region Ω_X:

$$p_X(z) = \frac{1}{|\Omega_X|} \sum_k h_X(k) \frac{1}{h} K\left(\frac{z - k}{h}\right). \tag{2.14}$$

Actually, we can see that both formulations are equivalent from the definition of the histogram:

$$h_X(k) = \int_{\Omega_X} \delta(k, I(\mathbf{x}))\, d\mathbf{x}, \tag{2.15}$$

where $\delta(i, j)$ is the usual kronecker function, equal to 1 if $i = j$ and 0 otherwise.

Several types of Kernels have been proposed. The most commons are spherical (uniform), Gaussian or Epanechnikov kernels [51]. This last one is shown to be asymptotically-optimal by minimizing the Mean Squared Error for an infinite number of samples. However, the choice of the scale parameter h is much more important than the type of kernel and in general, the easiest kernel to implement is preferred.

Then, using general *Kernel density estimates* to approximate region statistics, our objective function becomes:

$$E(\{\Omega_1, \ldots, \Omega_N\}) = -\sum_i \int_{\Omega_i} \log\left(\frac{1}{|\Omega_i|} \int_{\Omega_i} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}}\right) d\mathbf{x} + \nu|\mathcal{C}|. \tag{2.16}$$

Contrary to the parametric case, this functional depends only on the sub-domains Ω_i but nested region integrals appear. This type of functional can be minimized using the *Shape Gradient* formalism introduced in [4]. In particular, we will use the following theorem:

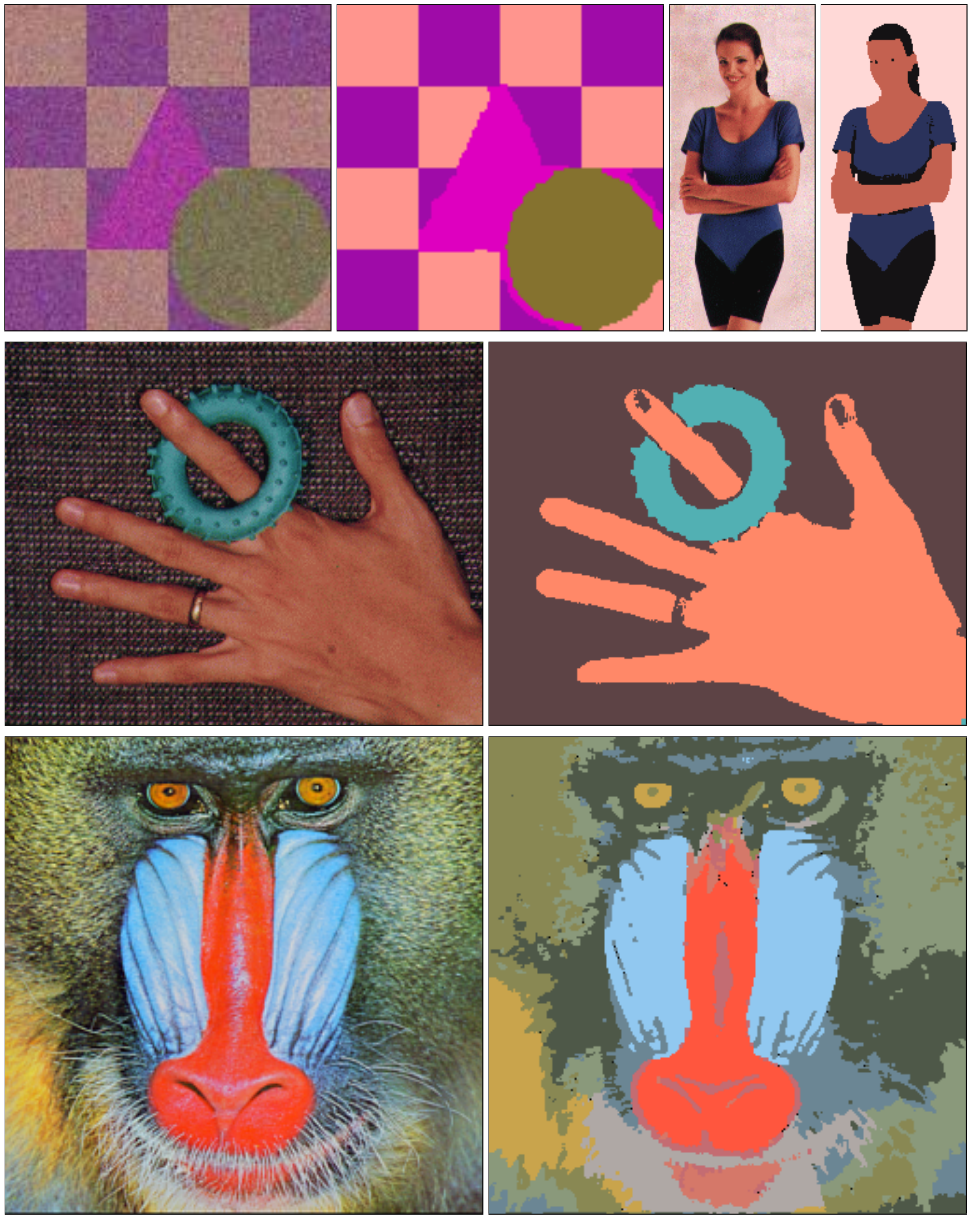


Figure 2.9: Unsupervised segmentation of color images in an arbitrary number of regions using the coupling method introduced in [20], jointly with our region-based evolution.

Theorem 2.1 *The Gâteaux derivative of the functional* $J(\Omega) = \int_\Omega g(\mathbf{x}, \Omega)\, d\mathbf{x}$ *in the direction of a vector field* $\mathbf{V}$ *is the following:*

$$< J'(\Omega), \mathbf{V} >= \int_\Omega f_s(\mathbf{x}, \Omega, \mathbf{V})\, d\mathbf{x} - \int_{\partial\Omega} f(x, \Omega)(\mathbf{V}(\mathbf{x}) \cdot \mathbf{N}(\mathbf{x}))\, d\mathbf{a}(\mathbf{x})$$

where $\mathbf{N}$ *is the unit normal to* $\partial\Omega$ *and* $d\mathbf{a}$ *its area element and* $f_s(\mathbf{x}, \Omega, \mathbf{V})$ *is the shape derivative of* $f(\mathbf{x}, \Omega)$.

Applied on our criteria, we obtain the

Proposition 2.2 *The Gâteaux derivative of a region-based term using a kernel density estimation with constant* bandwidth

$$F(\Omega_X) = \int_{\Omega_X} \log\left(\frac{1}{|\Omega_X|}\int_{\Omega_X} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}}\right) d\mathbf{x} \tag{2.17}$$

in the direction of a vector field $\mathbf{V}$ *is the following:*

$$\begin{aligned} < F'(\Omega_X), V >= -\int_{\partial\Omega_X} \Bigg(& \log p_X(I(\mathbf{x})) + \\ & \frac{1}{|\Omega_X|}\left(\int_{\Omega_X} \frac{\frac{1}{h}K\left(\frac{I(\mathbf{x})-I(\hat{\mathbf{x}})}{h}\right)}{p_X(I(\hat{\mathbf{x}}))}\, d\hat{\mathbf{x}} - 1\right)\Bigg)(V(\mathbf{x}) \cdot N(\mathbf{x}))d\mathbf{a}(\mathbf{x}) \end{aligned} \tag{2.18}$$

Proof : Let us consider the region-term (2.17). We introduce the following notations:

$$\begin{aligned} G(x, \Omega_X) &= \int_{\Omega_X} \frac{1}{h} K\left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h}\right) d\hat{\mathbf{x}} \\ V(\Omega_X) &= |\Omega_X| = \int_{\Omega_X} d\hat{\mathbf{x}} \end{aligned} \tag{2.19}$$

The functional (2.17) can then be rewritten:

$$F(\Omega_X) = \int_\Omega f(\mathbf{x}, \Omega_X)\, d\mathbf{x} = \int_{\Omega_X} \log \frac{G(\mathbf{x}, \Omega_X)}{V(\Omega_X)}\, d\mathbf{x} \tag{2.20}$$

Let $\partial\Omega_X$ be the boundary of Ω_X. According to Theorem 2.1, the Gâteaux derivative of this functional in the direction of a vector field V is given by:

$$< F'(\Omega_X), V >= \int_{\Omega_X} f_s(\mathbf{x}, \Omega_X, V)\, d\mathbf{x} - \int_{\partial\Omega_X} \log p_X(I(\mathbf{x}))(V(\mathbf{x}) \cdot N(\mathbf{x}))d\mathbf{a}(\mathbf{x}) \tag{2.21}$$

Using the chain rule on $G(\mathbf{x}, \Omega_X)$ and $V(\Omega_X)$, the shape derivative of $f(\mathbf{x}, \Omega_X)$ can

be expressed as follow:

$$\begin{aligned} f_s(\mathbf{x}, \Omega_X, V) &= f_G < G'(\mathbf{x}, \Omega), V > + f_V < V'(\Omega), V > \\ &= \frac{1}{G(\mathbf{x}, \Omega_X)} \left(- \int_{\partial\Omega_X} \frac{1}{h} K \left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h} \right) (V(\hat{\mathbf{x}}) \cdot N(\hat{\mathbf{x}}))\, d\mathbf{a}(\hat{\mathbf{x}}) \right) \\ &\quad - \frac{1}{V(\Omega_X)} \left(- \int_{\partial\Omega_X} (V(\hat{\mathbf{x}}) \cdot N(\hat{\mathbf{x}}))\, d\mathbf{a}(\hat{\mathbf{x}}) \right) \\ &= - \int_{\partial\Omega_X} \left(\frac{\frac{1}{h} K \left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h} \right)}{|\Omega_X| p_X(I(\mathbf{x}))} - \frac{1}{|\Omega_X|} \right) (V(\hat{\mathbf{x}}) \cdot N(\hat{\mathbf{x}}))\, d\mathbf{a}(\hat{\mathbf{x}}) \end{aligned} \tag{2.22}$$

Changing the order of the integrals and adding the second term, we obtain the final expression of the Gâteaux derivative of our functional (2.18). □

The first term of this derivative is similar to the one obtained for the Parametric evolution (2.5) but here a new term appears, accounting for histogram variations. Let us look closer to this new term:

$$q_X(z) = \frac{1}{|\Omega_X|} \int_{\Omega_X} \frac{\frac{1}{h} K \left(\frac{z - I(\mathbf{x})}{h} \right)}{p_X(I(\mathbf{x}))}\, d\mathbf{x}. \tag{2.23}$$

This term is quite similar to the density estimation but the convolution is weighted by $\frac{1}{p_X(I(\mathbf{x}))}$. Therefore, two convolutions are needed, the first one to estimate the densities and the second for this term.

If we consider the entire functional for the bi-partitioning case, the corresponding level-set evolution can be expressed as:

$$\begin{aligned} \frac{\partial \phi}{\partial t}(\mathbf{x}) = \delta(\phi) \Bigg(\nu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \log \frac{p_1(I(\mathbf{x}))}{p_2(I(\mathbf{x}))} \\ - \frac{1}{|\Omega_1|} + \frac{1}{|\Omega_2|} + q_1(I(\mathbf{x})) - q_2(I(\mathbf{x})) \Bigg). \end{aligned} \tag{2.24}$$

Obviously, the two terms in $\frac{1}{\Omega_X}$ can be neglected because of the size of the image. For a Gaussian kernel, the terms q_1 and q_2 can be also generally neglected. Then, the simplified evolution equation becomes very similar to the parametric one:

$$\begin{cases} p_i(I(\mathbf{x})) = \dfrac{1}{|\Omega_i|} \displaystyle\int_{\Omega_i} \frac{1}{h} K \left(\frac{I(\mathbf{x}) - I(\hat{\mathbf{x}})}{h} \right) d\hat{\mathbf{x}}, \quad i = 1, 2 \\ \dfrac{\partial \phi}{\partial t}(\mathbf{x}) \approx \delta(\phi) \left(\nu \operatorname{div} \left(\dfrac{\nabla \phi}{|\nabla \phi|} \right) + \log \dfrac{p_1(I(\mathbf{x}))}{p_2(I(\mathbf{x}))} \right). \end{cases} \tag{2.25}$$

1D case: scalar images

In the case of scalar or gray-valued images, density estimations can be done directly on the histogram as its dimension is generally much lower than the dimension of the

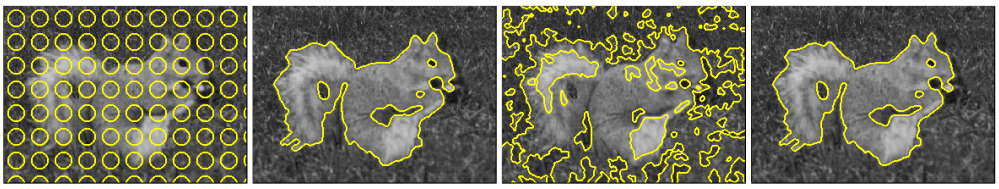


Figure 2.10: From left to right: contour initialization, result with Gaussian approximations, results using Parzen density estimates with $\sigma = 1$ and $\sigma = 10$ respectively.

image. On this type of data, intensity densities can be estimated very efficiently using a Gaussian kernel:

$$g_\sigma(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}. \tag{2.26}$$

A recursive implementation permits to convolve the histogram with the Gaussian kernel in a linear complexity [46]. The most important element of kernel density estimation is the choice of the bandwidth σ. It should be chosen high enough to remove noise but too large values would make details of the histogram disappear. Each different choice of σ will emphasize different aspects of the data and several works [134, 131] suggest to consider several density estimations with different bandwidths. However, we limit our study to a single choice of σ.

In [Figure 2.5.2], we compare results obtained on a gray version of the *squirrel* image, using either a Gaussian approximation or Parzen density estimates with different bandwidths. The density estimates, obtained at convergence for $\sigma = 10$, are shown in [Figure 2.5.2] . While a too small bandwidth is unable to capture region properties and gives a result highly dependent on the initialization, a pertinent choice of bandwidth gives a segmentation similar to the one obtained with a Gaussian approximation. The dependence on the initialization may be useful in applications where a good guess of the object of interest is possible. The nonparametric approximation can capture the distribution of a complex background like in medical images [Figure 2.5.2] or of a complex object like in [Figure 2.5.2].

Several similar nonparametric approaches can be found in [78, 69, 66] where slightly different criteria based on the information theory are considered.

General case: vector-valued images

The extension of kernel density estimation to vector-valued images straightforward, a similar Gaussian kernel is considered. However, the increase of complexity may be seen as a limiting factor. A simple solution is to quantify the image to a reduced number of color levels.

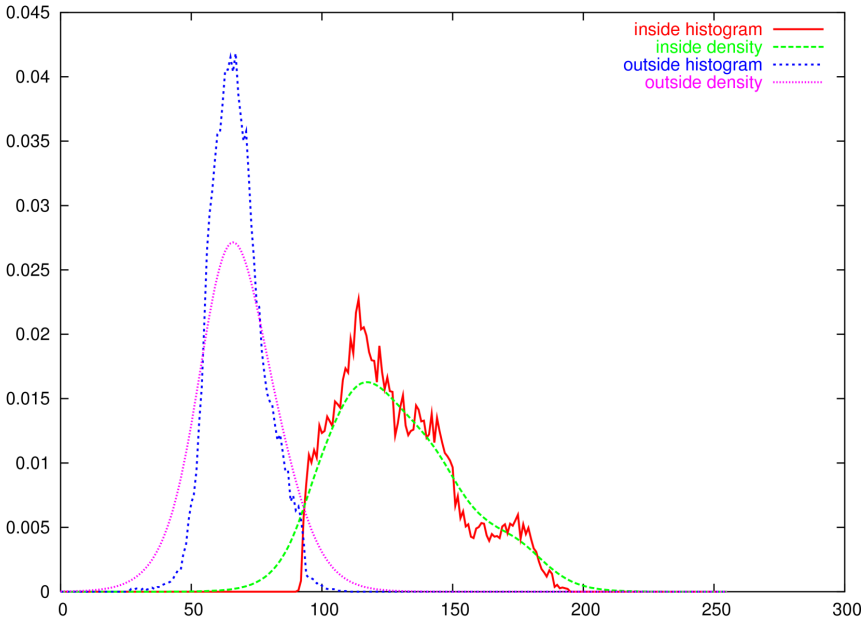


Figure 2.11: Histograms (after normalization) and density estimations of the inside and outside regions in the gray-valued squirrel image at convergence ($\sigma = 10$).

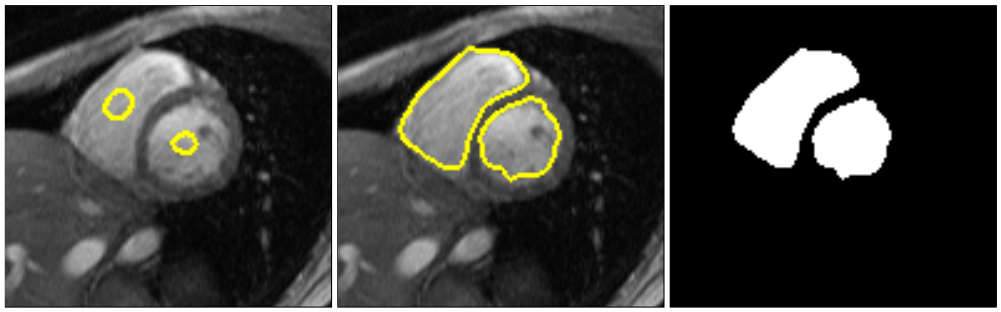


Figure 2.12: Cardiac image - Contour initialization and final segmentation using Parzen density estimates ($\sigma = 5$).

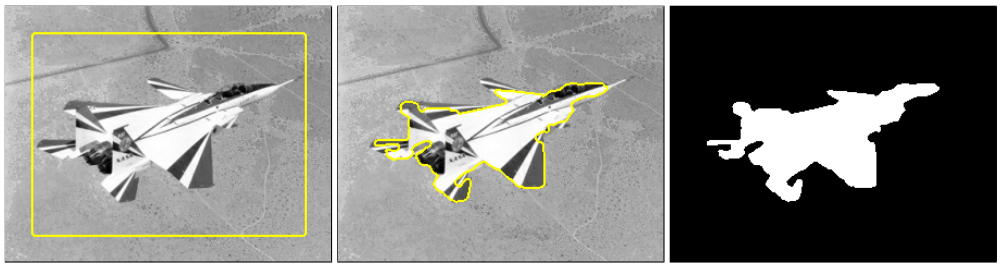


Figure 2.13: F15 image - Contour initialization and final segmentation using Parzen density estimates ($\sigma = 20$).

2.6 Implementation

2.6.1 The general case

The general algorithm is similar for both, parametric and nonparametric, models. Two steps alternate until convergence: (i) the estimation of region statistics, (ii) the front evolution.

While technical details for the update of region statistics have been discussed, the implicit front evolution needs some clarifications. First, as mentioned before, the level-set evolutions (2.5) and (2.25) are not to be applied to the whole level-set, i.e. the image domain. Thanks to the Dirac function, factor of these two evolutions, only pixels within the vicinity of the zero level of ϕ have to be updated. Then, working only a narrow-band is natural (early level-set methods made this restriction as an approximation since the narrow-band was not obtained from the equations). In practice, it can be done by using a list of points corresponding to this band which is modified after each iteration. This band is in fact naturally obtained from the reinitialization of the level-set to a distance function.

The implicit way of propagating fronts remains only valid if the level-set ϕ remains a distance function, or at least a Lipschitz function. This has been one of the important issues since there is no guarantee from the evolution equations that the level-set will keep these properties. The easier way of dealing with that is to reinitialize the level-set to the distance function when this is needed. Several methods have been proposed for this task, one of the most used aims at solving the following PDE [142]:

$$\frac{\partial \phi}{\partial t} = \left(1 - \text{sign}\left(\phi_0\right)\right)\left(1 - \left|\nabla \phi\right|\right),$$

where ϕ_0 is the level-set we wish to reinitialize. Another efficient re-distancing algorithm is based on the Fast Marching technique [132] and has a very low complexity in time but the interpolation around the zero level for the initialization makes it less accurate. From our experience, a mixed approach appears to be the most accurate and efficient: the PDE is used for reinitialization in a small neighborhood of the zero level while the Fast Marching permits to extend the distance function to a larger band. Let us also mention the elegant work of Gomes et al. [62] where the level-set evolution equation was rewritten so that the level-set remains a distance function. However, this modification may decrease the convergence speed and the re-distancing step is often preferred.

2.6.2 Multi-resolution implementation

Multi-resolution extensions are often employed to avoid local minima and to increase convergence speed. Our algorithm can be easily modified to a multi-resolution version. For this purpose, we solve successively the segmentation problem on down-scaled images while using the result of the lower scale as an initialization. Let $I^{(s)}$ be the

down-scaled version of the image I such that each side of the image has been divided by 2^s. For $s = n$ to $s = 0$, we evolve successively each level-set until convergence according to:

$$\begin{cases} \phi_{t=0}^{(s)} = \tilde{\phi}_{t=\infty}^{(s+1)}, \\ \dfrac{\partial \phi^{(s)}}{\partial t}(\mathbf{x}) = \delta(\phi^{(s)}) \left(\nu^{(s)} \operatorname{div} \left(\dfrac{\nabla \phi^{(s)}}{|\nabla \phi^{(s)}|} \right) + \log \dfrac{p_1(I^{(s)}(\mathbf{x}))}{p_2(I^{(s)}(\mathbf{x}))} \right). \end{cases}$$

where $\tilde{\phi}$ corresponds to an upper-scaled version of ϕ. $\nu^{(s)}$ is the regularization weight at a scale s. Since this term is factor of the regularity constraint which is proportional to the length of the evolving curve, the weight $\nu^{(s)}$ should be inversely proportional to this length so as to keep a similar ratio between regularity and data-based terms. Assuming the length of the curve to be proportional to the perimeter of the image, $\nu^{(s)}$ can be expressed according to $\nu^{(0)}$: $\nu^{(s)} = \frac{\nu^{(0)}}{4^s}$.

This implementation permits a drastic increase of the convergence speed. In [Figure 2.14], we show the segmentation obtained with three different scales on the *frog* image. The convergence time drops from 16 seconds to only 2 seconds. Additional results on other natural high-resolution images are presented in [Figure 2.15].

2.7 Connections with classical clustering methods

2.7.1 K-means

The K-means algorithm is frequently used for clustering tasks because it is fast and rather simple to implement. This approach aims at finding the partitions of a dataset X in a given number N of clusters. A simple mean-squared error functional is employed, summing the squared distances between each point of a cluster j to the corresponding center c_j:

$$KM(X) = \sum_{x_i \in X} \min_{j=1..N} d^2(x_i, c_j). \tag{2.27}$$

The definition of a pertinent distance depends on the type of data and often, a natural distance can be obtained from the geometry of the ambient space. For general vector-valued data sets with no particular geometry, L1, L2 and Mahalanobis distances can be considered. The optimization process consists in iterating two steps: (i) the points are clustered for given centers, (ii) new centers are estimated from the last partitioning. A critical point of such methods is the choice of the initial centers. Several heuristics have been proposed but no general strategy can be defined and a simple random choice is often preferred.

Actually, the *Active Contours Without Edges* model introduced by Chan and Vese in [30] can be seen like a topological K-means. Let us recall the CV functional for the partitioning of an image I in N regions (please refer to the previous chapter for

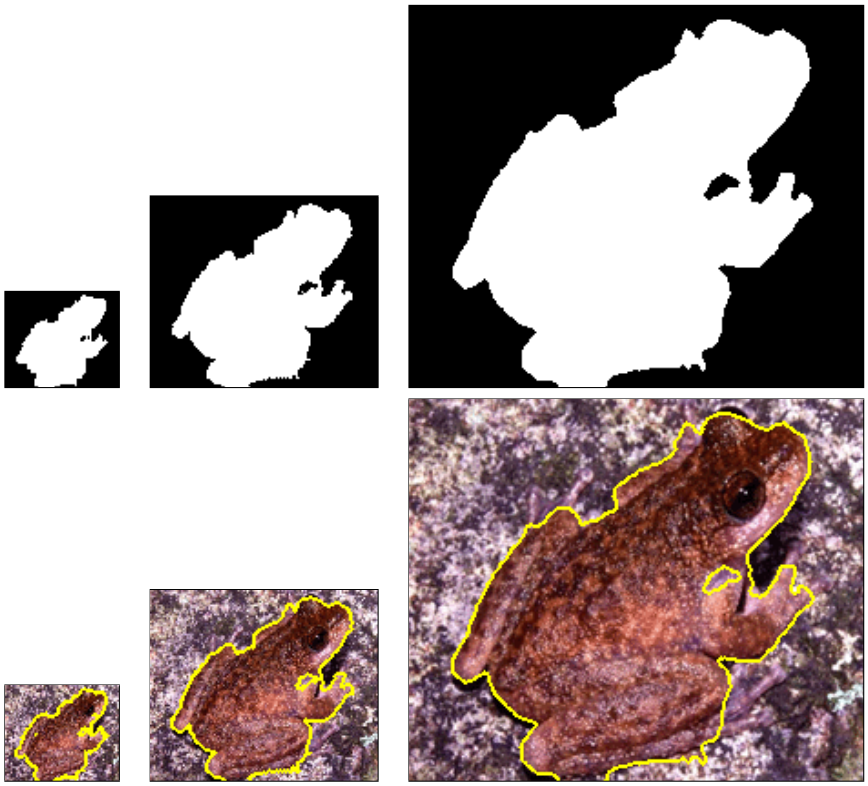


Figure 2.14: Multi-resolution segmentation of a *Frog* color image (250x210) - Curve evolution ($\nu = 5$, convergence time: **2 seconds**).

notation conventions):

$$E^0_{CV}(\mathcal{C}, c_i) = \int_\Omega \sum_{i=1}^{N} \chi_i(\mathbf{x})\, d^2(I(\mathbf{x}), c_i)\, d\mathbf{x} + \nu|\mathcal{C}| \tag{2.28}$$

If the regularization term $\nu|\mathcal{C}|$ is omitted, this energy reduces to the K-means criteria (2.27). However, the optimization process of the CV functional differs from the K-means. In classical K-means, the clustering is directly applied in the space of pixel values while the CV model is expressed on the image domain. Besides the possibility of imposing a spatial regularization permits to deal with noise and to remain consistent with the initialization. Consequently, the initialization is no more blind but it can be used to express prior spatial knowledge on the objects to be segmented.

2.7.2 Maximum-likelihood and the expectation-maximization (EM) algorithm

In this part, we show our optimization schemes of statistical parameters to be part of a very general statistical concept: the maximum-likelihood estimation. Referring to [13], we shortly summarize the maximum-likelihood estimation problem and the expectation-maximization algorithm [45]. A possible extension of our segmentation framework to mixture densities is also discussed.

Maximum-likelihood

The maximum-likelihood estimation problem consists in finding the set of parameters $\theta \in \Theta$ of a density function $p(\mathbf{x}|\theta)$ from an observed dataset $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ generated from this distribution. Assuming these data vectors independent and identically distributed, the density corresponding to $\mathcal{X}$ can be approximated as:

$$p(\mathcal{X}|\theta) = \prod_{i=1}^{N} p(\mathbf{x}_i|\theta) = \mathcal{L}(\theta|\mathcal{X})$$

The quantity $\mathcal{L}(\theta|\mathcal{X})$ is said to be the likelihood of θ given $\mathcal{X}$. The maximum-likelihood estimate (MLE) θ^* is then obtained by maximizing this quantity:

$$\theta^* = \operatorname*{Arg\,max}_{\theta \in \Theta} \mathcal{L}(\theta|\mathcal{X})$$

Going back to our segmentation functional, if we replace the dataset $\mathcal{X}$ by the pixel values in an subregion of the image $\Omega_\mathcal{X}$, the maximum-likelihood for this dataset gives the update rule of our statistical parameters (2.6). The optimal parameters can be easily obtained for simple densities like single Gaussian distributions (see (2.8) and (2.12)) but more complex densities like mixtures need more elaborate techniques.

Figure 2.15: Multi-resolution segmentation of high resolution color images - ($\nu = 5$, convergence time $\approx$ 20 seconds). Image resolutions are respectively (from left to right and top to bottom): 900x1200, 1400x1400 and 1280x960.

The EM algorithm

The EM algorithm is an iterative process that permits to solve the maximum-likelihood estimation problem for incomplete data. It can be used when observations are missing but it is also a technique to simplify the likelihood function when no analytical solution is available.

Assuming $\mathcal{X}$ to be an incomplete observation of a complete dataset $\mathcal{Z}$, a set of hidden or missing data is introduced such that $\mathcal{Z} = \{\mathcal{X}, \mathcal{Y}\}$. The complete-data likelihood function is then defined as $\mathcal{L}(\theta|\mathcal{Z}) = \mathcal{L}(\theta|\mathcal{X}, \mathcal{Y}) = p(\mathcal{X}, \mathcal{Y}|\theta)$.

The EM algorithm assume an initial guess of the MLE and then iterates two steps. The first step is called the E-step, it consists in estimating the expected value of the log-likelihood $\log p(\mathcal{X}, \mathcal{Y}|\theta)$ for a given observed data $\mathcal{X}$ and a previous estimate of the MLE $\theta^{(i-1)}$:

$$Q(\theta, \theta^{(i-1)}) = E\left[\log p(\mathcal{X}, \mathcal{Y}|\theta)|\mathcal{X}, \theta^{(i-1)}\right].$$

During the second step, the M-step, a new estimate of the MLE is estimated by maximizing the quantity estimated from the E-step:

$$\theta^i = \underset{\theta \in \Theta}{\operatorname{Arg\,max}}\, Q(\theta, \theta^{(i-1)}).$$

The two steps are repeated until convergence. The log-likelihood is guaranteed to increase at each iteration and the algorithm is proven to converge to local maxima. This algorithm is in particular well-adapted for mixture-densities. Analytical solutions of the M-step are available when considering a mixture of Gaussian densities in [13].

This approach can be used to extend our segmentation framework to Gaussian mixture models for each region. However, the rather small size of regions, i.e. the observed dataset, may be a limiting factor on the complexity of the density models.

2.8 Contributions and conclusions

In this chapter, we have presented a general method for the segmentation of scalar and vector-valued piecewise smooth images. An objective functional including region statistics and geometric constraints is obtained from a Bayesian formulation of the partitioning problem. Parametric and nonparametric densities are introduced to approximate pixel value distributions within each region. These two types of statistical models are presented for scalar and vector-valued images. The most important points of this chapter are the following ones:

1. Level-set functions are introduced directly in the objective functional to express region-dependent integrals and geometric constraints. The calculus of variations of the level-set leads naturally to an evolution equation equivalent to a front propagation. This makes our approach different from most approaches where the level-set representations are only used as a tool to implement the front evolution.

2. The new functional is minimized with respect to the level-set and region statistics simultaneously. Therefore, no supervised step is necessary to learn a priori region properties. A fast technique to update region statistics is proposed and the complexity of the complete algorithm remains linear in the size of the image.

Parts of this work have been presented in [122, 123]. This approach is well-adapted to smooth images but as soon as spatial structures are relevant to define regions of interest, the assumption of no-correlation between pixels, necessary to our model, is no longer valid. The next chapter proposes an extension to a wider range of images, in particular to images with texture.

Chapter 3

Cue Extraction & Integration

The approach presented so far is limited to smooth images. This is due to the important assumption that pixels inside a region are **independent** realizations of a random process:

$$p(I|\Omega_i) \approx \prod_{\mathbf{x}\in\Omega_i} p(I(\mathbf{x})|\Omega_i).$$

Consequently, only region histograms are used to model region statistics. This approximation is valid for smooth images but, as soon as texture is present in the image, the approach will fail to capture the pertinent information. By definition, a texture is related to a particular spatial distributions of pixel values within a given neighborhood. Since our approach does not integrate such spatial models, texture information has no influence on the segmentation.

Several possibilities are available to extend our approach to non-smooth images. Motivating our choice on a compromise between simplicity and genericity, a two-step unsupervised approach has been defined. The first phase consists in extracting texture features by considering a non-linear version of the structure tensor which is combined with the intensity to form a feature vector from which the image partition can be extracted. The region-based curve evolution technique of the previous chapter is considered to partition the image domain by considering this feature vector as input image. Experiments on a wide range of images show the potential of the approach, leading to an unsupervised segmentation of images with texture.

The second section is dedicated to the incorporation of motion information. A similar two-step strategy is employed. First, motion features are extracted from an image sequence using a robust estimation of the optical flow. Then, motion components are combined with intensity/color and texture characteristics to define a new feature vector. This vector can be used for the segmentation/detection of moving objects and their tracking along the sequence. A new coupling between level-sets is also introduced for the tracking of multiple objects to avoid overlaps between objects. Finally,

Figure 3.1: Examples of natural images with textures.

experiments on different image sequences show the genericity of the approach with the incorporation of gray/color, texture and motion features.

3.1 Static cue: intensity/color and structure tensor

3.1.1 Dealing with texture

When looking to natural images, texture is often one of the most characteristic visual properties [Figure 3.1]. Therefore, analyzing and modeling such information is important not only to understand natural scenes but also to mimic human perception. Despite numerous studies on this topic during the past decades [72, 43, 151, 90], the problem of texture segmentation remains unsolved. The first critical step is the definition of what a texture is. If we look in the literature, it appears that there is no unique definition. The main reason is the wide variety of textures [Figure 3.2]. A rather general but vague definition can be found in [116]: "texture images are specially homogeneous and consist of repeated elements, often subject to some randomization in their location, size, color, orientation, etc". In the absence of a more accurate definition, various models have been proposed, often targeting at a particular class of textures. Most of them can be classified into two areas [172]. The first one is "statistical modeling": high-order statistical models of textures are built from probability distributions on random fields. They use the theory of Markov random fields [64, 43]. The second one is "filtering theory": they are inspired by neurobiological studies of the visual cortex suggesting the presence of a multitude of oriented filters [94, 135]. The methods the most frequently used are based on Gabor filters [59, 151] and wavelet transforms [93, 136]. Recent works by Zhu et al. [172] and Simoncelli et al. [116] propose to combine both areas, applying statistical modeling on filter responses.

Unsupervised segmentation of textured images is different from texture analysis because no samples are available to learn texture parameters. Alternate techniques can be considered by iterating parameters estimation and segmentation. However, whichever method is used, Gabor filters [128, 130, 108] or Markov random fields, the estimation of the parameters remains difficult issue and lots of redundancy are often

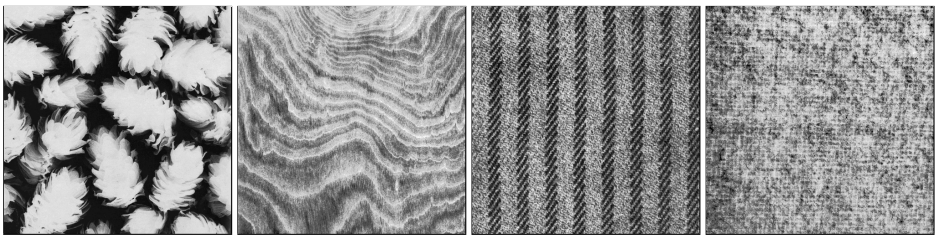


Figure 3.2: Different types of textures from the Brodatz dataset [16].

present in the extracted feature channels. A very interesting solution was proposed by Bigün et al. in [11] with the introduction of the structure tensor. It can be classified as a filtering approach but, in contrast to other methods, the structure tensor only yields three different feature channels per scale. This tensor has mainly been used to determine the intrinsic dimensionality of images in [11, 55] by providing a continuous measure to detect critical points like edges or corners. However, the structure tensor does not only give a scalar value reflecting the probability of an edge but it also includes the texture orientation. All these properties make this matrix a good descriptor for textures.

The only problem of the original work in [11] is the Gaussian smoothing used for the structure tensor which reduce spatial accuracy in the feature space by dislocating edges. To overcome this limitation, Weickert and Brox proposed in [162] to develop a non-linear version of the structure tensor based on a non-linear matrix-valued diffusion. We propose to extend this idea by also considering the intensity during the the nonlinear diffusion.

3.1.2 Nonlinear structure tensor

The nonlinear structure tensor introduced in [162] is based on the classic linear structure tensor [55, 11, 118, 87, 63]:

$$J_\rho = K_\rho * (\nabla I \nabla I^\top) = \begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2, \end{pmatrix} \tag{3.1}$$

where K_ρ is a Gaussian kernel with standard deviation ρ, I is the image, and subscripts denote partial derivatives. In the case of color images, all channels are taken into account by summing the tensor products of the particular channels [168]:

$$J_\rho = K_\rho * \left(\sum_{i=1}^{3} \nabla I_i \nabla I_i^\top \right). \tag{3.2}$$

The major problem of the classic structure tensor is the dislocation of edges due to the smoothing with Gaussian kernels. This leads to inaccurate results near discontinuities in the data. [Figure 3.3] shows the structure tensor obtained on a natural textured

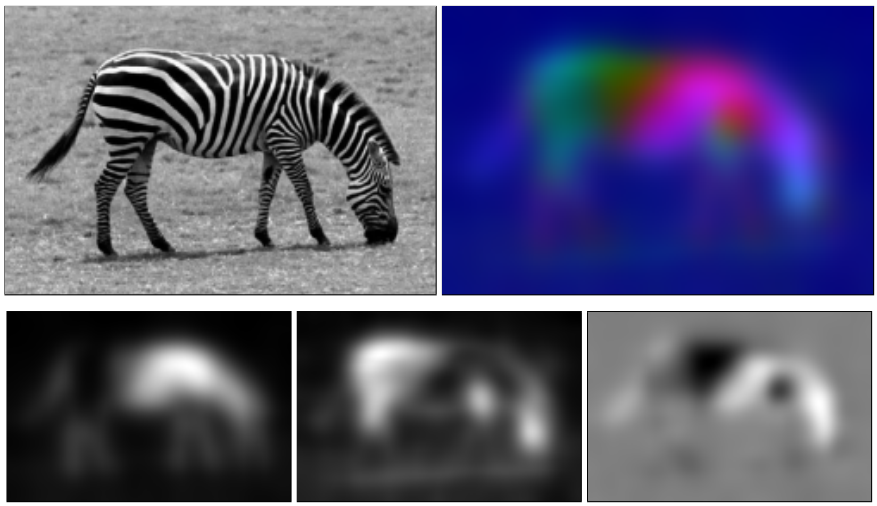


Figure 3.3: Classical structure tensor on the *zebra* image (TOP: original image and vector representation of the structure tensor, BOTTOM: I_x^2, I_y^2, $I_x I_y$ components of the structure tensor).

image (the color representation of the structure tensor is composed of the three different components of the matrix as RGB components). The structure tensor is clearly a good descriptor for the zebra texture but the contour of the zebra is quite blurry. First mentioned in [162], the basic idea to address this problem is to replace the Gaussian smoothing by a nonlinear diffusion. This can be done by applying the scheme of nonlinear matrix-valued diffusion introduced in [145, 146], and diffusivity functions without a contrast parameter. In the following, we describe the details of this procedure. Nonlinear diffusion is based on the early work of Perona and Malik [115]. For a review we refer to [160]. The main idea is to reduce the smoothing in the presence of edges. The resulting diffusion equation is

$$\partial_t u = \text{div}\,(g(|\nabla u|)\nabla u), \tag{3.3}$$

with $u(t=0)$ being the initial image and g a decreasing *diffusivity function*. Perona and Malik proposed two different diffusivity functions

$$g(|\nabla u|) = \frac{1}{1+|\nabla u|^2/\lambda^2}, \tag{3.4}$$

$$g(|\nabla u|) = \exp\Big(-\frac{|\nabla u|^2}{2\lambda^2}\Big), \tag{3.5}$$

where λ is a contrast parameter steering the transition from forward diffusion (i.e. edges are blurred by the process) to backward diffusion (i.e. edges are enhanced by the process).

Equation (3.3) can only be used with scalar-valued data like a gray-level image. Gerig

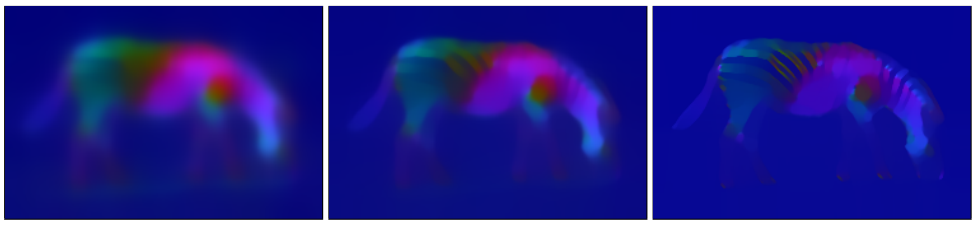


Figure 3.4: Nonlinear structure tensor on the *zebra* image for $p = 0.5$, 1 and 1.5.

et al. [61] introduced a version of nonlinear diffusion for vector-valued data

$$\partial_t u_i = \operatorname{div}\left(g\left(\sum_{k=1}^{N} |\nabla u_k|^2\right)\nabla u_i\right) \qquad \forall i = 1, \ldots, N, \tag{3.6}$$

where u_i is an evolving vector channel and N the total number of vector channels. Note that in this approach all channels are coupled by a joint diffusivity, so an edge in one channel also inhibits smoothing in the others. A recent survey on vector-valued diffusion can be found in [148].

When regarding the components of a matrix as components of a vector, what is reasonable, since the Frobenius norm of a matrix equals the Euclidean norm of the resulting vector, it is possible to diffuse a matrix, such as the structure tensor, with the above-mentioned scheme. In fact, this complies with the scheme proposed in [145].

A rather critical issue is the appropriate choice of the diffusivity function g. Applied to the structure tensor, the diffusivity functions proposed by Perona and Malik, for example, will not work properly. This is because the structure tensor contains first derivatives, which have very local responses of different magnitude. Thus it is mostly impossible to choose an appropriate contrast parameter. Instead we use diffusivity functions of type [149]

$$g(|\nabla u|) = \frac{1}{|\nabla u|^p}, \tag{3.7}$$

with $p \in \mathbb{R}$ and $p \geq 1$. These diffusivity functions include for $p = 1$ *total variation (TV) flow* [2, 49], a diffusion filter that is equivalent to TV regularization [127, 22]. For $p = 2$, one obtains the so-called *balanced forward backward diffusion* introduced in [75]. Their properties fit the requirements of the structure tensor very well: there is no contrast parameter, they remove oscillations, and experiments show that they yield piecewise constant results. Therefore they can preserve, or for $p > 1$ even enhance, important edges. Unfortunately, these diffusivity functions will lead to numerical problems when the gradient gets close to zero. This problem can be avoided by adding a small positive constant ϵ to the denominator.

$$g(|\nabla u|) = \frac{1}{|\nabla u|^p + \epsilon}. \tag{3.8}$$

Regarding the implementation, we apply the AOS scheme [164] which allows efficient computation of such flows also for small ϵ. For ϵ in the area of 0.01, where the approximation of such kinds of flow is much better than for larger ϵ, causing less blurring effects, the AOS scheme is around three orders of magnitude faster than a simple explicit scheme.

Our texture features consist of the three different components of the spatial structure tensor $u(t=0) = (I_x^2, I_y^2, 2I_xI_y)$ with I being the gray-level image. By applying equation (3.6) with the diffusivity function of equation (3.8) we obtain the smoothed components, i.e. the nonlinear structure tensor. In the case of color, the initial condition is extended to the sums $u(t=0) = (\sum_i (I_i)_x^2, \sum_i (I_i)_y^2, 2\sum_i (I_i)_x (I_i)_y)$. Note that in both cases the third vector component has to be weighted twice, since it appears twice in the matrix. In [Figure 3.4], the nonlinear structure tensor is estimated for different values of p. High values of p make the edges very accurate but some large texture structures remain visible.

3.1.3 Combination of intensity and texture cues

At this stage it becomes important to combine texture information with image intensity in a way that allows the simplification of the data, the removal of outliers, and the closing of structures, using the cues of *all* channels. For this purpose, vector-valued diffusion according to equation (3.6) is very well suited, as it couples all vector channels by a joint diffusivity. This way, the information of all channels is used to decide whether an edge is worth to be enhanced or not. However, for a balanced coupling, equation (3.6) assumes the values of all channels to have approximately the same range. Unfortunately, a simple normalization of all channels to the same range, which would solve this problem immediately, is not a good approach. If a channel contains no information, e.g. there is no texture, the data in this channel should be constant. A normalization, however, amplifies the noise in such a situation. Therefore, the problem of how all channels get approximately the same range has to be solved in a different manner.

For the texture channels this can be done by replacing the structure tensor by its square root. Given the eigenvalue decomposition of the structure tensor $J_0 = T(\lambda_i)T^\top$, the square root can be computed by

$$\tilde{J}_0 := \sqrt{J_0} = T(\sqrt{\lambda_i})T^\top.$$

In the case of a gray-level image, this comes down to

$$\tilde{J}_0 = \begin{pmatrix} \frac{I_x}{|\nabla I|} & -\frac{I_y}{|\nabla I|} \\ \frac{I_y}{|\nabla I|} & \frac{I_x}{|\nabla I|} \end{pmatrix} \begin{pmatrix} |\nabla I| & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{I_x}{|\nabla I|} & \frac{I_y}{|\nabla I|} \\ -\frac{I_y}{|\nabla I|} & \frac{I_x}{|\nabla I|} \end{pmatrix} = \begin{pmatrix} \frac{I_x^2}{|\nabla I|} & \frac{I_xI_y}{|\nabla I|} \\ \frac{I_xI_y}{|\nabla I|} & \frac{I_y^2}{|\nabla I|} \end{pmatrix} = \frac{J_0}{|\nabla I|}.$$

With color images, this simplification is not possible, due to the sum of matrices for each color channel. With a common range of values, only the question of the diffusivity

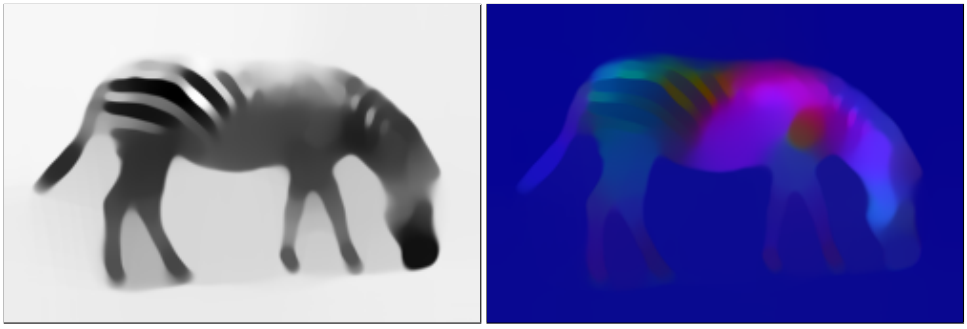


Figure 3.5: Image intensity and nonlinear structure tensor obtained after coupled diffusion with $p = 1.2$.

function remains open. As it was stressed in Section 2, diffusion processes based on the diffusivity functions in equation (3.8) yield some nice properties, like the removal of oscillations and piecewise constant results. Also the fact that there is no further contrast parameter is beneficial at this stage. It has also been mentioned that for $p > 1$ the diffusion process will be edge enhancing. This is very useful, since the feature combination should result in simplified data with precise common edges. It has to be noted that in the continuous case, backward diffusion is ill-posed. Discretization, however, has been shown to resolve this problem [161].

Hence, for our experiments we have chosen $p = 1.2$, though other values of $p > 1$ will also work. The larger p, the more important is the influence of the edge enhancement compared to the simplification effect. We want to stress that p has nothing to do with the contrast parameter λ in the Perona-Malik diffusivities, which is rather critical to choose. Conversely, p globally specifies the ratio between edge enhancement and simplification. As edge enhancement has basically a positive effect for our application, it would be best to use large p. However, this will considerably increase diffusion time necessary to obtain also an appropriate simplification effect.

The coupling of the intensity and the structure tensor components permits to improve the texture features [Figure 3.5]. The legs of the zebra are now visible in the texture components contrary to the simple nonlinear structure tensor of [Figure 3.4]. The smoothing of the intensity is also improved since region with a homogeneous texture have large diffusivity. This set of four features can now be considered as input for the segmentation.

3.1.4 Unsupervised segmentation of textured images

As mentioned before, the segmentation approach of the last chapter is restricted to smooth images because pixels inside a region are assumed to be independent. Markov random fields can be used to express directly these spatial links between pixel values. If these models are well-suited for texture analysis and synthesis, their high complex-

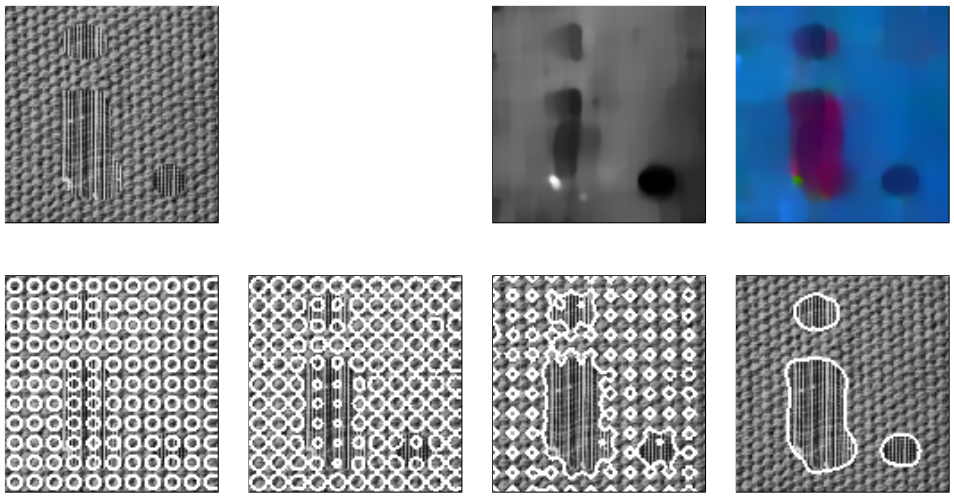


Figure 3.6: Result on a synthetic texture image - TOP, FROM LEFT TO RIGHT: original image, input features (smoothed image and structure tensor components), BOTTOM: curve evolution.

ity is a disadvantage in an unsupervised segmentation approach. The same problem for the learning texture parameters appears when a combination of linear filters is considered. However, when one wants to segment a textured image, accurate models of each texture may not be necessary since only a characteristic distinguishing the different textures would be sufficient. Based on the structure tensor, the features extraction gives us such characteristics where each pixel value reflects spatial properties. Therefore, it becomes possible to integrate texture information assuming pixels to be independent realizations of a random process within these channels:

$$p(I|\Omega_i) \approx p(u|\Omega_i) \approx \prod_{\mathbf{x}\in\Omega_i} p(u(\mathbf{x})|\Omega_i),$$

where u is a feature vector including the image intensity and the structure tensor component obtained after coupled diffusion. Following the framework developed in the previous chapter, the segmentation can be formulated as an energy minimization problem. The corresponding bi-partitioning energy with an additional regularization term is expressed by considering the level-set representation ϕ of the border between Ω_1 and Ω_2. ϕ is defined as the signed distance function to this border (we refer the reader to the last chapter for more details). The optimal segmentation is then obtained by minimizing

$$E(\phi) = \int_\Omega \Big(H(\phi) \log p(u(\mathbf{x})|\Omega_1) + (1 - H(\phi)) \log p(u(\mathbf{x})|\Omega_2) + \nu|\nabla H(\phi)\Big)| \, d\mathbf{x}.$$

The variational framework is not completely defined so far, since it still lacks the definition of the probability density function. Rather than modeling the joint probability

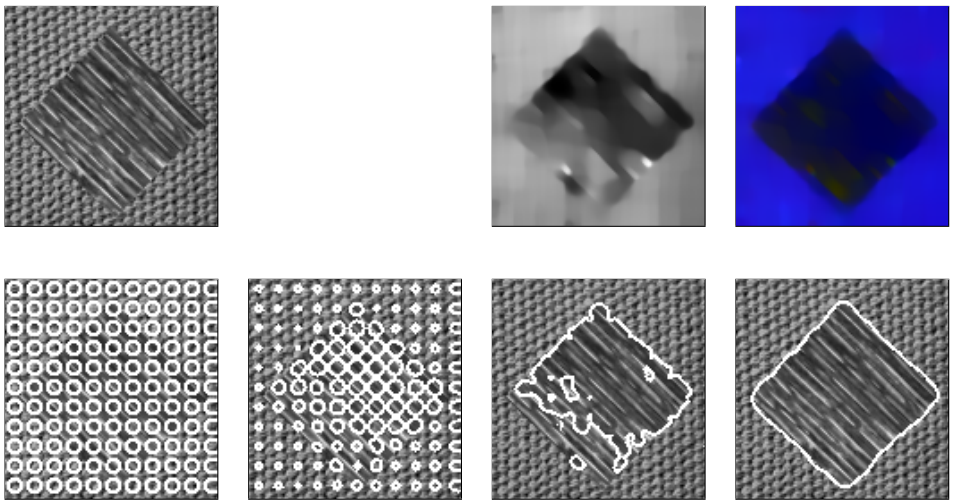


Figure 3.7: Result on a synthetic texture image - TOP, FROM LEFT TO RIGHT: original image, input features (smoothed image and structure tensor components), BOTTOM: curve evolution.

of the four channels, we assume the channels to be independent. Let $p_{ij}(\mathbf{x})$ the conditional probability density function of a value $u_j(x)$ to appear in region Ω_i, the energy becomes:

$$E(\phi) = \int_\Omega \left(H(\phi) \sum_{j=1}^{4} \log p_{1j}(u(\mathbf{x})) + (1 - H(\phi)) \sum_{j=1}^{4} \log p_{2j}(u(\mathbf{x})) \right) d\mathbf{x} + \int_\Omega \nu |\nabla H(\phi)| \, d\mathbf{x}.$$

A Gaussian approximation is a reasonable choice to represent region densities for each channel and their parameters are introduced as unknown parameters. The optimization process alternates between the minimization of the energy with respect to ϕ and the estimation of the optimal statistical parameters. The first variations of ϕ gives the following level-set evolution:

$$\frac{\partial \phi}{\partial t}(\mathbf{x}) = \delta(\phi) \left(\nu \, \mathrm{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \sum_{j=1}^{4} \log \frac{p_{1j}(I(\mathbf{x}))}{p_{2j}(I(\mathbf{x}))} \right) \quad \forall \mathbf{x} \in \Omega,$$

while the optimal statistical parameters are obtained directly from their empirical estimations in the corresponding regions. The whole approach introduces only two parameters: the number of iterations of the coupled diffusion for the features extraction and the weight of the regularization term of the curve evolution. The second one has a clear interpretation and can be fixed for regular objects ($\nu = 1$ is used for the experiments). The meaning of the first one is less obvious at first sight. The diffusion process permits to characterize local geometric structures. Small structures are characterized

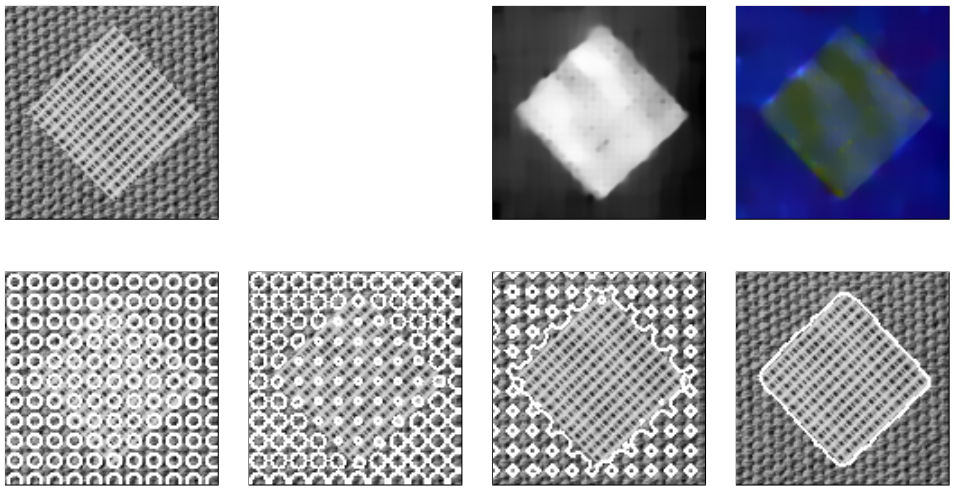


Figure 3.8: Result on a synthetic texture image - TOP, FROM LEFT TO RIGHT: original image, input features (smoothed image and structure tensor components), BOTTOM: curve evolution.

during the first iterations but coarser and more complex structures need more iterations. Therefore, two aspects have to be considered to set the number of iterations: the scale of the structures and their "complexity". A first empirical estimate can be obtained from the size of the image. For the coming experiments, the number of iterations was taken proportional to the perimeter of the image: $N = \alpha * (width + height)$. Using this conjecture, we are able to deal with a wide range of synthetic as well as real images with the same choice for α.

Segmentation of gray textured images

In order to verify the discrimination power of our segmentation approach with respect to texture, it was first applied to several synthetic gray images composed of the well-known Brodatz textures [16]. The complete process (features extraction and curve evolution) is presented in [Figures 3.6,3.7,3.8 and 3.9]. Extracted features are shown on the top of each figure and the corresponding curve evolution is shown on the bottom. The same parameters are used for each experiment: $\nu = 1$, $\alpha = 0.08$ and $p = 1.2$. Other synthetic results are presented in [Figure 3.10] where the same curve initialization with small circles was used on all examples. The results are quite satisfactory and they show the method to perform very well in delimiting the textured parts with a high accuracy. Contrary to classical kernel-based results where a loss of accuracy is generally observed because of the size of the kernels, our nonlinear features extraction avoids any blurring effect and the curve fits to the border of the objects at the pixel level. Even textures that are difficult to distinguish by humans, such those in [Figure 3.10f] and [Figure 3.10h], can be handled by the nonlinear structure tensor combined

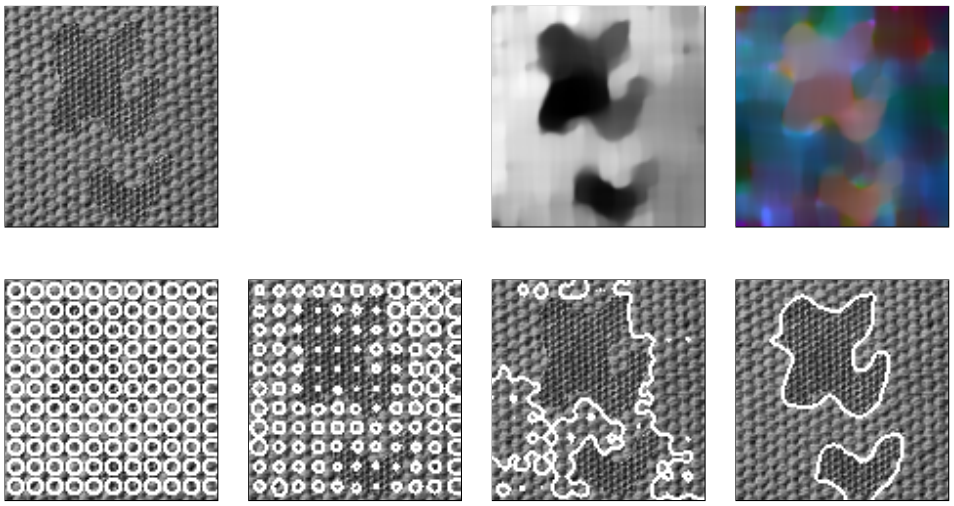


Figure 3.9: Result on a synthetic texture image - TOP, FROM LEFT TO RIGHT: original image, input features (smoothed image and structure tensor components), BOTTOM: curve evolution.

with the statistical region model.

To validate our approach on a larger set of images, we build a set of 100 synthetic images composed of two textures randomly taken from the Brodatz dataset. A complex symmetric mask was considered not to bias the experiment with a particular geometry [Figure 3.11(top)]. The twenty first samples of this test suite are shown in [Figure 3.11(bottom)]. Our approach has been tested on the whole suite with different parameters. Since the ground-truth image partition is available, we are able to estimate the ratio of well-classified pixels over all image pixels. In [Figure 3.12a], we show this ratio for the first twenty images for different values of α. The average percentage of well-classified pixels is 75% for $\alpha = 0.05$ and 81% for $\alpha = 0.1$. The performance of the algorithm for different values of α is better described in [Figure 3.12b] where the repartition of the images with respect to the classification rate is presented. Our approach performs very well for a wide range of parameters but $\alpha = 0.1$ gives clearly the best average ratio with a large majority of images including less than 20% of miss-classified pixels. However, a "bump" can be observed in [Figure 3.12b] for a ratio around 0.6, showing that several textures are not well captured with our algorithm. This limitation will be discussed in the next section.

Although synthetic test images are best suites to verify the potential and the limits of a method, results achieved for real images are more interesting. So we applied our method also to natural images. Contrary to synthetic images, a textured region may have large variation in a real image. This can be seen in the commonly used *zebra2* image in [Figure 3.13]. This make the segmentation of real images much more challenging. Our approach can deal with these large texture variations thanks to the

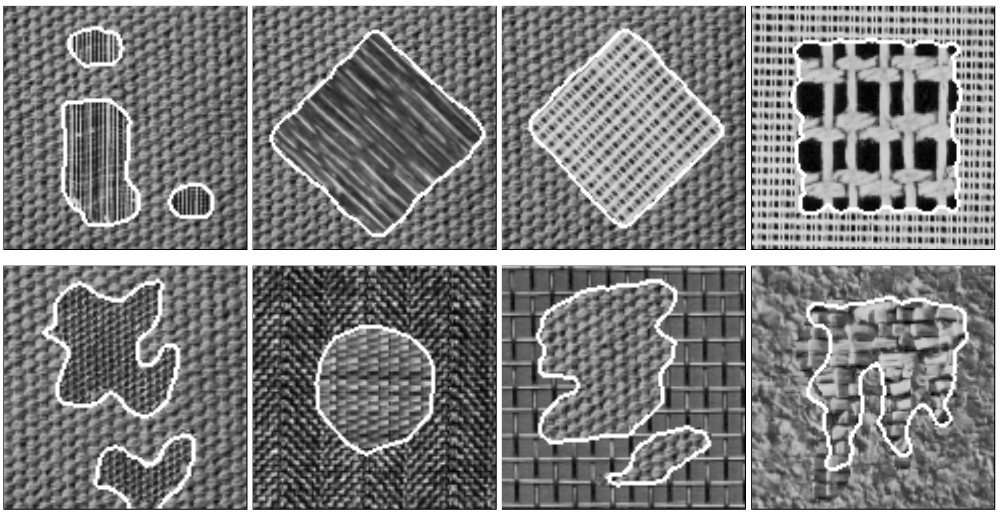


Figure 3.10: (a)-(h) FROM LEFT TO RIGHT, TOP TO BOTTOM: Segmentation results for synthetic texture images.

nonlinear regularization. Several convincing results are shown in [Figure 3.13, 3.14, 3.15 and 3.16]. All these results were obtained still using the same parameters: $\nu = 1$, $\alpha = 0.08$ and $p = 1.2$. To show the influence of the diffusion time on the the simplification of the image, results with two different values of α are shown in [Figure 3.14]. The first one is obtained for $\alpha = 0.05$ and the other one for $\alpha = 0.08$. More details are clearly captured for the smaller diffusion time.

Segmentation of color textured images

The structure tensor can also be expressed for color images by taking the sum of the gradients of each component. Color information can be combined with the corresponding structure tensor to extract particular structures. Like for gray images, a coupled diffusion of the color and texture components allows for the extraction of a feature vector from which the image partition can be obtained. We show several examples on real images where the addition of color improves the segmentation. In [Figures 3.17 and 3.18], we consider the same images as in [Figures 3.15 and 3.16] but with color. The result for the *frog* image is greatly improved and the frog's outline is almost extracted perfectly. Two more examples are shown in [Figures 3.19 and 3.20]. The fish of [Figures 3.19] has rather big structures which share a similar color with the background. Our approach is still able to extract correctly the fish with the default parameters. The last example is a giraffe with a complex background. The body of the giraffe is successfully discriminated from the background.

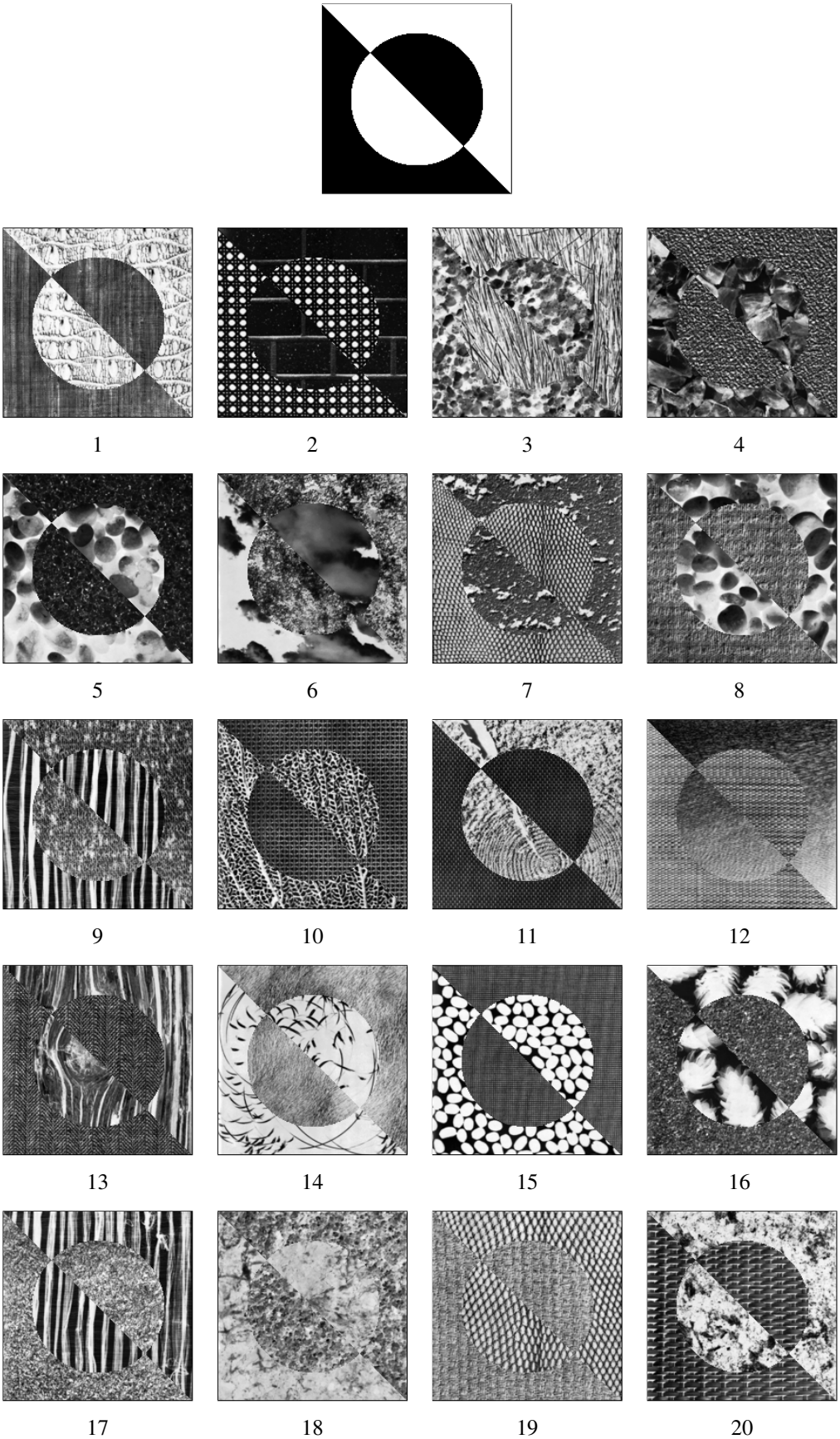
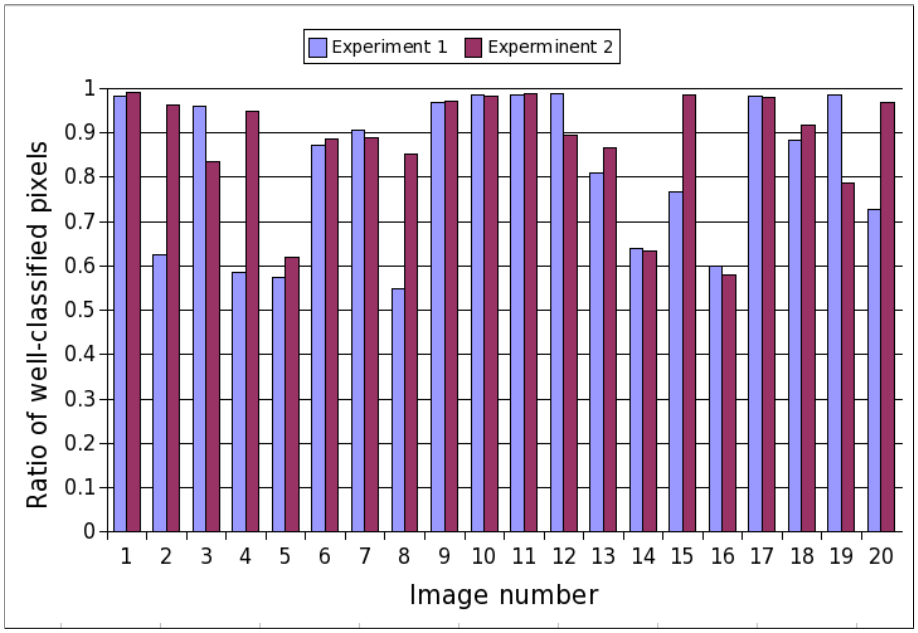
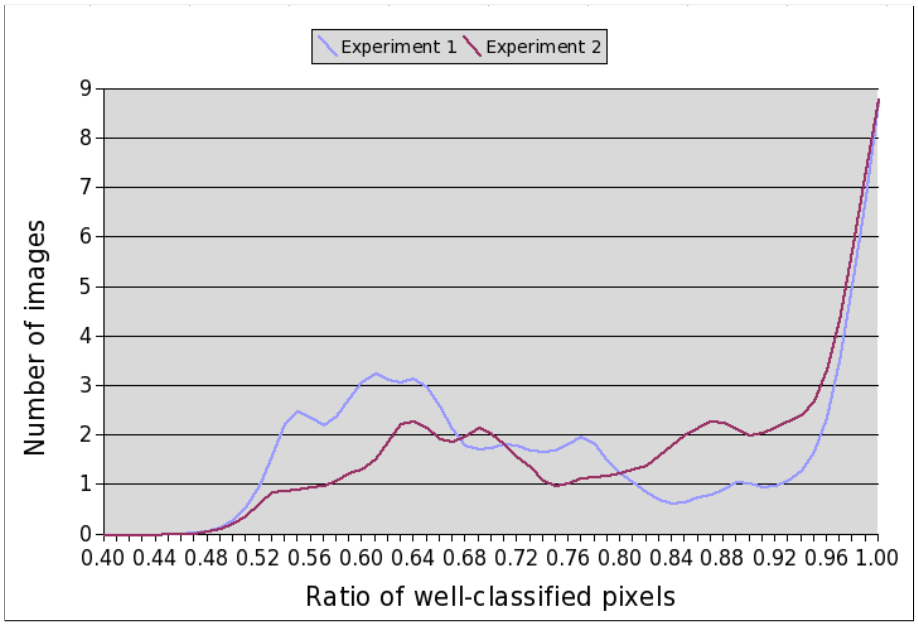


Figure 3.11: Brodatz texture test suite - TOP: region mask, BOTTOM: 20 samples of the 100 test images.



(a)



(b)

Figure 3.12: Experiments on the Brodatz texture test suite for *Experiment 1*: $\alpha = 0.05$, *Experiment 2*: $\alpha = 0.1$ - TOP: Ratio of well-classified pixels for the first 20 images, BOTTOM: Ratio repartition over the test suite.

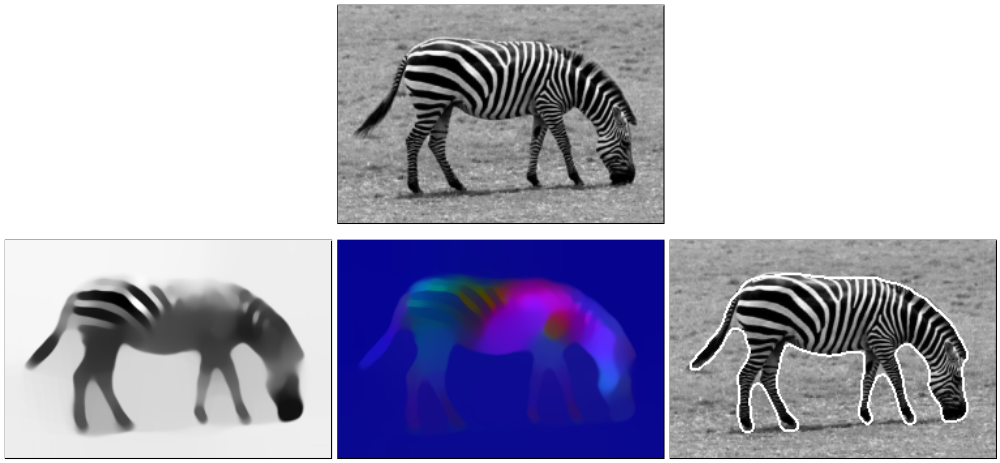


Figure 3.13: Result on the *zebra2* image - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

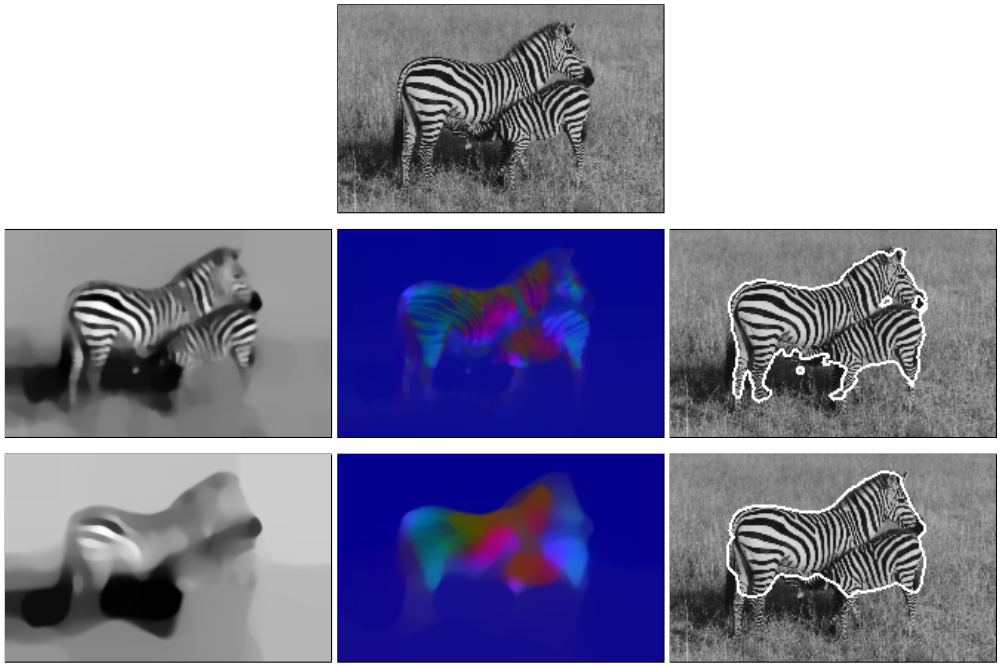


Figure 3.14: Results on the *zebra* image - TOP: Original image, BOTTOM ROWS, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation for different regularization times.

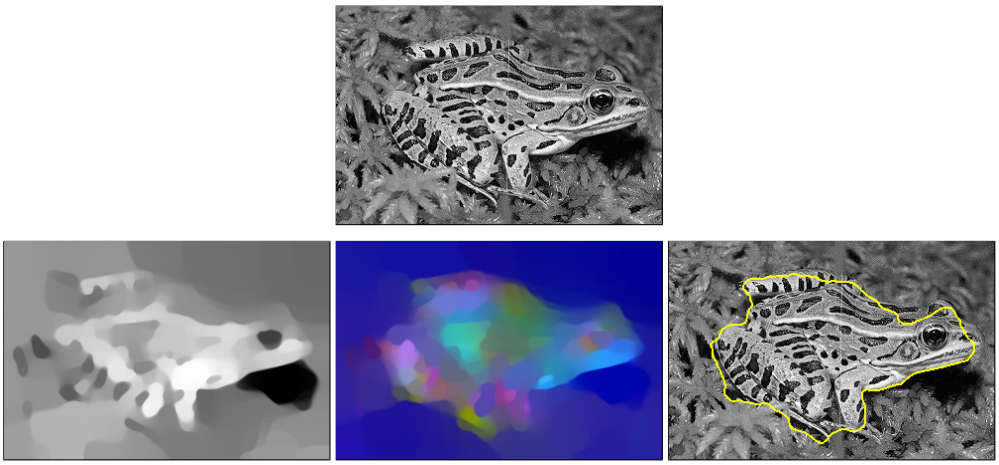


Figure 3.15: Result on the *frog* image - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

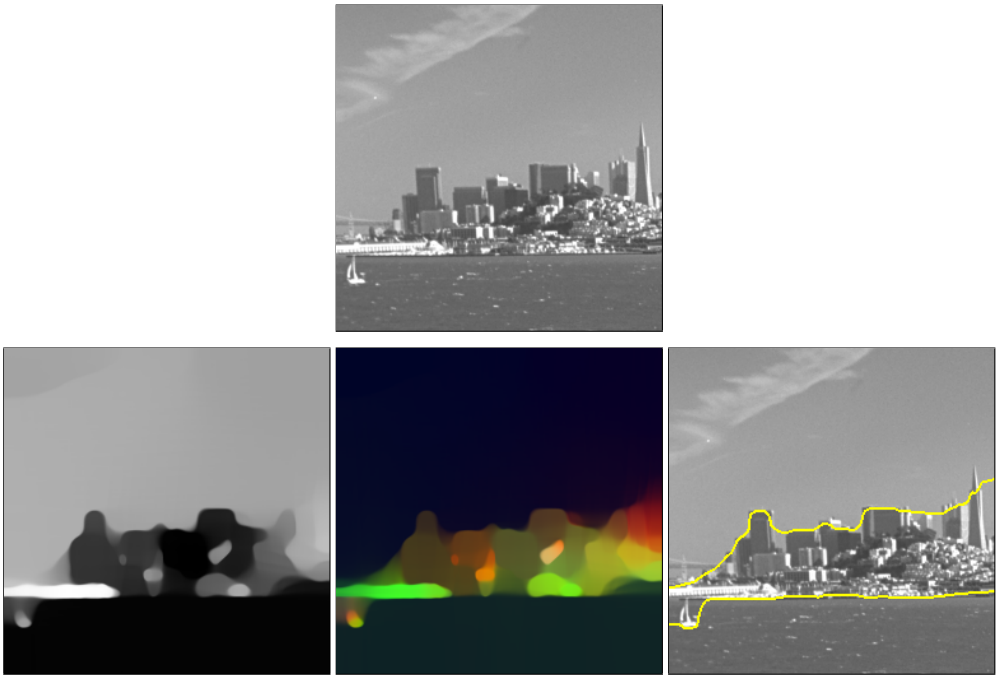


Figure 3.16: Result on the *city1* image - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

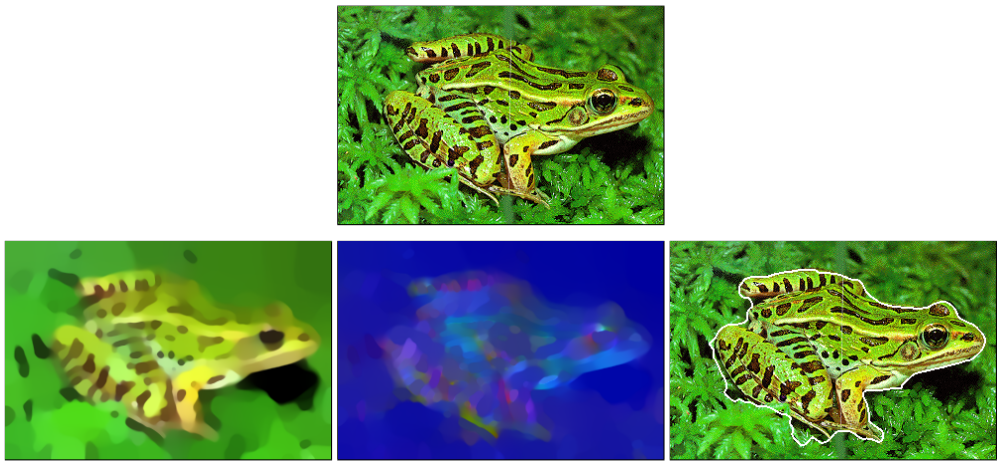


Figure 3.17: Result on the *frog* image with color - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

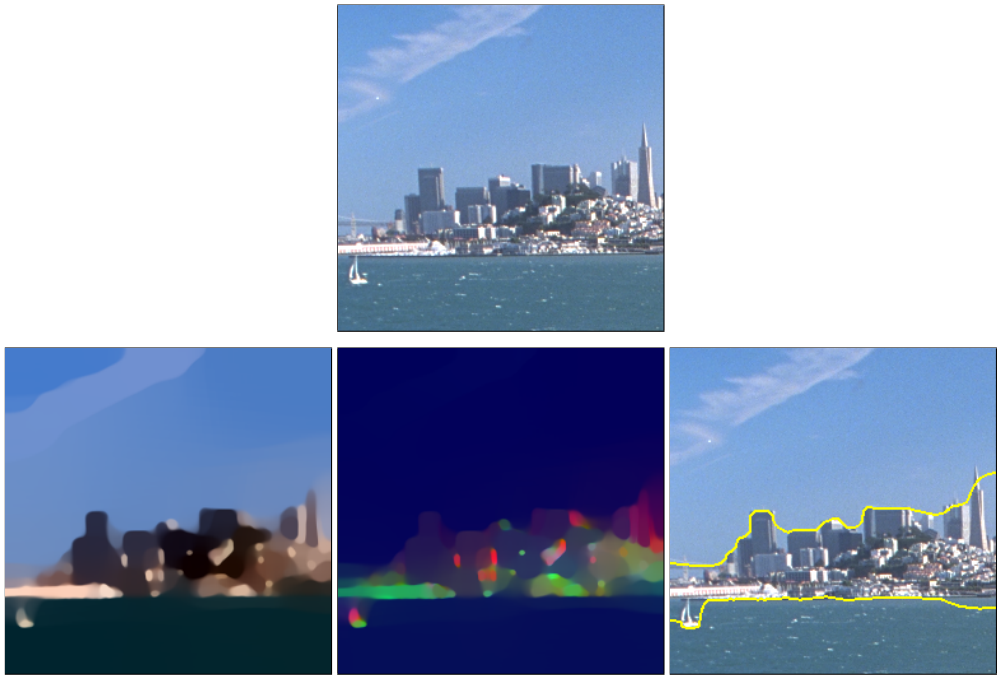


Figure 3.18: Result on the *city1* image with color - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

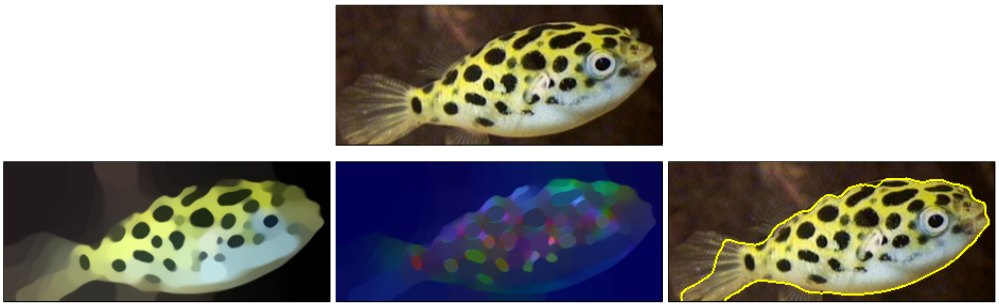


Figure 3.19: Result on the *fish2* image with color - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.

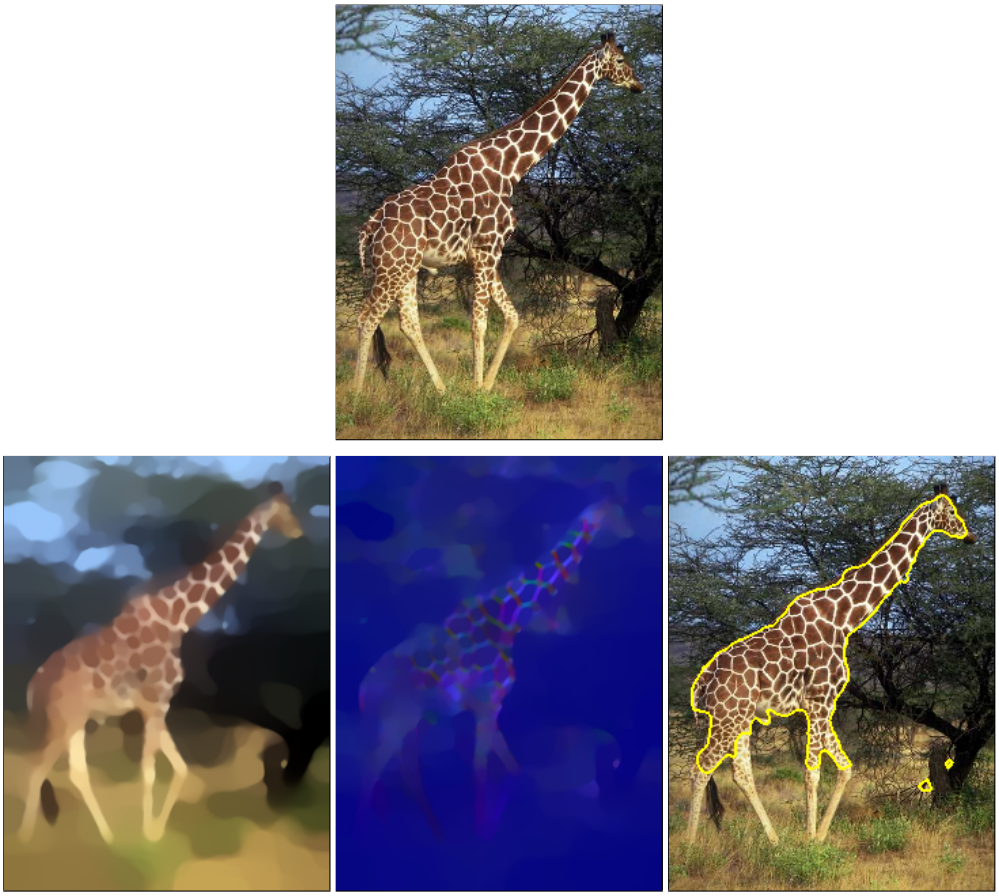
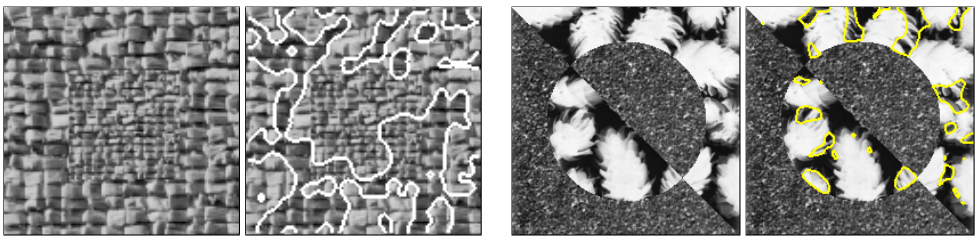


Figure 3.20: Result on the *giraffe* image with color - TOP: Original image, BOTTOM, FROM LEFT TO RIGHT: input features (smoothed image and structure tensor components), final segmentation.



(a) (b)

Figure 3.21: The method fails when texture structures have a large difference in scale.

3.1.5 Limitations and possible extensions

The feature extraction characterizes texture magnitude and orientation at a given scale. The scale is given by the number of iterations of the diffusion process and so, it is the same for all pixels. When in an image, textures are defined at a different scale, the extraction process is no longer adequate. A typical synthetic example is composed by the same texture at different scales [Figure 3.21(a)] but it may occur in more general images where the structures of the different textures are defined at various scales [Figure 3.21(c)]. A straightforward extension is to consider texture features at different scales by either introducing the nonlinear structure tensor obtained at different iterations or extracting these features at different scales of the image. If such modification includes the whole information, it also increases dramatically the number of channels and redundant information is introduced, making the second phase (the image partitioning) more difficult. With the same idea of a reduced feature space, Brox and Weickert proposed in [21] an elegant and efficient approach to tackle this problem by extending our framework. Similar texture features are considered but a local scale measure is also introduced. The scale component is deduced from the speed of a diffusion process based on the TV (total variation) flow. Another limitation is the restriction to a bi-partitioning of the image. This issue has been discussed for the segmentation of smooth images in the last chapter. Texture segmentations were restricted to the simplest case of two regions to concentrate only on texture discrimination but all the extensions presented earlier for more than two regions can be applied to this case.

3.1.6 Conclusion

In this section, we have presented an unsupervised approach to segment textured images. It is composed of two phases: the feature extraction and the image partitioning. Texture features are based on a nonlinear version of the structure tensor obtained from a coupled nonlinear diffusion of the second moment matrix components with the image intensity. The image is then segmented by applying a region-based curve evolution on the feature vector composed of the regularized version of the image intensity and

the texture features. This framework has proven to be able to automatically segment a wide range of synthetic and real images, with and without color. The quality of the results is comparable to recent supervised methods which is quite remarkable given that only two parameters need to be chosen and all the results were obtained with the same choices.

3.2 Motion cue: optical flow

Considering motion, the optical flow is the principal way of extracting this information. Optical flow estimation is a complete research area on its own, and there exist plenty of different techniques; see e.g. [7, 99, 141] for overviews. Interestingly, the nonlinear structure tensor, already applied for the texture, can also be used here [19]. Two different kinds of features for motion can be derived from the optical flow: first the optical flow vector as such, and second its magnitude. While the first case is more general and also allows the detection and tracking of moving objects in a scene where the camera is moving as well, the second case might be more attractive, if motion needs to be detected in a scene with a static camera.

3.2.1 Optical flow extraction

The structure tensor can also be used for optical flow estimation. In this case a spatio-temporal structure tensor is needed. It is obtained the same way as described in the last section, with only the initial condition being changed to $u(t = 0) = (I_x^2, I_y^2, I_z^2, 2I_xI_y, 2I_xI_z, 2I_yI_z)$ where z describes the time axis. In the case of color, these values will again be extended by using the corresponding sums. Although the explicit usage of the structure tensor for optical flow estimation has been proposed in [11] we will use the early approach of Lucas and Kanade [88] here, which implicitly also employs the structure tensor.

Optical flow estimation is based on the assumption that image structures do not alter their gray values during their motion. This is expressed by the optical flow constraint [67]

$$I_x u + I_y v + I_z = 0. \tag{3.9}$$

Again subscripts denote partial derivatives, and u and v are the unknown components of the optical flow vector. As this is only one equation for two flow components, the optical flow is not uniquely determined by this constraint *(aperture problem)*. A second assumption has to be made. Lucas and Kanade proposed to assume the optical flow vector to be constant within some neighborhood B_ρ of size ρ. The optical flow in some point (x_0, y_0) can then be estimated by the minimizer of the local energy function

$$E(u, v) = \frac{1}{2} \int_{B_\rho(x_0,y_0)} (I_x u + I_y v + I_z)^2 dxdy. \tag{3.10}$$

Figure 3.22: Two successive images of the *stau2* sequence.

A minimum (u, v) of E satisfies $\partial_u E = 0$ and $\partial_v E = 0$, leading to the linear system

$$\begin{pmatrix} \int_{B_\rho} I_x^2 dxdy & \int_{B_\rho} I_x I_y dxdy \\ \int_{B_\rho} I_x I_y dxdy & \int_{B_\rho} I_y^2 dxdy \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\int_{B_\rho} I_x I_z dxdy \\ -\int_{B_\rho} I_y I_z dxdy \end{pmatrix} \tag{3.11}$$

Instead of a sharp window for B_ρ often a convolution with a Gaussian kernel K_ρ is used yielding

$$\begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -K_\rho * I_x I_z \\ -K_\rho * I_y I_z \end{pmatrix}. \tag{3.12}$$

The linear system can only be solved if the system matrix is not singular. Such singular matrices appear in regions where the image gradient vanishes or the aperture problem remains present. In such situations the smaller eigenvalue of the system matrix is close to 0, and one may only compute the so-called normal flow (the optical flow component parallel to the image gradient). Using a sufficient amount of smoothing for the structure tensor, however, will greatly reduce such singular situations and dense results are obtained in most cases. A technique that combines the Lucas-Kanade method with dense flow fields can be found in [163].

Obviously the entries of the linear system are the five components of the spatio-temporal structure tensor J_ρ. Of course, instead of the classic linear structure tensor, the nonlinear structure tensor will be employed here, as proposed in [19]. For the diffusion of the structure tensor, TV flow will be applied, i.e. $p = 1$.

The idea to use a data adaptive structure tensor for optical flow estimation has already been proposed in [104]. Although it is still a linear method, this structure tensor is closely related to the nonlinear structure tensor. In [97] a method for how to find appropriate parameters for this technique has been proposed.

3.2.2 Combining spatial and temporal cues

In the case of the optical flow channels, a comparable range can only be achieved by some weighting. We weighted those channels with factor 64, so a displacement of

Figure 3.23: Detection of moving objects between frame 130 and 131 in the *stau2* sequence. LEFT: Using optical flow. RIGHT: Using intensity, optical flow, and texture.

4 pixels per frame corresponds to the possible maximum value in the other channels. This is a reasonable choice, since above a velocity of 4 pixels per frame differential optical flow estimation methods become more and more unreliable, so image sequences with such large displacements should not be used for input or a multi-scale approach should be considered.

3.2.3 Motion segmentation

With image sequences, motion information becomes available and the optical flow components can be added to the feature vector. For a gray value sequence the feature vector therefore consists of 6 components, namely the gray value, the 3 components of the spatial structure tensor, and the two components of the optical flow.

For the special case of motion *detection*, it is often useful to weight the optical flow channels higher than the other channels. Otherwise, the resulting segmentation might not split the moving objects from the non-moving background, but some bright, textured parts from dark parts. Experimentation pushed us to weight twice the optical flow channels. In [Figure 3.24] we show the results obtained for different combinations of the feature channels on the same test images (Figure 3.22) from the traffic scene *stau2* [1]. While the optical flow gives a robust information for the coarse shape and location of the objects, the addition of intensity and texture information increases the accuracy of the detection. The complete feature vector including all information is the only one that permits to separate the cars on the left, as shown in [Figure 3.24(iii)]. In [Figure 3.25], a colored and textured moving box is segmented with a very high accuracy when using all the features.

The complete feature vector also permits to capture small moving objects. A clip-

[1] The sequence *stau2* was kindly provided by the *Institut für Algorithmen und kognitive Systeme* at the University of Karlsruhe, Germany. The size of the whole sequence is $(700 \times 566 \times 1034)$

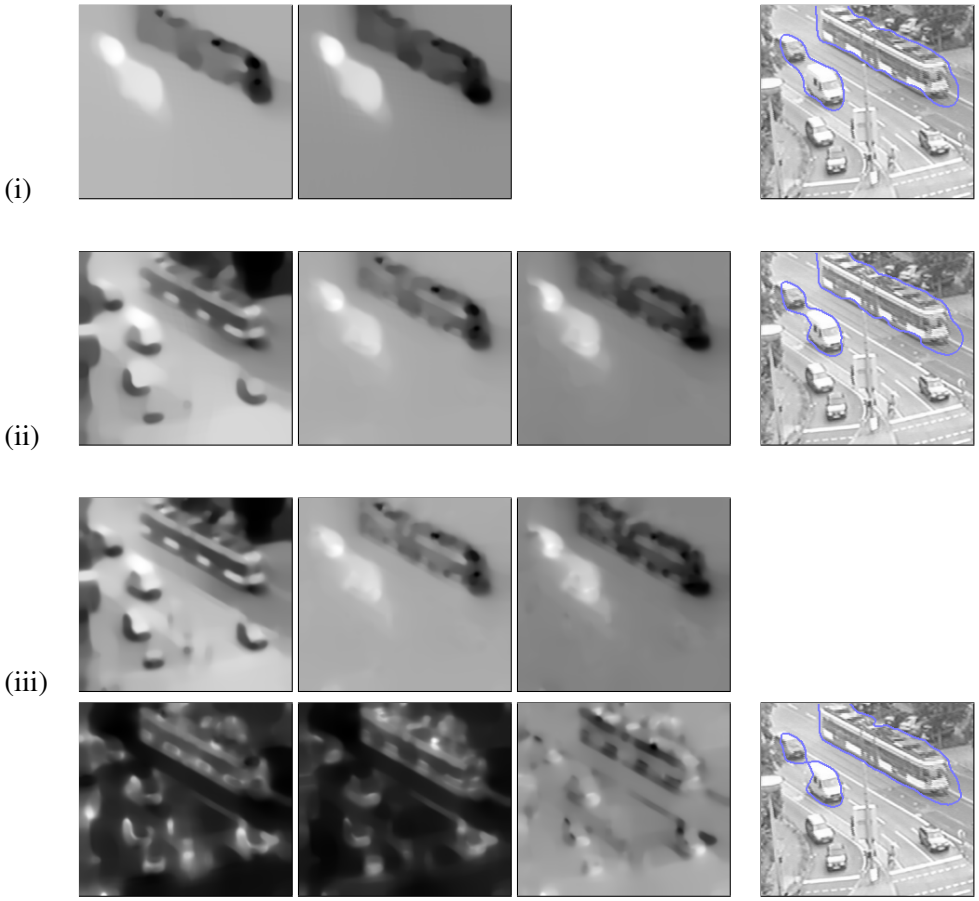


Figure 3.24: Motion segmentation integrating different cues. LEFT: Feature channels. RIGHT: Final Segmentation. (a) optical flow. (b) Intensity and optical flow. (x) Intensity, optical flow, and texture.

Figure 3.25: Object detection in the *box* sequence. LEFT: Using optical flow. RIGHT: Using color, optical flow, and texture.

ping from the *stau2* sequence clearly shows this benefit. Pedestrians and bicycles are rarely detected if only the optical flow is considered. [Figure 3.23] is an example where two more objects are detected thanks to spatial information (these objects are really moving in the sequence).

3.2.4 Tracking

When switching from segmentation to tracking, only a few things change. The main issue is that for tracking the initial position and number of objects are known by definition, so the initialization of the level-set is not a problem anymore. This allows for the assumption of a scene with one object and the background to be dropped, and it becomes possible to simultaneously track multiple objects. For this purpose, a new energy functional has to be introduced that allows the coupled evolution of an arbitrary number of level-sets, one for each object. The coupling is important in order to avoid overlaps between objects.

Before the general case of M objects, we consider the simplified case of only two objects Ω_1 and Ω_2, and the background Ω_b. One level-set is assigned to each object such that $\phi_i(\mathbf{x}) > 0$ if $\mathbf{x} \in \Omega_i$. Adding the non-overlapping constraint yields the characteristic functions

$$\begin{aligned} \chi_1 &= H(\phi_1)(1 - H(\phi_2)), \\ \chi_2 &= H(\phi_2)(1 - H(\phi_1)). \end{aligned} \tag{3.13}$$

The characteristic function assigned to the background is the remaining part of the image domain:

$$\chi_b = H(\phi_1)H(\phi_2) + (1 - H(\phi_1))(1 - H(\phi_2)) = 1 - (\chi_1 + \chi_2). \tag{3.14}$$

Let e_1, e_2, and e_b be the log-likelihoods of a pixel to be in Ω_1, Ω_2, and Ω_b. Considering a feature vector u of dimension N (gray-level, color, texture, motion...) and assuming no interaction between its components, these log-likelihoods can be expressed as:

$$e_i = \sum_{j=1}^{N} \log p_{ij}(u_j), \quad i = 1, 2, b. \tag{3.15}$$

According to our framework, the tracking of Ω_1 and Ω_2 at frame t is obtained by minimizing the region integrals of the negative log-likelihoods:

$$\begin{aligned} E(\phi_1, \phi_2) &= -\int_\Omega e_1\chi_1 + e_2\chi_2 + e_b\chi_b \\ &= -\int_\Omega (e_1 - e_b)\chi_1 + (e_2 - e_b)\chi_2 + e_b. \end{aligned} \tag{3.16}$$

By means of the Euler-Lagrange equations a system of evolution equations for the level-sets is obtained:

$$\begin{cases} \dfrac{\partial \phi_1}{\partial t} = \delta(\phi_1)\big((e_1 - e_b)(1 - H(\phi_2)) - (e_2 - e_b)H(\phi_2)\big) \\ \dfrac{\partial \phi_2}{\partial t} = \delta(\phi_2)\big((e_2 - e_b)(1 - H(\phi_1)) - (e_1 - e_b)H(\phi_1)\big). \end{cases} \tag{3.17}$$

This model is slightly different from the one we proposed in [17] where the coupling between the level-sets was weaker. Overlapping was avoided there, yet there has been no competition between objects when they are next to each other. Now in the case of overlapping objects, the system of evolution equations simplifies to:

$$\begin{cases} \dfrac{\partial \phi_1}{\partial t} = \delta(\phi_1)\,(e_1 - e_2)\,/2 \\ \dfrac{\partial \phi_2}{\partial t} = \delta(\phi_2)\,(e_2 - e_1)\,/2 \end{cases} \tag{3.18}$$

which effectively represents a competition between the two objects.

The extension to the general case of M objects is straightforward. Including the smoothness constraint for each level-set leads to the following energy:

$$\begin{aligned} E(\{\phi_k, k = 1..M\}) =& \nu \sum_{i=1}^{M} \int_\Omega |\nabla H(\phi_i)| d\mathbf{x} - \int_\Omega e_b(\mathbf{x})\, d\mathbf{x} \\ &+ \sum_{i=1}^{M} \int_\Omega (e_i(\mathbf{x}) - e_b(\mathbf{x}))H(\phi_i)\Pi_{j\neq i}(1 - H(\phi_j))\, d\mathbf{x}. \end{aligned} \tag{3.19}$$

A system of coupled evolution equations is obtained for the minimization of this energy:

$$\begin{aligned} \frac{\partial \phi_k}{\partial t} =& \delta(\phi_k)\Bigg((e_k - e_b)\prod_{j\neq k}(1 - H(\phi_j)) \\ &- \sum_{j\neq k}(e_j - e_b)H(\phi_j)\prod_{l\neq j, l\neq k}(1 - H(\phi_l)) + \nu\,\mathrm{div}\,\frac{\nabla\phi_k}{|\nabla\phi_k|}\Bigg), \quad k = 1..M \end{aligned} \tag{3.20}$$

while region statistics are still updated after each iteration.

The tracking algorithm was first tested on two gray-level sequences: the famous *Hamburg taxi* sequence[2] [Figure 3.26] as well as the *stau2* sequence already considered in the motion detection part [Figure 3.27]. In the first sequence, the objects are

[2]The sequence was created at the University of Hamburg and can be obtained from `ftp://csd.uwo.ca/pub/vision`

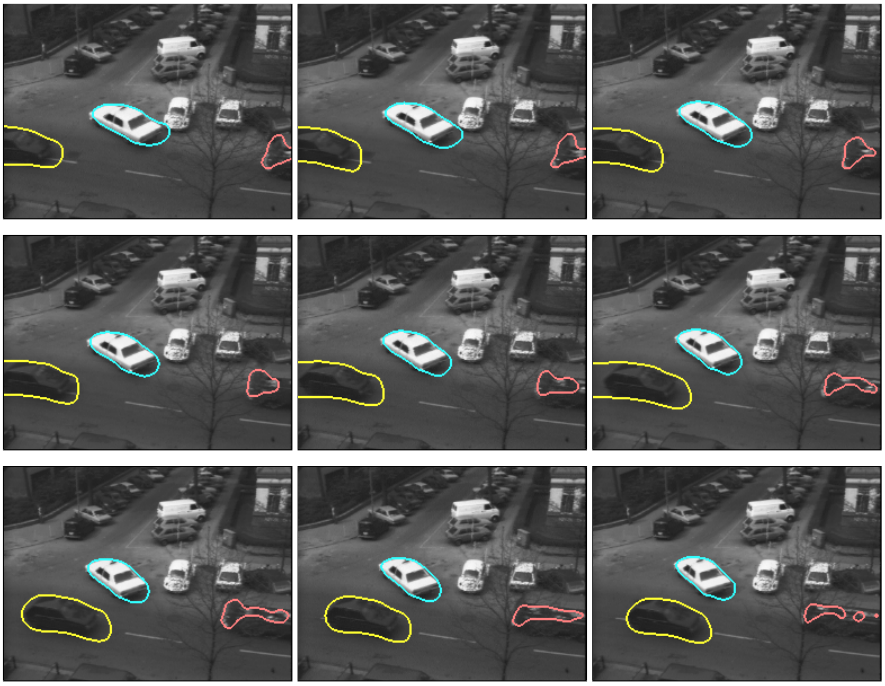


Figure 3.26: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result for the *Hamburg Taxi* sequence $(256 \times 190 \times 20)$.

relatively large, yet its quality caused by the camera hardware available at that time is rather challenging. Since the gray value and texture information is unreliable for the van on the right, only the motion information can provide for its rough tracking. Note that also in [Figure 3.27] the small cyclist can only be tracked because of its clear motion. In [Figure 3.28], only a small sub-part of the *stau2* sequence is considered. In this example, very small objects – one pedestrian and two bikes – are tracked along the sequence. In particular, the bike encircled in yellow is tracked despite very low intensity difference with the background. Finally, we have tested our approach in critical cases where objects overlap each other. The coupling between the level-sets permits to deal with this kind of situation. The *soccer* example depicted in [Figure 3.29] benefits from this behavior since three players very close to each other have to be tracked. This color sequence introduces also another complexity by having a moving camera, which actually does not change our approach.

3.3 Contributions and conclusions

The two-step approach proposed in this part permits to extend the Bayesian formulation of Chapter 2 to a wider range of images. Textured images have first been considered, including gray and color images. Then, an extension to image sequences incorporates the optical flow and allows for the extraction and tracking of moving objects.

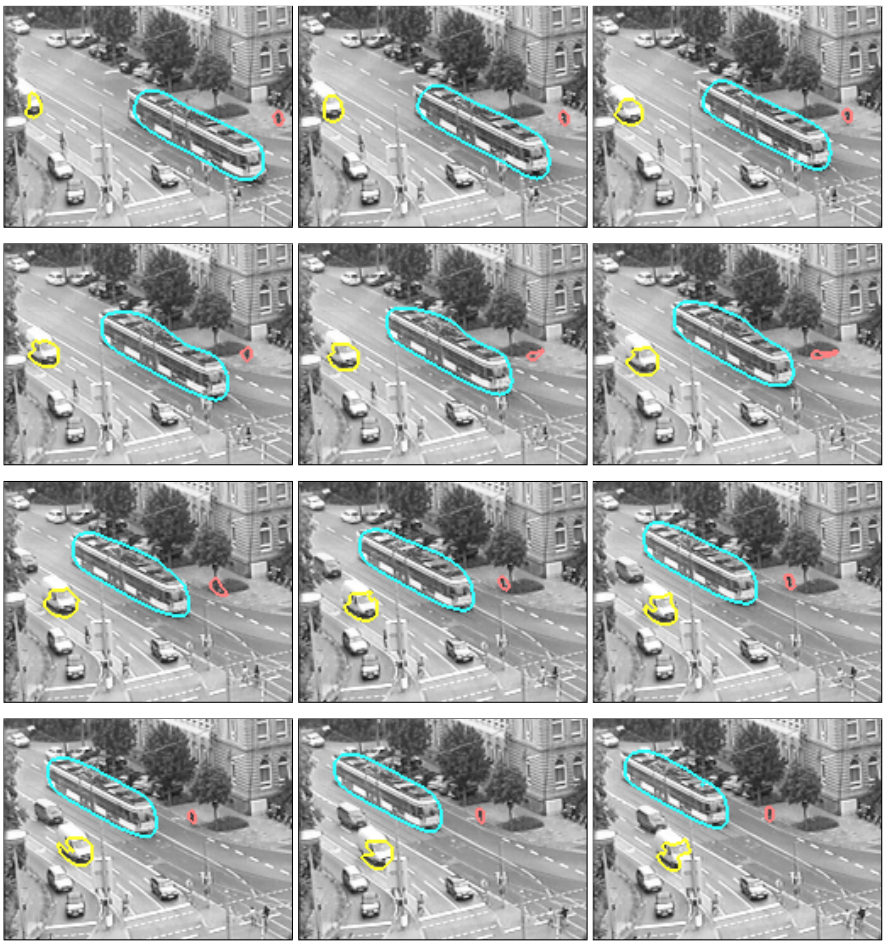


Figure 3.27: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result for a downsampled version of the *stau2* sequence $(170 \times 130 \times 76)$.

The generalization of our approach is based on several points:

1. The first one consists in extracting important characteristics of image regions. A reduced set of features is proposed to characterize texture and motion information. A nonlinear version of the structure tensor is used to extract texture properties from image derivatives while the optical flow is estimated to characterize moving objects in image sequences.

2. A coupled nonlinear diffusion of the vector formed by these characteristics is proposed to extend local features and remove noise while keeping an accurate separation between regions.

3. A new coupled level-set formulation is introduced to track multiple objects, allowing a natural competition between the level-sets representing each object.

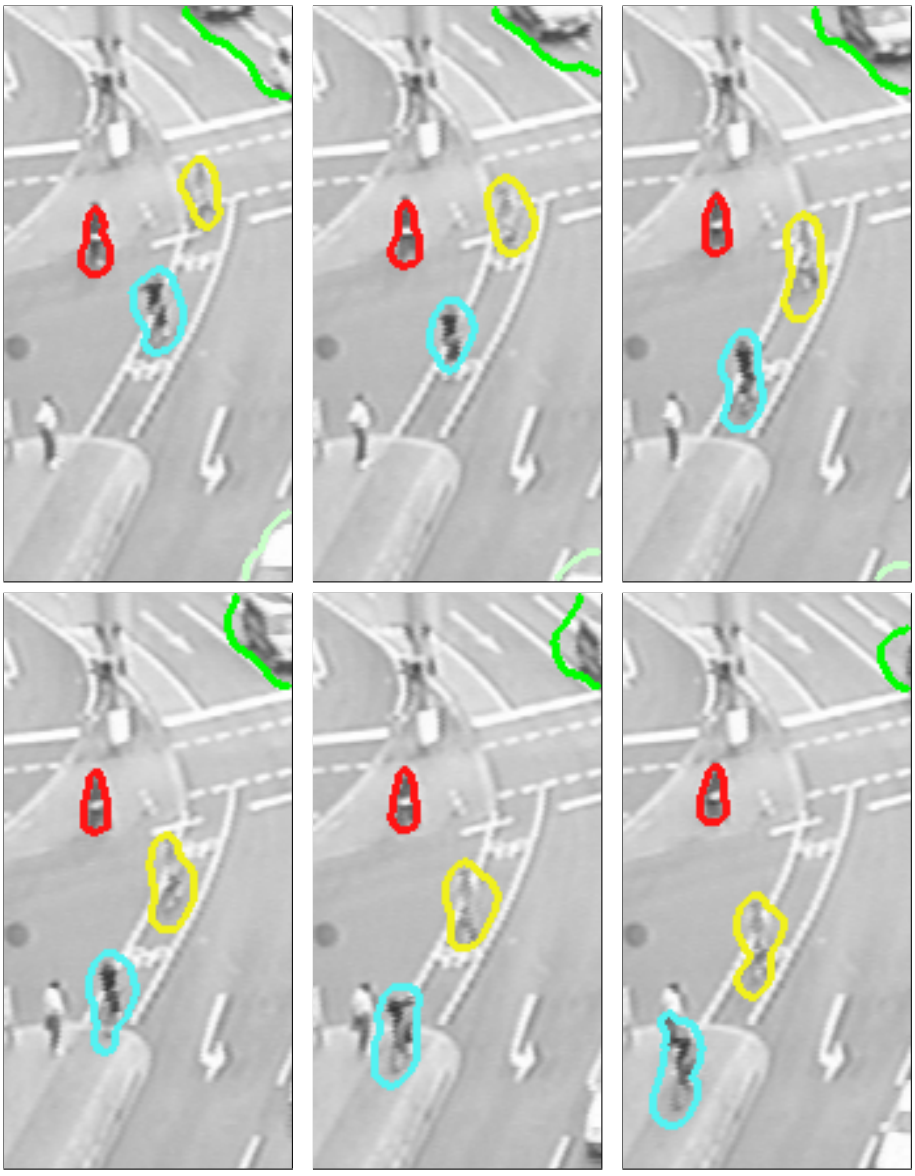


Figure 3.28: FROM LEFT TO RIGHT: Tracking of small objects in a sub-part of the *stau2* sequence $(110 \times 220 \times 90)$.

Figure 3.29: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking of 3 players in the *soccer* sequence $(180 \times 130 \times 40)$.

Parts of this work have been published in several conference papers [120, 17, 121] and research reports [89, 18].

Acknowledgments: I would like to thank Thomas Brox for his important contribution in the work presented in this chapter.

Chapter 4

Implicit Representation for Prior Shape Knowledge

Often, the extraction of a given object from an image is an ill-posed problem and data-based formulations as the ones presented so far will fail. The ill-posedness of the problem can be due to various perturbing factors like noise, occlusions, missing parts, cluttered data, etc... To overcome these limitations, one will try to make use of prior knowledge to constrain the extraction process. Different cues may be considered to include prior knowledge, such as the ones we have considered to define region terms with pixel-values distributions or texture properties. However, even when considering a particular application, these cues may vary a lot from one image to another and we will rather integrate prior knowledge about the shape of the objects which may remain similar. For example, this is the case for most organs extracted in medical imaging and in general, the shape of the object can be classified in a given shape family. Modeling this particular family is of high interest to help for the extraction of a new instance from another image. This strategy has been considered in many critical applications where the use of such prior information was mandatory and it is often referred to model-based. Such approaches consist of two steps: (i) the shape modeling from a training set and (ii), the constrained extraction in a new image. The Bayesian framework considered so far is well suited to integrate this kind of information using the partition probability term $p(\mathcal{P}(\Omega))$. Previously approximated by a simple regularization constraint, more complex distributions can be introduced. In the following, the distribution of the training shapes will be learnt during the shape modeling phase. This will give us a direct estimation of the probability of a new shape defined by a partition of the image $\mathcal{P}(\Omega)$, i.e $p(\mathcal{P}(\Omega))$.

Assuming the object(s) of interest to belong to a given class of shapes, the first step aims at building an accurate and compact model of this class from a set of training

samples. Then, the object extraction is an optimization problem where the best model parameters must be found so as to map best the data (possibly modulo some pose parameters). Both steps are quite critical: the shape model must be able to generate any object of the considered family while it should be restrictive/specific to this class, and then, appropriate features must be selected from the image to fit an instance of the modeled family to the object boundaries. Front evolution techniques have shown to be able to capture efficiently image features and the introduction of prior shape knowledge in such techniques will give a powerful tool to answer to the second issue, i.e. the attraction of the model toward image features.

The choice of the shape representation is quite a critical component when considering shape knowledge. Early works were considering rather simple geometric representations like straight segments [53] or ellipsoids [14]. While these models are efficient in terms of performance and low complexity when modeling simple geometric structures, they fail to account for local information and important variabilities of the object of interest. More general representations were later introduced with landmark representations [38], Fourier descriptors [139] or B-splines [41, 40]. A statistical study of these representations for a given training set is generally employed to capture its distribution and to obtain a compact model.

Fourier descriptors were considered by Staib and Duncan in [139] to approximate closed curves. This gives a very compact representation of the shapes and a shape distribution can be modeled by learning the distribution of the weights of each Fourier descriptor. However, this representation is limited to single closed curves and the compactness of the representation is only valid for smooth curves.

In [38], Cootes *et al.* proposed a more geometrical way of modeling a class of objects by simply considering a parametric representation of the contours, positioning landmark points at chosen locations. These control points need to be placed in a coherent way on all training contours, such that a one to one correspondence is obtained for each control point from one contour to another. A principal component analysis (PCA) is then applied on the landmark representations of the training shapes to reduce the dimension of the model by considering only the principal modes of variation with respect to a mean shape. Moreover, statistical analysis of the shape distribution can be also conducted in the PCA subspace, reducing again the degree of freedom of the model (less "false" shapes can be generated). Non-linear extensions of this approach have been proposed in [37] and [152]. This modeling approach is quite general and it has been applied on various types of shapes (hands, faces, medical images...). Nevertheless, as mentioned by the authors, the positioning of the landmark points is critical and different topologies in the learning set cannot be dealt with. Despite these limitations, this method remains well-adapted to represent numerous families of shapes with a fixed topology and where obvious and persistent shape features permit to easily set control points.

More recently, Klassen et al. [79] proposed a continuous representation of closed

curves, considering an arc length parameterization. A Riemmanian structure can be associated to the considered manifold, allowing the computation of geodesic distances between shapes, as well as statistical analysis. Several constraints are introduced to integrate natural invariance, first with respect to rotations but also removing the re-parameterization group. Advanced applications to shape clustering and object extractions demonstrate the power of this framework [71, 98]. However, the approach is restricted to 2D closed curves and the extension to higher dimensions does not seem straightforward.

Cremers *et al.* considered a B-Splines representation in [41, 40], easing the introduction of shape knowledge in classical front evolution techniques. The modeling phase relies on B-spline representations whose control points can be used for statistical analysis like the landmark points of the active shape model. Relying on a kernel-PCA analysis in [40], a non-linear reduction of the learning space is proposed, allowing multiple modes. Besides, contrary to active shape models, this representation can be used in a classical contour evolution approach. A variational formulation introduces the shape constraint in a B-spline-based Mumford-Shah functional. Nevertheless, B-splines still do not allow different topologies and the (re)parameterization of the contour remains a critical step.

The level-set representation has no such limitations: being implicit, different topologies can be handled and its intrinsic definition removes the constraining parameterization and re-parameterization issues. The first work to consider such representation to include prior shape knowledge is the one of Chen *et al.* in [33]. Distance functions are used to represent training shapes and the model is characterized by a mean distance function obtained from an averaging of the training distance functions after alignment. We proposed a direct extension of this method in [125], the same representation is considered but a stochastic model stands as a shape model by also considering the pixel-wise variance within the level-set space. A new shape module is defined to constrain classical curve evolution techniques. The final variational formulation integrates a prior shape module, jointly with the usual data terms. The stochastic model defines a region selective prior: the data terms will govern the curve evolution in regions of low confidence in the shape prior while the prior knowledge will be highly imposed when its variance is low. Moreover, pose invariance is directly integrated into the variational formulation.

Going further in the integration of prior shape knowledge and front evolution, the active shape model can be extended to the implicit representation. A similar PCA reduction can be considered but directly on the implicit representations. Such a modeling was used in [86] and [143]. In the following, we will review the principal steps of this modeling which is quite similar to the parametric case. The second step, the active integration of the model in object extractions, can be done in several ways. In [143], the authors considered the segmenting curve directly in the PCA subspace by expressing the segmenting level-set with the mean level-set of the model, plus a weighted combi-

nation of the principal modes of variation. Then, the optimization process is no more a curve evolution process but rather consists in finding the optimal mode weights. In [86], a variational framework integrates two terms. One is data-driven while the other one attracts the evolving contour to an optimal shape which is updated according to an other optimization process accounting for visual and prior shape knowledge.

We propose a more natural and adaptive way of introducing the shape constraint. We define a new energetic module that will attract the evolving contour to the specific class. The variational formulation permits to select actively a shape from the learning class that best approximates the evolving level-set. Contrary to [143], the evolving curve is not constrained to the learning space but only attracted to it. This allows for local variations and it is less subject to local minima inherent to the geometry of the learning space. Compared to [86] where a mixed approach alternates two optimization problems, our approach makes a clear separation between data-driven and shape prior terms.

This chapter is organized as follows: first, we present an approach to align the contours/surfaces of the training set since registration is a required step before shape modeling. The algorithm is also based on the level-set representation and it considers rigid transformations with possible additional scale factor. The second part is dedicated to the construction of two shape models: (i) a stochastic shape model and (ii), an active shape model. Then, the introduction of shape constraints within front evolution schemes is addressed and the efficiency and the flexibility of our approach are demonstrated through its integration with various data terms.

4.1 Implicit shape registration/alignment

Shape registration is a required step before any shape analysis. Shape alignment and registration is a complex issue in vision, graphics and medical imaging [10, 153]. The registration problem consists in matching a shape $\mathcal{C}$ to a target shape $\mathcal{C}_{\mathcal{T}}$ allowing a certain type of deformations/transformations. To solve this problem, a dissimilarity measure between shapes need to be defined. Then, an optimization procedure will try to minimize the dissimilarity between the transformed shape $\hat{\mathcal{C}}$ and the target shape $\mathcal{C}_{\mathcal{T}}$. This dissimilarity can be defined either along the contour (*shape-based*) or in the entire region (*area-based*) determined by the contour. Beside the dissimilarity, the nature of transformation is also a key component in registration. Global motion models or local deformations can be considered. In our particular problem, a global transformation is preferred since local deformations between shapes must be preserved to be captured during the modeling phase. Rigid (translation and rotation), similarity (translation, rotation and scale) and affine transformations are general global motion models. In the following, shape alignment is performed on distance transforms on the shapes considering rigid ans similarity transformations.

4.1.1 The level-set representation for global registration

We consider the general problem of aligning closed shapes of arbitrary dimension and topology. The registration is restricted to the transformation space of similarity transforms which are composed of a rigid transformation and a scale factor. The objective is to align a shape $\mathcal{C}$ to a target shape $\mathcal{C}_{\mathcal{T}}$. Distance transforms ϕ and $\phi_{\mathcal{T}}$ are used as implicit representations of the two shape according to the following definition:

$$\phi(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{C} \\ \mathcal{D}(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \mathcal{C}_{in} \\ -\mathcal{D}(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \mathcal{C}_{out} \end{cases} \tag{4.1}$$

where $\mathcal{D}(\mathbf{x}, \mathcal{C})$ refers to the minimum Euclidean distance between $\mathbf{x}$ and $\mathcal{C}$, while $\mathcal{C}_{in}$ and $\mathcal{C}_{out}$ are the image sub-domains delimited by $\mathcal{C}$ (the definition of the target level-set $\phi_{\mathcal{T}}$ is similar). In practice, the fast marching algorithm [132] or PDE-based techniques [142] can be used for the construction of these representations. We refer the reader to the first chapter of this thesis where the choice of the algorithm has been discussed. This representation defines a feature space with several nice properties for front evolutions: shapes of any dimension and topology can be considered, it is intrinsic and parameter free and it defines an optimization framework where the calculus of variations can be employed.

Furthermore, we can show easily that this representation is invariant to rigid transformations and that it can integrate easily a scale factor to deal with more general transformations. But first, we consider a rigid transformation $\mathcal{A}$ that can be decomposed in a rotation $\mathbf{R}$ and a translation $\mathbf{T}$: $\mathcal{A}(\mathbf{x}) = \mathbf{R}\,\mathbf{x} + \mathbf{T}$. Let ϕ' be the function obtained after the transformation of ϕ by $\mathcal{A}$. The zero-crossing of ϕ' gives a shape $\mathcal{C}'$ which is nothing but the shape $\mathcal{C}$ being transformed by $\mathcal{A}$. We want to show that ϕ' is also the distance transform of $\mathcal{C}'$. Let $\mathbf{x}'$ be the image of $\mathbf{x}$ by the transformation $\mathbf{T}$, for all $\mathbf{x}$ in the image domain Ω, we have: $\phi'(x') = \phi(x) = \mathcal{D}(\mathbf{x}, \mathcal{C})$. Then, we have to show that the following relation holds:

$$\forall \mathbf{x} \in \Omega, \mathbf{x}' = \mathbf{R}\,\mathbf{x} + \mathbf{T} \Rightarrow \mathcal{D}(\mathbf{x}', \mathcal{C}') = \mathcal{D}(\mathbf{x}, \mathcal{C}).$$

This equality is easily obtained from the definition of $\mathcal{D}$:

$$\begin{aligned} \mathcal{D}(\mathbf{x}', \mathcal{C}') &= \min_{\mathbf{y}' \in \mathcal{C}'} \left\{ \|x' - y'\|_2 \right\} \\ &= \min_{\mathbf{y} \in \mathcal{C}} \left\{ \|\mathbf{R}\,x + \mathbf{T} - (\mathbf{R}\,y + \mathbf{T})\|_2 \right\} \\ &= \min_{\mathbf{y} \in \mathcal{C}} \left\{ \|\mathbf{R}\,(x - y)\|_2 \right\} \\ &= \mathcal{D}(\mathbf{x}, \mathcal{C}) \end{aligned}$$

We can now also deduce the effect of adding a scale factor in the transformation: $\mathcal{A}(\mathbf{x}) = s\,\mathbf{R}\,\mathbf{x} + \mathbf{T}$. Using the same relations as above, the transformed level-set

appears to be equal to a scaled distance function:

$$\begin{aligned}\mathcal{D}(\mathbf{x}', \mathcal{C}') &= \min_{\mathbf{y}\in\mathcal{C}} \{\|s\,\mathbf{R}\,(x-y)\|_2\} \\ &= s\,\mathcal{D}(\mathbf{x}, \mathcal{C})\end{aligned}$$

which leads to $\phi'(x') = \mathcal{D}(\mathbf{x}, \mathcal{C}) = \frac{1}{s}\mathcal{D}(\mathbf{x}', \mathcal{C}')$. Therefore, ϕ' is simply multiplied by s to represent the distance transform of $\mathcal{C}'$.

Dealing with more general global deformations like affine transformations is not so straightforward. In general, the distance transform of the transformed shape is not available and it has to be recomputed. When considering iterative optimizations as the one of the following part, this reinitialization brings high computational costs.

4.1.2 Global registration

Registration is equivalent with finding a point-wise transformation between the current shape $\mathcal{C}$ and the target shape $\mathcal{C}_\mathcal{T}$ which minimizes a given dissimilarity measure. Distance transforms refer to a higher dimension space than the original one and increase the problem/solution potentials. One can seek for a transformation $\mathcal{A}$ that creates pixel-wise intensity correspondences between the source representation ϕ and the target representation $\phi_\mathcal{T}$. We have proved that distance transforms representations $[\phi; \phi_\mathcal{T}]$ are invariant to translation and rotation and we are able to predict the effect of scale variations. These conditions lead to the following constraint:

$$\begin{cases} (s, \mathbf{R}, \mathbf{T}) \\ \mathcal{A}(\mathbf{x}) = s\,\mathbf{R}\,\mathbf{x} + \mathbf{T} \\ \underline{\forall \mathbf{x} \in \Omega \,:\, [s\,\phi_\mathcal{T}(\mathbf{x}) = \phi(\mathcal{A}(\mathbf{x}))]} \end{cases}$$

With this constraint, we consider registration in a global optimization framework that involves all pixels in the image plane. The sum of squared differences, the optimization of the correlation ratio, the maximization of the mutual information, etc. can be used as as similarity measure between the source and the target representation:

$$E(s, \theta, \mathbf{T}) = \int_\Omega \rho(s\,\phi_\mathcal{T}(\mathbf{x}), \phi(\mathcal{A}(\mathbf{x})))\,d\mathbf{x}. \tag{4.2}$$

where ρ is the dissimilarity measure and θ the rotation angle: $\mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

A gradient descent can be used to recover the optimal registration parameters. Let ρ_1 and ρ_2 be respectively the derivatives of ρ with respect to its first and second

argument, we can write the gradient descent for a general dissimilarity measure:

$$\begin{cases} \dfrac{d}{dt}s = 2\displaystyle\int_\Omega \big(\rho_1(,)\phi_{\mathcal{T}}(\mathbf{x}) + \rho_2(,)\nabla\phi\cdot\mathbf{Rx}\big)\,d\mathbf{x} \\ \dfrac{d}{dt}\theta_i = 2\displaystyle\int_\Omega \rho_2(,)\nabla\phi\cdot\nabla_{\theta_i}(\mathcal{A}(\mathbf{x}))\,d\mathbf{x}, \quad 1\le i\le p \\ \dfrac{d}{dt}\mathbf{T} = 2\displaystyle\int_\Omega \rho_2(,)\nabla\phi\,d\mathbf{x} \end{cases} \tag{4.3}$$

where p is the number of rotation angles (1 in 2D and 3 in 3D).

In order to introduce and demonstrate the potential of the implicit shape representation and our method, we consider a simple well known and widely used criterion; the sum of squared differences:

$$\forall \mathbf{x}\in\Omega, \quad \rho(s\,\phi_{\mathcal{T}}(\mathbf{x}), \phi(\mathcal{A}(\mathbf{x}))) = (s\,\phi_{\mathcal{T}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x})))^2.$$

The corresponding gradient descent is obtained by replacing ρ_1 and ρ_2 in (4.3) according to:

$$\begin{cases} \rho_1(,) = 2(s\,\phi_{\mathcal{T}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x}))), \\ \rho_2(,) = -2(s\,\phi_{\mathcal{T}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x}))). \end{cases}$$

Domain restriction

The initial positions of the source $\mathcal{C}$ and the target $\mathcal{C}_{\mathcal{T}}$ can produce distance map representations that are not equally defined in a fixed image plane. To deal with this technical limitation, as well to decrease computational costs, we may restrict the integral to some parts of the image close to the contours. To this end, a region selective function $f()$ can be designed from pixel-values of the distance representations. After normalization, the following type of dissimilarity measure is obtained:

$$\rho(\phi_1,\phi_2) = \frac{f(\phi_1,\phi_2)(\phi_1-\phi_2)^2}{\int_\Omega f(\phi_1,\phi_2)\,d\mathbf{x}}.$$

Several choices have been proposed in the literature [111, 42, 32]. As proposed in [42], the comparison can be restricted to the inside parts of the contours, taking $f(\phi_1,\phi_2) = H_\epsilon(\phi_1) + H_\epsilon(\phi_2)$. In [111], the distance maps are used to consider the areas defined by two equal-distance contours (inwards, outwards) from the input shapes.

$$f(\phi_1,\phi_2) = N_\alpha(\phi_1,\phi_2) = \begin{cases} 0, \min(|\phi_1|,|\phi_2|) > \alpha \\ 1, \min(|\phi_1|,|\phi_2|) \le \alpha \end{cases}$$

with the interpretation that only pixels (isophotes) within a range of distance α from the shapes to be registered are considered in the optimization process. When $\alpha\to 0$, we obtain a simple geometric interpretation since our functional simplifies to:

$$E(\mathcal{A}) = \frac{1}{|\mathcal{C}|}\int_{\mathcal{C}}\left(\min_{\mathbf{y}\in\mathcal{C}_{\mathcal{T}}}\{\|\mathbf{y}-\mathcal{A}(\mathbf{x})\|_2\}\right)^2 d\mathbf{x} + \frac{1}{|\mathcal{C}_{\mathcal{T}}|}\int_{\mathcal{C}_{\mathcal{T}}}\left(\min_{\mathbf{y}\in\mathcal{C}}\{\|\mathbf{x}-\mathcal{A}(\mathbf{y})\|_2\}\right)^2 d\mathbf{x},$$

which can be related to [33, 60].

	ρ_a	ρ_b	ρ_c
$\rho(\phi_1, \phi_2)$	$N_\alpha(\phi_1, \phi_2)(\phi_1 - \phi_2)^2$	$(H(\phi_1) + H(\phi_2))(\phi_1 - \phi_2)^2$	$(H(\phi_1) - H(\phi_2))^2$
Meaning	band ϕ-diff.	inside ϕ-diff.	symmetric diff.

Figure 4.1: Several examples of dissimilarity measures

In [32], a purely area based functional was proposed. The level-set representations are used to express the symmetric difference between shapes: $\rho_($$\phi_1, \phi_2) = (H(\phi_1) - H(\phi_2))^2$. All these dissimilarities have different properties and the choice of a pertinent one is not obvious since it may also depend on the type of shape. The absence of local minima and the robustness to particular perturbing factors are some desirable properties one will seek. Despite the high dimension of the parameter space (4 in 2D), an empirical comparison between these functionals can be established on selected shapes, estimating the corresponding energy for a limited range of the transformation parameters. To this end, we can constrain the unknown parameter space in one or two dimensions. This comparison has been conducted on several 2D examples shown in [Figures 4.2 and 4.3]. For these empirical evaluations, we have quantized the search space using an uniform sampling rule (100 elements) for all unknown parameters in each case. Translation in (x, y) were in the range of $[-\frac{\text{width}}{2}, \frac{\text{width}}{2}]x[-\frac{\text{height}}{2}, \frac{\text{height}}{2}]$, scale in $[0.5, 1.5]$ and rotation in $[-\pi, \pi]$. Then, one can estimate the cost function in the space of two unknown parameters, by considering all possible combinations derived from the sampling strategy (the other two parameters are fixed).

In [Figure 4.2], a comparison between the different criteria of table 4.1 is shown where translation in x and rotation vary. Each energy plot shows a single global minima when the transformation is the identity (null translation and rotation). Several differences can be observed according to the chosen criteria. The complete evaluation of the mean square difference between the distance maps [Figure 4.2(b)] is particularly smooth and has only one minima which means that convergence will be obtained for any initialization within the considered transformation range. [Figure 4.2(c,d)] indicates the influence of the band approximation of the mean square criteria. A small band of 6 gives a flatter energy around the global minima while a choice of 50 is almost similar to the complete estimation. The evaluation of the criteria ρ_b and ρ_c are presented in [Figures 4.2(e,f)]. ρ_b shows a similar behavior to the previous ones but results in a flatter surface. The symmetric difference criteria ρ_c shows completely different properties. It is highly peaked around the global minima but many local minima are present. This shows that this criteria it quite sensitive to initial conditions. When considering mean square differences between signed distance functions [Figures 4.2(b-e)], the feature space includes more information. The empirical evaluations show a larger basin

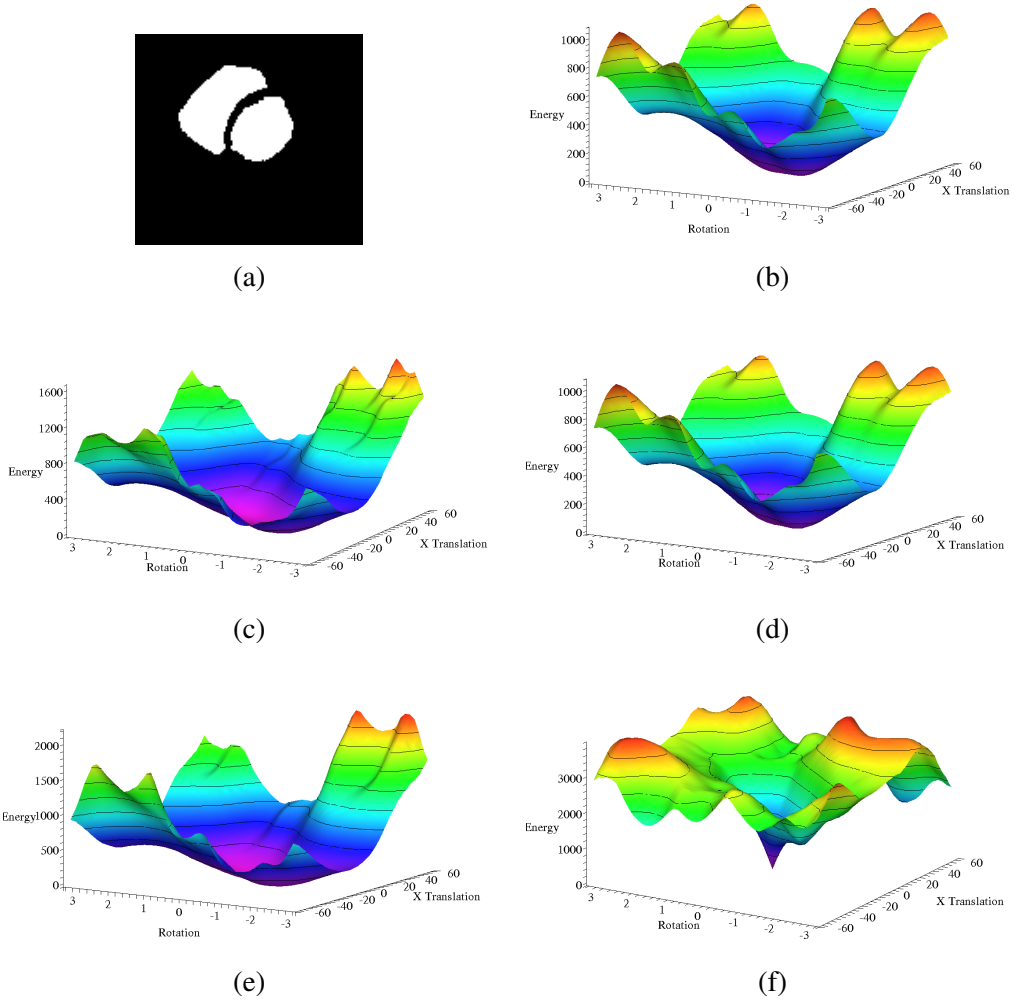


Figure 4.2: Evaluations of different cost functionals between the cardiac image (a) and its rigid transformation where translation in x and rotation vary: (b) $\rho(\phi_1, \phi_2) = (\phi_1 - \phi_2)^2$, (c) ρ_a with $\alpha = 6$, (d) ρ_a with $\alpha = 50$, (e) ρ_b, (d) ρ_c.

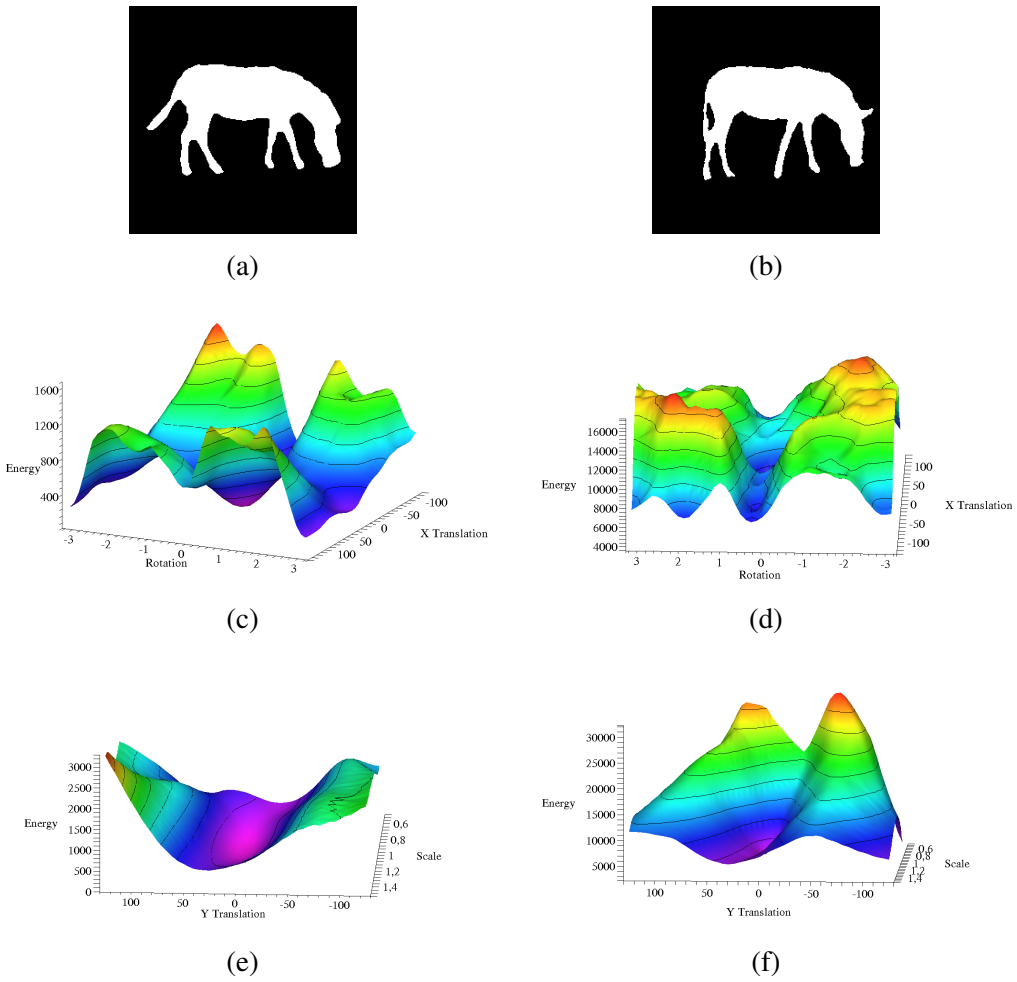


Figure 4.3: Evaluations of different cost functionals between the zebra images (a) and (b) with rigid transformations: 1) x translation and rotation vary: evaluation of ρ_a with $\alpha = 50$ in (c) and ρ_c in (d); 2) y translation and scale vary: evaluation of ρ_a with $\alpha = 50$ in (e) and ρ_c in (f).

of attraction around the minimum for this feature space, compared to the purely area based terms. Similar remarks can be made on the zebra shapes shown in [Figure 4.3] where scale variations are also considered. However, for this particularly complicated shape, several local minima are observed for far initializations.

Using either of the proposed functionals we are able to convert a geometry driven point-correspondence problem into an image-registration application where space as well feature-based (intensity) correspondences are considered. But we still need to make a choice. According to the empirical evaluations, which are only partial, ρ_a seems to give the best compromise between robustness with respect to the initial conditions and convergence speed.

Experimental results

Of course, when considering the full space of transformation parameters, the chosen criteria cannot be evaluated in the whole space, even with a coarse quantization. To solve this minimization problem, a gradient descent on each transformation parameter is considered. If we neglect the variation of the band/domain which is a reasonable approximation for small variations of the transformation parameters, the derivatives of ρ_b with respect to its first and second arguments are:

$$\begin{cases} \rho_{b,1}(,) = 2N_\alpha(s\,\phi_{\mathcal{T}}(\mathbf{x}), \phi(\mathcal{A}(\mathbf{x})))(s\,\phi_{\mathcal{T}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x}))) \\ \rho_{b,2}(,) = -2N_\alpha(s\,\phi_{\mathcal{T}}(\mathbf{x}), \phi(\mathcal{A}(\mathbf{x})))(s\,\phi_{\mathcal{T}}(\mathbf{x}) - \phi(\mathcal{A}(\mathbf{x}))) \end{cases}$$

The corresponding gradient descent is obtained by replacing $\rho_1()$ and $\rho_2()$ in the general gradient descent (4.3).

The performance of the proposed module under various initial conditions is shown in [Figures 4.4 and 4.5]. Given the rigid-invariant representation, one can claim that the method is suitable for rigid objects. Registrations of a hand shape are shown for important global transformations: a rotation of $\frac{\pi}{2}$ was applied in [Figure 4.4] while a more complex similarity transformation including translation, scaling and rotation, was considered in [Figure 4.5]. In both cases, our method is able to register the shapes. Our dissimilarity measure is also robust to small local variations as shown in [Figure 4.6] where the hand shape was registered with success to an altered and transformed version. Finally, in [Figure 4.7], we show the complete registration of a set of training contours from the SQUID dataset [1].

Even if the approach can deal with small deformations, important local variations will bias the registration process. The use of robust estimators may be considered to resolve this issue by reducing the influence of outliers. By outliers, we mean portions of the objects that are quite different in the training shapes and that should not be considered during the registration process. In the following part, we will present an original and automatic way of defining such a robust estimator that incorporates information from a shape modeling phase but before that we show an example of surface registration.

A nice advantage of the level-set representation is the possibility of being used to represent hyper-surfaces of any dimension. This is of good use in medical imaging where

[1] Center for Vision, Speech and Signal Processing Laboratory at the Univ. of Surrey, `http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html`

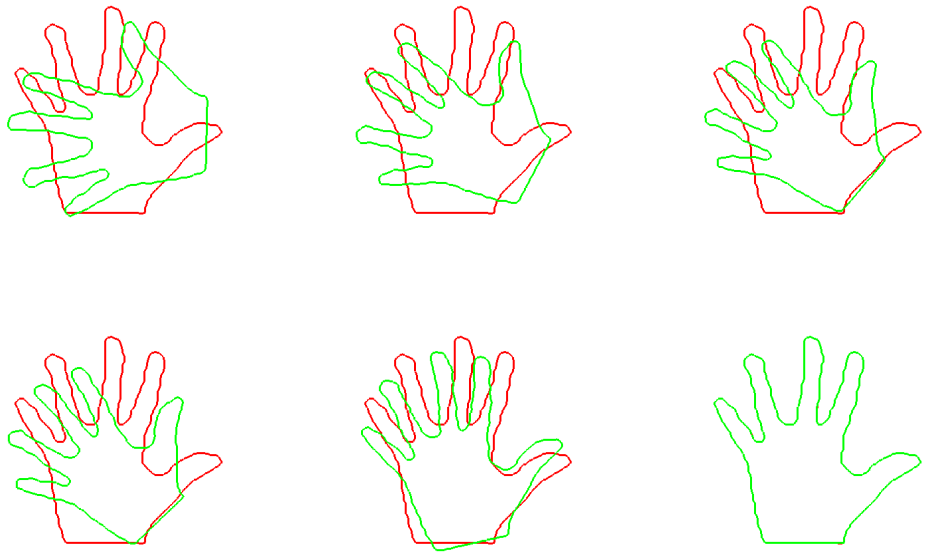


Figure 4.4: Hand registration for a synthetic rotation of $\frac{\pi}{2}$.

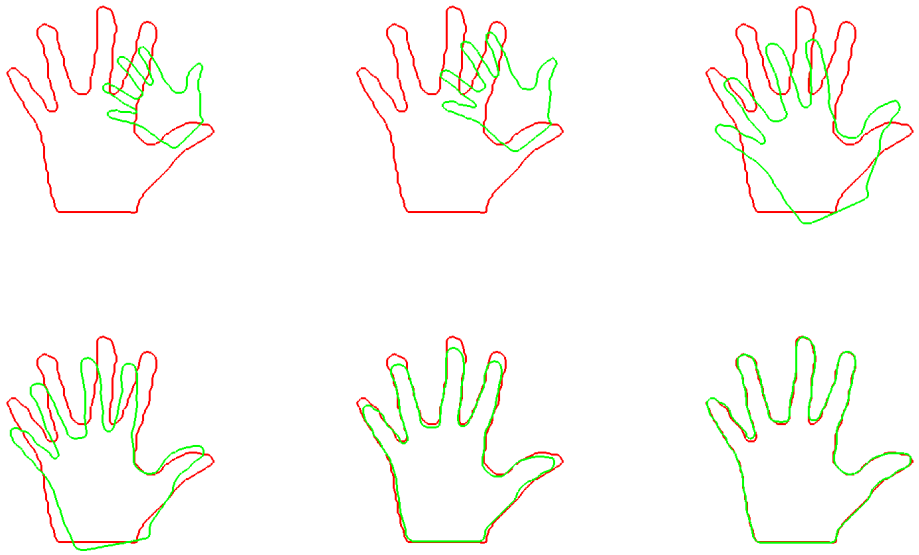


Figure 4.5: Hand registration for a synthetic similarity transformation: $\{s = 0.25, \theta = \frac{\pi}{4}, \mathbf{T} = (20, -10)\}$.

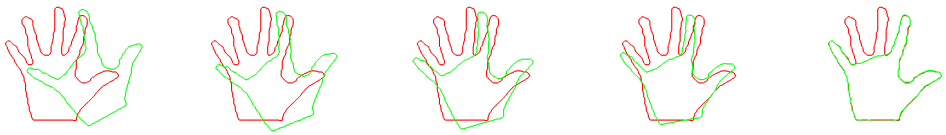


Figure 4.6: Hand registration for a synthetic similarity transformation with alteration: $\{s = 1, \theta = \frac{\pi}{6}, \mathbf{T} = (20, -10)\}$.

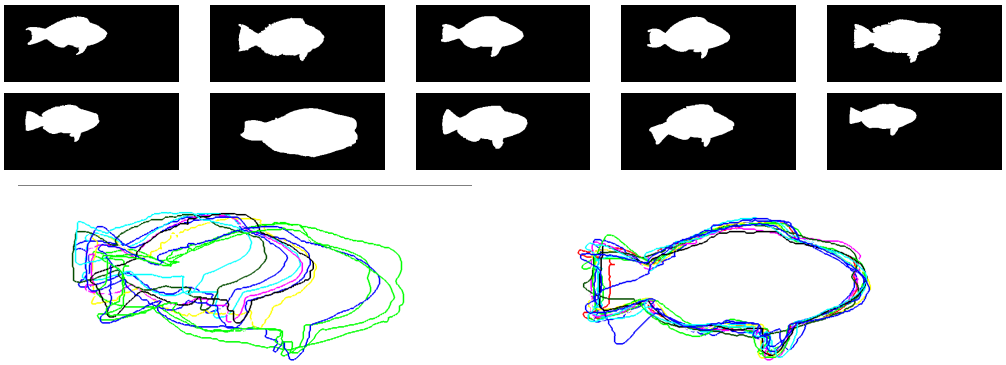


Figure 4.7: Fish registrations from the SQUID dataset - TOP: 10 of the 13 considered shapes, BOTTOM: before and after alignment.

surfaces have to be considered. In [Figure 4.8], we show an example of structures extracted from brain MR images. These surfaces represent the lateral brain ventricles. Twelve structures have been extracted semi-automatically from different patients. Of course, these surfaces have similar characteristics from one subject to another and generating a general shape model is of high interest. The registration is a required step before any modeling. Our registration approach has been applied on this training set and the obtained results are presented in [Figure 4.9]. For this 3D example, 7 pose parameters are estimated: $s, \mathbf{T}(T_x, T_y, T_z), \mathbf{R}(\theta_1, \theta_2, \theta_3)$.These results were obtained by setting the band size α to 2 and despite the higher dimension of the problem (by using implicit representations), the complexity of our registration approach is still linear in the size of the surfaces and convergence is generally obtained after less than 100 iterations. This 3D experiment is quite promising and it illustrates the power of the approach to deal with real problems.

4.2 Implicit shape modeling

Once the elements of the training set registered to a common pose, it becomes possible to analyze its properties, seeking for a meaningful compact representation that can encode prior knowledge for this particular pose.

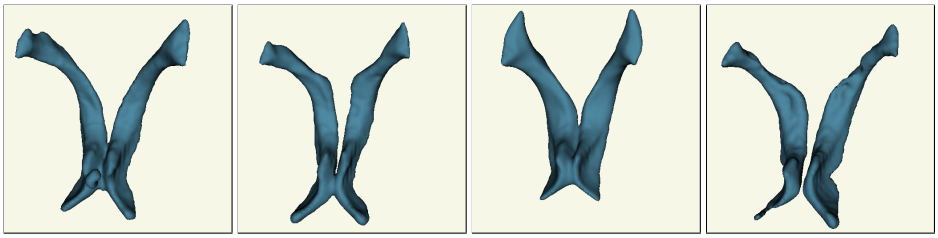


Figure 4.8: 4 of the 12 considered lateral brain ventricle surfaces.

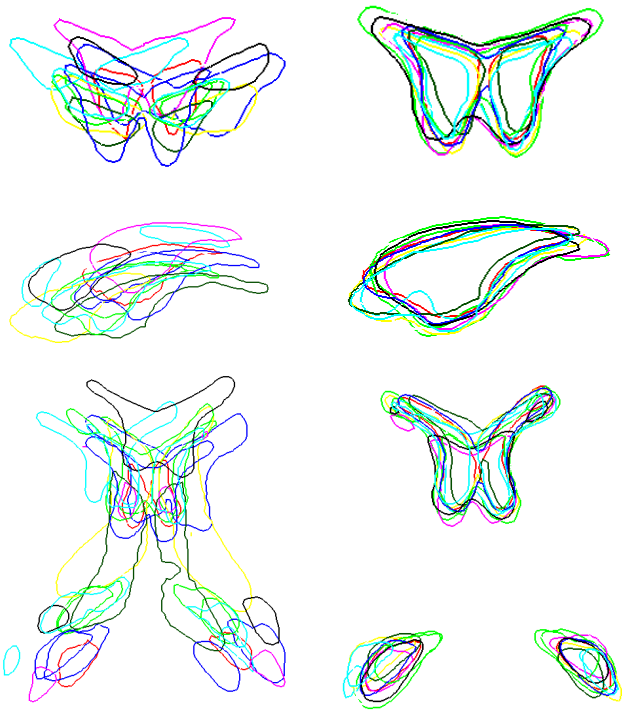


Figure 4.9: 3D registration of the lateral brain ventricles - LEFT: 2D cuts before alignment, BOTTOM: the same 2D cuts after alignment.

Introducing prior knowledge inside level-set methods requires the definition of a shape model. A cloud of points is a simple technique to represent such knowledge. However, such a technique is not convenient within a level-set framework where the evolving interface is not represented using points. A more natural selection is to consider the definition of the prior within level-set space [86, 33, 143, 125]. Consistency between the propagation technique/optimization framework and the form of the prior is meaningful. In other words, the objective is to recover a compact representation to encode the prior from a set of level-set examples $[\phi_1, \phi_2, \ldots, \phi_N]$. Building an average shape across the examples of the training set can be sufficient enough to represent such a prior [33] but such a technique cannot capture variability. However, local/pixel-wise variability as well as principal shape variations can be characterized to capture properties of the training set. We present two different shape models which capture respectively local variability and principal shape variations.

The first model combines the simple structure of average shape and the ability to capture the local variability of the learning set [125]. Such a model should consist of two components. The most prominent shape as well as the confidence along the shape parts. When agreement between the training examples for a particular part is present, the confidence should be high and the recovery of the object in the image should strongly respect the prior. On the other hand, when this is not the case, the prior constraint should be relaxed and the image information should be more important. Such a model will furnish a region selective shape constraint for the extraction of new structures.

The second model intends to capture properties of more complex family of shapes, not necessary close to a mean shape. When a significant number of samples is available in the training set, one can think of capturing principal shape variations. Principle component analysis (PCA) can be applied to capture the statistics of the corresponding elements across the training examples [86, 143]. PCA refers to a linear transformation of variables that retains - for a given number of operators - the largest amount of variation within the training data.

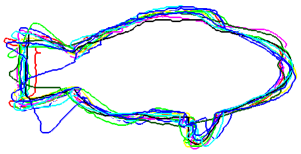
In the following, we present these two shape models and we show 2D and 3D modeling examples obtained for various shape sets. The technical question of reinitializing the mean shape representation to a distance function is also addressed.

4.2.1 Capturing local variability : a stochastic shape model

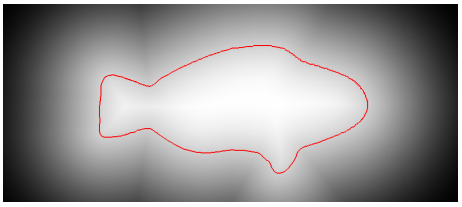
A stochastic framework with two unknown variables can be considered:

- The mean shape image, $\phi_s(\mathbf{x})$,
- The local degrees (variability) of shape deformations $\sigma_s(\mathbf{x})$.

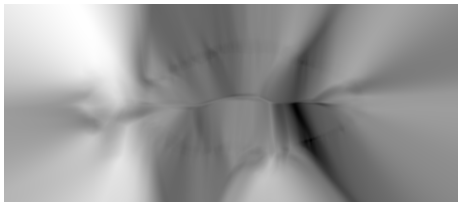
Similar models were proposed in a different context [156]. This model refers to a distance function $[\phi_s(\mathbf{x})]$ that is associated with some variability measurements



(a)



(b)



(c)

Figure 4.10: Stochastic shape model for the *Fish dataset*: (a) aligned training set, (b) model mean, (c) model confidence map.

$[\sigma_s(\mathbf{x})]$. Then, for a given pixel location $\mathbf{x}$ and a given value ϕ, the conditional probability of having this value at this location is given by:

$$p^s_{\mathbf{x}}(\phi) = \frac{1}{\sqrt{2\pi}\sigma_s(\mathbf{x})} e^{-\frac{(\phi-\phi_s(\mathbf{x}))^2}{2\sigma^2_s(\mathbf{x})}}. \tag{4.4}$$

The construction of such a model can be done through a training phase. One can assume that N contours/shapes of the target are available. Global registration of the training examples to a common pose provides N distance transforms representations (one for each training sample) $[\hat{\phi}_i]$. A sample - arbitrarily selected - shape from the training set can be used as common pose $[\phi_s]$. Such selection can introduce some bias to the model construction.

We consider a variational framework for the estimation of the most appropriate representation that aims at maximizing the local joint densities between the model (ϕ_s, σ_s) and the registered training samples:

$$\begin{gathered} E(\phi_s, \sigma_s) = -\sum_{i=1}^{n} \int_\Omega \log\left[p^s_{\mathbf{x}}(\hat{\phi}_i(\mathbf{x}))\right] d\mathbf{x}, \\ \text{subject to the constraint : } |\nabla\phi_s(\mathbf{x})|^2 = 1, \quad \forall \mathbf{x} \in \Omega. \end{gathered} \tag{4.5}$$

Additionally, it is natural to enforce spatial coherence on the variability estimates by adding a smoothness term. Since the constant term $(\sqrt{2\pi})$ does not affect the mini-

mization procedure, the following functional is used:

$$\begin{aligned} E(\phi_s, \sigma_s) =& \alpha \int_\Omega |\nabla \sigma_s(\mathbf{x})|^2 \, d\mathbf{x} \\ &+ (1-\alpha) \sum_{i=1}^{n} \int_\Omega \left(\log[\sigma_s(\mathbf{x})] + \frac{(\hat{\phi}_i(\mathbf{x}) - \phi_s(\mathbf{x}))^2}{2\sigma_s^2(\mathbf{x})} \right) d\mathbf{x}, \\ &\text{subject to the constraint : } |\nabla \phi_s(\mathbf{x})|^2 = 1, \quad \forall \mathbf{x} \in \Omega, \end{aligned} \tag{4.6}$$

where $[\alpha]$ balances the contribution between the data attraction and regularity of the variability field.

The constrained optimization of this functional can be done using Lagrange multipliers and a gradient descent method. However, given the form of the constraint (involvement of first and second order derivatives), we cannot obtain a closed form solution and prove that the conditions which guarantee the validity of Lagrange's theorem are satisfied. Moreover, the number of unknown variables of the system is too high and the system is quite unstable especially when a large variability among the training samples is present. An alternative selection refers to a two-step optimization process. During the first step, we obtain the "optimal" solution according to the data-driven terms, while during the second step we find the "optimal" projection of this solution on the manifold of acceptable solutions (distance functions). Thus, the unknown variables are obtained by minimizing the previously defined data-driven objective function that preserves some regularity conditions:

$$\begin{cases} \dfrac{d}{dt}\phi_s = (1-\alpha) \displaystyle\sum_{i=1}^{n} \frac{\hat{\phi}_i - \phi_s}{\sigma_s^2}, \\ \dfrac{d}{dt}\sigma_s = (1-\alpha) \displaystyle\sum_{i=1}^{n} \left[-\frac{1}{\sigma_s} + \frac{(\hat{\phi}_i - \phi_s)^2}{\sigma_s^3} \right] + 2\alpha\Delta\sigma_s, \end{cases} \tag{4.7}$$

while the projection/correction to the manifold space of accepted solutions (Euclidean distance maps) is done using a partial differential equation that does not require average shape extraction and was proposed in [142]. These two steps alternate until the system reaches a steady-state solution. Upon convergence of the system, we will obtain a distance map/transform representation model that optimally expresses the properties of the training set using degrees of variability constrained to be locally smooth. As far as the initial conditions of the system are concerned, we use the Euclidean distance map of a reference sample.

Several modeling examples are shown in [Figures 4.10, 4.11 and 4.12]. All these examples are well-adapted to our stochastic model since in each case the shapes are very close to each other and differ only on local parts. The stochastic shape model is able to retrieve a meaningful mean representation while it captures shape parts with high and low variability. For the *Fish* training set of [Figure 4.10], the mean level-set represents an "average" smooth fish without any particular detail and the confidence

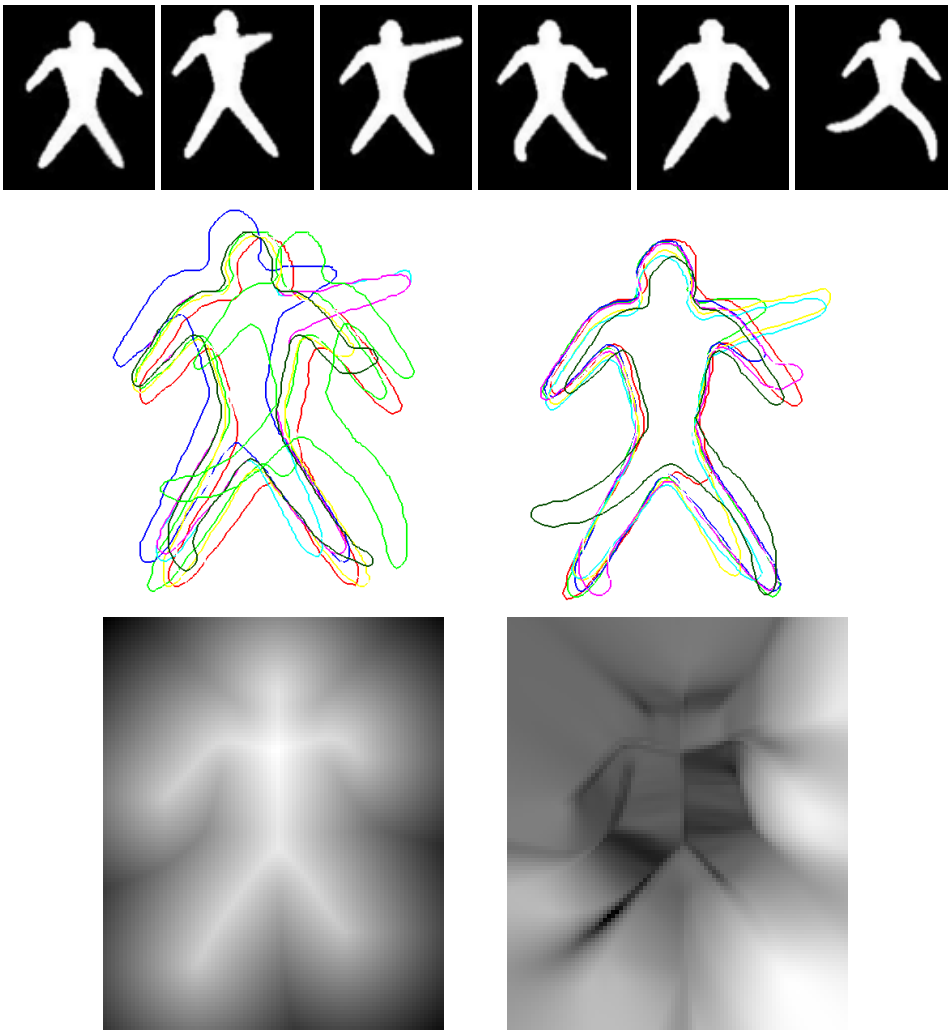
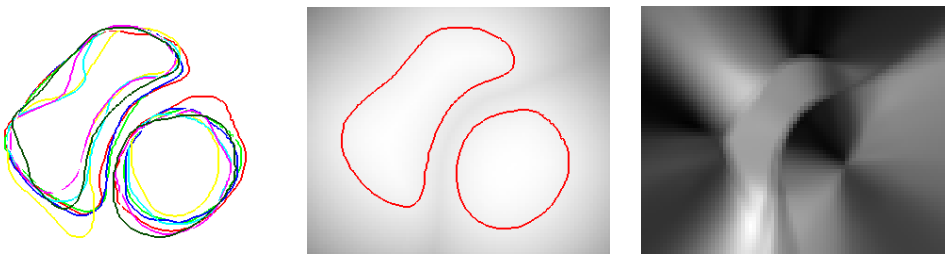
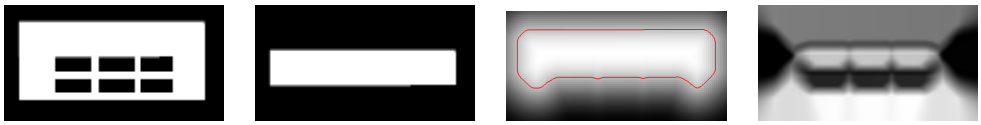


Figure 4.11: Stochastic shape model for the *Dude dataset* - TOP: Samples of the training set, BOTTOM: from left to right : dude shapes before alignment, after alignment, model mean and model confidence map.



(a) (b) (c)

Figure 4.12: Stochastic shape model for the *Cardiac Ventricles dataset*: (a) aligned training set, (b) model mean, (c) model confidence map.



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

Figure 4.13: Registration of highly different shapes - (a) and (b) synthetic shapes, (c) model mean, (d) model confidence.

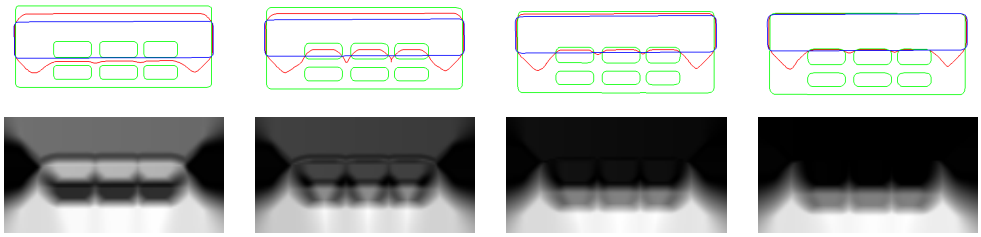


Figure 4.14: Alternate stochastic modeling/registration - TOP: Registration results at different iterations (in green and blue are the two shapes while the model mean is in red), BOTTOM: Corresponding model confidence map.

map shows a high variability around the tail. The *Dude* example in [Figure 4.11] is a synthetic example where the same shape was considered but with some modifications on local parts like the legs or the arms. We obtain a model in accordance with these local modifications since high variance is recovered in the parts which have endured some changes. The last example in [Figure 4.12] is taken from cardiac images. The left and right heart ventricles were previously extracted along a cardiac cycle. The shape model is able to capture the parts of the ventricles with important changes during the cardiac cycle. This can be pertinent to analyze whether a patient has breathing problems or not since a high variance will imply an important blood transfer.

The advantages of such a prior model are numerous. It encodes in a natural form prior knowledge within implicit representations, it provides straightforward techniques for the estimation of geometric properties, it can deal with multi-component objects and it can be determined from a small set of training examples. Moreover, the registration process can take benefits from this model.

Stochastic registration

This confidence map may also help to improve the registration phase of the training set by reducing the influence of variable parts of the objects and increasing the weight of highly similar parts during the registration process. Given this model, one can register a given shape $\mathcal{C}$ by maximizing the joint density between the model and the input shape representation ϕ. If we assume that the conditional densities of the model are indepen-

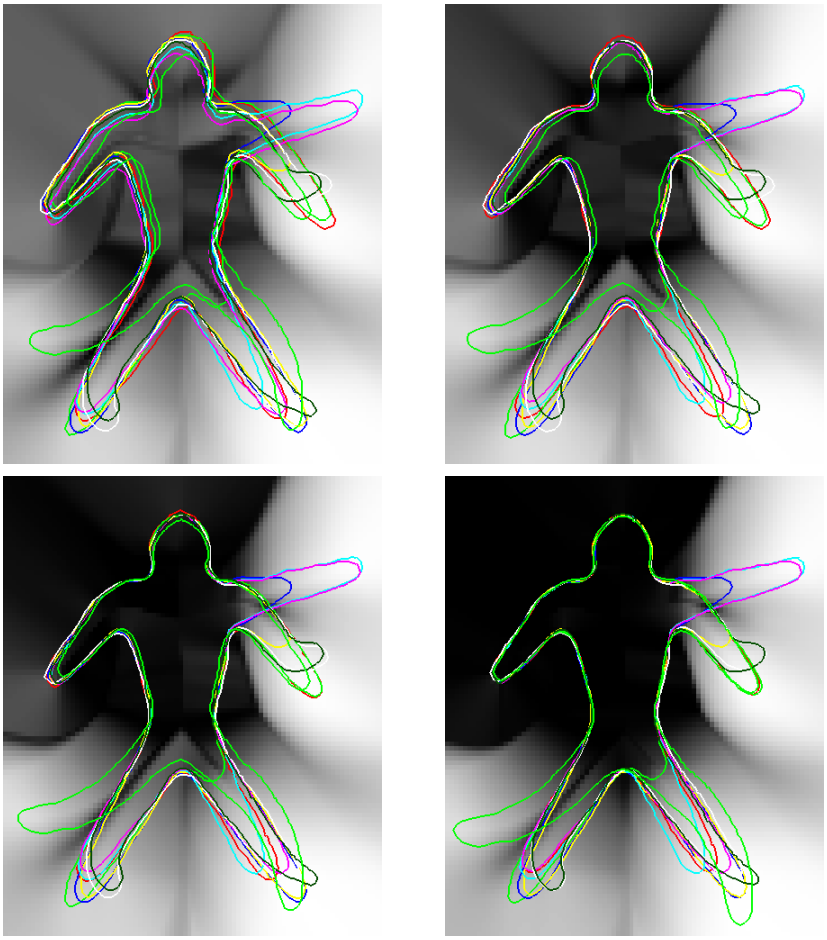


Figure 4.15: Alternate stochastic modeling/registration of the *Dude* dataset - Registration results at different iterations with superimposed confidence map.

dent across pixels, then the optimization criterion is equivalent to the maximization of

$$p(s,\theta,\mathbf{T}|\phi,\phi_s,\sigma_s) = \prod_{\Omega} p^s_{\mathcal{A}(\mathbf{x})}(s\,\phi).$$

As in the previous section, we consider the distance maps within a limited range of their zero crossings, leading to the following minimization criterion:

$$E(s,\theta,\mathbf{T}) = \int_{\Omega} N_\alpha(s\,\phi,\phi_s)\left(\log(\sigma_s(\mathcal{A}(\mathbf{x}))) + \frac{(s\,\phi(\mathbf{x}) - \phi_s(\mathcal{A}(\mathbf{x})))^2}{2\sigma_s(\mathcal{A}(\mathbf{x}))}\right) d\mathbf{x}.$$

A gradient descent can be used for the minimization of this functional. The resulting motion equations are similar to the ones obtained with the sum of squared differences but it also includes the variability of the shape model. As a consequence, the contribution of pixel locations with high variability are less significant than the ones with high confidence in the prior model. Both processes, the stochastic modeling and the stochastic registration, can be iterated, improving one another until stability is achieved.

In the experiments, this alternate registration/modeling approach converges toward a solution that appears to be the optimal stochastic registration and modeling.

The capabilities of this iterative approach are demonstrated on a synthetic example in [Figures 4.13 and 4.14]. In this example, the two shapes are highly different but they share a common part. The different parts bias the aligning process in [Figure 4.13]. Then, if we build the stochastic model, we get a high variance in these different parts and it is employed to improve the alignment. [Figure 4.14] shows the improvements of the alignment when both steps are iterated. The same approach is used to improve the registration of the *Dude* shapes [Figure 4.11]. The results are presented in [Figure 4.15] and our approach is able to capture regions of high confidence (head and left upper part) to improve the alignment. The final registration is close to perfect.

4.2.2 Capturing principal variations through principal component analysis

When a large training set is available, a more complex model can be built by seeking to model the shape sub-space corresponding to the training set. In a given application, the training shapes can be assumed to belong to a restricted family of shapes. Analyzing the training set, we wish to model this family. For this purpose, we consider the modeling approach introduced in [86, 143]. This model assumes the shape family to be generated from a linear combination of implicit representations. A principal component analysis is applied on the training shapes to retain only the principal modes of variation. First, the samples $\hat{\phi}_i$ are centered with respect to a mean representation ϕ_s, $[\psi_i = \hat{\phi}_i - \phi_s]$. Then, the most important modes of variation can be recovered through principal component analysis on the covariance matrix Σ defined as:

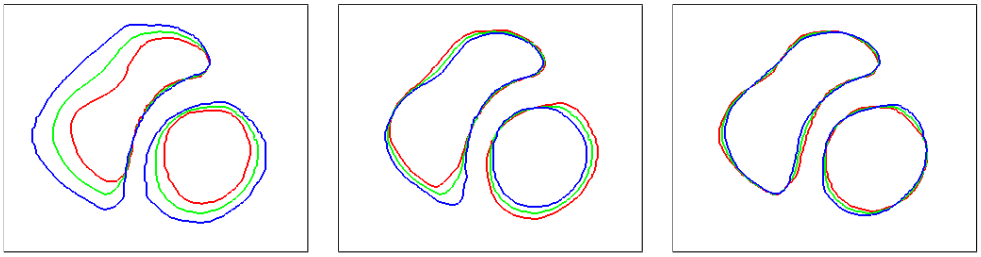
$$\Sigma_{i,j} = \frac{1}{|\Omega|} \int_{\Omega} \psi_i(\mathbf{x}) \psi_j(\mathbf{x}) \, d\mathbf{x}, \quad 1 \leq i \leq N, 1 \leq j \leq N.$$

Retaining the m principal modes of variation, a new shape belonging to the modeled family can be expressed as:

$$\phi = \phi_s + \sum_{j=1}^{m} \lambda_j \, U_j, \tag{4.8}$$

where U_j are the retained modes of variation (eigenvectors of Σ) and λ_j are weighting factors taking values within the allowable range defined by the eigenvalues.

In [Figure 4.16], we show the first three modes of variation obtained for the cardiac ventricles. The quantity of information r_j described by the mode of variation j is given directly by the corresponding eigen value: $r_j = \frac{\lambda_j}{\sum_{i=1}^{N} \lambda_i}$. Regarding the cardiac example of [Figure 4.16], the first mode represents a large majority of the shape variations, accounting for 81.7% of the information. Actually, this first mode captures most of the deformation of the ventricles during a cardiac cycle. The second mode cannot be neglected since its contribution is up to 13.5%. It incorporates more local



81.7% 13.5% 2.7%

Figure 4.16: Shape model of the cardiac ventricles - The most important modes of variation and their respective contributions are presented [principal three modes after rigid alignment (green: mean, blue: $+\sigma$, red:$-\sigma$)].

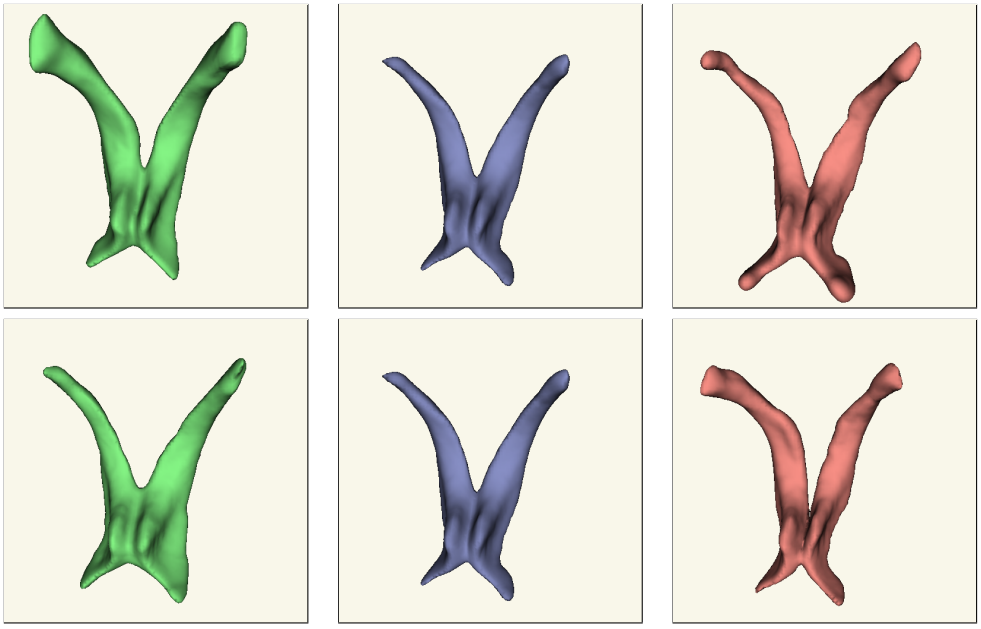


Figure 4.17: Shape model of the lateral brain ventricles model - The most important modes of variation are presented [principal two modes after rigid alignment (blue:mean, red: $+\sigma$, green:$-\sigma$)].

deformations of the ventricles. In this example, it appears that only two modes may be sufficient to describe the shape as it represents more than 95% of the information. Besides its use for constraining the extraction of new structures, this shape analysis may be a good tool to characterize possible abnormalities during the cardiac cycle.

Another example of such an analysis is shown in [Figure 4.17] for the 3D modeling of lateral brain ventricles. We consider the training set used in the last section for 3D rigid alignment [Figure 4.8]. The space reduction is conducted on the aligned shapes. [Figure 4.17] shows the mean level-set as well as the surfaces obtained from each of the first modes of variation, varying the weight of the corresponding eigen vectors. From only 12 surfaces, we are able to generate a whole family of brain ventricles whose shapes seem quite possible and natural. One interesting point is the compactness of the obtained representation since more than 70% of the information is contained in the first two modes, while the third one represents 10% and seems to account for non-symmetrical parts of the ventricles which can be observed in some samples of the training set. Moreover, contrary to most shape analysis techniques, the alignment and the modeling phases are entirely automatic and introduce no tuning parameter. The user only need to specify the percentage of information the model should capture and the number of retained modes will be set in consequence.

4.2.3 When should the mean representation be maintained to a distance function?

One can claim that constraining the mean level-set representation to a distance function is not necessary, that a simple average of the level-set representations of the training samples would be sufficient to give a mean representation. Let us consider a simple example composed of the two shapes of [Figure 4.18]. A simple average of the two implicit representations gives the image [Figure 4.19(a)] while constraining it to a distance function gives [Figure 4.19(b)]. The simple average gives more information since the legs of the horses can still be guessed while the other one completely removes this information. Therefore, a simple average seems to be a better approximation for the mean representation. However, in this case the information that is included in this mean representation is not clear anymore since it becomes something different from an implicit representation of a contour. Now, if we consider the full stochastic shape model, including the estimation of a confidence map jointly with the mean representation (constrained to a distance function), we obtain the model of [Figures 4.19(c,d)]. With this model, we include all the information corresponding to the legs and we have a clear interpretation. The mean level-set is the implicit representation of the mean contour while the confidence map gives a pixel-wise confidence in this representation. [Figures 4.19(c,d)] shows that the legs are included as areas of low confidence in the mean representations.

Regarding the second model, a simple average should be preferred. The reason

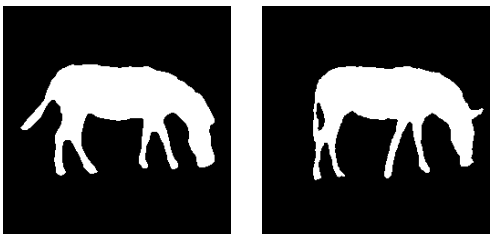
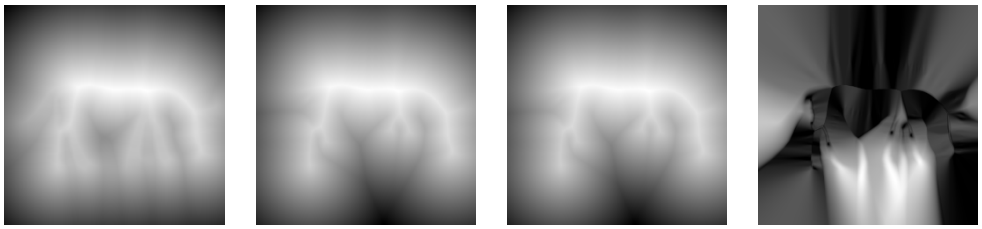


Figure 4.18: *Zebras* test shapes (see text).



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

Figure 4.19: *Zebras* shapes analysis - (a) simple average, (b) mean constrained to a distance function, (c) stochastic shape model mean, (c) stochastic shape model confidence map.

relies in the linear assumption of the model. A non-linear estimation of the mean would be in contradiction with the whole modeling process and training samples could not be recovered using the relation 4.8. According to what have been said before, it appears natural to investigate a non-linear extension of this principal component analysis in future works.

These two modeling approaches capture important information within the training sets which can be used as prior knowledge during the extraction of new structures. The introduction of this prior shape knowledge is addressed in the following section.

4.3 Shape constraints in object extraction

The next step is the introduction of the prior shape knowledge within a front evolution by defining the probability of a given partition of the image $p(\mathcal{P}(\Omega))$ when a shape model is available. Restricting us to the extraction of a single object (not necessary connex) and representing the evolving front by a level-set function ϕ, this is equivalent to defining the prior distribution $p(\phi|\text{“shape model”})$.

As the same representation is used for the modeling phase and the object extraction, the definition of the prior probability of a given shape is quite natural. Considering a Bayesian formulation, we first show how we can define a natural criteria that will constrain the evolving level-set, including dynamic estimation of pose parameters.

Then, we successively develop shape constraint modules for: (i) a simple static prior composed of one prior shape, (ii) the stochastic shape model introduced in Section 4.2.1 and (iii), the PCA-based model presented in Section 4.2.2. For this last model, the prior shape will be actively selected from the modeled family by optimizing with respect to the weights of the principal modes of variation. Finally, promising results will show the ability of our approach to tackle difficult problems including highly noisy data or objects with missing parts and occlusions.

4.3.1 Shape constrained surface evolution

Let us now consider an evolving interface represented by a level-set function ϕ as described in Chapters 2 & 3. We would like to evolve it while respecting some shape properties $\phi_{\mathcal{M}}$ modulo a given transformation $\mathcal{A}$. One can model the joint space of evolving interface and its transformation to the prior model using a probability density function. Let $[p(\phi, \mathcal{A}|\phi_{\mathcal{M}})]$ be the prior distribution of the transformation given the model $\phi_{\mathcal{M}}$. Such distribution is unknown, varies across different objects and cannot be recovered in the more general case. However, Monte-Carlo sampling or other techniques can be used to recover such distribution when empirical evidence is sufficient enough. In the absence of such knowledge, we can consider a Bayesian formulation for this density:

$$p(\phi, \mathcal{A}|\phi_{\mathcal{M}}) = \frac{p(\phi_{\mathcal{M}}|\phi, \mathcal{A})}{p(\phi_{\mathcal{M}})} p(\phi, \mathcal{A}) = \frac{p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)}{p(\phi_{\mathcal{M}})} p(\phi, \mathcal{A}).$$

The constant term $p(\phi_{\mathcal{M}})$ can be ignored and the joint space of interfaces and their transformations is assumed to be uniform. Such an assumption is often considered in the absence of knowledge on the properties of the object to be recovered. Then, recovering the optimal interface and the transformation is equivalent with finding the maximum posterior $p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)$ that is equivalent with finding the extremum of:

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x} \in \Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|\phi(\mathbf{x})),$$

where $\mathbf{x}$ is an image location and independence across pixels was assumed. The last thing to be accounted for is the conversion of the evolving interface ϕ to a similar pose to the one of the prior model $\phi_{\mathcal{M}}$ (rigid transformations are often considered for pose estimation). Including rotation, translation and scaling, we consider a similarity transformation: $\mathcal{A} = (\mathcal{T}, \theta, s)$. As explained in the registration part, scale variations cause predictive changes in the distance transform level-set representations. Therefore, the scale factor s is to be accounted for, leading to the following form for the posterior:

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x} \in \Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})).$$

The pixel-defined prior distributions $[p_{\mathbf{x}}()]$ are known from the modeling phase and solving the inference problem is equivalent with finding the lowest potential of the

$-\log$ function, or:

$$E(\phi, \mathcal{A}) = -\log\left[\prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x}))\right] = -\int_{\Omega} \log(p_{\mathbf{x}}(s\,\phi(\mathbf{x})))\,d\mathbf{x}.$$

Static prior knowledge

Considering the simplest shape model, i.e. a single prior shape $\phi_{\mathcal{M}}$, we would like to define the probability of the evolving level-set ϕ, modulo a given transformation $\mathcal{A}$. A pixel-wise Gaussian distribution can be considered:

$$p_{\mathbf{x}}(s\,\phi(\mathbf{x})) = p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(s\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2}{2\sigma^2}\right),$$

where σ is global confidence in the prior shape $\phi_{\mathcal{M}}$. Then, the evolving level-set ϕ and the transformation parameters should minimize the following energy:

$$E(\phi, \mathcal{A}) = \frac{1}{2\sigma^2}\int_{\Omega} (s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2\,d\mathbf{x}. \tag{4.9}$$

The constant factor in σ can be omitted or replaced by a weighting factor when this term is used as a module in a complete energy including data-based terms. Restricting the integral in the vicinity of the zero-level of ϕ, the following analytical expression for the objective function is obtained:

$$E(\phi, \mathcal{A}) = \int_{\Omega} \delta_{\epsilon}(\phi)(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2\,d\mathbf{x}, \tag{4.10}$$

where δ_{ϵ} is the usual smoothed version of the Dirac function. This energy is close to the dissimilarity measure used in Section 4.1.2 to register two level-sets ϕ_1 and ϕ_2:

$$\rho(\phi_1, \phi_2) = N_{\alpha}(\phi_1, \phi_2)(\phi_1 - \phi_2)^2,$$

which gives for small values of α: $\rho(\phi_1, \phi_2) = \delta_{\epsilon}(\phi_1)(\phi_1 - \phi_2)^2 + \delta_{\epsilon}(\phi_2)(\phi_1 - \phi_2)^2$. The geometric interpretation of this dissimilarity measure was discussed in the registration Section 4.1.2 and only the first term is sufficient to attract ϕ_1 toward ϕ_2 which corresponds to the energy (4.10).

In order to minimize this energy with respect to the evolving level-set representation and the global linear transformation, we use the calculus of variations. The current representation will evolve toward $\phi_{\mathcal{M}}$ modulo the rigid transformation $\mathcal{A}$. The equation of evolution for ϕ is given by:

$$\frac{\partial\phi}{\partial t} = -2\delta_{\epsilon}(\phi)s\left(s\,\phi - \phi_{\mathcal{M}}(\mathcal{A})\right) - \delta'_{\epsilon}(\phi)\left(s\,\phi - \phi_{\mathcal{M}}(\mathcal{A})\right)^2$$

The second term is a deflation force and does not affect the position of the zero level of ϕ, thus it can be neglected.

The rigid transformation $\mathcal{A}$ is also dynamically updated so as to map ϕ and $\phi_{\mathcal{M}}$ the best. The gradients of this functional with respect to the parameters of $\mathcal{A}$ drives to a system similar to the one obtained for the registration (4.3) replacing respectively $\rho_1(,)$ and $\rho_2(,)$ by:

$$\begin{cases} \rho_1(,) = 2\delta_\epsilon(\phi)(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))) \\ \rho_2(,) = -2\delta_\epsilon(\phi)(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))) \end{cases}$$

Stochastic prior knowledge

Rather than considering a single prior shape, we can consider the Gaussian model of Section 4.2.1 which gives us not only a mean shape $\phi_{\mathcal{M}}$ but also a pixel-wise confidence map $\sigma_{\mathcal{M}}$. The prior distribution of an evolving level-set ϕ can be expressed as:

$$p_{\mathbf{x}}(s\,\phi(\mathbf{x})) = p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})) = \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{M}}(\mathbf{x})} \exp\left(-\frac{(s\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2}{2\sigma_{\mathcal{M}}(\mathbf{x})^2}\right).$$

Still restricting the integral around the zero-level of ϕ, the corresponding energy is:

$$E(\phi, \mathcal{A}) = \int_\Omega \delta_\epsilon(\phi(\mathbf{x})) \left(\log(\sigma_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))) + \frac{(s\,\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2}{s\sigma_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))}\right) d\mathbf{x}, \quad (4.11)$$

where constant terms have been omitted. This objective function consists of two terms. The first one discourages the recovery of a transformation $\mathcal{A}$ that projects the evolving interface to model areas with low confidence (large $\sigma_{\mathcal{M}}(\mathcal{A})$). The second term couples local propagation and estimation with the following objectives: (i) to recover a transformation that aligns the evolving interface with the prior and (ii), to evolve the interface so it becomes close to the prior. Such terms have similar conceptual interpretation with the one used to introduce the static prior but it is also able to integrate the model confidence map. The projection error $(s\,\phi - \phi_{\mathcal{M}}(\mathcal{A}))^2$ is weighted according to the model confidence $\sigma_{\mathcal{M}}(\mathcal{A})$.

Within such an optimization framework, the error for deviating from the model in areas with low confidence (high $\sigma_{\mathcal{M}}(\mathcal{A})$) is downgraded (according to $\sigma_{\mathcal{M}}(\mathcal{A})$). Consequently, such areas become less important in the process of imposing the prior and recovering the transformation. Last, but not least, such a model deals in an implicit manner with outliers in the process of shape enforcement within the segmentation process.

Active shape model

During the model construction, we have also analyzed the principal modes of variation within the training set. Including this information, the ideal transformation will map each value of current representation to the "best" level set representation belonging to

the class of the training shapes. As mentioned in Section 4.2.2, if a shape representation $\phi_{\mathcal{M}}$ belongs to this class, then it can be derived from the m principal modes:

$$\phi_{\mathcal{M}} = \phi_s + \sum_{j=1}^{m} \lambda_j \, U_j.$$

The shape model $\mathcal{M}$ is composed of the mean representation ϕ_s and the principal modes of variation $\{U_j, j = 1..m\}$. We would like to estimate $[p(\phi, \mathcal{A}, \lambda|\mathcal{M})]$, the prior distribution of the transformation and mode weights given $\mathcal{M}$. A Bayesian formulation of this density gives

$$p(\phi, \mathcal{A}, \lambda|\mathcal{M}) = \frac{p(\mathcal{M}|\phi, \mathcal{A}, \lambda)}{p(\mathcal{M})} p(\phi, \mathcal{A}, \lambda) = \frac{p(\phi_{\mathcal{M}}(\mathcal{A})|\phi)}{p(\mathcal{M})} p(\phi, \mathcal{A}, \lambda).$$

The constant term $p(\mathcal{M})$ can be ignored and the joint distribution $p(\phi, \mathcal{A}, \lambda)$ is assumed to be uniform. Then, the optimal interface, transformation and mode weights are obtained by maximizing

$$p(\phi_{\mathcal{M}}(\mathcal{A})|\phi) = \prod_{\mathbf{x}\in\Omega} p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})).$$

A more complex model could be easily deduced from the modeling phase by learning the marginal distribution $p(\lambda)$. In the absence of such a model, we keep the assumption of a uniform distribution. Identically to the static prior, a pixel-wise Gaussian model centered in $\phi_{\mathcal{M}}$ is considered as the prior density estimate:

$$p_{\mathbf{x}}(s\,\phi(\mathbf{x})) = p(\phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))|s\phi(\mathbf{x})) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(s\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x})))^2}{2\sigma^2}\right).$$

Again, we apply the $-\log$ function to this density, and, restricting the integral around the zero-crossing of ϕ, the optimal parameters must minimize the new energy:

$$E(\phi, \mathcal{A}, \lambda) = \int_{\Omega} \delta_\epsilon(\phi) \left(s\,\phi(\mathbf{x}) - \left(\phi_s(\mathcal{A}(\mathbf{x})) + \sum_{j=1}^{m} \lambda_j \, U_j(\mathcal{A}(\mathbf{x}))\right)\right)^2 d\mathbf{x}.$$

The evolution equation of ϕ and the gradient descent of the transformation parameters are similar to the ones obtained with the static and stochastic shape models but this energy should also be minimized with respect to the mode weights λ. Actually, these new parameters are the easiest ones to estimate since their optimal values are solutions of the linear system

$$\bar{U}\lambda = b,$$

with

$$\begin{cases} \bar{U}(i,j) = \displaystyle\int_{\Omega} \delta_\epsilon(\phi) U_i(\mathcal{A}(\mathbf{x})) U_j(\mathcal{A}(\mathbf{x}))\, d\mathbf{x} \\ b(i) = \displaystyle\int_{\Omega} \delta_\epsilon(\phi)(s\phi(\mathbf{x}) - \phi_{\mathcal{M}}(\mathcal{A}(\mathbf{x}))) U_i(\mathcal{A}(\mathbf{x}))\, d\mathbf{x}, \end{cases}$$

where $\bar{U}$ is a $m \times m$ positive definite matrix and can be easily inverted.

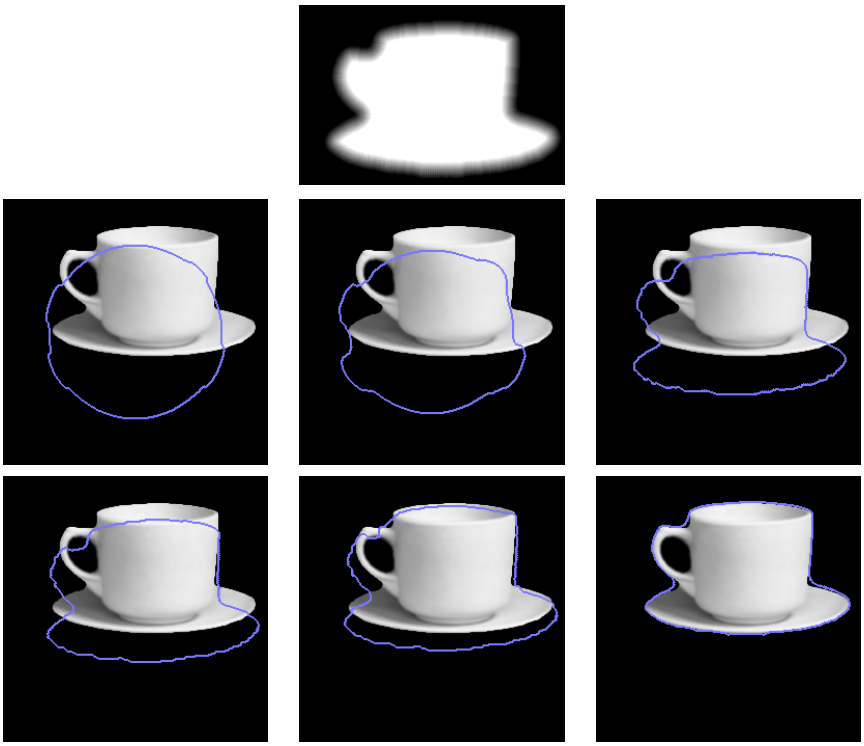


Figure 4.20: Constrained segmentation with a single static prior - TOP: Cup shape prior, BOTTOM: Constrained curve evolution.

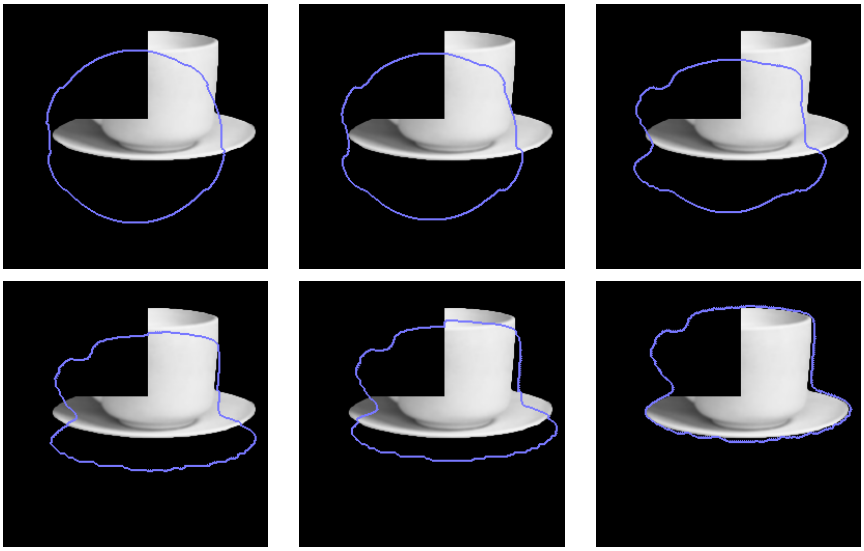


Figure 4.21: Constrained segmentation with a single static prior: Curve evolution for an object with missing parts.

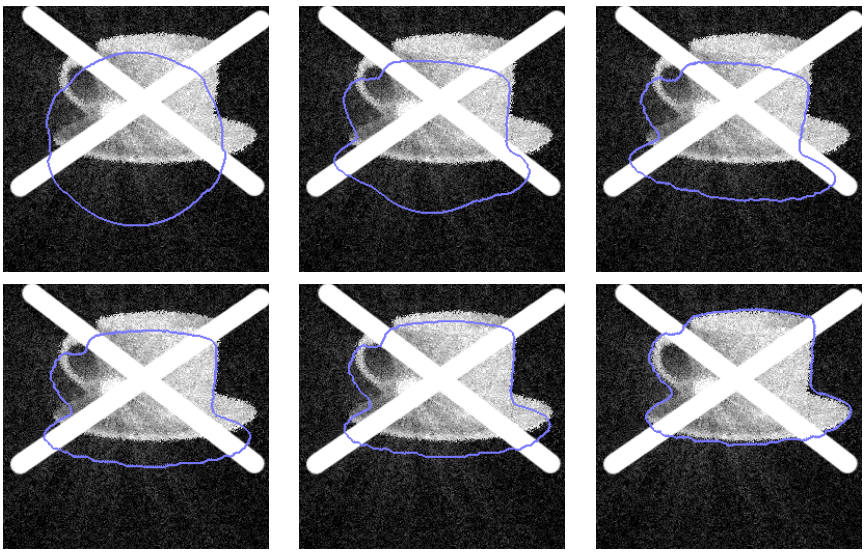


Figure 4.22: Constrained segmentation with a single static prior: Curve evolution for an object with occlusions.

4.3.2 Constrained object extraction

In this section, we integrate the three different shape models of the last section in the Bayesian segmentation method presented in Chapter 2. This permits us to constrain the object extraction from new images with three different types of prior shape knowledge:

- a single prior shape,
- the stochastic shape model of Section 4.2.1,
- the PCA-based model of Section 4.2.2.

Two different constrained object extractions are proposed. The first one assumes the prior shape constraint to be constant during the extraction process. It includes the single shape prior and the stochastic shape model. The second one is a form of an active shape model. It selects actively the "best" shape from the prior family while attracting the evolving level-set toward this new prior. Promising results are presented for both types of constrained extraction on 2D and 3D images.

Object extraction subject to a static shape constraint

Since prior shape knowledge can be introduced in a level-set evolution, this information can be used jointly with any of the data terms of Section 2 and 3. The data term will make the contour evolve toward the object of interest while the shape prior will maintain a global shape in accordance with the shape model. For this purpose, we can integrate our shape constraint in the Bayesian formulation of Chapter 2. In the case

of the extraction of a single object given by a level-set representation ϕ, the Bayesian formulation of Chapter 2 is:

$$p(\phi|I) \propto p(I|\phi)p(\phi).$$

While the first term of this expression has been extensively studied in Chapter 2 and 3, the second one has been only used as a regularization constraint. Now, we can define a much more complex constraint by integrating prior shape models. The corresponding variational formulation will be composed of two terms: one accounting for data information and one constraining the shape of the evolving level-set with respect to the shape model. In the case of a static shape prior, the following type of functional has to be minimized:

$$E(\phi, \mathcal{A}) = b\, E_{shape}(\phi, \mathcal{A}) + (1 - b)\, E_{data}(\phi).$$

where E_{shape} is the shape attraction term, E_{data} accounts for data information and the weight b permits to control the influence of the shape prior. By static prior shape knowledge, we mean a prior term that remains identical during the segmentation process. Only the rigid/similarity transformation that maps the evolving level-set to the prior will be updated. Such a prior can be a single shape or the stochastic shape model of Section 4.2.1 and Section 4.3.1.

First, we show some experimental results with a simple prior composed of a single shape. The corresponding energy is the one proposed in equation (4.10). In [Figures 4.20, 4.21 and 4.22], we show several experiments where the outline of a cup is used as shape prior. For the sake of simplicity, the transformation $\mathcal{A}$ was restricted to a simple translation in this experiment. In the first example ([Figure 4.20]), the complete cup is present in the image and the contour evolves toward the shape prior in a first before being translated to the object. [Figure 4.21] shows the robustness of the method after removing a large part of the object while in [Figure 4.21], the object was occluded and the method was still able to recover the object. The second set of experiments is more general since $\mathcal{A}$ is a similarity transformation and color information is included in the data term. The shape prior is a hand and two examples are shown. The first one ([Figure 4.23]) shows how the approach is able to recover the complete hand subject to translation, rotation and scaling. To increase the difficulty, noise and occlusion were also added in [Figure 4.24]. The shape constraint makes the curve capture the full hand.

Going further, we consider the stochastic shape model as prior. Including a pixel-wise confidence map, the model furnishes a region-selective constraint. In regions of low confidence in the shape model, the data term will drive the curve evolution while the shape term will be highly imposed in regions of high confidence. A first synthetic example illustrates the region selectivity of the prior. Two similar shapes are used to build a stochastic shape model in [Figure 4.25]. The confidence map separates

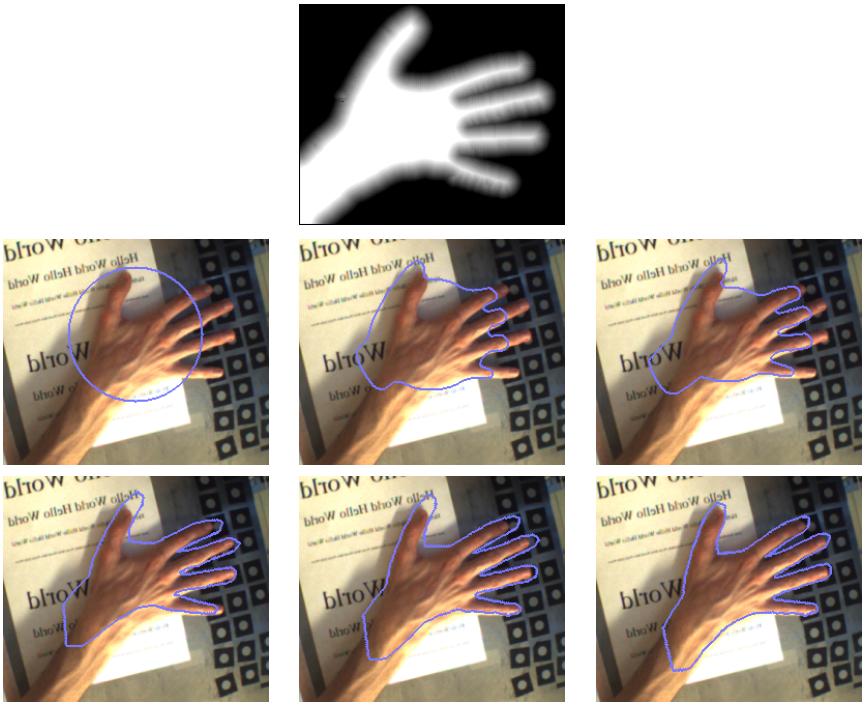


Figure 4.23: Constrained segmentation with a single static prior - TOP: Hand shape prior, BOTTOM: Constrained curve evolution.

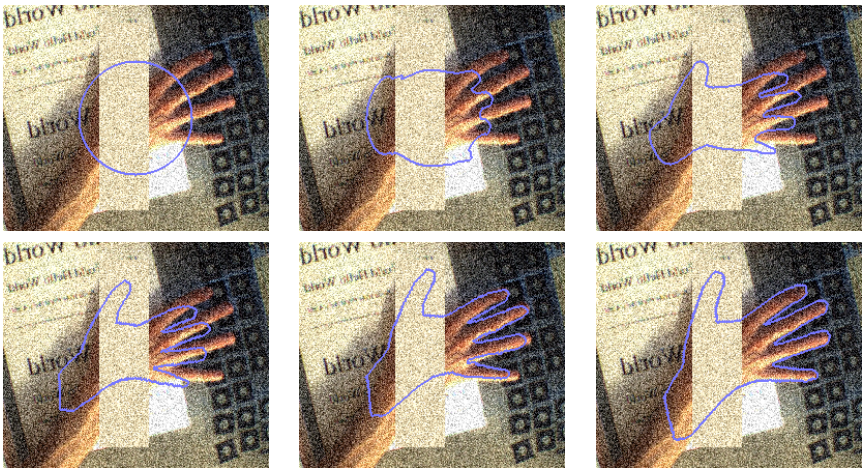


Figure 4.24: Constrained segmentation with a single static prior: Curve evolution for an object with occlusions and noise.

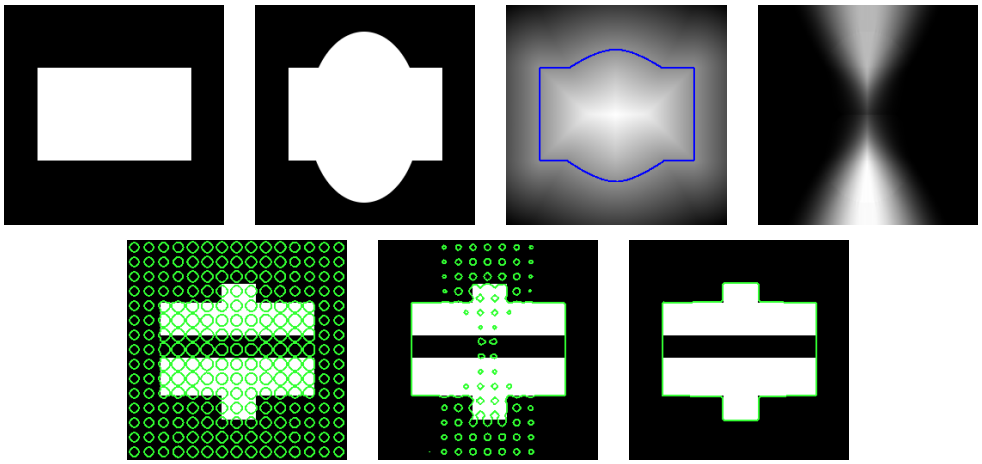


Figure 4.25: Constrained segmentation with the stochastic shape prior - TOP: Training contours and the corresponding stochastic shape model, BOTTOM: Constrained curve evolution on a synthetic example.

clearly the image in two parts, one of high confidence and one of low confidence. The constrained segmentation in [Figure 4.25] shows a very explicit example of this model. The shape prior constrains the curve on the left and right parts while the data term is favored in the top and bottom parts. This model is used on real medical data in [Figure 4.26]. The segmentation of cardiac ventricles is correct for the whole cardiac cycle. The use of prior shape knowledge makes the approach insensitive to dark parts inside the ventricles because the confidence in the prior is quite important in these regions. On the other hand, the data term will drive the curve evolution in parts of the ventricles with high variability.

Object extraction subject to an active shape constraint

As we have show in Section 4.3.1, the active shape model can be introduced dynamically in the segmentation process by optimizing w.r.t. the mode weights in parallel to the level-set evolution:

$$E(\phi, \mathcal{A}, \lambda) = b\, E_{shape}(\phi, \mathcal{A}, \lambda) + (1 - b)\, E_{data}(\phi).$$

This framework has been tested on the extraction of the lateral brain ventricles. [Figure (4.27)] show the robustness to noise brought by the prior shape knowledge (the image is one of the training images but with additional Riccian noise). In [Figure (4.28)], we show the ability of our approach to extract objects from new images (not used for building the model). The active shape model is able to approximate the surface with a similar one from the modeled class while the object extraction allows for small local variations with respect to the model. Finally, in [Figure (4.29)], we show the

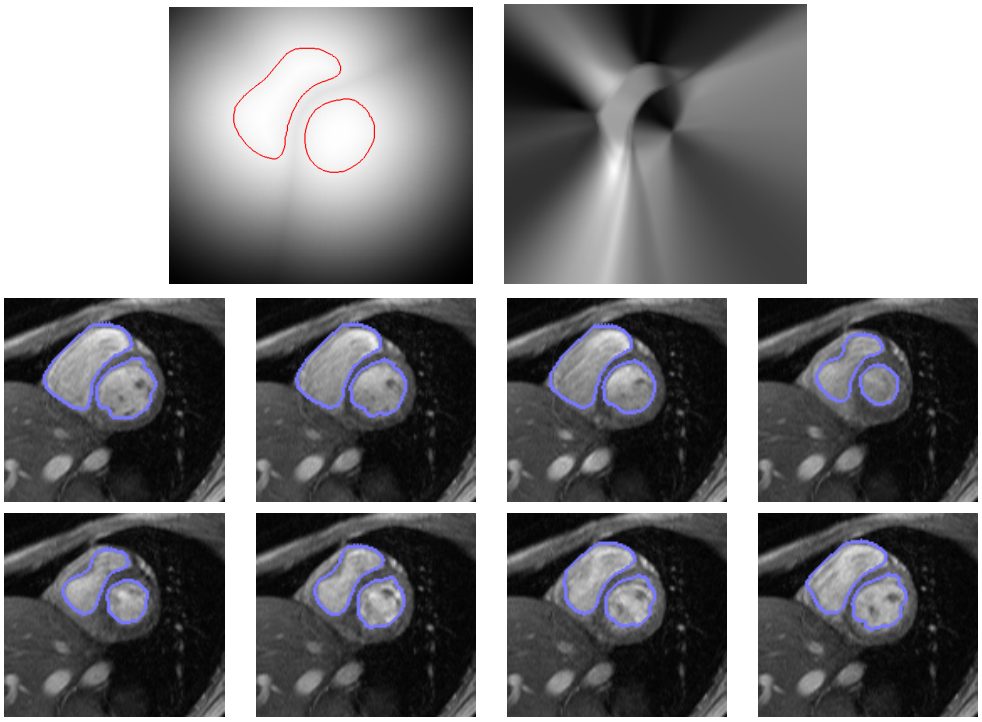


Figure 4.26: Constrained segmentation with the stochastic shape prior: Curve evolution for an object with occlusions and noise.

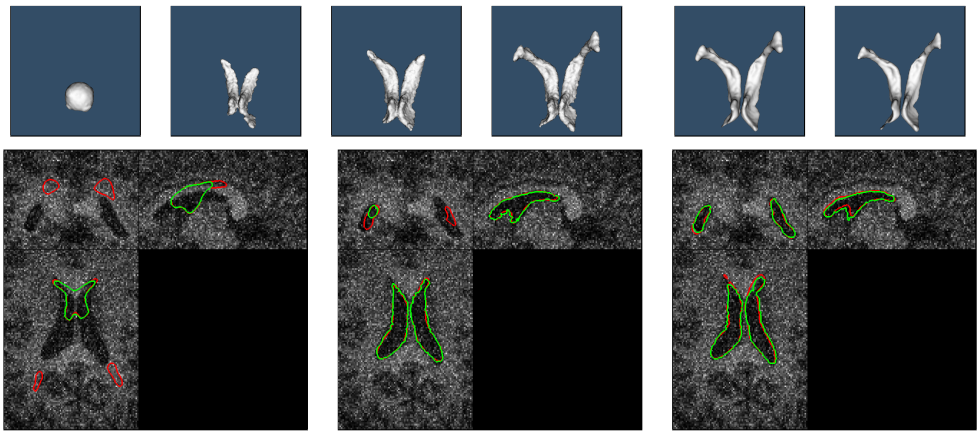


Figure 4.27: Segmentation of the lateral brain ventricles with Shape Prior ($b = 0.3$) of a noisy MR image. TOP LEFT: surface evolution, TOP RIGHT: projected surface in the learning space and ground-truth surface (from the training set), BOTTOM: surface cut (green) and its projection in the learning space (red) during surface evolution.

influence of the shape prior by changing its weight. While prior knowledge improves the quality of the object extraction, overweighting it will make object details to be missed. The possibility of tuning this parameter is an important advantage of our approach compared to [143].

4.4 Contributions and conclusion

In this chapter, we have presented a detailed study of prior shape knowledge in image segmentation. Two phases have been considered: (i) the shape modeling from a set of training samples, (ii) the shape constrained object extraction. Both steps make use of level-set functions as implicit representations of contours or surfaces. Several original aspects should be emphasized:

1. Two different modeling approaches have been proposed, both in the level-set space. The first one consists in a pixel/voxel-wise stochastic model which assumes the training shapes to be close to a mean representation and approximates pixel/voxel-wise distributions of the level-set representations by a Gaussian density. This model permits local variability/confidence within the training set to be captured with respect to the mean shape. The second model estimates the principal modes of variation through a principal component analysis over the training samples.

2. Considering the Bayesian formulation introduced in Chapter 2, we define a new term to account for prior shape model during the segmentation process. This new constraint also comes from a Bayesian inference and it is presented for three

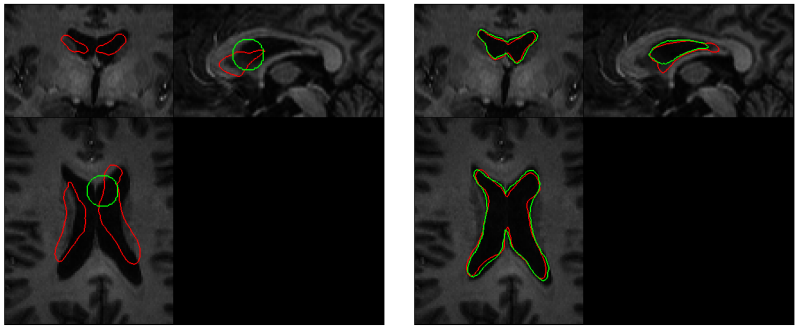


Figure 4.28: Segmentation of the lateral brain ventricles with Shape Prior ($b = 0.3$) in an MR image which was not used during the learning phase: surface cut (green) and its projection in the learning space (red) at initialization (LEFT) and after convergence (RIGHT).

different cases. The first one integrates a simple static shape prior composed of a single shape. The second model naturally includes the whole stochastic shape model, defining a region-selective prior constraint thanks to the learnt confidence map. Finally, the last model is actively integrated by introducing the modes of the principal modes of variation as additional unknown parameters in a variational formulation.

3. Several examples show the introduction of each of these shape prior modules with different data-terms. In addition to 2D images, several examples deal with the extraction of surfaces from 3D medical images. The use of implicit representations makes this possible.

Related publications can be found in [125, 110, 109, 111, 126].

Acknowledgments: I would like to thank Nikos Paragios who has been not only a perfect guide for this part but also very active in formalizing different ideas.

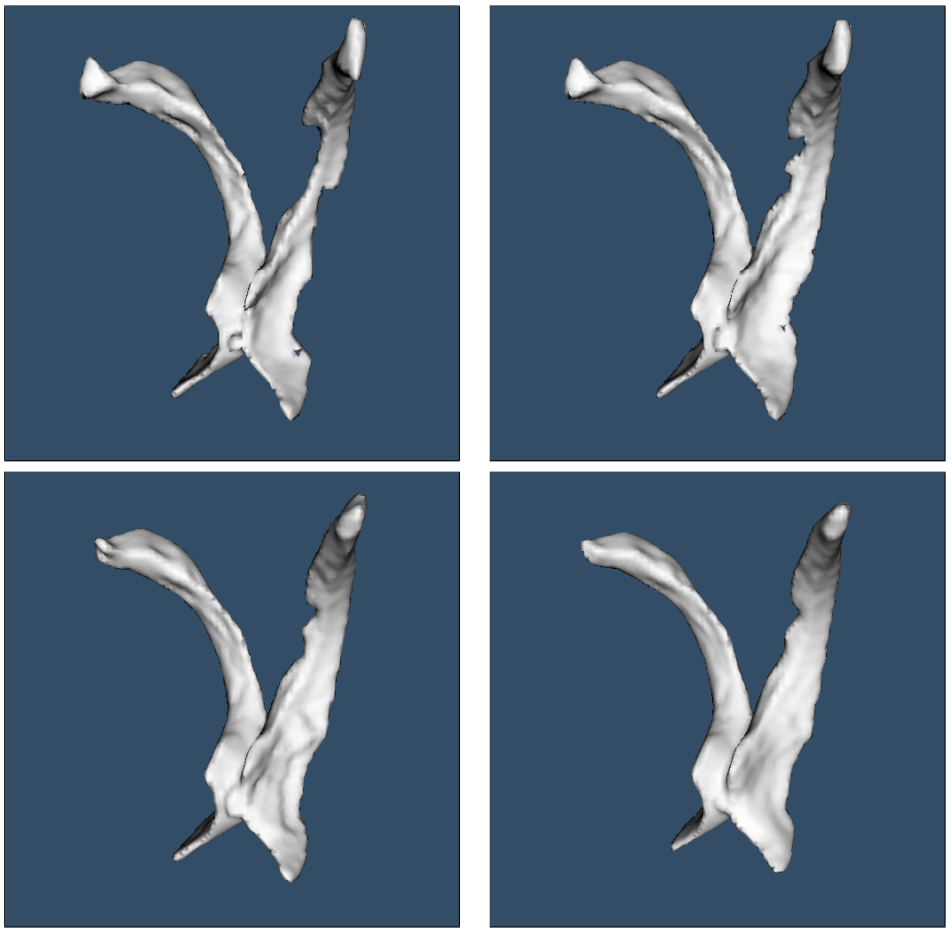


Figure 4.29: Segmentation of the lateral brain ventricles for different weights of the shape prior constraint - From left to right and top to bottom, the shape prior weight b is 0, 0.3, 0.4, 0.5.

Chapter 5

Segmentation of 3D Probability Density Fields: Application to Diffusion MRI

In this chapter, we propose to extend the region-based front evolution proposed in Chapter 2 for scalar and vector-valued images to more complex data. With the objective to segment diffusion magnetic resonance images, which represents a relatively new modality in medical imaging, we consider the problem of segmenting images where each voxel is assigned with a probability density function. In a first time, we propose a direct extension of the vector-valued approach developed so far by assuming the parameters of the densities to belong to a linear space. Then, we consider the general case by defining divergences between probability distributions. A first general approach is proposed by introducing the Kullback-Leibler divergence. Then, we consider the 6-dimensional statistical manifold defined by the parameters of the diffusion tensors. Region statistics are rigorously defined on this Riemannian manifold by computing geodesic distances and intrinsic means. We validate these approaches on synthetic data and show promising results on the extraction of the *corpus callosum* and of the lateral brain ventricles from a real dataset.

5.1 Introduction

In the previous chapters, we have proposed to integrate different cues within front evolutions for scalar and vector-valued images. While most images belong to these two cases, one may have to deal with different data, composed of more complex ”objects”. Medical imaging gives such data, in particular with diffusion magnetic resonance images where each voxel is assigned with a function describing the average motion of

water molecules [12, 96].

In 1994, P. Basser [8] proposed to model voxelwise the probability density function of the molecular motion $r \in \mathbb{R}^3$ by a Gaussian law whose covariance matrix is given by the diffusion tensor $\mathbf{D}$. Diffusion Tensor Imaging (DTI) then produces a volumic image containing, at each voxel, a 3×3 symmetric positive-definite tensor. The estimation of these tensors requires the acquisition of diffusion weighted images in different sampling directions together with a T2 image. Numerous algorithms have been proposed to perform a robust estimation and regularization of these tensors fields [147], [159]. More recently, Q-ball Imaging has been introduced by D. Tuch et al. [150] in order to reconstruct the Orientation Distribution Function (ODF). This ODF is the symmetric probability density function $S^2 \rightarrow \mathbb{R}$ giving the probability for a spin to diffuse in a given direction. Q-ball are then composed by a quantization on the sphere of this ODF at each voxel. This method provides a better angular contrast and is able to recover intra-voxel fiber crossings. Both, DTI and Q-ball imaging measure the same diffusion of water molecules but with a different accuracy. Assuming the diffusion process to follow a Gaussian law, DTI needs less measurements than the general non-parametric estimation obtained with Q-ball imaging which requires a large amount of data. In practice, the long acquisition time of Q-ball images (several hours) and the need of strong diffusion gradients restrict their use to exploration research. Therefore, DTI are naturally favored for human medical studies.

Diffusion MRI is particularly relevant to a wide range of clinical pathologies investigations such as acute brain ischemia detection [138], stroke, Alzheimer disease, schizophrenia [3], etc. It is also extremely useful in order to identify the neural connectivity of the human brain [82], [154], [26]. In the past, many techniques have been proposed to classify gray matter, white matter and cephlo-spinal fluid from T1-weighted MR [169] images but the issue of white matter internal structures segmentation is only feasible from diffusion images and only a few methods have been proposed [173, 70, 158, 157, 165].

In [173], Zhukov et al. defines an invariant anisotropy measure in order to drive the evolution of a level-set and isolate strongly anisotropic regions of the brain. The reduction of the full tensor to a single scalar gives a relatively low discrimination power to the method, potentially resulting in segmentation of mixed structures. On the other side, Wiegell et al. [165], Wang et al. [157, 158] and Jonassan et al. [70] propose different measures of dissimilarity between diffusion tensors. The first two methods use the Frobenius norm of the difference between tensors, together with a spatial coherence or regularity term respectively in a k-means algorithm or active contour model to perform the segmentation of the thalamus nuclei. The third method [158] relies on the symmetrized Kullback-Leibler divergence between Gaussian *pdf*s. The authors derived an affine invariant dissimilarity measure between diffusion tensors and applied it to the segmentation of 2D fields of diffusion tensor images. Finally, the fourth method introduces a geometric measure of dissimilarity by computing the normalized tensor

scalar product of two tensors, which can be interpreted as a measure of overlap. While these works show interesting results on particular structures, the issue of choosing a relevant dissimilarity measure between tensors appear clearly. Various ones are proposed in [165, 157, 158, 70], either assuming linearity [165, 157] or motivated by the geometry of the tensors [70] or by statistical divergence between the underlying probability distributions [158].

In the following, we extend the region-based font evolution technique presented in Chapter 2 to diffusion images. Of course, the main difficulty is to define statistical models over sets of diffusion tensors or more generally, of probability density fields. For this purpose, we need to start by defining a dissimilarity or distance between the considered objects: 3×3 symmetric positive-definite tensor or probability density functions. At first, we extend the works presented in Chapter 2 and in [165, 157] where the components of the diffusion tensors are assumed to belong to 6D linear space. Aiming at taking advantage of region statistics, multivariate Gaussian densities are defined on this 6D space to refine the segmentation process and hence recover irregular anatomical structures such as the minor and major forceps of the *corpus callosum*. Published in [124], this first approach also extend [157] to the 3D case. In the second part, we propose to introduce dissimilarity measures that are defined directly between the underlying probability density functions of the molecular motion. Our method is thus applicable not only to DTI but also to Q-ball data which should enable the proposed algorithm to catch even finer details. We start by defining region statistics making use of the symmetrized Kullback-Leibler (KL) divergence. Then, taking into account the Riemannian geometry of the space of Gaussian *pdf*s will allow us to precise the notion of intrinsic Gaussian law between Gaussian *pdf*s in order to improve our segmentation algorithm. Finally, we present and discuss experimental results both on synthetic and real DTI datasets to compare each method. Before staring with segmentation issue, we recall some details on the acquisition of diffusion images and for the estimation of diffusion tensor images.

5.2 Data acquisition, DTI

In the following, we describe the method used for the acquisition of our data and the robust estimation of the diffusion tensor. Our dataset consists of 30 diffusion weighted images $S_k : \Omega \to \mathbb{R},\ k = 1, ..., 30$ as well as 1 image S_0 corresponding to the signal intensity in the absence of a diffusion-sensitizing field gradient (ie. $b = 0$ in equation 5.1). The dataset was obtained on a GE $1.5\ T$[1]. We recall that the estimation of a field of 3×3 symmetric positive definite tensors $\mathbf{T}$ is done by using the Stejskal-Tanner equation [140] for anisotropic diffusion 5.1 at each voxel x.

$$S_k(\mathbf{x}) = S_0(\mathbf{x}) e^{-b \mathbf{g}_k^T \mathbf{T}(x) \mathbf{g}_k} \quad \forall \mathbf{x} \in \Omega. \tag{5.1}$$

[1] We would like to thank F. Mangin and J.B. Poline for providing us the data used.

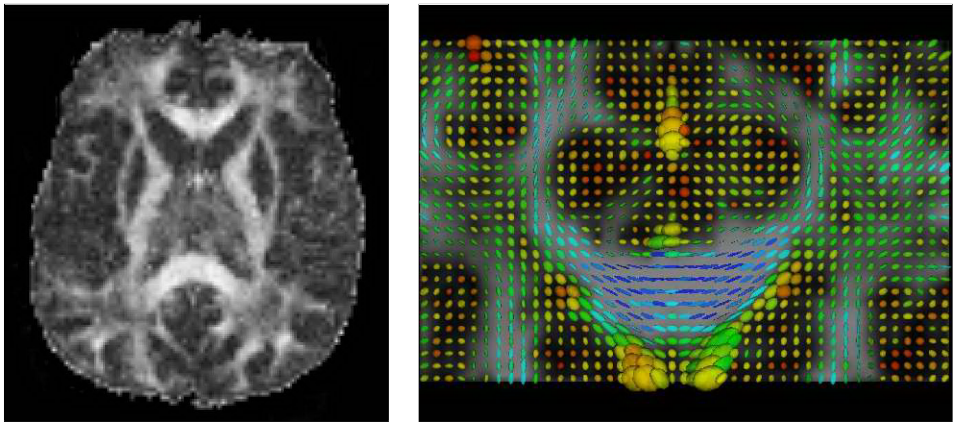


Figure 5.1: [left] Fractional Anisotropy map and [right] Corresponding tensors in the genu of the corpus callosum

where $\mathbf{g}_k$ are the normalized non-collinear sensitizing gradient and b the diffusion weighting factor. Various methods have been proposed for the estimation of the 6 elements of $\mathbf{T}(\mathbf{x})$ by using equation 5.1 (see figure 5.1). A survey of these approaches and a variational framework for the estimation and the regularization of of DTI data can be found in [147]. This last method provides a convenient mean to impose important constraints on the sought solution such as tensor positivity, orthonormality of the eigenvectors or some degree of smoothness of the result. This is performed by minimizing the following energy on the manifold of positive definite tensors $P(3)$:

$$\underset{\mathbf{T}\in P(3)}{\text{Arg min}} \int_{\Omega} \sum_{k=1}^{n} \psi(\| \ln(S_0/S_k) - b\mathbf{g}_k^T \mathbf{T} \mathbf{g}_k \|) + \alpha\rho(\|\nabla \mathbf{T}\|) d\Omega. \tag{5.2}$$

5.3 Diffusion tensor images segmentation

As mentioned in the previous section, the diffusion tensor is directly related to tissue properties. Thus, classical segmentation techniques may be applied on this type of images for the extraction of white matter structures of particular interest. The region-based front evolution developed along this thesis has shown to be efficient to extract structures in classical vector-valued images. We briefly recall the framework presented in Chapter 2. Let $\mathcal{S}$ be the optimal boundary between the 3D object to extract $\mathcal{S}_{in}$ and the 3D background $\mathcal{S}_{out}$, we introduce the level-set function $\phi : \Omega \rightarrow \mathbb{R}^3$, defined as follows:

$$\begin{cases} \phi(\mathbf{x}) = 0, & \text{if } \mathbf{x} \in \mathcal{S} \\ \phi(\mathbf{x}) = \mathcal{D}(\mathbf{x}, \mathcal{S}), & \text{if } \mathbf{x} \in \mathcal{S}_{in} \\ \phi(\mathbf{x}) = -\mathcal{D}(\mathbf{x}, \mathcal{S}), & \text{if } \mathbf{x} \in \mathcal{S}_{out} \end{cases}$$

where $\mathcal{D}(\mathbf{x}, \mathcal{S})$ stands for the Euclidean distance between $\mathbf{x}$ and $\mathcal{S}$ and $\Omega = \mathcal{S}_{in} \cup \mathcal{S}_{out}$. Furthermore, let $H_\epsilon(z)$ and $\delta_\epsilon(z)$ be regularized versions of the Heaviside and Dirac functions as defined in Chapter 1.

Let $q(\mathbf{x}, r)$ be the probability density function of our random vector r of $\mathbb{R}^3$ describing the water molecules average motion at a given voxel $\mathbf{x}$ of a diffusion MR image $\Omega \subset \mathbb{R}^3$ and for a given diffusion time τ imposed by the parameters of the PGSE (Pulsed Gradient Spin Echo) sequence. We are interested in characterizing the global coherence of that *pdf* field. According to the framework developed in Chapter 2, the optimal partitioning of image domain Ω in two regions $\mathcal{S}_{in}$ and $\mathcal{S}_{out}$ is obtained from the minimization of:

$$\begin{aligned} E(\phi, p_1, p_2) = &- \int_\Omega H_\epsilon(\phi) \log p_1(q(\mathbf{x}, .))\, d\mathbf{x} - \int_\Omega (1 - H_\epsilon(\phi)) \log p_2(q(\mathbf{x}, .))\, d\mathbf{x} \\ &+ \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x}, \end{aligned} \tag{5.3}$$

where p_1 and p_2 are respectively the distributions of the probability of $q(\mathbf{x}, .)$ to belong in $\mathcal{S}_{in}$ and $\mathcal{S}_{out}$. To define these two distributions p_1 and p_2, statistical measures must be defined for the considered data.

When dealing with Diffusion Tensor Images, the voxel-wise motion of molecules is described by a Gaussian distribution. The diffusion tensor, i.e. the covariance matrix of the diffusion process, can be used to express our segmentation criterion:

$$\begin{aligned} E(\phi, \hat{p}_1, \hat{p}_2) = &- \int_\Omega H_\epsilon(\phi) \log \hat{p}_1(\mathbf{T}(\mathbf{x}))\, d\mathbf{x} - \int_\Omega (1 - H_\epsilon(\phi)) \log \hat{p}_2(\mathbf{T}(\mathbf{x}))\, d\mathbf{x} \\ &+ \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x}. \end{aligned} \tag{5.4}$$

Then, we still need to define probability distributions over sets of 3×3 symmetric positive-definite matrices. A simple model can be obtained by assuming the 6 different parameters of these covariance matrices to belong to the linear space $\mathbb{R}^6$. This simplification leads to the segmentation of a vector-valued image. We recall our algorithm for this case in the next section.

It must be noted that such a simplification does not integrate the properties of the data and we propose to build more adequate models in Section 5.3.2 and 5.3.3. First, the Kullback-Leibler is introduced to define statistics over a general probability density field. Then, we propose to consider the Riemannian geometry of the space of Gaussian *pdf*s to define geodesic distances between Gaussian distributions, i.e. diffusion tensors, as well as intrinsic means. Multivariate Gaussian estimations are also introduced in this non-linear space, allowing us to capture regions with complex distributions of diffusion tensors.

5.3.1 Linear approximation

We consider a statistical distribution on linear spaces which overcomes the hypothesis of isotropic distribution. Hence, as done in Chapter 3 for texture images with the structure tensor, we consider a parametric approximation with a 6D Gaussian. Let u be the vector representation of a tensor $\mathbf{T}$, the likelihood of u in the region $s = 1, 2$ is given by:

$$p_s(u|\mu_s, \Sigma_s) = \frac{1}{(2\pi)^3|\Sigma_s|^{1/2}} e^{-\frac{1}{2}(u-\mu_s)^T\Sigma_s^{-1}(u-\mu_s)}.$$

By construction, the diagonal and non-diagonal components of a diffusion tensor are highly correlated and so, a full covariance matrix must be considered in the density of its vector representation u. Then, the Euclidean vector means and the covariance matrices of these densities are also supposed unknown. However, these parameters can be introduced as unknown in (5.4):

$$\begin{aligned} E(\phi, \{\mu_{1,2}, \Sigma_{1,2}\}) =& \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x} - \int_\Omega H_\epsilon(\phi) \log p_1(u(\mathbf{x})|\mu_1, \Sigma_1)\, d\mathbf{x} \\ & - \int_\Omega (1 - H_\epsilon(\phi)) \log p_2(u(\mathbf{x})|\mu_2, \Sigma_2)\, d\mathbf{x}. \end{aligned} \tag{5.5}$$

This type of energy was studied in Chapter 2. The Euler Lagrange equation for ϕ yields the following evolution equation for the level-set function:

$$\begin{aligned} \phi_t(\mathbf{x}) = \delta_\epsilon(\phi(\mathbf{x}))\Bigg(\nu \,\mathrm{div}\frac{\nabla\phi}{|\nabla\phi|} + \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(u(\mathbf{x}) - \mu_1)^T\Sigma_1^{-1}(u(\mathbf{x}) - \mu_1) \\ + \frac{1}{2}(u(\mathbf{x}) - \mu_2)^T\Sigma_2^{-1}(u(\mathbf{x}) - \mu_2)\Bigg) \end{aligned}$$

while the statistical parameters are updated as follows:

$$\begin{aligned} \mu_1(\phi) &= \frac{\int_\Omega u(\mathbf{x})H_\epsilon(\phi)d\mathbf{x}}{\int_\Omega H_\epsilon(\phi)d\mathbf{x}}, \\ \mu_2(\phi) &= \frac{\int_\Omega u(\mathbf{x})(1 - H_\epsilon(\phi))d\mathbf{x}}{\int_\Omega (1 - H_\epsilon(\phi))d\mathbf{x}}, \\ \Sigma_1(\phi) &= \frac{\int_\Omega (\mu_1 - u(\mathbf{x}))(\mu_1 - u(\mathbf{x}))^T H_\epsilon(\phi)d\mathbf{x}}{\int_\Omega H_\epsilon(\phi)d\mathbf{x}}, \\ \Sigma_2(\phi) &= \frac{\int_\Omega (\mu_2 - u(\mathbf{x}))(\mu_2 - u(\mathbf{x}))^T (1 - H_\epsilon(\phi))d\mathbf{x}}{\int_\Omega (1 - H_\epsilon(\phi))d\mathbf{x}}. \end{aligned}$$

Implementation schemes for this type of optimization are developed in Chapter 2. Two important details must be noted: (i) the explicit scheme is not stable for any time step because of regularization term, (ii) the level set function is reinitialized to the distance function at each iteration. If we restrict the covariance matrix to the identity matrix, these equations simplify and the log-likelihoods in equation (5.5) become simply the Euclidean distance between the vectors u and $\mu_{s=1,2}$, which is equivalent to the Frobenius norm of the difference between the corresponding tensors, as nicely studied in [157].

5.3.2 Probability density fields segmentation

General case

We consider a general probability density function $q(\mathbf{x}, r)$ of the random vector r of $\mathbb{R}^3$ describing the water molecules average motion at a given voxel $\mathbf{x}$. The classical symmetrized Kullback-Leibler divergence can be used to express the dissimilarity between diffusion processes at different voxels. With $q_1(\mathbf{x}, .), q_2(\mathbf{y}, .)\ \forall \mathbf{x}, \mathbf{y} \in \Omega$ two *pdfs* from $\mathbb{R}^3$ onto $\mathbb{R}^+$, their KL divergence is given by:

$$\mathcal{D}_{kl}(q_1, q_2) = \frac{1}{2} \int_{\mathbb{R}^3} \left(q_1(r) \log \frac{q_1(r)}{q_2(r)} + q_2(r) \log \frac{q_2(r)}{q_1(r)} \right) dr. \tag{5.6}$$

Assuming a partition of the data between the structure we try to segment $\mathcal{S}_{in}$ and the rest of the volume $\mathcal{S}_{out}$, we again seek the optimal separating surface $\mathcal{S}$ between those two subsets. We denote by $\overline{q}_1$ and $\overline{q}_2$ the mean *pdfs* over $\mathcal{S}_{in}$ and $\mathcal{S}_{out}$ verifying equation 5.9. It is then possible to model the regional distributions to $\overline{q}_1$ and $\overline{q}_2$ by suitable densities p_1^{kl}, p_2^{kl}. In the following, we make the assumption that p_1^{kl} and p_2^{kl} are centered Gaussians in $\overline{q}_1$ and $\overline{q}_2$ of variances σ_1^2 and σ_2^2. We then define the following energy in order to maximize the likelihood of these densities on their associated domain:

$$\begin{aligned} E(\phi, \{\sigma_{1,2}^2, \overline{q}_{1,2}\}) =& \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x} - \int_\Omega H_\epsilon(\phi) \log p_1^{kl}(q(\mathbf{x})|\overline{q}_1, \sigma_1^2)\, d\mathbf{x} \\ & - \int_\Omega (1 - H_\epsilon(\phi)) \log p_2^{kl}(q(\mathbf{x})|\overline{q}_2, \sigma_2^2)\, d\mathbf{x}, \end{aligned} \tag{5.7}$$

where

$$p_i^{kl}(q|\overline{q}_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \frac{-\mathcal{D}_{kl}^2(q, \overline{q}_i)}{2\sigma_i^2}, \quad i = 1, 2.$$

Of course, other models can easily be used for the p_i. Note that in the case where the σ_i are equal to 1, this energy will equal to the one proposed in [158]. In the experimental part, the importance of adding σ_i will be illustrated.

The derivation of the Euler-Lagrange equations for this class of energy was studied in Chapter 2 and yields the following evolution for ϕ:

$$\phi_t(\mathbf{x}) = \delta_\epsilon(\phi(\mathbf{x})) \left(\nu \text{div} \frac{\nabla \phi}{|\nabla \phi|} + \log \frac{p_1^{kl}(q(\mathbf{x})|\overline{q}_1, \sigma_1^2)}{p_2^{kl}(q(\mathbf{x})|\overline{q}_2, \sigma_2^2)} \right). \tag{5.8}$$

Moreover, the derivation of the energy with respect to σ_i^2 and $\overline{q}_i$ provide the update formulae for these statistical parameters. It can be shown that the variance must be updated with its empirical estimation (see equation 5.3.1) whereas some more work is needed for the $\overline{q}_i$. We indeed have to estimate:

$$\overline{q}_i = \underset{q_s}{\text{Arg min}} \int_{\Omega_i} \mathcal{D}_{kl}^2(q_s, q(\mathbf{x}))\, d\mathbf{x}, \quad i = 1, 2. \tag{5.9}$$

For a general *pdf* $q(\mathbf{x})$, for instance if we consider the ODF derived from Q-ball data, the variance is easily computed but the estimation of the $\overline{q}_i$ might require the use of numerical approximation techniques if no closed-form is available.

Application to Gaussian Distributions (DTI)

We now explicitly express the energy (5.7) for Gaussian *pdf*s and use a Gaussian law to model the distributions. The energy becomes

$$E(\Omega_i, \sigma_i^2, \overline{q}_i) = \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x} + \frac{1}{2}\int_\Omega H(\phi)\big(\log(2\pi\sigma_1^2) + \mathcal{D}_{kl}^2(q(\mathbf{x}), \overline{q}_1)\sigma_1^{-2}\big)\, d\mathbf{x} + \frac{1}{2}\int_\Omega (1 - H(\phi))\big(\log(2\pi\sigma_2^2) + \mathcal{D}_{kl}^2(q(\mathbf{x}), \overline{q}_2)\sigma_2^{-2}\big)\, d\mathbf{x}. \tag{5.10}$$

If we write, as in [158], the Kullback-Leibler distance between two Gaussian *pdf*s parameterized by their covariance matrices (*.ie* the Diffusion Tensor $\mathbf{T}$) as:

$$\mathcal{D}_{kl}^2(q(\mathbf{x}), \overline{q}_i) = \frac{1}{2}\left(\text{trace}\left[\mathbf{T}^{-1}(\mathbf{x})\overline{\mathbf{T}}_i + \overline{\mathbf{T}}_i^{-1}\mathbf{T}(\mathbf{x})\right] - 3\right), \tag{5.11}$$

then the Euler-Lagrange equations for our energy become:

$$\phi_t(\mathbf{x}) = \delta_\epsilon(\phi(\mathbf{x}))\left(\nu \text{div}\frac{\nabla\phi}{|\nabla\phi|} + \frac{1}{2}\left(\frac{1}{4}\left(\text{trace}\left[\mathbf{T}^{-1}(\mathbf{x})\overline{\mathbf{T}}_2 + \overline{\mathbf{T}}_2^{-1}\mathbf{T}(\mathbf{x})\right]\sigma_2^{-2} - \text{trace}\left[\mathbf{T}^{-1}(\mathbf{x})\overline{\mathbf{T}}_1 + \overline{\mathbf{T}}_1^{-1}\mathbf{T}(\mathbf{x})\right]\sigma_1^{-2}\right) + \log\frac{\sigma_2^2}{\sigma_1^2} + \frac{3(\sigma_1^2 - \sigma_2^2)}{2(\sigma_1^2\sigma_2^2)}\right)\right). \tag{5.12}$$

Notice that we obtain additional terms (the σ_i coefficients) in (5.12) if compared to the Euler-Lagrange equations proposed in [158].

For a given state of ϕ, closed-forms for the optimal covariance matrices of each region are available. In their recent paper, Wang et al. [158] showed that these covariance matrices are given by

$$\overline{\mathbf{T}}_i = \sqrt{\mathbf{B}_i^{-1}}\left(\sqrt{\sqrt{\mathbf{B}_i}\mathbf{A}_i\sqrt{\mathbf{B}_i}}\right)\sqrt{\mathbf{B}_i^{-1}},$$

where

$$\mathbf{A}_i = \int_{\Omega_i}\mathbf{T}(\mathbf{x})d\mathbf{x} \quad \text{and} \quad \mathbf{B}_i = \int_{\Omega_i}\mathbf{T}^{-1}(\mathbf{x})d\mathbf{x},$$

thus giving the update formula of the $\overline{q}_i$.

The Kullback-Leibler divergence is a common choice to measure dissimilarities between *pdf*s. However, the real meaning of this divergence is not always mentioned and for particular parametric densities like multivariate Gaussians with 0-mean, better

measures are naturally available. In the next paragraph, we see how a Riemmanian metric can be associated to the parameter space of these densities using the Fisher information matrix. Geodesic distances and intrinsic means can then be computed.

5.3.3 A Riemmanian approach to DTI segmentation

We now consider a Riemannian manifold $\mathcal{M}$ representing the family of three-dimensional Gaussian probability density functions through the 6 parameters of their covariance matrices. The *pdf*s are, again, assumed to be of 0-mean since this simply translates the fact that the average displacement of spins in a voxel is zero. Following the work by Rao [119] and Burbea-Rao [24], where a Riemannian metric was introduced in term of the information matrix, we wish to define geodesic distances and intrinsic means on this 6-dimensional manifold whose coordinate system is given by a real vector parameter $\theta = (\theta_1, ..., \theta_6) \in \mathbb{R}^6$ such that for all random vector $r \in \mathbb{R}^3$, $\mathcal{M} = \{q(r|\theta)\}$. In the following, we present the closed-form expression of the geodesic distance as well as an original algorithm to estimate the intrinsic mean of multivariate Gaussian densities with common mean.

Geodesic distance and intrinsic mean

We concentrate on the space $S^+(m, \mathbb{R})$, endowed with the information metric g. $S^+(m, \mathbb{R})$ denotes the set of $m \times m$ real symmetric positive-definite matrices. A detailed study on the definition of statistical models on this nonlinear space was presented in [84]. Another recent work by Pennec et al [114] relies on a comparable approach to derive tensor fields filtering techniques. We now remind some important results, necessary for our segmentation task. Following [84] and [114, 54, 101, 100, 56, 25, 23, 52, 137], S^+ can be characterized as an affine symmetric space for which closed-form expressions are available for the solution of the geodesic equations as well as for the geodesic distance (also known as Rao's distance). The geodesic distance $\mathcal{D}_g$ between any two elements Σ_1 and Σ_2 of S^+, whose components are given by the previously introduced vector parameter θ, was derived by Jensen in the following theorem:

Theorem 5.1: (S.T. Jensen, 1976)
Consider the family of multivariate Gaussian distributions with common mean vector μ but differing variance-covariance matrices Σ. The geodesic distance between two members of the family with variance-covariance matrices Σ_1 and Σ_2 is given by

$$\mathcal{D}_g(\Sigma_1, \Sigma_2) = \frac{1}{2}\text{trace}(\log^2(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})) = \frac{1}{2}\sum_{i=1}^{m}\log^2(\lambda_i),$$

where the λ_i denote the m eigenvalues of the determinantal equation $|\lambda\Sigma_2 - \Sigma_1| = 0$.

Properties of the geodesic distance:
The distance $\mathcal{D}_g$ on S^+ defined above exhibits some nice properties: positivity, symmetry, triangle inequality, invariance under congruent transformations and inversion. The interested reader can find more details about this in the technical report by Förstner and Moonen [56].

This distance can be used in the segmentation algorithm presented in the previous part, replacing the symmetrized Kullback-Leibler divergence by the new expression. For this purpose, we will need to estimate the empirical mean as proposed by Fréchet [57], Karcher [73] and Pennec [113]. A closed-form expression of the mean cannot be obtained [101] but a gradient descent algorithm was proposed in [84]. A flow is derived from an initial guess $\hat{\Sigma}_0$ toward the mean of a set of $S^+(m,\mathbb{R})$. The following evolution was obtained:

$$\hat{\Sigma}_{t+1} = \hat{\Sigma}_t^{1/2} \exp(-\frac{1}{N}\hat{\Sigma}_t^{1/2} \sum_{k=0}^{N} \log(\Sigma_k^{-1}\hat{\Sigma}_t)\hat{\Sigma_t}^{-1/2})\hat{\Sigma}_t^{1/2}. \tag{5.13}$$

The derivation of this intrinsic numerical scheme is detailed in [84].

A normal distribution on multivariate normal distributions

In order to derive the expression of a normal distribution on S^+, we make use of the various quantities derived up to that point. We thus proceed by plugging the appropriate quantities in the generalization of the normal distribution to Riemannian manifolds proposed in [113] for sufficiently concentrated probability density functions, e.g. for small covariance matrices. Following Theorem 4 proved in [113], we have the following theorem:

Theorem 5.1 *The normal distribution in $S^+(m,\mathbb{R})$ for an $m \times m$ covariance matrix Λ of small variance $\sigma^2 = \mathrm{tr}(\Lambda)$ is of the form*

$$k.\exp\frac{-\beta\gamma\beta^T}{2},$$

where

- *β is defined as $\nabla\mathcal{D}_g^2(\hat{\Sigma},\Sigma) = \hat{\Sigma}\log(\Sigma^{-1}\hat{\Sigma})$, as detailed in [101],*
- *The normalization constant: $k = \frac{1+O(\sigma^3)+\epsilon(\frac{\sigma}{r})}{\sqrt{(2\pi)^m|\Lambda|}}$,*
- *The concentration matrix: $\gamma = \Lambda^{-1} - \frac{Ricci}{3} + O(\sigma) + \epsilon(\frac{\sigma}{r})$, where $Ricci$ is the Riccian tensor whose expression can be found in [84].*

r is the injection radius at $\hat{\Sigma}$ and ϵ is such that $\lim_{o+} x^{-p}\epsilon(x) = 0\ \forall p \in \mathbb{R}^+$.

For further details on the Riemmanian geometry of the space multivariate normal distributions, we refer the reader to [84].

Segmentation of DTI assuming multivariate normal distributions

From this definition, we can express the probability of a diffusion tensor to belong to a region with a given normal distribution. For this purpose, the diffusion tensors $\mathbf{T}$ are considered as the covariance matrices Σ used so far. Assuming the data to follow a normal distribution in region Ω_s, the likelihood of a diffusion tensor $\mathbf{T}$ is given by $p(\mathbf{T}|\hat{\mathbf{T}}_s, \Lambda_s)$ of theorem 5.1. Our region-based functional can then be expressed as:

$$E(\phi, \hat{\mathbf{T}}_{s=1,2}, \Lambda_{s=1,2}) = \nu \int_\Omega |\nabla H_\epsilon(\phi)|\, d\mathbf{x} - \int_\Omega H_\epsilon(\phi) \log p(\mathbf{T}(\mathbf{x})|\hat{\mathbf{T}}_1, \Lambda_1)\, d\mathbf{x} - \int_\Omega (1 - H_\epsilon(\phi)) \log p(\mathbf{T}(\mathbf{x})|\hat{\mathbf{T}}_2, \Lambda_2)\, d\mathbf{x}. \tag{5.14}$$

We have introduced all the required tools for the minimization of this energy. The corresponding level-set evolution equation is:

$$\phi_t(\mathbf{x}) = \delta_\epsilon(\phi(\mathbf{x})) \left(\nu \,\mathrm{div} \frac{\nabla \phi}{|\nabla \phi|} + \log \frac{p(\mathbf{T}(\mathbf{x})|\hat{\mathbf{T}}_1, \Lambda_1)}{p(\mathbf{T}(\mathbf{x})|\hat{\mathbf{T}}_2, \Lambda_2)} \right), \tag{5.15}$$

while the optimal statistical parameters are their empirical estimates. The intrinsic means $\hat{\mathbf{T}}_1$ $\hat{\mathbf{T}}_2$ are estimated with the algorithm obtained from equation (5.13 with the gradient descent:

$$\hat{\mathbf{T}}_{s=1,2;t+1} = \hat{\mathbf{T}}_{s;t}^{1/2} \exp\left(-\frac{1}{|\Omega_s|} \hat{\mathbf{T}}_{s;t}^{1/2} \left(\int_{\Omega_s} \log(\mathbf{T}(\mathbf{x})^{-1} \hat{\mathbf{T}}_{s;t})\, d\mathbf{x} \right) \hat{\mathbf{T}}_{s;t}^{-1/2} \right) \hat{\mathbf{T}}_{s;t}^{1/2}. \tag{5.16}$$

The empirical covariance matrices are then easily obtained:

$$\Lambda_{s=1,2} = \frac{1}{|\Omega_s|} \int_{\Omega_s} \beta(\mathbf{x}) \beta(\mathbf{x})^T, \tag{5.17}$$

with $\beta(\mathbf{x}) = \nabla \mathcal{D}_g^2(\hat{\mathbf{T}}_s, \mathbf{T}(\mathbf{x})) = \hat{\mathbf{T}}_s \log\left(\mathbf{T}(\mathbf{x})^{-1} \hat{\mathbf{T}}_s \right)$.

Introducing multivariate statistics to represent the distribution of diffusion tensors inside image sub-regions, we are able to capture structures with more complex diffusivity properties than the univariate model of Section 5.3.2. We show this difference in the next section where several experimental comparisons are presented between the three approaches presented so far.

5.4 Experimental Results and Comparisons

We will respectively refer to the methods presented in Sections 5.3.1, 5.3.2 and 5.3.3 as Method 1, 2 and 3. We begin with a validation of Method 1 on synthetic data with illustrative examples where other approaches fail. Then, experiments are conducted on the extraction of the corpus callosum from DTI data by Methods 1, 2 and 3. We exhibit promising results of Method 3.

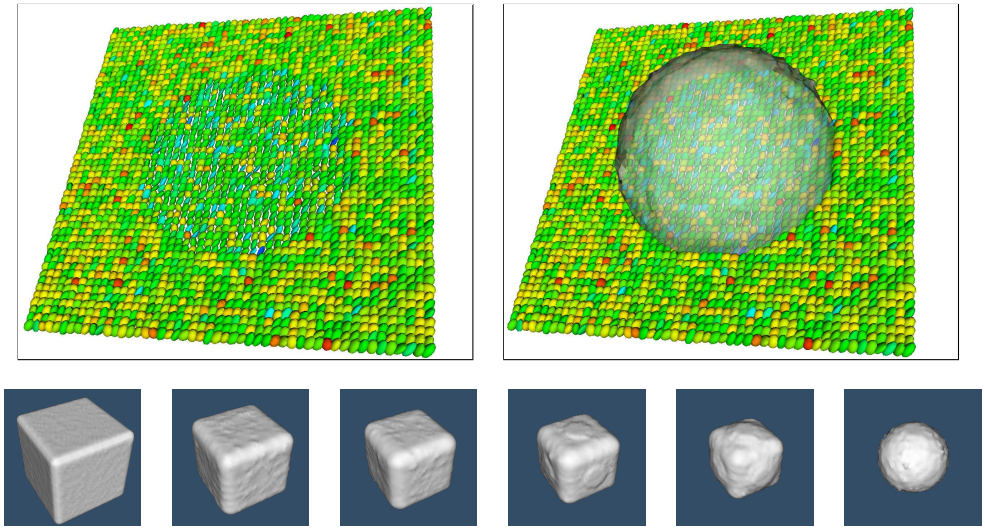


Figure 5.2: Segmentation (Method 1) of a noisy tensor field composed by two regions with same orientation but different scale ([TOP LEFT]: 2D-cut of the tensor field, TOP RIGHT: Final segmentation, BOTTOM: Surface evolution)

5.4.1 Synthetic data

Diffusion tensor images measure the average displacement of water molecules. This displacement characterizes different tissues and provides two different information: its intensities and its directions. When considering diffusion tensor images, these information are given respectively by the largest eigenvalue and the corresponding eigenvector. From this decomposition, we built one easy case to illustrate the general segmentation process (all methods will succeed in this example). It is made of a spherical inclusion [Figure 5.2] where the difference between the inside and the outside is only based on the eigenvalues. We also generate a more complicated case where the two regions differ only by the major orientation of the tensors. We also vary the main orientation of the tensors within the inside region by creating a junction as shown on [Figure 5.3]. In both cases, a Gaussian noise was added directly on the eigen-elements of each tensor. Initializing the surface with a bounding box, Method 1 is able to give the expected segmentations [Figure 5.2 and 5.3].

However, these examples do not show the necessity of including a statistical model for the distance distribution of each region and the approach proposed in [158] for 2D fields of *pdf*s gives a similar result. In order to show the advantages of our Method 2, which is more general, we have generated another test image. As shown on [Figure 5.4], it is composed by a torus whose internal tensors are oriented according to the tangential direction of the principal circle of the torus. A noise is also added to all the tensors of the image but with a different variance whether the tensor is inside or outside the torus. On [Figure 5.4], we compare the results obtained using [158] and our approach, for different initializations. The first method fails to segment the torus

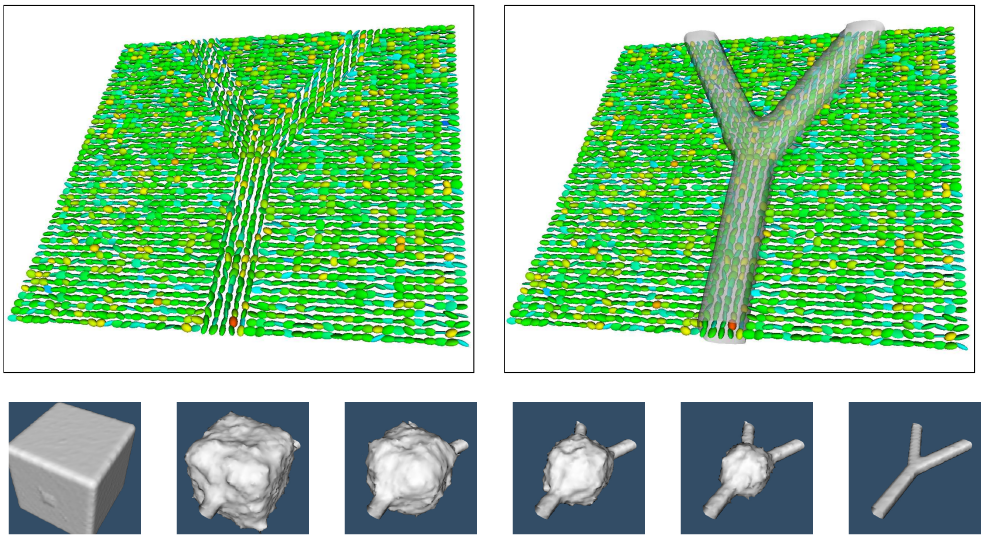


Figure 5.3: Segmentation (Method 1) of a noisy tensor field composed by two regions with same scale but different orientations (TOP LEFT: 2D-cut of the tensor field, TOP RIGHT: Final segmentation, BOTTOM: Surface evolution).

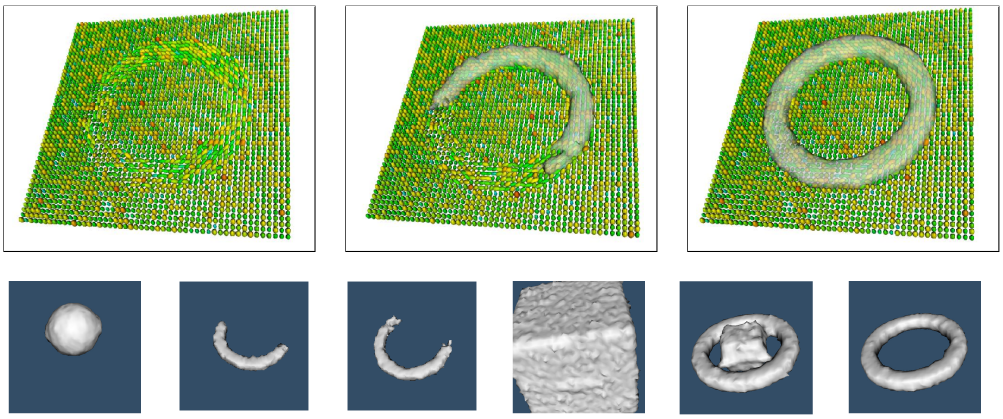


Figure 5.4: Segmentation of a noisy tensor field composed by two regions with different orientations (TOP LEFT: 2D-cut of the tensor field, TOP CENTER: Segmentation obtained from [158], TOP RIGHT: Segmentation obtained with Method 2, BOTTOM: Respective surface evolutions).

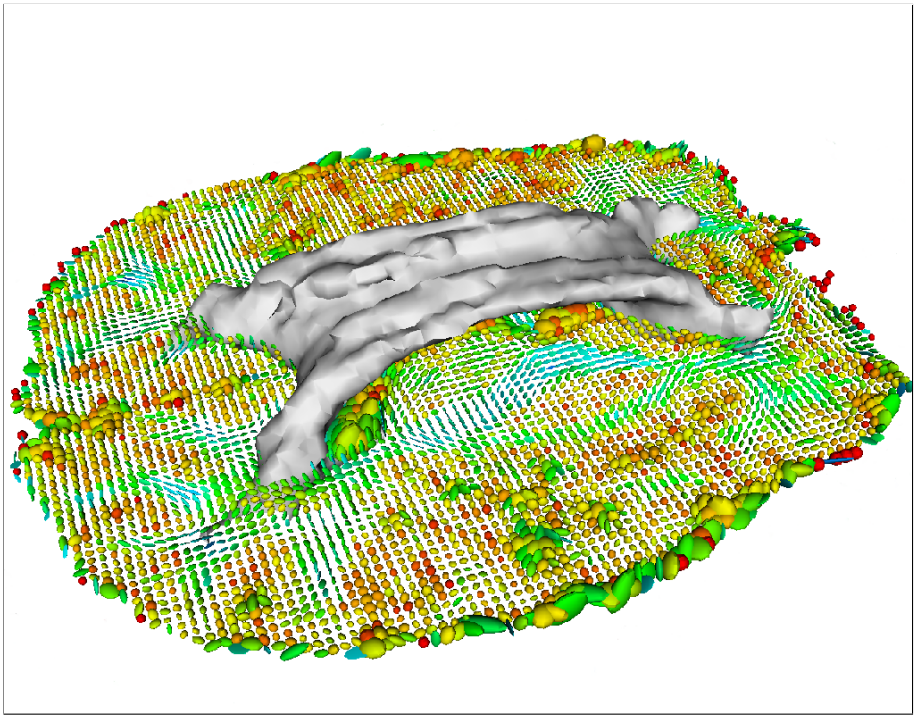


Figure 5.5: Segmentation of the corpus callosum obtained with Method 1

because the high variation of the orientations within each region. If we initialize with a bounding box, the surface shrinks until it disappears and if we start from a small sphere inside the torus, only a small part of the torus can be recovered. Using our Method 2, which models the variance of the distance between tensors, the torus is correctly extracted for the different initializations.

5.4.2 Real diffusion tensor data

Experiments for Methods 1 and 2

As mentioned in the introduction, the extraction of objects from DTI data is of great interest. This modality gives the opportunity to discriminate structures like the corpus callosum, much harder to characterize using other modalities. Before any processing, we need to roughly crop the image around the object of interest in order to respect the assumption of bi-partitioning imposed by our model. [Figure 5.5] shows the result obtained for the extraction of the corpus callosum with Method 1. The goal of the next experiment is to extract the lateral ventricles with Method 2. Two small spheres are manually set inside the ventricles to initialize the surface. The evolution and the final segmentation are shown on [Figure 5.6]. This result seems to be consistent with what we could expect from dissection data even though the validation for this type of data will have to be carefully addressed. However, Method 2 is unable to capture structures like the corpus callosum with complex diffusion tensor distributions because of the

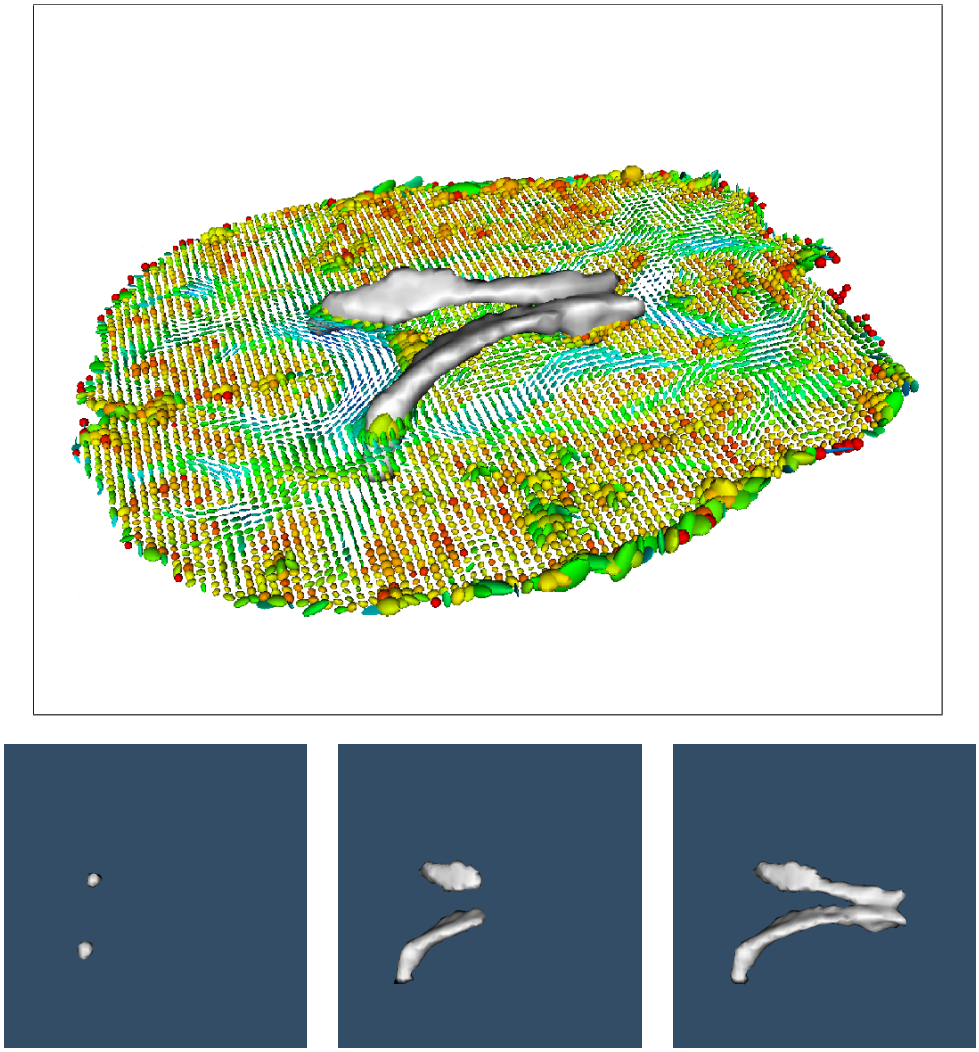


Figure 5.6: Segmentation (Method 2) of the lateral brain ventricles [TOP] in a real diffusion tensor image superimposed on the DTI field, [BOTTOM] Surface evolutions.

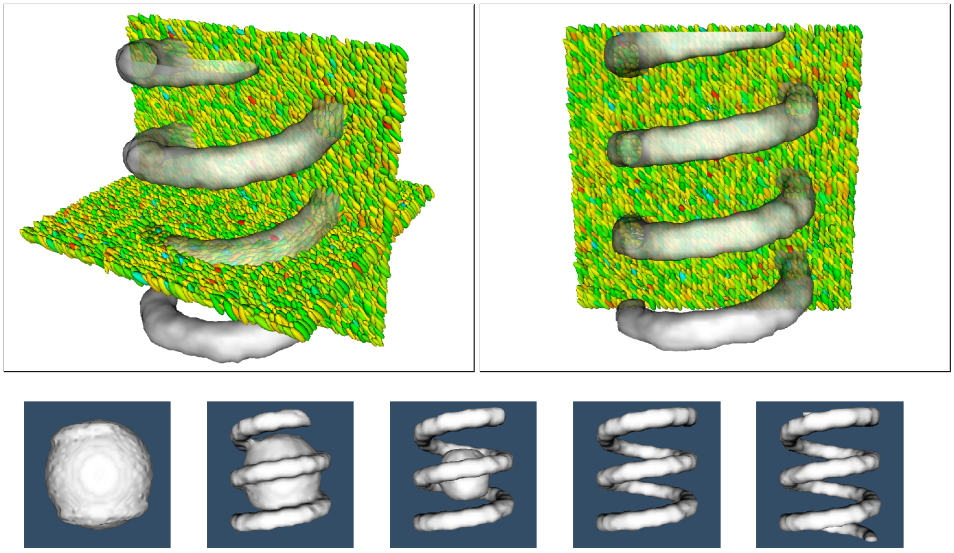


Figure 5.7: [TOP] Segmentation (Method 3) of a noisy tensor field composed by two regions with very different orientations, [BOTTOM] Surface evolution.

univariate model.

5.4.3 Experiments for Method 3

In order to test our Method 3, we show results on a much harder synthetic dataset on Figure 5.7. It is composed of an helix of tensors whose orientations follow the tangent to the helix at each position. All the tensors have the same eigenvalues. While method 1 and 2 do not capture the whole helix, considering the geodesic distance helps and gives the expected result. Results on real images are presented on [Figure 5.8] for the extraction of the lateral brain ventricles and the corpus callosum.

5.5 Contributions and conclusions

We have presented several extensions of the region-based front evolution of Chapter 2 to the case of magnetic diffusion images. After showing that diffusion images cannot me reduced the vector-valued case, we have proposed two original ways of including the properties of the considered data. First, the Kullback-Leibler divergence is employed to define region statistics over sets of *pdf*s. Then, considering the particular case of diffusion tensor images, a Riemmanian metric is associated to the multivariate 0-mean Gaussian densities observed at each voxel. Geodesic distance and intrinsic mean are formulated and statistics over samples of this manifold are deduced. This final approach gives us a segmentation method adapted to DTI with a strong mathematical justification. Several experimental results have shown that our approach can extract important structures from the brain white matter like the *corpus callosum* and

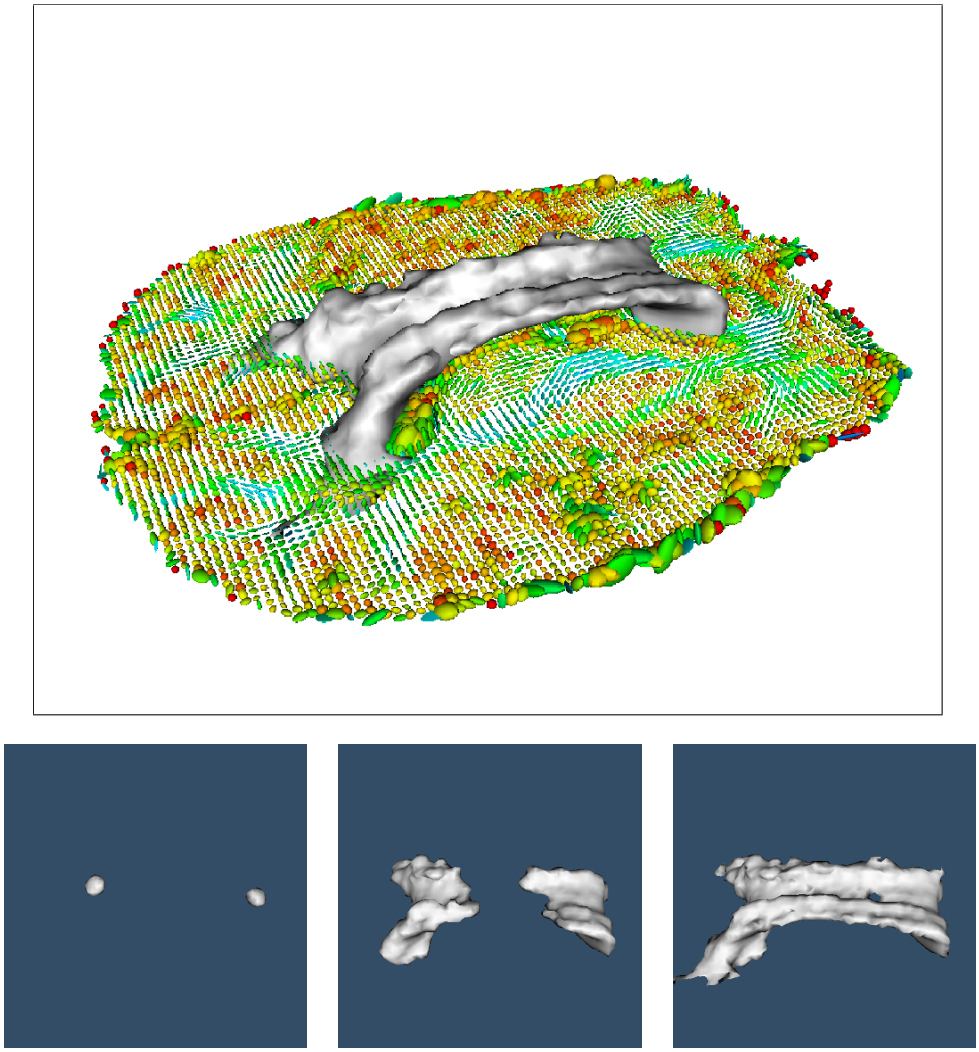


Figure 5.8: Segmentation (Method 3) of the corpus callosum [TOP] in a real diffusion tensor image superimposed on the DTI field, [BOTTOM] Surface evolution.

it seems promising for clinical applications. However, the validation of the extracted structures still need to be done which is a quite difficult task since no ground truth data are available yet. We can summarize the contributions of the is chapter with the following points:

1. Implementation of a direct extension of the vector-valued region-based front evolution,
2. Introduction of the Kullback-Leibler divergence to define statistics over probability density fields,
3. Definition of statistics on the statistical manifold representing the family of Gaussian probability density functions with 0-mean through the 6 parameters of their covariance matrices.

These contributions have been published in two separates research reports: [85] presents a detailed description of the Riemmanian geometry considered to define statistics over sets of multivariate 0-mean Gaussians while [84] is dedicated to the segmentation issue of DTI. Parts of this work have also been presented in two conferences [124] and [83].

Acknowledgments: This chapter mainly describes a joint work with Christophe Lenglet and I would to thank him for his expertise on diffusion images and for his strong involvement in the extension of vector-valued segmentation approaches to such complex data.

Conclusion

We have presented in this thesis a general framework for unsupervised image segmentation. A region-based front evolution has been obtained from a Bayesian formulation of the problem. Improving and extending existing works, we have been able to extract with good accuracy objects from a wide variety of natural and medical images. The main contributions of this thesis is the definition of a general framework that permits to integrate various cues to solve the problem of image segmentation in practical cases. This objective has been achieved using and combining several state of the art theories:

1. We introduce a Bayesian region-based front evolution which permits to integrate statistical models within a geometric approach,
2. The level-set representation has been employed to define our partitioning functional, driving naturally to an implicit front evolution restricted to the vicinity of the interface,
3. Parametric and non-parametric region densities have been estimated actively during the segmentation process in the case of scalar and vector-valued images,
4. Texture and motion features have been extracted and integrated in our framework,
5. An implicit approach has been presented to model and integrate prior shape knowledge,
6. Finally, this framework has been extended to more complex images like diffusion MR images.

Throughout these studies, we were concentrated on the unsupervised aspect of the approach. Rather than using complex statistical models, we have tried to identify and extract important information that defines regions of interest in an image. It has been rather successful since we have been able to obtain better results than numerous recent works and with a lower complexity. However, if we target a strong robustness, learning the few parameters of our approach and using more robust estimators may be necessary but we think that our framework is a good basis to elaborate efficient image analysis and recognition systems.

Of course, there is still place for improvement in most parts of this thesis. For example, the extraction of texture features may be extended to deal with a wider range of images. Several tasks remain to be done to model and integrate implicitly prior shape knowledge: definition of a ”natural” distance between shapes, implicit introduction of invariance to global transformations, integration of local variability in the second shape model... Also, the last part on diffusion MR images still lacks an in depth evaluation on a large real dataset and the results should be validated by experts.

When considering natural images, what we are trying to achieve with the integration of relevant cues and prior knowledge is to mimic the complex and not yet understood human system. The progress in brain imagery brings powerfull means to study the way we are doing these segmentation tasks ourseleves and the future of image analysis may be in bringing together computer scientist and neurologist to create efficient biologically motivated models like in [80].

Conclusion (version française)

Un cadre de travail général pour la segmentation non-supervisée d'images a été présenté dans cette thèse. Une évolution de front basée région a été obtenue à partir d'une formulation Bayésienne du problème. La méthode proposée améliore celles existantes et propose différentes extensions. Cela nous a permis d'extraire des structures avec une grande précision dans une grande variété d'images naturelles et médicales. La principale contribution de cette thèse est la définition d'un cadre général permettant d'incorporer différents attributs pour résoudre le problème de la segmentation dans des cas pratiques. Cet objectif a été atteint en combinant et en améliorant plusieurs techniques de l'état de l'art:

1. Nous avons introduit une évolution de front basée région issue d'une formulation Bayésienne qui permet d'intégrer des modèles statistiques dans une approche géométrique.

2. La représentation par ensembles de niveaux a été introduite pour définir la fonctionnelle de partitionnement, conduisant naturellement à une évolution de front implicite.

3. Des approximation paramétriques et non-paramétriques des densités de probabilité régions ont été estimées de manière dynamique en parallèle à l'évolution de courbe.

4. Des caractéristiques texture et mouvement ont été extraites et introduites dans ce cadre de travail afin de traiter une plus grande diversité d'images.

5. Une approche implicite a été présentée pour modéliser et intégrer une connaissance à priori sur la forme des objets à extraire.

6. Cette approche a finalement été étendue à des données plus complexes à valeurs matricielles, en particulier au cas des images de diffusion a résonance magnétique.

Tout au long de cette thèse, nous étions concentré sur l'aspect non-supervisé des approches présentées. Afin d'éviter des modèles statistiques trop compliqués dont l'estimation des paramètres aurait pu être un frein important, nous avons essayé

d'identifier et d'extraire les informations importantes qui définissent les régions d'intérêt dans un type d'image donné. Nous y sommes arrivé avec un certain succès puisque nous avons démontré l'efficacité de nos approches en fournissant de meilleurs résultats que de nombreux travaux récents et avec une complexité moindre. Cependant, en vu d'une robustesse accrue, l'apprentissage des quelques paramètres introduits et l'utilisation d'estimateurs plus robustes pourraient être considérés. Nous pensons que ce cadre de travail est une bonne base pour élaborer des systèmes efficaces d'analyse et de reconnaissance d'images.

Nous concluons sur une vision plus générale du problème de segmentation pour les images naturelles. Sur ce type d'images, ce que nous essayons d'accomplir en intégrant différents attributs et de la connaissance à priori, c'est d'imiter le complexe et toujours mystérieux système visuel humain. Les progrès effectués ces dernières années en imagerie cérébrale apportent des moyens inespérés pour étudier ce système complexe et pour comprendre comment nous, humains, effectuons ces taches d'extractions d'objets. Afin de franchir les limites des méthodes d'analyse d'images actuelles et de les rapprocher de nos propres performances, il pourrait être bénéfique de réunir informaticiens et neurologistes afin de créer des modèles motivés biologiquement de plus en plus proche de la réalité [80].

Bibliography

[1] D. Adalsteinsson and J.A. Sethian. A Fast Level Set Method for Propagating Interfaces. *Journal of Computational Physics*, 118(2):269–277, 1995.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, March 1998.

[3] B.A. Ardekani, J. Nierenberg, M.J. Hoptman, D.C. Javitt, and K.O. Lim. Mri study of white matter diffusion anisotropy in schizophrenia. *NeuroReport*, 14(16):2025–2029, November 2003.

[4] G. Aubert, M. Barlaud, O. Faugeras, and Jehan-Besson. Image segmentation using active contours: Calculus of variations or shape gradients? *SIAM Journal of Applied Mathematics*, 63(6):2128–2154, 2003.

[5] G. Aubert and L. Blanc-Féraud. Some remarks on the equivalence between 2D and 3D classical snakes and geodesic active contours. *The International Journal of Computer Vision*, 34(1):19–28, 1999.

[6] D.C. Ballard and C.M. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[7] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.

[8] P.j. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysica*, (66):259–267, 1994.

[9] M.O. Berger. *Les Contours Actifs: Modelisation, Comportement et Convergence*. PhD thesis, Institut National Polytechnique de Lorraine, Fev. 1991.

[10] Paul J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.

[11] J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, August

1991. Report LiTH-ISY-I-0828 1986 and Report LiTH-ISY-I-1148 1990, both at Computer Vision Laboratory, Linköping University, Sweden.

[12] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. Mr imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders. *Radiology*, pages 401–407, 1986.

[13] J. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report ICSI-TR-97-021, International Computer Science Institute, Berkeley, California, 1997.

[14] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 232–237, Santa Barbara, California, June 1998. IEEE Computer Society.

[15] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[16] P. Brodatz. *Textures: a Photographic Album for Artists and Designers*. Dover, New York, 1966.

[17] T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. In *10th International Computer Analysis of Images and Patterns*, number 2756 in Lecture Notes on Computer Science., pages 353–360. Springer Verlag, August 2003.

[18] T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion unsupervised segmentation incorporating colour, texture, and motion unsupervised segmentation incorporating colour, texture, and motion. INRIA Ressearch Report RR-4760, INRIA, March 2003.

[19] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In *DAGM-Symposium*, pages 446–453, 2002.

[20] T. Brox and J. Weickert. Level set based image segmentation with multiple regions. In *26th DAGM*, pages 415–423, Tübingen, Germany, August 2004. In Pattern Recognition, Springer LNCS 3175, C.-E. Rasmussen, H. Bülthoff, M. Giese, and B. Schölkopf (Eds.).

[21] T. Brox and J. Weickert. A tv flow based local scale measure for texture discrimination. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, volume 2, pages 578–590, Prague, Czech Republic, May 2004. Springer–Verlag.

[22] T. Brox, M. Welk, G. Steidl, and J. Weickert. Equivalence results for TV diffusion and TV regularisation. In L. D. Griffin, editor, *Scale-Space Theories in Computer Vision*, Lecture Notes in Computer Science, pages 86–100. Springer, Berlin, June 2003.

[23] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematica*, 4:347–378, 1986.

[24] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.

[25] M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9, 1991.

[26] J.S.W. Campbell, K. Siddiqi, B.C. Vemuri, and G.B Pike. A geometric flow for white matter fibre tract reconstruction. In *IEEE International Symposium on Biomedical Imaging Conference Proceedings*, pages 505–508, July 2002.

[27] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours. *Nümerische Mathematik*, 66:1–31, 1993.

[28] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *The International Journal of Computer Vision*, 22(1):61–79, 1997.

[29] T. Chan, B.Y. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11:130–141, 2000.

[30] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001.

[31] T. Chan and L. Vese. A level set algorithm for minimizing the Mumford–Shah functional in image processing. *1st IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 161–168, 2001.

[32] T. Chan and W. Zhu. Level set based shape prior segmentation. *UCLA Research Report 03-66*, November 2003.

[33] Y. Chen, H. Thiruvenkadam, H. Tagare, F. Huang, and D. Wilson. On the Incorporation of Shape Priors int Geometric Active Contours. In *IEEE VLSM*, pages 145–152, Vancouver, Canada, 2001.

[34] David L. Chopp. Computing minimal surfaces via level set curvature flow. *Journal of Computational Physics*, 106:77–91, 1993.

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

[36] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *pami*, 15(11):1131–1147, November 1993.

[37] T. Cootes and C. Taylor. Mixture model for representing shape variation. *Image and Vision Computing*, 17:567–574, 1999.

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

[39] M.G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton–Jacobi equations. *Trans. AMS*, 277:1–43, 1983.

[40] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, September 2003. Special Issue on Kernel and Subspace Methods in Computer Vision.

[41] D. Cremers, C. Schnoerr, and J. Weickert. Diffusion-snakes: Combining statistical shape knowledge and image information in a variational framework. In *1st IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 2001.

[42] D. Cremers and S. Soatto. A pseudo distance for shape priors in level set segmentation. In N. Paragios O. Faugeras, editor, *2nd IEEE Intl. Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, 2003.

[43] G. Cross and A. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[44] R. Curwen and A. Blake. Dynamic contours : Real-time active splines. In Andrew Blake and Alan Yuille, editors, *Active Vision*, chapter III, pages 39–58. The MIT Press, 1993.

[45] A.P. Dempster, N.M. Laird, and D.B. Rubin. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

[46] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *The International Journal of Computer Vision*, 1(2):167–187, May 1987.

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

[48] 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.

[49] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM Journal on Numerical Analysis*, 37(2):646–664, 2000.

[50] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification, 2nd edition*. Wiley Interscience, 2000.

[51] V. Epanechnikov. *Teoriya Veroyatnostej i Ee Primeneniya*, 14:156–162, 1969.

[52] P.S. Eriksen. Geodesics connected with the fisher metric on the multivariate manifold. Technical Report 86-13, Institute of Electronic Systems, Aalborg University, 1986.

[53] M. Fischler and R. Elschlager. The representation and matching of pictorial images. *IEEE Transactions on Computers*, 22:67–92, 1973.

[54] P.T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. Computer Vision Approaches to Medical Image Analysis*, Prague, May 2004.

[55] M. A. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centers of circular features. In *Proceedings of the Intercommission Workshop of the International Society for Photogrammetry and Remote Sensing*, Interlaken, Switzerland, 1987.

[56] W. Förstner and B. Moonen. A metric for covariance matrices. Technical report, Stuttgart University, Dept. of Geodesy and Geoinformatics, 1999.

[57] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, X(IV):215–310, 1948.

[58] P. Fua and Y.G. Leclerc. Model driven edge detection. *Machine Vision and Application*, 3:45–56, 1990.

[59] D. Gabor. Theory of communications. *IEE Proceedings*, 93(26), 1946.

[60] M. Gastaud, M. Barlaud, and G. Aubert. Combining shape prior and statistical features for active contour segmentation. *IEEE TCSVT special session on Audio and Video Analysis for Interactive Multimedia Services*, 14(5):726–734, May 2004.

[61] G. Gerig, O. Kubler, R. Kikinis, and F. Jolesz. Nonlinear anisotropic filtering of mri data. *IEEE TMI*, 11(2):221–231, 1992.

[62] J. Gomes and O. Faugeras. Reconciling distance functions and level sets. *Journal of Visual Communication and Image Representation*, 11:209–223, 2000.

[63] G.H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995.

[64] M. Hassner and J. Sklansky. The use of Markov random fields as models of texture. *Computer Graphics and Image Processing*, 12:357–370, 1980.

[65] R. J. Hathaway. Another interpretation of the em algorithm for mixture distributions. *Statistics and Probability Letters*, 4:53–56, 1986.

[66] A. Herbulot, S. Jehan-Besson, M. Barlaud, and G. Aubert. Shape gradient for image segmentation using information theory. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 21–24, Montreal, Canada, May 2004.

[67] B.K. Horn and B.G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[68] S. Jehan-Besson, M. Barlaud, and G. Aubert. Dream2s: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation. *ijcv*, 53(1):45–70, June 2003.

[69] Stéphanie Jehan-Besson, Michel Barlaud, Gilles Aubert, and Olivier Faugeras. Shape gradients for histogram segmentation using active contours. In *Proceedings of the 9th International Conference on Computer Vision*, pages 408–415, Nice, France, 2003. IEEE Computer Society, IEEE Computer Society Press.

[70] L. Jonasson, P. Hagmann, X. Bresson, R. Meuli, O. Cuisenaire, and J.P Thiran. White matter mapping in dt-mri using geometric flows. In *EUROCAST*, pages 585–596, 2003.

[71] S. Joshi, A. Srivastava, W. Mio, and X. Liu. Hierarchical organization of shapes for efficient retrieval. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, volume 3, pages 570–581, Prague, Czech Republic, May 2004. Springer–Verlag.

[72] B. Julesz. Information theory and statistical mechanisms. *IRE Trans. of Information Theory*, IT-8:620–630, 1962.

[73] H. Karcher. Riemannian centre of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[74] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision*, pages 259–268, London, June 1987.

[75] S. L. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for wide range edge sharpening. *Inverse Problems*, 18:175–190, January 2002.

[76] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proceedings of the 5th International Conference on Computer Vision*, pages 810–815, Boston, MA, June 1995. IEEE Computer Society Press.

[77] 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.

[78] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *IEEE Internation Conference on Image Processing*, pages 797–800, September 2002.

[79] E. Klassen, A. Srivastava, W. Mio, and S.H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[80] I. Kokkinos, R. Deriche, P. Maragos, and O. Faugeras. A biologically motivated and computationally tractable model of low and mid-level vision tasks. In *Proceedings Eighth European Conference on Computer Vision,*, Prague, May 2004.

[81] Y.G. Leclerc. Constructing simple stable description for image partitioning. *The International Journal of Computer Vision*, 3(1):73–102, 1989.

[82] C. Lenglet, R. Deriche, and O. Faugeras. Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech Republic, May 2004. Springer–Verlag.

[83] C. Lenglet, M. Rousson, and R. Deriche. Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. In *Proc. 7th Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, Saint-Malo, France, September 2004.

[84] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on multivariate normal distributions: A geometric approach and its application to diffusion tensor mri. Technical Report 5442, INRIA, June 2004.

[85] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Toward segmentation of 3d probability density fields by surface evolution: Application to diffusion mri. Technical Report 5243, INRIA, June 2004.

[86] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 316–323, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.

[87] Tony. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[88] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[89] T. Brox M. Rousson and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. INRIA Research Report RR-4695, INRIA, January 2003.

[90] J. Malik and J. Perona. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A*, 7:923–932, 1990.

[91] R. Malladi, J.A. Sethian, and B.C. Vemuri. A topology independent shape modeling scheme. *SPIE*, 2031:246, 1993.

[92] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, February 1995.

[93] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.*, 315:69–87, 1989.

[94] S. Marcelja. Mathematical description of the response of simple cortical cells. *J. Opt. Soc. Amer.*, 70:1297–1300, 1980.

[95] Tim McInerney and Demetri Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4:73–91, 2000.

[96] K.D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion nmr imaging using stimulated echoes. *J. Magn. Reson.*, 64:479–486, 1985.

[97] M. Middendorf and H.-H. Nagel. Empirically convergent adaptive estimation of grayvalue structure tensors. In *DAGM-Symposium*, pages 66–74, 2002.

[98] W. Mio, A. Srivastava, and X. Liu. Learning and bayesian shape extraction for object recognition. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, volume 4, pages 62–73, Prague, Czech Republic, May 2004. Springer–Verlag.

[99] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *The International Journal of Computer Vision*, 19(1):29–55, July 1996.

[100] M. Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.

[101] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications (to appear)*, 2004.

[102] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE.

[103] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.

[104] H.-H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of OF-fields. In Hans Burkhardt and Bernd Neumann, editors, *Proceedings of the 5th European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 86–102, Freiburg, Germany, June 1998. Springer–Verlag.

[105] S. Osher and N. Paragios, editors. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer-Verlag, August 2002.

[106] 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.

[107] N. Paragios and R. Deriche. Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 13(1/2):249–268, march/june 2002.

[108] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *The International Journal of Computer Vision*, 46(3):223, 2002.

[109] N. Paragios, M. Rousson, and V. Ramesh. Knowledge-based registration and segmentation of the left ventricle: A level set approach. In *Proc. IEEE Workshop on Applications of Computer Vision*, pages 37–43, Orlando, Florida, USA, December 2002.

[110] N. Paragios, M. Rousson, and V. Ramesh. Matching distance functions: A shape-to-area variational approach for global-to-local registration. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of the 7th European Conference on Computer Vision*, volume 2, pages 775–789, Copenhagen, Denmark, May 2002. Springer–Verlag.

[111] N. Paragios, M. Rousson, and V. Ramesh. Non-rigid registration using distance functions. *cviu*, 89(2-3):142–165, February 2003.

[112] E. Parzen. On the estimation of probability density function. *Ann. Math. Statist.*, 33:1065–1076, 1962.

[113] X. Pennec. Probabilities and statistics on riemannian manifolds: A geometric approach. Technical Report 5093, INRIA, January 2004.

[114] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A riemannian framework for tensor computing. Research Report 5255, INRIA, July 2004.

[115] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[116] J. Portilla and E.P. Simoncelli. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *The International Journal of Computer Vision*, 40(1):49–71, 2000.

[117] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[118] A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.

[119] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[120] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based space. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 699–704, Madison, Wisconsin, USA, June 2003.

[121] M. Rousson, T. Brox, and R. Deriche. Segmentation active et non supervisée d'images texturées basée sur le tenseur de structure non-linéaire. In *TAIMA'2003*, 2003.

[122] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 56–62, Orlando, Florida, December 2002.

[123] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. RR 4515, INRIA, July 2002.

[124] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor mri segmentation. In *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, Prague, May 2004.

[125] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of the 7th European Conference on Computer Vision*, volume 2, pages 78–92, Copenhagen, Denmark, May 2002. Springer–Verlag.

[126] M. Rousson, N. Paragios, and R. Deriche. Active shape models from a level set perspective. Technical Report 4984, INRIA, October 2003.

[127] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[128] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi. Texture segmentation via a diffusion-segmentation scheme in the gabor feature space. In *Proc. Texture 2002, 2nd International Workshop on Texture Analysis and Synthesis*, Copenhagen, June 2002.

[129] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zérubia. A variational model for image classification and restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):460–472, May 2000.

[130] B. Sandberg, T. Chan, and L. Vese. A level-set and Gabor-based active contour algorithm for segmenting textured images. Technical Report 39, Math. Dept. UCLA, Los Angeles, USA, July 2002.

[131] D. W. Scott. *Multivariate Density Estimation : Theory, Practice, and Visualization*. Wiley-Interscience, New York, 1992.

[132] J.A. Sethian. Theory, algorithms, and applications of level set methods for propagating interfaces. Technical Report PAM-651, Center for Pure and Applied Mathematics, University of California, Berkeley, August 1995. To appear Acta Numerica.

[133] K. Siddiqi, Y. Lauzière, A. Tannenbaum, and S. Zucker. Area and length minimizing flows for shape segmentation. *ieeetip*, 7:433–443, 1998.

[134] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[135] M. S. Silverman, D. H. Grosof, R. L. De Valois, and S. D. Elfar. Spatial-frequency organization in primate striate cortex. *Proceedings of the National Academy of Sciences of the USA*, 86(2):711–715, 1989.

[136] P. Simoncelli, W. Freeman, H. Adelson, and Heeger J. Shiftable multiscale transforms. *IEEE trans. on Information Theory*, 38:587–607, 1992.

[137] L.T. Skovgaard. A riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11:211–233, 1984.

[138] C. Sotak. The role of diffusion tensor imaging (dti) in the evaluation of ischemic brain injury. *NMR Biomed.*, 15:561–569, 2002.

[139] L.H. Staib and J.S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, November 1992.

[140] E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42:288–292, 1965.

[141] C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16:70–91, 1999.

[142] M. Sussman, P. Smereka, and S. Osher. A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow. *J. Computational Physics*, 114:146–159, 1994.

[143] A. Tsai, A. Yezzi, et al. Model-based curve evolution technique for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 463–468, December 2001.

[144] A. Tsai, A. Jr. Yezzi, and A.S. Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. *ieeetip*, 10(8):1169–1186, August 2001.

[145] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai Marriott, Hawaii, December 2001.

[146] D. Tschumperlé and R. Deriche. Regularization of orthonormal vector sets using coupled PDE's. In *IEEE Workshop on Variational and Level Set Methods*, pages 3–10, Vancouver, Canada, July 2001.

[147] D. Tschumperlé and R. Deriche. Variational frameworks for DT-MRI estimation, regularization and visualization. In *Proceedings of the 9th International*

Conference on Computer Vision, Nice, France, 2003. IEEE Computer Society, IEEE Computer Society Press.

[148] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE's : A common framework for different applications. In *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin (United States), June 2003.

[149] V. I. Tsurkov. An analytical model of edge protection under noise suppression by anisotropic diffusion. *Journal of Computer and Systems Sciences International*, 39(3):437–440, 2000.

[150] D.S. Tuch, T.G. Reese, M.R. Wiegell, and V.J. Wedeen. Diffusion mri of complex neural architecture. *Neuron*, 40:885–895, December 2003.

[151] M.R. Turner. Texture discrimination by Gabor functions. *Biological Cybernetics*, 55:71–82, 1986.

[152] C. J. Twining and C. J. Taylor. The Use of Kernel Principal Component Analysis to Model Data Distributions. *Pattern Recognition*, 36(1):217–227, 2004.

[153] R. Veltkamp and Hagedoorn M. State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands, 1999.

[154] B. Vemuri, Y. Chen, M. Rao, T. McGraw, T. Mareci, and Z. Wang. Fiber tract mapping from diffusion tensor mri. In *1st IEEE Workshop on Variational and Level Set Methods in Computer Vision (VLSM'01)*, July 2001.

[155] L. A. Vese and T. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *The International Journal of Computer Vision*, 50(3):271–293, 2002.

[156] Y. Wang and L. Staib. Elastic Model Based Non-rigid Registration Incorporating Statistical Shape Information. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 1162–1173, 1998.

[157] Z. Wang and B.C. Vemuri. An affine invariant tensor dissimilarity measure and its application to tensor-valued image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC., June 2004.

[158] Z. Wang and B.C. Vemuri. Tensor field segmentation using region based active contour model. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech Republic, May 2004. Springer–Verlag.

[159] Z. Wang, B.C. Vemuri, Y. Chen, and T. Mareci. Simultaneous smoothing and estimation of the tensor field from diffusion tensor mri. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 461–466, Madison, Wisconsin (United States), June 2003.

[160] J. Weickert. A review of nonlinear diffusion filtering. *Scale-Space Theory in Computer Vision, Lecture Notes in Comp. Science (Springer, Berlin)*, 1252:3–28, 1997. Invited Paper.

[161] J. Weickert and B. Benhamouda. *A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox*, pages 1–10. Advances in Computer Vision, Springer, Wien, 1997.

[162] J. Weickert and T. Brox. Diffusion and regularization of vector and matrix-valued images. Technical report, Universitat des Saarlandes, 2002.

[163] J. Weickert, A. Bruhn, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. Technical Report 82, Dept. of Mathematics, Saarland University, Saarbrücken, Germany, April 2003.

[164] J. Weickert, B.M. ter Haar Romeny, and M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, March 1998.

[165] M.R. Wiegell, D.S. Tuch, H.W.B. Larson, and V.J. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *NeuroImage*, 19:391–402, 2003.

[166] C. Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[167] A. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Proceedings of the 7th International Conference on Computer Vision*, volume II, pages 898–903, Kerkyra, Greece, September 1999. IEEE Computer Society, IEEE Computer Society Press.

[168] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33:116–125, 1986.

[169] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), January 2001.

[170] H.K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(0167):179–195, 1996.

[171] 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(9):884–900, September 1996.

[172] S.C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (frame). *The International Journal of Computer Vision*, 27(2):1–20, 1998.

[173] L. Zhukov, K. Museth, D. Breen, R. Whitaker, and A.H. Barr. Level set segmentation and modeling of dt-mri human brain data. *Journal of Electronic Imaging*, 2003.

[174] O.C. Zienkiewicz and R.L. Taylor. *The Finite element method, 4th ed.* Mac Graw Hill, 1987.