Prior Knowledge for Targeted Object Segmentation in Medical Images

by

Seyed Masoud Nosrati

M.Sc., Amirkabir University of Technology (Tehran Polytechnic), 2008
B.Sc., Shahrood University of Technology, 2005

Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2015

Approval

Name:	Seyed Masoud Nosrati
Degree:	Doctor of Philosophy (Computing Science)
Title:	*Prior Knowledge for Targeted Object Segmentation in Medical Images*
Examining Committee:	**Dr. Mark S. Drew** (chair) Professor, Computing Science, Simon Fraser University

Dr. Ghassan Hamarneh
Senior Supervisor
Professor,
Computing Science, Simon Fraser University

Dr. Greg Mori
Supervisor
Associate Professor,
Computing Science, Simon Fraser University

Dr. Ping Tan
Internal Examiner
Assistant Professor,
Computing Science, Simon Fraser University

Dr. Klaus D. Tönnies
External Examiner
Professor,
Department of Simulation and Graphics, Faculty of Computer Science, University of Magdeburg

Date Defended: March 25, 2015

Abstract

Medical image segmentation, the task of partitioning an image into meaningful parts, is an important step toward automating medical image analysis and is at the crux of a variety of medical imaging applications, such as computer aided diagnosis, therapy planning and delivery, and computer aided interventions. However, existence of noise, low contrast and objects' complexity in medical images preclude ideal segmentation. Incorporating prior knowledge into image segmentation algorithms has proven useful for obtaining more accurate and plausible results on targeted objects segmentation.

In this thesis, we develop novel techniques to augment optimization-based segmentation frameworks with different types of prior knowledge to identify and delineate only those objects (*targeted objects*) that conform to specific geometrical, topological and appearance priors. These techniques include employing prior knowledge to segment multi-part objects with part-configuration constraints and encoding priors based on images acquired from different imaging equipment and of differing dimensions. Our objective is to satisfy two important aspects in optimization-based image segmentation: (1) fidelity-optimizability trade-off, and (2) space and time complexity.

Particularly, in our first contribution, we adopt several prior information to build a faithful objective function unconcerned about its convexity to segment potentially overlapping cells with complex topology. In our second contribution, we improve the space and time complexity and augment the level sets framework with the ability to handle geometric constraints between boundaries of multi-region objects. In our first two contributions we opt for ensuring the objective function is flexible enough (even if it is non-convex) to accurately capture the intricacies of the segmentation problem. In our third contribution, we focus on optimizability. We propose a convex formulation to augment the popular Mumford-Shah model and develop a new regularization term to incorporate similar geometrical and distance prior as our second contribution while maintaining global optimality. Lastly, we efficiently incorporate different types of priors based on images acquired from different imaging equipment (different modalities) and of dissimilar dimensions to segment multiple objects in intraoperative multi-view endoscopic videos. We show how our technique allows for the inclusion of laparoscopic camera motion model to stabilize the segmentation.

Keywords: Prior knowledge; medical image segmentation; energy minimization; evolutionary computation; tribes-based genetic algorithm; level sets; geometric interaction; convex relaxation; 3D to 2D pose estimation; cell segmentation; microscopy; histology; endoscopy; augmented reality; robotic surgery

Dedication

To my parents Banafsheh and Hassan, my brother Kayvan, and my grandparents for their unconditional and immeasurable love and support.

Acknowledgements

This dissertation would not have been accomplished without the help and hard work of many wonderful people.

First and foremost, I would like to express my sincere gratitude to my senior supervisor, Dr. Ghassan Hamarneh, who mentored me in many different ways and taught me how to think critically. Without your patience, encouragement, support, constructive and illuminating discussions, and high scientific standards, this thesis would not have been completed.

I would also like to express my sincere appreciation to my supervisor, Dr. Greg Mori, who generously supported me on both professional and personal levels.

I deeply appreciate Dr. Klaus D. Tönnies, the defence external examiner, Dr. Ping Tan, the internal examiner, Dr. Mark Drew, the defence chair and Dr. Brian Funt the depth exam examiner for their valuable time as members of my defence committee.

I would like to thank my friends and collogues in the robotic surgery team, Dr. Rafeef Abugharbieh, Alborz Amir-Khalili, Jeremy Kawahara, and Dr. Jean-Marc Peyrat for their friendship, great discussions, and fruitful collaboration.

I have been fortunate to work in an amazing group and with wonderful friends in Medical Image Analysis Lab (MIAL) at SFU. Especially, I wish to express my gratitude to Dr. Lisa Tang for her full support and kindness in sharing her experiences with me and helping me to improve my writing. Many thanks go to Brian Booth and Jeremy Kawahara who generously took time out of their busy schedules patiently to read and correct my writing. I would like to thank Dr. Shawn Andrews for his helpful tips and suggestions and for the many inspirational discussions we had. I would also like to thank Aïcha bentaieb, Colin Brown, Payam Ahmadvand, Dr. Hengameh Mirzaalian, Ismail Khater, Mengliu Zhao and other former and current members of MIAL for their wonderful support, friendship, and great advice. Thank you all for being such great labmates. Many thanks to my former and current flatmates Ashgan Fararooy, Sepehr Attarchi, Ali Shagerdmootab, and Dr. Vahid Zakeri for their patience, wonderful tips, and continued support.

My Ph.D. has been supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Qatar robotic surgery centre (QRSC) and Qatar National Research Fund (a member of the Qatar Foundation). I thank these organizations for their funding support.

Finally, I would like to thank my parents, Banafsheh Nabavi and Hassan Nosrati, for their unconditional love, support, and sacrifices. Words are not enough to express all my gratitude. Without their encouragement, patience and care I could not have achieved this work.

Contents

List of Tables

List of Figures

List of Acronyms

ACWOE	Active contours without edges
CCMF	Combinatorial continuous max-flow
CFL	Courant-Friedrichs-Lewy
CT	Computed tomography
DCT	Discrete cosine transform
dPET	Dynamic positron emission tomography
DSC	Dice similarity coefficient
DT MRI	Diffusion tensor MRI
ES and ED	End-systole and end-diastole
FEM	Finite element method
FPS	Frame per second
GA	Genetic algorithm
GC	Graph cuts
GLOH	Gradient location and orientation histogram
GMM	Gaussian mixture models
GPU	Graphics processing unit
GVF	Gradient vector field
HOG	Histogram of oriented gradient
ICA	Independent component analysis
ILR	Isometric log-ratio
KDE	Kernel density estimation

Note: List of Acronyms to be continued on next page.

LBP	Local binary pattern
LogOdds	Logarithm of odds
LP	Linear programming
LV	Left ventricle
MIS	Medical image segmentation
MRA	Magnetic resonance angiography
MRF	Markov random field
MRI	Magnetic resonance imaging
PCA	Principal component analysis
PDE	Partial differential equations
PET	Positron emission tomography
PFA	Principal factor analysis
QPBO	Quadratic pseudo-boolean optimization
RMIS	Robotic minimally invasive surgery
RV	Right ventricle
SIFT	Scale invariant feature transform
SNR	Signal to noise ratio
TAC	Time activity curve
US	Ultrasound

List of Notations

Ω	Image domain in the Cartesian coordinate system
$I : \Omega \rightarrow \mathbb{R}^m$	Image function
I_i	Intensity/color value associated with pixel i in image I
$\boldsymbol{x} = (x, y)$ or $\boldsymbol{x} = (x, y, z)$	Spatial coordinate
$\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\}$	Set of spatial coordinate points
S_i	i^{th} segmentation
S_i^*	Optimal segmentation of i^{th} region
$\boldsymbol{S} = \{S_1, \cdots, S_k\}$	Segmentations of all regions/objects
$\boldsymbol{S}^* = \{S_1^*, \cdots, S_k^*\}$	Optimal segmentations of all regions/objects
E	Energy function
$\boldsymbol{\mathcal{R}}$, $\mathcal{R}$	Regularization term
$\boldsymbol{\mathcal{D}}$, $\mathcal{D}$	Data energy term
$\mathcal{G}(\mathcal{P}, \mathcal{E})$	Graphical model
$\mathcal{P}$	Set of nodes/vertices
$\mathcal{E}$	Set of edges
$p, q \in \mathcal{P}$	Indices to nodes representing spatial coordinate $\boldsymbol{x}$
f_p^i	Random binary variable associated with pixel p and label/region i
D	Dissimilarity function or unary penalty term
V_{pq}	Pairwise penalty term
$\mathcal{L} = \{l_1, \cdots, l_k\}$	Label set
h	Entropy
ϕ	Level set function
$\boldsymbol{J}$	Jacobian

Note: List of notations to be continued on next page.

u	Labeling function
ρ	Regional term
φ	Super level set function
$\boldsymbol{P}_i$	The i^{th} principal component
$\mathcal{N}$	Neighbourhood pixels
Du	Distributional derivative of u
δ	Dirac delta function
H	Heaviside function
$\gamma = [\gamma_{min}, \gamma_{max}]$	Continuous label set
$\boldsymbol{X}_i^{pre,\ell}$	i^{th} point of ℓ^{th} 3D model in the preoperative domain
$\boldsymbol{X}_i^{srg,\ell}$	i^{th} point of ℓ^{th} 3D model in the surgical domain
$\boldsymbol{x}_i^m$	i^{th} 2D point in m^{th} image (I_m)
T	Transformation matrix
π	3D to 2D projection function
$\boldsymbol{\xi} = \{\boldsymbol{\xi}_q, \boldsymbol{\xi}_w\}$	Set of pose and shape parameters
$\boldsymbol{\xi}_q = \{q_1, \cdots, q_n\}$	Pose parameters
$\boldsymbol{\xi}_w = \{w1, \cdots, w_m\}$	Shape parameters
$\hat{\boldsymbol{\xi}}$	Predicted pose parameters

Chapter 1

Introduction

1.1 Thesis Context

Image segmentation is the process of partitioning an image into smaller meaningful regions based in part on some homogeneity characteristics. The goal of segmentation is to delineate (extract or contour) target objects for further analysis. Medical image segmentation (MIS) is at the crux of a variety of medical imaging applications, such as computer aided diagnosis, therapy planning and delivery, and computer aided interventions. For example, segmentation of organs or tissue types in medical images is a necessary first step for measuring tumour burden (or volume) from positron emission tomography (PET) or computed tomography (CT) scans [Hatt et al., 2009, Bagci et al., 2013], analyzing vasculature from magnetic resonance angiography (MRA) (e.g. measuring tortuosity) [Bullitt et al., 2003, Yan and Kassim, 2006], grading cancer from histopathology images [Tabesh et al., 2007], performing fetal measurements from prenatal ultrasound (US) [Carneiro et al., 2008], performing augmented reality in robotic image guided surgery [Su et al., 2009, Pratt et al., 2012], building statistical atlas for population studies and voxel-based morphometry [Ashburner and Friston, 2000], etc.

Despite great advances in image segmentation, the accurate automatic (or even semi-automatic) partitioning of medical images with complex configurations of tissues and objects remains a challenging problem. Several traditional segmentation algorithms have been proposed for assigning labels to pixels; these include thresholding [Otsu, 1975, Sahoo et al., 1988], region-growing [Adams and Bischof, 1994, Pohle and Toennies, 2001, Pan and Lu, 2007], and watershed [Vincent and Soille, 1991, Grau et al., 2004, Hamarneh and Li, 2009]. Existence of noise, low contrast and objects complexity in medical images typically cause the aforementioned methods to fail. In addition, all these traditional methods assume that objects' entire appearance have some notion of homogeneity while this is not necessarily the case for complex objects. This is where more elaborate *prior information* about the targeted objects becomes helpful. Many objects in medical images consist of multiple regions, where each region has a meaningful geometric relationship, or interaction, with other regions of the object. For example, in histology and microscopy images, each cell consists of a cell membrane, nucleus and nucleolus, where the cell membrane contains the nucleus, and the nucleus contains nucleolus. These interactions between an object's regions have often been ignored in microscopic histology image segmentation or enforced via some ad-hoc post-processing (e.g. via parameter sensitive morphological operations or thresholding) [Wu et al., 2012, Mao et al., 2006, Al-Kofahi et al., 2010, Yang et al., 2008].

To overcome noise and handle complex objects in medical images, many attempts have been made to incorporate prior knowledge into the task of segmentation, since inclusion of shape, appearance and topological priors have proven useful for obtaining more accurate and plausible segmentation results. The majority of state-of-the-art image segmentation methods are formulated as optimization problems, i.e. energy minimization or maximum-a-posteriori estimation, mainly because of their: 1) formal and rigorous mathematical formulation, 2) availability of the mathematical tools for optimization, 3) capability to incorporate multiple (competing) criteria as terms in the objective function, 4) ability to quantitatively measure the extent by which a method satisfies the different criteria/terms, and 5) ability to examine the relative performance of different solutions.

In energy-based segmentation problems two important aspects need to be considered: (1) the trade off between *fidelity* and *optimizability* [McIntosh and Hamarneh, 2012, Ulén et al., 2013], and (2) space and time complexity. Fidelity describes how faithful the energy function is to the data and how accurate it can model intricate problem details. Optimizability determines how easily we can optimize the objective function and attain the global optimum. Generally, the better the objective function models the problem (e.g. by using more prior information), the more complicated the objective function becomes and the harder it is to optimize. If we instead sacrifice fidelity to obtain a globally optimizable objective function, the solution (albeit global) might not be accurate enough for our segmentation purpose. Ideally, an energy function is designed in a way that is faithful to the underlying segmentation problem and easy to be optimized at the same time.

The second aspect is the space and time complexity. In practice, the space and time complexity become important when we deal with very large images which is often the case in medical image analysis, e.g. 100s of millions pixels of microscopy images, and 3D volume scans.

1.2 Thesis Contributions

In this thesis, we will explore incorporating prior knowledge into medical image segmentation frameworks to better model the underlying segmentation problems (increase the fidelity) while considering the optimizability and space and time complexity. By incorporating prior expert information into segmentation frameworks, we are able to segment specific objects (targeted objects) solely and improve the accuracy and plausibility of the results especially when training data is unavailable. Our objective is to explore the two aforementioned aspects in optimization-based image segmentation: (1) fidelity-optimizability trade-off, and (2) space and time complexity.

Particularly, after giving a concrete overview of the different types of prior expert knowledge that have been utilized to improve image segmentation in Chapter 2, in our first contribution (Chapter 3), we adopt several prior information to segment *overlapping multi-region* cells in microscopy and histology images [Nosrati and Hamarneh, 2013]. Most existing methods have only considered simple structured cells and ignored their complex composition [Ali et al., 2011, Cheng et al., 2011, Bernardis et al., 2011]. In addition, previous works addressed cell overlapping, for single-region cells using post-processing techniques [Wu et al., 2012, Yang et al., 2008, Mao et al., 2006] (e.g. finding connected components and using parameter sensitive morphological operations [Mao et al., 2006]). Unlike previous works, in this work we build a faithful objective function, unconcerned about its convexity, to segment potentially *overlapping multi-region* cells with complex topology. Our framework allows for leveraging a variety of expert knowledge or priors like shape, appearance, edge polarity, pose, topology, adjacency, and user interaction. To optimize such complex objective function and to deal with imminent problems like initialization and local optima, we adopt a global optimization evolutionary computation method, genetic algorithm (GA), which can attain solutions close to the global optimum, does not require Euler Lagrangian or energy gradient calculations, is generally parallelizable, and allows for arbitrarily complex objective functions. Finally, to deal with the spatially recurring aspect in cell segmentation, we use genetic algorithms with *tribes* [Turner et al., 1996] to obtain multiple distinct solutions for our framework.

In our second contribution (Chapter 4), we improve the space and time complexity and augment the level sets framework with the ability to handle two important and intuitive geometric constraints, *containment* and *exclusion* of regions, along with *distance constraints* between boundaries of multi-region objects [Nosrati and Hamarneh, 2014, Nosrati and Hamarneh, 2015]. The distance constraints imposed in our level sets framework include minimum distance, maximum distance, and attraction/repulsion forces between different regions/surfaces. By applying our framework to diverse medical image segmentation tasks such as segmentation of cells in microscopic images, cardium in magnetic resonance images (MRI), brain in dynamic positron emission tomography (dPET) images, and lunge vessels in computed tomography images, we demonstrate the accuracy and generalizability of our method. In these two contributions we opt for ensuring the objective function is flexible enough (even if it is non-convex) to accurately capture the intricacies of the segmentation problem.

In our third contribution (Capter 5), we turn our attention to optimizability [Nosrati et al., 2013]. We propose a convex formulation to augment the popular Mumford-Shah model [Mumford and Shah, 1989] and develop a new regularization term to incorporate similar geometrical and distance prior as our second contribution while maintaining *global optimality*. Our method is able to handle multiple instances of multi-part objects defined by these geometrical constraints using a *single* labeling function. We show that the proposed convex continuous method is superior to other state-of-the-art methods, including its discrete counterparts, [Delong and Boykov, 2009] and [Ulén et al., 2013], in terms of memory usage and metrication errors[1]. Using a continuous framework provides several advantages over discrete methods: 1) no metrication error; 2) less memory usage; 3) efficient parallelizability, and 4) allowance for sub-pixel resolution.

While our previous contributions enforce geometrical priors into segmentation frameworks, in Chapter 6 we enforce priors based on images acquired from different imaging equipment (different modalities) and of dissimilar dimensions [Nosrati et al., 2014, Nosrati et al., 2015]. Specifically, we develop an efficient technique to segment *multiple objects* in intraoperative *multi-view* endoscopic videos based on priors captured from preoperative data. Our technique imitates the surgeon skill in leveraging information from 3D preoperative data into the analysis of visual cues in the 2D intraoperative data by formulating the problem of finding the 3D pose and *non-rigid* deformations of tissue models driven by features from 2D images. We present a *closed-form solution* for our formulation and demonstrate how it allows for the inclusion of laparoscopic camera *motion model* to stabilize the segmentation in the presence of a large objects occlusion.

We also created an interactive online database (`http://goo.gl/gy9pyn`) to categorize existing works based on the type of prior knowledge they use. Our online database is interactive so that researchers can contribute to keep it up to date.

1.3 Auto-Bibliography

The chapters of this thesis are largely based on the following publications.

Chapter 2

Masoud S. Nosrati and Ghassan Hamarneh. "Incorporating prior knowledge in medical image segmentation: a survey", **Submitted to** *Medical Image Analysis (**MedIA**)*, 2015.

Chapter 3

Masoud S. Nosrati and Ghassan Hamarneh. "Segmentation of cells with partial occlusion and part configuration constraint using evolutionary computation", *Medical Image Computing and Computer-Assisted Intervention (**MICCAI**)*, vol. 8149, pages 461-468, 2013.

---

[1] Metrication error is defined as the artifact which appear in graph-based segmentation methods due to penalizing region boundaries only across axis-aligned edges.

Chapter 4

Masoud S. Nosrati and Ghassan Hamarneh. "Local optimization based segmentation of spatially-recurring, multi-region objects with part configuration constraints", *IEEE Transactions on Medical Imaging (**IEEE TMI**)*, 33(9):1845-1859, 2014.

(**Joint submission**) Zhi Lu, Gustavo Carneiro, Andrew P. Bradley, Daniela Ushizima, Masoud S. Nosrati, Andrea G. C. Bianchi, Claudia M. Carneiro, and Ghassan Hamarneh. "Evaluation of three algorithms for the segmentation of overlapping cervical cells", **submitted to** *Medical Image Analysis (**MedIA**)*.

Masoud S. Nosrati and Ghassan Hamarneh. "Segmentation of overlapping cervical cells: a variational method with star-shape prior", *IEEE International Symposium on Biomedical Imaging (**IEEE ISBI**)*, pages 186-189, 2015.

Masoud S. Nosrati and Ghassan Hamarneh. "A variational approach for overlapping cell segmentation", Overlapping Cervical Cytology Image Segmentation Challenge, in conjunction with *IEEE International Symposium on Biomedical Imaging (**IEEE ISBI**)*, pages 1-2, 2014 (**2nd place among 6 competitors**).

Chapter 5

Masoud S. Nosrati, Shawn Andrews, and Ghassan Hamarneh. "Bounded labeling function for global segmentation of multi-part objects with geometric constraints", *IEEE International Conference on Computer Vision (**IEEE ICCV**)*, pages 2032-2039, 2013.

Chapter 6

Masoud S. Nosrati, Rafeef Abugharbieh, Jean-Marc Peyrat, Julien Abi-Nahed, Osama Al-Alao, Abdulla Al-Ansari, and Ghassan Hamarneh. "Simultaneous multi-tissue segmentation and 3D non-rigid pose estimation in image guided robotic surgery", **Accepted to be published in** *IEEE Transactions on Medical Imaging (**IEEE TMI**)*, 2015.

Masoud S. Nosrati, Jean-Marc Peyrat, Julien Abi-Nahed, Osama Al-Alao, Abdulla Al-Ansari, Rafeef Abugharbieh, and Ghassan Hamarneh. "Efficient multi-organ segmentation in multi-view endoscopic videos using pre-op priors", *Medical Image Computing and Computer-Assisted Intervention (**MICCAI**)*, volume 8674, pages 324-331, 2014.

Chapter 2

Incorporating prior knowledge in MIS: a review

In this chapter, we review works related to our contributions in this thesis. We begin by reviewing the fundamentals of optimization-based image segmentation techniques (Section 2.1). In Section 2.2, we give a concrete overview of the different types of prior knowledge devised to improve image segmentation. Prior information can take many forms: user interaction; appearance models; boundaries and edge polarity; shape models; topology specification; moments (e.g. area/volume and centroid constraints); geometrical interaction and distance prior between different regions/labels; and atlas or pre-known models. We compare the different methods utilizing prior information in image segmentation in terms of the type of prior information utilized, domain of formulation (continuous vs. discrete) and optimization techniques (global vs. local) used. Finally, in Section 2.3, we summarize our notes on the existing works.

2.1 Fundamentals of image segmentation

As mentioned earlier, formulating image segmentation as an optimization problem allows for incorporating multiple criteria and prior information as terms into an objective functional. In this section, we briefly review the fundamentals of optimization-based techniques for MIS.

2.1.1 Optimization-based image segmentation

Given an image $I : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, image segmentation partitions Ω into k disjoint regions $\boldsymbol{S} = \{S_1, \cdots, S_k\} \subset \mathcal{S}$ such that $\Omega = \cup_{i=1}^{k} S_i$ and $S_i \cap S_j = \emptyset$, $\forall i \neq j$. $\mathcal{S}$ is the solution space. The aforementioned partitioning is referred to as a crisp binary (when $k = 2$) or multi-region ($k > 2$) segmentation. In a fuzzy or probabilistic segmentation, each element in Ω (e.g. a pixel) is assigned a vector $\boldsymbol{p}$ of length k quantifying the memberships or probabilities of belonging to each of the k classes, $\boldsymbol{p} = [p_1, p_2, \cdots, p_k]$ where $p_i \geq 0,\ i = 1, \cdots, k$ and $\sum_{i=1}^{k} p_i = 1$. This task of image partitioning can be formulated as an energy minimization problem. An energy function, $E : \mathcal{S} \to \mathbb{R}$, usually consists of several objectives that are divided into two main categories *regularization terms* $\mathcal{R}_i : \mathcal{S} \to \mathbb{R}$ and *data terms* $\mathcal{D}_i : \mathcal{S} \to \mathbb{R}$. The regularization terms correspond to priors on the space of feasible solutions and penalize any deviation from the enforced prior such as shape, length, etc. The data terms measure how strongly a pixel should be associated with specific label/segment. These objectives (regularization and data terms) can then be scalarized as:

$$E(\boldsymbol{S}) = \lambda \sum_i \mathcal{R}_i(\boldsymbol{S}) + \sum_j \mathcal{D}_j(\boldsymbol{S}) \ . \tag{2.1}$$

The optimization problem is then formulated as:

$$\boldsymbol{S}^* = \arg\min_{\boldsymbol{S}} E(\boldsymbol{S}) = \arg\min_{\boldsymbol{S}} \lambda \boldsymbol{\mathcal{R}}(\boldsymbol{S}) + \boldsymbol{\mathcal{D}}(\boldsymbol{S}; I) \ , \tag{2.2}$$

where $\boldsymbol{S}^* = \{S_1^*, \cdots, S_k^*\}$ are the optimal solutions and, for simplicity, $\boldsymbol{\mathcal{R}}$ and $\boldsymbol{\mathcal{D}}$ represent all the regularization and data terms, respectively. λ is a constant weight that balances the contribution/importance of the data term and the regularization term in the minimization problem. One example of such energy is written as:

$$S_1^*, \cdots, S_k^* = \arg \min_{S_1, \cdots, S_k} \left\{ \lambda \sum_{i=1}^{k} \int_{\partial S_i} d\boldsymbol{x} + \sum_{i=1}^{k} \int_{S_i} D_i(\boldsymbol{x}) d\boldsymbol{x} \right\} , \tag{2.3}$$

where the first term (regularization term) measures the perimeter of the segmented regions S_i and penalizes large perimeters, thus favouring smooth boundaries. $D_i(\boldsymbol{x}) : \Omega \to \mathbb{R}$, associated with region S_i, measures how strongly pixel $\boldsymbol{x} \in \Omega$ should be associated with region S_i. In Section 2.2.3, we discuss different types of regularization terms used in image segmentation problems.

An optimization-based image segmentation problem can also be formulated as a maximization problem:

$$\boldsymbol{S}^* = \arg\max_{\boldsymbol{S}} P(\boldsymbol{S}|I) \ , \tag{2.4}$$

where $\boldsymbol{S}^*$ is the optimal segmentation. Using Bayes' theorem, (2.4) can be written as:

$$\boldsymbol{S}^* = \arg\max_{\boldsymbol{S}} \frac{P(I|\boldsymbol{S})P(\boldsymbol{S})}{P(I)} \equiv \arg\max_{\boldsymbol{S}} P(I|\boldsymbol{S})P(\boldsymbol{S}). \tag{2.5}$$

In (2.4) and (2.5), $P(\boldsymbol{S}|I)$ is the *posterior probability* that defines the degree of belief in $\boldsymbol{S}$ given the evidence I (or some features of I), $P(I|\boldsymbol{S})$ is the image *likelihood* measuring the probability of the evidence in I given the segmentations $\boldsymbol{S}$, and $p(\boldsymbol{S})$ is the *prior probability*

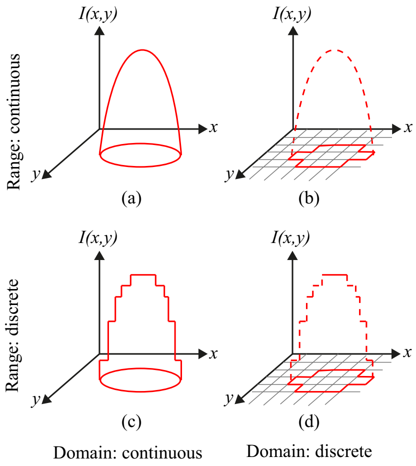


Figure 2.1: Image as a mapping. Continuous vs. discrete domain and image values.

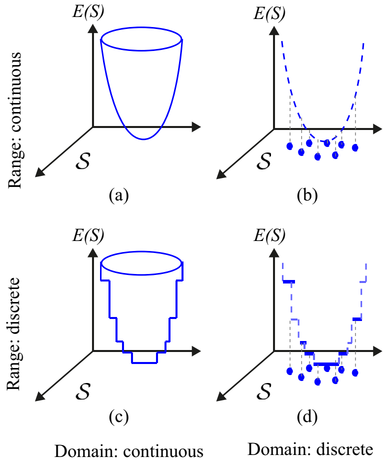


Figure 2.2: Energy function: continuous vs. discrete. $\mathcal{S}$ is the space of possible segmentations.

that indicates the initial (prior to observing I) degree of belief in $\boldsymbol{S}$. Maximizing the posterior probability (2.5) is equivalent to minimizing its negative logarithm:

$$\boldsymbol{S}^* = \arg\min_{\boldsymbol{S}} - \log P(I|\boldsymbol{S}) - \log P(\boldsymbol{S}). \tag{2.6}$$

The probability (2.6) and energy (2.2) notations are related via the Gibbs or Boltzmann distribution. Ignoring the Boltzmann's constant and thermodynamic temperature, as they do not affect the optimization, by substituting $P(I|\boldsymbol{S}) \propto e^{-\boldsymbol{\mathcal{D}}(\boldsymbol{S};I)}$ and $P(\boldsymbol{S}) \propto e^{-\lambda \boldsymbol{\mathcal{R}}(\boldsymbol{S})}$ into (2.6) we obtain (2.2).

To avoid terminological confusion, we emphasize that to improve a segmentation, prior knowledge can be incorporated into either the regularization term or the data term (or both). Hence, the prior knowledge should not be confused with the prior probability in (2.5).

2.1.2 Domain of formulation: continuous vs. discrete

In general, a segmentation problem can be formulated in a spatially discrete or continuous domain. In the community that advocates continuous methods, it is assumed that the world we live in is a continuous world (continuous Ω). However, images captured by digital cameras are discrete both in space and color/intensity. The discretization in space is called *sampling* (discrete Ω) and the discretization in color/intensity or value space is called *quantization*. Given this categorization, we have four different cases for image representation (Figure 2.1).

Energy optimization problems can also be formulated in a discrete or continuous domain. Depending on the solution space (discrete vs. continuous) and the energy values, four possible cases can be considered for an energy functional (Figure 2.2). In the spatially discrete setting, the energy function is defined over a set of finite variables (nodes $\mathcal{P} \subset \Omega$ and edges), leading to Markov random field (MRF) formulation. In MRF formulations, solution is often calculated using graph cut methods, e.g. max-flow/min-cut algorithms or graph partitioning methods. In the spatially continuous setting, the optimality conditions for the continuous energy functional are written in terms of a set of partial differential equations (PDE). The minimization problem in (2.3) is a continuous version of a multi-region segmentation functional, often called *minimal partition problem* in the PDE community [Nieuwenhuis et al., 2013]. Note that in Figure 2.2, the objective function is a cost or an energy function that has to be minimized. Nevertheless, an objective function can also be a fitness or utility function that has to be maximized.

In the discrete setting, the segmentation task begins with an undirected graph, $\mathcal{G}(\mathcal{P}, \mathcal{E})$, that is composed of vertices $\mathcal{P}$ and undirected edges $\mathcal{E}$. Each node of the graph ($p \in \mathcal{P}$) represents a random variable (f_p^i) taking on different labels ($i \in \mathcal{L} = \{l_1, \cdots, l_k\}$) and each edge encodes the dependency between neighbouring variables. The corresponding optimization problem in the discrete domain is

$$\begin{aligned} &\min_{\boldsymbol{f}} \left\{ \sum_{pq \in \mathcal{N}^i} V(f_p^i, f_q^i) + \sum_{p \in \mathcal{P}} D_p(f_p) \right\} \\ &\text{s.t.} \quad \sum_{i \in \mathcal{L}} f_p^i = 1, \quad \forall p \in \mathcal{P} \ , \end{aligned} \tag{2.7}$$

where V (the pairwise term) is the regularization term that encourages spatial coherence by penalizing discontinuities between neighbouring pixels, D (the unary term) is the data penalty term, $\boldsymbol{f} \in \mathbb{B}^{\mathcal{L} \times \mathcal{P}}$ are the binary variables ($f_p^i = 1$ if pixel $p \in \mathcal{P}$ belongs to region $i \in \mathcal{L}$ and $f_p^i = 0$ otherwise) and $\mathcal{N}^i$ is the neighbourhood which is typically defined as nearest neighbour grid connectivity.

There are several advantages and drawbacks associated with discrete and continuous methods:

- **Parameter tuning:** in the continuous domain, PDE-based approaches typically require setting a step size during the optimization procedure. In the PDE community, it is stated that the Euler-Lagrange equation provides a sufficient condition for a stationary point of the energy functional. Let u be a differentiable labeling function in a continuous domain and $E(u)$ be an energy functional. The Euler-Lagrange equation applied to E is:

$$\frac{\partial E}{\partial u} - \frac{d}{dx}\left(\frac{\partial E}{\partial u_x}\right) - \frac{d}{dy}\left(\frac{\partial E}{\partial u_y}\right) = 0 \ , \tag{2.8}$$

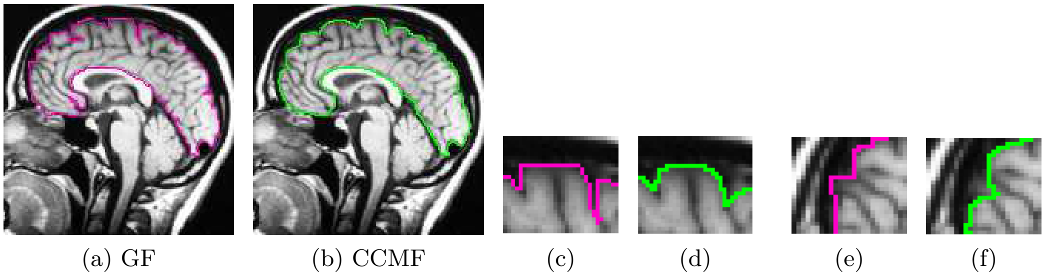

(a) GF (b) CCMF (c) (d) (e) (f)

Figure 2.3: Metrication artifacts. Brain segmentation using (a) classical max-flow algorithm or graph cuts (GC) and (b) combinatorial continuous max-flow (CCMF) [Couprie et al., 2011]. (c,e) Zoomed regions of (a). (d,f) Zoomed regions of (b).
(Images adapted from [Couprie et al., 2011] by permission)

where u_x and u_y are the derivatives of u in x and y directions, respectively. The minimizer of E may be computed via the steady state solution of the following update equation:

$$\frac{\partial u}{\partial t} = -\frac{\partial E}{\partial u} \,, \tag{2.9}$$

where ∂t is an artificial time step size. A step size too large leads to a non-optimal solution and numerical instability, while a step size too small increases the convergence time. One way to ensure numerical stability during the optimization is to place an upper bound on the time-step using the Courant-Friedrichs-Lewy (CFL) condition [Courant et al., 1967]. Also, optimal step sizes can be computed automatically as proposed by [Pock and Chambolle, 2011]. On the other hand, in discrete domain, graph cuts-based methods do not require such parameter tuning and have proven to be numerically stable.

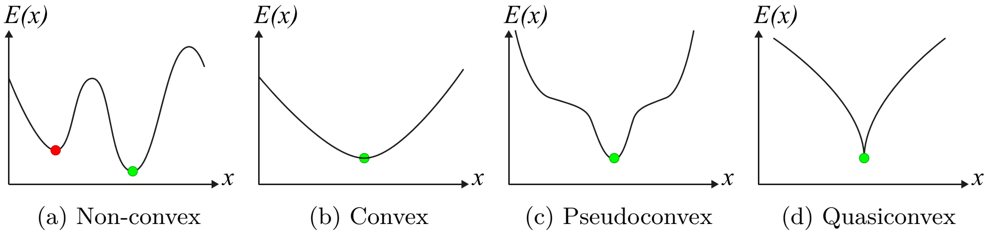
Note that other parameters in the segmentation energy function, including weighting parameters to balance the energy terms (e.g. λ in (2.2)) and hyper parameters within each energy term or objective (e.g. number of histogram bins in calculating the regional/data term) are common between continuous and discrete approaches. Setting parameters can be done based on training data (learning-based) [Gennert and Yuille, 1988, McIntosh and Hamarneh, 2007] or adaptively and based on the image content [Rao et al., 2010].

- **Termination criterion:** While graph based methods have an exact termination criterion, finding a general-purpose termination criteria for PDE-based methods is difficult. However, performing a fixed number of iterations and/or iterating until the change in the solution or energy is smaller than a predefined threshold are ways for stopping the optimization procedure.

- **Metrication error:** The computation results obtained from graph-based methods are often biased by the discrete graph setting and in case of a 4-connected neighbourhood graph, blocky structures (also known as metrication error) can be seen in the results. Figure 2.3 compares the discrete and continuous version of a max-flow algorithm. As seen in Figure 2.3, the contours obtained by graph cuts are noticeably blocky in the areas with weak regional cues (weak data term), while the contours obtained by the continuous method are smooth. The discrete nature of graph-based methods makes it difficult to efficiently implement a convex regularizer like total variation in the discrete domain. Metrication error can be reduced in graph-based methods by increasing the graph connectivity, e.g. [Boykov and Kolmogorov, 2003], but that also increases memory usage and computation time. Instead, in the continuous domain, there is no such limitation and regularizers can be implemented efficiently that makes the PDE approaches free from metrication error. Note that although approaches with continuous energy formulation do not have the limitation of discrete approaches, in the implementation stage, due to the discrete nature of digital images, all continuous operations are estimated by their discrete versions.

- **Parallelization:** Unlike PDE approaches that are easily parallelizable on GPUs, graph-based techniques are not straightforward to parallelize. As an example, the max-flow/min-cut, a core algorithm of many state-of-the-art graph-based segmentation methods, is a P-complete problem, which is probably not efficiently parallelizable [Goldschlager et al., 1982, Nieuwenhuis et al., 2013] due to two reasons: (1) augmenting path operations in min-cut/max-flow algorithms are interdependent as different augmentation paths can share edges; (2) the updates of the edge residuals have to be performed simultaneously in each augmentation operation as they all depend on the minimum capacity within the augmentation path [Nieuwenhuis et al., 2013]. Several attempts have focused on parallelizing the max-flow/min-cut computation. Push-relabel algorithms [Boykov et al., 1998, Delong and Boykov, 2008] relaxed the first issue mentioned above but the update operations are still interdependent. Other techniques split the graph into multiple parts and obtained the global optimum by iteratively solving sub-problems in parallel [Strandmark and Kahl, 2010, Liu and Sun, 2010] while [Shekhovtsov and Hlaváč, 2013] combined the path augmentation and push-relabel techniques.

- **Memory usage:** With respect to memory consumption, the continuous optimization methods are often the winner. While continuous methods require few floating point values for each pixel in the image, the graphical models require an explicit storage of edges as well as one floating value for each edge. This difference becomes important when we deal with very large images and when the large number of graph edges



(a) Non-convex (b) Convex (c) Pseudoconvex (d) Quasiconvex

Figure 2.4: One dimensional example of a (a) non-convex, (b) convex, (c) pseudoconvex and (d) a quasiconvex function. Red and green dots indicate the local minimum and the global solution, respectively. Red and green circles represent local and global optimum, respectively.

required to be implemented, e.g. 100s of millions pixels of microscopy images, and 3D volumes [Appleton and Talbot, 2006].

- **Runtime:** The runtime variance in graph-based methods is higher than PDE-based methods. For example, considering the α-expansion [Boykov et al., 2001] as a popular multi-label optimization technique, the number of max-flow problems that need to be solved highly depends on the input image and the chosen label order. In addition, the number of augmentation steps needed to solve a max-flow problem depends on the graph structure and edge residuals [Nieuwenhuis et al., 2013]. On the other hand, PDE-based methods have less runtime variance as they perform the same computation steps on each pixel.

For more qualitative and quantitative comparison between continuous and discrete domain, refer to [Nieuwenhuis et al., 2013, Couprie et al., 2011].

2.1.3 Optimization: convex (submodular) vs. non-convex (non-submodular)

In the continuous domain of energy, a function can be non-convex, convex, pseudoconvex or quasiconvex (Figure 2.4). We define each of these terms mathematically.

An energy function $E : \mathcal{S} \rightarrow \mathbb{R}$ is convex if

$$\begin{aligned} &\bullet \text{ the energy domain } \mathcal{S} \text{ (or the solution space) is a convex set and} \\ &\bullet\, \forall S_1, S_2 \in \mathcal{S} \text{ and } 0 \leq \lambda \leq 1 \\ &\quad E(\lambda S_1 + (1-\lambda)S_2) \leq \lambda E(S_1) + (1-\lambda)E(S_2). \end{aligned} \tag{2.10}$$

A set $\mathcal{S}$ is a convex set if $S_1, S_2 \in \mathcal{S}$ and $0 \leq \lambda \leq 1 \Rightarrow \lambda S_1 + (1-\lambda)S_2 \in \mathcal{S}$. If E is differentiable in $S_1 \in \mathcal{S}$, E is said to be pseudoconvex at S_1 if

$$\nabla E(S_1) \cdot (S_2 - S_1) \geq 0, S_2 \in \Omega \Rightarrow E(S_2) \geq E(S_1). \tag{2.11}$$

Pseudoconvex functions share the property of convex functions that, if $\nabla E(S) = 0$, then S is a global minimum of E. We call E a quasiconvex function if

$$\begin{aligned} &\bullet \text{ the energy domain } \mathcal{S} \text{ is a convex set and} \\ &\bullet \text{ the sub-level sets } \boldsymbol{S}_\alpha = \{S \in \mathcal{S} | E(S) \leq \alpha\} \text{ are convex for all } \alpha\ . \end{aligned} \tag{2.12}$$

The pseudoconvexity is strictly weaker than convexity. In fact, every convex function is pseudoconvex. For example, $E(S) = S + S^3$ is pseudoconvex and non-convex. Also, every pseudoconvex function is quasiconvex, but the relationship is not commutative, e.g. $E(S) = S^3$ is quasiconvex and not pseudoconvex.

In this thesis we focus on convex and non-convex optimization problems; more details on quasiconvex problems can be found in [dos Santos Gromicho, 1998]. In the continuous domain, an optimization problem must meet two conditions to be a convex optimization problem: 1) the objective function must be convex, and 2) the feasible set must also be convex. The drawbacks associated with non-convex problems are that, in general, there is no guarantee to find the global solution and results strongly depend on the initial guess/initialization. In contrast, for a convex problem, a local minimizer is actually a global minimizer and results are independent of the initialization. However, non-convex energy functional often give more accurate models (see Section 2.1.4).

The corresponding terminologies for convex and non-convex problems in the discrete domain are submodular and non-submodular (supermodular) problems, respectively. Let E be a function of n binary variables and $E(f_1, ..., f_n) = \sum_i E^i(f_i) + \sum_{i<j} E^{ij}(f_i, f_j)$. Then the discrete energy functional E is submodular if the following condition holds

$$E^{ij}(0,0) + E^{ij}(1,1) < E^{ij}(0,1) + E^{ij}(1,0). \tag{2.13}$$

For higher order energy terms, e.g. $E^{ijk}(f_i, f_j, f_k)$, E is submodular if all *projections* of E of two variables are submodular [Kolmogorov and Zabin, 2004]. To define *projection*, suppose E has n binary variables. If $m < n$ of these variables are fixed, then we get a new function E' of $n - m$ binary variables; E' is the projection of E.

Submodular energies can be optimized efficiently via graph cuts. [Greig et al., 1989] were the first to utilize min-cut/max-flow algorithms to find the globally optimal solution for binary segmentation in 1989. Later in 2003, [Ishikawa, 2003] generalizes the graph cut technique to find the exact solution for a special class of multi-label problems (more detail on Ishikawa's approach in Section 2.2.5).

In recent years, many efforts have been made to bridge the gap between convex and non-convex optimization problems in the continuous domain through convex approximations of non-convex models. Historically, the two-region segmentation problem (foreground and background) was convexified in 2006 by [Chan et al., 2006] and the multi-region seg-

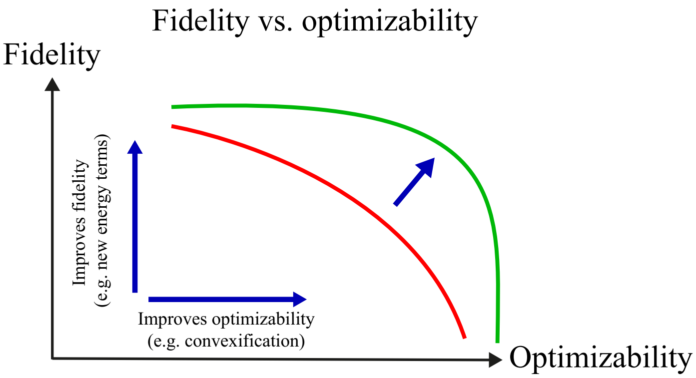


Figure 2.5: Fidelity vs. optimizibility. Ideally, an energy function is designed in a way that is faithful to the underlying segmentation problem and, at the same time, easy to optimized.

mentation problem was convexified in 2008 by [Chambolle et al., 2008] and [Pock et al., 2008] for the first time (more detail on continuous multi-region segmentation problem in Section 2.2.5).

2.1.4 Fidelity vs. Optimizibility

In energy-based segmentation problems there is a trade-off between *fidelity* and *optimizability* [Hamarneh, 2011, McIntosh and Hamarneh, 2012, Ulén et al., 2013]. Fidelity describes how faithful the energy function is to the data and how accurate it can model and capture intricate problem details. Optimizability determines how easily we can optimize the objective function and attain the global optimum.

Generally, the better the objective function models the problem, the more complicated it becomes and the harder it is to optimize. If we instead sacrifice fidelity to obtain a globally optimizable objective function, the solution might not be accurate enough for our segmentation purpose.

In the image segmentation literature, many works have focused on increasing the fidelity and improving the objective functions by (i) adding new energy terms, e.g. edge, region, shape, statistical overlap and area prior terms [Gloger et al., 2012, Shen et al., 2011, Andrews et al., 2011b, Bresson et al., 2006, Pluempitiwiriyawej et al., 2005, Ayed et al., 2009, Ayed et al., 2008]; (ii) extending binary segmentation methods to multi-label segmentation [Vese and Chan, 2002, Mansouri et al., 2006, Rak et al., 2013]; (iii) incorporating spatial relationships between labels, objects, or object regions [Felzenszwalb and Veksler, 2010, Liu et al., 2008, Rother et al., 2009, Colliot et al., 2006, Gould et al., 2008]; and (iv) learning objective function parameters [Alahari et al., 2010, Nowozin et al., 2010, Szummer et al., 2008, McIntosh and Hamarneh, 2007, Kolmogorov et al., 2007].

Other works chose to improve optimizibility by approximating non-convex energies with convex ones [Lellmann et al., 2009, Bae et al., 2011a, Boykov et al., 2001, Chambolle et al., 2008].

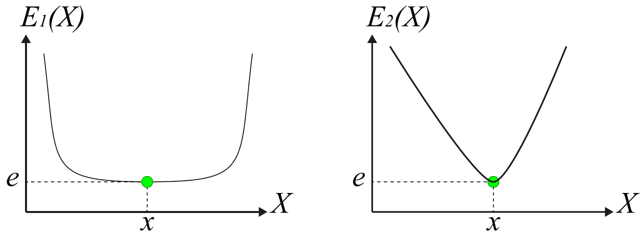
In an ideal case, we are interested in methods that improve both (optimizibility and fidelity), i.e. increase optimizibility without sacrificing the fidelity or even increase the fidelity at the same time, Figure 2.5.

2.1.5 Uncertainty and Fuzzy / probabilistic vs. crisp labelling

In an MIS problem, ideally, we are interested in finding an optimal ground truth labeling for an image, where each label represents a structure of interest. However, as medical images are approximate representations of physical tissues and also due to noise coming from other structures inside the body and/or imaging devices, sometimes it is difficult to precisely define a ground truth labeling. Even the manual segmentation of an image by several experts have some inter-expert (different experts) and intra-expert (same expert at different times) variability due to ambiguities in the image. Therefore, it is beneficial to encode uncertainty into segmentation frameworks [Koerkamp et al., 2010]. This information can be used to highlight the ambiguous image regions so to prompt users' attention to confirm or manually edit the segmentation of these regions.

Uncertainty in objects boundaries may arise from numerous sources, including graded decomposition, image acquisition artifacts, partial volume effects, and image segmentation methods intentionally designed to output probabilistic fuzzy results [Grady, 2006, Zhang et al., 2001]. Figure 2.6 demonstrates an example of how uncertainty information is encoded in an energy function. E_1 and E_2 in Figure 2.6 are two 1-D energy functions with the same optimal solution. However, segmentations near the minimal solution in E_1 have almost similar energy values (high uncertainty) as opposed to solutions near the same optimal point in E_2 (less uncertainty/more certain). In fact, using the energy E_1, a small perturbation in the image (e.g. an additional noise) may change the segmentation result noticeably. Given a probability distribution function, $P(x)$ in (2.4), over the label space, one way to calculate the uncertainty at pixel x is to use Shannon's entropy as: $h(x) = -\sum P(x)\log_2(P(x))$. The entropy can be used as an energy term in a segmentation energy function. In this case, lower entropy corresponds to larger certainty and vice versa.

As stated in Section 2.1.1, in addition to crisp labelling where each pixel is mapped to exactly one object label, we have probabilistic and fuzzy labelling that are two common ways to encode uncertainty into a segmentation framework. In probabilistic labelling a probability of each label at each pixel is reported [Wells III et al., 1996, Grady, 2006, Saad et al., 2008, Saad et al., 2010b, Changizi and Hamarneh, 2010, Andrews et al., 2011a, Andrews et al., 2011b] while fuzzy labelling reports a partial membership of each pixel belonging to each class of labels by a membership function [Bueno et al., 2004, Howing et al., 1997].



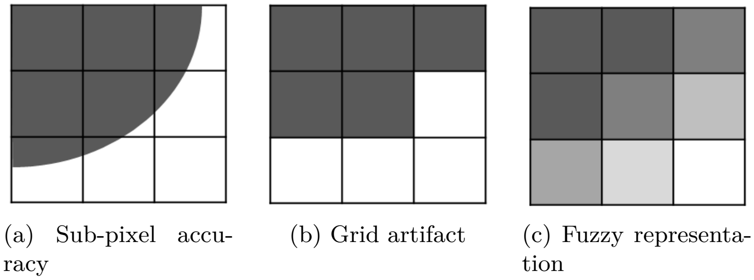
(a) Less certain solution (b) More certain solution

Figure 2.6: One dimensional example of two energy functions with (a) less vs. (b) more certain solutions.

One important issue with probabilistic methods is that most standard techniques for statistical shape analysis (e.g. principal component analysis (PCA)) assume that the probabilistic data lie in the unconstrained real Euclidean space, which is not valid as the sample space for a probabilistic data is the unit simplex. Neglecting this unit simplex in statistical shape analysis may produce invalid shapes. In fact, moving along PCA modes results in invalid probabilities that need to be projected back to the unit simplex. This projection discards uncertainty information. [Pohl et al., 2007] proposed a method based on the logarithm of odds (LogOdds) transform that maps probabilistic labels to an unconstrained Euclidean vector space and its inverse maps a vector of real values (e.g. values of the signed distance map at a pixel) to a probabilistic label. A shortcoming of the LogOdds transform is that it is asymmetric in one of the labels, usually chosen as the background, and changes in this label's probability are magnified in the LogOdds space. This issue was addressed by [Changizi and Hamarneh, 2010] and [Andrews et al., 2014] using the isometric log-ratio (ILR) transformation to isometrically and bijectively map the simplex to the Euclidean real space. They analyze data in the Euclidean real space and then back-transform the results to the unit simplex. More recently, [Andrews and Hamarneh, 2015] proposed a generalized log ratio transformation (GLR) that offers more refined control over the distances between different labels.

2.1.6 Sub-pixel accuracy

In the spatially discrete setting, objects are converted into a discrete graph. This discretization causes loss of spatial information, which causes the objects' boundaries to align with the axes or graph edges, Figure 2.7(b). On the other hand, the continuous domain does not have such shortcoming. In other words, sub-pixel accuracy allows for assigning a label to one part of a pixel and another label to the other part. This sub-pixel label assignment causes the segmentation accuracy to exceed the nominal pixel resolution of the image, Figure 2.7(a). However, as images are digitalized in computers, the accuracy of a


(a) Sub-pixel accuracy

(b) Grid artifact

(c) Fuzzy representation

Figure 2.7: A sample segmentation (a) with and (b) without sub-pixel accuracy. (c) Representing sub-pixel accuracy using a fuzzy representation.

crisp segmentation is always limited to the image pixel resolution. One way to represent sub-pixel accuracy is to use a fuzzy representation by defining the degree of membership to be proportional to the area covered by that label at each pixel, Figure 2.7(c).

We should emphasize that although in the continuous domain, image representation and energy formulations are continuous (Figure 2.2(a) and Figure 2.1(a)), implementation of these methods to image processing involves a discretization step (e.g. estimating a derivative by discrete forward difference). However, in the continuous setting it is possible to obtain a real pixel value (or label value), whereas in the discrete setting the value of labels are discrete, e.g. integer values. Nevertheless, theoretically continuous models correspond to the limit of infinitely fine discretization.

2.2 Prior knowledge for targeted image segmentation

In this section, we review the prior knowledge information devised to improve image segmentation. Table 2.1 presents some of these important priors have been used in the literature and compares them in terms of attaining a globally vs. locally optimal solution and metrication error, domain of action (continuous vs. discrete), and other properties. We also created an interactive online database (`http://goo.gl/gy9pyn`) to categorize existing works based on the type of prior knowledge they use. We made our website interactive so that researchers can contribute to keep the database up to date.

2.2.1 User interaction

Incorporating user input into a segmentation framework may be an intuitive and easy way to characterize the desired object and obtain usable results. In an interactive segmentation system, the user input is used to encode prior knowledge about the target object. The specific prior knowledge that the user is considering is unknown to the method, but only the implication of such prior knowledge (e.g. pixel x must be part of the object) is passed

Table 2.1: Some important prior information for targeted image segmentation

Method	**Multi-object**	**Shape**	**Topology**	**Moments**		**Geometrical/region interaction**			**Spatial distance**			**Adjacency**	**No. of regions/labels**	**Model/Atlas**	**No grid artifact**	**Guarantees on global solution**
				0^{th} (Area/Volume)	Higher orders	Containment	Exclusion	Relative position	min.	max. (Centroid)	Attraction					
[Cootes et al., 1995] [Cootes and Taylor, 1995] [Rousson and Paragios, 2002] [Chen et al., 2002] [Tsai et al., 2003]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
[Slabaugh and Unal, 2005] [Zhu-Jacquot and Zabih, 2007]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
[Veksler, 2008]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[Song et al., 2010]	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	✓
[Andrews et al., 2011b]	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✓	✓
[Han et al., 2003] [Zeng et al., 2008]	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[Vicente et al., 2008]	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[Foulonneau et al., 2006]	✗	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
[Ayed et al., 2008]	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
[Klodt and Cremers, 2011]	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
[Lim et al., 2011]	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[Wu et al., 2011]	✗	✗		✗	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✓
[Zhao et al., 1996]	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓
[Samson et al., 2000]	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗
[Li et al., 2006a]	✓	✗	✗	✗	✗	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗	✓
[Zeng et al., 1998]	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✓	✗
[Goldenberg et al., 2002]	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✓	✗
[Paragios, 2002]	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✓	✗
[Vazquez-Reina et al., 2009]	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗
[Ukwatta et al., 2012]	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	✓
[Rajchl et al., 2012]	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓
[Delong and Boykov, 2009]	✓	✗	✗	✗	✗	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓
[Ulén et al., 2013]	✓	✗	✗	✗	✗	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓
[Schmidt and Boykov, 2012]	✓	✗	✗	✗	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗
[Liu et al., 2008]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
[Felzenszwalb and Veksler, 2010]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓
[Strekalovskiy and Cremers, 2011] [Strekalovskiy et al., 2012] [Bergbauer et al., 2013]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓
[Zhu and Yuille, 1996] [Brox and Weickert, 2006]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗
[Delong et al., 2012a]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
[Yuan et al., 2012]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✓
[Iosifescu et al., 1997] [Collins and Evans, 1997] [Prisacariu and Reid, 2012]	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
[Sandhu et al., 2011] [Prisacariu et al., 2013]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗

on to the interactive algorithm. Given a high-level intuitive user interaction, the end-user need not know about the low-level underlying optimization energy function.

The work proposed by [Kass et al., 1988] is perhaps one of the early works to incorporate user interaction into the segmentation framework where they enforce spring-like forces between snake's control points to affect the energy functional and push the snake out of a local minima into another.

User input is incorporated in several ways, e.g. by mouse clicking (or even via eye gaze [Sadeghi et al., 2009]) and providing seed points, by specifying the subsets of object boundary or specifying sub-regions (bounding boxes) that contain the object of interest.

In the first form (providing *seed points*), the user specifies labels for some pixels inside and outside the target object by mouse-clicking or brushing. This allows a user to enforce a hard constraint on labeled pixels. For example, in a binary segmentation scenario in the discrete setting, one can enforce $f_p^{foreground} = 1$ if $p \in foreground$ and $f_p^{background} = 0$ if $p \in background$ in (2.7).

In the continuous domain, [Paragios, 2003], [Cremers et al., 2007] and [Ben-Zadok et al., 2009] proposed a level set-based method in which a user can correct the solution interactively by clicking on incorrectly labelled pixels. Let ϕ be the level set function (often is represented by the signed distance map of the foreground) where $\phi > 0$ and $\phi < 0$ represent inside and outside the object of interest, respectively. [Cremers et al., 2007] added the following user interaction term to their energy functional along with data and regularization terms:

$$E_{user}(\phi) = -\int_{\Omega} L(\boldsymbol{x})\text{sign}(\phi(\boldsymbol{x}))d\boldsymbol{x} \ . \tag{2.14}$$

$L : \Omega \to \{-1, 0, +1\}$ reflects the user input and is defined as:

$$L(\boldsymbol{x}) = \begin{cases} +1 & \text{if } \boldsymbol{x} \text{ is marked as 'object'} \\ -1 & \text{if } \boldsymbol{x} \text{ is marked as 'background'} \\ 0 & \text{if } \boldsymbol{x} \text{ is not marked} \end{cases} . \tag{2.15}$$

[Ben-Zadok et al., 2009] used an energy functional similar to [Cremers et al., 2007]. Assuming that $\{\boldsymbol{x}_i\}_{i=1}^n$ denotes the set of user input, which indicates the incorrectly labelled regions, they defined $M : \Omega \to \{0, 1\}$ as:

$$M(z) = \sum_{i=1}^{n} \delta(z - \boldsymbol{x}_i) \ , \tag{2.16}$$

where δ is the Dirac delta function. The function $L : \Omega \to \mathbb{R}$ is defined as:

$$L(\boldsymbol{x}) = H(\phi(\boldsymbol{x})) + (1 - 2H(\phi(\boldsymbol{x}))) \int_{z \in \lambda} M(z)dz \ , \tag{2.17}$$

where H is the Heaviside step function and λ is the neighbourhood of the coordinate $\boldsymbol{x}$. $L(\boldsymbol{x}) = 0$ if the user's click is within the segmented region and $L(\boldsymbol{x}) = 1$ if it is on the background. $L(\boldsymbol{x}) = H(\phi(\boldsymbol{x}))$ if $\boldsymbol{x}$ is not marked. The user interaction term proposed by [Ben-Zadok et al., 2009] is then defined as:

$$E_{user}(\boldsymbol{x}) = \int_{\boldsymbol{x}\in\Omega} \int_{\boldsymbol{x}'\in\Omega} \left(L(\boldsymbol{x}') - H(\phi(\boldsymbol{x}))\right)^2 K(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x}' d\boldsymbol{x} \ , \tag{2.18}$$

where $K(\boldsymbol{x}, \boldsymbol{x}')$ is a Gaussian kernel.

Another form of user input is *object boundary specification* where all or part of the object boundary is roughly specified by drawing a contour (in 2D) or initializing a surface (in 3D) around the object's boundary. This form of user input is more suitable for 2D images as providing a manual rough segmentation in 3D images (as is the case for most of medical images) is not straightforward. Examples that require the user to provide an initial guess close to objects' boundary include Wang et al.'s work [Wang et al., 2007] in the discrete setting, and edge-based active contours (e.g. gradient vector field (GVF) [Xu and Prince, 1997, Xu and Prince, 1998] and geodesic active contour [Goldenberg et al., 2001]) in the continuous setting. Live-wire, proposed by [Barrett and Mortensen, 1997], is another effective tool for 2D segmentation that benefits from user-defined seeds on the boundary of the desired object. The 2D live-wire uses the gradient magnitude, gradient direction, and canny edge detector to build cost terms. After providing an initial seed point on the boundary of the object, live-wire calculates the local cost for each pixel starting from the provided seed and finds the minimal path between the initial seed point p and the next point q chosen by the user. The 2D live-wire was extended to 3D by [Hamarneh et al., 2005].

Another form of user input, and probably the most convenient way for a user, is the *sub-region specification* where a user is asked to draw a box around the targeted object. This bounding box can be provided automatically using machine learning techniques in object detection. In the discrete setting, GrabCut proposed by [Rother et al., 2004] is one of the most well-known methods with this kind of initialization. [Lempitsky et al., 2009] proposed a method which shows how a bounding box is used to impose a powerful topological prior that prevents the solution from excessively shrinking and splitting, and ensures that the solution is sufficiently close to each of the sides of the bounding box. [Grady et al., 2011] performed a user study and showed that a single box input is in fact enough for segmenting the targeted object. In the continuous setting, this kind of user input (sub-region specification) is taken into account by methods like geodesic active contours [Caselles et al., 1997] in which the user initializes the active contour around the object of interest.

Similar interaction is utilized in 3D live-wire [Hamarneh et al., 2005] as implemented in the TurtleSeg software[1] [Top and et al., 2011, Top et al., 2011]. In 3D live-wire, few

---

[1] www.turtleseg.org

slices in different orientations of a 3D volume are segmented using 2D live-wire. Then, the segmented 2D slices are used to segment the whole 3D volume by generating additional contours on new slices automatically. The new contours are obtained by calculating optimal paths connecting the points of intersection between the new slice planes and the original contours provided semi-automatically by the user.

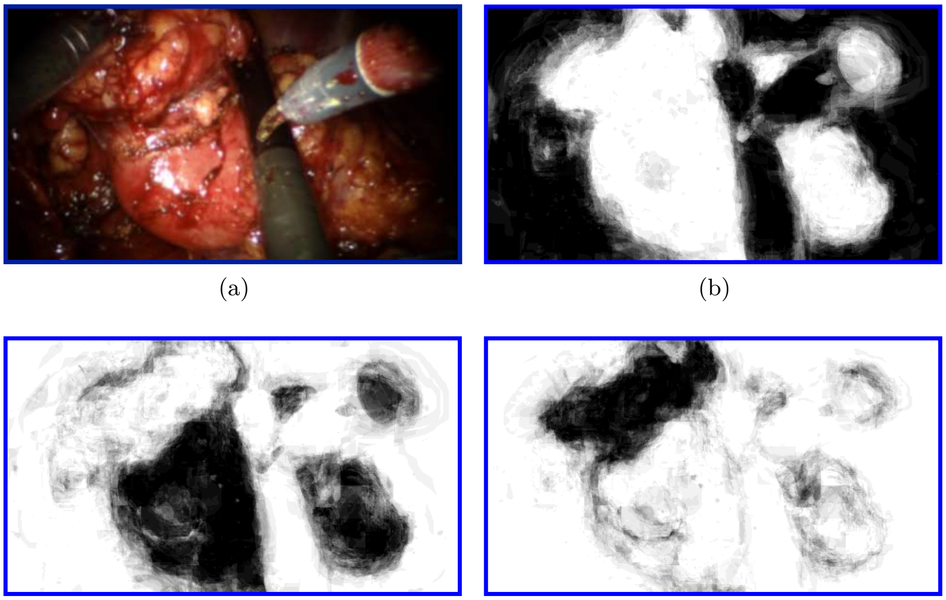
[Saad et al., 2010a] proposed another type of interactive image analysis in which a user is able to examine the uncertainty in the segmentation results and improve the results, e.g. by changing the parameters of their segmentation algorithm. For an expanded study on interaction in MIS, interested readers may refer to [Saad et al., 2010b, Saad et al., 2010a].

2.2.2 Appearance prior

Appearance is one of the most important visual cues to distinguish between different structures in an image. Appearance is described by studying the distribution of different features such as intensity values in gray-scale images, color, and texture inside each object. In most cases, appearance models are incorporated into the data term in (2.2) and (2.7). The purpose of incorporating appearance prior is to fit the appearance distribution of the segmented objects to the distribution of objects of interest, e.g. using Gaussian mixture model (GMM) [Rother et al., 2004]. In the literature, there are two ways to model the appearance: 1) adaptively learning the appearance during the segmentation procedure, and 2) knowing the appearance model prior to performing segmentation (e.g. by observing the appearance distribution of the training data). In the former case, the appearance model is learned as the segmentation is performed [Vese and Chan, 2002] (computed online). In the second case, it is assumed that the probability of each pixel belonging to particular label is known, i.e. if $F_i(\boldsymbol{x})$ represents a particular set of feature values (e.g. intensity/color) associated with each image location for i^{th} object, then it is assumed that $P(\boldsymbol{x}|F_i(\boldsymbol{x}))$ is known (or pre-computed offline). This probability is usually learned and estimated from the distribution of features inside small samples of each object.

Figure 2.8 illustrates the probability of different structures (the kidney, the tumour, and the background) in an endoscopic scene. A lower intensity in Figures 2.8(b-d) corresponds to higher probability.

To fit the segmentation appearance distribution to the prior distribution, a dissimilarity measure D is usually needed where $D(g_i, \hat{g}_i)$ measures the difference between the appearance distribution of i^{th} object (g_i) and its corresponding prior distribution $\hat{g}_i$. This dissimilarity measure can be encoded into the energy functional (2.2) directly as the data term or via a probabilistic formulation. For example, considering the appearance prior of an object in a scalar valued image I, $\hat{g}_i$ would be the mean (μ_i) and variance (σ_i^2) of the intensity of the desired object. Then, assuming a Gaussian approximation of the object's intensity I, the



(a) (b)

(c) (d)

Figure 2.8: Examples of regions probabilities. (a) Original endoscopic image. (b-d) Probability of background, kidney and tumour for the frame shown in (a). A lower intensity in (b-d) corresponds to higher probability.

corresponding probability distribution will be:

$$P(\boldsymbol{x}|\hat{g}_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(I(\boldsymbol{x})-\mu_i)^2}{2\sigma_i^2}} . \tag{2.19}$$

Other than scalar-valued medical images such as MR [Pluempitiwiriyawej et al., 2005] and US [Noble and Boukerroui, 2006]), appearance models can be extracted from other types of images like color image (e.g. skin [Celebi et al., 2009], endoscopy [Figueiredo et al., 2010]), other vector-valued images (dynamic postiron emission tomography, dPET, [Saad et al., 2008]), and tensor-valued or manifold-valued images [Feddern et al., 2003, Wang and Vemuri, 2004b, Weldeselassie and Hamarneh, 2007]. For the vector-valued images, one can use multivariate Gaussain density as an appearance model. The formulation is similar to (2.19) with the use of the covariance matrix Σ_i instead of σ_i^2. Regarding the tensor-valued images, several distance measures in the space of tensors have been proposed such as:

- Log-Euclidean tensor distance defined as:

$$D_{LE}(T_i, \hat{T}_i) = \sqrt{\text{trace}\Big((\text{logm}(T_i) - \text{logm}(\hat{T}_i))^2 \Big)},$$

where T_i and $\hat{T}_i$ are a tensor from i^{th} region and its corresponding prior tensor model, respectively.

- The symmetrized Kulback-Leibler (SKL) divergence (also known as J-convergence) [Wang and Vemuri, 2004b] defined as:

  $$D_{SKL}(T_i, \hat{T}_i) = \frac{1}{2}\sqrt{\text{trace}(T_i^{-1}\hat{T}_i + T_i\hat{T}_i^{-1}) - 2n} \,,$$

  where n is the size of the tensors T_i and $\hat{T}_i$ ($n = 3$ in DT-MRI). This measure is affine invariant.

- The Rao distance [Lenglet et al., 2004] defined as:

  $$D_R(T_i, \hat{T}_i) = \sqrt{\frac{1}{2}\sum_{i=1}^{n} \log^2(\lambda_i)} \,,$$

  where λ_i denotes the eigenvalues of $T_i^{-1/2}\hat{T}_i T_i^{-1/2}$ ($n = 3$ in DT-MRI).

Intensity and color information are not always sufficient to distinguish different objects. Hence, several methods proposed to model objects with more complex appearance using texture information as a complementary feature [Huang et al., 2005, Malcolm et al., 2007, Santner et al., 2009]. [Bigün et al., 1991] introduced a simple texture feature model consists of the Jacobian matrix smoothed by a Gaussian kernel (K_σ) that results in three different feature channels, i.e. in case of a 2D Image the features are $K_\sigma * (I_x^2, I_x I_y, I_y^2)$. However, these features ignore the non-textured object that might be of interest. Therefore, [Rousson et al., 2003] proposed to use the following texture features in order to segment objects with and without texture: $(I, \frac{I_x^2}{|\nabla I|}, \frac{I_x I_y}{|\nabla I|}, \frac{I_y^2}{|\nabla I|})$.

More advanced texture features such as those based on Haar and Gabor filter banks have shown many successes in medical image segmentation [Huang et al., 2005, Malcolm et al., 2007, Santner et al., 2009]. [Koss et al., 1999] and [Frangi et al., 1998] are two works that utilized advanced features to segment abdominal organs and to measure vesselness, respectively. In [Frangi et al., 1998], the eigenvalues of the Hessian matrix are used for measuring the vesselness of pixels in images. This measure is used for liver vessel segmentation both in a variational framework [Freiman et al., 2009] and in a graph-based framework [Esneault et al., 2010]. Statistical overlap prior is another strong appearance prior that has been proposed by [Ayed et al., 2009]. Their method embeds statistical information (e.g. histogram of intensities) about the overlap between the distributions within the object and the background in a variational image segmentation framework. They used the Bhattacharyya coefficient measuring the amount of overlap between two distributions, i.e. $D_B(g_i, \hat{g}_i) = \sum_z \sqrt{g_i(z)\hat{g}_i(z)} \quad \forall z \in \mathbb{Z}$ if $I : \Omega \to \mathbb{Z}$. [Ben Ayed et al., 2009] used this strong prior to segment left ventricle in MR images.

Other features such as frequency, bag of visual words, gradient location and orientation histogram (GLOH) [Mikolajczyk and Schmid, 2005], DAISY [Tola et al., 2008], GIST (spatial envelop) [Oliva and Torralba, 2001], local binary pattern (LBP) [Heikkilä et al., 2009], SURF [Bay et al., 2006], histogram of oriented gradient (HOG) [Dalal and Triggs, 2005], and scale invariant feature transform (SIFT) [Lowe, 2004] are sometimes helpful as appearance features [Bosch et al., 2007].

Sometimes the appearance of structures is too complicated that regular features cannot describe them accurately. To extract the appearance characteristics of such structures different machine learning techniques have been proposed. These machine learning techniques learn the appearance either by combining several features like texture, color, intensity, HOG, etc., and feed the combined feature vectors to a classifier like random decision forest (RF) or support vector machine (SVM) [Tu et al., 2006], or by learning a dictionary which describes the object of interest [Mairal et al., 2008, Nieuwenhuis et al., 2014, Nayak et al., 2013].

In general, appearance features can be extracted in the following domains based on the type of the medical data:

- **spatial domain:** several methods have been developed to segment 2D or 3D static images [Chan et al., 2000, Cootes et al., 2001, Vese and Chan, 2002, Feddern et al., 2003, Wang and Vemuri, 2004b, Huang et al., 2005, Malcolm et al., 2007, Santner et al., 2009];

- **time domain:** in dynamic medical images, it is beneficial to consider the temporal dimension along with the spatial dimensions. For example, extracting appearance features in temporal direction would be very informative in dPET images, where each pixel in the image represents a time activity curve (TAC) that describes the metabolic activity of a tissue as a result of tracer uptake [Saad et al., 2008]. Other examples include [Mirzaei et al., 2013] where spatio-temporal features are used to distinguish tumour regions in 4D lung CT (3D+time) and [Amir-Khalili et al., 2014] where the likelihood of vessel regions are calculated based on temporal and frequency analysis.

- **scale domain:** for some objects with more complex texture, it is useful to estimate the appearance model in different scales for more accurate results and ensure that the model is scale invariant [Han et al., 2009, Mirzaalian and Hamarneh, 2010].

Regardless of where the appearance information comes from, it is encoded into the data energy term ($\boldsymbol{\mathcal{D}}$) in order to assign each pixel a probability of belonging to each class of objects.

2.2.3 Regularization

The regularization term corresponds to priors on the space of feasible solutions. As an example, the regularization term in (2.3) and (2.7) ensures that the region boundaries are smooth. Several regularization terms have been proposed in the literature. The most famous one is the Mumford-Shah model [Mumford and Shah, 1989] that penalizes the boundary length of different regions in a spatially continuous domain, i.e. $\sum_{i=1}^{n} |\partial S_i|$. The corresponding regularization model in the discrete domain is Pott's model that penalizes any appearance discontinuity between neighbouring pixels and is defined as $\sum_{p,q \in \mathcal{N}} w_{pq} \delta(I_p \neq I_q)$.

The regularization term in the discrete setting is biased by the discrete grid and favours curves to orient along with the grid, e.g. in horizontal and vertical or diagonal directions in a 4-connected lattice of pixels. This produces grid artifacts, also known as metrication error (Figure 2.3). On the other hand, the regularization term in the continuous settings does not have such an issue and allows for accurately representing geometrical entities such as curve length (or surface area) without any grid bias.

Some other regularization terms in the level set notation (ϕ) are listed as follows :

- Length regularization: $\int_\Omega |\nabla H(\phi(\boldsymbol{x}))| d\boldsymbol{x}$. We recall that $H(.)$ is the Heaviside step function.

- Total variation (TV): $\int_\Omega |\nabla \phi(\boldsymbol{x})| d\boldsymbol{x}$ which smooths only the tangent direction of the level set curve. This term is used especially when a single function ϕ is used to segment multiple regions, i.e. ϕ is not necessarily a signed distance function. It is worth mentioning that there are two variants of the total variation term: the isotropic variant using ℓ_2 norm,

  $$\int_\Omega |\nabla \phi(\boldsymbol{x})|_2 d\boldsymbol{x} = \int_\Omega \sqrt{|\phi_{x_1}|^2 + \cdots + |\phi_{x_N}|^2}\, d\boldsymbol{x}\ , \tag{2.20}$$

  and the anisotropic variant using ℓ_1 norm,

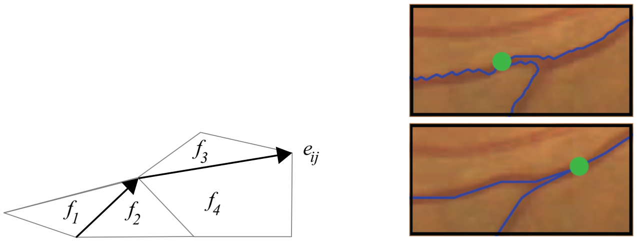
  $$\int_\Omega |\nabla \phi(\boldsymbol{x})|_1 d\boldsymbol{x} = \int_\Omega |\phi_{x_1}| + \cdots + |\phi_{x_N}| d\boldsymbol{x}\ . \tag{2.21}$$

  The anisotropic version is not rotationally invariant and therefore favours results that are aligned along the grid system. The isotropic version is typically preferred but cannot be properly handled by discrete optimization algorithms as the derivatives are not available in all directions in the discrete settings.

- $\mathcal{H}^1$ norm: $\int_\Omega |\nabla \phi(\boldsymbol{x})|^2 d\boldsymbol{x}$ which is a pure isotropic smoothing at every pixel $\boldsymbol{x}$.

A comparison of the above mentioned regularization terms can be found in [Chung and Vese, 2009].

Higher order regularization terms were also proposed to encode more constraints on the optimization problem. For example, [Duchenne et al., 2011] introduced the ternary term (along with unary and pairwise terms in the standard MRF) for graph matching



(a) (b)

Figure 2.9: Curvature regularization term. (a) Approximating the curvature term by tesselating the image domain into a cell complex. e_{ij} is the boundary variable and f_1, f_2, f_3 and f_4 are its corresponding region variables. (b) Vessel segmentation **top:** without curvature and **bottom:** with curvature regularization term. (Image (b) adapted from [Strandmark et al., 2013] by permission)

(and not image segmentation) application and [Delong et al., 2012b] proposed an efficient optimization framework to optimize sparse higher order energies in the discrete domain.

Curvature regularization is another useful type of regularization that has been shown to be able to capture thin and elongated structures [Schoenemann et al., 2009, El-Zehiry and Grady, 2010]. In addition, there is evidence that cells in the visual cortex are able to detect curvature [Dobbins et al., 1987].

Although curvature regularization term has been used in the local optimization frameworks easily, e.g. in a level set formulation [Leventon et al., 2000a] and Snakes model [Kass et al., 1988], however, it is much more difficult to incorporate such prior in a global optimization framework. [Strandmark and Kahl, 2011] proposed several improvements to impose curvature regularization within a global optimization framework. They defined the curvature term as: $\int_{\partial S} k(\boldsymbol{x})^2 d\boldsymbol{x}$, where ∂S is the boundary of the foreground region and k is the curvature function. They approximated the above mentioned curvature term with discrete computation techniques by tessellating the image domain into a cell complex, e.g. hexagonal mesh, a collection of non-overlapping basic regions whose union gives the whole domain. They recast the problem as an integer linear program (along with the data term and length/area penalty terms) and optimized the total energy via linear programming (LP) relaxation. Figure 2.9(a) shows how [Strandmark and Kahl, 2011] discretized the image domain by cells. If f_i, $i = 1, \cdots, m$ denote binary variables associated to each cell region and e_i be the boundary variable, then the curvature regularization term is written as a linear function: $\sum_{i,j} b_{ij} e_{ij}$, where e_{ij} denotes the boundary pairs and

$$b_{ij} = \min\{l_i, l_j\} \left(\frac{\alpha}{\min\{l_i, l_j\}} \right)^2 . \tag{2.22}$$

l_i is the length of edge i and α is the angle difference between two lines.

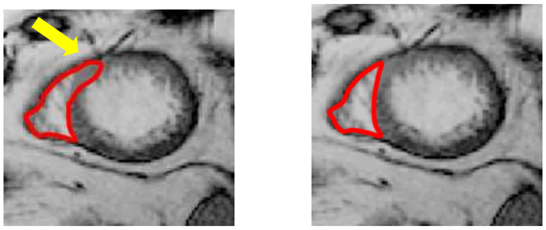
Later, [Strandmark et al., 2013] extended their previous work [Strandmark and Kahl, 2011] and proposed a globally optimal shortest path method that minimizes general functionals of higher-order curve properties, e.g. curvature and torsion. Figure 2.9(b) illustrate the usefulness of curvature prior on vessel segmentation.

2.2.4 Boundary information

Boundary and edge information is a powerful feature to delineate the objects of interest in an image. To incorporate boundary information, it is often assumed that the object's boundary is more likely to pass between pixels with large intensity/color contrast or, more generally, regions with different appearance (as captured by any of the measures in Section 2.2.2). As objects' boundaries are locations where we expect discontinuities in the labels, this information is usually linked to the regularization term in (2.2) such that the regularization penalty is decreased in high contrast regions (most likely objects' boundaries) to allow for discontinuity in labels. The functions $w_{ij} = exp(-\beta\|I_i - I_j\|_2^2)$ and $w'_{ij} = 1/1+\beta\|I_i - I_j\|_2^2$ are two examples of a boundary weighting function where I_i and I_j represent the intensity/color value associated with pixels i and j in image I, respectively [Grady, 2012]. These boundary weights are used as multiplication factors along with the regularization terms mentioned in Section 2.2.3. Geodesic active contour [Caselles et al., 1997], normalized-cut [Shi and Malik, 2000], and random walker [Grady, 2006] are three examples that employed such boundary weighting technique.

Boundary and edge detectors typically involve first and second order spatial differential operators. Several methods have been proposed to calculate first and second order differences in scalar images [Canny, 1986, Frangi et al., 1998] and color images [Shi et al., 2008, Tsai et al., 2002]. However, some medical images (e.g. diffusion tensor MRI or DT MRI) are manifold-valued. To address this, [Nand et al., 2011] extended the first order differential as $g(\boldsymbol{x}) = \sqrt{\lambda}\hat{e}$ where λ and $\hat{e}$ are respectively the largest eigenvalue and corresponding eigenvector of $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{J}(\boldsymbol{x})$ and $\boldsymbol{J}(\boldsymbol{x})$ is the Jacobian matrix generalizing the gradient of a scalar field to the derivatives of the 3D DT image. Similarly, the authors extended the second order differential as $G'(\boldsymbol{x}) = \frac{G(\boldsymbol{x})+G(\boldsymbol{x})^T}{2}$ where $G(\boldsymbol{x})$ is the Jacobian matrix of $g(\boldsymbol{x})$, i.e. $G_{ij} = \frac{\partial g_i}{\partial x_j}$. Similar approach has been proposed for boundary detection in color images, e.g. in color snakes [Sapiro, 1997] and in detecting boundaries of oral lesions in color images proposed by [Chodorowski et al., 2005].

Boundary polarity: A problem with the aforementioned boundary models is that they describe a boundary point that passes between two pixels with high image contrast without accounting for the direction of the transition [Boykov and Funka-Lea, 2006, Grady, 2012]. [Singaraju et al., 2008] considered the transition direction in boundary detection. For example, it is possible to distinguish between boundaries from bright to dark and from dark to bright (boundary polarity). This boundary polarity is incorporated into a graph based framework by replacing each undirected edge, e_{ij}, by two directed edges, e_{ij} and e_{ji}, such



(a) (b)

Figure 2.10: Cardiac right ventricle segmentation (a) without encoding edge polarity and (b) with encoding edge polarity by specifying the bright to dark edges as the desired ones. Note how the incorrect boundary transition (the yellow arrow) in (a) has been corrected by specifying boundary polarity in (b).

that their weight is calculated as:

$$w_{ij} = \begin{cases} \exp(-\beta_1 \|I_i - I_j\|_2^2) & if I_i > I_j \\ \exp(-\beta_2 \|I_i - I_j\|_2^2 & \text{otherwise} \end{cases}, \tag{2.23}$$

where $\beta_1 > \beta_2$. In (2.23), boundary transition from bright to dark is less costly than boundary transition from dark to bright. One example of encoding boundary polarity is shown in Figure 2.10, where the boundary ambiguity is resolved by specifying the boundary polarity, i.e. in this example, bright to dark boundary.

The assumption of high contrast in objects' boundaries might not be always valid in many medical images, e.g. soft tissue boundaries in CT. In addition, the two proposed contrast models, w_{ij} and w'_{ij}, are suitable for objects with smooth appearance and not for textured objects. One possible way to address these aforementioned issues (low contrast image and textured objects) is to utilize the piecewise constant case of Mumford-Shah model [Mumford and Shah, 1989] and replace I_i with $\tau(I_i)$, where τ is a function that maps the pixel content to a transformed space where the object appearance is relatively constant [Grady, 2012]. The Mumford-Shah model segments the image into a set of pairwise disjoint regions with minimal appearance variance and minimal boundary length. Among the most popular methods that adopted the Mumford-Shah model is the active contours without edges (ACWOE) method proposed by [Chan and Vese, 2001]. As an example [Sandberg et al., 2002] proposed a level set-based active contour algorithm to segment textured objects. Another example is the work proposed by [Paragios and Deriche, 2002] where boundary and region-based segmentation modules were exploited and unified into a geodesic active contour model to segment textured objects.

2.2.5 Extending binary to multi-label segmentation

In many medical image segmentation problems, we are interested in segmenting multiple objects (e.g. segmenting retinal layers from optical coherence tomography [Yazdanpanah et al., 2011]). Unlike a large class of binary labeling problems that can be solved globally, multi-label problems, on the other hand, cannot be globally minimized in general. In 2001, [Boykov et al., 2001] proposed two algorithms (α-expansion and α-β swap) based on graph cuts that efficiently find a local minimum of a multi-label problem. They consider the following energy functional:

$$E(\boldsymbol{f}) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{pq}(f_p, f_q) \ , \tag{2.24}$$

where $\mathcal{P}$ is the set of all pixels, $\boldsymbol{f} = \{f_p | p \in \mathcal{P}\}$ is a labeling of the image, $D_p(f_p)$ measures how well label f_p fits pixel p and V_{pq} is a penalty term for every pair of neighbouring pixels p and q and encourages neighbouring pixels to have the same label. The second term ensures that the segmentation boundary is smooth. Methods of proposed in [Boykov et al., 2001] require V_{pq} to be either a *metric* or *semimetric*. V is a metric on the space of labels $\mathcal{L}$ if it satisfies the following three conditions:

$$V(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta \tag{2.25}$$

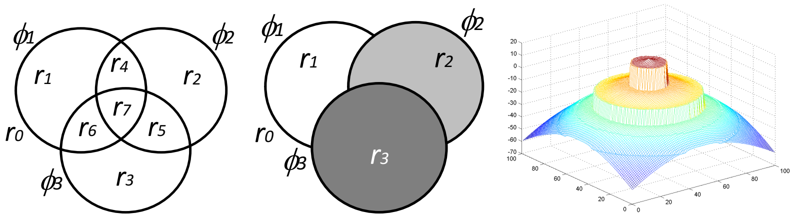
$$V(\alpha, \beta) = V(\beta, \alpha) > 0 \tag{2.26}$$

$$V(\alpha, \beta) \leq V(\alpha, \gamma) + V(\gamma, \beta) \ , \tag{2.27}$$

for any labels α, β,$\gamma \in \mathcal{L}$. If V only satisfies (2.25) and (2.26) then V is a semimetric. [Boykov et al., 2001] find the local minima by swapping a pair of labels (α-β-swap) or expanding a label (α-expansion) and evaluate the energy using graph cuts iteratively. Later in 2003, Ishikawa [Ishikawa, 2003] showed that, if $V_{pq}(f_p, f_q)$ is convex and symmetric in $f_p - f_q$, one can compute the exact solution of the multi-label problem. Ishikawa used the following formulation:

$$E(\boldsymbol{f}) = \sum_{p \in \mathcal{P}} D(f_p) + \sum_{(p,q) \in \mathcal{N}} g(\ell(f_p) - \ell(f_q)) \ , \tag{2.28}$$

where $D(.)$ in the first term (data term) is any bounded function that can be non-convex and $g(.)$ is a convex function and the function ℓ gives the index of a label, i.e. $\ell(\mathit{label}\ i) = i$. The term $g(\ell(f_p) - \ell(f_q))$ expresses that there is a linear order among the labels and the regularization depends only on the difference of their ordinal number. Ishikawa showed that if $g(.)$ is convex in terms of a linearly ordered label set, the problem of (2.28) can be exactly optimized by finding the min-cut over a specially constructed multi-layered graph in which each layer corresponds to one label.



(a) (b) (c)

Figure 2.11: Multi region level set methods proposed by (a) [Vese and Chan, 2002], (b) [Mansouri et al., 2006], and (c) [Chung and Vese, 2009].

In the continuous domain, [Vese and Chan, 2002] extended their level sets-based method to multiphase level sets. To segment N objects, their method needs $\lceil \log_2 N \rceil$ level set functions. The number of regions is upper-bounded by a power of two (Figure 2.11(a)). Therefore, the actual number of regions the method yields is sometimes not clear as it depends on the image and the regularization weights. This issue happens specifically when number of regions of interest is less than $2^{\lceil \log_2 N \rceil}$. [Mansouri et al., 2006] proposed to assign an individual level set function to each object of interest (excluding the background), i.e. their method needs $N-1$ non-overlapping level set functions to segment N objects (Figure 2.11(b)). [Chung and Vese, 2009] proposed another method that uses a single level set function for multi-object segmentation. They proposed to use different layers (or levels) of a level set function to represent different regions as opposed to just using the zero level set (Figure 2.11(c)). None of the aforementioned continuous methods guarantee the globally optimal solution for multi-label problems. [Pock et al., 2008] proposed a spatially continuous formulation of Ishikawa's multi-label problem. In their method, the non-convex variational problem is reformulated as a convex variational problem via a technique they called *functional lifting*. They used the following energy functional:

$$E(u) = \int_\Omega \rho(u(\boldsymbol{x}), \boldsymbol{x}) d\boldsymbol{x} + \int_\Omega |\nabla u(\boldsymbol{x})| d\boldsymbol{x} \ , \tag{2.29}$$

which can be seen as the continuous version of Ishikawa's formulation (2.24). $u : \Omega \to \Gamma$ in (2.29) is the unknown labeling function and $\Gamma = [\gamma_{min}, \gamma_{max}]$ is the range of u. The first term in (2.29) is the data term, which can be a non-convex function, and the second term is the total variation regularization term which is a convex term. In the functional lifting technique, the idea is to transfer the original problem formulation to a higher dimensional space by representing u in terms of its supper level sets φ defined as:

$$\varphi(\boldsymbol{x}, \gamma) = \begin{cases} 1 & \text{if } u(\boldsymbol{x}) > \gamma \\ 0 & \text{otherwise} \end{cases} . \tag{2.30}$$

Now, (2.29) can be re-written in terms of the supper level set function as:

$$E(\varphi) = \int_{\Sigma} \rho(\boldsymbol{x}, \gamma) |\partial_{\gamma} \varphi(\boldsymbol{x}, \gamma)| d\Sigma + \int_{\Sigma} |\nabla \varphi(\boldsymbol{x}, \gamma)| d\Sigma \ , \tag{2.31}$$

which is convex in φ and $\Sigma = [\Omega \times \Gamma]$. The minimization of $E(\varphi)$ is not a convex optimization problem since $\varphi : \Sigma \to \{0, 1\}$. Hence, φ is relaxed to vary in $[0, 1]$. We emphasize that the method of [Pock et al., 2008] cannot always guarantee the globally optimal solution of the original problem (before φ is relaxed and when φ is binary). [Brown et al., 2009] utilized functional lifting technique proposed by [Pock et al., 2008] and suggested a dual formulation for the multi-label problem. Their method guarantees a globally optimal solution. Recently, inspired by Ishikawa, [Bae et al., 2011b] proposed a continuous max-flow model for multi-labeling problem via convex relaxed formulations. Not only can their continuous max-flow formulations obtain exact and global optimizers to the original problem, but they also showed that their method is significantly faster than the primal-dual algorithm of [Pock et al., 2008].

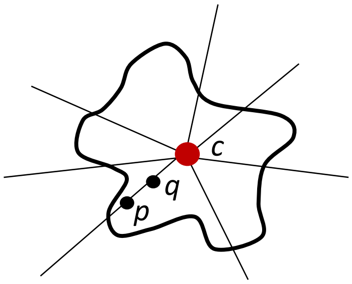
2.2.6 Shape prior

Shape information is a powerful semantic descriptor for specifying target objects in an image. In our categorization, shape prior can be modelled in three ways: geometrical (template), statistical and physical.

Geometrical model (template)

Sometimes the shape of the target object is known a priori (e.g. ellipse or cup-like shape). In this case, the shape can be modelled either by parametrization (e.g. an ellipse can be parametrized by its center coordinate, major and minor radius and orientation) or by a non-parametric way (e.g. by its level set representation) and incorporated into a segmentation framework.

One way to incorporate a geometrical shape model into a segmentation framework is to penalize any deviation from the model. In the continuous domain, given two shapes represented by their signed distance functions ϕ_1 and ϕ_2, a simple way to calculate the dissimilarity between them is given by $\int_{\Omega} (\phi_1 - \phi_2)^2 d\boldsymbol{x}$. The problem with this measure is that it depends on Ω, i.e. as the size of Ω is increased, the difference becomes larger. An alternative is to constrain the integral to the domain of ϕ_1, i.e. $\int_{\Omega} (\phi_1 - \phi_2)^2 H(\phi_1) dx$, as proposed in [Rousson and Paragios, 2002]. The aforementioned formulas are usable if the pose of the object of interest (location, rotation and scale) is known. If the pose of an object is unknown, one can include the pose parameters into the shape energy term and optimize the energy functional with respect to both pose parameters and the level set as has been done in [Chen et al., 2002, Pluempitiwiriyawej et al., 2005]. The authors in [Chen



(a)

Figure 2.12: A star shape with a given center c. p an q are two points on the line passing through c. If p is labeled as object, then q must also be labled as the abject.

et al., 2002] imposed the shape prior on the extracted contour after each iteration of the level set function. [Pluempitiwiriyawej et al., 2005] described the shape of an ellipse with five parameters including its pose parameters. They also optimized their energy functional by iterating between optimizing the shape energy term and the regional term.

In the discrete domain, the method of [Slabaugh and Unal, 2005] is one of the primary works to incorporate an explicit shape model into a graph-based segmentation framework. They proposed the following extra term (in addition to data and regularization terms) that constrained the segmentation to return an elliptical object

$$E_{ellipse}(\boldsymbol{f}, \theta) = \sum_{i \in \mathcal{P}} |M_i^{\theta} - f_i| \ , \tag{2.32}$$

where M^{θ} is the mask of an ellipse parametrized by θ. As minimizing such a term is not straightforward, the authors optimized the energy functional iteratively, i.e. by finding the best $\boldsymbol{f}$ for a fixed θ and then optimizing θ for a fixed $\boldsymbol{f}$. For complex shapes that are hard to parametrize, an alternative approach is to fit a shape template to the current segmentation as proposed in [Freedman and Zhang, 2005]. [Veksler, 2008] proposed to incorporate a more general class of shapes, known as *star shapes*, into graph-based segmentation. In Veksler's work, it is assumed that the center point (c) of the object is given. According to their definition, "an object has a star shape if for any point p inside the object, all points on the straight line between the center c and p also lie inside the object" (Figure 2.12). The following pairwise term was introduced to impose the star shape prior:

$$E_{pq}^{Star}(f_p, f_q) = \begin{cases} 0 & \text{if } f_p = f_q \\ \infty & \text{if } f_p = 1 \ \text{ and } \ f_q = 0 \\ \beta & \text{if } f_p = 0 \ \text{ and } \ f_q = 1 \end{cases} . \tag{2.33}$$

This prior is particularly useful for segmentation of convex objects, e.g. optic cum and disc segmentation [Bai et al., 2014].

Statistical model

In many practical applications, objects of the same class are not identical or rigid. For example, in medical images, the shape of organs vary from one subject to another or even over time and so, assuming a fixed shape template may be inappropriate. A typical way to capture the intra-class variation of shapes is to build a shape probability model, i.e. $P(shape)$. Now, two questions have to be investigated: 1) how to represent a shape; explicitly like point cloud, boundary-based (e.g. surface mesh), medial-based (e.g. m-reps [Pizer et al., 2003]), or implicitly (e.g. level set), and 2) what probability distribution model to adopt, e.g. Gaussian distribution, Gaussian mixture model, or kernel density estimation (KDE).

[Cootes et al., 1995] generated a compact shape representation and performed PCA (assuming Gaussian distribution) on a set of training shapes to obtain the main modes of variation. The idea is to model the plausible deformations of object's shape (S) by its principal modes of variation:

$$S = \overline{S} + \sum_{i=1}^{k} w_i P_i \ , \tag{2.34}$$

where $\overline{S}$ is the average shape, P_i is the i^{th} principal component and w_i is its corresponding weight (or shape parameter). [Cootes et al., 1995] used object's coordinates to represent S. Given an initial estimation of the position of an object, the segmentation is performed by directly optimizing an energy functional over the weights w_i. This model is later improved by [Tsai et al., 2001,Tsai et al., 2003,Leventon et al., 2000b] and [Van Ginneken et al., 2002]. For example, [Leventon et al., 2000b] represented S by its level sets to automatically handle topological changes during the contour evolution. [Tsai et al., 2003] used the same level set-based shape representation as [Leventon et al., 2000b] and incorporated the shape prior in a region-based energy functional as opposed to an edge-based energy proposed in [Cootes et al., 1995]. [Van Ginneken et al., 2002] proposed to use a general set of local image structure descriptors including the moments of local histograms instead of the normalized first order derivative profiles used in [Collins et al., 1995].

Similar to [Tsai et al., 2003] in the continuous domain, [Zhu-Jacquot and Zabih, 2007] employed an iterative approach that accounts for shape variability in a graph-based setting. At each iteration, they optimize the weights of principal modes of variations and the set of rigid transformation parameters given a tentative segmentation. Then, the segmentation is updated given the fitted shape template by minimizing an energy functional consisting a regional term. The procedure is repeated until convergence. Recently, [Andrews et al., 2014] proposed a probabilistic framework and incorporated shape prior to segment multiple anatomical structures. They utilized PCA in the isometric log-ratio space as PCA assumes that the probabilistic data lie in the unconstrained real Euclidean space. This is not a valid assumption as the sample space for a probabilistic data is the unit simplex and PCA may generate invalid probabilities, and hence, invalid shapes.

In the above mentioned PCA-based methods, aligning the shapes before computing the principal modes of variation is necessary and to do this alignment, it is often needed to provide point-to-point correspondence between landmarks of different subjects. This might be a tedious task. Hence, some methods proposed to capture shape variations in the frequency domain by representing shapes by the coefficients of its discrete cosine transform (DCT) [Hamarneh and Gustavsson, 2000], Fourier transform [Staib and Duncan, 1992] or spherical wavelet transform [Nain et al., 2006].

While PCA is a popular linear dimensionality reduction technique, it has the restrictive assumption that the input data is drawn from a Gaussian distribution. If the shape variation does not follow a Gaussian distribution, we might end up with invalid shapes or unable to represent valid shapes. In this case, a more accurate estimation of shape parameters might be obtained by GMM as proposed in [Cootes and Taylor, 1999]. In addition, PCA is only capable of describing global shape variations, i.e. changing a parameter corresponding to one eigenvector deforms the entire shape, which makes it difficult to obtain a proper local segmentation. To control the statistical shape parameters locally, [Davatzikos et al., 2003] presented a hierarchical formulation of active shape models, using the wavelet transform. In their method, a hierarchical representation of a deformable contour in terms of its wavelet transform is followed by PCA. The statistical properties extracted by PCA are used as priors in the contour's deformation. Some of these priors capture the global shape characteristics of the object boundaries, whereas, some of them capture local and high-frequency shape characteristics. [Hamarneh et al., 2004] also proposed a method to locally control the statistical shape parameters. They used the medial-based profile for shape representation and developed spatially-localized feasible deformations using hierarchical (multi-scale) and regional (multi-location) PCA and deform the medial profile at certain locations and scales. [Üzümcü et al., 2003] proposed to use independent component analysis (ICA) instead of PCA which does not assume a Gaussian distribution of the input data and can capture localized shape variations. However, ICA representation for shape variability is not as compact as PCA. [Ballester et al., 2005] proposed to use principal factor analysis (PFA) as an alternative to PCA and argued that PFA provides more "interpretable" modes of variations and is better suited for medical image analysis. PFA represents the observed D-dimensional data $\boldsymbol{O}$ as a linear function $\mathcal{F}$ of an L-dimensional ($L < D$) latent variable z and an independent Gaussian noise err as: $\mathcal{F}(\boldsymbol{O}) = \boldsymbol{\Lambda} z + \boldsymbol{\mu} + err$, where $\boldsymbol{\Lambda}$ is the $D \times L$ *factor loading matrix* defining the linear function $\mathcal{F}$, $\boldsymbol{\mu}$ is a D-dimensional vector representing the mean of the distribution of $\boldsymbol{O}$, and err is a D-dimensional vector representing the noise variability associated with each of the D observed variables. Authors in [Ballester et al., 2005] concluded that PFA is adequate for the study of shape variability and provides better "interpretability" than PCA as PFA models covariance between variables whereas PCA determines the factors which account for the total variance.

PCA and ICA are both linear factor analysis techniques. However, it is difficult for these approaches to cope with non-linear shape variations. Techniques such as kernel PCA [Schölkopf et al., 1998] and KDE are two alternatives to describe the non-linear data. The works proposed by [Cremers et al., 2006,Kim et al., 2007,Lu et al., 2012] are example papers that used non-linear dimensionality reduction techniques (e.g. kernel PCA and KDE) to incorporate shape prior into the image segmentation framework. For more information about other linear and non-linear factor analysis techniques refer to [Fodor, 2002,Bowden et al., 2000]

In addition to representation of shapes by a set of points (as usually done in e.g. PCA cf. (2.34)), a shape can be described by distance and angle information between different landmarks of an organ's shape [Wang et al., 2010,Nambakhsh et al., 2013]. [Wang et al., 2010] proposed a scale-invariant shape description by measuring the relative distances between pair of landmarks in a triplet, while [Nambakhsh et al., 2013] modeled the left ventricle (LV) shape in the cardium by calculating the distance between each point on the surface of the LV and a reference point in the middle of the LV provided by a user. More reviews on statistical shape models for 3D medical image segmentation can be found in [Heimann and Meinzer, 2009].

Beside the aforementioned statistical methods, some methods employed learning algorithms to impose a shape model into segmentation [Zhang et al., 2012,Kawahara et al., 2013]. In [Zhang et al., 2012], authors proposed a deformable segmentation method based on sparse shape composition and dictionary learning. In another work, [Kawahara et al., 2013] augmented the auto-context method [Tu and Bai, 2010] and trained sequential classifiers to learn what shape-features (e.g. volume of a segmentation) a good segmentation should have. Auto-context [Tu and Bai, 2010] is an iterative learning framework used for segmentation, which jointly learns the appearance and regularization distributions where the predicted labels from the previous iteration are used as input to the current iteration.

Physical model

In some medical applications, the biomechanical characteristics of tissues can be estimated. In this case, physical characteristics of a tissue can be incorporated into the segmentation framework as additional prior information to obtain more reliable segmentation.

Incorporating material elasticity property in image segmentation was first introduced in 1988 by [Kass et al., 1988] in which spring-like forces between snake's points is enforced. Following Kass' snakes model, several researchers also examined ways to extract vibrational (physical) modes of shapes based on finite element method (FEM); these include methods proposed by [Karaolani et al., 1989], [Nastar and Ayache, 1993] and [Pentland and Sclaroff, 1991]. In these frameworks, an object is modelled based on its vibrational modes similar to (2.34) where P_i represents vibrational modes instead of statistical modes.

When the physical characteristics of a tissue are known and several samples from the same tissue are available, one can take advantage of both statistical and physical models to obtain more accurate segmentation, as done by [Cootes and Taylor, 1995] and [Hamarneh et al., 2008] where statistical and vibrational modes of variation are combined into a single objective function.

[Schoenemann and Cremers, 2007] encoded an elastic shape prior into a segmentation framework by combining the shape matching and segmentation tasks. Given a shape template, they proposed an elastic shape matching energy term that maps the points of the evolving shape to the template based on two criteria: 1) points of similar curvature should be matched, and 2) a curve piece of the evolving shape should be matched to a piece of the template of equal size. Their method ensures the globally optimal solution.

2.2.7 Topological prior

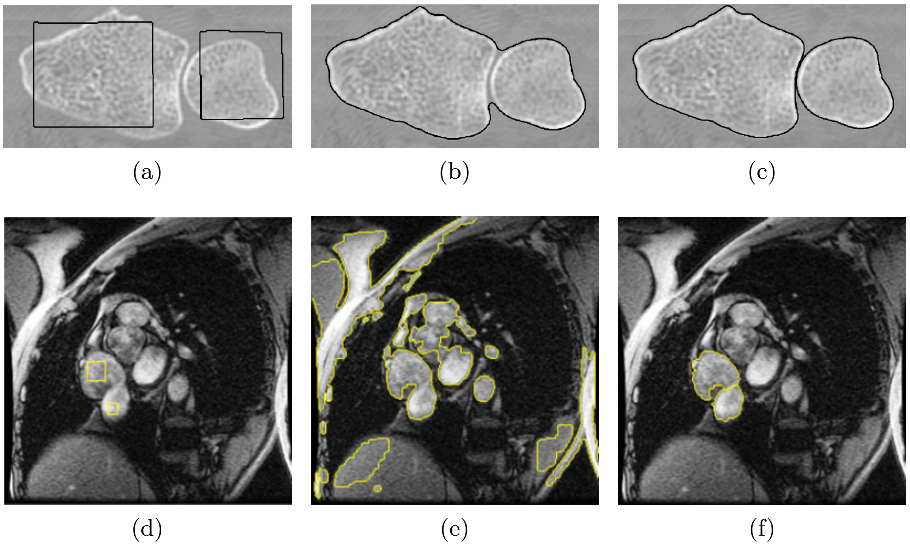
Many anatomical objects in medical images have a specific topology that has to be preserved after segmentation in order to obtain plausible results. There are two types of topology specification in the literature: *connectivity* and *genus*. Connectivity specification ensures that the segmentation of a single object is connected[2]. The genus information ensures that the final segmentation does not have any void region (if the object is known to be connected) or incorrectly fill void regions when the object is known to have internal holes [Grady, 2012]. For example, a doughnut-shape initial segmentation should keep its shape (doughnut) during the segmentation process.

[Han et al., 2003] proposed a level set-based method for segmenting objects with topology preservation. Their method is based on the *simple point* concept from digital topology [Bertrand, 1994]. A *simple point* is one that does not change the topology of the segmentation when it is added to or removed from a segmentation. Specifically, the proposed method checks the topological number at each iteration to detect any topology changes during the contour evolution. If the segmentation algorithm adds or removes only *simple points* from an initial segmentation, then the new segmentation will have the same genus as before.

Inspired by [Han et al., 2003], [Zeng et al., 2008] introduced *topology cuts* and cast the formulation of [Han et al., 2003] in a discrete setting. They showed that the optimization of their energy functional with topology preserving is NP-hard. In another work, [Vicente et al., 2008] proposed an interactive method in the discrete domain to segment objects with

---

[2] Formally, a segmentation S is connected if $\forall x, y \in S$, $\exists Path_{xy}$, s.t. if $z \in Path_{xy}$, then $z \in S$.


(a) (b) (c)

(d) (e) (f)

Figure 2.13: Incorporating topology constraint into medical image segmentation frameworks. **Top row:** Segmentation of a carpal bones CT image [Han et al., 2003]. **Bottom row:** Segmentation of cardiac ventricles [Zeng et al., 2008]. (a) Initialization. (b,c) Segmentation without (b) and with (c) topological constraint. (Images adapted from [Han et al., 2003] and [Zeng et al., 2008] by permission)

topology preserving. Their algorithm guarantees the connectivity between two designated points. The authors showed that their method will sometimes find the global optimum, while guaranteeing connectivity of the designated points. Figure 2.13 shows examples of encoding topological constraint in segmenting capral bones and cardiac ventricles.

2.2.8 Moment prior

In most segmentation methods that impose shape prior, deviations of the observed shape from training shapes are usually suppressed by the shape prior imposed. This is undesirable in medical image segmentation where pathological cases (abnormal cases that deviate from the training shapes of healthy organs) occur. Lower-order moments constraints seem to be an alternative to avoid this limitation.

- 0^{th} **order moment (size/area/volume):** The 0^{th} order moment corresponds to the size of an object. [Ayed et al., 2008] proposed to add the area prior into the level set framework to speed up the curve evolution and to prevent leakage in the final segmentation. Given an image I and the approximate area value of the target object ($\mathcal{A}$), their area energy term is defined as:

$$E^{Area}(\boldsymbol{x}) = \frac{1}{\mathcal{A}^2}\left(\int_{\Omega_{in}} d\boldsymbol{x} - \mathcal{A}\right)^2 \int_{\Omega_{in}} g(I(\boldsymbol{x}))d\boldsymbol{x} \ , \tag{2.35}$$

where Ω_{in} is the region inside the current segmentation and $g(.)$ attracts the evolving contour toward the high gradient regions (object boundaries).

- 1^{st} **order moment (location/centroid):** In case of having some rough information about the centroid of the target object, this valuable information can be encoded into a segmentation framework using the 1^{st} order moment as proposed in [Klodt and Cremers, 2011] (see below for more details).

- **Higher-order moment:** Generally, we can impose moment constraints of any order to refine the segmentation and capture fine-scale shape details. [Foulonneau et al., 2006] proposed to encode higher-order moments into a level sets framework using a local optimization scheme. Recently, [Klodt and Cremers, 2011] proposed a convex formulation to encode moment constraints. They used the objective function in the form of

$$E(u) = \int \rho(\boldsymbol{x})u(\boldsymbol{x})d\boldsymbol{x} + \int g(\boldsymbol{x})|Du(\boldsymbol{x})|d\boldsymbol{x} \ , \tag{2.36}$$

where $u \in BV : \mathbb{R}^d \to \{0,1\}$ is the labeling function and Du is the distributional derivative ($Du(\boldsymbol{x}) = \nabla u(\boldsymbol{x})$ for a differentiable u). Relaxing u to vary between 0 and 1, (2.36) becomes a convex optimization problem over the convex set $BV : \mathbb{R}^d \to [0,1]$. The global minimizer of the original problem ($E(u)$ before relaxing u) is obtained by finding the global minimum of the relaxed energy functional, u^*, and thresholding u^* by a value $\mu \in (0,1)$.

[Klodt and Cremers, 2011] imposed the 0^{th} order moment (i.e. area constraint in a 2D image) by bounding the area of u between c_1 and c_2 where $c_1 \leq c_2$ such that u lies in the set

$$\mathcal{C}_0 = \left\{ u | c_1 \leq \int_\Omega u d\boldsymbol{x} \leq c_2 \right\} . \tag{2.37}$$

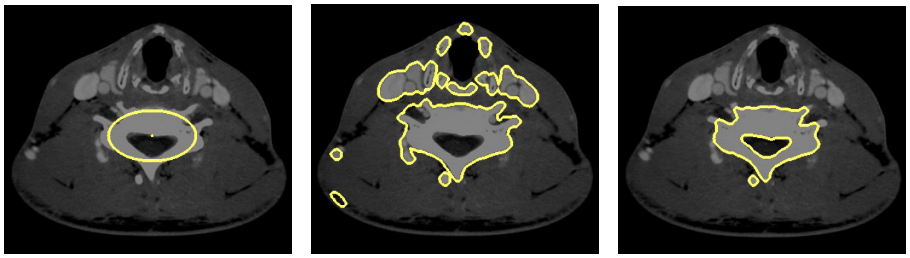
The exact area prior can be imposed by setting $c_1 = c_2$. The 1^{st} moment (i.e. centroid constraint) is imposed by constraining the solution u to the set $\mathcal{C}_1$ as:

$$\mathcal{C}_1 = \left\{ u | \mu_1 \leq \frac{\int_\Omega \boldsymbol{x} u d\boldsymbol{x}}{\int_\Omega u d\boldsymbol{x}} \leq \mu_2 \right\} \ , \tag{2.38}$$

where $\mu_1, \mu_2 \in \mathbb{R}^d$. The set $\mathcal{C}_1$ ensures that the centroid of the segmented object lies between μ_1 and μ_2. The centroid is fixed when $\mu_1 = \mu_2$.

In general, the n^{th} order moment constraint is imposed as:

$$\mathcal{C}_n = \left\{ u | A_1 \leq \frac{\int_\Omega (x_1 - \mu_1)^{i_1} \cdots (x_d - \mu_d)^{i_d} u d\boldsymbol{x}}{\int_\Omega u d\boldsymbol{x}} \leq A_2 \right\} \ , \tag{2.39}$$



(a) (b) (c)

Figure 2.14: CT segmentation (b) without and (c) with moments constraints. The area constraint limits the segmentation to the size of the ellipse that was clicked by the user (a) that resulting in more accurate result. (Images adapted from [Klodt and Cremers, 2011] by permission)

> where $i_1 + \cdots + i_d = n$, A_1,$A_2 \in \mathbb{R}^{d \times d}$ are symmetric matrices and $A_1 \leq A_2$ element wise. [Klodt and Cremers, 2011] proved that all these sets are convex. In their work, the above constraints (2.37)-(2.39) are all hard constraints. Alternatively, all of the aforementioned constraints can be enforced as soft constraints by including them into the energy functional using Lagrange multipliers. [Klodt and Cremers, 2011] mentioned that, in practice, imposing moments of more than the order of 2 is not very useful as users cannot interpret these moments visually and the improvements are very small.

In the discrete settings, [Lim et al., 2011] encoded area, centroid and covariance (2^{nd} order constraint) constraints into a graph based method. While their method does not guarantee a globally optimal solution, it can impose non-linear combinations of the aforementioned constraints as opposed to [Klodt and Cremers, 2011]. Figure 2.14 illustrates an example application of using moment constraints in CT segmentation.

2.2.9 Geometrical and region interactions prior

Anatomical objects often consist of multiple regions, each with a unique appearance model, and each has meaningful geometrical relationships or interactions with other regions of the object. Over the past decade, much attention has been given to incorporating geometrical constraints into the segmentation objective function.

In the continuous domain, several methods have been proposed based on coupled surfaces propagation to segment a single object in an image [Zeng et al., 1998, Goldenberg et al., 2002, Paragios, 2002]. [Vazquez-Reina et al., 2009] defined elastic coupling between multiple level set functions to model ribbon-like partitions. However, their approach was not designed to handle interactions between more than two regions. [Bloch, 2005] briefly reviewed the main fuzzy approaches that define spatial relationships including topological relations (set relationships and adjacency) as well as metrical relations (distances and directional relative position). None of the aforementioned methods guarantee the optimal solution.

[Wu et al., 2011] proposed a method for segmenting a region bounded by two coupled terrain-like surfaces by minimizing the intraclass variance. While their method yields globally optimal solution, it is limited to handling objects that can be “unfolded” into two coupled surfaces and can only segment a single object in an image.

[Ukwatta et al., 2012] also proposed a method that is based on coupling two surfaces for carotid adventitia and lumenintima segmentation. The advantage of their work over previous works is that they optimized their energy functional by means of convex relaxation. However, their method could only segment objects with coupled surfaces. Using the same framework as [Ukwatta et al., 2012], [Rajchl et al., 2012] presented a graphical model to segment the myocardium, blood cavities and scar tissue. Their method used seed points as hard constraints to distinguish the background from the myocardium. [Nambakhsh et al., 2013] proposed an efficient method for LV segmentation that iteratively minimizes a convex upper bound energy functional for a coupled surface. Their method implicitly imposes a distance between two surfaces by learning the LV shape.

In the discrete domain, [Li et al., 2006a] proposed a method to segment “nested objects” by defining distance constraints between the object’s surfaces with respect to a center point. As their formulation employed polar coordinates, their method could only handle star-shaped objects. Two containment and exclusion constraints between distinct regions have been encoded into a graph-cut framework by [Delong and Boykov, 2009] and [Ulén et al., 2013]. If only containment constraint is enforced, then both approaches guarantee the global solution. For a two-region object scenario (region A, B and background), the idea of [Delong and Boykov, 2009] is to create two graphs for A and B, i.e. $\mathcal{G}(\mathcal{P}^A, \mathcal{E}^A)$ and $\mathcal{G}(\mathcal{P}^B, \mathcal{E}^B)$. The segmentations of A and B are represented by the binary variables $\boldsymbol{f}^A$ and $\boldsymbol{f}^B$, respectively. The geometrical constraints between regions A and B are enforced by adding an additional penalty term W^{AB} defined in Table 2.2. This interaction term, W, is implemented in the graph construction by adding inter-layer infinity edges for each pixel (Figure 2.15(a)). Delong and Boykov employed what is known as the interaction term W

Table 2.2: Energy terms for encoding containment and exclusion constraints between regions A and B in (2.40) [Delong and Boykov, 2009].

A *contains* B		
f_p^A	f_q^B	W_{pq}^{AB}
0	0	0
0	1	∞
1	0	0
1	1	0

A *excludes* B		
f_p^A	f_q^B	W_{pq}^{AB}
0	0	0
0	1	0
1	0	0
1	1	∞

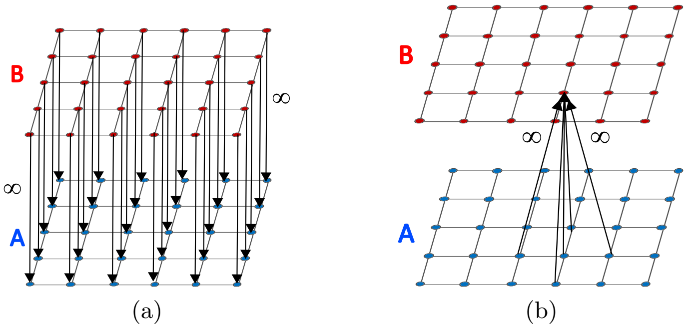


Figure 2.15: Enforcing containment constraint between objects A and B. (a) Graph construction to enforce A contains B proposed by [Delong and Boykov, 2009]. (b) Enforcing 1-pixel distance between A and B boundaries shown for one pixel only.

as follows:

$$\sum_{pq \in \mathcal{N}^{AB}} W_{pq}^{AB}(f_p^A, f_q^B) \ , \tag{2.40}$$

where $\mathcal{N}^{AB}$ is the set of all pixel pairs (p, q) at which region A is assigned some geometric interaction with region B. Table 2.2 lists energy terms for the region interaction constraints proposed in [Delong and Boykov, 2009]. The proposed graph-based method in [Delong and Boykov, 2009] guarantees the globally optimal solution for containment solution since its energy term is submodular. However, the energy for the exclusion constraint is nonsubmodular and thus harder to optimize. In some cases, because exclusion is supermodular everywhere, it is possible to make this term submodular by flipping the meaning of layer B's variables so that $f_p^B = 0$ designates the region's interior. Nonetheless, there are many useful interaction energy terms that cannot be modelled and optimized efficiently by [Delong and Boykov, 2009] and other approximation like quadratic pseudo-boolean optimization (QPBO) [Kolmogorov and Rother, 2007, Rother et al., 2007] or $\alpha\beta$-swap [Boykov et al., 2001] should be used for their optimization.

2.2.10 Spatial distance prior

In the literature, works that incorporate spatial distance prior may be categorized as follows:

- **Minimum distance:** In some applications the minimum distance between two structures must be enforced to ensure that sufficient separation between regions exists to obtain plausible results (e.g. distance between carotid adventitia and lumenintima). Examples of methods that employ this constraint include [Zeng et al., 1998, Goldenberg et al., 2002, Paragios, 2002] in the continuous settings and [Wu et al., 2011, Li et al., 2006a, Delong and Boykov, 2009, Ulén et al., 2013] in the discrete settings.

Looking at (2.40) for example, [Delong and Boykov, 2009] (and similarly, [Ulén et al., 2013]) enforce the minimum distance between two regions by defining the $\mathcal{N}^{AB}$ in (2.40). Figure 2.15(b) shows how 1-pixel margin between region boundaries is enforced by [Delong and Boykov, 2009, Ulén et al., 2013].

- **Maximum distance:** In other medical applications, maximum distance between regions is known a priori. For example, in cardiac LV segmentation, maximum distance between LV and its myocardium can be approximated. Enforcing a maximum distance between LV and its myocardium boundaries prevents the myocardium segmentation from growing too far from the LV. Maximum distance between two boundaries/surfaces is enforced as proposed in [Zeng et al., 1998, Goldenberg et al., 2002, Paragios, 2002] in the continuous settings. There is not much work on incorporating maximum distance between region boundaries in discrete settings except for [Wu et al., 2011] and [Schmidt and Boykov, 2012]. In [Wu et al., 2011] the maximum distance along with minimum distance prior for segmenting two-region ribbon-like objects can be enforced. To the best of our knowledge, the only work that solely focused on incorporating maximum distance in the discrete settings between regions for multi-region object segmentation is the approach proposed by [Schmidt and Boykov, 2012]. They modified the framework of [Delong and Boykov, 2009] by adding the Hausdorff distance prior to the MRF-based segmentation framework to impose maximum distance constraints. They showed that incorporating this prior into multi-surface segmentation is NP-hard due to the existence of supermodular energy terms.

- **Attraction/repulsion distance:** In applications like multi-region cell segmentation, distance between regions should be in a specific range. Maintaining a specific distance between different regions is enforced by enforcing attraction and repulsion forces between region boundaries as proposed in [Zeng et al., 1998, Goldenberg et al., 2002, Paragios, 2002] in the continuous settings. [Vazquez-Reina et al., 2009] specifically focused on attraction/repulsion interaction between two boundaries. They defined elastic couplings between level set functions using dynamic force fields to model ribbon-like partitions. Note that none of the above mentioned methods guarantee the globally optimal solution.

  In the discrete domain, [Wu et al., 2011] imposed attraction/repulsion forces between two surfaces by controlling the minimum and maximum distances between them. [Delong and Boykov, 2009] (similarly [Ulén et al., 2013, Schmidt and Boykov, 2012]) enforced the attraction/repulsion forces between pairs of regions, e.g. A and B, by penalizing the intersection of A and B (i.e. area/volume of $A - B$). Such constraint is encoded in (2.40) using the penalty terms defined in Table 2.3. In fact, containment and attraction are similar constraints but with different orientation. In the containment setting mentioned in Table 2.2, replacing infinity value with a positive

Table 2.3: Energy terms for encoding containment with attraction/repulsion between A and B regions.

A *attracts* B		
f_p^A	f_q^B	W_{pq}^{AB}
0	0	0
0	1	0
1	0	α
1	1	0

value for $W_{pq}^{AB}(0,1) > 0$ creates a spring-like repulsion force between inner and outer boundaries.

In graph-based methods, e.g. [Delong and Boykov, 2009, Ulén et al., 2013], increasing the distance (or thickness) between regions requires more edges to be added to the underlying graph, which increases the memory usage and computation time. In fact, to impose a distance constraint of w pixels between two regions, [Delong and Boykov, 2009] and [Ulén et al., 2013] need to add $O(w^2)$ extra edges *per pixel*. Therefore, although these graph-based methods are highly efficient in segmenting images with reasonable size and thickness constraint, they are not that efficient for large distance constraints.

In addition to the above mentioned approaches, methods based on the artificial life framework (deformable organisms) also employ spatial distance constraints to maintain the organism's structure [Hamarneh et al., 2009, Prasad et al., 2011]. In these models, the deformable organism evolves in a restricted way such that the distance between its skeleton and its boundary is restricted to be within a certain range.

2.2.11 Adjacency prior

Recently, several methods focused on ordering constraints and adjacency relationships on labels for semantic segmentation. As an example, "sheep" and "wolf" are unlikely to be next to each other and label transition from "sheep" to "wolf" should be penalized [Strekalovskiy et al., 2012].

In the discrete settings, [Liu et al., 2008] proposed a graph-based method to incorporate label ordering constraints in scene labeling and tiered[3] segmentation. They assumed that an image is to be segmented into five parts ("centre","left", "right", "above" and "bottom") such that a pixel labeled as "left" cannot be to the right of any pixel labeled as "center", etc. [Liu et al., 2008] encoded such constraints into the pair-wise energy term (regularization), i.e. $\sum_{(p,q)\in\mathcal{N}} V_{pq}(f_p, f_q)$. For example, if pixel p is immediately to the left of q, to prohibit $f_p =$"center" and $f_q =$"left", then one defines V_{pq}("center","left") $= \infty$. Generalizing this rule to other cases gives the following settings for $V_{p,q}$:

---

[3] Tiered labeling problem partitions an input image into multiple horizontal and/or vertical tiers.

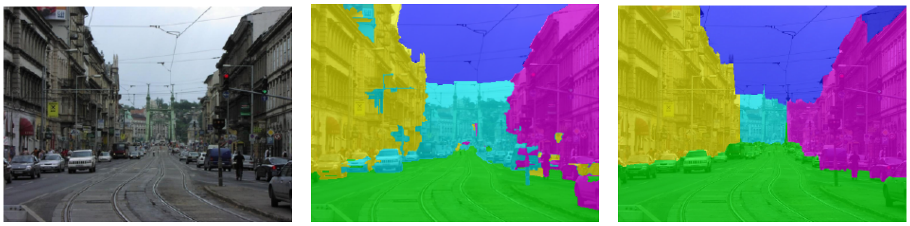

(a) Input image (b) Without label ordering constraint (c)

Figure 2.16: Tiered labelling. (a) Input image. Segmentation result (b) without and (c) with label ordering constraints. (Images adapted from [Strekalovskiy and Cremers, 2011] by permission)

$f_p\backslash f_q$	**L**	**R**	**C**	**T**	**B**
L	0	∞	w_{pq}	w_{pq}	w_{pq}
R	∞	0	∞	∞	∞
C	∞	w_{pq}	0	∞	∞
T	∞	w_{pq}	∞	0	∞
B	∞	w_{pq}	∞	∞	0

p is the left neighbour of q

$f_p\backslash f_q$	**L**	**R**	**C**	**T**	**B**
L	0	∞	∞	∞	w_{pq}
R	∞	0	∞	∞	w_{pq}
C	∞	∞	0	∞	w_{pq}
T	w_{pq}	w_{pq}	w_{pq}	0	∞
B	∞	∞	∞	∞	0

p is the top neighbour of q

Figure 2.16 illustrates a sample example of a tiered labelling. From optimization point of view and according to [Liu et al., 2008], α-expansion technique is more likely to get stuck in a local minimum when ordering constraints are used, as α-expansion acts on a single label (α) at each move. In order to improve on α-expansion moves, authors introduced two horizontal and vertical moves and allowed a pixel to have a choice of labels to switch to as opposed to just a single label α. Although their proposed optimization approach leads to better results (compared to α-expansion approach), the globally optimal solution is still not guaranteed. [Felzenszwalb and Veksler, 2010] proposed an efficient dynamic programming algorithm to impose similar constraints as [Liu et al., 2008] but with much less complexity. Their method computes the globally optimal solution in the class of tiered labelings.

In the continuous domain, [Strekalovskiy and Cremers, 2011] proposed a generalized label ordering constraint which can enforce many complex geometric constraints while maintaining convexity. This method requires that the constraint term obeys the triangle inequality, a requirement that was later relaxed by introducing a convex relaxation method for non-metric priors [Strekalovskiy et al., 2012]. To do so, authors enforce non-metric label distances in order to model arbitrary probabilities for label adjacency. The distance between different labels[4] operates only directly on neighbouring pixels. This often leads to artificial one pixel-wide regions between labels to allow the transition between labels with very high or infinite distance. For example, both the “wolf” and “sheep” labels can be next to “grass” but they cannot be next to each other [Strekalovskiy et al., 2012]. The method

---

[4] Note that this distance is between label classes and is not a spatial distance.

proposed in [Strekalovskiy et al., 2012] creates an artificial "grass" region between "wolf" and "sheep" to allow for this transition. Obviously this one-pixel wide distance between "wolf" and "sheep" will not make the sheep feel secure. Hence, larger neighbourhood than one pixel is needed to impose the proximity and propagate the co-occurrence penalty. [Bergbauer et al., 2013] addressed this issue and proposed a *morphological proximity* prior for semantic image segmentation in a variational framework. The idea is to consider pixels as adjacent if they are within a specified neighbourhood of arbitrary size. Consider two regions i and j and their indicator functions u_i and u_j, respectively. To see if i and j are close to each other, the overlap between the dilation of the indicator function u_i denoted by d_i and the indicator function of u_j is computed. The dilation of u_i is formulated as:

$$d_i(\boldsymbol{x}) = \max_{z \in \mathcal{S}e} u_i(\boldsymbol{x} + z), \quad \forall \boldsymbol{x} \in \Omega \tag{2.41}$$

with a structuring element $\mathcal{S}e$. For each pair of region i and j, the proximity penalty term is defined as:

$$\sum_{1 \leq i \leq j \leq n} \int_{\Omega} A(i,j) d_i(\boldsymbol{x}) u_j(\boldsymbol{x}) d\boldsymbol{x} \ , \tag{2.42}$$

where $A(i,j)$ indicates the penalty for the co-occurrence of label j in the proximity of label i such that $A(i,i) = 0$. This penalty term is relaxed and added to an energy functional (along with regional and regularization terms), which is then optimized with the help of Lagrange multipliers.

To the best of our knowledge, the adjacency and proximity priors as described above have not been utilized in medical image segmentation yet.

2.2.12 Number of regions/labels

In most segmentation problems, the number of regions is assumed to be known beforehand. However, it is not the case in many applications and considering a fixed number of labels in these cases often causes over-segmentation.

The intuitive way to handling this problem is to penalize the total number of labels. For the given maximum number of regions/labels (at most n labels), which is available in most applications, [Zhu and Yuille, 1996] proposed to partition images based on the following energy functional in the continuous domain:

$$\min_{\Omega_i} \sum_{i=1}^{n} \left\{ \int_{\Omega_i} \rho(\ell_i, \boldsymbol{x}) d\boldsymbol{x} + \int_{\partial \Omega_i} ds \right\} + \gamma M \ , \tag{2.43}$$

where Ω_i is the region corresponding to label ℓ_i; $\rho(\ell_i, \boldsymbol{x})$ is the data term that encodes the model of ℓ_i at pixel $\boldsymbol{x}$; the second term is the regularization term; and M in the third term is the number of non-empty partitions (known as label cost prior). [Zhu and Yuille, 1996] optimized the above energy functional using a local optimization technique which converges to a local minimum. This approach was later adapted in the level set formulation by [Kadir and Brady, 2003, Ben Ayed and Mitiche, 2008, Brox and Weickert, 2006] that allow region-merging. A convex formulation of such constraint was proposed by [Yuan et al., 2012]. They enforced the label cost prior into multi label segmentation by solving the following convex optimization problem:

$$\min_{u(\boldsymbol{x})} \sum_{i=1}^{n} \left\{ \int_{\Omega} u_i(\boldsymbol{x}) \rho(\ell_i, \boldsymbol{x}) d\boldsymbol{x} + \int_{\partial\Omega} |\nabla u_i(\boldsymbol{x})| d\boldsymbol{x} \right\} + \gamma \sum_{i=1}^{n} \max_{\boldsymbol{x} \in \Omega} u_i(\boldsymbol{x}) \ ,$$
$$s.t. \sum_{i=1}^{n} u_i(\boldsymbol{x}) = 1, \quad u_i(\boldsymbol{x}) \geq 0; \quad \forall \boldsymbol{x} \in \Omega. \tag{2.44}$$

In the discrete domain, [Delong et al., 2012a] developed an α-expansion method to optimize a general energy functional with incorporated label cost in a graph-based framework. Along with a unary (data) and pairwise (regularization) terms, [Delong et al., 2012a] penalized each unique label that appears in the image by introducing the following term:

$$\sum_{l \in \mathcal{L}} h_l \cdot \delta_l(f) \tag{2.45}$$

$$\delta_l(f) = \begin{cases} 1 & \exists p \in \Omega : f_p = l \\ 0 & otherwise \end{cases} , \tag{2.46}$$

where h_l is the non-negative label cost of label l.

2.2.13 Motion prior

The segmentation and tracking of moving objects in videos have a wide variety of applications in medical image analysis, e.g. echocardiography [Dydenko et al., 2006]. [Paragios and Deriche, 1999] used motion prior to constrain the evolution of a level set function. They integrated the motion estimation and tracking into the level set-based framework assuming that the motion is linear. [Dydenko et al., 2006] proposed a method to segment and track the cardiac structure in high frame rate echocardiographic images. Under an affine motion model, the motion field is estimated from the level set evolution. More complex motion prior is used in object tracking. For example, to track the LV in echocardiography, [Orderud et al., 2007] employed the Kalman filter, which is an optimal recursive algorithm that uses a series of measurements observed over time to estimate the desired variables (i.e. displacement in motion estimation).

As these methods focused more on tracking, which is outside the scope of this report, we refer interested readers to [Tang et al., 2012].

2.2.14 Model/Atlas

Atlas-based segmentation has also been particularly useful in medical image analysis applications. An atlas has the ability to encode (non-pathological) spatial relationships between multiple tissues, anatomical structures or organs. In atlas-based image segmentation, the image is non-rigidly deformed and registered with a model or atlas that has been labelled previously. Applying the inverse transformation of the labels to the image space gives the segmentation. However, atlas-based segmentation has so far been restricted to single (albeit multi-part or multi-region) object instance, i.e. it does not address spatially-recurring objects in the scene. Also, atlases usually are built from datasets of manually segmented images. These manual segmentations may not always be available, or it might not be straightforward to define a representative template for a given object. Several methods focused on such segmentation [Gee et al., 1993, Collins et al., 1995, Collins and Evans, 1997, Iosifescu et al., 1997].

The performance of atlas-based segmentation techniques relies on an accurate registration. Surveying registration methods is beyond the scope of this report. Interested readers may refer to [Sotiras et al., 2013], [Hill et al., 2001], and [Tang and Hamarneh, 2013] for more details on image registration.

In the field of computer vision (non-medical), a few techniques used 3D models of objects (more realistic but more complex) to segment 2D images. [Prisacariu and Reid, 2012] proposed a variational method to segment an object in a 2D image by optimizing a Chan-Vese energy functional with respect to six pose parameters of the object model in 3D. The idea is to transform the object's model in 3D so that its projection on the 2D image delineates the object of interest. Consider segmenting a single object in an image with the following energy functional:

$$E(\phi(\boldsymbol{x})) = \int_\Omega \rho_f(\boldsymbol{x}) H(\phi(\boldsymbol{x})) + \rho_b(\boldsymbol{x})(1 - H(\phi(\boldsymbol{x}))) d\Omega \, , \tag{2.47}$$

where ρ_f and ρ_b are two monotonically decreasing functions, measuring matching quality of the image pixels with respect to the foreground and background models, respectively. Instead of optimizing $E(\phi)$ with respect to the level set function ϕ, authors in [Prisacariu and Reid, 2012] proposed to minimize $E(\phi)$ with respect to the pose parameters (ξ_i) of the object of interest in 3D space:

$$\frac{\partial E}{\partial \xi_i} = (\rho_f - \rho_b) \frac{\partial H(\phi)}{\partial \xi_i} = (\rho_f - \rho_b) \delta(\phi) \begin{bmatrix} \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial \xi_i} \\ \frac{\partial y}{\partial \xi_i} \end{bmatrix} \, . \tag{2.48}$$

Unlike [Prisacariu and Reid, 2012], [Sandhu et al., 2011] derived a gradient flow for the task of non-rigid pose estimation for a single object and used kernel PCA to capture the variance in the space of shapes. Later, [Prisacariu et al., 2013], introduced non-rigid pose parameters into the same optimization framework. They captured 3D shape variance by learning non-linear probabilistic low dimensional latent spaces, using the Gaussian process latent variable dimensionality reduction technique. All three aforementioned works ([Prisacariu and Reid, 2012, Prisacariu et al., 2013, Sandhu et al., 2011]) assume that the camera parameters (for 3D to 2D projection) are given.

2.3 Chapter summary

In this chapter we reviewed and summarized previous works that incorporated different types of prior information in their segmentation framework. This review form the basis for this thesis. Specifically, Chapter 3 will build upon the review of prior information in Section 2.2.1-2.2.10, presenting a method for segmentation of overlapping multi-region cells in microscopy/histology images. Chapter 4 and 5 present two new segmentation frameworks to delineate multi-region objects with part configuration constraints outlined in Sections 2.2.9 and 2.2.10 with the focus on space and time complexity as well as optimizability. Chapter 6 presents an efficient method to segment multi-region structures in multi-view endoscopic videos using 3D pre-operative model and laparoscopic camera motion model, the priors discussed in Sections 2.2.13 and 2.2.14.

Chapter 3

Evolutionary computation technique for targeted spatially-recurring multi-part object segmentation

Histology and microscopy image analysis plays a crucial role in studying diseases such as cancer and in obtaining reference diagnosis (e.g. biopsy histopathology). Automatically segmenting cells in such images is one of the preliminary steps toward automatic image analysis and computer-aided diagnosis. In spite of recent advances in segmenting cells based on some homogeneity and smoothness characteristics, segmenting complex cells with a non-homogeneous appearance (with multiple internal regions) remains challenging. This problem becomes even more challenging when these complex cells overlap. Previous works addressed cell overlapping, for single-region cells, using post-processing [Wu et al., 2012, Yang et al., 2008,Mao et al., 2006] (e.g. finding connected components and using parameter sensitive morphological operations [Mao et al., 2006]). However, cells in histology and microscopy images typically consist of multiple regions (e.g. membrane, nucleus, nucleolus), each with a unique appearance model (intensity, color or texture) and unique geometric characteristics (e.g. cell size and shape prior). Furthermore, well defined spatial interactions usually exist between different regions of a cell (e.g. membrane contains nucleus, and nucleus contains nucleolus). Most existing methods have only considered simple structured cells and ignored their complex composition [Ali et al., 2011,Cheng et al., 2011,Bernardis et al., 2011].

There are many types of priors that benefit the segmentation of spatially-recurring cells with appearance inhomogeneity along with cell-overlapping. Incorporating several energy terms enables us to describe the problem in more detail and thus obtain a more accurate formulation. On the other hand, adding more terms to the objective function generally makes it more complicated and harder to optimize.

In this chapter, we opt for ensuring the objective function is flexible enough (even if it is nonconvex) to accurately capture the intricacies of the cell segmentation problem. To optimize such objective function and to deal with imminent problems like initialization and local optima, we adopt a global optimization evolutionary computation method, genetic algorithm, which can attain solutions close to the global optimum, does not require Euler Lagrangian or energy gradient calculations, is generally parallelizable, and allows for arbitrarily complex objective functions. Our framework allows us to leverage a variety of expert knowledge or priors by adding them as additional terms in the objective function without being overly concerned about convexification. Finally, to deal with the spatially recurring aspect in cell segmentation, we use genetic algorithms with *tribes* [Turner et al., 1996] to obtain multiple distinct solutions for our framework.

3.1 Methods

Given an n-channel 2D image $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n$, the goal is to segment the objects of interest (cells) in I. We represent the boundary of each object (or each part of a multi-region object) by $\boldsymbol{X}_i = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \in \Omega$, where i indicates the i^{th} part/region. Next, we review the useful priors in microscopy images that we can leverage.

- **Shape:** When an object has a specific geometrical shape (e.g. circle, ellipse, rectillipse, etc.) we model it by shape parameters such as $\boldsymbol{b} = \{\text{radius, major axis, eccentricity, etc.}\}$. When no clear geometrical representation exists, we model a shape (e.g. i^{th} region's shape) by its statistical (from m training samples) and vibrational properties as:

$$\boldsymbol{X}_i \approx \bar{\boldsymbol{X}}_i + \boldsymbol{P}_i^c \boldsymbol{b}_i \ , \tag{3.1}$$

where $\bar{\boldsymbol{X}}_i$ is the average of a set of pose-normalized training shapes and

$$\boldsymbol{P}_i^c = \boldsymbol{P}_{stat} + \beta \boldsymbol{P}_{vib} \tag{3.2}$$

is the combined (statistical P_{stat} and vibrational P_{vib}) covariance matrix [Hamarneh et al., 2008, Cootes and Taylor, 1995], $\beta \propto 1/m$ is the balancing parameter and $\boldsymbol{b}_i = (b_i^1, \cdots, b_i^t)^T$ is a vector of shape parameters. We use the Mahalanobis distance to measure the validity of a novel shape $\boldsymbol{X}_j$ by:

$$F_i^{sh}(\boldsymbol{X}_j) = e^{-\sqrt{(\boldsymbol{X}_j - \bar{\boldsymbol{X}}_i)^T (\boldsymbol{P}_i^c)^{-1} (\boldsymbol{X}_j - \bar{\boldsymbol{X}}_i)}} \ . \tag{3.3}$$

- **Appearance:** Histology/microscopy images typically have different discriminative **color** channels, $\boldsymbol{c} = \{c_1, \cdots, c_q\}$, where c_i is a color channel, e.g. R, G, B, etc. Further, cells (and their constitutive regions) might also have different discriminative **texture**, $\boldsymbol{t} = \{t_1, \cdots, t_r\}$, where t_i is a texture channel, e.g. multi-scale Gabor or Haar-like features. To leverage cell appearance (color+texture), we concatenate $\boldsymbol{c}$ and $\boldsymbol{t}$ into a regional appearance vector $\boldsymbol{r}$ calculated within inner and outer bands around $\boldsymbol{X}_j$, $\Omega^j_{in,d}$ and $\Omega^j_{out,d}$, with thickness d. This band-localization is important since cells can contain inner parts (e.g. nucleus) and can be adjacent to other objects (e.g. other cells), both of which can pollute the regional appearance measures if a band is not used. In addition, by using an inner versus outer band, we are encoding the boundary polarity (e.g. dark to bright). We define the appearance fitness function for object i as:

$$F_i^{ap}(\boldsymbol{X}_j) = \frac{1}{2}\Big(\frac{1}{|\Omega^j_{in,d}|}\int_{\Omega^j_{in,d}} P(\boldsymbol{x} \in O_i)d\boldsymbol{x} + \frac{1}{|\Omega^j_{out,d}|}\int_{\Omega^j_{out,d}} P(\boldsymbol{x} \in B_i)d\boldsymbol{x}\Big) \tag{3.4}$$

where $P(\boldsymbol{x} \in O_i)$ and $P(\boldsymbol{x} \in B_i)$ are the probabilities of a given pixel $\boldsymbol{x} \in \Omega^j_{in,d} \cup \Omega^j_{out,d}$, belonging to object i (O_i) and its background (B_i), respectively, and are estimated by training a random forest (RF) consisting of N_b binary decision trees. To segment an R-region object in I, $R+1$ patches within R different regions of the object plus background are selected (i.e. regions $\mathcal{L} = \{0, \cdots, R\}$) to train the RF. After training, for each pixel $\boldsymbol{x}$, the feature channels, $\boldsymbol{r}(\boldsymbol{x})$, are propagated through each tree resulting in the probability $P_j(\boldsymbol{x} \in k|\boldsymbol{r}(\boldsymbol{x}))$, for the j^{th} tree, where $k \in \mathcal{L}$. These probabilities are combined into a forest's joint probability:

$$p(\boldsymbol{x} \in k|\boldsymbol{r}(\boldsymbol{x})) = \frac{1}{N_b}\sum_{j=1}^{N_b} p_j(\boldsymbol{x} \in k|\boldsymbol{r}(\boldsymbol{x})) \tag{3.5}$$

to determine the probability of $\boldsymbol{x}$ belonging to class k. Note that $O_i, B_i \in \mathcal{L}$.

- **Edge:** Since boundaries of cells and their parts exhibit appearance discontinuities, we incorporate edge information in the image by defining the following edge fitness term:

$$F^{ed}(\boldsymbol{X}_j) = \frac{1}{|\boldsymbol{X}_j|}\oint_{\boldsymbol{X}_j} e^{-g(\boldsymbol{X}_j)}\ , \tag{3.6}$$

where $g(.) = 1/(\epsilon+\lambda)$, λ is the maximum eigenvalue of the structure tensor J^TJ (generalizes scalar field gradients to those of vector fields), where J is the Jacobian matrix of the weighted feature channels, $\boldsymbol{w}^T\boldsymbol{r}$, and the vector $\boldsymbol{w}$, resulting from training the RF, is the importance of each feature channel in discriminating inside versus outside of an object (i.e. maximizes boundary edge response).

- **Pose:** Each cell has a specific size, orientation and position in the image. Given the training data, we estimate the average area ($\bar{A}$) and the principal orientation ($\bar{\theta}$) of cells and use them for imposing constraints on the solutions. We specify the position of a cell by its centroid, $\boldsymbol{x}^c = (x^c, y^c)$.

- **Geometrical constraints:** In addition to the shape and appearance properties of an object (color, texture, edge, shape and pose), in multi-region objects, meaningful geometrical relationships typically exist between different object's regions, e.g. regions *contain/exclude* others. To enforce containment and exclusion between two regions, e.g. $\boldsymbol{X}_j$ is contained in $\boldsymbol{X}_i$, or, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are excluded from one another, the following constraints are imposed:

$$D(\boldsymbol{x}_j) \underset{exclude}{\overset{contain}{\overset{\geq}{\leq}}} 0 \ , \quad \forall \boldsymbol{x_j} \in \boldsymbol{X}_j \ , \quad D(x) = SDM(\boldsymbol{X}_i) \ , \tag{3.7}$$

where $SDM(\boldsymbol{X}_i)$ is the signed distance map of $\boldsymbol{X}_i$ and is positive inside and negative outside $\boldsymbol{X}_i$. Eq. (3.7) is a general constraint for convex and non-convex shapes. However, for convex shapes, as we typically have in microscopy images, we adopt the following simplification for containment (i contains j):

$$||\boldsymbol{x}_j^c - \boldsymbol{x_i}|| - ||\boldsymbol{x}_j^c - \boldsymbol{x_j}'|| \geq 0 \ , \tag{3.8}$$

and exclusion (i and j are excluded from one another):

$$||\boldsymbol{x}_j^c - \boldsymbol{x_i}|| - ||\boldsymbol{x}_j^c - \boldsymbol{x_j}'|| < 0, \ \forall \boldsymbol{x_i} \in \boldsymbol{X}_i \ , \tag{3.9}$$

for faster computation, where $\boldsymbol{x}'_j \in \overrightarrow{\boldsymbol{x}_j^c \boldsymbol{x}_i} \cap \boldsymbol{X}_j$ and $\boldsymbol{x}_j^c = (x_j^c, y_j^c)$ is the spatial position of $\boldsymbol{X}_j$.

- **Inter-part adjacency:** In biomedical applications, the minimum (d^{min}) and maximum (d^{max}) distances between two adjacent regions of an object are sometimes known. Bounding the minimal and maximal distances between two adjacent boundaries (e.g. i and j) from below and above, respectively, prevents segmentation leakage and improves the results. We impose these constraints by:

$$\min(f_{ij}, f_{ji}) \geq d_{ij}^{min} \quad , \quad \max(g_{ij}, g_{ji}) \leq d_{ij}^{max}, \tag{3.10}$$

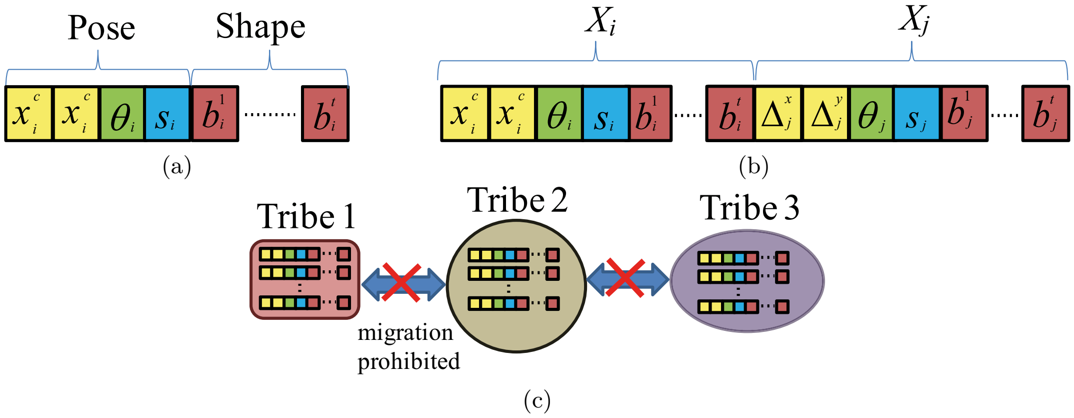
where

$$f_{ij} = \min_{\boldsymbol{x}_i \in \boldsymbol{X}_i} \min_{\boldsymbol{x}_j \in \boldsymbol{X}_j} ||\boldsymbol{x}_i - \boldsymbol{x}_j|| \text{ and} \tag{3.11}$$

$$g_{ij} = \max_{\boldsymbol{x}_i \in \boldsymbol{X}_i} \min_{\boldsymbol{x}_j \in \boldsymbol{X}_j} ||\boldsymbol{x}_i - \boldsymbol{x}_j|| \ . \tag{3.12}$$

For efficiency, we only calculate and restrict f_{ij} and g_{ij} (not f_{ji} and g_{ji}).

- **User interaction:** User interaction is another useful prior. This prior can be applied on the boundary and/or the region of an object by providing corresponding seed points,



(a) (b) (c)

Figure 3.1: Chromosome structure for (a) a single-region and (b) a two-region object. The position of the second region, (Δ_j^x, Δ_j^y), is computed relative to the first object's position. (c) Tribe-based GA. No migration is allowed between tribes.

$\boldsymbol{s}_i^b$ and $\boldsymbol{s}_i^r$, and force the solution to satisfy the following constraints $\boldsymbol{s}_i^r \in \Omega^i_{in,d=\infty}$ and $\boldsymbol{s}_i^b \in \Omega^i_{in,d=\epsilon} \cup \Omega^i_{out,d=\epsilon}$.

Fitness function: The overall fitness function (for an R-region cell) is constructed by integrating all above mentioned information as:

$$F_{total}(\boldsymbol{X}) = \sum_{i=1}^{R} \Big(F_i^{sh}(\boldsymbol{X}) + F_i^{ap}(\mathcal{M}_i(\boldsymbol{X})) + F^{ed}(\mathcal{M}_i(\boldsymbol{X})) \Big), \tag{3.13}$$

subject to

$$\begin{array}{llll} \boldsymbol{\mathit{Shape\ and\ pose:}} & |b_i^j| \le 3\sqrt{\lambda_i^j} & |\theta_i - \bar{\theta}_i| \le 3\sqrt{\lambda_i^\theta} & |\mathrm{Area}(\boldsymbol{X}_i) - \bar{A}_i| \le 3\sqrt{\lambda_i^A} \\ \boldsymbol{\mathit{User\ interaction:}} & \boldsymbol{s}_i^r \in \Omega^i_{in,\infty} & \boldsymbol{s}_i^b \in \Omega^i_{in,\epsilon} \cup \Omega^i_{out,\epsilon} & \\ \boldsymbol{\mathit{Geometry:}} & eq.(3.7) & & \\ \boldsymbol{\mathit{Adjacency:}} & eq.(3.10), & & \end{array} \tag{3.14}$$

where

$$\boldsymbol{X} = \overline{\boldsymbol{X}} + \boldsymbol{P}^c \boldsymbol{b}\,, \tag{3.15}$$

$$\mathcal{M}_i(\boldsymbol{X}) = s_i \boldsymbol{R}_i \boldsymbol{X} + \boldsymbol{T}_i\,. \tag{3.16}$$

$\mathcal{M}_i$ is a similarity transformation with rotation $\boldsymbol{R}_i$, scaling s_i, and translation $\boldsymbol{T}_i$. λ_i^j in (3.13) is the j^{th} eigenvalue of i^{th} region's covariance matrix and λ_i^A and λ_i^θ are the area and orientation variance of i^{th} region, respectively, obtained from the training data.

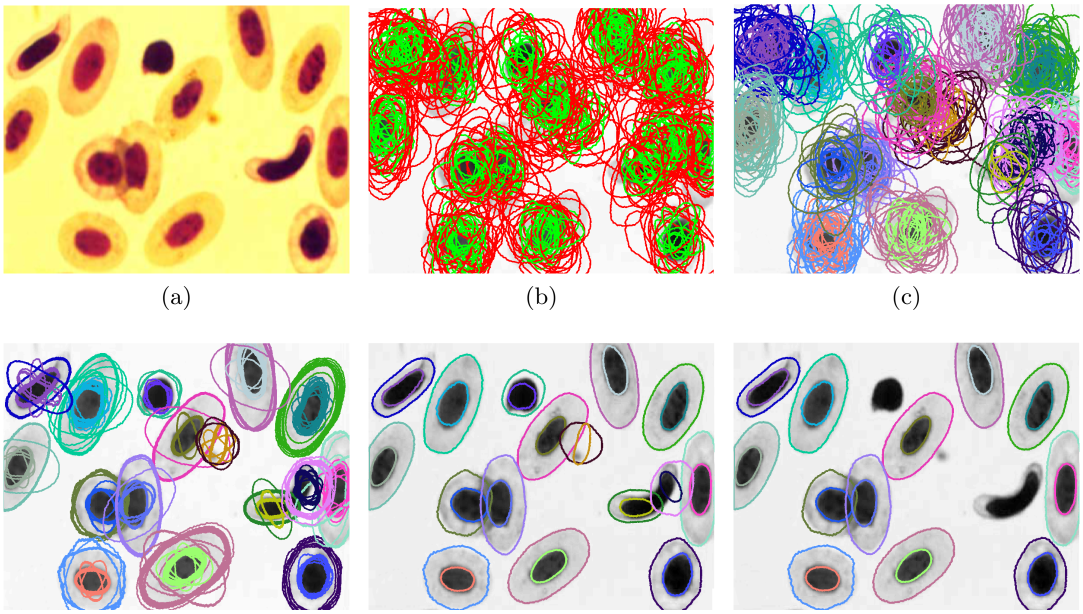

(a) (b) (c) (d) (e) (f)

Figure 3.2: Fish blood cells segmentation. (a) Original image. (b) Initial population. (c) Tribes formation. (d) Converged population. (e) Individuals with best fitting score in each tribe. (f) Final result.

To find the best fit for such a complex fitness function (3.13), we adopt GA as a global optimization tool. Although GA does not strictly guarantee the global solution, our results confirm the ability of this approach to accurately segment the spatially-recurring, multi-region cells with partial overlap. In GA, each individual solution is represented by a chromosome consisting of several genes (Figure 3.1(a)). The first four genes describe the individual's pose information, x_i^c and y_i^c are the spatial position of i^{th} region and θ_i, s_i and $b_i^1, \cdots, b_i^t$ are its orientation, scale and shape parameters.

Encoding multi-region object's information into GA: For simplicity and to conserve space, here we consider a two-region object scenario. Assuming a cell consists of two regions: $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, where $\boldsymbol{X}_1$ contains $\boldsymbol{X}_2$, we compute $\bar{A}$ and $\bar{\theta}$, as well as the shape parameters, $\boldsymbol{b}$, for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, separately, as described before. We represent the cell while encoding the interaction between its regions by concatenating the two chromosomes of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. However, the position of $\boldsymbol{X}_2$, (x_2^c, y_2^c), is computed relative to $\boldsymbol{X}_1$'s position, $(x_1^c$, $y_1^c)$, and its corresponding genes are replaced by Δ_x and Δ_y (Figure 3.1(b)). Δ_x and Δ_y allow $\boldsymbol{X}_2$ to move in small distances around its relative position to $(x_1^c$, $y_1^c)$. The average relative distance between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as well as limits on Δ_x and Δ_y are learned from the training data.

Each object (cell) typically recur in different parts of the image domain. To deal with such spatially recurring aspect of cell segmentation, we use GA with *tribes* to obtain multi-

Table 3.1: Comparison against state-of-the-art methods on counting lymphocytes on HIMA dataset

Method	m_d	s_d	m_N	s_N
[Kuse et al., 2010]	3.04	3.40	14.01	4.4
[Panagiotakis et al., 2010]	2.87	3.80	14.23	6.3
[Graf et al., 2010]	7.60	6.30	24.50	16.2
[Cheng et al., 2010]	8.10	6.98	26.67	12.5
[Kuse et al., 2011]	3.14	0.93	4.30	3.09
[Bernardis et al., 2011]				
($\rho = 5$)	3.22	3.92	5.40	3.68
($\rho = 4$)	2.84	2.89	8.20	4.75
($\rho = 2$)	1.12	0.71	16.75	7.47
Our method	1.40	0.77	6.30	4.20

ple distinct solutions (i.e. cells). In tribes-based GA, the whole population is grouped into several tribes. During the GA evolution and in the gene crossover phase, any two selected parents must be from the same tribe. In fact, tribes are too choosy about who is allowed to join them (Figure 3.1(c)); they do not accept any stranger (no migration is allowed) and even children who are not similar to the tribe's population are rejected. This tribes-based GA allows for the desired multiple distinct solutions. We choose the tribes' membership based on the spatial position of each member (cell), i.e. (x^c, y^c).

Initialization and implementation: We used 6 channels of colors (RGB+ HSV) and 3-channel Gabor features as our regional cues. Gabor filters were calculated in 8 different orientations and 3 different scales and were summed up across orientations to obtain rotational-invariance texture features. For RF, we used $N_b = 50$ binary trees. We randomly spread 10,000 random chromosomes over the image wherever the probability of existing cells obtained from RF is large enough, i.e. $P(\boldsymbol{x} \in O|\boldsymbol{r}(\boldsymbol{x})) > 0.6$ (Figure 3.2(b)). d_{ij}^{min} and d_{ij}^{max} were set based on the training dataset. Although our method can handle user interaction, none was used in our experiments. The crossover and mutation rates were fixed to 0.7 and 0.01, respectively, in all of our experiments. Individuals that are within a distance of ℓ pixels from each other establish a tribe (Figure 3.2(c)). We implemented our method in MATLAB in a way that all individuals are evaluated simultaneously in parallel. After convergence, Figure 3.2(d), the best solution in each tribe is examined (Figure 3.2(e)). We use the final fitness measure as a confidence measure, where the user can request displaying e.g. the top 10% confident segmentation. According to our fitness function, the ideal fitness score is 3. In all of our experiments we kept the solutions that are higher than 2.4 (top 20% confident segmentation) as the final solution (Figure 3.2(f)).

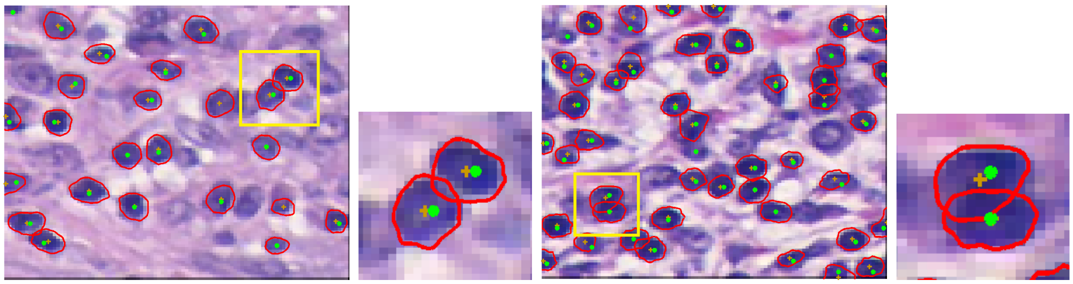


Figure 3.3: Sample results on HIMA dataset [Gurcan et al., 2010]. Red contours: our segmentation result; gold+: our segmentation centroid; green dots: ground truth.

3.2 Experiments

3.2.1 Single-region overlapping cells segmentation

In our first experiment, we evaluated our method on stained breast cancer tissue images used in ICPR 2010's HIMA contest on 'Counting Lymphocytes on Histopathology Images' [Gurcan et al., 2010]. We benchmarked our results against the state-of-the-art methods, including the contest's finalists. We used the centroid of the segmented cells to compare our results against the expert annotated ground truth (GT). Table 3.1 quantitatively compares our results against the competing methods. The evaluation criteria are based on the Euclidean distance, d_E, between the GT and centroid of the segmented lymphosytes, as well as the absolute difference, N, between the true number of cells in GT and detected cells. m and s in Table 3.1 are mean and standard deviation, respectively. In Table 3.1, [Bernardis et al., 2011] reported results for different thresholds, ρ, on the same dataset. While for some cases (e.g. $\rho = 2$) they achieved better distance accuracy, d_E, than our method, they found less true cells (bigger N). On the other hand, they obtained better detection rate (smaller N) for $\rho = 5$ but with less accuracy, d_E. We emphasize that their method has been designed for single-region cells only. Figure 3.3 demonstrates how our method distinguishes the merged cells. Our method can not only detect and segment the single-region cells but also delineate the different boundaries of a multi-region cell.

3.2.2 Multi-region overlapping cells segmentation

To further showcase our method, we ran a second experiment on another dataset, MICR, consisting of 20 different histology and microscopy images with multi-region cells. Our results in Figure 3.4 verify the use of proposed constraints (geometrical, thickness and shape) as compared to ubiquitous unconstrained image segmentation methods; graph cuts (GC), and constrained methods; Delong and Boykov [Delong and Boykov, 2009] (DB). While GC is not designed to segment cells, its results show the issues and difficulties involved in segmenting complex, multi-region cells. Delong and Boykov's method incorporates con-

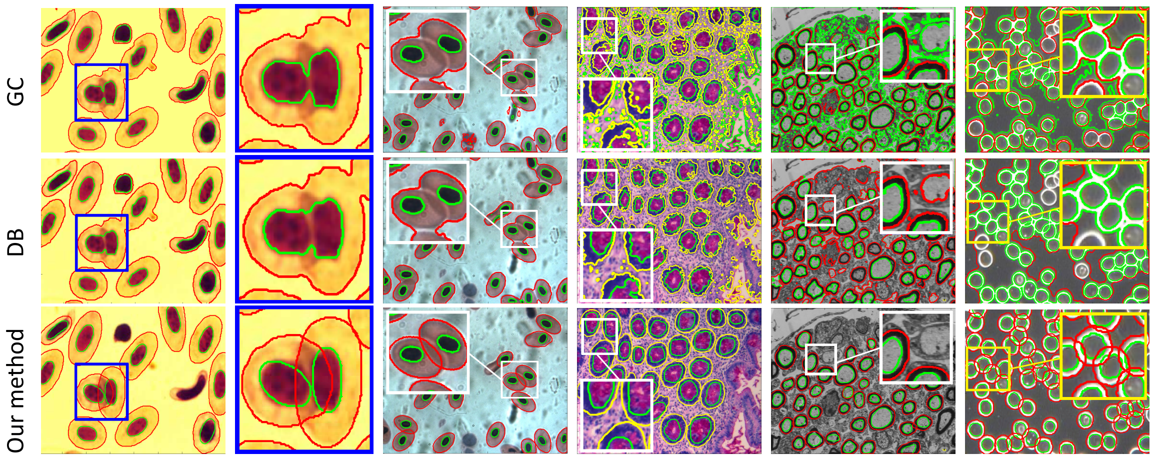


Figure 3.4: Sample results on MICR dataset. Note how the proposed method segments only the targeted cells. Same data term was used for all experiments.

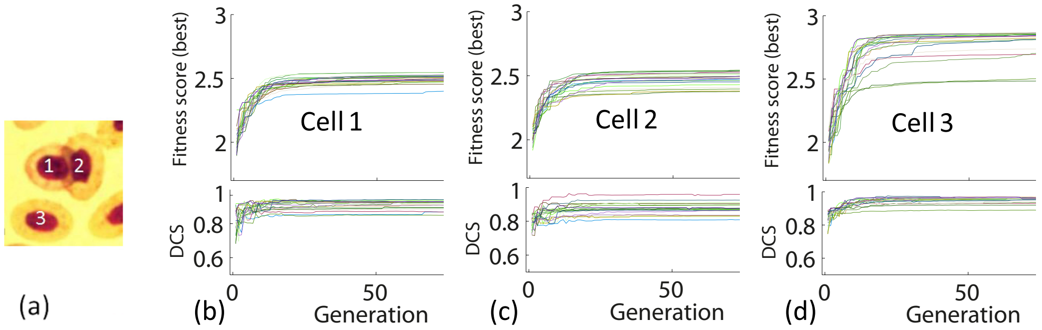


Figure 3.5: Sample segmentation shown in Figure 3.4. (b-d) Fitness and DSC of the best individual of each tribe vs. generation number for the three tribes (1^{st}, 2^{nd} and 3^{rd}) corresponding to the three cells in (a), for 20 different runs on (a).

tainment constraint in the GC framework, however, their method is unable to segment the targeted objects solely (Figure 3.4). Table 3.2 quantitatively compares our method with watershed (WS), GC and DB on both HIMA and MICR datasets using Dice similarity coefficient (DSC).

Using non-optimized MATLAB code on standard 2.3 GHz CPU, the running time of our algorithm ranged between 60-300 s/image, which depended primarily on the number of cells per image, which varied between 2-60 for both HIMA and MICR.

3.2.3 Reproducibility test

Due to the random initialization and evolution in GA, they may not always produce the same result. To examine the reproducibility of the proposed approach, in our third experiment, we ran our method 20 times on sample images and monitored the fitness and DSC vs.

Table 3.2: Accuracy comparison using DSC ($mean \pm std$)

Dataset	Watershed	Graph cuts	[Delong and Boykov, 2009]	Our method
HIMA	0.68 ± 0.12	0.72 ± 0.09	0.72 ± 0.08	0.81 ± 0.03
MICR	0.65 ± 0.13	0.69 ± 0.20	0.76 ± 0.18	0.91 ± 0.01

generation (Figure 3.5). The results confirm that our method converges to almost similar results (similar DSC) although we randomly initialized the population at each run. From Figure 3.5, the fitness values for the 1^{st} and 2^{nd} cells are lower than the 3^{rd} simply because the first two cells overlap. The variations between different runs can be reduced by increasing the size of initial population but at the cost of computational complexity.

3.3 Chapter Summary

Segmenting spatially recurring complex objects consisting of different regions with varying shapes, colors and textures remains a challenging problem in biomedical image segmentation. Another layer of complexity is added once these multi-region objects overlap. In this chapter we showed how to address this complexity holistically by incorporating several intuitive priors into an objective function without being overly concerned about its optimization. The proposed high level priors help us to segment only the targeted objects (cells) in an image. The proposed method has its own pros and cons:

- **Pros:**
  The proposed method
  - segments overlapping multi-region cells (unlike many existing works that deal with single region cells)
  - attains solutions close to the global optimum
  - does not require Euler-Lagrangian or energy gradient calculations
  - is generally parallelizable
  - allows for arbitrarily complex objective functions
- **Cons:**
  On the other hand, our method
  - does not guarantee global optimal solution
  - has many parameters to tune
  - has a slow runtime; runtime depends primarily on number of cells per image

In Chapter 4, we focus on improving the the time and space complexity in solving the similar segmentation problem.

Chapter 4

Augmentation of level set segmentation with geometric and distance constraints

As our previous method is not efficient on large images with many cells, in this work, we aim to improve the time and space complexity by moving away from GA and explicit shape representation to a PDE-based approach and implicit shape representation. We choose to use the level set framework. The level set-based methods have been widely used in computer vision over the years and proven to be useful for medical image segmentation [Toennies, 2012, Mitiche and Ayed, 2010, Paragios, 2002, Angelini et al., 2004, Li et al., 2006b] due to their several advantages such as 1) parametrization independence, 2) the ease of implementation, 3) their ability to deal with topological changes, 4) the ease of extendibility from a curve in 2D to higher dimensions (e.g. surfaces and hyper-surfaces), and 5) their ability to impose different image data and prior knowledge terms and control their contributions in segmentation tasks. However, we believe that the formulations based on level sets have not reached its full potential yet. In this contribution, we augment the level set framework with the ability to handle two intuitive geometric relationships, *containment* and *exclusion*, along with a *distance constraint* between boundaries of multi-region objects. We chose these two constraints due to their descriptive power in segmenting compound objects.

Level set's important property of automatically handling topological changes of evolving contours/surfaces enables us to segment spatially-recurring objects (e.g. multiple instances of multi-region cells in a large microscopy image) while satisfying the two aforementioned constraints. Also, using level sets embedding functions that are based on distance transforms (as is usually done) enables us to naturally enforce optional distance constraints between different regions. Our framework can enforce *attraction* forces as well as *minimum* and *maximum* distances between regions' boundaries. The downside, however, is a local optimization framework in which the final segmentation solution depends on the initialization. In fact, here, we sacrifice the optimizibility (local instead of global solution) in exchange for lower space complexity (less memory usage) and faster runtime (especially for large microscopic images) as well as no grid artifacts. Nevertheless, the result from validating our method on synthetic and several biomedical applications, mainly on multi-region cell segmentation in microscopy images and cardiac segmentation in MR images, showed the utility and advantages of this augmented level set framework (even with fully automatic or rough initialization that is distant from the desired boundaries).

Table 4.1 compares certain features of our work with popular and state-of-the-art methods that encoded geometrical and distance constraints in their segmentation framework.

The remaining of this chapter is organized as follows. We start describing our constraints and geometric interaction terms for segmenting a two-region object in Section 4.1. The extension to multi-region is described in Section 4.1.4. Section 4.1.5 details the optimization procedure. Section 4.2 presents our experimental validation and evaluation in different applications. Following a note and example results of a special case of distance constraints in Section 4.3, we conclude this chapter in Section 4.4.

4.1 Methods

We first introduce the concepts of containment and exclusion, the two intuitive geometrical constraints, we use in this work to segment an image with two-region objects.

Containment: We say region A contains region B if B is completely encapsulated by A, i.e. $A \cap B = B$. We also add an optional distance constraint to this term; allowing us to specify that B is inside A with a distance of d pixels between their boundaries, perhaps with an attraction/repulsion force between their boundaries.

Exclusion: We say region A and B are excluded from one another if they are disjoint, i.e. $A \cap B = \emptyset$. We also add an optional distance constraint to this term; allowing us to specify that A and B are disjoint with a minimum distance of d pixels between their boundaries. For clarity, we first formulate the containment and exclusion energy terms for two-region objects (surrounded by the background).

4.1.1 Containment energy

Let Ω be a bounded open subset of $\mathbb{R}^n$ where n is the image dimension (in this work $n = 2$ or 3) and $I : \Omega \to \mathbb{R}$ is a given image (scalar field)[1]. In our formulation, C_i represents the boundary (or surface) of the i^{th} region in a multi-region object. We also define $\phi_i(x) : \Omega \to \mathbb{R}$ as the signed distance function corresponding to C_i, where $\phi_i(x) > 0$ is inside and $\phi_i(x) < 0$ is outside the region i. C_i and ϕ_i are related by the zero level set

---

[1] Our method extends directly to non scalar fields, e.g. color images, vector fields, or tensor fields, by modifying the data terms, as in [Wang and Vemuri, 2004a, Chan et al., 2000]

Table 4.1: A comparison of certain features of commonly used state-of-the-art methods and our proposed method.

Method	**Containment**	**Exclusion**	**Attraction**	**Min. distance**	**Max. distance**	**No grid-bias**	**≥ 3 regions**	**Guarantees‡**
[Wu et al., 2011]	✓	✗°	✓	✓	✓	✗	✗	✓
[Zhao et al., 1996]	✗	✓	✗	✗	✗	✓	✓	✓
[Samson et al., 2000]	✗	✓	✗	✗	✗	✓	✓	✗
[Li et al., 2006a]	✓	✗	✓	✓	✓	✗	✓	✓
[Zeng et al., 1998]	✓	✗	✓	✓	✓	✓	✗	✗
[Goldenberg et al., 2002]	✓	✗	✓	✓	✓	✓	✗	✗
[Paragios, 2002]	✓	✗	✓	✓	✓	✓	✗	✗
[Vazquez-Reina et al., 2009]	✓	✗	✓	✗	✗	✓	✗	✗
[Ukwatta et al., 2012]	✓	✗	✗	✓†	✗	✓	✗	✓
[Rajchl et al., 2012]	✓	✓	✗	✗	✗	✓	✓	✓
[Delong and Boykov, 2009]	✓	✓	✓*	✓	✗	✗	✓	✓
[Ulén et al., 2013]	✓	✓	✓*	✓	✗	✗	✓	✓
[Schmidt and Boykov, 2012]	✓	✗	✓**	✓	✓	✗	✓	✗
our work	✓	✓	✓	✓	✓	✓	✓	✗

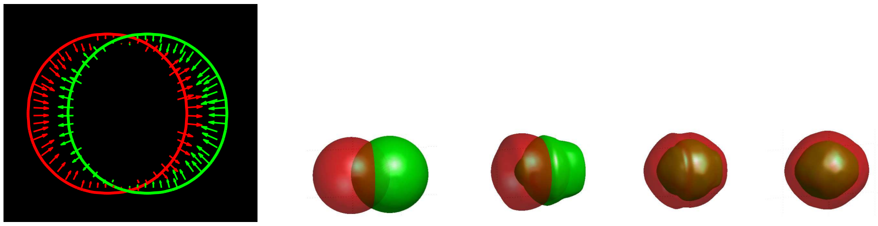
° Works for special cases only.

† The distance constraint is imposed after pre-segmenting the inner region.

‡ Exclusion constraint is a non-submodular constraint and the specified methods cannot guarantee the global solution for such constraint.

* Changing the distance constraints requires reconstructing the graph. Further, increasing the thickness increases the storage memory requirements. In contrast, in our work we can change the thickness by simply changing a parameter, i.e. d in (4.1) or $|\mathcal{C}_{ij}|$ in (4.7), and memory usage in our case is independent of thickness constraint.

** The attraction force between regions has not been discussed in [Schmidt and Boykov, 2012]. However, since they include containment similar to [Delong and Boykov, 2009], it seems that they are able to enforce attraction between regions' boundaries.



(a) (b)

Figure 4.1: Containment constraint between two red and green regions. (a) Attraction and repulsion applied to the red and green boundaries when the red region contains the green one. (b) Surface (zero level set) evolution (from left to right): red contains green with a predefined distance between them.

of ϕ_i; i.e., $C_i = \{x|\phi_i(x) = 0\}$. Using the signed distance functions enables us to efficiently control the relative distance d between the objects surfaces. Given the two regions i and j, our containment energy, such that i *contains* j with a distance of d pixels between their surfaces is

$$E_C(\phi_i, \phi_j; d) = \int_\Omega \parallel \phi_i(x) - \phi_j(x) - d \parallel^2 dx. \tag{4.1}$$

If $d = 0$, minimizing E_C encourages ϕ_i and ϕ_j to be identical. Adding a positive or negative constant d to a signed distance function ϕ, dilates and shrinks the region specified by ϕ, respectively. To better illustrate (4.1), we re-write (4.1) as:

$$\int_\Omega \parallel (\phi_i(x) - d) - \phi_j(x) \parallel^2 dx \ , \tag{4.2}$$

that penalizes the difference between $\phi_j(x)$ and the shrunk ϕ_i, i.e., $\phi_i(x) - d$. This causes the surface i to contain j while maintaining the distance of d pixels between them. E_C in (4.1) creates *attraction* and *repulsion* between i and j surfaces, by penalizing the area in which ϕ_i and ϕ_j are different.

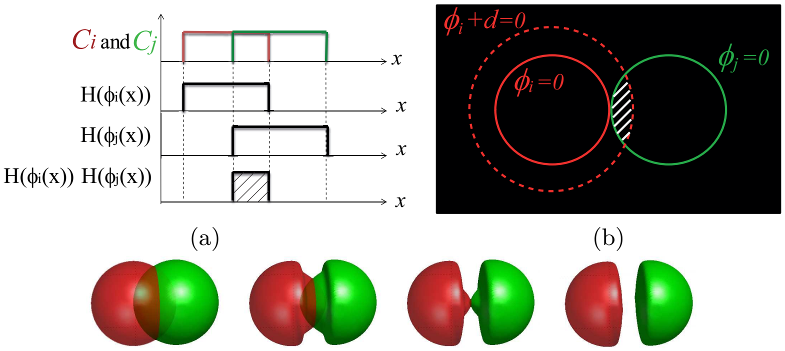
Figure 4.1(a) shows the attraction and repulsion along the boundaries of i and j when i is encouraged to contain j. Figure 4.1(b) shows how the two surfaces i and j attract each other while maintaining the predefined distance d between them by solving the Euler-Lagrange equation for (4.1):

$$\frac{\partial \phi_i}{\partial t} = -2(\phi_i - \phi_j - d), \qquad \frac{\partial \phi_j}{\partial t} = 2(\phi_i - \phi_j - d), \tag{4.3}$$

where t is an artificial time variable.

4.1.2 Exclusion energy

In many segmentation applications, there might be a need to exclude objects from one another; e.g., we might be interested in segmenting the regions that reside outside another



(a) (b)

(c)

Figure 4.2: Exclusion (i *excludes* j): (a) Shared area between regions i and j (the shaded area) is penalized by (4.4). (b) Adding a positive constant d to ϕ_i dilates region i and penalizes any region j that is within distance d pixels from i (shaded area). (c) Surface evolution (from left to right): Green and red are excluded from one another with a predefined distance between them.

region. Following [Zhao et al., 1996], we enforce an exclusion constraint on two regions by penalizing the area that the two regions share. The proposed energy term for excluding regions i and j from one another is

$$E_E(\phi_i, \phi_j) = \int_\Omega H(\phi_i(x))H(\phi_j(x))dx, \tag{4.4}$$

where $H(.)$ is the regularized Heaviside function. Figure 4.2(a) shows how the shared area between two regions is penalized (without any distance between the regions). Similar to (4.1), the distance condition between two surfaces can be easily added for the exclusion term as well. The distance d can be added to either ϕ_i or ϕ_j in (4.4). Adding a positive constant value d to ϕ_i dilates region i and ensures that any region j within the distance of d pixels from i is penalized (white shaded area in Figure 4.2(b)). Similarly if $d > 0$ is added to ϕ_j, ϕ_i will be pushed away from j by d pixels. Note that adding d to both ϕ_i and ϕ_j causes a final distance of $2d$ pixels between the two surfaces. Figure 4.2(c) shows a 3-D evolution of surfaces i and j by solving the Euler-Lagrange equation for (4.4):

$$\frac{\partial \phi_i}{\partial t} = -\delta(\phi_i)H(\phi_j), \quad \frac{\partial \phi_j}{\partial t} = -\delta(\phi_j)H(\phi_i), \tag{4.5}$$

where $\delta(z) = dH(z)/dz$ is the regularized Dirac delta function.

4.1.3 Regional and regularization terms

The above mentioned energy terms control the geometrical interactions between regions of an object and are independent of image data. Assuming piecewise constant regional intensities and adopting the total variation regularization, we employ the standard Chan-Vese formulation [Chan and Vese, 2001] to minimize the intra-region variance considering the geometric, e.g. containment, between regions. For a three-region (i, j and background) scenario in which i *contains* j, the regional term is:

$$\begin{aligned} E_I(\phi_i, \phi_j; \rho_i, \rho_j, \rho_{bg}) = \int_\Omega \Big(& \rho_j(x) H(\phi_j(x)) \\ & + \rho_i(x) H(\phi_i(x)) H(-\phi_j(x)) \\ & + \rho_{bg}(x) H(-\phi_i(x)) \Big) dx \\ & + \int_\Omega |\nabla H(\phi_i(x))| dx \\ & + \int_\Omega |\nabla H(\phi_j(x))| dx, \end{aligned} \tag{4.6}$$

where $\rho_i = |I(x) - \mu_i|^2$, μ_i is the intensity prior for region i and $H(\phi_i)$ and $H(-\phi_i)$ indicate the inside and outside region i, respectively. The last two terms in (4.6) are the regularization terms that smooth the surfaces i and j by penalizing their surface area.

So far, we introduced our constraints for segmenting an image with two-region objects and background. In the next section, we extend our framework to segment objects with more than two-regions.

4.1.4 Extension to multi-region objects

To extend our framework to multi-region objects segmentation, we introduce an $R \times R$ *constraints matrix*, $\mathcal{C}$, that encodes the containment and exclusion constraints, where R is the number of regions of an object in the image to be segmented. Our motivation to introduce $\mathcal{C}$ is to provide a framework where one can encode geometrical constraints (containment/exclusion) easily and intuitively. Table 4.2 shows how we encode containment and exclusion constraints into matrix $\mathcal{C}$.

Table 4.2: Encoding containment and exclusion into matrix $\mathcal{C}$

Constraint	**Matrix $\mathcal{C}$**
i contains j	$\mathcal{C}_{ij} > 0$
i and j are disjoint	$\mathcal{C}_{ij} < 0$ AND $\mathcal{C}_{ji} < 0$
i and j have no constraint	$\mathcal{C}_{ij} = \mathcal{C}_{ji} = 0$

(4.7)

In addition to encoding the containment and exclusion constraints, we also encode the distance constraint into $\mathcal{C}$ such that $S_{ij} = sign(\mathcal{C}_{ij})$ defines the containment or exclusion of objects i and j ($S_{ij} > 0$ means i contains j and $S_{ij} = S_{ji} < 0$ means i and j are excluded from one another), and $|\mathcal{C}_{ij}|$ indicates the distance between these two regions. Note that $sign(\mathcal{C})$ is symmetric with respect to exclusion, i.e., if i excludes j, j also excludes i.

We extend E_C in (4.1) to more than two regions by identifying all positive entries in $\mathcal{C}$ (i.e. containment) as follows:

$$E_C^{total}(\mathbf{\Phi};\mathcal{C}) = \\ \sum_{i=1}^{R} \int_\Omega \left(\prod_{j|\mathcal{C}_{ij}>0} \| \phi_i(x) - \phi_j(x) - \mathcal{C}_{ij} \|^2 \right) dx, \tag{4.8}$$

where $\mathbf{\Phi} = \{\phi_1, ...\phi_R\}$.

Exclusion energy can also be extended to multi-region objects by adding an exclusion term (as describes in (4.4)) for any pair of objects that must adhere to an exclusion constraint. Given matrix $\mathcal{C}$, the total exclusion energy is:

$$E_E^{total}(\mathbf{\Phi};\mathcal{C}) = \\ \sum_{i=1}^{R} \sum_{j>i}^{R} \frac{S_{ij}(S_{ij}-1)}{2} \int_\Omega H(\phi_i - \mathcal{C}_{ij}) H(\phi_j - \mathcal{C}_{ji}) dx. \tag{4.9}$$

where the term $\frac{S_{ij}(S_{ij}-1)}{2}$ is equal to one, only if $\mathcal{C}_{ij} = \mathcal{C}_{ji} < 0$, i.e., i excludes j, and it is zero otherwise. Since matrix $\mathcal{C}$ is symmetric with respect to exclusion, it is sufficient to incorporate only the upper triangle of $\mathcal{C}$ in (4.9).

Finally, to extend the regional energy term and to simplify the formulation, we note that regions not contained by any other region are contained by the background. So we add a dummy background row (say row 0) to matrix $\mathcal{C}$ that is positive in the columns that do not have any positive value (i.e. are not contained by other regions). Recalling that $\rho_i = |I - \mu_i|^2$ and defining $\boldsymbol{\rho} = \{\rho_0, \rho_1, ..., \rho_R\}$, we extend (4.6) to multi-region by searching each row of $\mathcal{C}$ for positive values (containments) as follows:

$$\begin{aligned} E_I^{total}(\mathbf{\Phi};\mathcal{C},\boldsymbol{\rho}) = & \sum_{i=0}^{R} \int_\Omega \left(\rho_i H(\phi_i(x)) \prod_{j|\mathcal{C}_{ij}>0} H(-\phi_j(x)) \right) dx \\ & + \sum_{i=1}^{R} \int_\Omega |\nabla H(\phi_i(x))| dx, \end{aligned} \tag{4.10}$$

where $\rho_0 = \rho_{bg}$ and $\phi_0 = 1$. The first and second terms in (4.10) are the multi-region extension of the regional and regularization terms in (4.6), respectively.

Combining (4.8-4.10), the total energy functional becomes:

$$E_{total} = \lambda_1 E_I^{total} + \lambda_2 E_C^{total} + \lambda_3 E_E^{total}, \tag{4.11}$$

where the positive constants λ_1, λ_2 and λ_3 control the contribution of each term in the segmentation.

4.1.5 Optimization

To minimize the functional in (4.11), we follow the approach of Chan and Vese [Chan and Vese, 2001] and derive the Euler-Lagrange equation. The objective function (4.11) is a weighted sum of non-negative terms and it will become zero if and only if all its terms are zero. To minimize E_C^{total} in (4.8), note that a region, represented by ϕ_ℓ, can contain other regions and can be contained by other ones (i.e., ϕ_ℓ can appear as either $\phi_i = \phi_\ell$ or $\phi_j = \phi_\ell$ in (4.8)). For region ϕ_ℓ, the Euler-Lagrange equation associated to E_C^{total} in (4.8) is calculated as:

$$\begin{aligned} F_\ell^C = & \\ & \sum_{j|\mathcal{C}_{\ell j}>0} 2(\phi_\ell - \phi_j - \mathcal{C}_{\ell j}) \prod_{k\neq j|\mathcal{C}_{\ell k}>0} \| \phi_\ell(x) - \phi_k(x) - \mathcal{C}_{\ell k} \|^2 \\ & - \sum_{i=1}^{R} S_{i\ell}(S_{i\ell}+1) \Bigg(\\ & \qquad (\phi_i - \phi_\ell - \mathcal{C}_{i\ell}) \prod_{j\neq\ell|\mathcal{C}_{ij}>0} \| \phi_i(x) - \phi_j(x) - \mathcal{C}_{ij} \|^2 \Bigg) = 0, \end{aligned} \tag{4.12}$$

where the term $\frac{S_{i\ell}(S_{i\ell}+1)}{2}$ is equal to one, only if $\mathcal{C}_{i\ell} > 0$, i.e., ℓ is contained by i, and is zero otherwise.

Deriving the Euler-Lagrange for E_E^{total} in (4.9) for the level set ϕ_ℓ results in the following equation:

$$F_\ell^E = \sum_{i=1}^{R} \frac{S_{\ell i}(1 - S_{\ell i})}{2} \delta(\phi_\ell - \mathcal{C}_{\ell i}) H(\phi_i - \mathcal{C}_{i\ell}) = 0. \tag{4.13}$$

To minimize E_I^{total} in (4.10) for a specific level set, ϕ_ℓ, it should be noted that ϕ_ℓ can appear in the product term in (4.10) (like ϕ_j in (4.10)) when $\mathcal{C}_{i\ell} > 0$, i.e., $\frac{S_{i\ell}(S_{i\ell}+1)}{2} = 1$. For the level set ϕ_ℓ, the Euler-Lagrange equation associated to (4.10) is:

$$
\begin{aligned}
F_\ell^I = {} & \rho_\ell \delta(\phi_\ell) \prod_{j|\mathcal{C}_{\ell j}>0} H(-\phi_j) \\
& - \sum_{i=0}^{R} \frac{S_{i\ell}(S_{i\ell}+1)}{2} \left(\rho_i H(\phi_i) \delta(\phi_\ell) \prod_{j \neq \ell, \mathcal{C}_{ij}>0} H(-\phi_j) \right) \\
& - div \left(\frac{\nabla \phi_\ell}{|\nabla \phi_\ell|} \right) = 0. \qquad (4.14)
\end{aligned}
$$

Having (4.12), (4.13) and (4.14), the update equation of region ϕ_ℓ is calculated as:

$$\frac{\partial \phi_\ell}{\partial t} = -(\lambda_1 F_\ell^I + \lambda_2 F_\ell^C + \lambda_3 F_\ell^E). \qquad (4.15)$$

Finally, in implementing the proposed method the level sets ϕ should remain signed distance maps during the optimization procedure, otherwise the distance priors do not work properly. To do so, the level sets functions are re-initialized to signed distance functions after every few (~ 5) iterations.

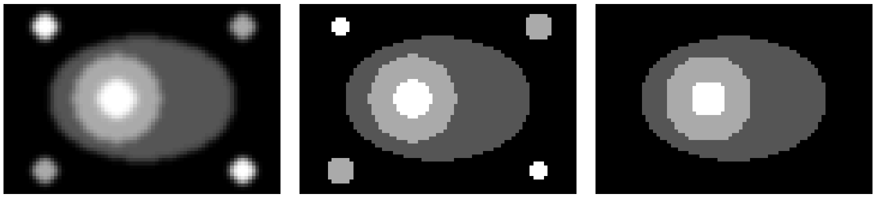
To ensure that our level set-based framework is numerically stable, we place an upper bound for the time-step ∂t, using the CFL condition [Courant et al., 1967]. The stability condition is:

$$F_{max} \partial t \leq min(h_x, h_y, h_z), \qquad (4.16)$$

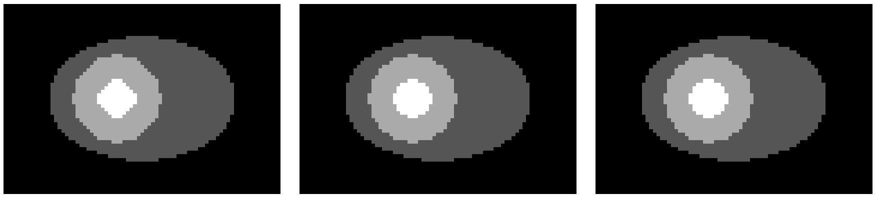
where h_x, h_y and h_z are the grid spacing in the x, y and z direction, respectively, and F_{max} is the maximum absolute force (also known as speed function) applied to the level set at each iteration and is calculated from (4.12),(4.13),(4.14). For $h_x = h_y = h_z = 1$, at each iteration we make sure that $\partial t \leq 1/F_{max}$. Violating the CFL condition results in instabilities.

4.2 Experiments

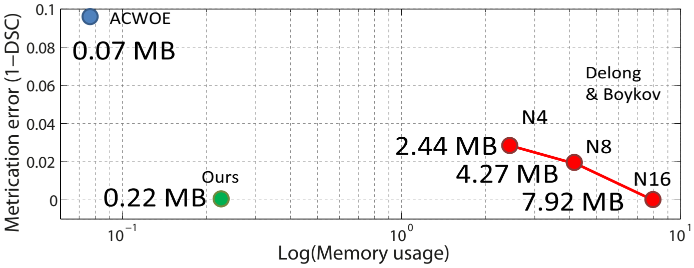
In this section, we present several experiments and show the applicability and utility of our framework on different biomedical applications. However, our comprehensive quantitative validation mainly focuses on two important medical applications: (a) histology and microscopy image segmentation and (b) left and right cardiac ventricles segmentation. We also compare our method to the analogous discrete works of Delong and Boykov (DB) [Delong and Boykov, 2009] and Ulén et al. (USK) [Ulén et al., 2013] on both synthetic and real data and analyze the metrication error, running time and memory usage.



(a) Original image (b) ACWOE (c) GC: 4- connected



(d) GC: 8- connected (e) GC: 16- connected (f) Our method



(g) Metrication error vs. memory usage

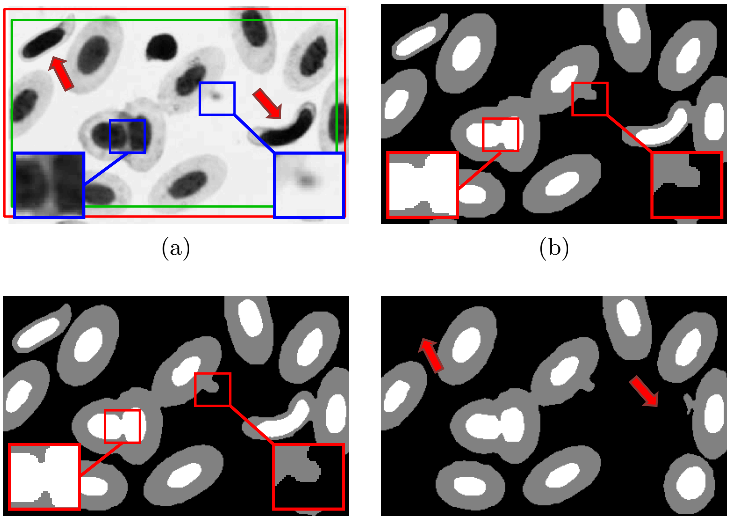
Figure 4.3: Synthetic three-region object segmentation. (b) ACWOE's result. (c-e) DB graph cuts based method [Delong and Boykov, 2009] with different connectivities. (f) Our segmentation results. (g) Metrication error vs. memory usage: *red curve*: GC-based method; *blue circle*: ACWOE; *green circle*: our method.

4.2.1 Synthetic data

In our first experiment, we compare our method with DB in terms of metrication error and memory usage on a simple synthetic example.

Metrication error: We compared our method with its counterpart graph-based method [Delong and Boykov, 2009] in terms of metrication artifact. In Figure 4.3(a), we are interested in segmenting the 3-region elliptic object. Figure 4.3(b) shows a segmentation without imposing any geometrical constraint (here we used multiphase ACWOE [Vese and Chan, 2002]). DB's results for 4, 8 and 16 graph connectivity are shown in Figures 4.3(c-e). Note the metrication artifacts in (c) and (d).

In this experiment, we quantify the metrication error by *metrication error* $= (\text{DSC}_c - \text{DSC}_d)$, where DSC_c and DSC_d are Dice similarity coefficient for continuous and discrete methods, respectively. The DSC measures the segmented regions overlap and is given by $2|A \cap B|/(|A| + |B|)$, where A and B are the ground truth and the segmentation result regions, respectively. Since we used the same data term for DB and our method, and since



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

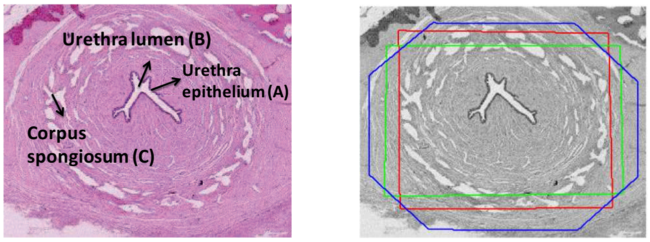
Figure 4.4: Cell segmentation in a microscopy image. (a) Original image, 250 × 395 pixels. Arrows show abnormal cells. (b) Result of DB, 33.90 MB. (c) Our result (thickness= 2 pixels, 1.50 MB). (d) Our result when segmenting only normal looking (elliptical) cells (thickness=10 pixels, 1.50 MB). Note that DB needs 313.41 MB extra memory (347.31 MB in total) to impose a thickness constraint of 10 pixels while the memory usage of our method is independent of thickness constraint.

the DSC of our continuous method is one for this synthetic data ($\mathrm{DSC}_c = 1$), the only source of error in DB's work is due to the gridding bias. The metrication error in graph based methods can be reduced by increasing the graph connectivity at the expense of increasing the memory usage.

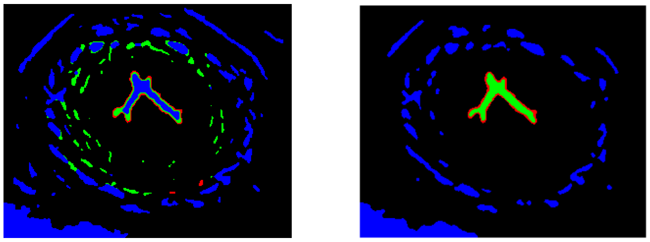
Memory usage: The memory consumption of our method and the graph-based methods are compared in Figure 4.3(g). The red curve in Figure 4.3(g) illustrates the metrication error vs. memory usage of DB for 4, 8 and 16 connectivity, while the blue and green circles represent ACWOE and our method, respectively. To remove the metrication error, DB needed 16 connectivity, Figure 4.3(e), requiring 36 times more memory than our method (7.92 MB vs. 0.22 MB). We emphasize that DB needs more memory largely due to exploring the whole search space to find the global solution. The larger the size of the image (e.g. higher resolution 2D microscopy images that can be in the order of 100s of mega-pixels or 3D volumes), the more important it is to pay attention to this increased memory consumption. In this work, we sacrificed the optimality (i.e. via our local optimization with level sets) for memory efficiency.

$$\begin{array}{c|ccc} & A & B & C \\ \hline A & 0 & +d & -d \\ B & 0 & 0 & 0 \\ C & -d & 0 & 0 \end{array}$$

(a) $\mathcal{C}$ matrix



(b) Original image (c) Initialization



(d) ACWOE (e) Our method

Figure 4.5: Urethra segmentation in a histology image. The constraint matrix is set such that the urethra epithelium (A) contains the urethra lumen (B) and excludes the other regions with similar intensity with B, i.e. the corpus spongiosum (C). Here A, B and C are represented by red, green and blue colors, respectively.

4.2.2 2D histological and microscopic images

Histology and microscopy image analysis is becoming increasingly crucial for studying diseases such as cancer and for obtaining reference diagnoses. Two important features that can be seen frequently in many histology and microscopy images are (a) the existence of multiple objects of the same class (e.g. cells) in a single image, and (b) the geometric interactions between the objects' regions. The latter includes containment and exclusion between the objects and different parts of the objects (e.g. a cell and its sub-cellular components, such as nucleus and nucleolus).

Figure 4.4 compares our method with DB in fish blood cell segmentation. Figures 4.4(b) and (c) show DB's and our results, respectively. Note the gridding artifacts in DB's work (red boxes). In both ours and DB's work, we can **adjust the distance (thickness)** between different regions to control which objects are segmented. To exclude the abnormal cells (red arrows in Figure 4.4(a)) from the segmentation, we increase the imposed thickness constraint between two regions from 2 pixels to 10 pixels, Figure 4.4(d). In DB, increasing the thickness requires graph reconstruction with additional edges. The new graph needs an extra ∼313.41 MB memory, an almost 10-fold increase. In contrast, we emphasize that in our method, thickness can be increased by simply changing the value of d in (4.1) or equivalently $|\mathcal{C}_{ij}|$ in (4.7) and it does not affect the memory usage.

Figure 4.5(a) shows how the containment and exclusion constraints are encoded into matrix $\mathcal{C}$ to segment and distinguish the urethra from other regions with similar intensity/color in a histological image of the ureter (Figure 4.5(b)). Note how our method is able to distinguish the urethra from other similar regions by forcing it to be contained in the urethral epithelium (Figure 4.5(e)), while the conventional active contours without edges, or any other method without containment and exclusion constraint, is incapable of distinguishing the urethra from other similar regions based solely on image pixel values (Figure 4.5(d)).

We tested our algorithm on 20 histology and microscopy images and used the DSC to evaluate the performance of our method. Figure 4.6 presents qualitative results on histology and microscopy data. The first and the second rows in Figure 4.6 show the original image and initialization overlaid on the original image. The third row shows multi-phase ACWOE results. The fourth and fifth rows show USK results with 4-connected and 8-connected graphs, respectively, and the bottom row shows the proposed method's results. These results illustrate the importance of geometrical constraints in histology/microscopy image segmentation and also show the effects of metrication error (4^{th} and 5^{th} rows). We note that we tried to get the best results as we could for USK's method by exhaustively searching for the best regularization weight and thickness (minimum distance) constraint. As is seen in Figure 4.6, the metrication error in the 8-connected graph (5^{th} row) is improved compared to the 4-connected graph (4^{th} row) but not completely resolved. Penalizing boundaries of objects only across axis aligned edges in graph-based methods makes it difficult for a convex regularizer like TV to be implemented in the discrete domain efficiently. Boykov and Kolmogorov [Boykov and Kolmogorov, 2003] proposed a method to roughly overcome this issue. However, their method requires extra memory and computational time due to adding extra edges to the graph. On the other hand, the continuous frameworks, including level set framework, can efficiently encode general convex regularizers like TV.

We emphasize that in this experiment there was no need to initialize too close to the solution and we performed fully automatic initialization for almost all of the cases, Figure 4.6(a-e, g). Indeed, using the regions close to the image boundary was sufficient for initialization. However, to show the effect of the initialization on the results, we ran our algorithm starting from three different initializations (Figure 4.7). Note that the obtained results from the first two initializations, Figure 4.7(a,b), are almost identical, despite significant difference in their initialization. Yet the third result is affected by the initialization, Figure 4.7(c).

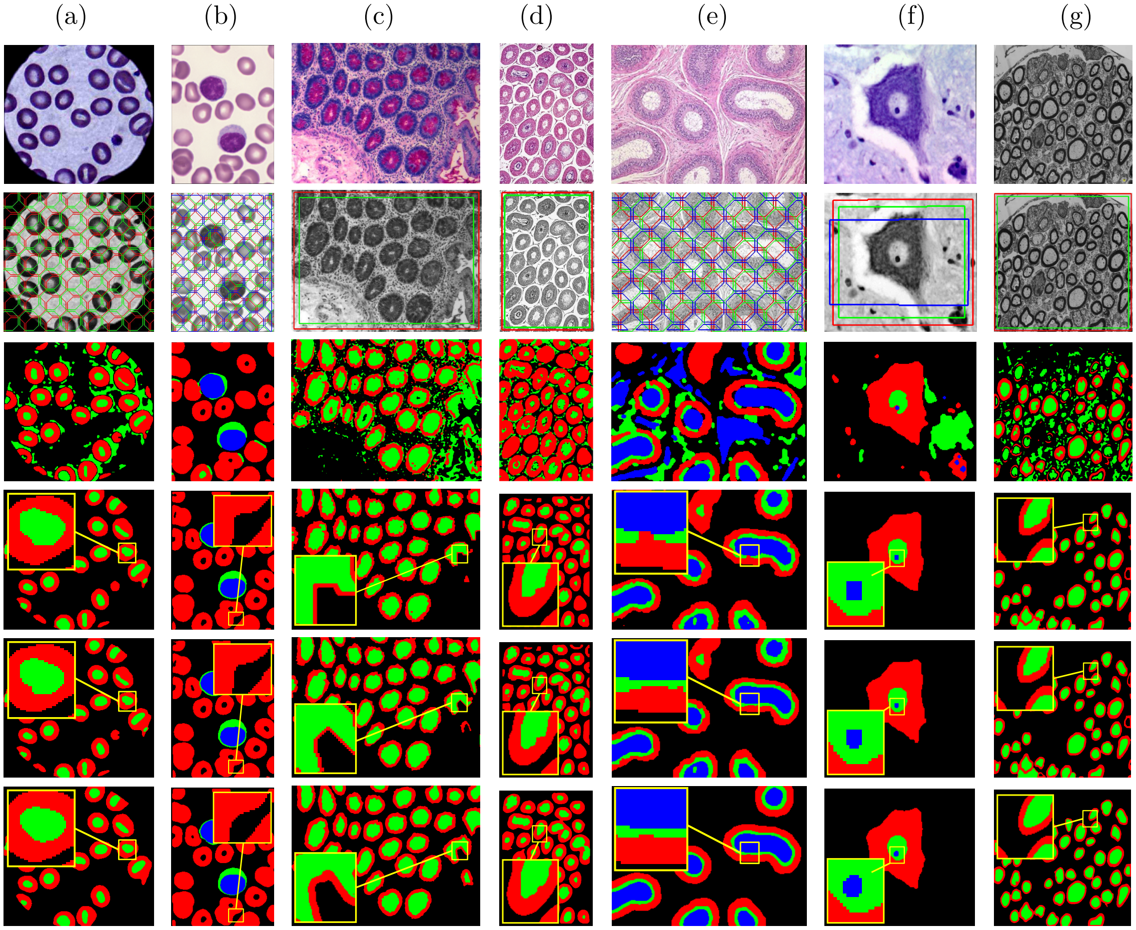


Figure 4.6: Segmentation of histology and microscopy images. 1^{st} row: Original image. 2^{nd} row: Initialization. 3^{rd} row: ACWOE results. 4^{th} row: USK's method with 4-connectivity. 5^{th} row: USK's method with 8-connectivity. 6^{th} row: Our results. The images from left to right are: (a,b) Microscopy images of blood cells, (c,d,f) histological cross sections of testes histology and (e,g) neuron histology. Note that in case of no exclusion constraint USK is equivalent to DB.

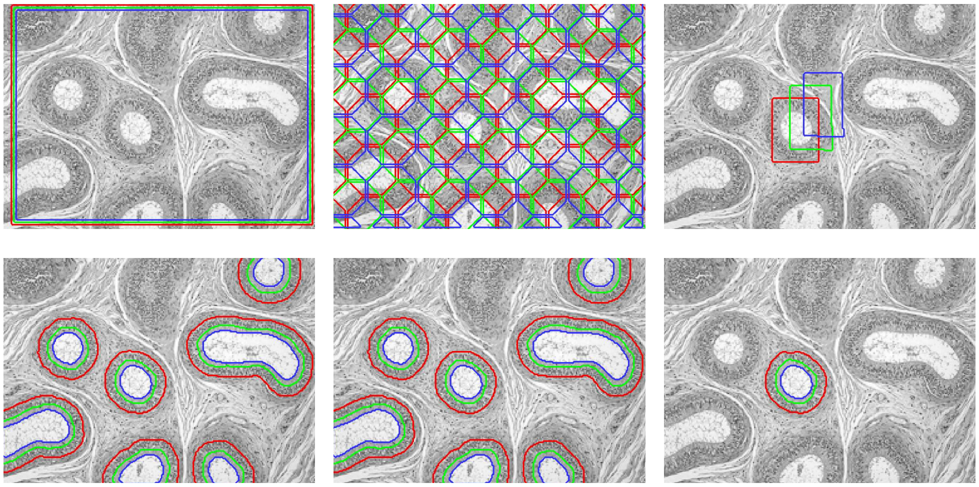

(a) (b) (c)

Figure 4.7: Effect of initialization on the results. Top row: initializations. Bottom row: results after convergence. Despite significant different initializations in (a) and (b), the obtained results are almost identical. Yet the third result (c) is clearly affected by the initialization.

Table 4.3: DSC and memory usage comparison

	ACWOE [Vese and Chan, 2002]	USK [Ulén et al., 2013]	Our method
Regions that have not been contained by other regions	0.88 ± 0.05	4-C: 0.89 ± 0.04 8-C: 0.90 ± 0.05	0.91 ± 0.02
Regions that contained by (or excluded from) others	0.54 ± 0.14	4-C: 0.89 ± 0.05 8-C: 0.90 ± 0.04	0.90 ± 0.04
Overall	0.68 ± 0.07	4-C: 0.89 ± 0.04 8-C: 0.90 ± 0.05	0.91 ± 0.03
	Memory usage (MB)	4-C: 120 ± 99.30 8-C: 167 ± 101.60	1.86 ± 1.37

We evaluated our segmentation method by measuring the overlap between segmented and ground truth regions using DSC. We report the DSC for contained regions and not-contained regions separately in Table 4.3. As expected, our results for contained regions improved dramatically over the conventional ACWOE as the latter only considers the image intensity/color for its external energy term. In many histology and microscopy images, regions of interest and the background might have similar intensities making ACWOE insufficient for this task. For the same reason, other methods that do not enforce such constraints are not able to segment the contained regions properly. For the regions that have not been contained by other regions, both methods have a similar accuracy. However, our method indirectly improves the performance for these regions as well due to the attraction and repulsion created by the containment energy terms. The average memory usage to segment this microscopic histology dataset is 120 MB and 1.86 MB for USK (4-connected graph) and our method, respectively.

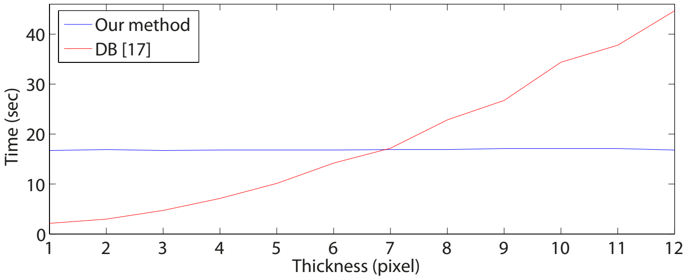


Figure 4.8: CPU runtime versus thickness/distance constraint.

Comparing the computation time of our method with a global graph based method, e.g. DB, is critical as we use local optimization and convergence time depends on how close we initialize the contours. In addition, the computation time for graph based methods, e.g. DB, highly depends on the thickness/distance constraint and largely varies from one image to another. To have a fair run time comparison, we created a 500 × 500 synthetic image consisting of several two-region nested objects with different distance between their regions. To initialize our method, we place the initial contours at the border of the image, e.g. similar to Figure 4.6(d,g). Figure 4.8 compares the runtime between our method and DB for different thickness constraints. To impose a distance constraint of T pixels between two regions, DB and USK need to add $O(T^2)$ extra edges *per pixel*. Therefore, these graph based methods are highly efficient in segmenting images with reasonable size and thickness constraint. On the other hand, for large distance constraints DB and USK are not that efficient (considering that they still provide us with the global solution) while in our framework the runtime is almost constant with respect to different distance constraints.

4.2.3 Cardiac ventricles segmentation

We also evaluated our framework on left and right cardiac ventricles segmentation. To this end, we used two different 3D datasets: 1) The Sunnybrook Health Science Centre dataset for left ventricle segmentation used in the MICCAI 2009 challenge [Centre, 2009], and 2) The Rouen University Hospital data for right ventricle segmentation used in the MICCAI 2012 challenge [Hospital, 2012].

Left ventricle segmentation: LV segmentation is an important step for the diagnosis of cardiovascular diseases. Accurate calculation of key clinical parameters such as ejection fraction, myocardium mass, and stroke volume depends on accurate segmentation of endocardial and epicardial boundaries of the left ventricle. We used our framework to segment the left ventricle. To model the LV, we encode the constraint "myocardium contains the left ventricle" into our framework.

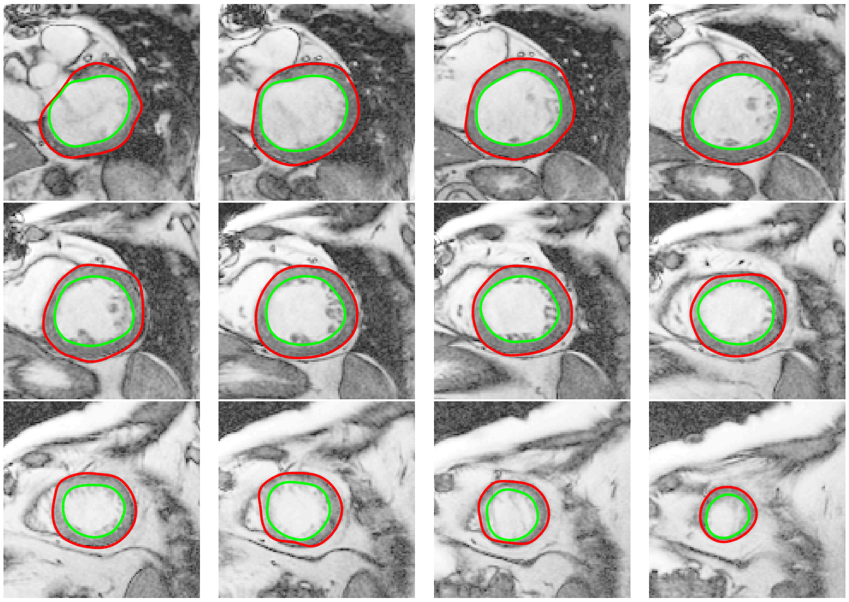


Figure 4.9: A representative sample of LV segmentation using our method for one subject at different slice levels: Basal (first row), mid-cavity (second row) and apical (third row).

We evaluated our method on the Sunnybrook Health Science Centre dataset. This dataset consists of 30 short-axis cardiac cine-MR images (15 volumes for training and 15 volumes for validation) obtained by a 1.5T GE Signa MRI. All the images were obtained during 10-15 second one breath-hold with a temporal resolution of 20 cardiac phases over the heart cycle[2].

The ground truth of endocardial and epicardial contours have been provided by an experienced cardiologist in all slices at end-diastole (ED) and end-systole (ES) phases.

Figure 4.9 shows the result for one subject over different slice levels: basal, mid-cavity and apical. The red and green curves indicate the epi- and endocardial boundaries, respectively. The distance between the epi- and endocardial boundaries (myocardium thickness) is not fixed but decreases from basal to apical level. Hence, we cannot choose a fixed distance prior d for LV segmentation (c.f. (4.1)). Instead, we allow d to vary linearly from the first slice, at the basal level, to the last slice, at the apical level from 9 mm to 4 mm. Figure 4.9 shows that although the pixel intensity of the papillary muscles (the darker regions inside the green contours) and the myocardium is similar, the proposed method is able to exclude the papillary muscles from the myocardium. This is because of the attraction between the two epi- and endocardium contours enforced by the containment energy term. Figure 4.10 displays the result of ACWOE as well as the effect of the containment term on LV

---

[2] Due to different breath-holds between slice acquisition there could be misalignments between different short axis slices. Ideally, these slices must be properly aligned in a pre-processing stage prior to segmentation. However, to use the provided ground truth and for fair comparison with other methods, we use the cardiac data as provided with simple noise reduction and smoothing.

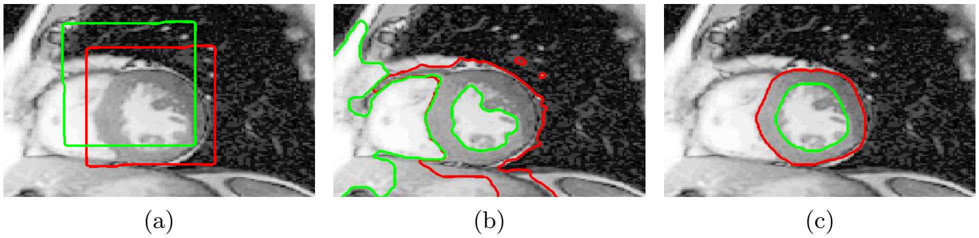


Figure 4.10: Effect of the containment term on LV segmentation. (a) Endocardium (green) and epicardium (red) initialization, (b) ACWOE result, (c) Our result with containment energy term. Note how the containment constraint improves the LV segmentation by creating attraction and repulsion on epi- and endocardial boundaries.

segmentation. Without the containment constraint, the level set cannot segment the left ventricle properly, Figure 4.10(b), while this issue has been addressed in Figure 4.10(c) via the attraction between the two red and green contours.

Since there is no applicable exclusion constraint for LV segmentation, we set $\lambda_3 = 0$ in (4.11). Hence, we have two free parameters to set for LV segmentation: λ_1 and λ_2 control the contribution of the regional intensity term and the containment term. From an energy minimization point of view, one of the parameters can be fixed and we have only one free parameter to set. We set $\lambda_1 = 1$ and varied λ_2 from 0 to 10 to find the best value in the training set provided in the Sunnybrook dataset. The obtained optimum value for λ_2 is 3.8.

For initialization, we provided the initial contours (similar to Figure 4.10(a)) in the mid-axial cardiac slice of the 3D scan of each subject. The level sets then evolve in 3D.

We quantitatively evaluate our segmentation method based on two measures:

1) **The average distance error:** measures the perpendicular distance between the resultant contour and the corresponding manually drawn expert contour, averaged over all contour points. The smaller the average distance value implies that the two contours match more closely.

2) **The DSC** described earlier.

Tables 4.4 and 4.5 compare the proposed method's performance with other competing methods that were evaluated on the same dataset. Figure 4.11 visualizes the average distance error (in mm) obtained from the proposed method for all cases in the Sunnybrook validation dataset. One of the important clinical parameters for cardiac diagnosis is the left ventricular volume. The LV volume determined by the proposed method and by manual expert segmentation have been compared in Figures 4.12(a) and (b) for 16 different subjects (volumes) over the two ES and ED phases of the cardiac cycle.

Figure 4.12(b), illustrates the Bland-Altman plot [Martin Bland and Altman, 1986], which is used to compare two clinical measurements and shows the difference between the

Table 4.4: LV segmentation results (Sunnybrook dataset): DSC

Method	LV endo.	LV epi.
[Marak et al., 2009]	0.86 ± 0.04	0.93 ± 0.02
[Casta et al., 2009]	? ± ?	0.93 ± ?
[Lu et al., 2009]	0.89 ± 0.03	0.94 ± 0.02
[Wijnhout et al., 2009]	0.89 ± 0.03	0.93 ± 0.01
[OŠBrien et al., 2009]	0.81 ± ?	0.91 ± ?
[Constantinides et al., 2009]	0.89 ± 0.04	0.92 ± 0.02
[Huang et al., 2009]	0.89 ± 0.04	0.94 ± 0.01
[Jolly, 2009]	0.88 ± 0.04	0.93 ± 0.02
[Ulén et al., 2013]	0.86 ± 0.05	0.92 ± 0.02
Our method	0.90 ± 0.03	0.94 ± 0.01

"?": Not reported in the corresponding paper

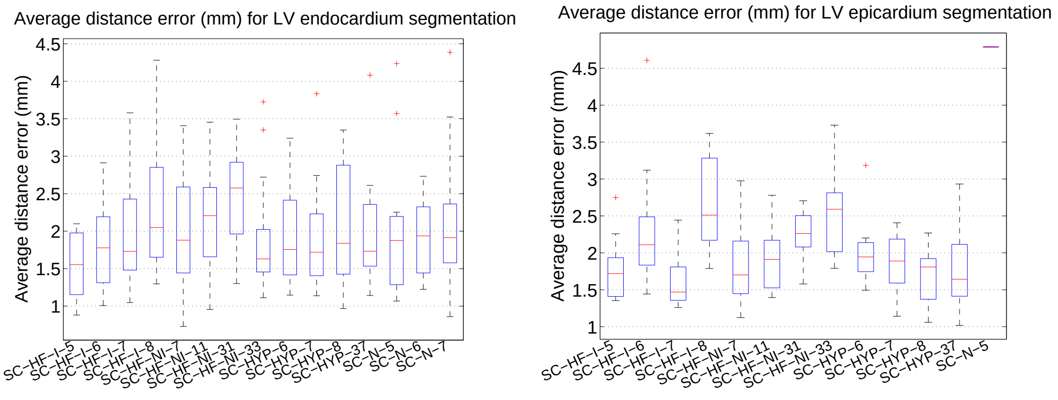
Table 4.5: LV segmentation results (Sunnybrook dataset): Average distance error

Method	LV endo.	LV epi.
[Marak et al., 2009]	2.60 ± 0.38	3.00 ± 0.59
[Casta et al., 2009]	? ± ?	2.72 ± ?
[Lu et al., 2009]	1.91 ± 0.63	2.07 ± 0.61
[Wijnhout et al., 2009]	2.28 ± ?	2.29 ± ?
[OŠBrien et al., 2009]	3.16 ± ?	3.73 ± ?
[Constantinides et al., 2009]	2.35 ± 0.57	2.04 ± 0.47
[Huang et al., 2009]	2.11 ± 0.41	2.06 ± 0.39
[Jolly, 2009]	1.97 ± 0.48	2.26 ± 0.59
[Ulén et al., 2013]	? ± ?	? ±?
Our method	1.89 ± 0.29	1.98 ± 0.33

"?": Not reported in the corresponding paper

two measurements versus their average value. The Bland-Altman plot is useful for detecting any systematic bias between the two measurements and identifying possible outliers. The *limits of agreement* in a Bland-Altman analysis is usually specified as $mean(difference) \pm 1.96 \times std(difference)$, where $mean(.)$ and $std(.)$ are the average and the standard deviation of the data, respectively. If the difference is within $mean \pm 1.96 std$ then it is deemed not clinically important, i.e. the two methods (our proposed method and the expert segmentation) can be used interchangeably. The average time for segmenting one phase (ES or ED) in a single volume on a 3.4 GHz Intel(R) CPU with 16 GB RAM is about 65 seconds.

Left and right ventricles segmentation: Studies show that the right ventricular (RV) function may be effective for diagnosing cardiovascular diseases such as pulmonary hypertension, congenital heart disease, and coronary artery disease [Noseworthy et al., 2008]. Myocardial left and right ventricular segmentation is a suitable application for our framework since it exhibits both containment and exclusion geometrical constraints. In the cardiac model, the myocardium contains both the left and right ventricles while left and right ventricles are excluded from one another (Figure 4.13(a)). We use our framework to



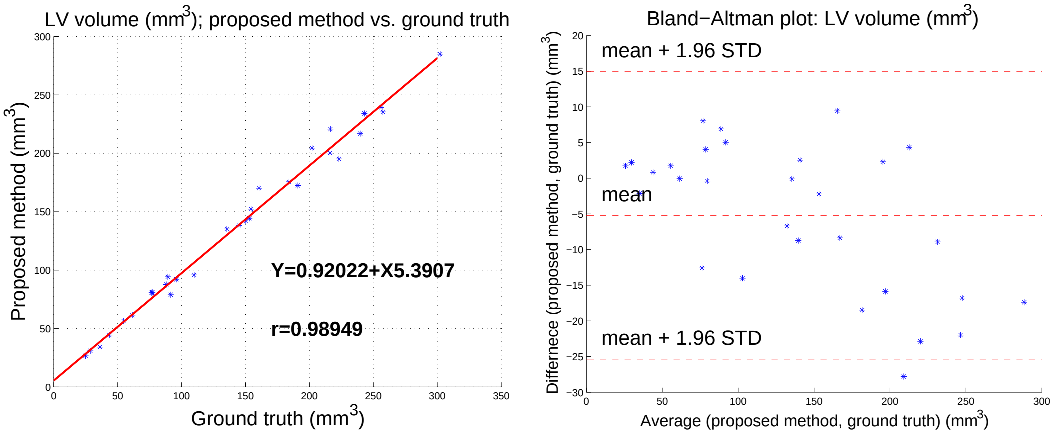
(a) (b)

Figure 4.11: Box plots showing the average distance error (mm) between the obtained results and the ground truth for LV, (a) endocardium, and (b) epicardium segmentation of the Sunnybrook dataset. Results are shows for 16 different cases (along the x-axis). The top and bottom line of each box indicate the first and third quartiles of the measurements, respectively. The red line in the middle of each box shows the median. The whiskers from each box show the largest and smallest observation and the "+" symbol show the outliers.

encode these geometrical constraints and segment the whole heart as an object consisting of multiple parts.

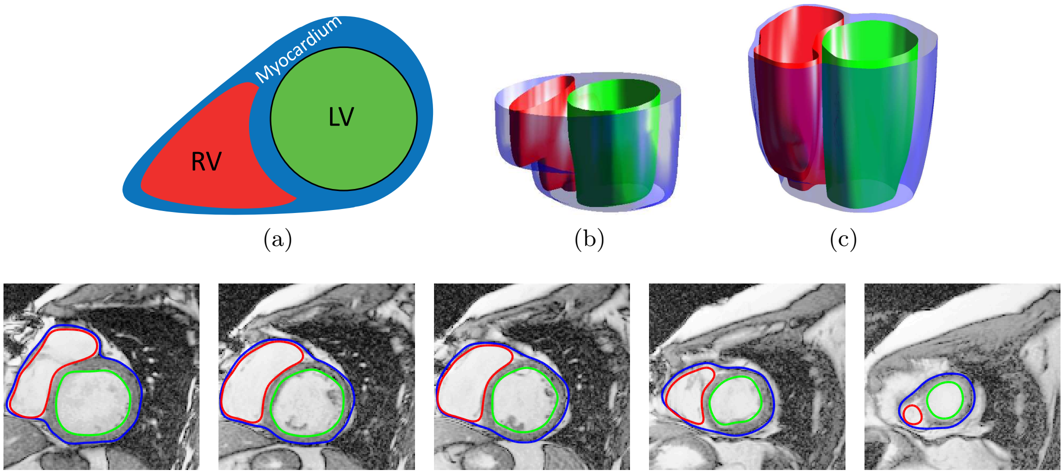
We evaluated our method on both the Sunnybrook and the Rouen datasets. The Rouen dataset consists of 16 short-axis cardiac MR volumes obtained using a 1.5T MRI. All the images were obtained during 10-15 second one breath-hold with a temporal resolution of 20 cardiac phases over the heart cycle. The ground truth of the right ventricle endocardial and epicardial contours have been provided by Rouen University Hospital. In both these datasets, we segmented the myocardium and the left and right ventricles simultaneously. Since the Sunnybrook and the Rouen datasets provide the ground truth segmentation only for LV and RV, respectively, we only report the results for the parts for which ground truth has been provided, i.e. the LV for the Sunnybrook and RV for the Rouen dataset.

We use three simple elliptic cylinders as initialization surfaces for LV, RV and myocardium. The centres and radii of these elliptic cylinders are different for three basal, mid-cavity and apical slice levels (Figure 4.13(b)). In this experiment we have three distances between the endo- and epicardium in the left ventricle (d_L), the endo- and epicardium in the right ventricle (d_R) and, the distance between the left and right ventricles (d_{LR}). Similar to LV segmentation in Section 4.2.3, we allow d_L to vary from 9 mm to 4 mm from the basal to the apical level. In this experiment we set $d_R = 4$ mm and $d_{LR} = 6$ mm. To have a suitable estimation of the weights λ_1, λ_2 and λ_3 in (4.11), we tune these weights using the leave-one-out cross validation technique over the dataset. Figure 4.13(d)-(h) illustrates qualitative cardiac segmentation results for the case SC-HF-I-5 in the Sunnybrook dataset



(a) (b)

Figure 4.12: Scatter plots of the proposed method results against the ground truth for LV volume in mm^3. (a) Regression analysis for LV volume measurement. (b) Bland-Altman plot comparing the proposed method and the ground truth on LV volume measurement.

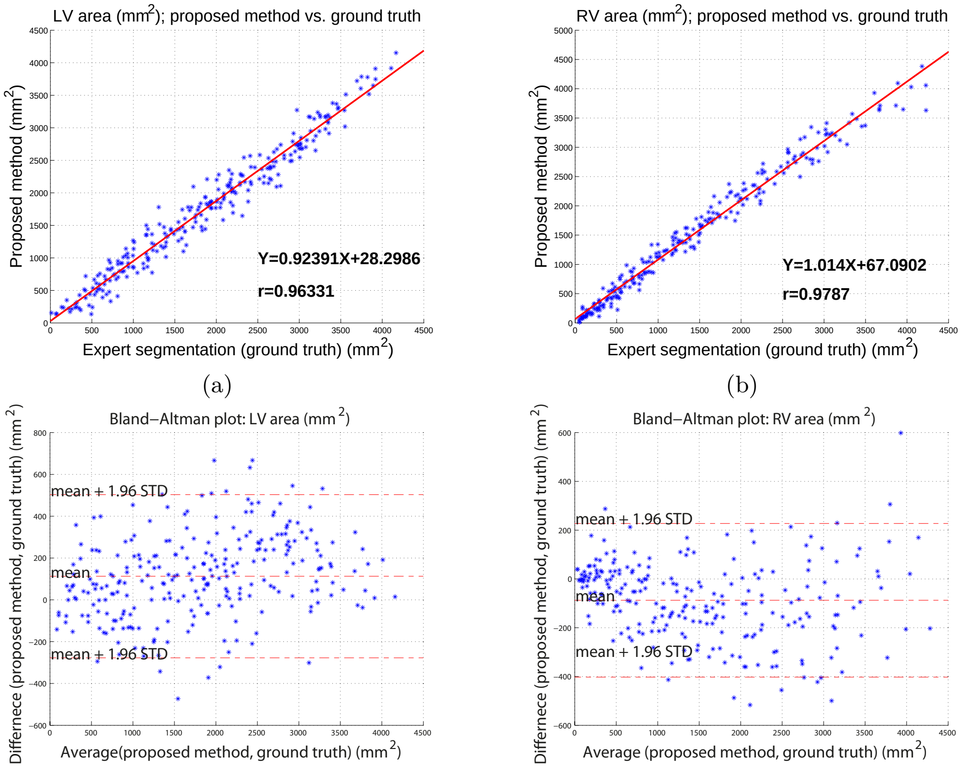


(a) (b) (c)

(d) (e) (f) (g) (h)

Figure 4.13: Myocardium and left and right ventricles simultaneously segmented using the proposed geometrical constraints. (a) Cardiac model. Myocardium contains left and right ventricles, while left and right ventricles are excluded from one another. (b) 3D model used for initialization. (c) 3D rendering of the segmentation of case SC-HF-I-5 from Sunnybrook dataset. (d)-(h) 2D cross sections of the segmentation result of (c).

and Figure 4.13(c) shows its corresponding 3D rendering. Due to the regularization term in our level-set formulation, the RV insertion points might be over smoothed. This issue can be addressed by post-processing or by imposing a spatially-varying regularization into the level set framework. Figures 4.14(a) and (b) show the linear regression analysis for LV



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

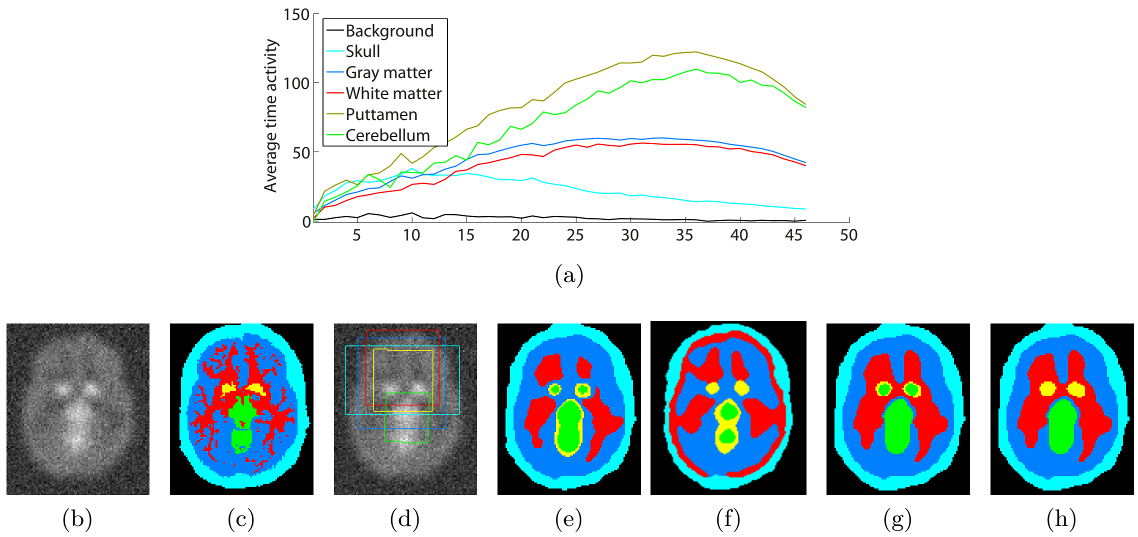
Figure 4.14: Scatter plots of the proposed method against the manual segmentation (ground truth) for LV and RV area measurement in mm^2. (a,b) Regression and correlation analysis of the area of (a) LV from the Sunnybrook dataset, and (b) RV from Rouen training dataset in two ED and ES cardiac phases. (c,d) Bland-Altman plot for (c) LV area and (d) RV area measurement.

Table 4.6: Simultaneous LV and RV endocardium segmentation results

Region	DSC	Average distance error (mm)
Left ventricle	0.89 ± 0.03	2.15 ± 0.41
Right ventricle	0.87 ± 0.02	1.79 ± 0.41

and RV area (mm^2) respectively, obtained by the proposed method and the ground truth for each single slice of ED and ES phases in the Sunnybrook and Rouen volumes. The correlation value between the proposed method and the ground truth is 0.963 and 0.978 for LV and RV segmentation, respectively. Also, the Bland-Altman plot for LV and RV segmentation are shown in Figures 4.14(c) and (d).

Note that while in our segmentation approach we segment the full myocardium (Figure 4.13), we only compare the LV and RV endocardium with the ground truth and not the epicardium, since the two Sunnybrook and Rouen datasets have provided ground truth



(a)

(b) (c) (d) (e) (f) (g) (h)

Figure 4.15: Brain dPET segmentation. (a) Average TAC for each functional region in the ground truth. (b) Raw image (last frame of the dynamic sequence, which is typically visualized by clinicians). (c) Ground truth. (d) Initialization. (e) Multi-phase ACWOE (no constraints). (f) Saad et al. [Saad et al., 2008]. (g) Our result with containment constraint but without enforcing any exclusion constraint. (h) Our result with containment and exclusion constraint. Note how the putamen is contained by the white matter (red) as it should be whereas (e) and (f) are anatomically incorrect. Also, note how the putamen (yellow) and cerebellum (green) are properly detached in (h) as opposed to (e-g).

for only the LV and RV epicardium, respectively. Table 4.6 reports the DSC and average distance error for simultaneous LV and RV segmentation.

4.2.4 Brain dynamic-PET segmentation

To test our framework on more complex application, we applied our method to dPET images, where, at each pixel in the image, a time activity curve describes the metabolic activity of a tissue as a result of tracer uptake, Figure 4.15(a). Figure 4.15 shows an example of segmenting a dPET image, $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^{40}$. Note the low signal-to-noise ratio (SNR) of the dPET image (Figure 4.15(b)), which is the result of not having enough time to collect a large number of photons within the short time intervals needed to capture the tracer kinematics. Our spatial relationships include: 1) the skull contains gray matter; 2) gray matter contains white matter; 3) white matter contains putamen; 4) putamen and cerebellum must be excluded from one another. From Figure 4.15, the problems of putamen surrounding cerebellum (yellow around green), mentioned in [Saad et al., 2008], are now clearly solved (Figure 4.15(h)). Despite the low signal to noise ratio in the dPET image and with a not great initialization (Figure 4.15(d)), our method's ability to incorporate geometric constraints results in an improved and anatomically plausible segmentation compared to

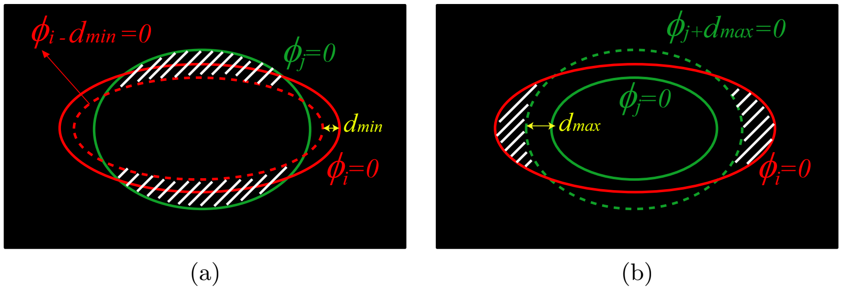


Figure 4.16: Enforcing (a) minimum and (b) maximum distance between regions i and j using (4.18) while i contains j. The shaded area shows the region that is penalized by (4.18).

the results reported in (e) and (f). We also compared our method with and without the exclusion constraint. As shown in Figure 4.15(g), without the exclusion constraint, we still can get the incorrect result of putamen surrounding cerebellum (yellow around green) due to their TAC similarity. By enforcing the exclusion constraint between putamen and cerebellum, we ensure that the final result is anatomically plausible (Figure 4.15(h)). We emphasize that a bad and irrational initialization will result in a wrong segmentation due to our local optimization framework. We empirically set λ_1, λ_2 and λ_3 to 0.68, 0.04 and 0.15, respectively, to balance between data, containment and exclusion terms. In fact, here, a very small weight for geometric constraint was enough to place the contours in the correct ordering to satisfy the geometric constraints.

4.3 A note on containment constraint without attraction forces

In this section, we discuss a special case of containment constraint in our framework with a corresponding result on real data. In the proposed framework, by having equation (4.1) (and its extended version (4.8)), there is always interactions (attraction/repulsion) between the surfaces with containment constraint. There are cases in which the attraction between two regions with containment constraint is not important, e.g. there might be several small disjoint regions j contained by region i. To address this case, we replace (4.1) with the following modified energy term for the case in which i contains j:

$$E_C^{ND}(\phi_i, \phi_j) = \int_\Omega H(-\phi_i(x))H(\phi_j(x))dx. \tag{4.17}$$

The way the above equation works is similar to the exclusion equation (4.4). E_C^{ND} penalizes region j that falls outside region i. Thanks to the level set's nature, we can enforce *minimum* and *maximum* distance (but without attraction/repulsion forces) between regions' boundaries by modifying (4.17) as follows

$$\begin{aligned} E_C^{ND}(\phi_i, \phi_j) = & \int_\Omega H(-\phi_i(x) + d_{min}) H(\phi_j(x)) dx \\ & + \int_\Omega H(\phi_i(x)) H(-\phi_j(x) - d_{max}) dx, \end{aligned} \tag{4.18}$$

where the first term enforces minimum distance of d_{min} pixels between two i and j regions and the second term ensures that i does not grow farther than d_{max} pixels from j's boundary. The term $H(\phi_i - d_{min})$ corresponds to the shrunk version of i by d_{min} pixels. The first term in (4.18) penalizes region j that falls outside the shrunk i. On the other hand, the term $H(\phi_j + d_{max})$ expands the zero level set of ϕ_j by d_{max} pixels. The second term in (4.18) penalizes region i that falls outside the expanded j. Figure 4.16 shows the mechanism of how (4.18) works.

One practical example of encoding containment between two regions but without any attraction/repulsion forces between their boundaries is lung blood vessels segmentation. In this case, blood vessels have to be contained in the lungs probably without any specific distance constraint between them. Here, we set $d_{min} = 1$ pixel. In this example, the lung stands out with high contrast and it is unlikely that the corresponding surface grow irrationally far from the blood vessels. Hence, we ignored the maximum distance (one may set d_{max} arbitrary large). Figure 4.17 shows 3D blood vessel segmentation in a lung. Figure 4.17(a) shows an unbiased initialization. Note how the incorrect segmentation in Figures 4.17(b) and (c) performed by ACWOE (without containment constraint) has been improved by our framework using (4.18) as the containment energy term (Figures 4.17(d) and (e)).

4.4 Chapter summary

In this chapter, we augmented the level set framework with two important geometric constraints, containment and exclusion, along with a distance prior for segmenting spatially-recurring multi-region objects. We showed that only adding the containment and exclusion terms into the level set framework can improve the segmentation results in a number of applications, even when only a simple intensity/color-based data term is used.

By comparing our local optimization-based framework in the continuous domain with its counterpart methods in the discrete domain [Delong and Boykov, 2009, Ulén et al., 2013], we draw the following conclusions:

1) Metrication error: This issue might not be an overwhelming issue in many medical applications, however, it remains a known issue with discrete (graph-based) methods, which makes it difficult to efficiently implement a convex regularizer like total variation in the discrete domain. On the other hand, due to our continuous formulation, our method is free from metrication error.

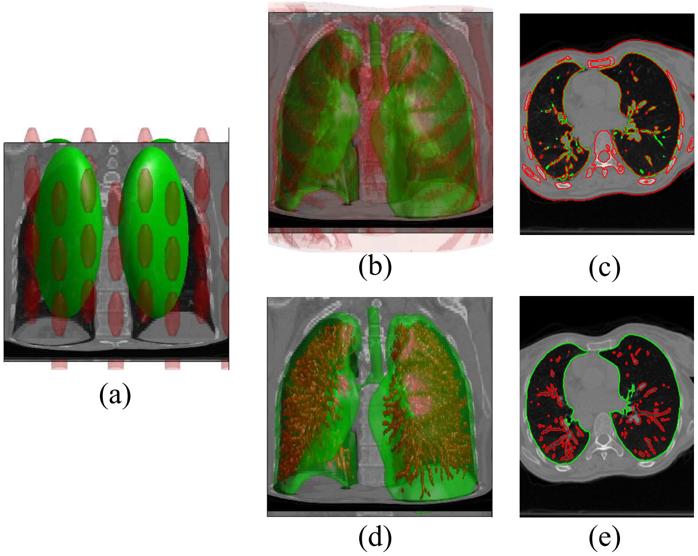


Figure 4.17: Lung and its blood vessels segmentation. (a) Initialization. (b) 3D ACWOE result. (c) ACWOE segmentation result in a 2D slice. (d) The proposed method result using (4.18) as the containment term. (e) The proposed method's result shown in a 2D slice.

2) Memory usage: Due to the graphical representation of an image, the graph based methods, e.g. [Delong and Boykov, 2009] and [Ulén et al., 2013], consume more memory compared with our framework. This also is due to the fact that [Delong and Boykov, 2009] and [Ulén et al., 2013] explore the whole search space to find the global solution as opposed to our method that only finds a locally optimal solution.

3) Runtime: In general, global graph based methods are highly efficient in finding the optimal solution. However, the computation time and memory usage in these methods, [Delong and Boykov, 2009] and [Ulén et al., 2013], depend intricately on the distance (thickness) constraint between regions due to the need for adding extra edges *per pixel*. The runtime and memory usage in our framework, on the other hand, is almost constant with respect to different distance constraints rendering our method suitable for segmenting very large microscopy images (100s of mega-pixels).

4) Initialization and numerical stability: While our method's results depend on the initialization and needs to satisfy the CFL condition for numerical stability, graph based methods avoid such requirements and have proven to be numerically stable. Nevertheless, we showed favourable results even with fully automatic or rough initialization that are distant from the desired boundaries.

The summary of pros and cons of our method proposed in this chapter are as follows:

- **Pros:**
  The proposed method
  - is parameterization-independent (defined over image domain)
  - can easily encode knowledge as energy terms
  - is implemented via update equations that can be implemented easily
  - handles topological changes (split and merge)
  - is easily extendible from curve in 2D to surfaces and higher dimensions
  - (compared to discrete methods)
    - has no metrication error
    - consume less memory
    - has less variation in computation time
- **Cons:** On the other hand, our method
  - has initialization dependency issue
  - does not guarantee the globally optimal solution

In Chapter 5, we turn our attention to optimizability and propose a continuous convex formulation to augment the popular Mumford-Shah model [Mumford and Shah, 1989] and develop a new regularization term to incorporate similar geometrical and distance prior as our second contribution while maintaining global optimality.

Chapter 5

Continuous convex formulation for multi-region object segmentation with geometric constraints

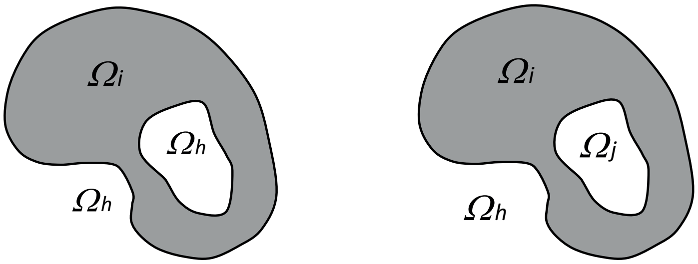
In our previous work (Chapter 4), results depended on initialization and globally optimal solution was not guaranteed. In this chapter we address these issues by proposing a convex formulation and augment the popular Mumford-Shah model [Mumford and Shah, 1989] to incorporate *containment* and *detachment* constraints between different regions with a specified *minimum distance* between their boundaries. Our method is able to handle multiple instances of multi-part objects defined by these geometrical constraints using a *single* labeling function while maintaining global optimality. We demonstrate the utility and advantages of these two constraints and show that our proposed convex continuous method is superior to other state-of-the-art methods, including its discrete counterparts, [Delong and Boykov, 2009] and [Ulén et al., 2013], in terms of memory usage, and metrication errors.

The remaining of this chapter is organized as follow. We introduce the containment and detachment constraints in Section 5.1.1. We show how to encode these two constraints in a continuous segmentation framework and show how our formulation can be convexified by functional lifting technique in Section 5.1.2. Section 5.1.3 explains how the energy is optimized. Different examples as well as comparisons with other popular state-of-the-art methods are given in Section 5.2, followed by our conclusions in Section 5.3.

5.1 Methods

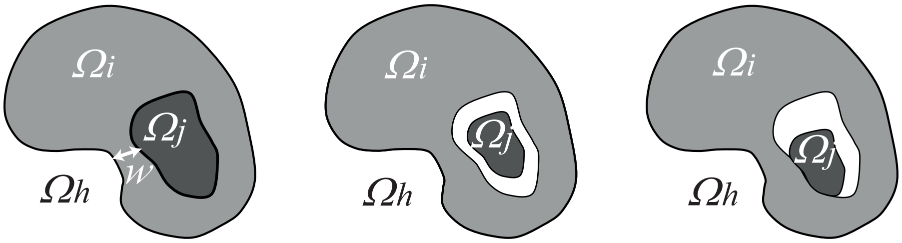
5.1.1 Problem formulation

In this section we explicitly define *containment* and *detachment* and show how we encode them in a Mumford-Shah based model while maintaining global optimality.



(a) Standard labeling function setting (b) Our setting

Figure 5.1: The inside vs. outside ambiguity in (a) is resolved by our *containment* constraint in (b).



(a) (b) (c)

Figure 5.2: Containment vs. similar configurations (h: background). According to (5.3), "object i contains object j" in (a) with $Th(\Omega_h, \Omega_i, \Omega_j) \geq w$, but the relationship between i and j in (b) and (c) is *not* containment.

We first consider a containment constraint in a 3 region segmentation. We divide the image domain, $\Omega \subset \mathbb{R}^2$, into three non-overlapping parts (Figure 5.1(b)): the outside or background region Ω_h, the outer region Ω_i, and the contained region Ω_j, where $\Omega = \Omega_h \cup \Omega_i \cup \Omega_j$. In many binary segmentation applications that use relaxed labeling functions, label values below 1/2 correspond to background and values above 1/2 correspond to foreground. We extend this definition as follows. Given a label set $\Gamma = [0, 1]$, we define our labeling function $u : \Omega \to \Gamma$, such that

$$\begin{aligned} 0 \leq u(\boldsymbol{x}) < 1/3 &\iff \boldsymbol{x} \in \Omega_h \\ 1/3 \leq u(\boldsymbol{x}) < 2/3 &\iff \boldsymbol{x} \in \Omega_i \\ 2/3 \leq u(\boldsymbol{x}) < 1 &\iff \boldsymbol{x} \in \Omega_j \ . \end{aligned} \tag{5.1}$$

To precisely define containment, we introduce a function that measures the *thickness* of the outer region Ω_i:

$$Th(\Omega_h, \Omega_i, \Omega_j) = \min_{\boldsymbol{x}_1 \in \Omega_j} \ \min_{\boldsymbol{x}_2 \in \Omega_h} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \ . \tag{5.2}$$

We define containment for 3 regions as:

Definition 1 (Containment). We say object i contains object j with thickness w if and only if

$$Th(\Omega_h, \Omega_i, \Omega_j) \geq w \ . \tag{5.3}$$

We note that $\Omega = \Omega_h \cup \Omega_i \cup \Omega_j$ is assumed here.

An example is shown in Figure 5.2(a) where the light gray object, i, contains the dark gray object, j, with a minimum thickness of w. The related configurations between i and j seen in Figure 5.2(b) and (c) are *not* containment based on our definition in (5.3). However, (b) can be seen as containment in a 4 region segmentation: i contains the interior white region, and the interior white region contains j.

Given an input image[1] $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, for objects i and j and the background h, let μ_i, μ_j and μ_h be constant approximations of the regional intensities and define $g_k(\boldsymbol{x}) = |I(\boldsymbol{x}) - \mu_k|^2$ for $k = \{h, i, j\}$. To segment I such that i contains j we solve the following energy minimization problem:

$$\begin{aligned} &\operatorname*{argmin}_{u \in D} E(u, \mathbf{g}) \\ &= \operatorname*{argmin}_{u \in D} \int_\Omega |\nabla_{\boldsymbol{x}} u(\boldsymbol{x})| + \rho(\boldsymbol{x}, u(\boldsymbol{x}), \mathbf{g}) d\boldsymbol{x} \ , \\ &D = \{u \mid u(\boldsymbol{x}) = 0 \text{ for } x \in \partial\Omega \text{ and} \\ &\qquad Th(\Omega_h, \Omega_i, \Omega_j) \geq w\} \ . \end{aligned} \tag{5.4}$$

Here $\nabla_{\boldsymbol{x}}$ is the gradient in x and y directions, $\mathbf{g} = (g_h, g_i, g_j)$, and $\rho(\boldsymbol{x}, u(\boldsymbol{x}), \mathbf{g}) : \Omega \rightarrow \mathbb{R}^+$ is a non-negative data term that encourages u to satisfy (5.1), e.g.

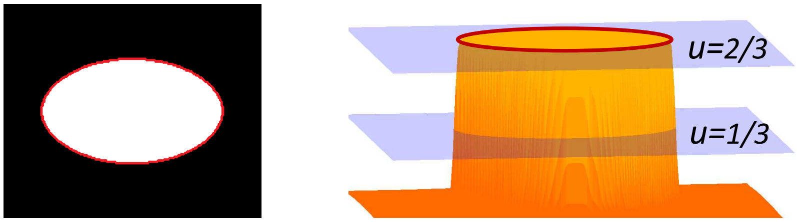
$$\rho(\boldsymbol{x}, u(\boldsymbol{x}), \mathbf{g}) = \begin{cases} g_h(\boldsymbol{x}) & \text{if } 0 \leq u(\boldsymbol{x}) < 1/3 \\ g_i(\boldsymbol{x}) & \text{if } 1/3 \leq u(\boldsymbol{x}) < 2/3 \\ g_j(\boldsymbol{x}) & \text{if } 2/3 \leq u(\boldsymbol{x}) < 1 \end{cases} \ . \tag{5.5}$$

For convenience, we often let ρ be a function of $\mathbf{g}$ implicitly, and write $\rho(\boldsymbol{x}, u(\boldsymbol{x}))$.

Constraining u to D, $E(u, \mathbf{g})$ ensures that object j and object h have no shared boundaries, resulting in j being contained in i. In other words, the segmentation corresponding to u cannot abruptly change from object j to object h, and thus the value of u cannot change from $u \geq 2/3$ to $u \leq 1/3$ in a distance less than w.

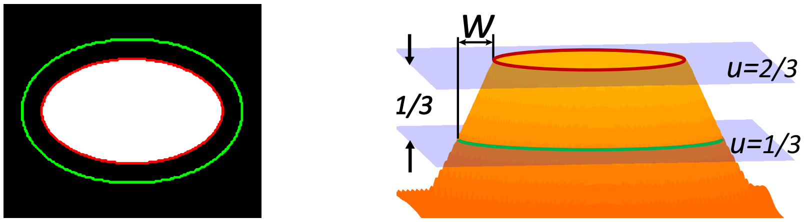
---

[1] Our method can be extended to vector valued images, e.g. color images or tensor fields, by modifying the data terms, as in [Chan et al., 2000], [Wang and Vemuri, 2004a]



(a) Unconstrained segmentation

(b) u corresponding to (a)



(c) Constrained segmentation

(d) u corresponding to (c), $|\nabla_{\boldsymbol{x}} u| \leq 1/3w$

Figure 5.3: Constrained vs. unconstrained labeling function u. In (a), the 3-region labeling function u, as defined in (5.1), is used to segment the white object from the black background, while the intermediate region has zero thickness. In (b), we see the u corresponding to (a). Without a thickness constraint, u is allowed to become discontinuous, skipping over the interval corresponding to the intermediate region, i.e. Ω_i in (5.1). In (c), we see how the segmentation changes when the thickness constraint (5.3) is enforced, with the intermediate region being hallucinated around the white object. In (d), we see the u corresponding to (c). By bounding the rate of change of u, a band of thickness w must be assigned to the intermediate region.

This leads us to the fact that the constraint $Th(\Omega_h, \Omega_i, \Omega_j) \geq w$ can be replaced by the following more convenient constraint:

$$|\nabla_{\boldsymbol{x}} u| \leq \frac{1}{3w} \ , \tag{5.6}$$

which limits the rate that u can change spatially. This lets us rewrite D as:

$$D = \left\{ u \;\middle|\; u(\boldsymbol{x}) = 0 \text{ for } x \in \partial\Omega,\ |\nabla_{\boldsymbol{x}} u| \leq \frac{1}{3w} \right\} \ . \tag{5.7}$$

To better understand the equivalence of these constraints, let's consider the example shown in Figure 5.3. Here, a black and white image is segmented into three regions, with μ_h corresponding to black pixels, μ_i corresponding to (non-existent) gray pixels, and μ_j corresponding to white pixels. Figure 5.3(b) illustrates the labeling function u corresponding to the segmentation in Figure 5.3(a), with no thickness constraint. Here, u becomes discontinuous (unbounded $|\nabla_{\boldsymbol{x}} u|$) in order to avoid assigning any pixels to the exterior object i.

By enforcing the restriction $|\nabla_{\boldsymbol{x}} u| \leq \frac{1}{3w}$, u is not able to jump from $< 1/3$ (background) to $\geq 2/3$ (white object) in less than distance w (Figure 5.3(d)). By restricting u, we force the white object to be contained by an intermediate region of thickness of w (cf. Figure 5.3(c)). We note that if w is large enough, the energy increase from hallucinating the intermediate region in Figure 5.3(c) will become greater than the energy increase from not segmenting the inner white region, and the result will be $u = 0$ across the image.

Using a similar formulation to containment, we can incorporate a constraint ensuring two regions are detached.

Definition 3 (Detachment). Object i and object j are detached with thickness w if and only if

$$Th(\Omega_i, \Omega_h, \Omega_j) \geq w. \tag{5.8}$$

In other words, object j does not share a boundary with object i, thereby ensuring a separation between i and j by enforcing the labeling function u to pass through the background, h, as it travels from i to j. To encode detachment, we simply swap g_h and g_i in (5.4).

Now, we note that the data term (5.5) is not convex, making standard gradient descent based optimization schemes prone to local minima. In the next section, we discuss how to convexify (5.4) and thus find a globally minimizing segmentation.

5.1.2 Function convexification

We use a "functional lifting" technique similar to the one proposed by [Pock et al., 2008] (motivated by Ishikawa's work in the discrete Markov random field setting [Ishikawa, 2003]) to transfer our energy functional to a higher dimensional space, where it becomes convex. The objective is to solve the following minimization problem:

$$\underset{u \in D}{\operatorname{argmin}} \left\{ \int_{\Omega} |\nabla_{\boldsymbol{x}} u(\boldsymbol{x})| d\boldsymbol{x} + \int_{\Omega} \rho(\boldsymbol{x}, u(\boldsymbol{x})) d\boldsymbol{x} \right\}. \tag{5.9}$$

The first term is a convex TV term, but the second term $\rho(\boldsymbol{x}, u(\boldsymbol{x}))$ can be non-convex. To lift the original energy to a higher dimensional space, we represent u in terms of its *super level set*, $\varphi : \Omega \times \Gamma \to \{0, 1\}$ by:

$$\varphi(\boldsymbol{x}, \gamma) = \mathbf{1}_{\{u \geq \gamma\}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } u(\boldsymbol{x}) \geq \gamma \\ 0 & otherwise \end{cases}. \tag{5.10}$$

u can be recovered from φ via the layer-cake formula:

$$u(\boldsymbol{x}) = \int_{\Gamma} \varphi(\boldsymbol{x}, \gamma) d\gamma. \tag{5.11}$$

The TV term in (5.9) can be re-written with respect to φ using the generalized co-area formula [Fleming and Rishel, 1960]:

$$\int_\Omega |\nabla_{\boldsymbol{x}} u(\boldsymbol{x})| d\boldsymbol{x} = \int_\Omega \int_\Gamma |\nabla_{\boldsymbol{x}} \varphi(\boldsymbol{x}, \gamma)| d\gamma d\boldsymbol{x}. \tag{5.12}$$

By observing[2] that $\delta(u(\boldsymbol{x}) - \gamma) \equiv |\partial_\gamma \varphi(\boldsymbol{x}, \gamma)|$, where $\delta(\cdot)$ is the Dirac delta function, the data term in (5.9) can then be re-written as [Pock et al., 2008]:

$$\begin{aligned} \rho(\boldsymbol{x}, u(\boldsymbol{x})) &= \int_\Gamma \rho(\boldsymbol{x}, \gamma) \delta(u(\boldsymbol{x}) - \gamma) d\gamma \\ &= \int_\Gamma \rho(\boldsymbol{x}, \gamma) |\partial_\gamma \phi(\boldsymbol{x}, \gamma)| d\gamma. \end{aligned} \tag{5.13}$$

Now, using (5.12) and (5.13), the equivalent form of (5.9) is

$$\underset{\{\varphi | u \in D\}}{\operatorname{argmin}} \int_\Omega \int_\Gamma |\nabla_{\boldsymbol{x}} \varphi(\boldsymbol{x}, \gamma)| + \rho(\boldsymbol{x}, \gamma) |\partial_\gamma \varphi(\boldsymbol{x}, \gamma)| d\gamma d\boldsymbol{x}. \tag{5.14}$$

Note that the non-convex function ρ does not depend on φ any more and (5.14) is convex in φ. As the last stage of convexification, φ in (5.10), is relaxed so it varies continuously between zero and one, i.e. $\varphi \in \Omega \times \Gamma \rightarrow [0, 1]$. To recover u, we threshold φ and apply (5.11).

Now, it is not immediately clear what form the constraint $\{\varphi | u \in D\}$ will take. If we ignore the thickness constraint from D, we could use $\varphi \in \mathcal{D}_1$ where

$$\mathcal{D}_1 = \{\varphi | \varphi(\boldsymbol{x}, 0) = 1,\ \varphi(\boldsymbol{x}, 1) = 0,\ \partial_\gamma \varphi \leq 0\}. \tag{5.15}$$

It is evident from (5.10) that every φ constructed from the super level sets of some u are in $\mathcal{D}_1$.

We now present a theorem describing how to enforce the thickness constraint in φ. Let $\nabla_3 \varphi$ be the gradient of φ in all components, i.e. $\nabla_3 \varphi = (\partial_x \varphi, \partial_y \varphi, \partial_\gamma \varphi)$.

Theorem 1. *If*

$$\frac{|\nabla_{\boldsymbol{x}} \varphi|}{|\partial_\gamma \varphi|} \leq \frac{1}{3w}, \tag{5.16}$$

then constraint $|\nabla_{\boldsymbol{x}} u| \leq \frac{1}{3w}$ *is satisfied by any* u *constructed by thresholding* φ *at some value and applying* (5.11).

Proof: The gradient of $\nabla_3 \varphi$ at point (x, y, γ) is perpendicular to its level set surface at that point. That is, if we let L_ν to be the ν-level surface of φ, let v_1 be tangent to L_ν at (x, y, γ), and let $v_2 = \nabla_3 \varphi(x, y, \gamma)$, then $\langle v_1, v_2 \rangle = 0$.

Based on (5.10), u is a level set of φ no matter where φ is thresholded. The standard technique for finding a vector tangent to the surface defined by u is to choose a unit vector

---

[2] From (5.10), it is observed that the derivative of φ with respect to γ is zero everywhere except where φ changes, i.e. $u(x) = \gamma$.

in x and y and set the γ component equal to the rate of change of u in the chosen direction. Specifically, this means

$$v = \left(\frac{\partial_x u}{|\nabla_{\boldsymbol{x}} u|}, \frac{\partial_y u}{|\nabla_{\boldsymbol{x}} u|}, |\nabla_{\boldsymbol{x}} u| \right) \tag{5.17}$$

is tangent to the surface u, and thus is orthogonal to $\nabla_3 \varphi$ at any point $(x, y, u(x, y))$. Thus

$$\begin{align} 0 &= \langle v, \nabla_3 \varphi \rangle \tag{5.18} \\ &= \frac{\langle \nabla_{\boldsymbol{x}} u, \nabla_{\boldsymbol{x}} \varphi \rangle}{|\nabla_{\boldsymbol{x}} u|} + \partial_\gamma \varphi |\nabla_{\boldsymbol{x}} u| \tag{5.19} \\ |\nabla_{\boldsymbol{x}} u| &= -\frac{\langle \nabla_{\boldsymbol{x}} u, \nabla_{\boldsymbol{x}} \varphi \rangle}{|\nabla_{\boldsymbol{x}} u|} \cdot \frac{1}{\partial_\gamma \varphi} \tag{5.20} \\ |\nabla_{\boldsymbol{x}} u| &\leq \frac{|\nabla_{\boldsymbol{x}} \varphi|}{|\partial_\gamma \varphi|} \leq \frac{1}{3w} \ , \tag{5.21} \end{align}$$

where the last step uses $\frac{\langle a,b \rangle}{|a|} \leq |b|$. □

The objective now is to solve:

$$\begin{align} \underset{\varphi \in \mathcal{D}_2}{\operatorname{argmin}} \int_\Omega \int_\Gamma |\nabla_{\boldsymbol{x}} \varphi| + \rho |\partial_\gamma \varphi| d\gamma d\boldsymbol{x}, \tag{5.22} \\ \mathcal{D}_2 = \left\{ \varphi \in \mathcal{D}_1 \middle| |\nabla_{\boldsymbol{x}} \varphi| \leq \frac{|\partial_\gamma \varphi|}{3w} \right\}. \tag{5.23} \end{align}$$

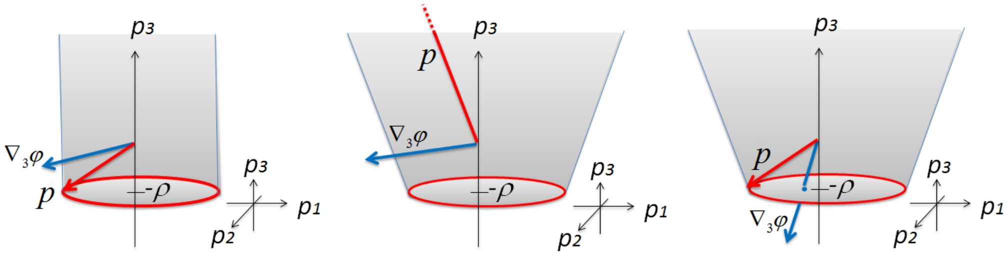
5.1.3 Optimization

Due to the discontinuity in the Euler-Lagrange equation of (5.22), we use a primal-dual algorithm, [Chan et al., ,Chambolle, 2005], to obtain the global solution. Defining the dual variable $\boldsymbol{p} = (p_1, p_2, p_3)^T$, we can write the total variation part of (5.22) as:

$$|\nabla_{\boldsymbol{x}} \phi(\boldsymbol{x}, \gamma)| = \max_{|\boldsymbol{p}'| \leq 1} \langle \boldsymbol{p}'(\boldsymbol{x}, \gamma), \nabla_{\boldsymbol{x}} \phi(\boldsymbol{x}, \gamma) \rangle \ , \tag{5.24}$$

where $\boldsymbol{p}' = (p_1, p_2)$. In (5.22), recalling that $-1 \leq \partial_\gamma \phi \leq 0$, it can be easily shown

$$\rho(\boldsymbol{x}, \gamma) |\partial_\gamma \phi(\boldsymbol{x}, \gamma)| = \max_{-p_3 \leq \rho} p_3(\boldsymbol{x}, \gamma) \partial_\gamma \varphi(\boldsymbol{x}, \gamma). \tag{5.25}$$



(a) (b) (c)

Figure 5.4: Valid sets for dual variables $\mathbf{p}$. (a) Set $\mathcal{C}_1$ [Pock et al., 2008] (without any geometrical constraint). (b-c) Set $\mathcal{C}_2$; the truncated cone that impose our constraint (5.16). (b) When (5.16) is not satisfies $\mathbf{p}$ becomes ∞. (c) When (5.16) is satisfied we obtain the same solution as (a).

Using these dual variables, the optimization problem of (5.22) becomes a min-max problem:

$$\underset{\varphi\in\mathcal{D}_2}{\operatorname{argmin}}\left\{\int_\Omega\int_\Gamma\left(\max_{|\boldsymbol{p}'|\le 1}\langle\boldsymbol{p}'(\boldsymbol{x},\gamma),\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\gamma)\rangle+\right.\right.$$
$$\left.\left.\max_{-p_3\le\rho}p_3(\boldsymbol{x},\gamma)\partial_\gamma\varphi(\boldsymbol{x},\gamma)\right)d\gamma d\boldsymbol{x}\right\} \tag{5.26}$$
$$=\underset{\varphi\in\mathcal{D}_2}{\operatorname{argmin}}\left\{\int_\Omega\int_\Gamma\max_{\boldsymbol{p}\in\mathcal{C}}\langle\nabla_3\varphi,\boldsymbol{p}\rangle\right\}, \tag{5.27}$$

Combining the constraints in (5.24) and (5.25) gives [Pock et al., 2008]:

$$\mathcal{C}=\mathcal{C}_1=\left\{\boldsymbol{p}\in\mathbb{R}^3\middle|\sqrt{p_1^2+p_2^2}\le 1,\quad p_3\ge-\rho\right\}, \tag{5.28}$$

a cylinder with radius one, open on one end (cf. Figure 5.4(a)). Recalling that $\partial_\gamma\varphi\le 0$, p_3 that maximizes (5.27) is always as negative as possible, i.e. $p_3=-\rho$. Thus for $\mathcal{C}=\mathcal{C}_1$, the maximizing $\boldsymbol{p}$ is always on the circle at the base of the cylinder, highlighted in red in Figure 5.4(a).

Unfortunately, while minimizing φ over $\mathcal{D}_1$ can be done efficiently, the gradient magnitude constraint $\mathcal{D}_2$ imposes (in order to enforce the thickness constraint) makes the minimization of (5.27) for φ difficult.

We will show that by moving the burden of enforcing the thickness constraint to the dual space, the optimization problem becomes much easier. Specifically, we will constrain φ to $\mathcal{D}_1$, and introduce a new space for the dual variables, $\mathcal{C}_2$, satisfying the following properties:

1. $\mathcal{C}_2$ is convex;

2. if φ satisfies (5.16) then

$$\max_{\boldsymbol{p}\in\mathcal{C}_2}\langle\nabla_3\varphi,\boldsymbol{p}\rangle=\max_{\boldsymbol{p}\in\mathcal{C}_1}\langle\nabla_3\varphi,\boldsymbol{p}\rangle\ ; \tag{5.29}$$

3. if φ does not satisfy (5.16) then (5.27) becomes arbitrarily large and that choice of φ will be disallowed:

$$\max_{\boldsymbol{p}\in\mathcal{C}_2}\langle\nabla_3\varphi,\boldsymbol{p}\rangle=\infty\ . \tag{5.30}$$

Note that since (5.16) must be enforced at each spatial location, conditions 2 and 3 must also hold at each spatial location, thus the integrals are dropped from (5.27).

Theorem 2. *A set that satisfies the three above conditions is the truncated cone defined by:*

$$\mathcal{C}_2=\left\{\boldsymbol{p}\middle|\sqrt{p_1^2+p_2^2}\leq 3w(p_3+\rho)+1,\ p_3\geq-\rho\right\}. \tag{5.31}$$

Proof. As a truncated cone, $\mathcal{C}_2$ is convex, so condition 1 is satisfied. To show 2 and 3 are satisfied, we will determine, for a given $\nabla_3\varphi$, the value of

$$\max_{\boldsymbol{p}\in\mathcal{C}_2}\langle\nabla_3\varphi,\boldsymbol{p}\rangle\ . \tag{5.32}$$

We first note that the circle at the base of the truncated cone is the same as the circle at the base of the cylinder defined by $\mathcal{C}_1$. Thus, if the $\boldsymbol{p}$ maximizing (5.32) lies on this circle (the red set in Figure 5.4(c)), then (5.29) is satisfied. We now simplify this problem by reducing the possible forms for $\nabla_3\varphi$ and $\boldsymbol{p}$.

Since the condition (5.16) and $\mathcal{C}_2$ are rotationally symmetric with respect to the first two components, we can assume the second component of $\nabla_3\varphi$ is 0, i.e. rotating $\nabla_3\varphi$ does not change (5.32). Further, if $\nabla_3\varphi$ is scaled by some constant, the maximizing $\boldsymbol{p}$ is not changed. Thus, we let $\nabla_3\varphi=(\ell,0,a)$, where $\ell=|\nabla_{\boldsymbol{x}}\varphi|$ is held constant and $\partial_\gamma\varphi=a\leq 0$ is the only free parameter.

The maximizing $\boldsymbol{p}$ is always on the boundary of $\mathcal{C}_2$, since scaling $\boldsymbol{p}$ until it is on the boundary increases (5.32). Further, the first two components of the maximizing $\boldsymbol{p}$ must align with the first two components of $\nabla_3\varphi$, so $p_2=0$.

If $\boldsymbol{p}$ is on the circle at the base of the cone, it would take the form $\boldsymbol{p}=(b,0,-\rho)$, where $0\leq b\leq 1$ is a free variable, and (5.32) becomes

$$\max_{\boldsymbol{p}}\ \langle\boldsymbol{p},\nabla_3\varphi\rangle=\max_{b}\ (b\ell-a\rho)\ , \tag{5.33}$$

which is maximized for $b=1$, i.e. a $\boldsymbol{p}$ on the edge of the circle (the red set in Figure 5.4(c)), so no maximizing $\boldsymbol{p}$ is on the inner part of the circle.

If $\boldsymbol{p}$ is on the surface of the cone, it would take the form $\boldsymbol{p} = (3w(b+\rho)+1, 0, b)$ (from (5.31)), where $b \geq -\rho$ is a free variable. This gives

$$\max_{\boldsymbol{p}} \langle \boldsymbol{p}, \nabla_3\varphi\rangle = \max_{b} \left(\ell(3wb + 3w\rho + 1) + ab\right) \tag{5.34}$$

$$= \max_{b} \left(b(3w\ell + a) + 3w\rho\ell + \ell\right). \tag{5.35}$$

If $(3w\ell + a) \leq 0$, (5.35) is maximized by minimizing b, i.e. $b = -\rho$. This corresponds to $\boldsymbol{p}$ on the edge of the circle (the red set in Figure 5.4(c)), which implies (5.29) is satisfied. If $(3w\ell + a) > 0$, (5.35) is maximized by $b \to \infty$, and (5.32) gets arbitrarily large, satisfying (5.30) (cf. Figure 5.4(b)).

To complete the proof, we note that $(3w\ell + a) \leq 0$ is equivalent to the thickness constraint (5.16):

$$(3w\ell + a) \leq 0$$

$$3w|\nabla_{\boldsymbol{x}}\varphi| \leq -\nabla_\gamma\varphi \tag{5.36}$$

$$|\nabla_{\boldsymbol{x}}\varphi| \leq \frac{|\nabla_\gamma\varphi|}{3w}\ . \tag{5.37}$$

□

To find the optimal solution for $\varphi \in \mathcal{D}_1$ and $\boldsymbol{p} \in \mathcal{C}_2$, we perform the following primal and dual steps [Chambolle, 2004]:

Primal step Find the minimum φ for a fixed $\boldsymbol{p}$:

$$\varphi^{k+1} = \underset{\varphi\in\mathcal{D}_1}{\operatorname{argmin}} \int_{\Omega\times\Gamma} \langle \nabla_3\varphi, \boldsymbol{p}^k\rangle + \frac{(\varphi - \varphi^k)^2}{2t_\varphi}. \tag{5.38}$$

Dual step Find the maximum $\boldsymbol{p}$ for a fixed φ:

$$\boldsymbol{p}^{k+1} = \underset{\boldsymbol{p}\in\mathcal{C}_2}{\operatorname{argmax}} \int_{\Omega\times\Gamma} \langle \nabla_3\varphi^{k+1}, \boldsymbol{p}\rangle + \frac{(\boldsymbol{p} - \boldsymbol{p}^k)^2}{2t_p}. \tag{5.39}$$

t_φ and t_p are the step sizes of the primal and dual update equations (here we used $t_\varphi = 0.01$ and $t_p = 5$). The solutions of (5.38) and (5.39) are derived from the Euler-Lagrange equations and projecting the obtained solutions to their valid sets. Algorithm 1 shows the alternating minimization scheme that is used to update φ and $\boldsymbol{p}$ in (5.38) and (5.39).

The update equations in lines 1 and 3 in Algorithm 1 are easily obtained by deriving the Euler-Lagrange equation for (5.38) and (5.39), respectively. We use Euclidean projector to reproject φ^{k+1} and $\boldsymbol{p}^{k+1}$ onto the sets $\mathcal{D}_1$ and $\mathcal{C}_2$, respectively.

In our numerical implementation and discretization, assuming that Δx and Δy are spatial step sizes in x and y directions and $\Delta\gamma$ is the label discretization step, we approximate $(x, y, \gamma) \in \Omega \times \Gamma$ by $(i\Delta x, j\Delta y, k\Delta\gamma)$, where (i, j, k) is the discrete location on the following

```
Algorithm 1: Finding the global optimizer of (5.9).
  Input: $I$: Image; $w$: thickness constant; $\mu_i$: average regional intensities, $t_p$ and $t_\varphi$:
         time steps.
  Output: Labeling function $u$.
  - Initialize labeling function, $u^0$ (the output does not depend on the initialization).
  - Initialize dual variable $\boldsymbol{p}$ to zero.
  - Form the super-level set function $\varphi^0$ from $u^0$ using (5.10).
  repeat
      1. Update $\varphi$: $\varphi^{k+1} = \varphi^k + t_\varphi \mathrm{div}_3 \boldsymbol{p}^k$.
      2. Project $\varphi^{k+1}$ onto the set $\mathcal{D}_1$.
      3. Update $\boldsymbol{p}$: $\boldsymbol{p}^{k+1} = \boldsymbol{p}^k + t_p \nabla_3 \varphi^{k+1}$.
      4. Project $\boldsymbol{p}^{k+1}$ onto the set $\mathcal{C}_2$.
  until convergence;
  Recover $u$ by thresholding $\varphi$ at 0.5 and using (5.11).
```

Cartesian grid:

$$\{(i,j,k)\ |1 \leq i \leq M, 1 \leq j \leq N, 1 \leq k \leq O\},$$

and M, N and O denote the size of the grid. We use forward and backward finite differences to approximate $\nabla_3\varphi$ and $\mathrm{div}_3\boldsymbol{p}$, respectively:

$$(\nabla_3\varphi)_{i,j,k} =$$
$$\left(\frac{\varphi_{i+1,j,k} - \varphi_{i,j,k}}{\Delta x}, \frac{\varphi_{i,j+1,k} - \varphi_{i,j,k}}{\Delta y}, \frac{\varphi_{i,j,k+1} - \varphi_{i,j,k}}{\Delta\gamma}\right)^T$$
$$(\mathrm{div}_3\boldsymbol{p})_{i,j,k} =$$
$$\frac{p^1_{i,j,k} - p^1_{i-1,j,k}}{\Delta x} + \frac{p^2_{i,j,k} - p^2_{i,j-1,k}}{\Delta y} + \frac{p^3_{i,j,k} - p^3_{i,j,k-1}}{\Delta\gamma}$$

Here, we set $\Delta x = \Delta y = 1$ and $\Delta\gamma = 0.05$.

After finding the global solution for the relaxed optimization problem, the labeling function u is recovered by thresholding φ at 0.5 and applying (5.11). It can be proven that thresholded minimizers of the relaxed problem are the minimizers of the binary problem (5.14). We follow Theorem 2 of [Pock et al., 2008] to prove that the thresholded minimizers of the relaxed problem ($\varphi \in \mathcal{D}_1$) are the minimizers of the binary problem (5.14).

Theorem 3. *Let $\varphi^* \in \mathcal{D}_1$ be the solution of the relaxed problem. Then for almost any threshold $\mu \in [0,1]$ the characteristics function $\boldsymbol{1}_{\{\varphi^* \geq \mu\}} \in \{0,1\}$ is also a minimizer of the binary variational problem* (5.14).

Proof by contradiction [Pock et al., 2008]. Since the relaxed problem is homogeneous of degree one, we make use of the generalized co-area formula to decompose it by means of the level sets of φ.

$$\begin{aligned} E(\varphi) &= \int_\Omega \int_{\boldsymbol{\Gamma}} |\nabla \varphi(\boldsymbol{x}, \gamma)| + \rho(\boldsymbol{x}, \gamma)|\partial_\gamma \varphi(\boldsymbol{x}, \gamma)| d\gamma d\boldsymbol{x} \\ &= \int_0^1 \left(\int_\Omega \int_{\boldsymbol{\Gamma}} |\nabla \mathbf{1}_{\{\varphi \geq \mu\}}| + \rho(\boldsymbol{x}, \gamma)|\partial_\gamma \mathbf{1}_{\{\varphi \geq \mu\}}| d\gamma d\boldsymbol{x} \right) d\mu \\ &= \int_0^1 E(\mathbf{1}_{\{\varphi \geq \mu\}}) d\mu. \end{aligned} \tag{5.40}$$

Assume to the contrary that $\mathbf{1}_{\{\varphi^* \geq \mu\}} \in \{0, 1\}$ is not a global minimizer of the binary problem (5.14), i.e. there exists a binary function $\varphi' \in \{0, 1\}$ with $E(\varphi') < E(\mathbf{1}_{\{\varphi^* \geq \mu\}})$ for a set of $\mu \in [0, 1]$. This directly implies that:

$$E(\varphi') = \int_0^1 E(\varphi') d\mu < \int_0^1 E(\varphi^* \geq \mu) d\mu = E(\varphi^*), \tag{5.41}$$

which means that φ^* is not a global minimizer of $E(.)$, contradicting our assumption. □

We note that our framework can be extended to multiple nested regions by dividing Γ into more than 3 intervals in (5.1). The thickness constraint between consecutive regions can be set by adjusting w and the interval length for each region in (5.1).

5.2 Experiments and discussion

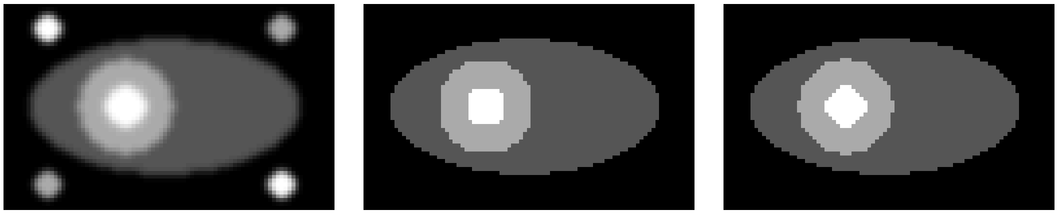
In this section, we demonstrate advantages of our work over popular state-of-the-art segmentation methods and compare our framework with the analogous discrete work of Delong and Boykov (DB) [Delong and Boykov, 2009].

5.2.1 Synthetic data

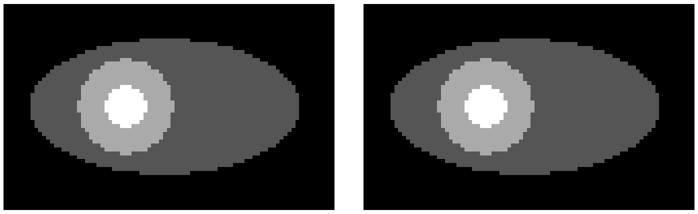
In our first experiment, we compare our method with DB in terms of memory usage and metrication error on a simple synthetic example.

Metrication error is defined as the artifacts which appear in graph-based segmentation methods due to penalizing region boundaries only across axis aligned edges. In Figure 5.5, the goal is to segment the three-region object from the background. Figure 5.5(b-d) resulted from DB's method for 4, 8 and 16 graph connectivity. Note the metrication artifacts in Figure 5.5(b-c). Increasing the graph connectivity reduces metrication error, but also increases memory usage.

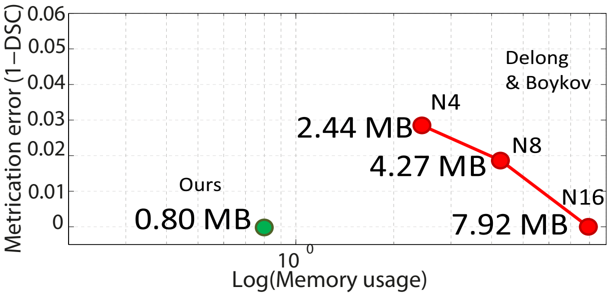
Memory usage of our method and the graph-based methods is seen in Figure 5.5(f). The red curve in Figure 5.5(f) illustrates the metrication error (1 - Dice similarity coefficient(DSC)) vs. memory usage of [Delong and Boykov, 2009] for 4, 8 and 16 connectivity, while the green circle represents our method. Here, removing the metrication error in the graph-based method requires 16 connectivity, even for these smooth objects. This requires ∼10 times more memory than our method (0.80 vs. 7.92 MB).



(a) Original image (b) GC: 4-connected (c) GC: 8-connected



(d) GC: 16-connected (e) Our result



(f) Metrication error vs. memory usage

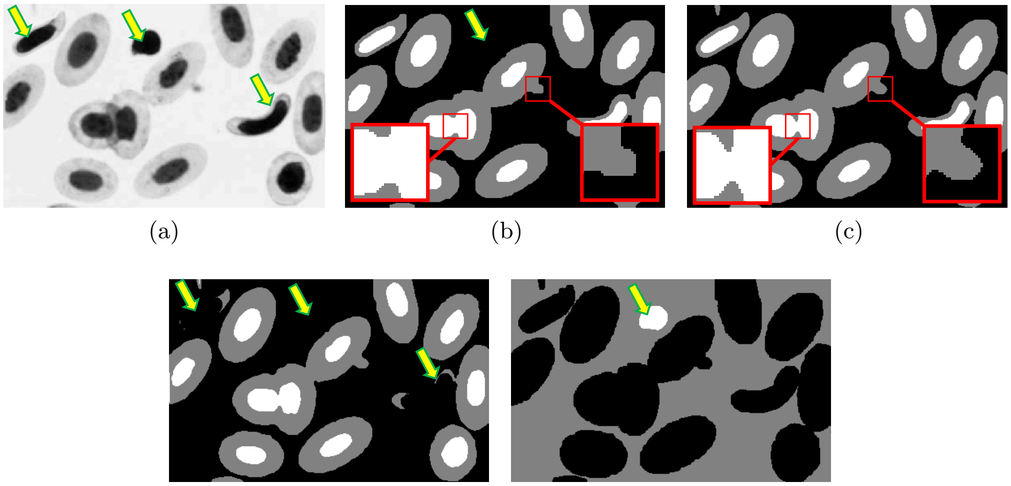
Figure 5.5: Synthetic three-region object+background segmentation. (b-d) DB graph cuts based method [Delong and Boykov, 2009] with different connectivities. (e) Our segmentation results. (f) Metrication error vs. memory usage: red curve: GC-based method; green circle: our method.

5.2.2 Microscopy/Histology cell segmentation

We applied our method to a set of 20 different histology and microscopy images consist of multiple instances of multi-region cells. In these experiments, we show how containment and detachment with thickness constraints are useful for cell segmentation and compare our method with DB [Delong and Boykov, 2009] in terms of memory usage and metrication error.

Figure 5.6 shows a typical microscopy image with multiple cells, where nuclei are typically contained inside a cell membrane. Figure 5.6(b) and (c) show the segmentation results with a thickness of $w = 2$ pixels for DB (with 4-connectivity) and our method, respectively. Metrication error can be clearly seen in Figure 5.6(b). Our method also requires less memory (7.91 MB vs. 33.90 MB).

Changing the thickness of the containing region allows us to control which objects are segmented. In Figure 5.6(a), to exclude the abnormal cells (arrows) from the segmentation,


(a) (b) (c)

(d) (e)

Figure 5.6: Cell segmentation in a microscopy imagery. (a) Original image, 250 × 395 pixels. Arrows show abnormal cells. (b) Result of [Delong and Boykov, 2009], 33.90 MB. (c) Our result (thickness= 2 pixels, 7.91 MB) . (d) Our result to segment only normal looking (elliptical) cells (thisckness=10 pixels, 7.91 MB). Note that [Delong and Boykov, 2009] needs ∼ 313.41 MB extra memory (347.31 MB in total) to impose thickness of 10 pixels while the memory usage of our method is independent of thickness constraint. (e) Segmenting isolated nuclei by imposing *detachment* constraint (5.8).

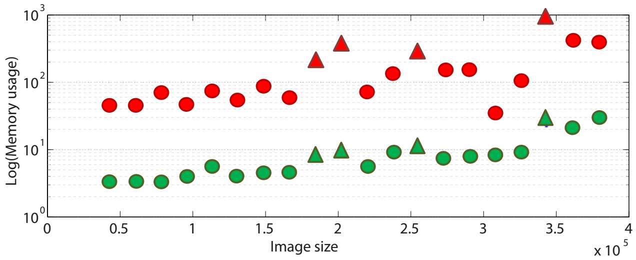


Figure 5.7: Memory efficiency: DB (in red) vs. ours (in green). ○: 3-region segmentation; Δ: 4-region segmentation. Memory usage ratio (DB/ours) for 3 regions: 14.63 ± 4.52 and for 4 regions: 32.40 ± 8.35.

we increase the thickness of the outer region (membrane) from 2 to 10 pixels, resulting in Figure 5.6(d). In DB, increasing the thickness requires more edges be added to the underlying graph, and increasing the thickness from 2 to 10 pixels requires an extra ∼313.41 MB memory, an almost 10-fold increase. Thickness can be increased in our method by simply changing the value of w. Figure 5.6 (e) demonstrates the usage of a detachment constraint (cf. (5.8)), identifying nuclei that are not surrounded by a cell membrane.

Figure 5.8 shows results for 7 other images, comparing segmentations generated using a continuous method without a containment constraint [Pock et al., 2008], DB with 4 connectivity, and our method. These results illustrate the importance of a containment

Table 5.1: DSC and memory usage comparison (20 images).

Method	**DSC**	**Memory (MB)**
No containment [Pock et al., 2008]	0.6478 ± 0.06	8.9 ± 7.2
DB [Delong and Boykov, 2009]	0.9065 ± 0.08	180.0 ± 204.7
Ours	0.9158 ± 0.07	8.9 ± 7.2

constraint (first column) and also show the effects of metrication error (second column). Table 5.1 summarizes the mean accuracy and memory usage of the 3 methods across all 20 images, and Figure 5.7 compares the memory usage vs. image size of our method and DB across all 20 images. We note that some of the images (e.g. the bottom two rows in Figure 5.8) have 4 regions segmented, and Figure 5.7 shows that the memory usage in graph-based methods tends to increase proportionally more than our continuous method in these cases. On average, our method converges after 200 iterations for a 256×256 image. Using non-optimized MATLAB code on a standard 2.3 GHz CPU with 6GB RAM, the graph cuts-based method [Delong and Boykov, 2009] tends to run 2-3 times faster than our method but with more memory usage.

5.3 Chapter summary

We introduced a variational framework to augment the conventional Mumford-Shah model for segmenting multi-region objects. We proposed a labeling function that allows us to enforce useful geometric constraints such as containment and detachment. By using this framework, a user can easily apply high level intuitive geometric constraints to improve segmentation results without the need for a deep understanding of how the method works. Our method compares favourably with analogous graph-based methods in terms of metrication error and memory efficiency.

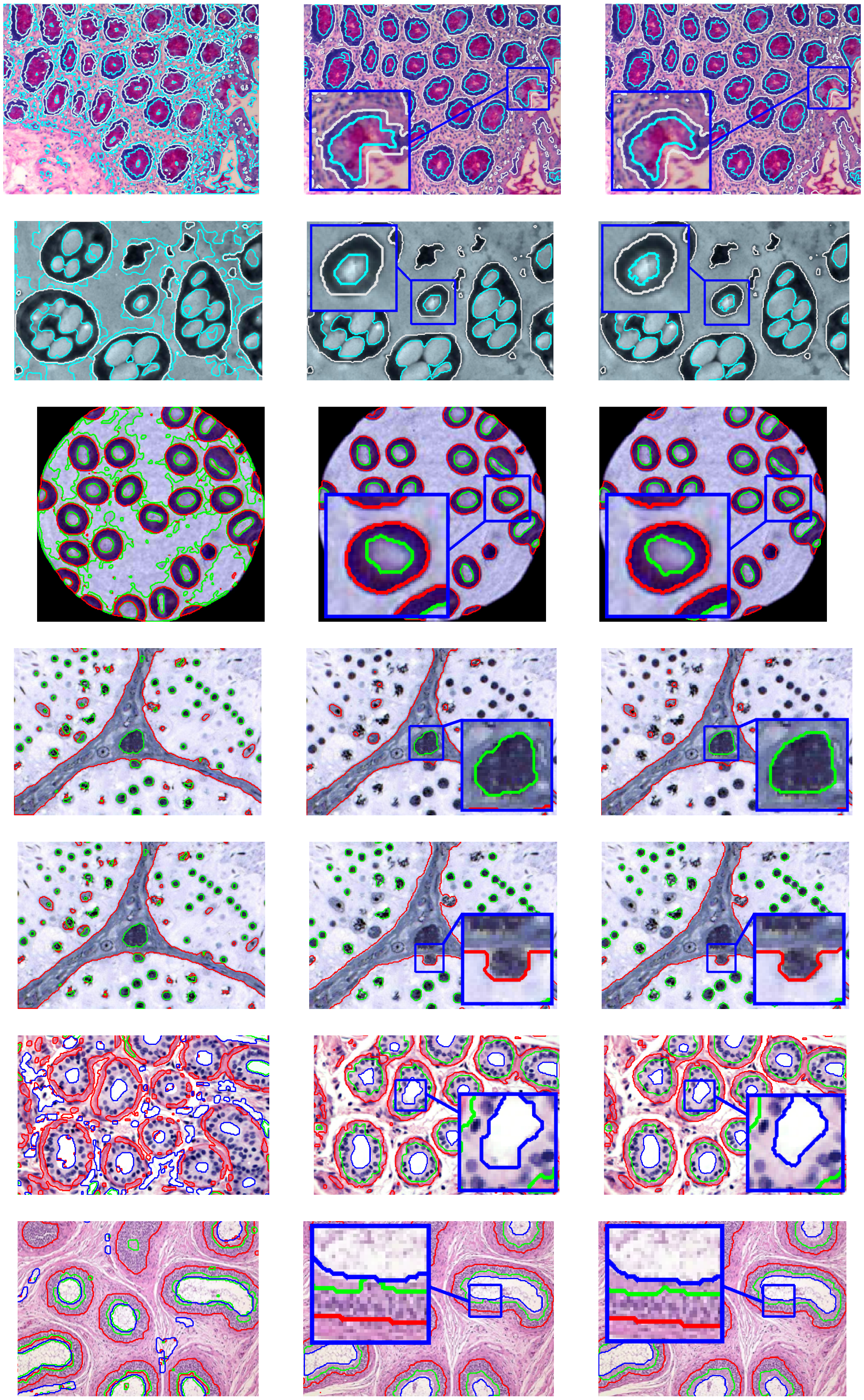


Figure 5.8: Incorporating geometrical constraint into the segmentation of histology and microscopy images. **Left column:** Pock's method [Pock et al., 2008] (without geometrical constraint); **middle column:** graph-based method [Delong and Boykov, 2009]; **right column:** our method.

Chapter 6

Endoscopic video segmentation in robot-assisted minimally invasive surgeries using prior information

While our previous contributions enforce geometrical priors into segmentation frameworks, in this chapter we enforce priors based on images acquired from different imaging modalities. Particularly, we employed 3D models as well as camera motion prior to segment multiple objects in a 2D multi-view endoscopic video.

Robotic minimally invasive surgery (RMIS) systems have been gaining popularity due to their many advantages compared to traditional minimally invasive surgery and open surgeries including greater precision, improved dexterity and enhanced 3D immersive visualization for surgeons [Pratt et al., 2012]. In RMIS, image-guided localization and delineation of tissues, e.g. tumour and kidney in partial nephrectomy, is an important step that can significantly enhance the surgeon's perception of the surgical scene and facilitate their decision-making. However, accurate identification of various tissues in an intraoperative video is by no means an easy task due to the limited viewing area, presence of occluding objects (e.g. surgical tools), data acquisition noise (e.g. specular light reflection, blood and smoke), as well as similarity in the visual appearance of different tissues. To localize different tissues, surgeons rely on (mentally) combining the information they recall from preoperative scans, often CT, with the information they see in the intraoperative stereo endoscopic video feed, a task that requires exceptional skill. Advances in intraoperative imaging has introduced some other modalities into the operating room, e.g. ultrasound and X-ray [Estépar and Vosburgh, 2014]. However, the feasibility, quality, and information content of such data still markedly lags behind the typically high resolution 3D preoperative data, and endoscopic imaging remains the staple modality in minimally invasive surgeries. Using efficient 2D and 3D computer vision techniques to support the analysis of endoscopic images alleviates the need for using additional equipment, e.g. fiducial markers, and helps to relax the handling of multiple intraoperative data streams. Yet, the aforementioned complications related to noise and clutter pose many challenges in endoscopic video segmentation.

Some recent works in endoscopic image segmentation proposed the use of level set-based approaches [Figueiredo et al., 2010, Figueiredo et al., 2012] while others focused on parameter-sensitive morphological operations and thresholding techniques [Dhandra and et al., 2006, Mewes et al., 2011]. However, practical success of such methods is limited as they mostly rely on image color/intensity, do not use preoperative information, and focus on segmenting a single object. Other contributions focused on feature tracking, e.g. [Puerto-Souza and Mariottini, 2013], while [Mountney and Yang, 2010] proposed a method to estimate laparoscopic camera and *periodic* organ motion. Both methods cannot handle free-form deformation of organs and the latter method assumed the camera has (on average) constant velocity, which is often not the case in minimally invasive surgery.

In this chapter, we propose a technique that imitates surgeons skill in leveraging preoperative information into the analysis of intraoperative endoscopic visual cues. Our approach encodes the fused information within an energy optimization process to efficiently segment multiple structures in endoscopic videos. Our work is inspired by the work of [Prisacariu and Reid, 2012] in the (non-medical) computer vision area where the segmentation of an object in an image was obtained by finding the six pose parameters of its 3D model. In minimally invasive surgery, six degrees of freedom are not enough as tissues non-rigidly deform. Unlike [Prisacariu and Reid, 2012], [Sandhu et al., 2011] used kernel PCA to capture the shape variation and estimated the non-rigid pose of a single object. Their method, however, segments a single object in a single view image. Applying the methodology of [Prisacariu and Reid, 2012] and [Sandhu et al., 2011] to robotic surgery applications is also not straightforward as images in endoscopic videos are highly noisy and cluttered. Moreover, [Sandhu et al., 2011] does not leverage stereo vision and is incapable of handling *large* occlusions (large portions of objects are occluded, e.g. by tools), both of which are common in minimally invasive surgery. In our formulation, we provide a *closed-form solution* (unlike [Prisacariu and Reid, 2012, Sandhu et al., 2011]) to segment *multiple tissues* in a *multi-view* endoscopic video (here we use stereo video) based on prior knowledge extracted from preoperative data. Our approach thus simultaneously estimates the 3D pose of tissues in the preoperative domain as well as their *non-rigid* deformations from their preoperative state. Furthermore, our framework allows for the inclusion of *motion priors* on laparoscopic camera motion to stabilize the segmentation/pose tracking in the presence of a large occlusion. Such feature is especially useful in RMIS as camera motion signals can be easily read using the robot's API and incorporated into our formulation to obtain even more accurate and robust results. In this work, we enforce a motion prior on the cameras only; nonetheless, our flexible mathematical formulation allows for incorporating of general non-rigid motion or temporal deformation priors on various tissues. Our method runs in *real-time on a single CPU core* which makes it suitable for robotic surgical systems as they are typically limited in computational resources.

We validate the utility of our technique on *ex vivo* data as well as *in vivo* clinical data from laparoscopic partial nephrectomy surgery and demonstrate its robustness in segmenting stereo endoscopic videos.

6.1 Methods

For ease of exposition, we start by describing our method for segmenting a **single object in a single image**, given its 3D segmentation in the preoperative domain. Later, we show how we extend our framework to segment multiple objects in multiple 2D images (multi-view data). Let $\boldsymbol{X}_i^{pre} \in \mathbb{R}^3$ be point i of a 3D model obtained from segmented preoperative 3D data. The goal is to spatially transform and deform the model non-rigidly in 3D such that its silhouette on the 2D color image, $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^3$, delineates the object of interest in I, i.e. the silhouette encapsulates the foreground. The silhouette of a 3D model is obtained by projecting the model from 3D to 2D given the projection function π and the corresponding camera's focal point f_m and principal point c_m. We represent the foreground by the level set function ϕ such that, for a pixel $\boldsymbol{x}_i \in \Omega$ in I, $\phi(\boldsymbol{x}_i) > 0$ if $\boldsymbol{x}_i$ belongs to the foreground, $\phi(\boldsymbol{x}_i) < 0$ if $\boldsymbol{x}_i$ belongs to the background, and $\phi(\boldsymbol{x}_i) = 0$ if $\boldsymbol{x}_i$ is on the object's boundary. Every 2D point $\boldsymbol{x}_i$ on the foreground is related to its corresponding 3D point $\boldsymbol{X}_i^{pre}$ by

$$\boldsymbol{x}_i = \pi(T(\boldsymbol{X}_i^{pre}; \boldsymbol{\xi})) \ , \tag{6.1}$$

where T deforms $\boldsymbol{X}_i^{pre}$ and transforms it from the preoperative domain to $\boldsymbol{X}_i^{srg} \in \mathbb{R}^3$ in the surgical domain. $\boldsymbol{\xi} = \{\boldsymbol{\xi}_q, \boldsymbol{\xi}_w\}$ is the set of pose $\boldsymbol{\xi}_q = \{q_1, \cdots, q_n\}$ and shape $\boldsymbol{\xi}_w = \{w_1, \cdots, w_m\}$ parameters. We use the weights of shape variation modes as $\boldsymbol{\xi}_w$. To segment a tissue in the 2D image I given its 3D model, we define the following residual for the i^{th} pixel, $\boldsymbol{x}_i \in \Omega$ as:

$$e_i = -\rho_f(\boldsymbol{x}_i) H(\phi(\boldsymbol{x}_i)) + \rho_b(\boldsymbol{x}_i)(1 - H(\phi(\boldsymbol{x}_i))), \tag{6.2}$$

where ρ_f and ρ_b are the regional terms measuring the agreement of the image pixel $\boldsymbol{x}_i$ with the foreground and background statistical models and $H(.)$ is the Heaviside step function. Ideally e_i would be zero for a perfect model-to-data fit, however, due to noise, e_i will have a distribution $P(e_i|\boldsymbol{\xi})$, which can be modelled as $\mathcal{N}(0, \sigma)$ when $\boldsymbol{\xi}$ is close to the solution (Figure 6.1(a)). The residual value for all pixels $\boldsymbol{e}$ is calculated assuming that the noise is independent across pixels. The objective is to find the most likely transformation parameters

$\boldsymbol{\xi}$ given the residual $\boldsymbol{e}$ by maximizing the following posterior probability:

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}} \ P(\boldsymbol{\xi}|\boldsymbol{e}) = \arg\min_{\boldsymbol{\xi}} \ -\sum_i \log P(e_i|\boldsymbol{\xi}) - \log P(\boldsymbol{\xi}) \ . \tag{6.3}$$

The second term in (6.3) is the prior on the transformation. Here, we enforce a prior only on pose parameters ($\boldsymbol{\xi}_q$) which can also be considered as camera motion parameters (due to their relative rigid motion); however, one can use this term to enforce a prior on the tissues' non-rigid deformation. The camera motion prior can be obtained either from the robotic surgical system or the prediction of a Kalman filter. The uncertainty in the motion estimation is modelled with a Gaussian meaned around the predicted pose parameter $\hat{\boldsymbol{\xi}}$ with covariance Σ_ξ, i.e. $P(\boldsymbol{\xi}) \sim \mathcal{N}(\hat{\boldsymbol{\xi}}, \Sigma_\xi)$. We minimize (6.3) by taking its derivative and setting it to zero:

$$\sum_i \frac{\partial \log P(e_i|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} + \frac{\partial \log P(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \sum_i \frac{-1}{\sigma} e_i \frac{\partial e_i}{\partial \boldsymbol{\xi}} - \frac{1}{\Sigma_\xi}(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}) = 0. \tag{6.4}$$

As e_i is not linear in $\boldsymbol{\xi}$, to solve (6.4) efficiently we linearize e_i with respect to $\boldsymbol{\xi}$ using the first order Taylor approximation:

$$e_i^{lin}(\boldsymbol{\xi}) = e_i(\mathbf{0}) + \frac{\partial e_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\Big|_{\boldsymbol{\xi}=0} \Delta\boldsymbol{\xi} \ = \ e_i(\mathbf{0}) + J_i \Delta\boldsymbol{\xi}, \tag{6.5}$$

where J_i is the Jacobian of the i^{th} pixel's error with respect to $\boldsymbol{\xi}$. Substituting (6.5) into (6.4) and using matrix notation, we have:

$$(\frac{-1}{\sigma}\boldsymbol{J}^T\boldsymbol{J} + \Sigma_\xi^{-1})\Delta\boldsymbol{\xi} = \frac{1}{\sigma}\boldsymbol{J}^T\boldsymbol{e}(0) + \Sigma_\xi^{-1}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{t-1}) \ , \tag{6.6}$$

where $\boldsymbol{J}$ is the stacked matrix of all J_i pixel-wise Jacobians and $\boldsymbol{\xi}_{t-1}$ is the pose in the previous frame. At each frame of video, the linear system of equations (6.6) is solved efficiently for $\Delta\boldsymbol{\xi}$ (by Choleskey decomposition) with which we update the transformation/deformation parameters $\boldsymbol{\xi}$. For this linearization, we assumed $\Delta\boldsymbol{\xi}$ is small. This assumption is valid given the high video capture rate of current surgical systems, e.g. daVinci with ∼30 FPS. To handle larger transformations, one may apply a coarse-to-fine scheme. The gradient of the i^{th} pixel's error with respect to ℓ^{th} component of $\boldsymbol{\xi}$ (ℓ^{th} element of J_i) is calculated as:

$$J_i(\ell) = \frac{\partial e_i}{\partial \xi_\ell} \ = \ (-\rho_f - \rho_b)\frac{\partial H}{\partial \phi}\frac{\partial \phi}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \xi_\ell} \ = \ (-\rho_f - \rho_b)\delta(\phi)\nabla_x \phi \frac{\partial \boldsymbol{x}}{\partial \xi_\ell} \ , \tag{6.7}$$

where $\delta(.)$ is the Dirac delta function. Given the camera parameters f_c and c_c, every 2D point $\boldsymbol{x} = (x, y)$ in I is related to at least one corresponding 3D point $\boldsymbol{X}^{srg} =$

$(X^{srg}, Y^{srg}, Z^{srg})$ by:

$$x = \frac{f_c}{Z^{srg}} X^{srg} + c_c \;, \tag{6.8}$$

$$y = \frac{f_c}{Z^{srg}} Y^{srg} + c_c \;. \tag{6.9}$$

Hence,

$$\frac{\partial x}{\partial \xi_\ell} = \frac{f_c}{Z^{srg2}} \big(Z^{srg} \frac{\partial X^{srg}}{\partial \xi_\ell} - X^{srg} \frac{\partial Z^{srg}}{\partial \xi_\ell} \big) \;. \tag{6.10}$$

We similarly calculate $\frac{\partial y}{\partial \xi_\ell}$. 3D point $\boldsymbol{X}^{srg}$ is related to $\boldsymbol{X}^{pre}$ by the transformation function T such that:

$$\boldsymbol{X}^{srg} = T(\boldsymbol{X}^{pre}; \boldsymbol{\xi}) = \boldsymbol{R}(\boldsymbol{\xi}_q)\boldsymbol{X}^{pre} + \boldsymbol{t}(\boldsymbol{\xi}_q) \;, \tag{6.11}$$

where $\boldsymbol{R}$ and $\boldsymbol{t}$ are rotation (linear) and translation matrices and can represent any rigid (linear) transformation upon the choice of pose parameters in T, $\boldsymbol{\xi}_q = \{q_1, \cdots, q_n\}$. Therefore, $\frac{\partial X^{srg}}{\partial q_i}$, $\frac{\partial Y^{srg}}{\partial q_i}$ and $\frac{\partial Z^{srg}}{\partial q_i}$ in (6.10) are easily calculated upon the choice of transformation function.

To handle non-rigid deformation, we generate a catalog of realistic 3D deformed shapes for the organs of interest (Figure 6.1(b)) using DeformIt [Hamarneh et al., 2008]. Having this catalog, we obtain the modes of variation through principal component analysis such that a novel 3D shape of an organ/tissue can be estimated as:

$$\boldsymbol{X}^{pre} = \overline{\boldsymbol{X}^{pre}} + \boldsymbol{U}\boldsymbol{w} \;, \tag{6.12}$$

where $\overline{\boldsymbol{X}^{pre}}$ is the coordinates of the average shape of the organ of interest, $\boldsymbol{U} = \{u_1, \cdots, u_K\}$ are the K principal modes of variation and $\boldsymbol{w} = \boldsymbol{\xi}_q = \{w_1, \cdots, w_K\}$ are their weights. We chose the number of principal modes explaining 97% of the variance in the training set. The derivative of a 3D point $\boldsymbol{X}^{srg}$ in the surgical domain with respect to w_ℓ is calculated as:

$$\frac{\partial \boldsymbol{X}^{srg}}{\partial w_\ell} = \boldsymbol{R} \frac{\partial \boldsymbol{X}^{pre}}{\partial w_\ell} = \boldsymbol{R} \cdot u_\ell \;, \tag{6.13}$$

where u_ℓ is the ℓ^{th} mode of variation in $\boldsymbol{U}$.

We now describe the extension to segmenting **multi-object in multi-view** images. Having N tissues of interest in the preoperative data and M views in the intraoperative domain, we extend our framework to segment multiple objects in multi-view images by computing a level set function for each object. Also, all the pixels from all views contribute toward calculating the residual. The residual of pixel $\boldsymbol{x}_i$ belonging to the n^{th} tissue is calculated as:

$$e_i = \sum_{m=1}^{M} \Big(-\rho^n_{f,m} \, H(\phi^n_m(\boldsymbol{x}_i)) + \rho^n_{b,m} \, (1 - H(\phi^n_m(\boldsymbol{x}_i))) \Big) \,, \tag{6.14}$$

where $\rho^n_{f,m}$ and $\rho^n_{b,m}$ are the regional terms and ϕ^n_m is the level set of the n^{th} object in the m^{th} image. Note that in the multi-view scenario, the extrinsic camera parameters $(\boldsymbol{R}^{ext}_m, \boldsymbol{t}^{ext}_m)$ have to be considered in calculating the Jacobian, i.e. $\boldsymbol{R}$ in the above equations is replaced by $\boldsymbol{R}^{ext}_m \boldsymbol{R}$.

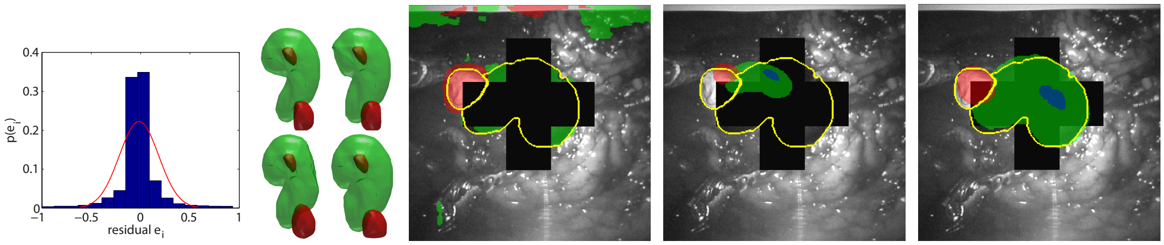
To make our method more robust, we leverage a variety of image features (normalized RGB and YCbCr and local color histogram) to calculate the regional terms, $\rho^n_{f,m}$ and $\rho^n_{b,m}$ in (6.14), for different tissues in all 2D views. We estimate the probability of a given pixel $\boldsymbol{x}$ belonging to the n^{th} object (O_n), $P(\boldsymbol{x} \in O_n | I_m)$, and its background (B_n), $P(\boldsymbol{x} \in B_n | I_m)$, in the m^{th} image, I_m, by training a random forest consisting of N_b binary decision trees (here $N_b = 20$). To train a RF, we select few 20 × 20 patches on different tissues in 2% of all frames from the *same* patient, i.e. ∼10 frames out of ∼600 frames. In practice, surgeons may select these patches with the help of surgical tools. The regional terms are calculated as:

$$\rho^n_{f,m}(\boldsymbol{x}) = -\log P(\boldsymbol{x} \in O_n | I_m) \tag{6.15}$$

$$\rho^n_{b,m}(\boldsymbol{x}) = -\log P(\boldsymbol{x} \in B_n | I_m) \,. \tag{6.16}$$

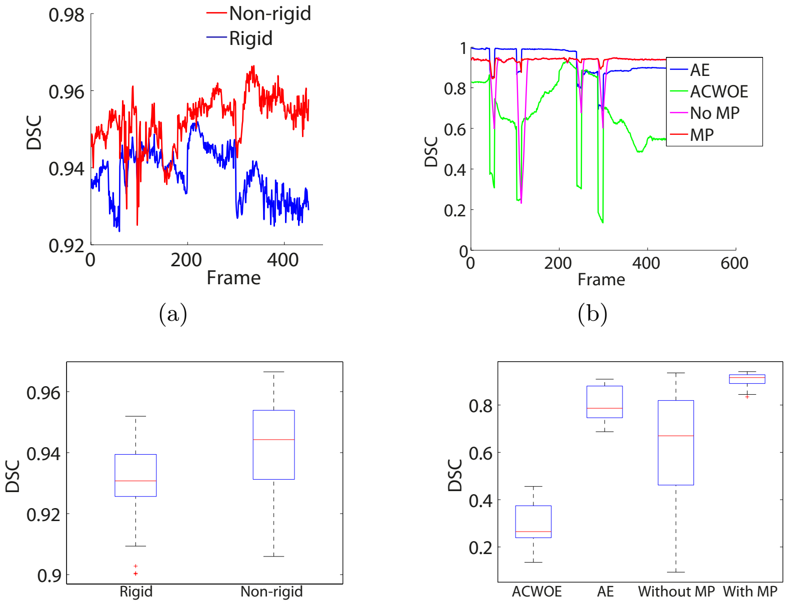
6.2 Materials

We evaluated our framework on 10 *ex vivo* lamb kidney datasets as well as three *in vivo* clinical partial nephrectomy data. We constructed the set of *ex vivo* phantoms using lamb kidneys and implanted artificial tumours outside and inside each kidney to emulate a partially exophytic and completely endophytic tumour, respectively. CT volumes and stereo video sequences of our phantoms were captured by a Somatom CT scanner (Siemens, Germany) and a daVinci S system (Intuitive Surgical, USA), respectively. We segmented the kidney and tumours in each CT using the TurtleSeg software [Top and et al., 2011]. We simulated deformations of each kidney and tumour in respectively ∼40 and ∼15 different ways using DeformIt [Hamarneh et al., 2008] (Figure 6.1(b)). To obtain ground truth segmentation for stereo videos, we used the “Rotobrush” tool of After Effect (Adobe Systems Inc., USA) as a semi-automatic segmentation tool allowing for visual inspection and correction. For initialization, we manually aligned the preoperative model with the intraoperative image by choosing ∼6-8 landmarks in the CT and the first frame of the stereo video.



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

Figure 6.1: (a) Residual distribution. (b) Few samples from our kidney and tumour catalog. Segmentation result in the presence of artificial tools (black cross) largely occluding the kidney and tumour phantoms using (c) ACWOE, (d) our method without any motion prior and (e) our method with motion prior. Green: kidney; Red: exophytic tumour; Blue: endophytic tumour; Yellow: ground truth.



(a) (b)

(c) (d)

Figure 6.2: DSC vs. frame of a phantom for (a) rigid vs. non-rigid transformation and, for (b) ACWOE, AE [Bai and et al., 2009] and our method with and without motion prior (MP vs. No MP). Box plot representation of DSC for the whole dataset is presented in (c) for rigid vs. non-rigid transformation and in (d) for ACWOE, AE and our method with/without motion prior over the occlusion periods.

6.3 Experiments

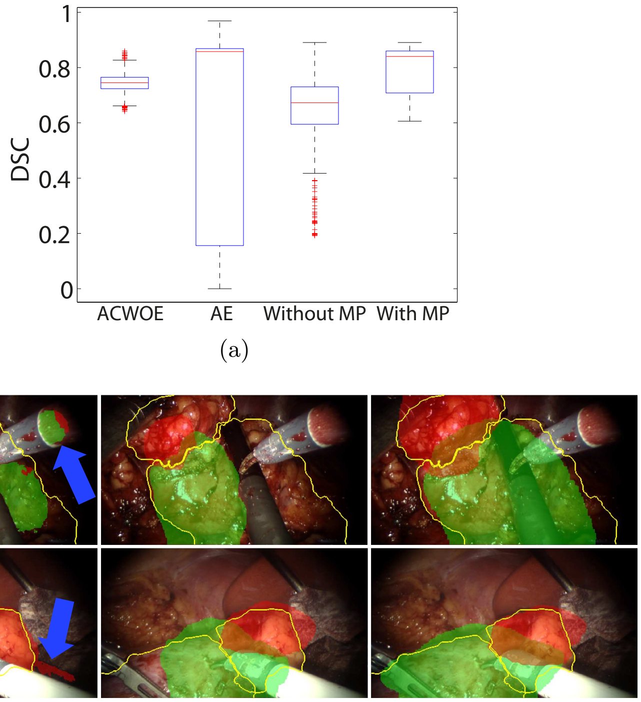
Our **first experiment** on *ex vivo* data focused on evaluating our method w.r.t. using a rigid vs. deformable transformation model. Figure 6.2(a) shows an example DSC vs. time for one of our phantoms. It is seen in Figure 6.2(a) that, as expected, incorporating the non-rigid deformation of tissues improves the final results. Note that the method in [Puerto-Souza and Mariottini, 2013] does not handle non-rigid deformation and [Mountney and Yang, 2010] can only estimate the non-rigid deformation for organs with *periodic* motion, whereas our method does not pose any such constraints. Results for all the *ex vivo* phantom datasets are shown in Figure 6.2(c).

In our **second experiment**, we artificially occluded large portions of the kidney and tumour in the videos (Figure 6.1(c-e)) and contrasted the performance of our method with and without the motion prior. We also compared with the state-of-the-art video segmentation method proposed by [Bai and et al., 2009], implemented in the After Effect software (AE), and with the level-sets based ACWOE [Vese and Chan, 2002] as considered in [Figueiredo et al., 2010, Figueiredo et al., 2012] for endoscopic video segmentations. In this experiment, we used the constant velocity model as our motion prior. Note that there is no explicit limitation on our motion prior term and more complicated motion priors can be seamlessly deployed into our framework, e.g. using the surgical robot's API signals. To compare our method with AE, we provided AE with an accurate segmentation for the first frame and, since this software is only able to segment a single object in a single view image, we used it multiple times to segment the kidney and tumours in the left and right view channels. Figure 6.2(b) illustrates the DSC vs. time when using ACWOE, AE and our method with and without motion prior for a phantom case. The dips in Figure 6.2(b) are caused by the occluding objects. The figure demonstrates how incorporating a motion prior stabilizes the segmentation and pose tracking and helps overcome large tissue occlusions that may occur in minimally invasive surgery. Obviously, AE and ACWOE are not able to show the internal tissues, e.g. the endophytic tumours shown in blue in Figure 6.1(e), as they do not use any preoperative information. They also are both fragile when a large occlusion occurs. Figure 6.2(d) compares the results of ACWOE, AE and our method with/without motion prior for all the *ex vivo* phantom datasets during the occlusion periods.

In our **third experiment** we tested our method on three different clinical cases of partial nephrectomy. For each patient, we prepared a ∼20-second stereo endoscopic video from the surgical system with a frame rate of 30 FPS, i.e. ∼600 frames. Segmenting CT and preparing the ground truth was performed similar to the *ex vivo* phantom data. Each stereo video took ∼3.5 hours to segment semi-automatically using AE to create the ground truth. Quantitative and qualitative results on the real *in vivo* cases are illustrated in Figure 6.3. Despite existing clutter and tool crossings, our method was able to achieve a DSC close to 0.85 on these challenging real *in vivo* cases with an average runtime of 0.045 seconds per frame using non-optimized MATLAB code on a single core 3.40 GHz CPU.

6.4 Chapter summary

Leveraging both preoperative data as well as endoscopic visual cues, we proposed a novel formulation with closed form solution for segmenting multiple tissues in multi-view en-



(a)

(b)

Figure 6.3: Clinical partial nephrectomy results. (a) DSC for three clinical cases. (b) Qualitative results of (1^{st} column) ACWOE (blue arrows indicate errors), (2^{nd} column) our method without motion prior and, (3^{rd} column) our method with motion prior. Green: kidney; Red: tumour; Yellow: ground truth.

doscopic videos. Our formulation further incorporated a motion prior in our optimization framework to stabilize the segmentation and pose tracking. One shortcoming of our method is the use of statistical deformable model which might not represent patient-specific tissue deformations. In addition, sudden camera motion will cause our tracking to fail. However, in cases where access to the surgical robots' API is available, one can easily feed in the camera motion signals into our formulation, as a camera motion constraint, which would enable handling the complex camera motion.

Chapter 7

Conclusion and Future Work

7.1 Thesis summary

Existence of noise, low contrast and objects' complexity in medical images are main reasons to preclude ideal segmentation. Incorporating prior expert knowledge into segmentation frameworks has proven useful for obtaining more accurate and plausible results. In this thesis, we proposed several techniques to bridge the gap between experts' and clinicians' knowledge and computer vision approaches by incorporating different prior information into segmentation frameworks. Incorporating such information (e.g. geometry and topology of objects) helps clinicians to segment targeted object solely and more accurately especially when training data is unavailable. In our proposed techniques, we moved toward improving both fidelity and optimizability in MIS and proposed new formulations to enforce useful topological and geometrical constraints by which a user can easily apply high level intuitive geometric constraints to improve segmentation results without the need for a deep understanding of how the method works.

In Chapter 3, unlike existing methods that have only considered simple structured (single-region) cells in overlapping cell segmentation problems, we adopted several prior information to build a faithful objective function unconcerned about its convexity to segment potentially overlapping multi-region cells with complex topology. In Chapter 4, we moved toward improving the space and time complexity and augmented the level sets framework with the ability to handle geometrical and different spatial distance constraints. In Chapter 5, we focused on the optimizability aspect of the segmentation task and improved our work on Chapter 4 by proposing a continuous convex formulation to augment the popular Mumford-Shah model. More specifically, we developed a new regularization term to incorporate similar geometrical and distance prior as our proposed method in Chapter 4 while maintaining global optimality. Our continuous methods, proposed in Chapters 4 and 5, compared favourably with analogous graph cuts-based methods in terms of metrication error and memory and time efficiency. In Chapter 6, we proposed a novel and efficient formulation with closed-form solution for segmenting multiple structures in multi-view endoscopic videos. We showed how to leverage both preoperative data as well as endoscopic visual cues to obtain more accurate and robust segmentations in the highly noisy, cluttered, and occasionally occluded environment of endoscopic video. We also encoded a motion prior in our optimization framework to stabilize the segmentation and pose tracking process. In cases where access to the surgical robots' API is available, we can easily feed in the camera motion signals into our formulation, as a camera motion constraint, which may increase the accuracy even further.

7.2 Discussion and future work

In this section, we summarize the take-home lessons from our experiments that can be helpful in guiding the methods and future research in the area of medical image segmentation.

Segmentation techniques are aimed at partitioning (crisply or fuzzily) an image into meaningful parts (two or more). Traditional segmentation approaches (e.g. thresholding, watershed, or region growing) proved incapable of robust and accurate segmentation due to noise, low contrast and complexity of objects in medical images. By incorporating prior knowledge of objects into rigorous optimization-based segmentation formulations, researchers developed more powerful techniques capable of segmenting specific (targeted) objects.

It is important to appreciate that, although incorporating richer prior into an objective function may increase the fidelity of the energy functional (by better modelling the underlying problem), this typically comes at the expense of complicating its optimization (lower optimizibility). On the other hand, focusing on optimizibility by simplifying the energies might decrease the fidelity of the energy functions. In other words, be wary of segmentation algorithms that always converge to the globally optimal but inaccurate solution, or ones that rely heavily on intricate initialization or meticulously tweaked parameters. Consequently, recent research surveyed has focused on developing methods that increase the optimizibility of energy functions (e.g. by proposing convex or submodular energy terms) without sacrificing the fidelity.

In addition to the optimizability-fidelty tradeoff that is impacted by the choice of priors, it is important to observe the runtime and memory efficiency of proposed medical image segmentation algorithms. For example, graph-based approaches may not be very efficient in handling very large images and they often produce artifacts like grid-bias errors (also known as metrication error) due to their discrete nature.

Despite the great advances that have been made in terms of increasing the fidelity and optimizibility of various segmentation energy models, there is still more to be done. We believe that through ongoing research, new methods will be proposed that allow for models that are faithful to the underlying problems, while being globally optimizable, memory- and time-efficient regardless of image size, and are free from any artifacts like metrication error.

In extending prior information in medical image segmentation, there are several directions to explore. One direction may focus on consolidating all of these previously mentioned priors such that a user (or an automatic system) can add one or more of these priors as a module to the segmentation task at hand. Such system is expected to minimize user inputs like manual initialization. Although many efforts have been made to convexify energy terms, many priors (especially when combined together) are non-convex (non-submodular) and hard to optimize. As convex relaxation and convex optimization techniques are becoming popular recently, research emphasis that focuses on convexification of energy functions with as many priors as needed would be an important step toward automatic image segmentation.

In optimization-based segmentation that encodes a set of desired priors, it is important to consider how to combine their respective energy (or objective) terms. The most common approach for dealing with such a multi-objective optimization is to scalarize the energies (via a linear sum of terms). Aside from choosing which priors are relevant and which mathematical formulae encode them, how to learn and set the contribution weight of each term needs to be explored carefully especially when there is not enough training data. When large sets of training data are available, machine learning techniques have been used to discover a good set of weights that adapt to image class, weights that change per image, and spatially adaptive weights.

The priors we reviewed and introduced in this thesis have been specifically and carefully designed to address particular segmentation problem. Another potential complementary approach that is worthy of future exploration is to attempt to learn the priors (not only their weight in the objective function) from available training data.

Future research directions could also focus on combining the hand-crafted features with machine learning techniques in case of availability of training data. For example it makes sense to use expert knowledge when the training data is not available and increase the contribution of machine learning techniques as more data becomes available and/or expert knowledge is harder to collect.

Finally, we mentioned that segmenting a medical image allows for easier (and automatic) analyzing the data (e.g. measuring the volume of different structures). Whether we absolutely need to segment an image in order to analyze it, remains an open question. One may consider developing machine learning-based techniques that, for example, predict an object's volume directly from image features.

Bibliography

[Adams and Bischof, 1994] Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 16(6):641–647. 1

[Al-Kofahi et al., 2010] Al-Kofahi, Y., Lassoued, W., Lee, W., and Roysam, B. (2010). Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 57(4):841–852. 2

[Alahari et al., 2010] Alahari, K., Russell, C., and Torr, P. (2010). Efficient piecewise learning for conditional random fields. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 895–901. 14

[Ali et al., 2011] Ali et al., S. (2011). Adaptive energy selective active contour with shape priors for nuclear segmentation and gleason grading of prostate cancer. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 661–669. 3, 49

[Amir-Khalili et al., 2014] Amir-Khalili, A., Peyrat, J.-M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Hamarneh, G., and Abugharbieh, R. (2014). Auto localization and segmentation of occluded vessels in robot-assisted partial nephrectomy. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 407–414. Springer. 24

[Andrews et al., 2014] Andrews, S., Changizi, N., and Hamarneh, G. (2014). The isometric log-ratio transform for probabilistic multi-label anatomical shape representation. *IEEE Transactions on Medical Imaging (IEEE TMI)*. 16, 33

[Andrews and Hamarneh, 2015] Andrews, S. and Hamarneh, G. (2015). The generalized log-ratio transformation: Learning shape and adjacency priors for simultaneous thigh muscle segmentation. *IEEE Transactions on Medical Imaging (IEEE TMI)*. 16

[Andrews et al., 2011a] Andrews, S., Hamarneh, G., Yazdanpanah, A., HajGhanbari, B., and Reid, W. D. (2011a). Probabilistic multi-shape segmentation of knee extensor and flexor muscles. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 651–658. 15

[Andrews et al., 2011b] Andrews, S., McIntosh, C., and Hamarneh, G. (2011b). Convex multi-region probabilistic segmentation with shape prior in the isometric log-ratio transformation space. *IEEE International Conference on Computer Vision (IEEE ICCV)*. 14, 15, 18

[Angelini et al., 2004] Angelini, E. D., Song, T., Mensh, B. D., and Laine, A. (2004). Multiphase three-dimensional level set segmentation of brain mri. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 318–326. 59

[Appleton and Talbot, 2006] Appleton, B. and Talbot, H. (2006). Globally minimal surfaces by continuous maximal flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 28(1):106–118. 12

[Ashburner and Friston, 2000] Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometryŮthe methods. *Neuroimage*, 11(6):805–821. 1

[Ayed et al., 2008] Ayed, I. B., Li, S., Islam, A., Garvin, G., and Chhem, R. (2008). Area prior constrained level set evolution for medical image segmentation. In *SPIE Medical Imaging*, volume 6914, page 691402. 14, 18, 37

[Ayed et al., 2009] Ayed, I. B., Li, S., and Ross, I. (2009). A statistical overlap prior for variational image segmentation. *International Journal of Computer Vision (IJCV)*, 85(1):115–132. 14, 23

[Bae et al., 2011a] Bae, E., Yuan, J., and Tai, X.-C. (2011a). Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision (IJCV)*, 92(1):112–129. 15

[Bae et al., 2011b] Bae, E., Yuan, J., Tai, X.-C., and Boykov, Y. (2011b). A fast continuous max-flow approach to non-convex multilabeling problems. *Efficient Global Minimization Methods for Variational Problems in Imaging and Vision*. 31

[Bagci et al., 2013] Bagci, U., Udupa, J. K., Mendhiratta, N., Foster, B., Xu, Z., Yao, J., Chen, X., and Mollura, D. J. (2013). Joint segmentation of anatomical and functional images: Applications in quantification of lesions from pet, pet-ct, mri-pet, and mri-pet-ct images. *Medical Image Analysis (MedIA)*, 17(8):929–945. 1

[Bai et al., 2014] Bai, J., Miri, M. S., Liu, Y., Saha, P., Garvin, M., and Wu, X. (2014). Graph-based optimal multi-surface segmentation with a star-shaped prior: Application to the segmentation of the optic disc and cup. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 525–528. IEEE. 33

[Bai and et al., 2009] Bai, X. and et al. (2009). Video SnapCut: robust video object cutout using localized classifiers. *ACM Transactions on Graphics (TOG)*, 28(3):70:1–70:11. xviii, 108, 109

[Ballester et al., 2005] Ballester, M. Á. G., Linguraru, M. G., Aguirre, M. R., and Ayache, N. (2005). On the adequacy of principal factor analysis for the study of shape variability. In *Medical Imaging*, pages 1392–1399. International Society for Optics and Photonics. 34, 35

[Barrett and Mortensen, 1997] Barrett, W. A. and Mortensen, E. N. (1997). Interactive live-wire boundary extraction. *Medical Image Analysis (MedIA)*, 1(4):331–341. 20

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *European Conference on Computer Vision (ECCV)*, pages 404–417. Springer. 24

[Ben Ayed et al., 2009] Ben Ayed, I., Li, S., and Ross, I. (2009). Embedding overlap priors in variational left ventricle tracking. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 28(12):1902–1913. 23

[Ben Ayed and Mitiche, 2008] Ben Ayed, I. and Mitiche, A. (2008). A region merging prior for variational level set image segmentation. *IEEE Transactions on Image Processing (IEEE TIP)*, 17(12):2301–2311. 46

[Ben-Zadok et al., 2009] Ben-Zadok, N., Riklin-Raviv, T., and Kiryati, N. (2009). Interactive level set segmentation for image-guided therapy. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1079–1082. IEEE. 19, 20

[Bergbauer et al., 2013] Bergbauer, J., Nieuwenhuis, C., Souiai, M., and Cremers, D. (2013). Morphological proximity priors: Spatial relationships for semantic segmentation. 18, 45

[Bernardis et al., 2011] Bernardis et al., E. (2011). Pop out many small structures from a very large microscopic image. *Medical Image Analysis (MedIA)*, 15(5):690–707. 3, 49, 55, 56

[Bertrand, 1994] Bertrand, G. (1994). Simple points, topological numbers and geodesic neighborhoods in cubic grids. *Pattern Recognition Letters*, 15(10):1003–1011. 36

[Bigün et al., 1991] Bigün, J., Granlund, G. H., and Wiklund, J. (1991). Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI*, 13(8):775–790. 23

[Bloch, 2005] Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89–110. 39

[Bosch et al., 2007] Bosch, A., Zisserman, A., and Munoz, X. (2007). Image classification using random forests and ferns. 24

[Bowden et al., 2000] Bowden, R., Mitchell, T. A., and Sarhadi, M. (2000). Non-linear statistical models for the 3d reconstruction of human pose and motion from monocular image sequences. *Image and Vision Computing*, 18(9):729–737. 35

[Boykov and Funka-Lea, 2006] Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision (IJCV)*, 70(2):109–131. 27

[Boykov and Kolmogorov, 2003] Boykov, Y. and Kolmogorov, V. (2003). Computing geodesics and minimal surfaces via graph cuts. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 26–33. 11, 71

[Boykov et al., 1998] Boykov, Y., Veksler, O., and Zabih, R. (1998). Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 648–655. IEEE. 11

[Boykov et al., 2001] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 23(11):1222–1239. 12, 15, 29, 41

[Bresson et al., 2006] Bresson, X., Vandergheynst, P., and Thiran, J.-P. (2006). A variational model for object segmentation using boundary information and shape prior driven by the mumford-shah functional. *International Journal of Computer Vision (IJCV)*, 68:145–162. 14

[Brown et al., 2009] Brown, E., Chan, T., and Bresson, X. (2009). Convex formulation and exact global solutions for multi-phase piecewise constant Mumford-Shah image segmentation. *UCLA CAM Report*, pages 09–66. 31

[Brox and Weickert, 2006] Brox, T. and Weickert, J. (2006). Level set segmentation with multiple regions. *IEEE Transactions on Image Processing (IEEE TIP)*, 15(10):3213–3218. 18, 46

[Bueno et al., 2004] Bueno, G., Martínez-Albalá, A., and Adán, A. (2004). Fuzzy-snake segmentation of anatomical structures applied to ct images. In *Image Analysis and Recognition*, pages 33–42. Springer. 15

[Bullitt et al., 2003] Bullitt, E., Gerig, G., Pizer, S. M., Lin, W., and Aylward, S. R. (2003). Measuring tortuosity of the intracerebral vasculature from mra images. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 22(9):1163–1171. 1

[Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, (6):679–698. 27

[Carneiro et al., 2008] Carneiro, G., Georgescu, B., Good, S., and Comaniciu, D. (2008). Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 27(9):1342–1355. 1

[Caselles et al., 1997] Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision (IJCV)*, 22(1):61–79. 20, 27

[Casta et al., 2009] Casta, C., Clarysse, P., Schaerer, J., and Pousin, J. (2009). Evaluation of the dynamic deformable elastic template model for the segmentation of the heart in mri sequences. *The MIDAS Journal, Cardiac MR Left Ventricle Segmentation Challenge*. 77

[Celebi et al., 2009] Celebi, M. E., Iyatomi, H., Schaefer, G., and Stoecker, W. V. (2009). Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics*, 33(2):148–153. 22

[Centre, 2009] Centre, S. H. S. (2009). Cardiac MR left ventricle segmentation challenge. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 74

[Chambolle, 2004] Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97. 95

[Chambolle, 2005] Chambolle, A. (2005). Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 136–152. 92

[Chambolle et al., 2008] Chambolle, A., Cremers, D., and Pock, T. (2008). A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany. 14, 15

[Chan et al.,] Chan, T., Golub, G., and Mulet, P. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*. 92

[Chan et al., 2000] Chan, T., Sandberg, B., and Vese, L. (2000). Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2):130–141. 24, 60, 88

[Chan and Vese, 2001] Chan, T. and Vese, L. (2001). Active contours without edges. *IEEE Transaction son Image Processing (IEEE TIP)*, 10(2):266–277. 28, 64, 66

[Chan et al., 2006] Chan, T. F., Esedoglu, S., and Nikolova, M. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648. 13

[Changizi and Hamarneh, 2010] Changizi, N. and Hamarneh, G. (2010). Probabilistic multi-shape representation using an isometric log-ratio mapping. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 563–570. 15, 16

[Chen et al., 2002] Chen, Y., Tagare, H. D., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K. S., Briggs, R. W., and Geiser, E. A. (2002). Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328. 18, 31

[Cheng et al., 2010] Cheng et al., J. (2010). Identifying cells in histopath. images. *Recognizing Patterns in Signals, Speech, Images and Videos - ICPR*, pages 244–252. 55

[Cheng et al., 2011] Cheng et al., L. (2011). Discriminative segmentation of microscopic cellular images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 637–644. 3, 49

[Chodorowski et al., 2005] Chodorowski, A., Mattsson, U., Langille, M., and Hamarneh, G. (2005). Color lesion boundary detection using live wire. In *Medical Imaging*, pages 1589–1596. International Society for Optics and Photonics. 27

[Chung and Vese, 2009] Chung, G. and Vese, L. A. (2009). Image segmentation using a multilayer level-set approach. *Computing and Visualization in Science*, 12(6):267–285. xiii, 25, 30

[Collins and Evans, 1997] Collins, D. and Evans, A. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(8):1271–1294. 18, 47

[Collins et al., 1995] Collins, D., Holmes, C., Peters, T., and Evans, A. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208. 33, 47

[Colliot et al., 2006] Colliot, O., Camara, O., and Bloch, I. (2006). Integration of fuzzy spatial relations in deformable models–application to brain MRI segmentation. *Pattern Recognition*, 39(8):1401–1414. 14

[Constantinides et al., 2009] Constantinides, C., Chenoune, Y., Kachenoura, N., Roullot, E., Mousseaux, E., Herment, A., and Frouin, F. (2009). Semi-automated cardiac segmentation on cine magnetic resonance images using gvf-snake deformable models. *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*. 77

[Cootes et al., 2001] Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 23(6):681–685. 24

[Cootes and Taylor, 1995] Cootes, T. F. and Taylor, C. J. (1995). Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5):403–409. 18, 36, 50

[Cootes and Taylor, 1999] Cootes, T. F. and Taylor, C. J. (1999). A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567–573. 34

[Cootes et al., 1995] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding (CVIU)*, 61(1):38–59. 18, 33

[Couprie et al., 2011] Couprie, C., Grady, L., Talbot, H., and Najman, L. (2011). Combinatorial continuous maximum flow. *SIAM Journal on Imaging Sciences*, 4(3):905–930. xii, 10, 12

[Courant et al., 1967] Courant, R., Friedrichs, K., and Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2):215–234. 10, 67

[Cremers et al., 2007] Cremers, D., Fluck, O., Rousson, M., and Aharon, S. (2007). A probabilistic level set formulation for interactive organ segmentation. In *Medical Imaging*, pages 65120V–65120V. International Society for Optics and Photonics. 19

[Cremers et al., 2006] Cremers, D., Osher, S. J., and Soatto, S. (2006). Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision (IJCV)*, 69(3):335–351. 35

[Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, volume 1, pages 886–893. IEEE. 24

[Davatzikos et al., 2003] Davatzikos, C., Tao, X., and Shen, D. (2003). Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 22(3):414–423. 34

[Delong and Boykov, 2008] Delong, A. and Boykov, Y. (2008). A scalable graph-cut algorithm for nd grids. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. IEEE. 11

[Delong and Boykov, 2009] Delong, A. and Boykov, Y. (2009). Globally optimal segmentation of multi-region objects. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 285–292. xi, xiii, xiv, xvii, 4, 18, 40, 41, 42, 43, 56, 58, 61, 67, 68, 83, 84, 86, 97, 98, 99, 100, 101

[Delong et al., 2012a] Delong, A., Osokin, A., Isack, H. N., and Boykov, Y. (2012a). Fast approximate energy minimization with label costs. *International Journal of Computer Vision (IJCV)*, 96(1):1–27. 18, 46

[Delong et al., 2012b] Delong, A., Veksler, O., Osokin, A., and Boykov, Y. (2012b). Minimizing sparse high-order energies by submodular vertex-cover. In *Advances in Neural Information Processing Systems (NIPS)*, pages 971–979. 26

[Dhandra and et al., 2006] Dhandra, B. and et al. (2006). Analysis of abnormality in endoscopic images using combined HSI color space and watershed segmentation. In *IEEE International Conference on Pattern Recognition (ICPR)*, volume 4, pages 695–698. 103

[Dobbins et al., 1987] Dobbins, A., Zucker, S. W., and Cynader, M. S. (1987). End-stopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438–441. 26

[dos Santos Gromicho, 1998] dos Santos Gromicho, J. A. (1998). *Quasiconvex optimization and location theory*. Number 9. Springer. 13

[Duchenne et al., 2011] Duchenne, O., Bach, F., Kweon, I.-S., and Ponce, J. (2011). A tensor-based algorithm for high-order graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 33(12):2383–2395. 25

[Dydenko et al., 2006] Dydenko, I., Jamal, F., Bernard, O., DŠhooge, J., Magnin, I. E., and Friboulet, D. (2006). A level set framework with a shape and motion prior for segmentation and region tracking in echocardiography. *Medical Image Analysis (MedIA)*, 10(2):162–177. 46

[El-Zehiry and Grady, 2010] El-Zehiry, N. Y. and Grady, L. (2010). Fast global optimization of curvature. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 3257–3264. IEEE. 26

[Esneault et al., 2010] Esneault, S., Lafon, C., and Dillenseger, J.-L. (2010). Liver vessels segmentation using a hybrid geometrical moments/graph cuts method. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 57(2):276–283. 23

[Estépar and Vosburgh, 2014] Estépar, R. and Vosburgh, K. (2014). Multimodality guidance in endoscopic and laparoscopic abdominal procedures. In *Intraoperative Imaging and Image-Guided Therapy*, pages 767–778. 102

[Feddern et al., 2003] Feddern, C., Weickert, J., and Burgeth, B. (2003). Level-set methods for tensor-valued images. In *IEEE Workshop on Geometric and Level Set Methods in Computer Vision*, pages 65–72. 22, 24

[Felzenszwalb and Veksler, 2010] Felzenszwalb and Veksler (2010). Tiered scene labeling with dynamic programming. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 3097–3104. 14, 18, 44

[Figueiredo et al., 2010] Figueiredo, I. N., Figueiredo, P. N., Stadler, G., Ghattas, O., and Araujo, A. (2010). Variational image segmentation for endoscopic human colonic aberrant crypt foci. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 29(4):998–1011. 22, 103, 109

[Figueiredo et al., 2012] Figueiredo, I. N., Moreno, J. C., Prasath, V. S., and Figueiredo, P. N. (2012). A segmentation model and application to endoscopic images. In *Image Analysis and Recognition*, pages 164–171. 103, 109

[Fleming and Rishel, 1960] Fleming, W. and Rishel, R. (1960). An integral formula for total gradient variation. *Archiv der Mathematik*, 11(1):218–222. 91

[Fodor, 2002] Fodor, I. K. (2002). A survey of dimension reduction techniques. 35

[Foulonneau et al., 2006] Foulonneau, A., Charbonnier, P., and Heitz, F. (2006). Affine-invariant geometric shape priors for region-based active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 28(8):1352–1357. 18, 38

[Frangi et al., 1998] Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 130–137. Springer. 23, 27

[Freedman and Zhang, 2005] Freedman, D. and Zhang, T. (2005). Interactive graph cut based segmentation with shape priors. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 755–762. 32

[Freiman et al., 2009] Freiman, M., Joskowicz, L., and Sosna, J. (2009). A variational method for vessels segmentation: algorithm and application to liver vessels visualization. In *SPIE Medical Imaging*, pages 72610H–72610H. International Society for Optics and Photonics. 23

[Gee et al., 1993] Gee, J., Reivich, M., and Bajcsy, R. (1993). Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225. 47

[Gennert and Yuille, 1988] Gennert, M. A. and Yuille, A. L. (1988). Determining the optimal weights in multiple objective function optimization. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 87–89. Citeseer. 10

[Gloger et al., 2012] Gloger, O., Toennies, K. D., Liebscher, V., Kugelmann, B., Laqua, R., and Volzke, H. (2012). Prior shape level set segmentation on multistep generated probability maps of mr datasets for fully automatic kidney parenchyma volumetry. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 31(2):312–325. 14

[Goldenberg et al., 2001] Goldenberg, R., Kimmel, R., Rivlin, E., and Rudzsky, M. (2001). Fast geodesic active contours. *IEEE Transactions on Image Processing (IEEE TIP)*, 10(10):1467–1475. 20

[Goldenberg et al., 2002] Goldenberg, R., Kimmel, R., Rivlin, E., and Rudzsky, M. (2002). Cortex segmentation: A fast variational geometric approach. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 21(12):1544–1551. 18, 39, 41, 42, 61

[Goldschlager et al., 1982] Goldschlager, L. M., Shaw, R. A., and Staples, J. (1982). The maximum flow problem is log space complete for p. *Theoretical Computer Science*, 21(1):105–111. 11

[Gould et al., 2008] Gould, S., Rodgers, J., Cohen, D., Elidan, G., and Koller, D. (2008). Multi-class segmentation with relative location prior. *International Journal of Computer Vision (IJCV)*, 80(3):300–316. 14

[Grady, 2006] Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 28(11):1768–1783. 15, 27

[Grady, 2012] Grady, L. (2012). Targeted image segmentation using graph methods. *Image Processing and Analysis with Graphs*. 27, 28, 36

[Grady et al., 2011] Grady, L., Jolly, M.-P., and Seitz, A. (2011). Segmentation from a box. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 367–374. IEEE. 20

[Graf et al., 2010] Graf et al., F. (2010). Counting lymphocytes in histopathology images using connected components. *Recognizing Patterns in Signals, Speech, Images and Videos - ICPR*, pages 263–269. 55

[Grau et al., 2004] Grau, V., Mewes, A., Alcaniz, M., Kikinis, R., and Warfield, S. K. (2004). Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 23(4):447–458. 1

[Greig et al., 1989] Greig, D., Porteous, B., and Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279. 13

[Gurcan et al., 2010] Gurcan et al., M. (2010). Pattern recognition in histopathological images: An ICPR 2010 contest. *Recognizing Patterns in Signals, Speech, Images and Videos - ICPR*, pages 226–234. xiii, 56

[Hamarneh, 2011] Hamarneh, G. (2011). The optimizability-fidelty trade-off in image analysis. Technical Report SFU-Summit-10897, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. 14

[Hamarneh et al., 2004] Hamarneh, G., Abugharbieh, R., and McInerney, T. (2004). Medial profiles for modeling deformation and statistical analysis of shape and their use in medical image segmentation. *International Journal of Shape Modeling*, 10(2):187–210. 34

[Hamarneh and Gustavsson, 2000] Hamarneh, G. and Gustavsson, T. (2000). Statistically constrained snake deformations. In *EEE International Conference on Systems, Man, and Cybernetics*, volume 3, pages 1610–1615. IEEE. 34

[Hamarneh et al., 2008] Hamarneh, G., Jassi, P., and Tang, L. (2008). Simulation of ground-truth validation data via physically-and statistically-based warps. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 459–467. 36, 50, 106, 107

[Hamarneh and Li, 2009] Hamarneh, G. and Li, X. (2009). Watershed segmentation using prior shape and appearance knowledge. *Image and Vision Computing*, 27(1):59–68. 1

[Hamarneh et al., 2009] Hamarneh, G., McIntosh, C., McInerney, T., and Terzopoulos, D. (2009). Deformable organisms: An artificial life framework for automated medical image analysis. *Computational Intelligence in Medical Imaging: Techniques and Applications*, page 433. 43

[Hamarneh et al., 2005] Hamarneh, G., Yang, J., McIntosh, C., and Langille, M. (2005). 3d live-wire-based semi-automatic segmentation of medical images. In *Medical Imaging*, pages 1597–1603. International Society for Optics and Photonics. 20

[Han et al., 2009] Han, S., Tao, W., Wang, D., Tai, X.-C., and Wu, X. (2009). Image segmentation based on grabcut framework integrating multiscale nonlinear structure tensor. *IEEE Transactions on Image Processing (IEEE TIP)*, 18(10):2289–2302. 24

[Han et al., 2003] Han, X., Xu, C., and Prince, J. L. (2003). A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 25(6):755–768. xiii, 18, 36, 37

[Hatt et al., 2009] Hatt, M., Cheze le Rest, C., Turzo, A., Roux, C., and Visvikis, D. (2009). A fuzzy locally adaptive bayesian segmentation approach for volume determination in pet. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 28(6):881–893. 1

[Heikkilä et al., 2009] Heikkilä, M., Pietikäinen, M., and Schmid, C. (2009). Description of interest regions with local binary patterns. *Pattern recognition*, 42(3):425–436. 24

[Heimann and Meinzer, 2009] Heimann, T. and Meinzer, H.-P. (2009). Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis (MedIA)*, 13(4):543–563. 35

[Hill et al., 2001] Hill, D. L., Batchelor, P. G., Holden, M., and Hawkes, D. J. (2001). Medical image registration. *Physics in Medicine and Biology*, 46(3):R1. 47

[Hospital, 2012] Hospital, R. U. (2012). RV segmentation challenge in cardiac MRI, http://www.litislab.eu/rvsc. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 74

[Howing et al., 1997] Howing, F., Wermser, D., and Dooley, L. (1997). Fuzzy snakes. In *International Conference on Image Processing and Its Applications*, volume 2, pages 627–630. IET. 15

[Huang et al., 2009] Huang, S., Liu, J., Lee, L., Venkatesh, S., Teo, L., Au, C., and Nowinski, W. (2009). Segmentation of the left ventricle from cine MR images using a comprehensive approach. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 339–347. 77

[Huang et al., 2005] Huang, X., Qian, Z., Huang, R., and Metaxas, D. (2005). Deformable-model based textured object segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 119–135. Springer. 23, 24

[Iosifescu et al., 1997] Iosifescu, D., Shenton, M., Warfield, S., Kikinis, R., Dengler, J., Jolesz, F., and McCarley, R. (1997). An automated registration algorithm for measuring MRI subcortical brain structures. *Neuroimage*, 6(1):13–26. 18, 47

[Ishikawa, 2003] Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 25(10):1333–1336. 13, 29, 90

[Jolly, 2009] Jolly, M. (2009). Fully automatic left ventricle segmentation in cardiac cine mr images using registration and minimum surfaces. *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*. 77

[Kadir and Brady, 2003] Kadir, T. and Brady, M. (2003). Unsupervised non-parametric region segmentation using level sets. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 1267–1274. IEEE. 46

[Karaolani et al., 1989] Karaolani, P., Sullivan, G. D., Baker, K. D., and Baines, M. (1989). A finite element method for deformable models. In *Alvey Vision Conference*, pages 1–6. Citeseer. 36

[Kass et al., 1988] Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision (IJCV)*, 1(4):321–331. 19, 26, 35

[Kawahara et al., 2013] Kawahara, J., McIntosh, C., Tam, R., and Hamarneh, G. (2013). Augmenting auto-context with global geometric features for spinal cord segmentation. In *Machine Learning in Medical Imaging*, pages 211–218. Springer. 35

[Kim et al., 2007] Kim, J., Çetin, M., and Willsky, A. S. (2007). Nonparametric shape priors for active contour-based image segmentation. *Signal Processing*, 87(12):3021–3044. 35

[Klodt and Cremers, 2011] Klodt, M. and Cremers, D. (2011). A convex framework for image segmentation with moment constraints. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 2236–2243. IEEE. xiii, 18, 38, 39

[Koerkamp et al., 2010] Koerkamp, B. G., Weinstein, M. C., Stijnen, T., Heijenbrok-Kal, M. H., and Hunink, M. M. (2010). Uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making*. 15

[Kolmogorov et al., 2007] Kolmogorov, V., Boykov, Y., and Rother, C. (2007). Applications of parametric maxflow in computer vision. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 1–8. 14

[Kolmogorov and Rother, 2007] Kolmogorov, V. and Rother, C. (2007). Minimizing nonsubmodular functions with graph cuts-a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 29(7):1274–1279. 41

[Kolmogorov and Zabin, 2004] Kolmogorov, V. and Zabin, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 26(2):147–159. 13

[Koss et al., 1999] Koss, J. E., Newman, F., Johnson, T., and Kirch, D. (1999). Abdominal organ segmentation using texture transforms and a hopfield neural network. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 18(7):640–648. 23

[Kuse et al., 2010] Kuse et al., M. (2010). A classification scheme for lymphocyte segmentation in H&E stained histology images. *Recognizing Patterns in Signals, Speech, Images and Videos - ICPR*, pages 235–243. 55

[Kuse et al., 2011] Kuse et al., M. (2011). Local isotropic phase symmetry measure for detection of beta cells and lymphocytes. *Journal of Pathology Informatics*, 2. 55

[Lellmann et al., 2009] Lellmann, J., Kappes, J., Yuan, J., Becker, F., and Schnörr, C. (2009). Convex multi-class image labeling by simplex-constrained total variation. In *Scale Space and Variational Methods in Computer Vision*, pages 150–162. 15

[Lempitsky et al., 2009] Lempitsky, V., Kohli, P., Rother, C., and Sharp, T. (2009). Image segmentation with a bounding box prior. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 277–284. IEEE. 20

[Lenglet et al., 2004] Lenglet, C., Rousson, M., and Deriche, R. (2004). Segmentation of 3d probability density fields by surface evolution: Application to diffusion mri. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 18–25. Springer. 23

[Leventon et al., 2000a] Leventon, M. E., Faugeras, O., Grimson, W. E. L., and Wells, W. M. (2000a). Level set based segmentation with intensity and curvature priors. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 4–11. IEEE. 26

[Leventon et al., 2000b] Leventon, M. E., Grimson, W. E. L., and Faugeras, O. (2000b). Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, volume 1, pages 316–323. IEEE. 33

[Li et al., 2006a] Li, K., Wu, X., Chen, D., and Sonka, M. (2006a). Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 28(1):119–134. 18, 40, 41, 61

[Li et al., 2006b] Li, S., Fevens, T., Krzyżak, A., Jin, C., and Li, S. (2006b). Fast and robust clinical triple-region image segmentation using one level set function. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 766–773. 59

[Lim et al., 2011] Lim, Y., Jung, K., and Kohli, P. (2011). Constrained discrete optimization via dual space search. 18, 39

[Liu and Sun, 2010] Liu, J. and Sun, J. (2010). Parallel graph-cuts by adaptive bottom-up merging. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 2181–2188. IEEE. 11

[Liu et al., 2008] Liu, X., Veksler, O., and Samarabandu, J. (2008). Graph cut with ordering constraints on labels and its applications. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. 14, 18, 43, 44

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110. 24

[Lu et al., 2012] Lu, C., Chelikani, S., Jaffray, D. A., Milosevic, M. F., Staib, L. H., and Duncan, J. S. (2012). Simultaneous nonrigid registration, segmentation, and tumor detection in mri guided cervical cancer radiation therapy. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 31(6):1213–1227. 35

[Lu et al., 2009] Lu, Y., Radau, P., Connelly, K., Dick, A., and Wright, G. (2009). Automatic image-driven segmentation of left ventricle in cardiac cine MRI. *The MIDAS Journal*, 49. 77

[Mairal et al., 2008] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2008). Discriminative learned dictionaries for local image analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. 24

[Malcolm et al., 2007] Malcolm, J., Rathi, Y., and Tannenbaum, A. (2007). A graph cut approach to image segmentation in tensor space. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. IEEE. 23, 24

[Mansouri et al., 2006] Mansouri, A., Mitiche, A., and Vázquez, C. (2006). Multiregion competition: A level set extension of region competition to multiple region image partitioning. *Computer Vision and Image Understanding (CVIU)*, 101(3):137–150. xiii, 14, 30

[Mao et al., 2006] Mao, K., Zhao, P., and Tan, P. (2006). Supervised learning-based cell image segmentation for p53 immunohistochemistry. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 53(6):1153–1163. 2, 3, 49

[Marak et al., 2009] Marak, L., Cousty, J., Najman, L., Talbot, H., et al. (2009). 4D Morphological segmentation and the MICCAI LV-segmentation grand challenge. In *MICCAI Workshop on Cardiac MR Left Ventricle Segmentation Challenge. MIDAS Journal.* 77

[Martin Bland and Altman, 1986] Martin Bland, J. and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310. 76

[McIntosh and Hamarneh, 2007] McIntosh, C. and Hamarneh, G. (2007). Is a single energy functional sufficient? adaptive energy functionals and automatic initialization. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 503–510. 10, 14

[McIntosh and Hamarneh, 2012] McIntosh, C. and Hamarneh, G. (2012). Medial-based deformable models in nonconvex shape-spaces for medical image segmentation. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 31(1):33–50. 2, 14

[Mewes et al., 2011] Mewes, P., Neumann, D., Licegevic, O., Simon, J., Juloski, A. L., and Angelopoulou, E. (2011). Automatic region-of-interest segmentation and pathology detection in magnetically guided capsule endoscopy. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 141–148. 103

[Mikolajczyk and Schmid, 2005] Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 27(10):1615–1630. 24

[Mirzaalian and Hamarneh, 2010] Mirzaalian, H. and Hamarneh, G. (2010). Vessel scale-selection using mrf optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 3273–3279. IEEE. 24

[Mirzaei et al., 2013] Mirzaei, H., Tang, L., Werner, R., and Hamarneh, G. (2013). Decision forests with spatio-temporal features for graph-based tumor segmentation in 4d lung ct. In *Machine Learning in Medical Imaging*, pages 179–186. Springer. 24

[Mitiche and Ayed, 2010] Mitiche, A. and Ayed, I. B. (2010). *Variational and level set methods in image segmentation*, volume 5. Springer. 59

[Mountney and Yang, 2010] Mountney, P. and Yang, G. (2010). Motion compensated SLAM for image guided surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 496–504. 103, 109

[Mumford and Shah, 1989] Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685. 4, 25, 28, 85, 86

[Nain et al., 2006] Nain, D., Haker, S., Bobick, A., and Tannenbaum, A. (2006). Shape-driven 3d segmentation using spherical wavelets. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 66–74. Springer. 34

[Nambakhsh et al., 2013] Nambakhsh, C., Yuan, J., Punithakumar, K., Goela, A., Rajchl, M., Peters, T., and Ayed, I. B. (2013). Left ventricle segmentation in MRI via convex relaxed distribution matching. *Medical Image Analysis (MedIA)*, 17(8):1010–1024. 35, 40

[Nand et al., 2011] Nand, K. K., Abugharbieh, R., Booth, B. G., and Hamarneh, G. (2011). Detecting structure in diffusion tensor mr images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 90–97. Springer. 27

[Nastar and Ayache, 1993] Nastar, C. and Ayache, N. (1993). Non-rigid motion analysis in medical images: a physically based approach. In *Information Processing in Medical Imaging (IPMI)*, pages 17–32. Springer. 36

[Nayak et al., 2013] Nayak, N., Chang, H., Borowsky, A., Spellman, P., and Parvin, B. (2013). Classification of tumor histopathology via sparse feature learning. In *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 410–413. 24

[Nieuwenhuis et al., 2014] Nieuwenhuis, C., Hawe, S., Kleinsteuber, M., and Cremers, D. (2014). Co-sparse textural similarity for interactive segmentation. In *European Conference on Computer Vision (ECCV)*, pages 285–301. 24

[Nieuwenhuis et al., 2013] Nieuwenhuis, C., Töppe, E., and Cremers, D. (2013). A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *International Journal of Computer Vision (IJCV)*, 104(3):223–240. 9, 11, 12

[Noble and Boukerroui, 2006] Noble, J. A. and Boukerroui, D. (2006). Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 25(8):987–1010. 22

[Noseworthy et al., 2008] Noseworthy, P., Newton-Cheh, C., Schoen, F., Naik, M., Kim, D., O'Brien, F., Axel, L., and Srichai, M. (2008). Contemporary reviews in cardiovascular medicine. *Circulation*, 118(18). 77

[Nosrati and Hamarneh, 2014] Nosrati, M. and Hamarneh, G. (2014). Local optimization based segmentation of spatially-recurring, multi-region objects with part configuration constraints. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 33(9):1845–1859. 3

[Nosrati et al., 2015] Nosrati, M. S., Abugharbieh, R., Peyrat, J.-M., Abinahed, J., Al-Alao, O., Al-Ansari, A., and Hamarneh, G. (2015). Simultaneous multi-structure segmentation and 3d non-rigid pose estimation in image-guided robotic surgery. *IEEE Transactions on Medical Imaging (IEEE TMI)*. 4

[Nosrati et al., 2013] Nosrati, M. S., Andrews, S., and Hamarneh, G. (2013). Bounded labeling function for global segmentation of multi-part objects with geometric constraints. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 2032–2039. IEEE. 3

[Nosrati and Hamarneh, 2013] Nosrati, M. S. and Hamarneh, G. (2013). Segmentation of cells with partial occlusion and part configuration constraint using evolutionary computation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 461–468. Springer. 3

[Nosrati and Hamarneh, 2015] Nosrati, M. S. and Hamarneh, G. (2015). Segmentation of overlapping cervical cells: a variational method with star-shape prior. In *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 1–4. Springer. 3

[Nosrati et al., 2014] Nosrati, M. S., Peyrat, J.-M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Abugharbieh, R., and Hamarneh, G. (2014). Efficient multi-organ segmentation in multi-view endoscopic videos using pre-operative priors. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 324–331. Springer. 4

[Nowozin et al., 2010] Nowozin, S., Gehler, P., and Lampert, C. (2010). On parameter learning in crf-based approaches to object class image segmentation. *European Conference on Computer Vision (ECCV)*, pages 98–111. 14

[OŠBrien et al., 2009] OŠBrien, S., Ghita, O., and Whelan, P. (2009). Segmenting the left ventricle in 3d using a coupled asm and a learned non-rigid spatial model. *The MIDAS Journal*, 49. 77

[Oliva and Torralba, 2001] Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision (IJCV)*, 42(3):145–175. 24

[Orderud et al., 2007] Orderud, F., Hansgård, J., and Rabben, S. I. (2007). Real-time tracking of the left ventricle in 3d echocardiography using a state estimation approach. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 858–865. Springer. 46

[Otsu, 1975] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27. 1

[Pan and Lu, 2007] Pan, Z. and Lu, J. (2007). A bayes-based region-growing algorithm for medical image segmentation. *Computing in science and Engineering*, 9(4):32–38. 1

[Panagiotakis et al., 2010] Panagiotakis et al., C. (2010). Lymphocyte segmentation using the transferable belief model. *Recognizing Patterns in Signals, Speech, Images and Videos - ICPR*, pages 253–262. 55

[Paragios, 2002] Paragios, N. (2002). A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision (IJCV)*, 50(3):345–362. 18, 39, 41, 42, 59, 61

[Paragios, 2003] Paragios, N. (2003). User-aided boundary delineation through the propagation of implicit representations. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 678–686. Springer. 19

[Paragios and Deriche, 1999] Paragios, N. and Deriche, R. (1999). Geodesic active regions for motion estimation and tracking. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, volume 1, pages 688–694. IEEE. 46

[Paragios and Deriche, 2002] Paragios, N. and Deriche, R. (2002). Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision (IJCV)*, 46(3):223–247. 28

[Pentland and Sclaroff, 1991] Pentland, A. and Sclaroff, S. (1991). Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 13(7):715–729. 36

[Pizer et al., 2003] Pizer, S. M., Fletcher, P. T., Joshi, S., Thall, A., Chen, J. Z., Fridman, Y., Fritsch, D. S., Gash, A. G., Glotzer, J. M., Jiroutek, M. R., et al. (2003). Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision (IJCV)*, 55(2-3):85–106. 33

[Pluempitiwiriyawej et al., 2005] Pluempitiwiriyawej, C., Moura, J., Wu, Y., and Ho, C. (2005). Stacs: New active contour scheme for cardiac MR image segmentation. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 24(5):593–603. 14, 22, 31, 32

[Pock and Chambolle, 2011] Pock, T. and Chambolle, A. (2011). Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 1762–1769. 10

[Pock et al., 2008] Pock, T., Schoenemann, T., Graber, G., Bischof, H., and Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, pages 792–805. xvii, 14, 30, 31, 90, 91, 93, 96, 99, 100, 101

[Pohl et al., 2007] Pohl, K., Fisher, J., Bouix, S., Shenton, M., McCarley, R., Grimson, W., Kikinis, R., and Wells, W. (2007). Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis (MedIA)*, 11(5):465–477. 16

[Pohle and Toennies, 2001] Pohle, R. and Toennies, K. D. (2001). Segmentation of medical images using adaptive region growing. In *Medical Imaging*, pages 1337–1346. International Society for Optics and Photonics. 1

[Prasad et al., 2011] Prasad, G., Joshi, A. A., Feng, A., Barysheva, M., Mcmahon, K. L., De Zubicaray, G. I., Martin, N. G., Wright, M. J., Toga, A. W., Terzopoulos, D., et al.

(2011). Deformable organisms and error learning for brain segmentation. In *International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 135–147. 43

[Pratt et al., 2012] Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., and Yang, G.-Z. (2012). An effective visualisation and registration system for image-guided robotic partial nephrectomy. *Journal of Robotic Surgery*, 6(1):23–31. 1, 102

[Prisacariu and Reid, 2012] Prisacariu, V. and Reid, I. (2012). PWP3D: Real-time segmentation and tracking of 3d objects. *International Journal of Computer Vision (IJCV)*, 98(3):335–354. 18, 47, 48, 103

[Prisacariu et al., 2013] Prisacariu, V. A., Segal, A. V., and Reid, I. (2013). Simultaneous monocular 2D segmentation, 3D pose recovery and 3D reconstruction. In *Asian Conference on Computer Vision (ACCV)*, pages 593–606. 18, 47, 48

[Puerto-Souza and Mariottini, 2013] Puerto-Souza, G. and Mariottini, G. (2013). Toward long-term and accurate augmented-reality display for minimally-invasive surgery. In *IEEE International Conference on Robotics and Automation (IEEE ICRA)*, pages 5384–5389. 103, 109

[Rajchl et al., 2012] Rajchl, M., Yuan, J., White, J. A., Nambakhsh, C. M., Ukwatta, E., Li, F., Stirrat, J., and Peters, T. M. (2012). A fast convex optimization approach to segmenting 3D scar tissue from delayed-enhancement cardiac MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 659–666. 18, 40, 61

[Rak et al., 2013] Rak, M., König, T., and Toennies, K.-D. (2013). An adaptive subdivision scheme for quadratic programming in multi-label image segmentation. *Proceedings of the British Machine Vision Conference*. 14

[Rao et al., 2010] Rao, J., Abugharbieh, R., and Hamarneh, G. (2010). Adaptive regularization for image segmentation using local image curvature cues. In *European Conference on Computer Vision (ECCV)*, pages 651–665. Springer. 10

[Rother et al., 2009] Rother, C., Kohli, P., Feng, W., and Jia, J. (2009). Minimizing sparse higher order energy functions of discrete variables. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1382–1389. 14

[Rother et al., 2004] Rother, C., Kolmogorov, V., and Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 309–314. ACM. 20, 21

[Rother et al., 2007] Rother, C., Kolmogorov, V., Lempitsky, V., and Szummer, M. (2007). Optimizing binary mrfs via extended roof duality. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. IEEE. 41

[Rousson et al., 2003] Rousson, M., Brox, T., and Deriche, R. (2003). Active unsupervised texture segmentation on a diffusion based feature space. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, volume 2, pages II–699. IEEE. 23

[Rousson and Paragios, 2002] Rousson, M. and Paragios, N. (2002). Shape priors for level set representations. In *European Conference on Computer Vision (ECCV)*, pages 78–92. Springer. 18, 31

[Saad et al., 2010a] Saad, A., Hamarneh, G., and Moller, T. (2010a). Exploration and visualization of segmentation uncertainty using shape and appearance prior information. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1366–1375. 21

[Saad et al., 2008] Saad, A., Hamarneh, G., Möller, T., and Smith, B. (2008). Kinetic modeling based probabilistic segmentation for molecular images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 244–252. xvi, 15, 22, 24, 81

[Saad et al., 2010b] Saad, A., Möller, T., and Hamarneh, G. (2010b). ProbExplorer: Uncertainty-guided exploration and editing of probabilistic medical image segmentation. In *Computer Graphics Forum*, volume 29, pages 1113–1122. Wiley Online Library. 15, 21

[Sadeghi et al., 2009] Sadeghi, M., Tien, G., Hamarneh, G., and Atkins, M. S. (2009). Hands-free interactive image segmentation using eyegaze. In *SPIE Medical Imaging*, pages 72601H–72601H. International Society for Optics and Photonics. 19

[Sahoo et al., 1988] Sahoo, P. K., Soltani, S., and Wong, A. K. (1988). A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260. 1

[Samson et al., 2000] Samson, C., Blanc-Féraud, L., Aubert, G., and Zerubia, J. (2000). A level set model for image classification. *International Journal of Computer Vision (IJCV)*, 40(3):187–197. 18, 61

[Sandberg et al., 2002] Sandberg, B., Chan, T., and Vese, L. (2002). A level-set and gabor-based active contour algorithm for segmenting textured images. In *UCLA Department of Mathematics CAM report*. Citeseer. 28

[Sandhu et al., 2011] Sandhu, R., Dambreville, S., Yezzi, A., and Tannenbaum, A. (2011). A nonrigid kernel-based framework for 2D-3D pose estimation and 2D image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 33(6):1098–1115. 18, 47, 48, 103

[Santner et al., 2009] Santner, J., Unger, M., Pock, T., Leistner, C., Saffari, A., and Bischof, H. (2009). Interactive texture segmentation using random forests and total variation. In *BMVC*, pages 1–12. Citeseer. 23, 24

[Sapiro, 1997] Sapiro, G. (1997). Color snakes. *Computer Vision and Image Understanding (CVIU)*, 68(2):247–253. 27

[Schmidt and Boykov, 2012] Schmidt, F. R. and Boykov, Y. (2012). Hausdorff distance constraint for multi-surface segmentation. In *European Conference on Computer Vision (ECCV)*, pages 598–611. 18, 42, 61

[Schoenemann and Cremers, 2007] Schoenemann, T. and Cremers, D. (2007). Globally optimal image segmentation with an elastic shape prior. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 1–6. IEEE. 36

[Schoenemann et al., 2009] Schoenemann, T., Kahl, F., and Cremers, D. (2009). Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 17–23. IEEE. 26

[Schölkopf et al., 1998] Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319. 35

[Shekhovtsov and Hlaváč, 2013] Shekhovtsov, A. and Hlaváč, V. (2013). A distributed mincut/maxflow algorithm combining path augmentation and push-relabel. *International Journal of Computer Vision (IJCV)*, 104(3):315–342. 11

[Shen et al., 2011] Shen, T., Li, H., and Huang, X. (2011). Active volume models for medical image segmentation. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 30(3):774–791. 14

[Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 22(8):888–905. 27

[Shi et al., 2008] Shi, L., Funt, B., and Hamarneh, G. (2008). Quaternion color curvature. In *Color and Imaging Conference*, volume 2008, pages 338–341. Society for Imaging Science and Technology. 27

[Singaraju et al., 2008] Singaraju, D., Grady, L., and Vidal, R. (2008). Interactive image segmentation via minimization of quadratic energies on directed graphs. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. IEEE. 27

[Slabaugh and Unal, 2005] Slabaugh, G. and Unal, G. (2005). Graph cuts segmentation using an elliptical shape prior. In *IEEE International Conference on Image Processing (IEEE ICIP)*, volume 2, pages II–1222. IEEE. 18, 32

[Song et al., 2010] Song, Q., Wu, X., Liu, Y., Sonka, M., and Garvin, M. (2010). Simultaneous searching of globally optimal interacting surfaces with shape priors. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 2879–2886. IEEE. 18

[Sotiras et al., 2013] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 32(7):1153–1190. 47

[Staib and Duncan, 1992] Staib, L. H. and Duncan, J. S. (1992). Deformable fourier models for surface finding in 3-d images. In *Visualization in Biomedical Computing*, pages 90–104. International Society for Optics and Photonics. 34

[Strandmark and Kahl, 2010] Strandmark, P. and Kahl, F. (2010). Parallel and distributed graph cuts by dual decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 2085–2092. IEEE. 11

[Strandmark and Kahl, 2011] Strandmark, P. and Kahl, F. (2011). Curvature regularization for curves and surfaces in a global optimization framework. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 205–218. Springer. 26, 27

[Strandmark et al., 2013] Strandmark, P., Ulén, J., Kahl, F., and Grady, L. (2013). Shortest paths with curvature and torsion. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 2024–2031. IEEE. xii, 26, 27

[Strekalovskiy and Cremers, 2011] Strekalovskiy, E. and Cremers, D. (2011). Generalized ordering constraints for multilabel optimization. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, pages 2619–2626. IEEE. xiii, 18, 44

[Strekalovskiy et al., 2012] Strekalovskiy, E., Nieuwenhuis, C., and Cremers, D. (2012). Nonmetric priors for continuous multilabel optimization. In *European Conference on Computer Vision (ECCV)*, pages 208–221. Springer. 18, 43, 44, 45

[Su et al., 2009] Su, L.-M., Vagvolgyi, B. P., Agarwal, R., Reiley, C. E., Taylor, R. H., and Hager, G. D. (2009). Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration. *Urology*, 73(4):896–900. 1

[Szummer et al., 2008] Szummer, M., Kohli, P., and Hoiem, D. (2008). Learning CRFs using graph cuts. *European Conference on Computer Vision (ECCV)*, pages 582–595. 14

[Tabesh et al., 2007] Tabesh, A., Teverovskiy, M., Pang, H.-Y., Kumar, V. P., Verbel, D., Kotsianti, A., and Saidi, O. (2007). Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 26(10):1366–1378. 1

[Tang et al., 2012] Tang, L., Bressmann, T., and Hamarneh, G. (2012). Tongue contour tracking in dynamic ultrasound via higher-order mrfs and efficient fusion moves. *Medical Image Analysis (MedIA)*, 16(8):1503–1520. 46

[Tang and Hamarneh, 2013] Tang, L. and Hamarneh, G. (2013). Medical image registration: A review (chapter 22). *Medical Imaging: Technology and Applications*, pages 619–660. 47

[Toennies, 2012] Toennies, K. D. (2012). Guide to medical image analysis. 59

[Tola et al., 2008] Tola, E., Lepetit, V., and Fua, P. (2008). A fast local descriptor for dense matching. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. IEEE. 24

[Top and et al., 2011] Top, A. and et al. (2011). Active learning for interactive 3D image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6893, pages 603–610. Springer. 20, 107

[Top et al., 2011] Top, A., Hamarneh, G., and Abugharbieh, R. (2011). Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3d image segmentation. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pages 204–213. Springer. 20

[Tsai et al., 2003] Tsai, A., Yezzi Jr, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E., and Willsky, A. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 22(2):137–154. 18, 33

[Tsai et al., 2001] Tsai, A., Yezzi Jr, A., Wells III, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E., and Willsky, A. (2001). Model-based curve evolution technique for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, volume 1, pages I–463. IEEE. 33

[Tsai et al., 2002] Tsai, P., Chang, C.-C., and Hu, Y.-C. (2002). An adaptive two-stage edge detection scheme for digital color images. *Real-Time Imaging*, 8(4):329–343. 27

[Tu and Bai, 2010] Tu, Z. and Bai, X. (2010). Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 32(10):1744–1757. 35

[Tu et al., 2006] Tu, Z., Zhou, X. S., Comaniciu, D., and Bogoni, L. (2006). A learning based approach for 3d segmentation and colon detagging. In *European Conference on Computer Vision (ECCV)*, pages 436–448. 24

[Turner et al., 1996] Turner et al., A. (1996). Obtaining multiple distinct solutions with genetic algorithm niching methods. *Parallel Problem Solving from NatureŮPPSN IV*, pages 451–460. 3, 50

[Ukwatta et al., 2012] Ukwatta, E., Yuan, J., Rajchl, M., and Fenster, A. (2012). Efficient global optimization based 3D carotid AB-LIB MRI segmentation by simultaneously evolving coupled surfaces. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 377–384. 18, 40, 61

[Ulén et al., 2013] Ulén, J., Strandmark, P., and Kahl, F. (2013). An efficient optimization framework for multi-region segmentation based on lagrangian duality. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 32(2):178–188. 2, 4, 14, 18, 40, 41, 42, 43, 61, 67, 73, 77, 83, 84, 86

[Üzümcü et al., 2003] Üzümcü, M., Frangi, A. F., Sonka, M., Reiber, J. H., and Lelieveldt, B. P. (2003). Ica vs. pca active appearance models: Application to cardiac mr segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 451–458. 34

[Van Ginneken et al., 2002] Van Ginneken, B., Frangi, A. F., Staal, J. J., ter Haar Romeny, B. M., and Viergever, M. A. (2002). Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 21(8):924–933. 33

[Vazquez-Reina et al., 2009] Vazquez-Reina, A., Miller, E., and Pfister, H. (2009). Multiphase geometric couplings for the segmentation of neural processes. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 2020–2027. 18, 39, 42, 61

[Veksler, 2008] Veksler, O. (2008). Star shape prior for graph-cut image segmentation. In *European Conference on Computer Vision (ECCV)*, pages 454–467. Springer. 18, 32

[Vese and Chan, 2002] Vese, L. and Chan, T. (2002). A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision (IJCV)*, 50(3):271–293. xiii, 14, 21, 24, 30, 68, 73, 109

[Vicente et al., 2008] Vicente, S., Kolmogorov, V., and Rother, C. (2008). Graph cut based image segmentation with connectivity priors. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 1–8. 18, 36

[Vincent and Soille, 1991] Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 13(6):583–598. 1

[Wang et al., 2010] Wang, C., Teboul, O., Michel, F., Essafi, S., and Paragios, N. (2010). 3d knowledge-based segmentation using pose-invariant higher-order graphs. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 189–196. Springer. 35

[Wang et al., 2007] Wang, J., Agrawala, M., and Cohen, M. F. (2007). Soft scissors: an interactive tool for realtime high quality matting. In *ACM Transactions on Graphics (TOG)*, volume 26, page 9. ACM. 20

[Wang and Vemuri, 2004a] Wang, Z. and Vemuri, B. (2004a). Tensor field segmentation using region based active contour model. *European Conference on Computer Vision (ECCV)*, pages 304–315. 60, 88

[Wang and Vemuri, 2004b] Wang, Z. and Vemuri, B. C. (2004b). An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, volume 1, pages I–228. IEEE. 22, 23, 24

[Weldeselassie and Hamarneh, 2007] Weldeselassie, Y. T. and Hamarneh, G. (2007). Dt-mri segmentation using graph cuts. In *Medical Imaging*, pages 65121K–65121K. International Society for Optics and Photonics. 22

[Wells III et al., 1996] Wells III, W. M., Grimson, W. E. L., Kikinis, R., and Jolesz, F. A. (1996). Adaptive segmentation of mri data. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 15(4):429–442. 15

[Wijnhout et al., 2009] Wijnhout, J., Hendriksen, D., Assen, H., and Geest, R. (2009). LV challenge LKEB contribution: Fully automated myocardial contour detection. *Midas Journal*, page 683. 77

[Wu et al., 2012] Wu, X., Amrikachi, M., and Shah, S. (2012). Embedding topic discovery in conditional random fields model for segmenting nuclei using multispectral data. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 59(6):1539–1549. 2, 3, 49

[Wu et al., 2011] Wu, X., Dou, X., Wahle, A., and Sonka, M. (2011). Region detection by minimizing intraclass variance with geometric constraints, global optimality, and efficient approximation. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 30(3):814–827. 18, 40, 41, 42, 61

[Xu and Prince, 1997] Xu, C. and Prince, J. L. (1997). Gradient vector flow: A new external force for snakes. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 66–71. IEEE. 20

[Xu and Prince, 1998] Xu, C. and Prince, J. L. (1998). Generalized gradient vector flow external forces for active contours. *Signal Processing*, 71(2):131–139. 20

[Yan and Kassim, 2006] Yan, P. and Kassim, A. A. (2006). Segmentation of volumetric mra images by using capillary active contour. *Medical Image Analysis (MedIA)*, 10(3):317–329. 1

[Yang et al., 2008] Yang, L., Tuzel, O., Meer, P., and Foran, D. (2008). Automatic image analysis of histopathology specimens using concave vertex graph. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 833–841. 2, 3, 49

[Yazdanpanah et al., 2011] Yazdanpanah, A., Hamarneh, G., Smith, B. R., and Sarunic, M. V. (2011). Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 30(2):484–496. 29

[Yuan et al., 2012] Yuan, J., Bae, E., Boykov, Y., and Tai, X.-C. (2012). A continuous max-flow approach to minimal partitions with label cost prior. In *Scale Space and Variational Methods in Computer Vision*, pages 279–290. Springer. 18, 46

[Zeng et al., 1998] Zeng, X., Staib, L. H., Schultz, R. T., and Duncan, J. S. (1998). Volumetric layer segmentation using coupled surfaces propagation. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 708–715. 18, 39, 41, 42, 61

[Zeng et al., 2008] Zeng, Y., Samaras, D., Chen, W., and Peng, Q. (2008). Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in n–d images. *Computer Vision and Image Understanding (CVIU)*, 112(1):81–90. xiii, 18, 36, 37

[Zhang et al., 2012] Zhang, S., Zhan, Y., and Metaxas, D. N. (2012). Deformable segmentation via sparse representation and dictionary learning. *Medical Image Analysis (MedIA)*, 16(7):1385–1396. 35

[Zhang et al., 2001] Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 20(1):45–57. 15

[Zhao et al., 1996] Zhao, H., Chan, T., Merriman, B., and Osher, S. (1996). A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1):179–195. 18, 61, 63

[Zhu and Yuille, 1996] Zhu, S. C. and Yuille, A. (1996). Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 18(9):884–900. 18, 45

[Zhu-Jacquot and Zabih, 2007] Zhu-Jacquot, J. and Zabih, R. (2007). Graph cuts segmentation with statistical shape priors for medical images. In *International IEEE Conference on Signal-Image Technologies and Internet-Based System*, pages 631–635. 18, 33