

UNIVERSITY OF NICE-SOPHIA ANTIPOLIS

Doctoral School of Information and Communication Sciences and Technologies

THESIS

for the degree of

DOCTOR OF SCIENCE

OF THE UNIVERSITY OF NICE-SOPHIA ANTIPOLIS

Specialized in Automatic Processing of Signal and Images

by

Ting PENG

NEW HIGHER-ORDER ACTIVE CONTOUR MODELS, SHAPE PRIORS, AND MULTISCALE ANALYSIS

THEIR APPLICATION TO ROAD NETWORK EXTRACTION FROM VERY HIGH RESOLUTION SATELLITE IMAGES

Supervised by Ian H. JERMYN¹, Véronique PRINET², Josiane ZERUBIA¹ and Baogang Hu²
and prepared in the Project-Team ARIANA¹ (INRIA Sophia Antipolis)
and at LIAMA & NLPR² (Institute of Automation, Chinese Academy of Sciences)

Defended in Beijing on 18 November 2008, in front of the committee composed of

Patrick Louis COMBETTES	Professor, UPMC	Examiner
Jianwen MA	Professor, CEODE, CAS	Reviewer
Nikos PARAGIOS	Professor, ECP	Reviewer
Zengfu WANG	Professor, USTC	Examiner
Baogang HU	Professor, CASIA	Director
Ian H. JERMYN	Full-Time Researcher, INRIA	Co-supervisor
Véronique PRINET	Associate Professor, CASIA	Co-supervisor
Josiane ZERUBIA	Research Director, INRIA	Director

Acknowledgements

I would like to express my gratitude to all those who made this thesis possible.

First and foremost, I want to thank my supervisors Véronique Prinnet, Ian Jermyn, Josiane Zerubia and Baogang Hu. I have been in the project of Véronique since 2004 when I started my Master studies. During these years, she was always available for me in order to help me and guide me in my work. I am really glad that I have come to know her in my life. I am deeply grateful to Ian for his continuous guidance and support. He provided encouragement, sound advice, good teaching, and lots of good ideas. He checked carefully the final version of the thesis for English style and grammar, correcting both and offering suggestions for improvement. Thanks to Josiane for accepting me as a co-supervised PhD student and providing many constructive comments. Professor Hu also gave me many perceptive remarks and suggestions.

I very warmly thank the members of my PhD committee, Professors Patrick Louis Combettes, Jianwen Ma, Nikos Paragios, and Zengfu Wang for evaluating my manuscript.

I would like to thank Bruno Serra from Thales Alenia Space, and Haitao Zhang and Lebin Sun from Beijing Institute of Surveying and Mapping for their support and help.

I greatly appreciate Véronique Prinnet, Josiane Zerubia and many other people for their efforts to enter into the cotutelle agreement between the CASIA and the UNSA.

I wish to thank my colleagues at Ariana: Marie, Guillaume, Florent, Peter, Avik, Dan, Pierre, Alex, Alexis, Olivier, Praveen, Maria and Aymen; and my colleagues at LIAMA: Jinghui Duan, Timothée, Wei Liu, Jun Bai, Fengfeng Tang, Cyril, Vincent, Liangliang He, Régis, Fei Yuan and Dongmin Ma. All of them contributed to an excellent and inspiring working atmosphere. Special thanks to Guillaume, not only for his kindness and friendship, but also for introducing another Guillaume to me. I am grateful to Corinne, Régine and Bizhen Hong for their warmth and responsiveness. I also wish to thank my good friends Huizhi Zhao, Hua Zhang, Trung and Junbo for all the emotional support and care they provided. Sincere thanks to Trung who gave me invaluable help when I was a newcomer to the French riviera.

Lastly, I owe big thanks to my parents, my fiancée, my uncle and my aunt for their love, endless patience and encouragement during the 3-year PhD period.

Contents

Acknowledgements	i
List of Figures	vii
List of Tables	ix
Notations	xi
Introduction	1
Résumé détaillé (version française)	7
1 State-of-the-Art	13
1.1 Active Contours	13
1.1.1 Edge-Based Active Contours	13
1.1.2 Region-Based Active Contours	18
1.1.3 Active Contours with Shape Constraints	22
1.1.4 Summary	24
1.2 Road Extraction	25
1.2.1 Road Characteristics	26
1.2.2 Active Contours	27
1.2.3 Multiscale and Multiresolution Analysis	28
1.2.4 Dynamic Programming	29
1.2.5 Morphological Methods	30
1.2.6 Markov Random Fields and Marked Point Processes	31
1.2.7 Recursive Filtering	32
1.2.8 Support Vector Machines	33
1.2.9 Utilization of New Sensor Data	33
1.2.10 Other Methods	34
1.2.11 Summary	35
1.3 Conclusion	36

2	Higher-Order Active Contours and Phase Fields	37
2.1	A General Framework for Image Segmentation	37
2.2	Summary of Higher-Order Active Contours and Phase Fields	38
2.2.1	Higher-Order Active Contours	38
2.2.2	Phase Field Models	41
2.3	Stability Analysis of the Standard HOAC Total Prior Model	44
2.3.1	Definition of a Bar	44
2.3.2	Basic Phase Field Term of a Bar	45
2.3.3	Standard Phase Field HOAC Term of a Bar	46
2.3.4	Standard HOAC Total Prior Model of a Bar	46
2.4	Overall Model for Linear Network Extraction	52
2.4.1	Data Energy	52
2.4.2	Optimization and Parameter Setting	53
2.5	Conclusion	54
3	Multiresolution Analysis of the Primary Model	57
3.1	Multiresolution Analysis and Wavelets	57
3.1.1	Definition of Wavelet Based Multiresolution Analysis	58
3.1.2	Haar Multiresolution Analysis	61
3.2	Model Definition at Multiple Resolutions	61
3.2.1	Multiresolution Data Model	63
3.2.2	Multiresolution Framework	64
3.3	Experimental Results and Comparisons	66
3.3.1	Results Using the Single-Resolution Model	66
3.3.2	Results Using the Multiresolution Model	68
3.3.3	Results Using the Multiresolution Framework and Comparisons	68
3.4	Conclusion	77
4	GIS Specific Prior for Map Updating	79
4.1	Introduction	79
4.2	Specific Prior Energy	81
4.3	Experimental Results and Comparisons	82
4.3.1	Results Using the GIS-Based HOAC Model	82
4.3.2	Evaluation and Comparison	86
4.4	Conclusion	88
5	Modeling Shape for Network Extraction	89
5.1	Introduction	89
5.2	Nonlinear Nonlocal HOAC Prior Energy	91
5.2.1	Contour Model Definition	91
5.2.2	Phase Field Model Definition	92
5.2.3	Stability Analysis of the Nonlinear Nonlocal HOAC Total Prior Model	93
5.3	Linear Nonlocal HOAC Prior Energy	96
5.3.1	Contour Model Definition	97

5.3.2	Phase Field Model Definition	98
5.3.3	Stability Analysis of the Linear Nonlocal HOAC Total Prior Model	99
5.4	Experimental Results and Comparisons	106
5.4.1	Nonlinear Nonlocal Overall Model	106
5.4.2	Linear Nonlocal Overall Model	110
5.5	Conclusion	118
Conclusion		119
A Stability Calculations		123
A.1	Energy Terms of a Bar	123
A.1.1	Basic Phase Field Term	123
A.1.2	Standard Phase Field HOAC Term	124
A.1.3	Nonlinear Nonlocal Phase Field Term	124
A.1.4	Linear Nonlocal Phase Field Term	125
A.2	Model Energy per Unit Length	126
A.2.1	Standard HOAC Total Prior Model	126
A.2.2	Nonlinear Nonlocal HOAC Total Prior Model	127
A.2.3	Linear Nonlocal HOAC Total Prior Model	128
B Evolution Equations of New HOAC Prior Energies		131
B.1	Nonlinear Nonlocal HOAC Prior Energy	131
B.2	Linear Nonlocal HOAC Prior Energy	132
C Another Nonlinear Nonlocal HOAC Prior Term		135
C.1	Definition of \tilde{E}_{NL}	135
C.2	Derivative of \tilde{E}_{NL}	136
D Another Type of Interaction Function		139
E Summary of Other Methods Used in Our Comparisons		143
E.1	Method by Wang	143
E.2	Method by Yu	143
E.3	Method by Bailloeuil	144
F Publications and Scientific Activities of the Author		145
Bibliography		147

List of Figures

1	Input QuickBird data and its zooms.	3
2.1	Interaction function Ψ	39
2.2	Effects of the quadratic HOAC term.	40
2.3	Examples for road detection on a SPOT satellite image (© CNES).	40
2.4	Example of the potential function U	43
2.5	Bar of length $L \rightarrow \infty$ and width W	45
2.6	Diagram of $\hat{\beta}$ and \hat{W} for $e_{P,S}$	48
2.7	Prior energy $e_{P,S}$ for different $\hat{\beta}$	49
2.8	Geometric evolutions using different $\hat{\beta}$	50
2.9	Geometric evolutions using the stable parameter setting, with other initial shapes.	51
2.10	Example of histograms of the intensity and the variance, and their fitted models.	53
3.1	Multiresolution representation of the original data via a wavelet transform.	62
3.2	Flowchart of the multiresolution framework.	65
3.3	Experiment on Figure 1(a) at level 3, using the single-resolution primary model.	67
3.4	Experiments on Figure 1(a) at finer resolutions, using the single-resolution primary model.	67
3.5	Experiment on Figure 1(a) at full resolution, using the multiresolution model.	68
3.6	Data, and experiment at 1/4 resolution, using the single-resolution primary model.	70
3.7	Experiments on Figure 3.6(a) at full resolution.	71
3.8	Comparisons on Figure 3.6(a) at 1/4 resolution.	72
3.9	Comparisons on Figure 3.6(a) at full resolution.	73
3.10	Experiment on Figure 1(a) at full resolution, using the multiresolution framework.	73
3.11	More experiments on the QuickBird image.	74
3.12	More experiments on the QuickBird image.	75
3.13	More experiments on the QuickBird image.	76
4.1	Two pairs of QuickBird panchromatic images.	80

4.2	Experiments on the later pair of images in Figure 4.1 at full resolution, using the GIS-based HOAC model.	83
4.3	Ground truth and simulated GIS map.	84
4.4	Experiments on Figure 1(a) at full resolution, using the GIS-based HOAC model.	84
4.5	Road extraction results from Figure 1(a) at full resolution.	86
4.6	Road extraction results from Figure 4.1(c) at full resolution.	87
4.7	Road extraction results from Figure 4.1(d) at full resolution.	87
5.1	Example of small roads.	90
5.2	Interaction functions $\Psi_{+ }$ and $\Psi_{- }$	91
5.3	Diagram of $\hat{\beta}$, $\hat{\beta}_2$ and \hat{W} , for $e_{P,NL}$ ($w = 2$).	95
5.4	Effects of E_L	98
5.5	Stability analysis of $e_{P,L}$, for $\hat{d}_2 = 2 < \hat{D}_2$	101
5.6	Stability analysis of $e_{P,L}$, for $\hat{d}_2 = 5.5 > \hat{D}_2$	102
5.7	Graphs of $e_{P,L}$ when the parameter setting is moved from left to right along the dotted path P shown in Figure 5.6.	103
5.8	Graphs of $e_{P,L}$ when the parameter setting is chosen in the two intersection points of curves shown in Figure 5.6.	104
5.9	Data, and experiment at 1/4 resolution, using the nonlinear model (with E_{NL}).	107
5.10	Experiments on Figure 5.9(a) at 1/4 resolution and ground truth.	108
5.11	More experiments at reduced resolutions.	109
5.12	Data, experiments using the linear model (with E_L), and analysis on the effect of the different terms in the energy.	111
5.13	Comparisons on Figure 5.12(a) at full resolution.	112
5.14	More experiments at a reduced resolution and at full resolution, using the new linear model (with E_L).	113
5.15	A large image.	114
5.16	Road extraction result on this large image obtained using our models, at full resolution.	115
5.17	Result of river extraction.	116
5.18	Extraction of a road network containing two different widths, at 1/4 resolution.	116
D.1	New interaction function $\tilde{\Psi}$	139
D.2	For different given values of $\hat{d}_2 > \hat{D}_2$, diagrams of $\hat{\beta}$ and $\hat{\beta}_2$, and corresponding graphs of $\tilde{e}_{P,L}$ at the upper swallowtail corner.	141

List of Tables

3.1	Quality measures for the different methods tested on Figure 3.6(a) at 1/4 resolution.	69
3.2	Quality measures for the different methods tested on Figure 3.6(a) at full resolution.	72
4.1	Average quality measures for our different models.	85
4.2	Average quality measures for the different methods.	88
5.1	Quantitative evaluation criteria for the different methods tested on Figure 5.9(a) at 1/4 resolution.	106
5.2	Quantitative criteria tested on Figure 5.12(a) at full resolution.	117

Notations

I	image
Ω	image domain
K	prior knowledge
γ	contour
$\dot{\gamma}$	tangent vector of γ
$\Delta\gamma(t, t')$	$\gamma(t) - \gamma(t')$
R	region
\bar{R}	region complement, <i>i.e.</i> $\Omega \setminus R$
R_C	interface region
∂R	boundary of the region R
R_0	a reference shape
ϕ	level set function or phase field function
ϕ_+, ϕ_-	characteristic functions of R and \bar{R}
ϕ_{R_0}	phase field function of a reference shape
ϕ_{R_0+}, ϕ_{R_0-}	characteristic functions of R_0 and \bar{R}_0
z	threshold of the phase field
U	potential function
$A(R)$	interior area of the region
$L(\partial R), L(C)$	length of the boundary contour
κ	curvature
Ψ	interaction function
$\Psi_{+ }$	interaction function along the branch
$\Psi_{- }$	interaction function across the branch
d, d_2	interaction range
w	interface width
W	branch width
\hat{W}	scaled width, <i>i.e.</i> W/d
V	variance of the image I
P_{\pm}	density functions of the grey level I on-road (+) and off-road (-)
Q_{\pm}	density functions of the variance V on-road (+) and off-road (-)

E_P	prior model
E_0	basic phase field term
e_0	basic phase field term per unit length of bar
E_S	standard phase field HOAC term
e_S	standard phase field HOAC term per unit length of bar
$E_{S,CON}$	standard phase field HOAC term with a weight constrained by a reduced resolution result
E_{NL}	nonlinear nonlocal phase field HOAC term
e_{NL}	nonlinear nonlocal phase field HOAC term per unit length of bar
E_L	linear nonlocal phase field HOAC term
e_L	linear nonlocal phase field HOAC term per unit length of bar
E_{GIS}	specific prior term derived from a GIS map or a result at reduced resolution
E_D	data model
$E_{D,MUL}$	multiresolution data term
$E_{C,S}$	standard HOAC total prior model in the contour representation
$E_{P,S}$	standard HOAC total prior model in the phase field representation, <i>i.e.</i> $E_0 + E_S$
$e_{P,S}$	standard HOAC total prior model in the phase field representation per unit length of bar, <i>i.e.</i> $e_0 + e_S$
$E_{P,NL}$	nonlinear nonlocal HOAC total prior model in the phase field representation, <i>i.e.</i> $E_0 + E_S + E_{NL}$
$e_{P,NL}$	nonlinear nonlocal HOAC total prior model in the phase field representation per unit length of bar, <i>i.e.</i> $e_0 + e_S + e_{NL}$
$E_{P,L}$	linear nonlocal HOAC total prior model in the phase field representation, <i>i.e.</i> $E_0 + E_S + E_L$
$e_{P,L}$	linear nonlocal HOAC total prior model in the phase field representation per unit length of bar, <i>i.e.</i> $e_0 + e_S + e_L$
E_{primary}	primary model, <i>i.e.</i> $\theta(E_0 + E_S) + E_D$
$E_{\text{primary,mul}}$	primary model with the multiresolution data term, <i>i.e.</i> $\theta(E_0 + E_S) + E_{D,MUL}$
E_{HR}	primary model with the weight constraint quadratic prior term, <i>i.e.</i> $\theta(E_0 + E_{S,CON}) + E_D$
GIS	Geographical Information System
GVF	Gradient Vector Flow
HOAC	Higher-Order Active Contour
LIDAR	LIght Detection And Ranging
MAP	Maximum A Posteriori
MLE	Maximum Likelihood Estimation
MRA	Multiresolution Analysis
MRFs	Markov Random Fields
PCA	Principal Component Analysis
PDE	Partial Differential Equation
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
VHR	Very High Resolution

Introduction

Scientific Goal and Applications

The need to segment, from an image, entities that have the form of a ‘network’, *i.e.* branches joining together at junctions, arises in a variety of domains. Examples include the segmentation of road and river networks in remote sensing imagery, and of vascular networks in medical imagery. However, extracting automatically the network region in the image is a difficult task, because images are usually complex, containing much noise and confounding elements having similar local properties to the entity of interest. For this reason, techniques that include no prior knowledge about the region containing the network cannot succeed. In order to solve this problem, such prior knowledge must be injected somehow, either through the intervention of a user, or by incorporating it into a model. Human experts possess very specific prior knowledge about the shape of regions corresponding to networks, and in most applications, this level of knowledge is necessary rather than merely sufficient. Unfortunately, current methods, based on manual extraction, are time and labor intensive. On the other hand, the most generic prior knowledge alone, for example concerning boundary smoothness, is not enough. The need to include more specific prior knowledge of a class of shapes raises a difficult methodological issue, however. The set of network-like regions is complicated to model, because they may have arbitrary topology. More concretely, it consists of a large (in principle infinite) number of connected components, corresponding to the different possible topologies of a network (number of connected components in the network, number of loops in each connected component), or equivalently to the set of planar graphs (for 2D data). To this is added a geometric superstructure corresponding to an embedding of the graph in the plane, and to its ‘fattening’ into a region. The construction of a model that favors regions lying in this set as opposed to those outside it is a non-trivial problem.

The incorporation into models of prior knowledge about a region to be segmented from an image has a long history. The earliest and still most widely used models incorporate local knowledge about the boundary, essentially smoothness: active contours (Kass et al., 1988) are one example, the Ising model (Ising, 1925; Geman and Geman, 1984) another. This degree of prior knowledge is almost never enough to segment an entity of interest automatically, even in relatively simple images. More recent work has focused on the inclusion of more specific prior knowledge into the active contours, as we will see in subsection 1.1.3.

This work involves shape priors saying that the region sought must be ‘close’ to an exemplar region or regions. This type of model is useful for many applications, when a topologically correct reference shape is given. An intermediate level of prior knowledge is to model families of shapes. To model network regions, Rochery et al. (2006) introduced ‘higher-order active contours’ (HOACs). HOACs incorporate not only local, differential knowledge about the boundary, but also nonlocal, long-range interactions between tuples of contour points. Via such interactions, they favor regions with particular geometric characteristics without constraining the topology via use of a reference region.

Detecting roads from remotely sensed imagery is critical for many applications, for example cartographic data updating, intelligent navigation, environmental monitoring, disaster management, and so on. Recently, the commercial availability of very high resolution (VHR) optical satellite images (QuickBird, Ikonos, and in the near future Pléiades), with sub-metric resolutions, provides new opportunities for the extraction of information from remotely sensed imagery. Figure 1(a) shows one example of our input QuickBird panchromatic images ($\sim 0.61\text{m/pixel}$). At this resolution, qualitatively new categories of information are available, and the accuracy of previously extracted categories of information can be quantitatively improved. For example, road networks can be extracted as two-dimensional regions rather than as one-dimensional structures, and the geometric accuracy of the extracted road network can be greatly improved. As a result, road extraction from VHR images has become an increasingly important research topic in remote sensing.

Higher resolution brings with it new challenges however. The appearance of details invisible in lower resolution images can easily disrupt the recognition process. Figure 1(b) shows a full resolution zoom on Figure 1(a), illustrating the complexity existing in VHR images; and Figure 1(c) shows a full resolution zoom on a $1/8$ resolution image, illustrating that even after three levels of smoothing and down-sampling, the data is still rather complex. The difficulties lie in the following factors. First, much ‘noise’ exists in the road region due to cars, road markings, shadows, etc., while the background is very diverse, containing many features that are locally similar to roads. Second, rather than being simple lines as in images at low resolutions, *i.e.* more than 5m/pixel , roads appear as elongated, more or less homogeneous surfaces with different widths and curvatures. It is a hard task to retrieve road surfaces. Last but not least, in a dense urban environment, the contrast between roads and background is relatively poor, as opposed to a rural or semi-urban area. All the above factors result in the relative failure of existing road extraction approaches. The development of a reliable automatic algorithm for road extraction in dense urban areas from VHR images is thus a necessity if the increasing demand is to be met.

In this thesis, we propose new variational models for network modeling. In particular, first, to overcome a serious limitation of the previous model introduced by Rochery (2005), we develop two new HOAC shape prior energy terms, both of which allow separate control of road straightness and width, and thus better modeling of elongated structures. Second, we combine together all three different types of prior geometric knowledge. They are: generic boundary smoothness constraints; intermediate prior knowledge for modeling a family of shapes; and specific shape prior knowledge of the road network derived from Geographical Information System (GIS) data created at an earlier date. Even if errors exist in the GIS data, we can perform GIS updating. In addition, we take advantage of multiresolution analysis

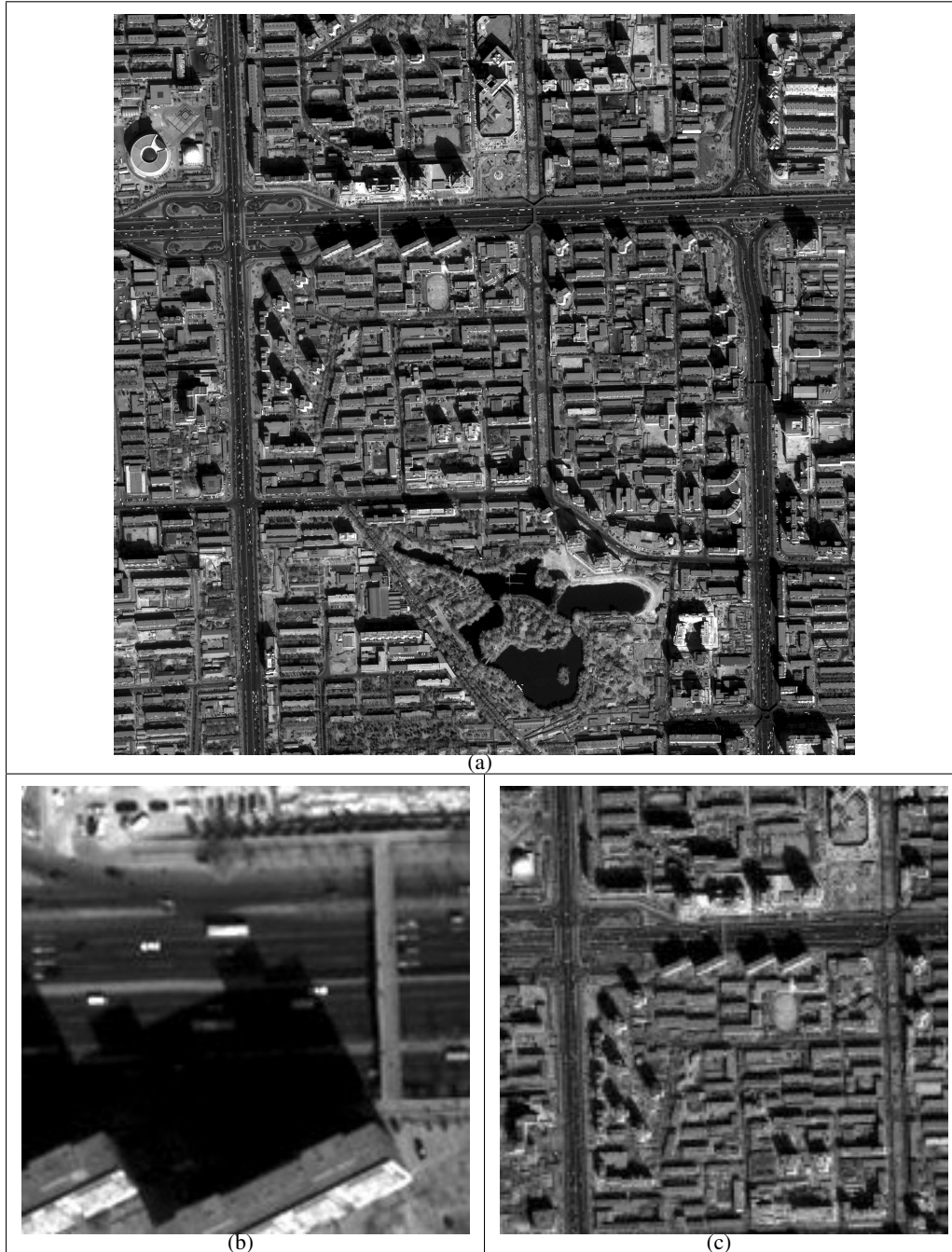


Figure 1. Input data. (a): a QuickBird image (size: 2560×2560); (b): a full resolution zoom on the image; (c): a full resolution zoom on $1/8$ resolution image, *i.e.* scaling coefficients after performing a wavelet transform three times.

(MRA) to alleviate data complexity. We test and evaluate our models on the problem of road network extraction in a dense urban area from VHR QuickBird images of Beijing.

Organization of the Thesis Manuscript

This thesis is organized as follows:

Chapter 1

In the first part, we review existing variational approaches for segmentation based on curve evolution. We classify the various models into two main classes of evolution: edge-based and region-based approaches. Moreover, special attention is paid to active contours incorporating prior shape constraints. In the second part, we describe representative approaches to road extraction from remotely sensed imagery, organized according to the different techniques used. The advantages and drawbacks of each type of model are discussed.

Chapter 2

We describe the theoretical foundations of the thesis. First, we briefly introduce the different energy terms of HOACs and its corresponding phase field equivalent (Rochery et al., 2005b; Rochery, 2005; Rochery et al., 2006). We show the equivalence between phase field modeling and standard active contours. Due to the incorporation of long-range interactions between a pair of contour points, HOACs are a powerful tool for modeling networks. We define our data energy term, and then formulate our primary model adapting ‘phase field HOACs’ to the problem of urban road extraction from VHR images. However, one of the problems we face is that there are a number of parameters in the HOACs that need to be estimated in a more systematic way. For this purpose, we calculate each energy term on an *ansatz* of a straight long bar. Based on this calculation, we explore stability conditions and possible behaviors of the model, and describe how to choose the model parameters properly.

Chapter 3

In order to overcome the great complexity of the information existing in VHR images, we focus on an MRA of the primary model. We first introduce a multiresolution statistical data model. The use of several resolutions allows the combination of coarse resolution data, in which the detail that can disrupt recognition has been eliminated, with fine resolution data to increase precision. Subsequently, we propose a two-step robust multiresolution framework. We segment at first a low resolution image, and obtain an approximate pre-segmentation of the object of interest. This preliminary result is then incorporated into the segmentation at full resolution, to force the final segmentation to lie close to the pre-segmentation, thus diminishing the local minima of the total energy at full resolution.

Parts of this work have been published by Peng et al. (2007a, 2008b).

Chapter 4

In the context of updating an outdated digital map, we propose a specific prior term to incorporate the information available from a GIS map into our primary model. This map dates from before the image was acquired, and thus represents a different road network: the accuracy needs to be improved, and the network topology needs to be corrected. In our model, we include all three different types of prior geometric knowledge characterized by their level of generality. We show all three types of prior knowledge to be essential for a successful segmentation at full resolution. In addition, in the case of extracting the road network at full resolution with GIS data unavailable, we can use a result obtained at a reduced resolution, where GIS information appears not to be necessary in the presence of the other two types of prior knowledge, to replace GIS information in the specific prior term. Thus, we can free ourselves from the need for GIS data. This results in another MRA approach.

Parts of this work have been published by Peng et al. (2007b, 2008b).

Chapter 5

The previous chapters addressed the extraction of main road networks. Now we turn to the issue of extracting the secondary road network. Compared to the main roads, the difficulties with secondary roads lie on the one hand in the low discriminative power of the grey-level distributions of road regions and background, and on the other hand, in the greater effect of occlusions and other types of noise on narrower roads. This suggests that strong geometric prior information is needed. The model developed previously for main road extraction contains such prior knowledge, but it suffers from a limitation that is severe in the case of secondary roads: the scale on which the road is expected to be straight is the same as the width of the road, whereas in fact road width gives only an (approximate) upper bound on the radius of curvature of the road. Therefore, the solution is to separate the interaction function along one side of a road from the one across a road. To this end, we propose respectively two nonlocal HOAC prior energies: a nonlinear term and a linear term. The former increases the magnitude of the interactions along the road; and the latter provides longer-range interactions along the road. In fact, both of them achieve similar effects. As before, we analyze two new total prior models with respect to a straight long bar, and establish the parametrical constraints. Through this analysis of the linear term, we show that in addition, it enables the modeling of two widely separated widths simultaneously, and opens up new possibilities in applications.

Parts of this work have been published by Peng et al. (2008a,c).

Chapters 3, 4 and 5 are each illustrated with experimental results which are analyzed and compared with several other techniques in the literature.

To summarize, the contributions of this thesis are as follows:

1. to analyze stability conditions of the phase field HOAC model proposed by Rochery (2005) (chapter 2),

2. to propose a multiresolution statistical data term and a two-step multiresolution framework (chapter 3),
3. to propose a GIS prior term and combine all the three types of different geometric prior knowledge together (chapter 4),
4. to propose two nonlocal HOAC shape prior terms for modeling elongated structures, and to conduct the stability analysis on each of the new total prior models (chapter 5),
5. to make progress towards an automatic road extraction system for VHR optical satellite images (chapters 3-5).

Context of the PhD Study

This PhD was supervised jointly by LIAMA/NLPR (Institute of Automation, Chinese Academy of Sciences) and the Project-Team Ariana (INRIA/I3S). The PhD was funded by grants from MAE, LIAMA, and Thales Alenia Space.

Résumé détaillé (version française)

But Scientifique et Applications

La nécessité de segmenter, à partir d'une image, des entités qui ont la forme d'un 'réseau', c'est-à-dire des branches se joignant à des jonctions, apparaît dans de nombreux domaines applicatifs. Cela inclut, par exemple, l'extraction des réseaux routiers et fluviaux en télédétection, ou encore celle des réseaux vasculaires en imagerie médicale. Cependant, l'extraction automatique des réseaux d'une image est une tâche difficile. Les images sont, en général, complexes, très bruitées et certains éléments ayant localement des propriétés similaires à l'entité d'intérêt peuvent être aisément confondus avec cette dernière. C'est pourquoi, les techniques qui n'incluent aucune connaissance a priori sur la région du réseau ne peuvent pas aboutir. Afin de résoudre ce problème, une telle connaissance a priori doit être injectée d'une manière ou d'une autre : soit par l'intervention d'un utilisateur, soit en l'incorporant au sein d'un modèle. Les êtres humains possèdent une connaissance a priori très spécifique au sujet de la forme des régions correspondant aux réseaux, et dans la plupart des applications, ce niveau de connaissance est nécessaire. Malheureusement, les méthodes actuelles, basées sur l'extraction manuelle, sont très coûteuse en temps et en main-d'oeuvre. Par ailleurs, une seule connaissance a priori générique, par exemple concernant le degré de lissage d'une frontière, n'est pas suffisante. La nécessité d'inclure une connaissance a priori plus spécifique décrivant une classe de formes soulève cependant une question méthodologique difficile. Un ensemble de régions ressemblant à des réseaux est compliqué à modéliser parce que ces réseaux peuvent présenter des topologies arbitraires. Plus concrètement, il se compose d'un grand nombre (en principe infini) de composantes interconnectées, correspondant aux différentes topologies possibles d'un réseau (nombre de composantes interconnectées dans le réseau, nombre de boucles dans chaque composante interconnectée), ou d'une manière équivalente à un ensemble de graphes planaires (pour des données 2D). À ceci s'ajoute une superstructure géométrique correspondant au prolongement du graphe dans le plan, et à son 'épaississement' au sein d'une région. La construction d'un modèle qui favorise des régions se situant dans cet ensemble, par opposition à celles en dehors de lui est un problème non trivial.

L'incorporation au sein d'un modèle d'une connaissance a priori pour segmenter une région d'une image a une longue histoire. Les modèles les plus anciens et encore les plus largement utilisés incorporent une connaissance locale des frontières, essentiellement la

douceur : les contours actifs (Kass et al., 1988) en sont un exemple, le modèle d'Ising (Ising, 1925; Geman and Geman, 1984) un autre. Ce degré de connaissance a priori n'est presque jamais suffisant pour segmenter l'entité d'intérêt automatiquement, même avec des images relativement simples. Des travaux plus récents se sont concentrés sur l'inclusion d'une connaissance a priori plus spécifique dans les contours actifs, comme nous le verrons dans la partie 1.1.3. Ces travaux utilisent des termes de connaissance a priori indiquant que la forme de la région recherchée doit être 'proche' de celle de la région de référence. Ce type de modèle est utile pour beaucoup d'applications, quand une forme de référence topologiquement correcte est connue. Un niveau intermédiaire de connaissance a priori est de modéliser des familles des formes. Afin de modéliser des zones de réseaux, Rochery et al. (2006) ont introduit les 'contours actifs d'ordre supérieur' (CAOS). Les CAOS incorporent non seulement une connaissance différentielle locale sur les frontières, mais également des interactions nonlocales, à longue distance, entre des tuplets des points de contour. Par de telles interactions, ils favorisent des régions ayant des caractéristiques géométriques particulières sans contraindre la topologie du réseau à ressembler à celle d'une région de référence.

La détection des routes basée sur l'imagerie de télédétection est critique dans beaucoup d'applications, par exemple, dans les cas de la mise à jour des données cartographiques, la navigation intelligente, le contrôle de l'environnement, ou encore la gestion des catastrophes. La disponibilité commerciale des images satellitaires optiques à très haute résolution (THR) (QuickBird, Ikonos, et dans un avenir proche Pléiades), avec des résolutions sub-métriques, a fourni récemment de nouveaux moyens pour l'extraction d'information à partir d'images de télédétection. La Figure 1(a) montre un exemple d'images panchromatiques QuickBird utilisées comme données d'entrée ($\sim 0.61\text{m/pixel}$). À cette résolution, des catégories d'information nouvelles du point de vue qualitatif sont disponibles, et la précision des catégories précédemment extraites peut être quantitativement améliorée. Par exemple, les réseaux routiers peuvent être extraits en tant que régions bidimensionnelles plutôt qu'en tant que structures unidimensionnelles, et la précision géométrique du réseau routier extrait peut ainsi être considérablement améliorée. Par conséquent, l'extraction de réseaux routiers à partir d'images THR est devenue un sujet de recherche de plus en plus important dans le domaine de la télédétection.

Néanmoins, une résolution plus haute apporte avec elle de nouveaux défis. L'apparition des détails invisibles dans des images à plus basse résolution peut facilement perturber le processus d'identification. La Figure 1(b) représente un agrandissement à pleine résolution de la Figure 1(a), illustrant la complexité des images THR. La Figure 1(c) montre un gros plan de l'image à la résolution $1/8$, illustrant le fait que même après trois niveaux de lissage et de sous-échantillonnage, les données sont encore relativement complexes. Les difficultés résident dans les facteurs suivants : tout d'abord, beaucoup de 'bruit' existe sur la route elle-même, dû aux voitures, marquages routiers, ombres, etc. De plus, le fond est très diversifié, contenant beaucoup de traits qui sont localement similaires aux routes. Par ailleurs, les routes apparaissent en tant que surfaces allongées, et plus ou moins homogènes, ayant différentes largeurs et courbures, et non pas en tant que rubans fins tels que dans les images basse résolution, c'est-à-dire de plus de 5m/pixel . L'extraction de surfaces routières est une tâche complexe. Enfin, dans un environnement urbain dense, le contraste entre les routes et le fond est relativement faible, par opposition à une région rurale ou semi-urbaine. Tous

les facteurs ci-dessus concourent à l'échec relatif des approches existantes d'extraction de routes. Le développement d'un algorithme automatique fiable pour l'extraction de routes en zone urbaine dense à partir d'images THR devient ainsi une nécessité si l'on veut satisfaire la demande croissante de tels algorithmes.

Dans cette thèse, nous proposons de nouveaux modèles variationnels pour la modélisation de réseau. En premier lieu, afin de surmonter une limitation sérieuse du modèle précédemment introduit par Rochery (2005), nous proposons deux nouveaux termes d'énergie de forme a priori de CAOS, qui permettent le contrôle séparé de la courbure et de la largeur de la route, constituant ainsi une meilleure modélisation des structures allongées. D'autre part, nous combinons ensemble trois types différents de connaissance géométrique a priori : des contraintes génériques de douceur de frontière ; une connaissance a priori intermédiaire pour modéliser une famille des formes ; et une connaissance a priori de la forme spécifique du réseau routier obtenue à l'aide des données issues d'un Système d'Information Géographique (SIG). Même si des erreurs existent dans les données du SIG, nous pouvons effectuer sa mise à jour. En outre, nous profitons de l'analyse multi-résolution (AMR) pour réduire la complexité de données. Nous appliquons et évaluons nos modèles au problème d'extraction des réseaux routiers en zone urbaine dense à partir d'images QuickBird THR sur la ville de Pékin.

Organisation du Manuscrit de thèse

Cette thèse est organisée comme suit.

Chapitre 1

Dans la première partie de ce chapitre, nous passons en revue les approches variationnelles existantes de segmentation basées sur l'évolution de courbe. Nous classifions les divers modèles en deux classes principales d'évolution : les approches basées sur les frontières et celles basées sur les régions. De plus, une attention particulière est prêtée aux contours actifs incorporant des contraintes a priori de forme. Dans la deuxième partie, nous décrivons des approches représentatives d'extraction de routes à partir d'images de télédétection, organisées selon les différentes techniques utilisées. Les avantages et les inconvénients de chaque type de modèle sont discutés.

Chapitre 2

Nous posons les fondements théoriques du travail de la thèse. Nous présentons tout d'abord les différents termes de la fonctionnelle d'énergie et leurs transposés dans une modélisation par champs de phase (Rochery et al., 2005b; Rochery, 2005; Rochery et al., 2006). Nous montrons l'équivalence entre le modèle par champs de phase et les contours actifs standards. L'incorporation additionnelle d'un terme de CAOS permet de prendre en compte une interaction à longue portée entre paires de points du contour, fournissant ainsi un outil robuste pour la modélisation des réseaux. Nous définissons également le terme d'attache aux données et formulons ainsi notre modèle primaire, adaptant les CAOS par champs de

phase au problème d'extraction de route urbaine à partir d'images THR. Cependant, l'un des problèmes auxquels nous sommes confrontés est que les paramètres du modèle CAOS doivent être estimés d'une manière plus systématique. À cette fin, nous calculons chaque terme d'énergie sur un *ansatz* d'une longue barre droite. À partir de ce calcul, nous explorons les conditions de stabilité et les comportements possibles du modèle, et décrivons ainsi comment choisir correctement les paramètres du modèle.

Chapitre 3

Afin de surmonter la grande complexité d'information existant dans les images THR, nous nous concentrons, dans ce chapitre, sur une AMR du modèle primaire. Nous introduisons d'abord une modélisation statistique multi-résolution des données. L'utilisation de plusieurs résolutions permet la combinaison des données de résolution brute, dans lesquelles le détail qui pourrait perturber l'identification a été éliminé, avec des données de résolution fine, pour augmenter la précision. Nous proposons alors un cadre robuste multi-résolution en deux étapes. Nous segmentons tout d'abord une image à basse résolution, et obtenons une pre-segmentation approximative de l'objet d'intérêt. Ce résultat préliminaire est alors incorporé à la segmentation à pleine résolution, afin de forcer la segmentation finale à se trouver proche de la pre-segmentation, diminuant ainsi les minima locaux de l'énergie totale à pleine résolution.

Des parties de ce travail ont été publiées par Peng et al. (2007a, 2008b).

Chapitre 4

Dans le cadre de la mise à jour d'une ancienne carte numérique, nous proposons un terme a priori spécifique pour incorporer les informations disponibles d'une carte de SIG à notre modèle primaire. Cette carte est antérieure à l'acquisition de l'image, et représente ainsi un réseau routier différent : la précision doit être améliorée, et la topologie de réseau doit être corrigée. Dans notre modèle, nous incluons trois types différents de connaissance géométrique a priori, caractérisés par leur niveau de généralité. Nous montrons que les trois types de connaissance a priori sont essentiels pour une segmentation réussie à pleine résolution. En outre, dans le cas de l'extraction du réseau routier à pleine résolution avec des données SIG indisponibles, nous pouvons utiliser un résultat obtenu à une résolution réduite, où l'information de SIG semble ne pas être nécessaire en présence des deux autres types de connaissance a priori, pour remplacer l'information de SIG dans le terme a priori spécifique. Ainsi, nous pouvons nous libérer du besoin de données de SIG. Ce qui nous donne une autre approche d'AMR.

Des parties de ce travail ont été publiées par Peng et al. (2007b, 2008b).

Chapitre 5

Les chapitres précédents ont traité de l'extraction de réseaux routiers principaux. Nous nous tournons maintenant vers la question de l'extraction du réseau routier secondaire. Comparées au cas des routes principales, les difficultés liées aux routes secondaires se situent d'une part dans la faible discrimination entre la répartition en niveaux de gris des routes et

celle du fond, et d'autre part dans l'effet plus important des occlusions et des autres types de bruits sur les routes étroites. Ceci suggère qu'une information a priori géométrique forte est nécessaire. Le modèle développé précédemment pour l'extraction des routes principales contient une telle connaissance a priori, mais il souffre d'une limitation qui est critique dans le cas des routes secondaires : l'échelle à laquelle la route est supposée rectiligne est identique à la largeur de la route, alors qu'en réalité la largeur de route donne seulement une borne supérieure (approximative) sur le rayon de courbure de la route. Par conséquent, la solution est de séparer la fonction d'interaction suivant la direction de la route de celle traversant la route. À cet effet, nous proposons deux énergies nonlocales a priori de CAOS : un terme non-linéaire et un terme linéaire. Le terme non-linéaire augmente l'intensité des interactions le long de la route, tandis que le terme linéaire fournit des interactions à plus grande portée. En fait, les deux termes produisent des effets similaires. Comme précédemment, nous analysons deux nouveaux modèles a priori par rapport à une longue barre droite, et en établissons les contraintes de paramétrage. Par cette analyse du terme linéaire, nous montrons, en outre, qu'il permet simultanément la modélisation de deux largeurs clairement éloignées, et ouvre ainsi de nouvelles possibilités d'applications.

Des parties de ce travail ont été publiées par Peng et al. (2008a,c).

Les chapitres 3, 4 et 5 sont, chacun, illustrés par des résultats expérimentaux qui sont analysés et comparés à des résultats obtenus par plusieurs autres techniques issues de la littérature.

Pour résumer, les contributions de cette thèse sont les suivantes :

1. Analyser les conditions de stabilité du modèle de CAOS de champs de phase proposé par Rochery (2005) (chapitre 2),
2. Proposer un terme statistique multi-résolution des données et un cadre multi-résolution en deux étapes (chapitre 3),
3. Proposer un terme a priori de SIG et combiner ensemble les trois types différents de connaissance géométrique a priori (chapitre 4),
4. Proposer deux termes nonlocaux de CAOS de forme a priori, pour modéliser des structures allongées, et réaliser l'analyse de stabilité sur chacun des nouveaux modèles a priori résultants (chapitre 5),
5. Progresser vers un système automatique d'extraction de routes pour des images satellitaires optiques à THR (chapitres 3-5).

Conclusion

Le but de cette thèse était de développer de nouveaux modèles variationnels pour la segmentation, à partir d'une image d'entités qui prennent la forme d'un 'réseau', c'est-à-dire des branches se joignant ensemble à des jonctions. Plus particulièrement, nous nous

sommes concentrés sur l'incorporation de différents types de géométrie a priori des régions de réseau, en profitant d'une analyse multi-résolution de l'image. Nous avons appliqué les différents modèles que nous avons proposés à la segmentation des réseaux routiers en se basant sur des images panchromatiques QuickBird à THR ($\sim 0.61\text{m/pixel}$) en zone urbaine dense. Ceci a constitué une problématique ardue, étant donné la complexité existante dans les données, ainsi que la complexité de modélisation des régions de réseau à la topologie arbitraire.

Nous avons d'abord commencé par un modèle primaire formulé en champs de phase. Il se décompose en un modèle a priori standard de CAOS et un modèle probabiliste de ressemblance de régions. Grâce aux interactions à longue portée entre les pixels du CAOS, ce modèle nous a permis d'inclure la connaissance a priori sophistiquée de la géométrie de région. Les champs de phase ont également apporté beaucoup d'avantages par rapport à d'autres méthodes conventionnelles de modélisation de région. Pour surmonter les difficultés issues de la complexité du contenu de l'image à haute résolution, nous avons introduit un modèle statistique multi-résolution des données et un modèle a priori de contraintes multi-résolution. Ces deux modèles ont permis l'intégration efficace de l'information des images à différentes résolutions. L'analyse multi-résolution peut considérablement augmenter la robustesse de l'algorithme. Par la suite, dans le cadre de l'actualisation de données cartographiques, nous avons inclus un terme de forme a priori spécifique de la région, dérivé d'une carte de SIG, et nous l'avons combiné avec les autres termes a priori plus génériques. L'information issue du SIG, bien qu'ancienne, peut fournir des informations partiellement correctes de la scène considérée, et ainsi éliminer des fausses détections sur le fond de l'image. Nous avons montré que notre modèle pouvait améliorer la précision des objets inchangés, extraire les objets nouvellement apparus, et enlever les objets disparus de la carte.

Pour faciliter l'extraction des structures allongées et plus étroites, nous avons proposé deux nouveaux modèles a priori de CAOS. Ces deux modèles ont un effet similaire sur la modélisation de forme, c'est-à-dire qu'ils permettent de contrôler indépendamment la courbure et la largeur de la route, conjointement avec le terme standard de CAOS. Par conséquent, l'incorporation de ces deux termes a permis la génération de plus longues branches en forme de bras ainsi qu'une meilleure prolongation de celles-ci. De plus, le terme linéaire a plusieurs avantages par rapport au terme non-linéaire : il est plus efficace d'un point de vue calculatoire, et il peut modéliser des largeurs multiples simultanément. Dans chaque modèle a priori de CAOS, il y a un ensemble de paramètres de contrôle. Il a été d'une grande importance d'établir les contraintes internes entre ces paramètres, afin d'en fixer la valeur de certains en fonction des autres. Pour ces modèles a priori, utilisant une représentation de champs de phase, nous avons calculé leurs conditions de stabilité, et avons décrit les contraintes associées. Nous avons montré que les relations entre les paramètres de contrôle et le comportement de l'énergie suivaient certains cas particuliers de la théorie des catastrophes.

Nous avons testé et évalué les modèles proposés sur des images QuickBird de la ville de Pékin, et nous les avons comparés à de nombreuses autres techniques publiées dans la littérature. Les résultats expérimentaux et les comparaisons effectuées démontrent la supériorité des modèles que nous avons proposés.

Chapter 1

State-of-the-Art

In this chapter, we present a brief state-of-the-art for active contours and road extraction. Both fields have been widely studied for a long time. A significant number of approaches exist in the literature. Here we highlight only the most representative techniques of each of them. At the end of each section, we give a brief summary.

1.1 Active Contours

Active contours provide a framework for segmenting an object from an image. In principle, one has to define an appropriate energy functional associated to the current contour. During the minimization of such an energy functional, the boundary contour is evolved dynamically from some initialization in the direction of the negative of the energy gradient. Active contours can be divided into *edge-based* and *region-based* approaches. In this section, we first review the typical techniques of these two main categories. Thereafter, we focus on a special case: active contours with shape constraints.

1.1.1 Edge-Based Active Contours

Originally, active contours were edge-based methods. The models are driven towards the edges of an image. The evolution equation is computed from an energy functional that only relies on some *local* measurements around the boundary of the object to be segmented. The boundary is usually described by the local image gradient, or the field of the gradient vectors in the image domain. Edge-based active contours can be classified as *parametric active contours* or *geometric active contours*, according to whether the energy is parameterization-independent or not. More concretely, the functional of parametric active contours is not intrinsic, since one could obtain different solutions by changing the parameterization while preserving the same initial curve. The evolution of the parametric contour is normally implemented by a set of ordinary differential equations acting on the sample contour points. On the other hand, geometric active contours overcome the above drawback by constructing

an energy functional independent on the parameterization of the contour. This formulation also enables performance of more efficient level set schemes.

Parametric Active Contours

Starting from the vision theory of Marr (1982), typical low-level vision tasks have been considered as an autonomous *bottom-up* process. The three independent levels represent the processing of visual information, but up to the 2.5D sketch, no higher-level information has yet been brought to bear: the computations proceed by utilizing only what is available in the image itself. This rigidly sequential approach, which manages vision tasks as different independent steps, propagates mistakes made at a low level without opportunity for correction. However, due to the effect of many complex factors, such as noise, projection, etc., many low-level visual tasks are inherently ill-posed.

Kass et al. (1988) challenged this vision theory, and argued that in many image interpretation tasks, the correct interpretation of low-level events can require high-level knowledge. They sought to design energy functions whose local minima comprise the set of alternative solutions available to higher-level processes. The choice among these alternatives is accomplished by adding suitable energy terms together. In this way, high-level mechanisms can interact with the contour model by pushing image features toward an appropriate local minimum.

Based on the above idea, Kass et al. (1988) proposed an active contour model called *snakes*. Let $I : \Omega \rightarrow \mathbb{R}^+$; they defined the following energy functional along a parameterized contour $C : [0, 1] \rightarrow \Omega \subset \mathbb{R}^2$:

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

Snakes are dynamic curves defined in the image domain that can move under the influence of internal forces and external data forces. The first two terms are the internal energies coming from within the curve itself. They define geometric prior knowledge of the contour: rigidity and elasticity respectively. The last term is the external energy computed from the data. This term attracts the contour toward large image gradients. α, β and λ are relative weights describing the importance of each energy term. For minimizing the total energy, the evolution equation is derived by a simple gradient descent algorithm, to update the active contour until convergence.

The snake model has several advantages over conventional methods:

- Snakes provide a unified extraction procedure. They include image data, initial estimation, contour boundary and shape constraints, which had been treated separately in the past.
- After a proper initialization, snakes converge iteratively to an energy minimum state.
- Energy minimization from coarse to fine in scale space (Nielsen et al., 1997) can greatly enlarge the capture region around features of interest, and decrease the computational complexity.

Although this basic model has had a profound influence on the image processing and computer vision field, it suffers from several drawbacks:

- In general, the initial contour must be close enough to the true contour, which requires careful initialization.
- Since the contour is represented explicitly by sampled points, the model cannot handle topological changes. It is impossible to extract multiple objects.
- Points tend to bunch up on strong portions of an edge contour.
- Snakes are very sensitive to noise. In addition to the boundaries in an image, the noise is able to attract the active contour, and make parts of the contour adhere to it.

Many people have developed improvements to this model. Amini et al. (1990) proposed an algorithm for the active contour model using dynamic programming. This approach allows the inclusion of hard constraints that cannot be violated in addition to soft constraints inherent in the formulation of the functional. However, it improves the stability at the expense of an increase in computational complexity. Williams and Shah (1992) presented a greedy algorithm which has performance comparable to dynamic programming and variational calculus approaches. The distance between any two contiguous points is compared to the average distance between contiguous points. The difference is minimized for the internal terms so that points in the contour are more evenly spaced. It retains the properties of stability and flexibility. To simplify the curve representation, Menet et al. (1990) proposed the B-snakes model. The curve is represented by B-splines (Unser et al., 1993a,b), which provide an attractive formalism for parametric curves. McInerney and Terzopoulos (2000) tackled the issue of topological change for snakes at the expense of a computationally expensive and complex implementation.

Considering the internal forces of snakes, we can see that every curve tends to shrink and vanish. To fix this inclination, Cohen (1991) introduced the *balloon model*. In such a model, the force on the contour is modified by normalizing the associated potential force, and by adding a constant pressure force. The contour behaves like a balloon which is inflated or deflated by this additional internal force. The capture range of the contour is then enlarged, and the initial curve need no longer be very close to the ideal boundary. The curve can pass over weak edges and be stopped only if the edge is strong. In fact, the balloon force is the same as an interior area constraint. Then, Cohen and Cohen (1993) discussed the 3D generalization of the balloon model, and solved the snakes model using the finite-element method. This yields faster convergence and better stability. Leroy et al. (1996) proposed a multiresolution version of the balloon model. However, a problem with the balloon model is that the prior knowledge about whether the initial contour is inside or outside the ideal object boundary is required.

Another important improvement over snakes is the work of Xu and Prince (1998). They presented a new external force, called *gradient vector flow* (GVF). GVF is computed as a diffusion of the gradient vectors of an edge map derived from the image. Let us write f an edge map derived from the image having the property that it is larger near the edges, and

(u, v) the GVF field parameterized by (x, y) . The diffusion of the gradient information is realized by minimizing the energy functional:

$$E(u, v) = \int_{\Omega} \left\{ \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |(u, v)^T - \nabla f|^2 \right\} dx dy, \quad (1.2)$$

where u_x, u_y, v_x, v_y are the spatial derivatives of the field. This formulation keeps the vector field nearly equal to the gradient of the edge map when the norm $|\nabla f|$ is large, and forces the field to be slowly varying in homogeneous regions. The constant μ is set according to the amount of image noise. The GVF field can be found by updating

$$\begin{cases} u_t = \mu \nabla^2 u - (u - f_x) |\nabla f|^2, \\ v_t = \mu \nabla^2 v - (v - f_y) |\nabla f|^2. \end{cases} \quad (1.3)$$

Thus, the external force of the evolution equation in the snakes model is replaced by the GVF field. The GVF model allows for flexible initialization and encourages convergence to boundary concavities.

Geometric Active Contours

Osher and Sethian (1988) presented an intrinsic representation for evolving curves, called *level sets*. The curve C is represented implicitly via a Lipschitz function ϕ , by $C = \{x \in \Omega \mid \phi(x) = 0\}$. The evolution of the curve is given by the zero-level curve at time t of the function $\phi(x, t)$. Level sets have been extensively used, because there are many advantages to working with this representation:

- Level sets can naturally handle topology changes, cusps, and corners as ϕ evolves. This is the main advantage of level sets.
- The level set formulation can be easily extended to interfaces of higher dimension.
- The computations are made on a fixed rectangular grid over the entire image domain, so that numerical computation can be facilitated.

On the other hand, level sets also suffer from some drawbacks:

- Level sets bring greater computational complexity, which arises from using a one higher dimension representation of the original interface.
- The level set function needs to be initialized at the beginning, and to be reinitialized once every few iterative steps, as a distance function, which increases the calculation cost.

The motion of curve C along its normal direction with speed F is equivalent to solving the following partial differential equation (PDE):

$$\frac{\partial \phi}{\partial t} = F |\nabla \phi|, \quad (1.4a)$$

$$\phi(0, x) = \phi_0(x), \quad (1.4b)$$

where $t \in \mathbb{R}^+$, and the set $\{x \mid \phi_0(x) = 0\}$ defines the initial contour. This equation only specifies the speed F and the evolution of ϕ on the zero level set. As a consequence, the speed F has to be well-defined off the curve, *i.e.* in the whole image domain, so that ϕ can be calculated on $\mathbb{R}^+ \times \Omega$. Once this calculation is done, one just needs to extract the zero level set of ϕ to obtain the curve. A particular case is motion by mean curvature, when

$$F = \operatorname{div} \left(\frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right), \quad (1.5)$$

is the curvature of the level-curve of ϕ passing through pixel x (Osher and Sethian, 1988).

Caselles et al. (1993) proposed a *geometric active contour* model based on the theory of curve evolution and geometric flows. A general edge detector is defined by a positive and regular monotonic decreasing function g , depending on the image gradient. The function g can be chosen as

$$g(|\nabla I|) = \frac{1}{1 + |\nabla G * I|^2}, \quad (1.6)$$

where $G * I$ is the convolution of the image I with the Gaussian G . Thus, $g(0) = 1$ and $\lim_{s \rightarrow \infty} g(s) = 0$.

Then equation (1.4) can be modified as

$$\frac{\partial \phi}{\partial t} = g(|\nabla I|) |\nabla \phi| \left(\operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \nu \right), \quad (1.7a)$$

$$\phi(0, x) = \phi_0(x), \quad (1.7b)$$

where ν is a balloon force (Cohen, 1991), and ϕ_0 is the initial level set function. Hence, the geometric active contour model benefits from the flexibility of level sets in detecting objects with arbitrary topology. Note that if $g = 1$ and $\nu = 0$, equation (1.7) reduces to the simple case controlled only by mean curvature.

The geometric approach described above constructs directly the evolution equation, or equivalently the force on the curve. In practice, g is never zero on the edges, and therefore the evolving curve may not be stopped on the desired boundary. To overcome this problem, as an extension of the previous geometric approach, Caselles et al. (1997) proposed *geodesic active contours*. The geometric flow is obtained indirectly from the minimization of an energy functional. Starting from the basic snakes (Kass et al., 1988), Caselles et al. (1997) neglected the second-order internal term (*i.e.* $\beta = 0$), and replaced the minus image gradient $-|\nabla I|$ by the more general edge detector $g(|\nabla I|)^2$, where $g(|\nabla I|)$ has been defined in equation (1.6). The resulting energy functional is

$$E(C) = \alpha \int_0^1 |C'(p)|^2 dp + \lambda \int_0^1 g(|\nabla I(C(p))|)^2 dp. \quad (1.8)$$

Equation (1.8) is still not intrinsic, because it depends on the parameterization p of the curve. Caselles et al. (1997) defined a new parameterization of the curve and substituted p

in equation (1.8). The energies can be then changed in any arbitrary form. The equivalent functional to be minimized becomes

$$E(C) = \int_0^1 g(|\nabla I(C(p))|) |C'(p)| dp = \int_0^{L(C)} g(|\nabla I(C(s))|) ds . \quad (1.9)$$

This means that the boundary is a geodesic curve in a Riemannian space with a metric derived from the image intensity. Hence, this approach is named geodesic active contours. Note that equation (1.9) has a bad property, *i.e.* it is always non-negative, so the global minimum is always zero, which cannot be achieved unless the contour vanishes completely. By means of a gradient descent algorithm, the curve evolution equation is

$$\frac{\partial C}{\partial t} = g(|\nabla I|) \kappa \vec{N} - (\nabla g(|\nabla I|) \cdot \vec{N}) \vec{N} , \quad (1.10)$$

where κ is the Euclidean curvature, and \vec{N} is the unit inward normal vector. Equation (1.10) can easily be reformulated in terms of the level set function:

$$\frac{\partial \phi}{\partial t} = g(|\nabla I|) |\nabla \phi| \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nabla g(|\nabla I|) \cdot \nabla \phi . \quad (1.11)$$

This evolution equation of the geodesic curve via a geometric flow is very similar to the one obtained in equation (1.7) by the geometric approach mentioned earlier. However, this geodesic flow includes a new component in the curve velocity. This new term increases the attraction of the evolving curve toward the object boundary, and is of special help when the boundary has high variations, including gaps, in its gradient values. Similarly to geometric active contours, the balloon force ν can also be incorporated into equation (1.11) to increase the convergence speed. Similar flows were also proposed independently by Kichenassamy et al. (1995). To combine the advantages of the geodesic active contour and the GVF external force, Paragios et al. (2004) integrated them within a level set formulation, and proposed a geometric boundary-based flow for boundary extraction and image segmentation.

1.1.2 Region-Based Active Contours

In spite of the great efforts dedicated to edge-based active contours, they are still limited to relatively simple images. The main difficulty in this approach is that the model incorporates purely *local* characteristics of the data. More specifically, it models only a tiny fraction of the data: the data near the region boundary; and it fails to distinguish between the interior and the exterior of regions, and thus information about pixel positions is lost except when the pixels are close to a boundary. As a result, the initial contour has to be very near the ideal boundary, and the model is very sensitive to noise. In order to introduce more *global* information about the different regions to be segmented, region-based active contours were proposed. Regions are characterized by *global* descriptors associated with each separate region. The image domain is partitioned by progressively fitting statistical models to the descriptors. Generally, a descriptor is a statistical feature of a region, such as the intensity, color, histogram, texture, or motion in each of a set of regions. In the case of unsupervised

segmentation, these descriptors can be recalculated each time the active contour evolves, and thus they vary during the propagation of the active contour. Here, we present the two main region-based segmentation models: *the Mumford-Shah functional based model*, and *the Bayesian approach based model*.

Mumford-Shah Functional

In order to solve the segmentation problem, Mumford and Shah (1985, 1989) proposed to compute the optimal approximation of an image as a combination of regions of piecewise smooth intensities. In a variational framework, after concerning boundary smoothness, for a given image I , the authors searched for a pair (u, C) such that u is a piecewise smooth approximation of I and $C \subset \Omega$ is a 1D subset of edges. Thus, they defined the functional as

$$E(u, C) = \lambda \int_{\Omega} |I(x) - u(x)|^2 dx + \int_{\Omega \setminus C} |\nabla u(x)|^2 dx + \mu L(C), \quad (1.12)$$

where $L(C)$ denotes the boundary length. The first term causes u to approximate the image I ; the second term makes u (and hence I) not varying very much on each disjoint region; and the third term forces the boundaries C to have minimal length. The solution image u obtained by minimizing the above functional is formed by smooth disjoint regions with sharp boundaries C . Mumford and Shah (1989) also pointed out that a reduced version of this model is simply the restriction of equation (1.12) to piecewise constant functions, *i.e.* u is a constant on each disjoint region.

Chan and Vese (2001a) proposed a particular case of the Mumford-Shah model. They considered binary segmentation, and reduced the best approximation u of the image I to a function taking only two values, *e.g.* the average of I inside and outside the segmented region, and with one edge C . Thus, the energy functional in the level set representation is written as

$$\begin{aligned} E(c_{\text{in}}, c_{\text{out}}, \phi) = & \mu \int_{\Omega} |\nabla H(\phi(x))| dx \\ & + \lambda_{\text{in}} \int_{\Omega} |I(x) - c_{\text{in}}(\phi(x))|^2 H(\phi(x)) dx + \lambda_{\text{out}} \int_{\Omega} |I(x) - c_{\text{out}}(\phi(x))|^2 (1 - H(\phi(x))) dx, \end{aligned} \quad (1.13)$$

where H is the Heaviside function. The values of c_{in} and c_{out} , depending on the evolving curve, are the averages of I inside and outside the active contour. The minimization of E at each iteration is done in two steps. First, keeping ϕ fixed and minimizing E with respect to c_{in} and c_{out} , these two constants can be expressed by

$$c_{\text{in}}(\phi(x)) = \frac{\int_{\Omega} I(x) H(\phi(x)) dx}{\int_{\Omega} H(\phi(x)) dx}, \quad (1.14a)$$

$$c_{\text{out}}(\phi(x)) = \frac{\int_{\Omega} I(x) (1 - H(\phi(x))) dx}{\int_{\Omega} (1 - H(\phi(x))) dx}. \quad (1.14b)$$

Second, keeping c_{in} and c_{out} fixed, and minimizing E with respect to ϕ , the level set function is updated by

$$\frac{\partial \phi}{\partial t} = \left\{ \mu \operatorname{div} \left(\frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) - \lambda_{\text{in}} \left(I(x) - c_{\text{in}}(\phi(x)) \right)^2 + \lambda_{\text{out}} \left(I(x) - c_{\text{out}}(\phi(x)) \right)^2 \right\} \delta(\phi(x)) , \quad (1.15)$$

where δ is the 1D Dirac measure. Compared to the previous edge-based approaches, the initial curve of this model does not necessarily have to start around the objects to be detected and instead can be placed anywhere in the image. Chan et al. (2000) applied a similar model to vector-valued images, by considering each component of the image.

Chan and Vese (2001b); Vese and Chan (2002) generalized the active contour model without edges two-phase segmentation (Chan and Vese, 2001a) to piecewise smooth optimal approximation and/or multiphase segmentation. For the former extension, a general smooth function is considered instead of a constant to model image intensity inside each region. Thus, different regions of distinct intensities can be represented and detected with the correct intensities. For the latter extension, the idea is to use a combination of level set functions to represent the boundaries. In the piecewise constant case, only $\log_2 n$ level set functions are needed to represent n phases or segments with complex topologies, such as triple junctions. While in the piecewise smooth case, using only two level set functions, producing up to four phases, any general case can be considered and represented thanks to the Four-Color Theorem (Appel and Haken, 1989). This brings a big advantage that knowing a priori how many segments the image has is not necessary.

Region-Based ‘Bayesian’ Formulation

Paragios and Deriche (2002) presented a unified approach, namely the *geodesic active region* model to deal with frame partition problems. This model combines the edge-based functional from the geodesic active contour model, and the region-based functional derived from Maximum a Posteriori (MAP). The boundary term is similar to geodesic active contour proposed by Caselles et al. (1997). The boundary is regular, of minimal length, and attracted by the real region boundary. The region term aims at finding a contour C that maximizes two posterior probabilities of image intensities inside C and outside C . To simplify the problem, Paragios and Deriche (2002) made two assumptions: 1) all the contours are equally possible, *i.e.* $P(C) = 1/Z$, where Z is the total number of possible contours; 2) there is no correlation between the regions labeling, and all the pixels within each region are identically and independently distributed. Taking these into account, based on the Bayes’ theorem, the *a posteriori* segmentation probability for a contour C given the observed image I is determined by

$$P(C|I) = \prod_{x \in \Omega_{\text{in}}} P_{\text{in}}(I(x)) \prod_{x \in \Omega_{\text{out}}} P_{\text{out}}(I(x)) . \quad (1.16)$$

Maximization of the *a posteriori* segmentation probability is equivalent to minimizing its negative logarithm. The geodesic active region objective function is defined as

$$E(C) = - \int_{\Omega_{\text{in}}} \log(P_{\text{in}}(I(x))) dx - \int_{\Omega_{\text{out}}} \log(P_{\text{out}}(I(x))) dx + \mu L(C) . \quad (1.17)$$

This functional is minimized using a gradient descent method. Using the level set representation, the curve evolution for a given point x is

$$\frac{\partial \phi}{\partial t} = \left\{ \log \left(\frac{P_{\text{out}}(I(x))}{P_{\text{in}}(I(x))} \right) + \mu \kappa(x) \right\} |\nabla \phi(x)|. \quad (1.18)$$

For any given pixel x on the boundary, if the true state of x belongs to Ω_{out} , then $P_{\text{out}}(I(x)) > P_{\text{in}}(I(x))$, resulting in a positive force that aims at shrinking the curve to pass through this pixel. On the other hand, if the true state of x belongs to Ω_{in} , the region force aims at expanding the curve to include this pixel.

In (Paragios and Deriche, 2002), the probability densities are supposed to follow a normal distribution¹:

$$P(I(x)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(I(x)-c)^2}{2\sigma^2}}, \quad (1.19)$$

where c is the average, and σ^2 is the variance. These parameters are learned in advance through a supervised way. In (Rousson, 2004), P also follows the normal distributions, but the parameters are dynamically calculated during the contour evolution. Using the form of the Heaviside function in the Chan-Vese model, the energy functional becomes

$$\begin{aligned} E(c_{\text{in}}, c_{\text{out}}, \sigma_{\text{in}}^2, \sigma_{\text{out}}^2, \phi) = & \int_{\Omega} \left\{ \frac{(I(x) - c_{\text{in}}(\phi(x)))^2}{2\sigma_{\text{in}}^2(\phi(x))} + \log((2\pi)^{1/2}\sigma_{\text{in}}(\phi(x))) \right\} H(\phi(x)) dx \\ & + \int_{\Omega} \left\{ \frac{(I(x) - c_{\text{out}}(\phi(x)))^2}{2\sigma_{\text{out}}^2(\phi(x))} + \log((2\pi)^{1/2}\sigma_{\text{out}}(\phi(x))) \right\} (1 - H(\phi(x))) dx + \mu L(C), \end{aligned} \quad (1.20)$$

where c_{in} and c_{out} are updated following equation (1.14), and σ_{in}^2 and σ_{out}^2 are estimated by

$$\sigma_{\text{in}}^2(\phi(x)) = \frac{\int_{\Omega} (I(x) - c_{\text{in}}(\phi(x)))^2 H(\phi(x)) dx}{\int_{\Omega} H(\phi(x)) dx}, \quad (1.21a)$$

$$\sigma_{\text{out}}^2(\phi(x)) = \frac{\int_{\Omega} (I(x) - c_{\text{out}}(\phi(x)))^2 (1 - H(\phi(x))) dx}{\int_{\Omega} (1 - H(\phi(x))) dx}. \quad (1.21b)$$

The evolution equation derived from the gradient descent method is

$$\frac{\partial \phi}{\partial t} = \left\{ -\frac{(I(x) - c_{\text{in}}(\phi(x)))^2}{2\sigma_{\text{in}}^2(\phi(x))} + \frac{(I(x) - c_{\text{out}}(\phi(x)))^2}{2\sigma_{\text{out}}^2(\phi(x))} + \frac{1}{2} \log \left(\frac{\sigma_{\text{out}}^2(\phi(x))}{\sigma_{\text{in}}^2(\phi(x))} \right) + \mu \kappa(x) \right\} \delta(\phi(x)). \quad (1.22)$$

¹Instead of modeling $P(I(x))$ as parametric distributions, non-parametric density estimation, like the Parzen window method (Kim et al., 2002) can be also used, which we do not detail further.

For $\sigma_{\text{in}}^2 = \sigma_{\text{out}}^2 = \text{constant}$, equation (1.22) reduces to the evolution equation of the Chan-Vese model in equation (1.15). In the early *region competition* approach proposed by Zhu and Yuille (1996), they obtained a similar functional as equation (1.22), but there, the energy functional was deduced from the minimum description length (MDL) method, rather than MAP.

1.1.3 Active Contours with Shape Constraints

In the previous subsections, all the presented active contours, whether edge-based or region-based, drive the curve evolution based on the information from the image, except perhaps for having some regularization term to smooth boundary. In this case, little prior knowledge about the object exists in the model, and hence the data quality has a large impact on the segmentation accuracy. Unfortunately, in many real applications, the image data is often damaged by noise or partial occlusion; or the contrast between the object and the background is poor. These factors may lead to unsuccessful segmentation. Recently, a large amount of work on the inclusion of shape knowledge in active contour models has been presented. Having prior information about the expected shape of the object can significantly increase the robustness of the segmentation algorithm. In general, in a variational framework, a prior energy E_P related to the shape knowledge is integrated together with a data energy E_D describing the image information. A constant weight is normally used to balance the contributions of these two terms. Note that Bailloeuil (2005) converted the constant weight into a space function, to relax the influence of the constraints near the zero level set of the prior shape, while keeping a strong and uniform constraint far from the prior shape. This makes the active contours resemble the prior shape globally, but with some variations locally.

Chen et al. (2001, 2002) considered a non-probabilistic model that minimizes an energy functional depending on the image gradient and the shape of interest. The basic idea is to influence the geometric active contour with a vector field that depends on the shape prior. The shape prior energy is formulated as

$$E_P(C, T) = \int_0^1 d^2(T(C(p))) |C'(p)| dp, \quad (1.23)$$

where $d(x) = d(C^*, x)$ is the distance of the point x from the shape reference C^* . Here T represents scale, rotation and translation transform. Therefore, this term evaluates the similarity of the shape of the contour C to C^* , but making use of a single shape reference (though with different size, orientation and translation), this model cannot handle a large variation of shapes. To solve this limit, the authors suggested grouping the sampled curves into k clusters for some k , and finding the average shape C_i^* ($i = 1, \dots, k$) in each group. However, this model is still sensitive to the initialization because of the inclusion of an edge-based data term.

Leventon et al. (2000) extended geometric active contours (Caselles et al., 1997; Kichenasamy et al., 1995) by incorporating shape information into the evolution process. They computed a statistical shape model over a training set of curves, in which principal component analysis (PCA) is used to derive the principal deformation modes. The segmentation

process is implemented in the level set framework, and the level set surface is driven solely by a data term in the first step. At each step of the evolution, the MAP position and shape of the object in the image are estimated, based on the prior shape information and the image information. In the second step, a correction of the previous segmentation is performed locally based on image gradients and curvatures, and globally towards the MAP estimate. Thus, an additional shape influence term is introduced to the original evolution equation of geometric active contours.

Rousson and Paragios (2002) also constructed a statistical shape model using a collection of samples. The differences with (Leventon et al., 2000) lie in two aspects. First, the shape constraint is expressed by a shape energy functional. Thus, the complete evolution equation can be derived directly from the overall energy. Second, since the learning shapes are aligned with a reference shape selected arbitrarily from samples, the model can handle shapes with a large variety of global transformations as well as scale variations. The shape constraint energy is defined, in the level set representation, as

$$E_P(\phi, T, s) = \int_{\Omega} \left\{ \frac{(s\phi(x) - \phi_M(T(x)))^2}{2\sigma_M^2(T(x))} + \log(\sigma_M(T(x))) \right\} H(\phi(x)) dx, \quad (1.24)$$

where $\phi_M(x)$ and $\sigma_M^2(x)$ denote a representative shape and a confidence map respectively; T and s are a global transformation and a scale factor. The first term constrains the surface to be likewise the prior, and weighs their difference by the confidence; and the second term penalizes the shape with low confidence.

Cremers and Soatto (2003) pointed out the following drawbacks of the shape constraints in (Rousson and Paragios, 2002) (equation (1.24)). First, the restriction of the energy integral to the positive part of the level set function induces a bias toward shrinking area of the shapes. Second, all shape discrepancies outside the evolving shape are neglected. Such priors are therefore not well-suited to encode multicomponent shapes. To address these problems, they proposed a symmetric dissimilarity measure between the evolving contour ϕ and the given reference shape ψ_0 :

$$E_P(\phi, T) = \int_{\Omega} (\phi(x) - \psi_0(T(x)))^2 \frac{h(\phi(x)) + h(\psi_0(T(x)))}{2} dx, \quad (1.25)$$

where h is the normalized Heaviside function. This symmetrized dissimilarity measure averages the squared deviation of shapes over both areas $\phi(x)$ and $\psi_0(T(x))$. The authors proved that the constructed dissimilarity measure is in fact a pseudo-distance. A limitation of this approach is that the shape prior encodes a single shape. Moreover, pose optimization may result in numerous instabilities. Chan and Zhu (2003) proposed another symmetric constraint energy:

$$E_P(\phi, T) = \int_{\Omega} (H(\phi(x)) - H(\psi_0(T(x))))^2 dx. \quad (1.26)$$

This term is the area of the symmetric set difference. It does not enforce the evolving level set to resemble the shape reference, instead it demands similarity of the regions within

the respective contours. Riklin-Raviv et al. (2004) integrated perspective (a six-parameter transformation) between a single prior image and the image to be segmented, by slicing the signed distance function at various angles. Riklin-Raviv et al. (2007) further generalized this algorithm to an eight-parameter projective model.

To resolve the issue of pose estimation, Cremers et al. (2006) evaluated the evolving level set function not in global coordinates, but in coordinates of an intrinsic reference frame attached to the evolving surface, thereby obtaining invariant shape distances. Such a closed-form solution removes the need to iteratively update local estimates of pose parameters. For the explicit contour representation, an analogous intrinsic alignment with respect to similarity transformations was proposed in (Cremers et al., 2002). Rousson and Paragios (2007) also addressed the construction of a similarity invariant shape constraint. The similarity is considered not between the evolving level set and the reference shape as usual, but alternatively between the evolving level set and the image. The evolving level set remains in the same pose as the prior. The evolution equations of pose parameters are only guided by the image term, resulting in a much faster and robust estimation of the transformation parameters.

So far, statistical shape priors in the above models are *static* in time. It is worth to mention that Cremers (2006, 2007) took into account *dynamical* statistical models for implicitly represented shapes, in the context of image sequence segmentation. The dynamical prior is based on the fact that the probability of a contour depends on which contours have been observed in previous frames. The respective shape models capture the temporal correlations among silhouettes which characterize many deforming shapes in the presence of noise and occlusion.

1.1.4 Summary

In this section, we have given a brief survey of active contours in segmentation. In general, active contours can be classified into two different categories: *edge-based* and *region-based*. Edge-based active contours consider some *local* measurements around the boundary of the object to be segmented. This is responsible for many limitations of this approach. Between the two types of edge-based active contours, *geometric active contours* have many advantages over *parametric active contours*, such as computational simplicity and the ability to change curve topology during evolution. Note that the precise relationship between the two was derived by Xu et al. (2000). On the other hand, thanks to the introduction of *global* information, region-based active contours tend to be more robust to noise. Correspondingly, the energy functional tends to have less local minima, which makes them particular well-suited for local optimization methods such as the level set framework. Cremers et al. (2007) presented a comprehensive survey of the class of region-based level set segmentation methods, and detailed how they can be derived from a general statistical framework. Rousson (2004) pointed out that the minimal partition limit of the *Mumford-Shah functional* is a particular example of the *region-based 'Bayesian' formulation*. In addition, we have investigated a special case: active contours with shape constraints, because it is where our interests lie. Active contours with shape constraints have drawn great attention from the computer vision community, due to their potential for increasing algorithm robustness. They can

be classified into *non-statistical* approaches and *statistical* approaches. The difference is whether the shape reference is learned from a training set or not. Many researchers have also developed different ways to tackle the problem of pose estimation (Cremers et al., 2002, 2006; Rousson and Paragios, 2007).

1.2 Road Extraction

For the last few decades, road detection from remotely sensed imagery has been extensively studied, due to the variety and importance of the potential applications of an automatic extraction method. A great number of approaches have been proposed (Fortier et al., 1999; Mena, 2003). They consider different image resolutions (low, medium and high resolutions), and different image complexities (rural area, semi-urban area, and urban area). Furthermore, a wide range of sensors have been utilized: the image data includes monochromatic imagery, infrared band, color imagery (RGB), multi- and hyper-spectral imagery (HYDICE), synthetic aperture radar imagery (SAR), light detection and ranging imagery (LIDAR), and so on.

Despite all these works, there is no widely recognized solution to the road extraction problem. The development of reliable procedures is still a challenge. The difficulty stems from two aspects:

- The large variety of objects in the data: the appearance of road networks differs according to the road type, their surroundings and the construction layout of different countries. Moreover, the same network appears differently according to the view angle, the satellite sensor, spatial resolution, spectral wavebands, etc.
- The complexity of the imaged scene: it contains many different objects, such as roads, buildings, trees, etc., with differences in shape, brightness, and texture.

To classify and analyze existing algorithms, several criteria can be chosen: 1) Some methods are semi-automatic, in which interaction with a human operator is more or less necessary. Others aim at being “automatic”, but until now, automatic methods for road extraction seem to be far from mature. As far as we know, at least, model parameters still need to be adjusted manually. 2) Previous work deals with different resolutions of the image, or uses a multiresolution analysis (MRA) to combine road extraction at different resolutions. 3) We can also characterize previous work into ‘bottom-up’ and ‘top-down’ strategies, which are elaborated in the following.

A *bottom-up* strategy attempts to answer the question: what is where? In a hierarchical scheme, we first detect basic features in different locations, and then progressively add constraints, such as the nature of the object, up to higher-level recognition. As opposed to a bottom-up strategy, a *top-down* strategy is shifted to answer the question: where is what? Using such a strategy, we first define what the object to search for should be, and then try to find out its location in the image. In fact, since most methods make use of a combined control strategy, in which the bottom-up process provides local information, and the top-down process provides global knowledge of the properties of objects, the borderline between the two strategies is not clear. Moreover, if one concentrates on models instead

of algorithms as described above, the distinction disappears in principle, because the only question left is: is the algorithm good to solve the model?

In the rest of this section, we first describe the common physical properties of roads. We provide a quite detailed survey of work on road extraction and updating, and classify them according to the applied extraction techniques.

1.2.1 Road Characteristics

In spite of the great variety in road shape and appearance, we can summarize several “universal” characteristics that might be used for the extraction process. The different characteristics are as follows (Fortier et al., 1999):

1. Geometric properties

- The width of a road is almost constant locally.
- The curvature of roads in urban areas has some changes globally, but it is small locally.
- The steepness of roads has an upper bound. In urban areas, it is rather small locally even near bridges; in mountainous regions, roads tend to be winding. However, the steepness is not easily measurable unless stereoscopic pairs or other 3D information are available.

2. Spectral properties

- For high resolution images, the surface of a main road is not smooth. There are some objects, such as zebra crossings, over-bridges, vehicles, shadows, road signs, etc.
- It is more complex in the case of small roads. Some parts of the road may be sheltered completely by the trees and building shadows beside the road.
- Roads are mainly built using materials such as concrete or asphalt; consequently unsheltered roads have a spectral signature corresponding to these materials.
- Other objects may have a spectral signature similar to roads. A sheltered road has the spectral signature of the shelters.

3. Topological properties

- Roads are built to link certain places together and neighboring roads are connected to form networks.
- Roads emerged recently are inevitably linked with the previous network, which provides some hints for finding new roads.

4. Contextual properties

- The width of a road has an upper bound. This upper bound is highly dependent on the importance of the roads. For instance, highways are much wider than rural or semi-urban roads.

- The density of road networks and junctions depends on the surrounding context. It is higher in urban areas than in rural areas.

The above properties are more or less embodied in the techniques proposed in the literature for road extraction.

1.2.2 Active Contours

Active contours provide a convenient way of embedding geometric properties of the features in the form of constraints to guide the recognition process. Hence, they have been widely used for the extraction of man made structures.

Klang (1998) developed an automatic procedure for detecting changes between an existing road database and a new 10m resolution satellite image in a forestry area. First he used the existing data to initialize an optimization process, using the ‘ziplock snakes’ approach to correct road location. Then, he ran a line-following process using a statistical approach to detect new roads, starting from the existing network. Fortier et al. (2001) improved this approach by considering road junctions. Road junctions are generally reliable information and since roads form networks, they are very relevant in this context. In the scheme of road position correction, line junctions are added to re-localize road location from the road database, in order to bring the initial snakes closer to the real contour. To generate hypotheses for new roads, a road following algorithm is applied in the image, starting from line junctions near the known road network.

Mayer et al. (1998); Laptev et al. (2000) proposed an approach for automatic road extraction in aerial imagery. The approach takes advantage of the scale-space behavior of roads in combination with geometrically constrained edge extraction by means of snakes. Based on the scale-space behavior of roads, the procedure starts by extracting lines at coarse scale, which are less precise but also less disturbed than features at fine scale. The lines are used to initialize ‘ribbon snakes’ at fine scale, which describe the roads as bright, more or less homogeneous elongated areas. Ribbons with constant width are accepted as salient roads. The connections between adjacent ends of salient roads are checked if they correspond to non-salient roads. For the sections of non-salient roads which include high noise, additional constraints and a special strategy are exploited to close gaps that arise from shadows and partially occluded areas. However, the proposed approach is mostly restricted to rural areas.

Bonnefon et al. (2002) presented a complete process to update – by matching –, and upgrade – via detection – geographic linear features in a Geographical Information System (GIS). In its first step, linear objects on SPOT images are matched with GIS vectors using a combination of dynamic programming and a deformable contour model. The dynamic programming process first finds quickly the optimal path, but with some erratic sections due to the lack of curvature constraints. Then the snakes smooth the erratic sections and control the curvature. As the optimal path is very close to the objects, the process converges fast without being attracted to surrounding local minima.

A modification of the classic concept of snakes was presented in (Agouris et al., 2001). In this paper, a novel framework, comprising change detection and versioning (which means

in the absence of changes, the pre-existing information may be improved in terms of accuracy), is introduced. Based on fuzzy logic, a criterion of quality is defined. The standard snake method is extended to function in a differential mode by introducing an energy term to describe the discrepancy between the present extracted result and the prior information. Change is detected if and only if the new image supports the notion that the object has moved beyond the stochastic range of the older information. Otherwise, the object is treated with versioning. The system has the ability to differentiate change detection from the recording of numerous slightly different versions of objects that may remain unchanged. This framework is applied to 5m resolution images in rural areas.

Péteri and Ranchin (2003) addressed the problem of extracting the road network from Ikonos satellite images in dense urban areas. Based on a given topologically correct graph of the network, roads are reconstructed as surface elements using two different types of active contours. For extracting road portions with parallel sides, geometric knowledge is introduced via a parallelism constraint on the contours representing the road borders. An MRA is used to minimize geometric noise. Once road segment extraction is finished, the junctions are extracted by simple active contours, which are initialized by extremity points on parallel road borders. This method can provide precise road areas, but the retrieval of a topologically correct graph is critical.

Youn and Bethel (2004) assumed that the road network and block pattern in a city have a semi-regular grid pattern. Based on this, the image is segmented according to dominant road directions. Then the road lines are detected with the ‘acupuncture method’. These detected lines are used to construct initial approximations for the subsequent snake refinement. Finally, the road corners are rectified by applying adaptive snakes.

1.2.3 Multiscale and Multiresolution Analysis

A multiscale approach offers an appropriate framework to fuse the information from different scales. This is particularly important for high resolution images, because the road extraction task can be greatly facilitated with the aid of road detection results at reduced resolution, which needs less effort at the cost of probable mistakes.

Baumgartner et al. (1997) presented a multiresolution approach for automatic extraction of roads from aerial images. For different context regions, *i.e.* rural, forest, and urban areas, the model describes relations between background objects and road objects. The approach to road detection is based on the extraction of edges in a high resolution image and the extraction of lines in a reduced resolution image. Using both resolution levels and explicit prior knowledge about roads, hypotheses for road borders are generated. Quadrilaterals and polygons are constructed to represent road-parts and intersections. Neighboring road-parts are chained to form road-segments. Road-links are built by grouping road-segments and closing gaps.

For high resolution images, such as airborne images, Benharrosh (1998) aimed to obtain a multiresolution image with some parts at full resolution and others at coarser resolution. They proposed three methods of adaptive filtering that smooth an image while preserving cartographic structuring features or interesting regions of the image. The first one relies on an anisotropic diffusion method introduced by Saint-Marc et al. (1991). The second is based

on wavelet theory. And the third method is based on an analogy between the image and an electrical network. These methods are applied to the generation of an adaptive quick-look and adaptive compression.

Couloigner and Ranchin (2000) defined a hierarchical method to extract urban road networks semi-automatically from 2m resolution space-borne images. This method is performed in two steps. First, given two endpoints on both sides of the road, the two sides of each road are located from the original image and the scaling coefficients of a wavelet transform at different spatial resolutions. In the second step, the strip(s), if existing, is extracted using the detected road borders and the wavelet coefficients at characteristic scales.

Prinet et al. (2000) addressed scale selection for curvilinear structure detection. This approach is based on the study of the existence and stability properties of the crest-lines that localize the center axis of the structures to be retrieved. The authors showed that the crest-lines exist and are stable over a band of width scale values. The selection of different scales according to different local resolution of the structure is followed by a fusion scheme that enables the merging of crest-lines.

Zhang and Couloigner (2003) also made use of an analysis in the wavelet domain to detect road features; they aimed at providing an automatic method. In their framework for road network change detection, road sides, centerlines and junctions are extracted based on the observation of wavelet-based road features. Along the road gradient direction, the road centerline and the road borders correspond to respectively a local maximum and two zeros of wavelet coefficients; a road junction is localized by a local maximum of wavelet coefficients in a 7×7 neighborhood; the linked road branches are identified as the local maximum pixels along the border of the above neighborhood. However, this wavelet-based method is restricted to simple scenes such as countryside.

Note that the work of Mayer et al. (1998), Laptev et al. (2000) and Péteri and Ranchin (2003), introduced in subsection 1.2.2 on active contours, also used a multiscale or multiresolution strategy.

1.2.4 Dynamic Programming

Dynamic programming formulates an extraction problem as the minimization of a cost function defined on a graph. Given knowledge about the presence of roads in the form of endpoints, usually fixed manually, the algorithm is forced to track a connecting path between them through the image which best fits the model. Since this method restricts the topology of networks, dynamic programming is potentially semi-automatic. In this sense, a more reliable handling of obstacles is possible.

Fischler et al. (1981) aimed to extract roads from low-resolution aerial images. Different low level operators are classified into two types: Type I operators identify roads with a high accuracy, but some roads may be missing; Type II operators extract all roads completely, but may yield some false detections. By combining them, a cost array is defined. Between two given endpoints, the best path, which gives the lowest cost, is chosen as the road using the F^* algorithm.

Merlet and Zerubia (1996) extended the F^* algorithm of Fischler et al. (1981) to cliques and to neighborhoods larger than one. By means of the cliques of more than two points,

contrast information is introduced into the calculation of the minimum cost path, and the larger neighborhoods allow for the consideration of the curvature of the final path. All the needed information are synthesized in a unique cost function. Thanks to the curvature constraint, even if the initial endpoints are incorrect, or are located on different roads, a correct path may still be obtained. The method is applied to detect roads and valleys from low-resolution SPOT satellite images.

Barzohar and Cooper (1996) presented an automatic approach to find main roads in aerial images. Geometric-stochastic models are built for representing road images, and then MAP estimation is used for estimating the road boundaries in an image. First, the image is partitioned into windows. They searched the small windows to obtain initial road candidates by dynamic programming. Then, starting with the windows containing high confidence estimates, they obtained optimal global estimates of the roads using dynamic programming again. The algorithm produces two boundaries for each road.

Bonnefon et al. (2002) proposed a semi-automatic tracking method using the F^* algorithm. From a starting point, they used a small search window in which the algorithm tries to find the optimal path. Costs for the F^* algorithm depend on the difference of radiometry values between the original SPOT image and the detection image computed from it. The last point of the optimal path in the search window becomes the new starting point. The algorithm identifies the detected linear features using texture information from high-resolution images.

Dal Poz and do Vale (2003) presented a dynamic programming approach for road centerline extraction from medium- and high-resolution images. This approach is a modification of a pre-existing dynamic programming approach, proposed by Gruen and Li (1997), for road extraction from low-resolution images. A constraint that the gradients at road edges are antiparallel is incorporated into the cost function. This allows the approach to treat the road as a ribbon feature.

Bucha et al. (2006) also used dynamic programming to extract road centerlines, but the weights of the edges in the graph are provided by a force field, drawn such that at each pixel, a two-dimensional vector defines interactions between pixels. In such an image representation, the vector is oriented to the center of the region composed of pixels having the same qualitative property, such as color and gray-scale level.

1.2.5 Morphological Methods

Mathematical morphology is a theory for the analysis and processing of signals in terms of geometrical structures, developed by Matheron (1975) and Serra (1982). Mathematical morphology can characterize topological and geometrical concepts such as size, shape, convexity, connectivity, and geodesic distance, on both continuous and discrete spaces. Morphological image processing is based on this theory. It consists of a set of operators, *e.g.* union, intersection and complementation, as well as dilation, erosion, openings, closings, thinning and other derived operators. These operators transform the image according to the above characterizations. Mathematical morphology was originally proposed to process binary images, and was later extended to gray-scale images and multi-band images.

Zhang et al. (1999) proposed an approach for detecting road networks from high resolution images using a combination of mathematical morphology operations. In the pre-processing stage, the image is segmented and road network regions are separated from their surroundings. Morphological trivial opening is then adopted. A criterion T is defined as the long axis of the minimal ellipse which encloses an object. Only the connected components that satisfy the criterion T are retained after morphological reconstruction. Thus, this process preserves the elongated road areas, and filters out almost all the houses and small clusters of noise as well. The result is further refined by filling holes, removing small paths and recovering shadowed areas. In the high resolution simulation images and aerial photos, the approximate road centerline network is finally extracted. However, road gaps may still exist, if road surfaces are completely broken and there is no other information supporting the linkage.

Chanussot and Lambert (1998) introduced a simple and fast unsupervised method for the automatic extraction of road networks in SAR images. A series of morphological operators is used in order to retain elongated structures with a specific width. The sequence of morphological filtering consists of an opening by reconstruction, a directional closing in 40 successive directions, an opening, and a closing top-hat operator. At every step, flat structuring elements are used, and their size is specified according to *a priori* information about the road's maximum width and curvature. In the final stage, the roads are extracted by a simple thresholding applied to the response of the morphological operators. The drawback of this method is that the lack of contextual knowledge results in partial detection of the road network together with several spurious detections.

Katartzis et al. (2001) combined and extended two earlier approaches for road detection in SAR satellite images, and described a model based method for the automatic extraction of linear features from aerial imagery. In fact, the authors made use of both bottom-up and top-down processing, *i.e.* a combined strategy. During its first local analysis step, to detect elongated structures, a set of morphological operators proposed by Chanussot and Lambert (1998) is modified to enhance the performance of the morphological filtering in the case of heavily noisy environments and partially disconnected roads. In its second global analysis step, a segment linking process is performed with some prior information, based on the Bayesian framework of Tupin et al. (1998) (which will be introduced in subsection 1.2.6).

In his survey of previously existing applications of mathematical morphology in geoscience and remote sensing, Soille and Pesaresi (2002) analyzed the suitability of morphological operators for the processing of Earth observation data, detailed some new advances in the theory of mathematical morphology, and demonstrated their efficiency for extracting structural information from Earth observation data.

1.2.6 Markov Random Fields and Marked Point Processes

Contextual constraints are a general and powerful way to model spatial properties. Markov random fields (MRFs) provide a convenient and consistent way to model context-dependent entities such as image pixels and correlated features. MRF based models have been widely used to identify road networks.

Tupin et al. (1998) used both local and global techniques for linear feature extraction. First, they performed a local detection of linear structures based on the fusion of the results from two line detectors, both taking into account the statistical properties of speckle in SAR images. The masks of the line detectors have widths ranging from one to a maximal number of five pixels. The produced candidate road segments are then organized as a graph, together with an additional set of segments that correspond to all possible connections between them. The road identification is solved by the extraction of the best graph labeling based on an MRF model for road like structures and a MAP criterion. As a result, this algorithm allows complex topology and *a priori* detection; but its drawback is that the road network has to be represented as a graph, and the number of nodes is fixed.

Stoica et al. (2004) modeled thin networks as ensembles of line segments embedded in the image domain. A point process controls network parameters such as connectivity and curvature. The road network is approximated by connected line segments under constraints enforced by the interaction model. The specific properties of the road network in the image are described by the data term. The probabilistic model is solved by simulated annealing based on a Reversible Jump Markov Chain Monte Carlo algorithm. The main advantage is that during the optimization process, new segments can be created, and their location and orientation can be changed. Unlike the method of Tupin et al. (1998), the dimension of the space is not fixed. Results are shown on SPOT, ERS and aerial images. Lacoste et al. (2005) developed an extension of this model.

Negri et al. (2006) proposed a general processing framework for urban road network extraction in high-resolution SAR images. First, road candidates are extracted with two detectors. Then, road network topology is optimized with an MRF model, incorporating the prior knowledge about road junctions. A final network regularization step is based on perceptual grouping concepts.

1.2.7 Recursive Filtering

Filtering is a recursive procedure to estimate the parameters, *i.e.* the state, of a dynamic system. In the case of road tracking, although the state is time independent, the progressive growing of the estimated contour along the road is treated as if it were the time variable, and the state to be estimated is the parameters that describe the position and shape of roads.

Vosselman and de Knecht (1995) presented a road tracing algorithm by Kalman filtering. The first road segment given by an operator is used to initialize the parameters of the Kalman filter and to extract a template road intensity profile. Since roads are elongated regions, adjacent profiles taken perpendicularly to the road axis usually have a very similar shape. The position and the width of the next adjacent road profile can be predicted based on the current segment. The profile at the predicted position is matched with the model profile. The shift between the two profiles is used by the Kalman filter recursively to update the parameters. The prediction step of the Kalman filter enables continued following the road despite temporary failures of the profile matching due to road crossings, exits and cars. Baumgartner et al. (2002) developed a prototype system based on this method. An interaction interface is designed to enable human interaction during the tracking process.

In the ‘JetStream’ method proposed by Pérez et al. (2001), the contour of the road is

seen as the path of a stochastic process driven by both an inner stochastic dynamics, and a statistical data model. The prior dynamics retains expected properties of the contours to be extracted; the data model provides evidence about whether a measurement is, or is not, in the vicinity of the ‘true’ contour. A Monte Carlo technique, based on sequential importance sampling/resampling, is used to realize the recursion. Given the current position, particle filtering manipulates the filtering density to estimate the most probable next position according to different criteria. The filtering density is represented by a set of random samples with associated weights. If the state space is extended to include a width variable, the algorithm can deal with ribbon features. It is applied to road extraction from aerial images.

Zhou et al. (2005) employed a similar model as (Vosselman and de Knecht, 1995). The main difference is that a particle filter is used in place of the Kalman filter for prediction. Zhou et al. (2006) gave an overview of a road tracking system based on human-computer interaction and two kinds of Bayesian filters, *i.e.* extended Kalman filters and particle filters.

1.2.8 Support Vector Machines

Support Vector Machine (SVM) is a powerful classification technique based on the principles of statistical learning theory. SVM works by finding the hyperplane with the largest margin in the feature space that separates the positive and negative training samples.

Yager and Sowmya (2003) applied SVM to road extraction from rural areas using edge-based features. In the first level, the edges are found using a Canny edge detector. The edge length and the edge gradient are the features used to classify edges as road edges or non-road edges. After a learning stage from a training data set, an SVM is used to classify all the edges. In the second level, opposite road edges are paired to create road segments. They are classified as positive and negative edge pairs, using the pair width and the enclosed intensity. Another SVM is trained from the pair samples to classify all possible edge pairs. An extension of this work can be found in (Lai et al., 2005).

Song and Civco (2004) performed a two-step approach for road extraction from rural and urban areas. In the first step, SVM is employed to classify the input image into a road group and a non-road group. In the second step, the road group image is segmented into geometrically homogeneous objects using a region growing technique based on a similarity criterion, with higher weighting on shape factors over spectral criteria. Finally, a thresholding on the shape index and density features derived from these objects is applied to extract road features, which are further processed by thinning and vectorization to obtain road centerlines.

1.2.9 Utilization of New Sensor Data

The utilization of new sensor data is a new tendency for research in road extraction to overcome the difficulties from panchromatic satellite and aerial images.

Zhu et al. (2004) proposed an automatic road extraction technique that combines information from LIDAR data and aerial optical images. Firstly, the method obtains height and edges of high objects from LIDAR data. Secondly, digital images are analyzed at

these edges for road detection. Finally, shadowed parts are reconstructed by a spline-approximation algorithm. Hu et al. (2004) also combined information from LIDAR data and aerial optical images, taking advantage of deriving multiple clues and constraints to significantly minimize the uncertainty in the extraction process.

To extract roads from LIDAR, Rottensteiner et al. (2005) used a two step process. First, roads are detected by a hierarchical technique classifying the LIDAR points into road or non-road. Each step in the classification hierarchy addresses the appearance of roads in the sensor with respect to spectral, geometric, topological, and contextual characteristics. Secondly, road vectorization is performed. The road centerline, orientation and width are detected.

1.2.10 Other Methods

Many other techniques have also been employed for the application of road extraction.

Active testing. This approach was proposed by Geman and Jedynak (1996), originally for road tracking from SPOT satellite images with 10m resolution. It also offers a general computational strategy for tracking linear structures and other recognition tasks in computer vision. Given a start point and an orientation, the algorithm searches the road by constructing a decision tree. Candidate road pixels are selected in a limited number of directions using hypothesis tests. At each step, the best test is that which removes as much as possible the uncertainty in the road position, given the results of the previous tests. This choice is made online based on a statistical model for the joint distribution of tests and hypotheses. The uncertainty is measured by entropy. At each iteration, new image data are examined, and a new entropy minimization problem is solved, resulting in a new image location to inspect, and so forth. Dal Poz and Silva (2002) extended the work of Geman and Jedynak (1996) to medium- and high-resolution images. They defined road segments as rectangles, and performed the active testing strategy to track road segments. Finally, the road centerline is obtained from the extracted road segments. However, if the road borders are incomplete, or the road width changes, the centerline position might not be accurate.

Perceptual grouping. Mangin et al. (1992) proposed an iterative algorithm for edge intensity image enhancement. It uses local cooperation-inhibition processes to produce an edge image in which the most relevant contours have reached maximal activation, and small gaps and junctions have been filled in. The algorithm is robust in complex edge image context, and is stable under any number of iterations. Urago et al. (1994) proposed an algorithm to restore images of incomplete contours using a Markovian model. To complete the boundaries, a criterion is defined and introduced in an energy function. The minimization of this energy generates a configuration, in which the contours are reconstructed. This algorithm is tested on real SPOT images to fill up large gaps and to get a better contour grouping. Gamba et al. (2006) detected urban road networks from high resolution optical and SAR images. An adaptive filtering procedure is performed to capture the predominant directions of roads, and enhance the extraction results. Then, to both discard redundant segments and avoid gaps, a perceptual grouping algorithm is devised, exploiting colinearity and proximity concepts. Finally, the road network topology is considered, checking for road intersections, and regularizing the overall pattern.

Region competition. Amo et al. (2006) proposed a combined approach consisting of region growing and region competition to extract road centerlines and sides. The initial seeds are given manually. Then, a first simple region growing based model is applied to obtain a rough road approximation. This model is refined by the region competition algorithm.

Shape classification. Hu et al. (2007) presented a two-step approach, *i.e.* detecting and pruning, for the automatic extraction of road networks from aerial images. The road detection step is based on the shape classification of a local homogeneous region around a pixel. This step involves detecting road footprints, tracking roads, and growing a road tree. The road tree pruning step makes use of a Bayesian decision model based on the area-to-perimeter ratio of the footprint to prune the paths that leak into the surroundings.

Neural networks. Mokhtarzade and Zoej (2007) treated the possibility of using artificial neural networks for road detection from high-resolution satellite images. Attempts are also made to verify the impacts of different input parameters on the network's ability to find the optimal input vector for the problem. A variety of network structures are used to determine the best network structure and termination condition in the training stage.

Normalized cut. Grote et al. (2007) dealt with the segmentation of images of sub-urban scenes with the normalized cut algorithm. The similarity matrix necessary for normalized cuts is built up using similarity criteria, such as edges, color, hue and road surface color, which are suitable for the separation of road and non-road segments. The results are used for the verification of an existing road database, which is also introduced into the model as prior information for the segmentation.

Differential geometry. Jin and Davis (2003) presented an automatic road extraction strategy from pan-sharpened multispectral Ikonos images. First, initial road seeds are generated based on shape and structural analysis through segmentation and grouping. Based on the fact that roads are 1D bar-shaped or parabola-shaped intensity profiles perpendicular to road directions, a line detection method is used to extract additional road seeds, which are integrated with the initial seeds using graph theory. Missing pieces of road networks are added using a road tracker based on profile matching.

1.2.11 Summary

In this section, we have presented a wide range of techniques applied to road extraction. The early work mainly focused on extracting thin linear features from low resolution data, where the data input is some relative simple scene, *i.e.* rural or semi-urban areas. With the commercial availability of very high resolution (VHR) satellite images, recent research tends to extract belt or ribbon structures from VHR data. Furthermore, more and more attention has been directed to urban areas due to a variety of urban applications. The existing algorithms are very varied due to different goals, available data, scene complexity, etc. Comparatively, bottom-up methods, such as mathematical morphology and dynamic programming, are less computationally expensive, but due to their high sensitivity to nuisance factors, they show strong limitations, and in general, low robustness. Top-down methods, such as active contours, MRFs and marked point processes, are relatively less sensitive to incomplete and ambiguous information, but the computations needed are usually expensive, and the generic definition of a specific object is usually difficult.

1.3 Conclusion

In this chapter, we have briefly reviewed the state-of-the-art in the areas of road extraction and active contours. In the former, we classified road extraction algorithms according to the different techniques applied. Although a great number of approaches exist, few algorithms can give satisfactory results for VHR data; most of them utilize a series of processing steps to extract roads progressively. We also classified the various variational models into edge-based and region-based approaches. We focused on active contours with shape constraints, because having prior shapes is especially useful in segmenting high noise remote sensing images. Almost all of this work includes shape knowledge by using a reference shape or shapes. These methods are powerful when a topologically correct reference shape is given, but unfortunately fail when the topology is unknown *a priori*, such as in the problem of road extraction. In this thesis, we choose to make use of active contours because they provide a unified, analyzable and solid mathematical framework. Different types of prior knowledge can be easily integrated into such a framework. The introduction of geometric prior constraints in the absence of a particular reference shape, *i.e.* generic shape modeling problem, has been first tackled by Rochery (2005). For the flexibility and robustness that it provides, we decide to adopt this particular family of models in our work. In the next chapter, we will introduce the fundamentals of our used framework.

Chapter 2

Higher-Order Active Contours and Phase Fields

This chapter presents the phase field higher-order active contour (HOAC) framework for image segmentation. HOACs are a new generation of active contours, recently developed by Rochery et al. (2003, 2005b, 2006). They can encode complex prior knowledge about geometry. Thus, they are more robust and can be initialized generically, hence automatically, outperforming conventional active contours. Later, ‘phase fields’ were introduced for region modeling in (Rochery et al., 2005a), where HOACs are reformulated as (non-local) phase field models. Phase field modeling is the approach favored in this thesis, because it has several advantages over other traditional methods: a linear representation space; neutral initialization; ease of implementation; and greater topological freedom. On the other hand, parameter settings play an important role in the phase field HOAC model, and greatly affect the model performance, so we need to choose the parameters properly. In section 2.1, we first present a general probabilistic framework for image segmentation. In section 2.2, we briefly review the HOAC model, and its application to road network extraction from low to medium resolution ($\sim 10\text{m/pixel}$) remote sensing images. We recall the essentials of phase field models in the same section. In section 2.3, a stability analysis of a long straight bar of a fixed width is performed as an effective way to understand model behavior and the internal dependence among these parameters. In section 2.4, after adding a region based likelihood term, we define our overall primary model and specify the implementation details.

2.1 A General Framework for Image Segmentation

In this section, we introduce the Bayesian formulation of image segmentation, and show its connection with energy minimization methods.

Image segmentation refers to the process of partitioning on the image domain $\Omega \subset \mathbb{R}^2$ into multiple meaningful regions, given image data $I : \Omega \rightarrow \mathbb{R}$. In particular, the aim of image binary segmentation is to find the region R that corresponds to objects of interest,

from the image to be segmented. Thanks to Bayes' theorem, different knowledge, such as intensity, variance, color, texture, and shape information, can be integrated easily into this framework. Knowledge of R can be taken into account to define the *a posteriori* probability distribution $P(R|I, K)$, given the image data I , and given all the prior knowledge K we may have. From this formulation, the optimal partition of the image can be computed by finding the region R with maximum posterior probability. Using Bayes' rule, this probability can be expressed as

$$P(R|I, K) = \frac{P(I|R, K)P(R|K)}{P(I|K)}, \quad (2.1)$$

where $P(R|K)$ and $P(I|K)$ are respectively the probability of a segmented region R and the probability of an image I , given the knowledge K . The third term, $P(I|R, K)$ represents the likelihood of the image I , given the region R and the knowledge K . When segmenting a given image using *maximum a posteriori* (MAP) estimation, the denominator $P(I|K)$ does not depend on the estimated quantity R , and can thus be neglected in the maximization. Equation (2.1) simplifies to

$$P(R|I, K) \propto P(I|R, K)P(R|K). \quad (2.2)$$

Equivalently, the MAP estimate can be found by minimizing the negative logarithm of the probability, which leads to the 'energy'. The total energy to be minimized, up to an additive constant, can be defined by

$$E(R; I) = \theta E_P(R) + E_D(I, R), \quad (2.3)$$

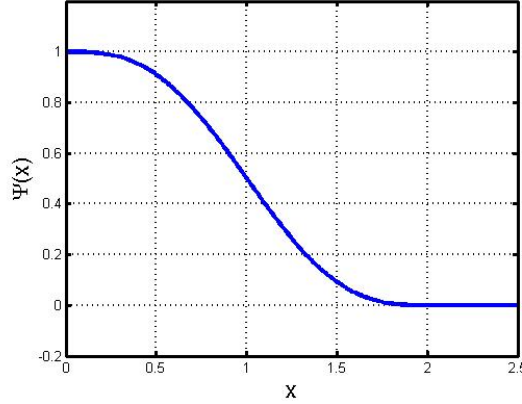
where E_P is the prior energy, and E_D is the data energy. θ is a constant that balances the contribution of the two energies.

2.2 Summary of Higher-Order Active Contours and Phase Fields

In this section, we give a brief introduction to the work on HOACs in (Rochery et al., 2003, 2005b, 2006), and then to the work on phase field models in (Rochery et al., 2005a).

2.2.1 Higher-Order Active Contours

Conventional active contours (Chan and Vese, 2001a; Chen et al., 2002; Cremers et al., 2002; Kass et al., 1988; Mayer et al., 1998) are defined by linear functionals on a certain space containing the space of curves. They are expressed in terms of single integrals over the contour, and can incorporate only local interactions between contour points, and hence only very weak prior information about the region geometry. Examples of linear functionals include boundary length and region area. In contrast, HOACs (Rochery et al., 2003, 2005b, 2006) are defined by arbitrary polynomial functionals on this space. Expressed as multiple integrals over the contour, they are associated to nonlocal interactions. Thus, via these arbitrarily long-range interactions between subsets of points along the region boundary, HOACs allow the inclusion of complex prior geometrical constraints.

Figure 2.1. The interaction function Ψ .

Rochery et al. (2003) proposed a Euclidean invariant HOAC prior energy for modeling network regions. It is the sum of two linear terms and a quadratic HOAC term:

$$E_{C,S}(R) = \lambda_C L(\partial R) + \alpha_C A(R) - \frac{\beta_C}{2} \iint_{(\partial R)^2} \dot{\gamma}(t) \cdot \dot{\gamma}(t') \Psi\left(\frac{|\Delta\gamma(t,t')|}{d}\right) dt dt' , \quad (2.4)$$

where ∂R is the boundary of region R ; $\gamma : S^1 \rightarrow \Omega$ is a map representing ∂R , parameterized by t ; dots represent differentiation with respect to t ; L is the boundary length; A is the region area; $\Delta\gamma(t,t') = \gamma(t) - \gamma(t')$; and d is a constant that controls the range of the interaction. The long range interaction between t and t' is modulated by Ψ , the interaction function (Figure 2.1):

$$\Psi(x) = \begin{cases} \frac{1}{2} \left(2 - |x| + \frac{1}{\pi} \sin(\pi|x|) \right) & \text{if } |x| < 2 , \\ 0 & \text{else .} \end{cases} \quad (2.5)$$

It is a smoothly decreasing function from 1 at $x = 0$ to 0 for $x \geq 2$ ¹. In Appendix D, another type of interaction function will be discussed, for a new model proposed in chapter 5.

In equation (2.4), $L(\partial R)$ acts as a regularizer and encourages the smoothness of the boundary, while $A(R)$ controls the expansion of the region. The quadratic HOAC term has two effects, as illustrated in Figure 2.2. It prevents pairs of points with antiparallel tangent vectors from coming too close; they keep a certain distance related to the value of d . On the other hand, it encourages pairs of points with parallel tangent vectors to attract each other, and thus the growth of arm-like structures. Consequently, the effect is to assign low energy to (and hence favor) regions composed of long arms of a certain width and with roughly parallel sides that join together at junctions. In other words, it models network structures. Note that this prior knowledge is more *generic* compared to those methods reviewed in subsection 1.1.3, because it permits modeling a family of shapes, rather than a given *specific* reference shape.

¹In this thesis, if the interaction function is not specially indicated, equation (2.5) is the default definition of Ψ .

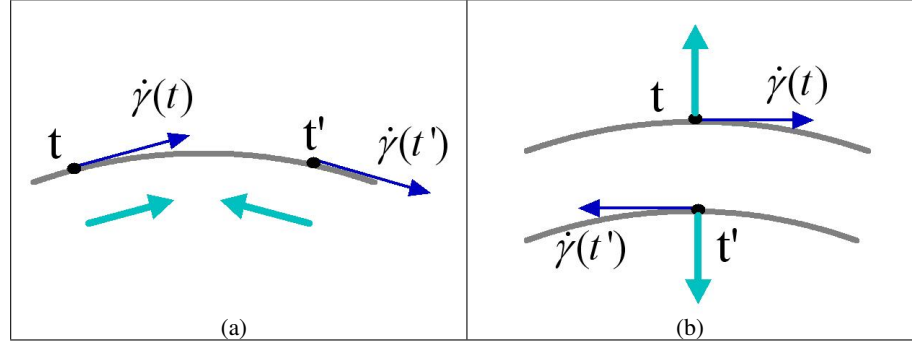


Figure 2.2. The effects of the quadratic HOAC term in equation (2.4) (thin blue arrow: tangent vector; thick cyan arrow: interaction force; black dot: interacting point). (a): when the tangent vectors of two interacting points are nearly aligned, it encourages them to attract each other; (b): when the tangent vectors of two interacting points are nearly anti-aligned, it prevents them from coming too close.



Figure 2.3. Evolutions for road detection on a SPOT satellite image (Rochery et al., 2005b).

Adding an edge-based likelihood term, Rochery et al. (2003) used the entire model to detect road structure from low to medium resolution ($\sim 10\text{m/pixel}$) remote sensing images. An example is shown in Figure 2.3. The road network has been detected in the SPOT image. The prior term favors network shapes with respect to other shapes, by reducing the energy of network configurations. In addition, the data term defines privileged directions, and drives the production, growth, and branching of the network in such directions. Perturbations that do not fit the data will be rapidly damped, but this is dependent on the relative magnitudes of the likelihood and prior models. In this case, the presence of HOACs penalizes a large number of false contour configurations, and eliminates many local energy minima, thanks to the incorporation of more sophisticated prior knowledge. For this reason, HOACs are more robust to noise than conventional active contours, get rid of the pose estimation of the shape, and permit a generic initialization that renders automatic the network extraction.

2.2.2 Phase Field Models

Phase field models have been widely used for modeling many phenomena in physics. Reviews can be found for example in (Chen, 2002) and (González-Cinca et al., 2004). Suppose an interface moves with a velocity proportional to the gradient of some field, which could correspond to temperature or impurity concentration in solidification problems, pressure or another appropriate potential in viscous fingering, etc.; this is a class of non-equilibrium pattern formation problems. The motion of the interface is controlled by the associated mechanical principles, coupled with a boundary condition. The shape is evolving in time, so it has proved to be very difficult to find analytic solutions of practical use for these moving boundary problems. Alternatively, the phase field model provides a powerful computational way to study such problems. It enables the conversion of a moving boundary problem into a set of partial differential equations (PDEs), enabling easier numerical treatment. In phase field modeling, no distinction is made between the different phases or the interface. All regions are described in terms of an additionally introduced phase field function ϕ . The phase field ϕ is continuous in space, but takes distinct constant values in each phase. The physical interface is then located in the region where ϕ changes its value. The range over which the value changes is the width of the interface w . This representation permits the whole domain to be computed simultaneously. In particular, the need for an explicit boundary representation is eliminated, and the interface is then given implicitly by the chosen value of ϕ as a function of time and space. Not having to track the boundary greatly simplifies topological changes of the interface and extensions to higher dimensions.

Recently, phase field models have attracted increasing interest in the image processing community. For image inpainting, Grossauer and Scherzer (2003) solved a complex Ginzburg-Landau equation, with boundary conditions depending on image data. Beneš et al. (2004) presented an algorithm based on the modified Allen-Cahn equation for image segmentation. A special forcing term is imposed on mean curvature to drive the level set to the desired object boundary. Aubert et al. (2005) modified the complex Ginzburg-Landau equation by a diffusion coefficient depending on the image data and by the addition of a data term, and applied it to detect codimension two objects in the image domain. Samson et al. (2000) used a phase field model with Gamma convergence to construct piecewise constant approximations to an image.

In the rest of this subsection, we explain two phase field terms introduced in (Rochery et al., 2005a): the basic phase field term, and the standard phase field HOAC term, which is an equivalent reformulation of the quadratic term described in the previous subsection.

Region Representation

Phase fields represent a region R using a function $\phi : \Omega \rightarrow \mathbb{R}$ defined over the entire image, rather than using its bounding contour. By definition, ϕ defines a region R in the space of regions \mathcal{R} via the map $\zeta_z(\phi) \in \mathcal{R} = \{x \in \Omega : \phi(x) > z\}$, where $z \in \mathbb{R}$ is some threshold. R can be regarded as a level set of ϕ . In fact, when the interface width $\omega \rightarrow 0_+$, the phase field model with diffuse-interface becomes identical to a sharp-interface formulation. Compared to other region modeling methods, *e.g.* parametric active contours or standard level sets,

phase fields provide a more convenient framework, for the following reasons:

- The lack of any hard constraints on the phase field function, *e.g.* that it should be a distance function, means that the set of functions considered is a linear space. This is responsible for a great simplification in model description and analysis.
- A neutral initialization can be used for evolutions². The initial value of the phase field function is set equal to a suitable value z (a constant depending on the mathematical definition of the model) everywhere in the image domain. Hence, the bias which the initialization may create can be removed.
- They provide a unified, analyzable variational framework, rather than a series of processing steps. The evolutions are based solely on the PDE that results from an energy functional using gradient descent. The complicated procedure of boundary extraction, velocity extension, and *ad hoc* regularization or reinitialization in the level set framework is thereby avoided. Moreover, the functional derivative of HOAC terms takes the form of a convolution, and can be computed linearly in the Fourier domain. As a consequence, numerical implementation is simple.
- They allow greater topological freedom, *i.e.* components of a region can be created or destroyed everywhere in the image domain; holes can be created or destroyed inside existing regions. This characteristic is critical when the topology of the region is not known *a priori*. In the context of the present application, the objects to be segmented, *i.e.* road networks, may have several connected components and many loops. Dealing with this topological complexity is arguably one of the most difficult aspects of automatic road network extraction; phase fields handle it “naturally”.

Because of the above advantages, the phase field representation and phase field modeling is the favored methodology in this thesis. Later, we will use ϕ instead of R to denote the region in the energy functionals. For each term of the HOAC model in equation (2.4), to a very good approximation, a counterpart can be constructed using phase fields.

Basic Phase Field Term

To model the region of interest with the phase field function, the *basic* phase field term E_0 is given by the Ginzburg-Landau energy (Ginzburg and Landau, 1950) plus an odd parity term:

$$E_0(\phi) = \int_{\Omega} \left\{ \frac{1}{2} \nabla \phi(x) \cdot \nabla \phi(x) + U(\phi(x)) \right\} dx. \quad (2.6)$$

The ‘potential’ U is

$$U(y) = \lambda \left(\frac{1}{4} y^4 - \frac{1}{2} y^2 \right) + \alpha \left(y - \frac{1}{3} y^3 \right), \quad (2.7)$$

²Except in the case of purely geometric evolution, where ϕ is initialized to some particular shape, or to the neutral value plus random noise, like the experiments in subsection 2.3.4.

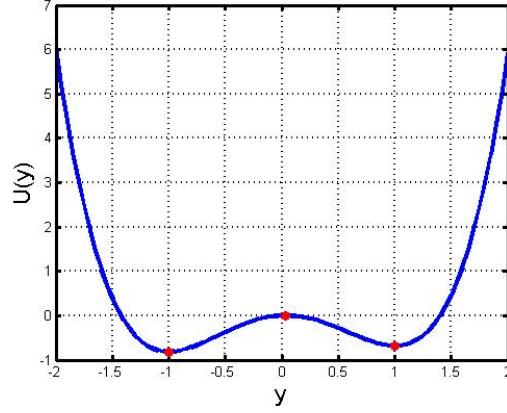


Figure 2.4. An example of the potential function U ($\lambda = 3, \alpha = 0.1$). Red dot: extremum.

where λ and α are constants. As shown in Figure 2.4, for $\lambda \geq \alpha > 0$, U has two minima, at $y = -1$ and $y = 1$, and a maximum at $y = \alpha/\lambda$. The threshold z we mentioned before for the neutral initialization is associated to this maximum.

Define $\phi_R = \arg \min_{\phi: \zeta_z(\phi)=R} E_0(\phi)$. If we ignore the gradient term in equation (2.6), and set $z = \alpha/\lambda$, we clearly find that $\phi_R(x) = 1$ for $x \in R$ and $\phi_R(x) = -1$ for $x \in \bar{R} = \Omega \setminus R$. As a result, the quantities $\phi_{\pm} = (1 \pm \phi)/2$ are equal to the characteristic functions of R and \bar{R} . Adding the gradient term results in a smooth transition from 1 to -1 over an interface region R_C around the boundary ∂R . It has been proved that $E_0(\phi_R) \simeq \lambda_C L(\partial R) + \alpha_C A(R)$, *i.e.* E_0 corresponds to the two linear terms in $E_{C,S}$ (equation (2.4)). The detailed proof can be found in (Rochery et al., 2005a). Therefore, E_0 ensures region stability, boundary smoothness, and the properties of the functions ϕ_{\pm} .

Standard Phase Field HOAC Term

To introduce sophisticated prior knowledge of geometric information into the phase field framework, the quadratic term in $E_{C,S}$ (equation (2.4)) should also be reformulated in terms of an equivalent phase field energy (Rochery et al., 2005a). The tangent vector of the contour is then replaced by the normal vector of the phase field surface at the threshold level. The *standard* phase field HOAC term E_S ³ becomes

$$E_S(\phi) = -\frac{\beta}{2} \iint_{\Omega^2} \nabla \phi(x) \cdot \nabla \phi(x') \Psi\left(\frac{|x-x'|}{d}\right) dx dx', \quad (2.8)$$

where β is constant, and Ψ denotes the interaction function as defined in equation (2.5). Note that $\nabla \phi$ is very close to zero except in the interface region, so the integrand in E_S is non-zero solely inside the interface R_C . In other words, E_S describes long-range interactions between pairs of points on R_C , and it also favors the configurations shown in

³In order to distinguish this term from the other phase field HOAC terms we will propose in chapter 5, we name the one in (Rochery et al., 2005a) the standard HOAC prior term.

Figure 2.2. Consequently, the sum $E_0 + E_S$ is equivalent to the HOAC prior model $E_{C,S}$ in equation (2.4) (Rochery et al., 2005a).

2.3 Stability Analysis of the Standard HOAC Total Prior Model

So far, the *standard* HOAC total prior model takes the form $E_{P,S} = E_0 + E_S$. There are four parameters $(\alpha, \lambda, \beta, d)$, and unfortunately, not all the parameter values can create the desired stable network structures. The behavior of the prior energy indeed depends on the parameter settings, and can vary significantly. If we wish to model networks with this energy, it is therefore very important to analyze the stability conditions, and to deduce the resulting parameter dependence. In this section, a long straight bar of a given width is first defined. We calculate the terms E_0 and E_S for this bar. By considering the total prior energy $E_{P,S}$, we establish the constraints on the parameters $(\alpha, \lambda, \beta, d)$ that ensures that a long network branch of the desired width is a stable configuration of the energy functional. An important side-effect is that some of the (rather abstract) model parameters are effectively replaced by ‘physical’ quantities, such as bar and interface width, which we can reasonably fix from numerical or application considerations. Appendix A furnishes details of intermediate steps in the derivation of certain results used in this section. Note that, in addition, to guarantee the Turing stability of the model, Rochery et al. (2005a) introduced another constraint, which we will not detail here.

2.3.1 Definition of a Bar

Since network branches are locally like straight bars, we can, to a good approximation, analyze the stability of a long straight bar, of length L and width $W \ll L \rightarrow \infty$. This allows us to ignore boundary effects. Such a bar, oriented arbitrarily, is shown in Figure 2.5.

Ideally, we should minimize the prior energy under the constraint that $\zeta_z(\phi) = R_{\text{bar}}$, and then expand around that point to test stability, but this is very difficult. Instead, we take a simple *ansatz* for $\phi_{R_{\text{bar}}}$, and study its stability in a low-dimensional subspace of function space; the results may be justified *a posteriori* by numerical experiments. In (Rochery et al., 2005a), a similar procedure was followed, the results comparing favorably to those obtained by more sophisticated ‘matched asymptotics’.

The *ansatz* is defined as follows. The phase field is given by $\phi(x) = 1$ for $x \in R \setminus R_C$; $\phi(x) = -1$ for $x \in \bar{R} \setminus R_C$, while in R_C , ϕ changes linearly from 1 to -1 . To denote more clearly the “bottom” interface and the “top” interface during the calculation, they are denoted respectively by R_{CB} and R_{CT} . Using Cartesian coordinates and rotation invariance, ϕ at each pixel $x = (u, v) \in \Omega$ takes the values:

$$\phi(x) = \begin{cases} \frac{2}{w}v - 1 & \text{if } -\infty < u < \infty \text{ and } 0 < v < w, \\ 1 & \text{if } -\infty < u < \infty \text{ and } w < v < W, \\ \frac{2}{w}(W - v) + 1 & \text{if } -\infty < u < \infty \text{ and } W < v < W + w, \\ -1 & \text{else.} \end{cases} \quad (2.9)$$

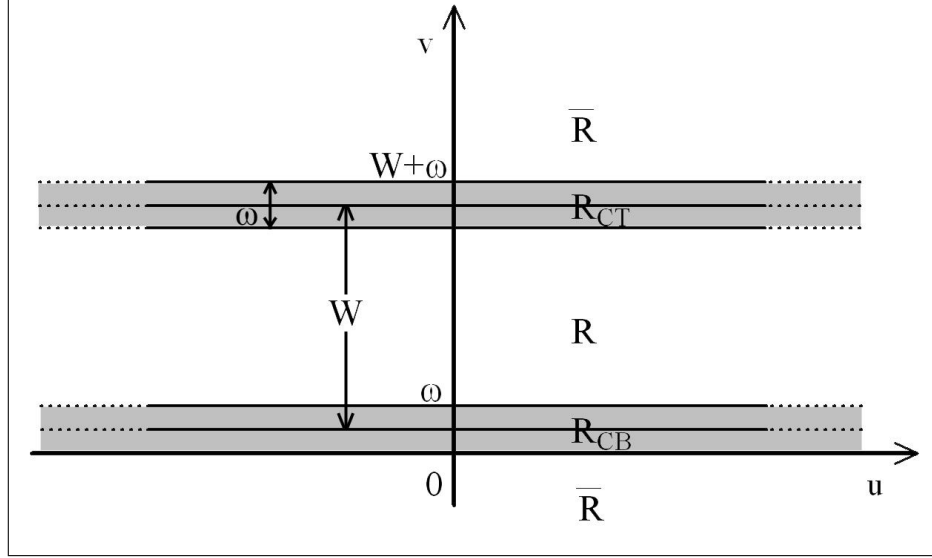


Figure 2.5. A bar of length $L \rightarrow \infty$ and width W in the $u - v$ plane.

Now we evaluate the prior energy on this *ansatz*, per unit length of bar. The constraint that the network branch is a local energy extremum requires the first derivatives with respect to w and W to be zero, while the constraint that it is a local minimum requires that the second derivatives must be positive semi-definite (the semi-positive condition comes from the fact that in some applications, the inflection point can achieve a particular effect). These requirements provide the stability conditions of the model.

2.3.2 Basic Phase Field Term of a Bar

For the basic phase field term E_0 (equation (2.6)), there are three contributions from the region R , the region \bar{R} and the interface R_C .

First, the contributions from the two regions are:

$$E_{0,R} = \left(-\frac{1}{4}\lambda + \frac{2}{3}\alpha\right)(WL - wL), \quad (2.10)$$

$$E_{0,\bar{R}} = \left(-\frac{1}{4}\lambda - \frac{2}{3}\alpha\right)(A_0 - WL - wL), \quad (2.11)$$

where A_0 is the area of Ω , and W is the distance between the threshold points.

The “bottom” and “top” interfaces are symmetric, so we only need to calculate the contribution of one of them, for example, R_{CB} . The gradient of ϕ with respect to x in R_C is given by $\nabla\phi = -(2/w)\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the outward normal vector to the boundary, extended to R_C . The third contribution, from the interface, is thus (see details in Appendix A.1.1)

$$E_{0,R_C} = \frac{4L}{w} - \frac{7}{30}\lambda wL. \quad (2.12)$$

Therefore, summing equations (2.10) to (2.12), the basic phase field energy, per unit length of bar, e_0 is

$$e_0(w, \hat{W}) = \frac{4}{3}\alpha\hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + C_1, \quad (2.13)$$

where the scaled width $\hat{W} = W/d$, and $C_1 = -(\lambda/4 + 2\alpha/3)A_0/L$ is independent of \hat{W} or w . We can therefore neglect it in the following discussion.

2.3.3 Standard Phase Field HOAC Term of a Bar

Since $\nabla\phi$ is zero in the regions R and \bar{R} , the double integral of the standard phase field HOAC term E_S in equation (2.8) needs to be calculated only across the interface R_C . After a lengthy calculation (see Appendix A.1.2), we have

$$E_S = -8\beta L \left\{ \int_0^\infty \Psi\left(\frac{z}{d}\right) dz - d \int_{\hat{W}}^\infty \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \Psi(\eta) d\eta \right\}. \quad (2.14)$$

According to the definition of the interaction function Ψ (equation (2.5)), the first integral is equal to d . For the second integral, there are two cases: $\hat{W} > 2$, and $\hat{W} < 2$. The first case just gives zero. Since the stable width of a network branch is expected to lie in the interval $\hat{W} < 2$, we obtain the standard phase field HOAC energy, per unit length of bar, e_S :

$$\begin{aligned} e_S(w, \hat{W}) &= -8\beta d + 4\beta d \int_{\hat{W}}^2 \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \left(2 - \eta + \frac{1}{\pi} \sin(\pi\eta)\right) d\eta \\ &= 4\beta d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta + C_2, \end{aligned} \quad (2.15)$$

where $C_2 = -8\beta d$ is also independent of \hat{W} or w .

2.3.4 Standard HOAC Total Prior Model of a Bar

Combining equations (2.13) and (2.15), and omitting the terms independent of \hat{W} and w , we have the standard HOAC total prior energy, per unit length of bar, $e_{P,S}$:

$$\begin{aligned} e_{P,S}(w, \hat{W}) &= e_0 + e_S \\ &= \frac{4}{3}\alpha\hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + 4\beta d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta. \end{aligned} \quad (2.16)$$

There are two unknown variables w and \hat{W} . The energy $e_{P,S}$ is now minimized with respect to w and \hat{W} by setting its first derivatives to zero, while ensuring that the second derivatives are positive semi-definite. The first derivatives are

$$\frac{\partial e_{P,S}}{\partial \hat{W}} = \frac{4}{3}\alpha d - 4\beta \hat{W} d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta, \quad (2.17a)$$

$$\frac{\partial e_{P,S}}{\partial w} = \frac{4}{15}\lambda - \frac{4}{w^2}. \quad (2.17b)$$

Since the interface width $w > 0$, the derivative of order two with respect to w is always positive. On the other hand, we need to check that the second derivative with respect to \hat{W} is non-negative. It is given by (see details in Appendix A.2.1)

$$\frac{\partial^2 e_{P,S}}{\partial \hat{W}^2} = -4\beta d \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right) + 4\beta d \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta. \quad (2.18)$$

In fact, stability depends only on the scaled parameter $\hat{\beta} = \beta/\alpha$, and the scaled width $\hat{W} = W/d$. By tidying up the expressions derived in equations (2.17) to (2.18), we find constraints on the parameters that ensure that a network branch with the given width is stable:

$$\hat{\beta} = \frac{1}{3\hat{W}I_1(\hat{W})}, \quad (2.19a)$$

$$I_2(\hat{W}) \geq \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right), \quad (2.19b)$$

$$\lambda = \frac{15}{w^2}, \quad (2.19c)$$

where

$$I_1(\hat{W}) = \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta, \quad (2.20a)$$

$$I_2(\hat{W}) = \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta. \quad (2.20b)$$

Based on the constraint in equation (2.19a), if we plot $\hat{\beta}$ against the possible position of the energy local extrema of a straight bar, we obtain the curve in Figure 2.6. This is an example of a *fold catastrophe* (Zeeman, 1977). The solid curve represents the extremum as a local minimum, *i.e.* the second constraint in equation (2.19b) is satisfied; while the dashed curve represents the extremum as a local maximum, *i.e.* the second constraint is not satisfied. Indeed, for any given choice of \hat{W} along this curve, there is a unique $\hat{\beta}$. This means that for α and β , one of them can be eliminated thanks to their dependence. Fixing $\hat{\beta}$ by a value on the curve, for example, suppose that $\hat{\beta} = 0.2125$ (the two associated local extrema are at $\hat{W}_{\text{MAX}} = 0.584$ and $\hat{W}_{\text{MIN}} = 1.173$), and also suppose that units have been chosen such that $d = 1$, we can plot the energy $e_{P,S}$ of a straight bar against its width W , as shown in Figure 2.7(a). As expected, the energy has a local minimum at $W = \hat{W}_{\text{MIN}}$, and a local maximum at $W = \hat{W}_{\text{MAX}}$. If $\hat{\beta}$ is decreased, the local maximum and the local minimum will move closer and closer together. At the critical point ($\hat{W}_0 = 0.8798, \hat{\beta}_0 = 0.1732$, the red dot in Figure 2.6), the two local extrema completely merge, and become a single inflection point at \hat{W}_0 . For $\hat{\beta} < \hat{\beta}_0$, no local extrema exist (see Figure 2.7(b)): the energy function

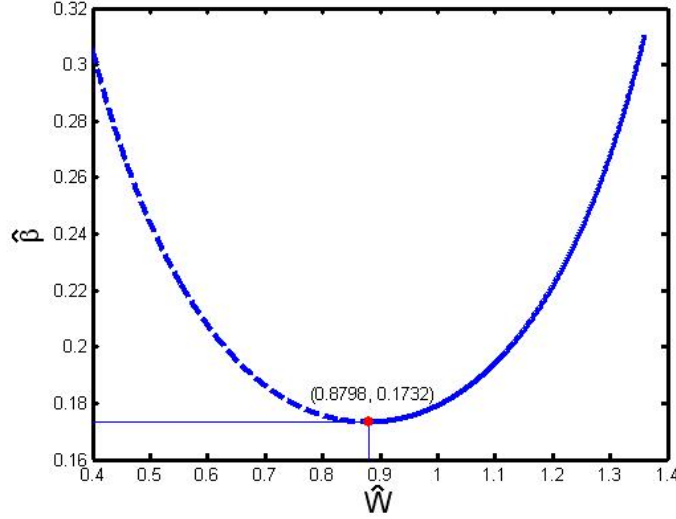


Figure 2.6. The relationship between the scaled parameters $\hat{\beta}$ and \hat{W} . The solid curve represents a local minimum; the dash curve represents a local maximum; and the red dot ($\hat{W}_0 = 0.8798, \hat{\beta}_0 = 0.1732$) represents the inflection point.

is monotonically increasing because the quadratic term is not strong enough to compensate for the shrinking effect of the basic phase field term. In this case, the region will entirely vanish under gradient descent. So when setting the parameters, we have to ensure that the chosen $\hat{\beta}$ is on the right branch of Figure 2.6.

Note that when $d = 1$, so that $W = \hat{W}$, the minimum width that can be achieved is $\hat{W}_0 = 0.8798$. By choosing the pixel units d , we can model a network branch with any given width $\hat{W}d$. More importantly, the $\hat{\beta}$ value remains the same as in the case $d = 1$.

To determine the parameter values of the standard phase field HOAC model for a prescribed network branch width W , we take the following steps:

- Choose w . If w is too small, very fine and/or adaptive discretization is required, which is computationally expensive; if w is too big, a reasonable approximation by the phase field model of the HOAC model cannot be achieved (Rochery et al., 2005a). Normally, the model works well with $w = 1 \sim 4$. Then λ in equation (2.7) is fixed using equation (2.19c).
- The critical point \hat{W}_0 determines an upper bound on d . Choose d in equation (2.8) less than W/\hat{W}_0 .
- Find the $\hat{\beta}$ value corresponding to the current W/d on the curve in Figure 2.6. Choose α or β properly, so the other is then fixed.
- Make sure the parameter setting satisfies the more general Turing stability condi-

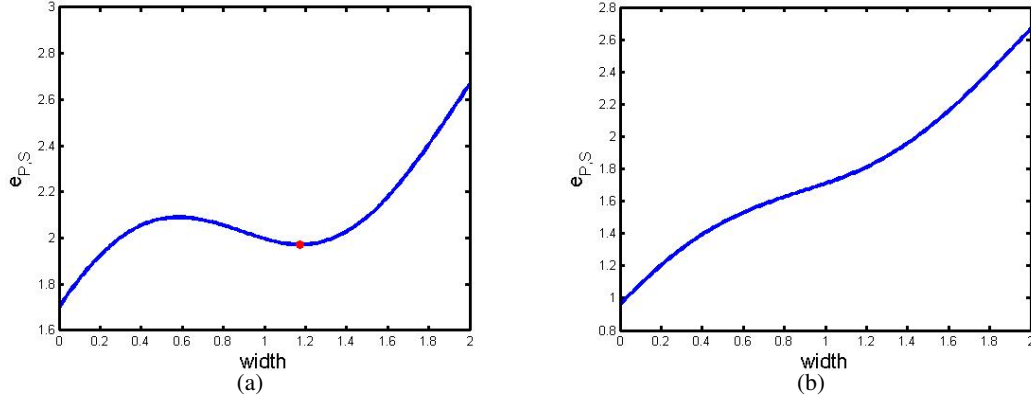


Figure 2.7. The prior energy $e_{P,S}$ for different $\hat{\beta}$. (a): $e_{P,S}$ has one local minimum at $W = 1.173$ ($\hat{\beta} = 0.2125 > \hat{\beta}_0$); (b): $e_{P,S}$ has no local minimum ($\hat{\beta} = 0.12 < \hat{\beta}_0$).

tion (Rochery et al., 2005a). Moreover, since we haven't expanded the energy around the stable point to test stability to perturbations with non-zero frequency, *a posteriori* numerical experiments are required to verify the results.

Figure 2.8 shows geometric evolutions using different $\hat{\beta}$. The initial shape is composed of four straight bars of various widths: 3, 9, 15, and 29 pixels. The other parameters are $(\lambda, d) = (2, 6)$. In the top row, we fix $\hat{\beta} = 0.14$, which is less than the critical value $\hat{\beta}_0$. The straight bars disappear one after another due to the weak force of the quadratic energy. In the bottom row, we set $\hat{\beta} = 1.5 > \hat{\beta}_0$, a relatively strong value to exaggerate the effect of the quadratic term in order to make clear the information contained in it. Note that since these initial widths are between the local maximum and the local minimum of the prior energy, the situation that a straight bar slides down to the zero width, will not arise. By either expanding or shrinking, all the straight bars eventually stop at a common stable width as desired.

Figure 2.9 represents the phase field function for geometric evolutions using the stable parameter setting ($\hat{\beta} > \hat{\beta}_0$), but with other initial shapes. The parameters used in this group of experiments are $(\lambda, \alpha, \beta, d) = (2, 0.2, 0.3, 6)$. In the top row, the initial shape is a circle of radius 50 pixels. Arms with a certain width are formed from the circle, and continue growing in the whole domain. In the bottom row, the function ϕ is initialized at the local maximum of the potential function α/λ , plus Gaussian random noise of very small variance. The evolution shows that different pixels agglomerate into regions, whose borders become smoother and smoother, and a line network is finally generated. Although different initial conditions result in convergence to possibly very different shapes, these shapes have many qualitative and quantitative properties in common. Therefore, the designed prior energy incorporates geometric knowledge suitable for network structure modeling.

Note that this value $\hat{\beta} = 1.5$ is not the same as the one used in the presence of a data term. In the absence of data, $\hat{\beta}$ is adjusted to a somewhat large value so that each unit length

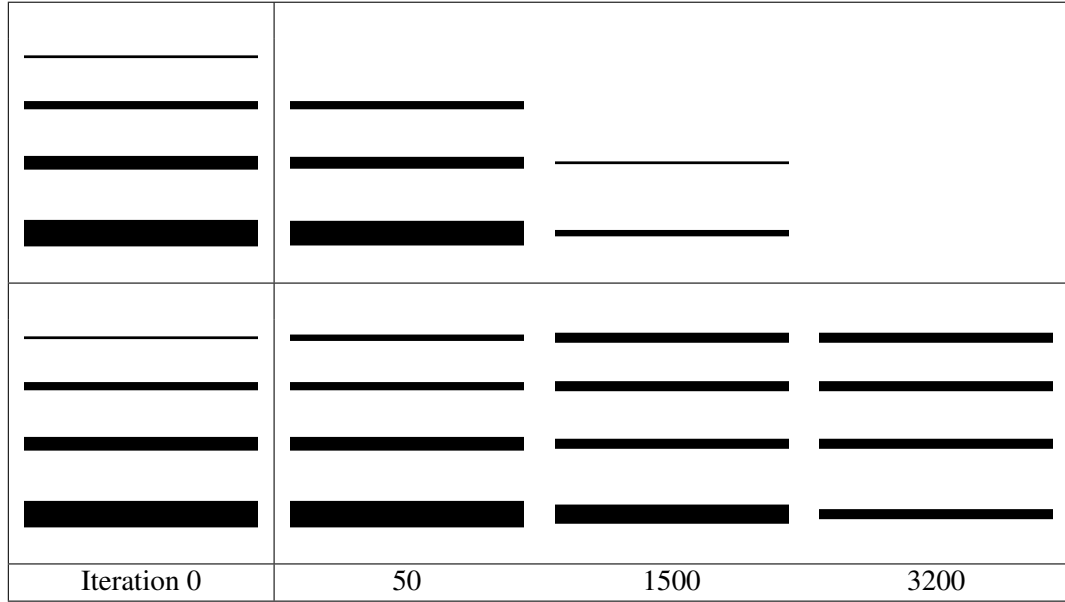


Figure 2.8. Geometric evolutions using different $\hat{\beta}$ ($\lambda = 2, d = 6$). Left column: the initial shape composed of four straight bars of various widths (3, 9, 15, and 29 pixels); three rightmost columns: the thresholded phase field function at different iteration steps (time runs from left to right). The precise behavior of the prior energy depends on the size of $\hat{\beta}$. Top row: the straight bars disappear one after another ($\hat{\beta} = 0.14 < \hat{\beta}_0$); bottom row: by either expanding or shrinking, all the straight bars eventually stop at a common stable width as desired ($\hat{\beta} = 1.5 > \hat{\beta}_0$).

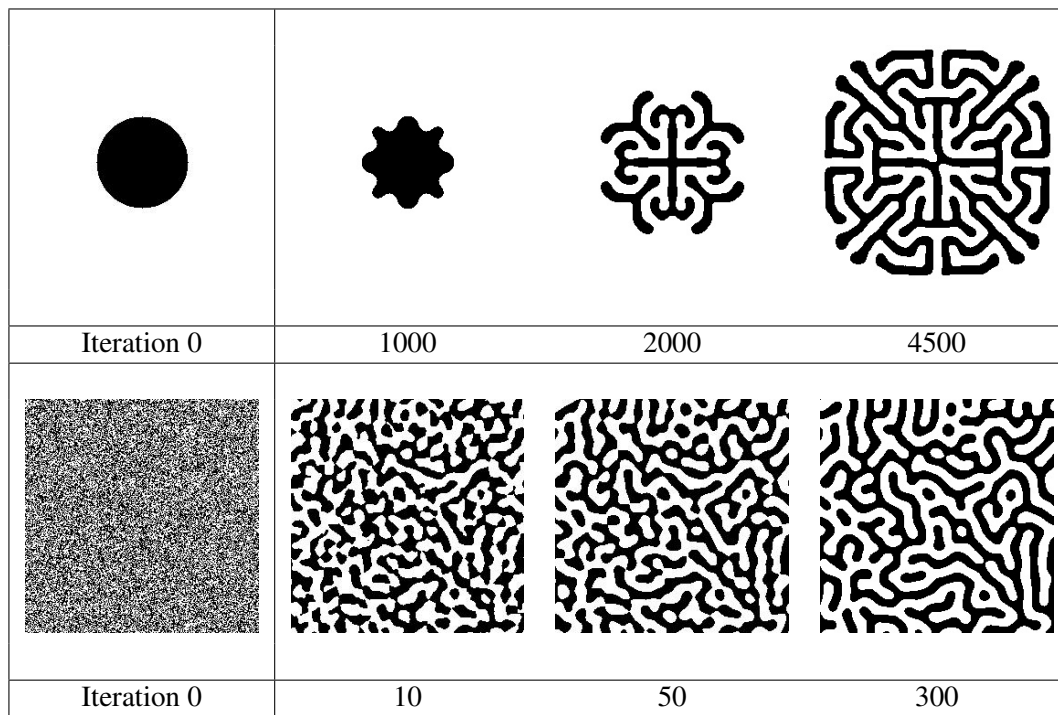


Figure 2.9. Geometric evolutions using the stable parameter setting ($\hat{\beta} > \hat{\beta}_0$), with other initial shapes. The parameters used are $(\lambda, \alpha, \beta, d) = (2, 0.2, 0.3, 6)$. Left column: the initial shape; three rightmost columns: the thresholded phase field function at different iteration steps (time runs from left to right). The initial ϕ is a circle of radius 50 pixels (top row); or the local maximum of the potential function α/λ , plus Gaussian random noise of very small variance (bottom row).

of a straight bar adds a negative amount of energy to the total. Once created, the arms are maintained, and furthermore, elongate. In the experimental results on the real images that we will show later, $\hat{\beta}$ is tuned so that each unit length of a straight bar adds a small but positive amount to the total energy, because the data can indicate the preferred regions.

2.4 Overall Model for Linear Network Extraction

Building upon the previous work on HOACs and phase field models, in this section, we set up our *primary* model. This model integrates the geometric constraints as already seen in section 2.2, and the data-based constraint, which we introduce hereafter. We also specify some of the implementation details.

2.4.1 Data Energy

In addition to the prior energy, we need a likelihood energy E_D linking the region R , which in our case corresponds to the road network, to the data, *i.e.* a very high resolution (VHR) optical satellite image. The proposed data energy E_D takes into account the following radiometric properties of dense urban areas, which discriminate roads from the background:

- Roads are mainly built from the same materials (concrete, asphalt) and thus tend to have somewhat homogeneous spectral properties. In contrast, the background (*i.e.* the non-road region) has no single photometric characteristic.
- The surfaces of main roads are not entirely uniform due to the presence of deterministic noise, such as zebra crossings, over-bridges, vehicles, shadows, road signs, etc. Nevertheless, they still show much less variability than the background.

E_D is the negative logarithm of $P(I|R, K)$ in equation (2.2). After making the reasonable assumption of independence between the image data in R and \bar{R} given R and K , $P(I|R, K)$ can be factorized as $P(I_R|R, K)P(I_{\bar{R}}|\bar{R}, K)$, where subscripts indicate ‘restricted to’. We use the same parameterized model for I_R and $I_{\bar{R}}$, the choice of model being based on a study of the image statistics. We model both the one point statistics of the image intensity, *i.e.* the distribution of $I(x)$, and the two-point statistics, which we characterize by the variance $V(x)$ of the image in a small window around each pixel. Because of the factorization, the data energy becomes the sum of two pieces (R is replaced by its representation ϕ):

$$E_D(I, \phi) = - \int_{\Omega} \left\{ [\ln P_+(I(x)) + \theta_v \ln Q_+(V(x))] \phi_+(x) + [\ln P_-(I(x)) + \theta_v \ln Q_-(V(x))] \phi_-(x) \right\} dx, \quad (2.21)$$

where θ_v is the weight of the two-point statistics. $\phi_{\pm} = (1 \pm \phi)/2$, as we discussed before, are approximately equal to the characteristic functions of R and \bar{R} . P_+ and P_- are the density functions of the grey level I (one-point statistics model), and Q_+ and Q_- are the density functions of the variance V (two-point statistics model) respectively. The parameters for this model are learned *a priori*, in a supervised way, from image samples of road and non-road.

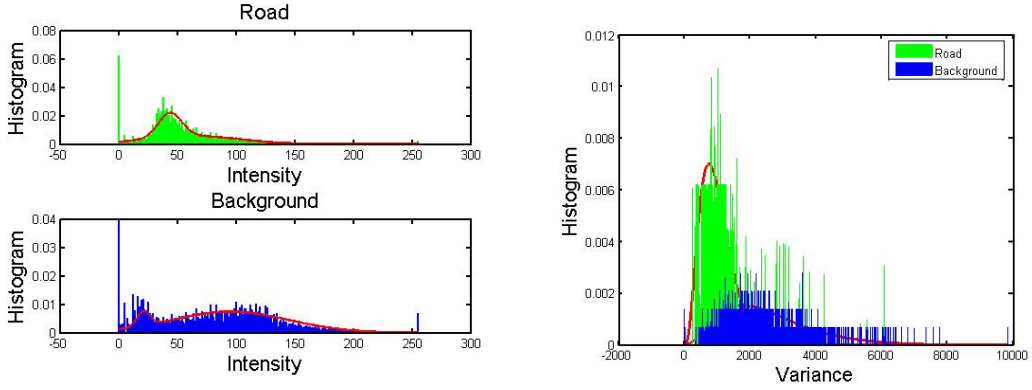


Figure 2.10. Left: histograms of the pixel intensity I on-road (top) and off-road (bottom); right: histograms of the variances V on-road (green/light grey) and off-road (blue/dark grey), and the models fitted to them (solid lines).

For each class (road and background), we compute the histograms of the pixel intensities I and the variances V , from which we estimate the parameters of two-component Gaussian mixture distributions $(a_{\pm}; \mu_{1\pm}, \sigma_{1\pm}^2; \mu_{2\pm}, \sigma_{2\pm}^2)$ and of Gamma distributions $(b_{\pm}, c_{\pm}, d_{\pm})$:

$$P_{\pm}(I) = a_{\pm}N(I; \mu_{1\pm}, \sigma_{1\pm}^2) + (1 - a_{\pm})N(I; \mu_{2\pm}, \sigma_{2\pm}^2), \quad (2.22a)$$

$$Q_{\pm}(V) = \frac{V^{b_{\pm}}}{d_{\pm}} e^{-\frac{V}{c_{\pm}}}, \quad (2.22b)$$

where $+$ denotes the road and $-$ denotes the background, $a_{\pm} \in [0, 1]$, and N is the normal distribution. Examples of histograms and fitted models are shown in Figure 2.10.

The goal of introducing a region-based data term, rather than an edge-based data term, is to make use of all the information in the VHR image. Because in the edge-based data term, the *local* descriptor, such as the gradient vector, considers neighborhoods that are very small compared to the size of the image, the algorithm may easily get trapped in one of the many local minima of the energy. Although in (Rochery et al., 2005a), this problem is partially alleviated by introducing nonlocal interactions between the phase field and the data in an edge-based data term, the low and medium resolution image data used in that application is less complicated than the VHR image data. A data term based on high gradients along the edges of roads is clearly insufficient in VHR images, where the edges have rapid changes of contrast. So it is of great importance to introduce more *global* information about the data into the model.

2.4.2 Optimization and Parameter Setting

We use the standard prior model $E_0 + E_S$ plus the above data model E_D as a primary model E_{primary} for extracting road networks from VHR images. To minimize the total energy E_{primary} with respect to ϕ , we perform gradient descent with the neutral initialization: the

initial value ϕ_{init} is set equal to the threshold $z = \alpha/\lambda$ everywhere in Ω (Rochery et al., 2005a). In this sense, it corresponds to the maximum of the potential, and hence is not biased towards one minimum or the other; moreover, ϕ_{init} is not biased towards interior or exterior, since both interior and exterior regions are empty. Because the initialization is automatic, the algorithm is quasi-automatic.

The functional derivative of E_{primary} is given by

$$\begin{aligned} \frac{\delta E_{\text{primary}}(\phi)}{\delta \phi(x)} = & \theta \left\{ -\nabla^2 \phi(x) + \lambda(\phi^3(x) - \phi(x)) + \alpha(1 - \phi^2(x)) + \beta \int_{\Omega} \phi(x') \nabla^2 \Psi\left(\frac{|x - x'|}{d}\right) dx' \right\} \\ & + \frac{1}{2} \ln \frac{P_-}{P_+} + \frac{\theta_v}{2} \ln \frac{Q_-}{Q_+} . \end{aligned} \quad (2.23)$$

The derivative $\delta E_S / \delta \phi$ is nonlocal. To avoid performing explicit convolutions, it is calculated in the Fourier domain. The resulting evolution equation is

$$\begin{aligned} \frac{\partial \phi(x)}{\partial t} = & \theta \left\{ \nabla^2 \phi(x) - \lambda(\phi^3(x) - \phi(x)) - \alpha(1 - \phi^2(x)) + \beta \mathcal{F}^{-1} \left\{ k^2 d \hat{\Psi}(kd) \hat{\phi}(k) \right\} \right\} \\ & + \frac{1}{2} \ln \frac{P_+}{P_-} + \frac{\theta_v}{2} \ln \frac{Q_+}{Q_-} , \end{aligned} \quad (2.24)$$

where \mathcal{F} and \mathcal{F}^{-1} denote the Fourier and the inverse Fourier transform respectively, and a hat ($\hat{\cdot}$) indicates the Fourier transform of a variable. In the discretized implementation, all derivatives are computed in the Fourier domain, while the time evolution uses the forward Euler method. At convergence, the region where $\phi_{\text{final}} > z$ belongs to the road network.

The parameters of the prior energy, *i.e.* λ, α, β , and d are constrained by the stability conditions discussed in section 2.3, and the general Turing stability condition (Rochery et al., 2005a). Only the weight θ of the prior energy remains. θ is a key parameter in tuning the behavior between the prior energy and the data energy. It reflects the confidence we have in the prior knowledge. There are also certain intuitions that can be used to give order of magnitude settings of this parameter: the forces due to the prior term should not be much greater than or much less than the forces due to the data term, otherwise the prior will dominate in one case and have little effect in the other. In our application on ‘noisy’ images of dense urban areas, θ is quite large, to prevent the model from being too sensitive to the ‘noise’ in the data. These intuitions are common to any energy-based method, however.

2.5 Conclusion

‘Phase field HOACs’ are powerful models for region-based and shape-based image segmentation. They have many advantages with regard to conventional methods in both model analysis and implementation. Most importantly, they can include sophisticated prior knowledge of region geometry. In this chapter, we have established the foundations of our models. Following a brief review of image segmentation from a probabilistic point of view, we have recalled the main principles of the HOAC model (Rochery et al., 2006), and subsequently the phase field HOAC model (Rochery et al., 2005a). The problem of stability and parameter tuning is a crucial issue in this model. We have shown that the precise performance

of the model greatly depends on the parameter values, and in particular on the ratio of the parameter describing the strength of the quadratic term and the parameter describing the strength of the area term. We have provided an efficient way to fix some of the parameters in terms of network width. We have presented an overall primary model adapting ‘phase field HOACs’ to the problem of urban road extraction from VHR images. In the following chapters, we will develop various new models based on this primary model for the purpose of solving different problems.

Chapter 3

Multiresolution Analysis of the Primary Model

A multiresolution representation provides a very effective hierarchical framework for analyzing data content. It has been widely used in both applied mathematics and signal processing. In this chapter, we examine the extraction of the main road network from a single QuickBird image. We show that the difficulty of performing this task from very high resolution (VHR) images of dense urban areas suggests naturally a multiresolution analysis (MRA). In section 3.1, following an explanation of our motivations for this idea, we give a brief review of wavelet based MRA: the general mathematical background, and one of its specific forms, *i.e.* Haar MRA. In section 3.2, we introduce a multiresolution statistical data model, which integrates the data information at different resolution levels. We then describe a two-step robust multiresolution framework. In this framework, we first segment a low resolution image. Since reducing the resolution eliminates high frequency noise, the resulting pre-segmentation gives an approximate detection of the objects of interest. This rough result is then incorporated into the model, to force the final segmentation to lie ‘close’ to the pre-segmentation. In section 3.3, we present experiments on a VHR QuickBird image at several different resolutions using the primary model introduced in the previous chapter, and the preliminary improvement using the multiresolution data model. Finally, we show experimental results using the multiresolution framework. Comparisons with other methods demonstrate quantitatively its robustness and efficiency.

3.1 Multiresolution Analysis and Wavelets

With the aim of tackling the complexity of the information existing in VHR images, we propose to take advantage of MRA. The motivation for this idea originates in the following three observations:

- Real-world objects, in particular the objects observed in VHR images, *e.g.* roads,

buildings, are composed of different structures of different scales. In order to extract objects of different scales, we should analyze an image at several resolutions.

- The large dimension of VHR images compels us to “optimize” computation speed. Computational efficiency may be improved by coarse-to-fine strategies.
- Real-world objects may appear in different ways depending on the scale of observation. Some relevant details in images exist only over a limited range of scales (as opposed to certain ideal mathematical entities such as ‘point’ or ‘line’, which appear in the same way at all scales of observation). Specifically, in our image data, at low resolutions, the background can be viewed as noise, while the larger roads are still clearly distinguished as homogeneous regions. In contrast, at higher resolutions, a large amount of noise appears, but on the other hand, high resolution images can provide a more precise location and width for the roads. Thereby, road segmentation at low resolutions, compared to that at high resolutions, is facilitated, but is also less precise. The use of several resolutions thus allows the combination of coarse data – in which details in the image that can disrupt the recognition process have been eliminated – with fine data to increase precision.

The concept of MRA using a wavelet decomposition was initiated by Meyer (1992) and Mallat (1989, 1999). The original signal/data is represented using an orthonormal wavelet basis, and can be interpreted as a decomposition using a set of independent frequency channels having a spatial orientation. We explain, in the rest of this section, the formal definition of MRA, and show how to use the wavelet decomposition. Then, we describe the canonical form of Haar MRA.

3.1.1 Definition of Wavelet Based Multiresolution Analysis

A function or signal $f(t)$ can be viewed as composed of a smooth background and fluctuations or details on top of it. The distinction between the smooth part and the details is determined by the resolution, *i.e.* by the scale below which the details of the given signal cannot be discerned. We label the resolution level by j . The scale associated with the level $j = 0$ is set to unity, and that with the level j is $1/2^j$ ¹. At a given resolution j , an approximated signal $f_j(t)$ is obtained by ignoring all details below that scale. Let $d_j(t)$ be the details included at the next level of resolution $j + 1$. We can imagine progressively increasing the resolution; at each stage of increasing the resolution from j to $j + 1$, finer details $d_j(t)$ are added to the coarser description $f_j(t)$, providing $f_{j+1}(t) = f_j(t) + d_j(t)$, a successively better approximation to the original signal. Eventually, when the resolution goes to infinity, the exact original signal $f(t)$ is recovered:

$$f(t) = f_j(t) + \sum_{k=j}^{+\infty} d_k(t) . \quad (3.1)$$

¹To facilitate this explanation, here we adopt the general definition of the resolution. Note that the resolution ‘level’ in our problem, which we will describe later, has a different meaning.

The $L^2(\mathbb{R})$ space can be decomposed into a sequence of subspaces $\{W_j\}_{j \in \mathbb{Z}}$ and $\{V_j\}_{j \in \mathbb{Z}}$, such that for a specific level of resolution j , the details $d_j(t)$ are found in W_j (detail space), and correspondingly $f_j(t)$ in V_j (approximation space).

Definition of multiresolution analysis. An MRA of $L^2(\mathbb{R})$ is a nested sequence of embedded closed subspaces $\{V_j\}_{j \in \mathbb{Z}}$ (\mathbb{Z} is the set of integers) with the following conditions (Wolfram, 1999):

1. $\dots \subset V_{-1} \subset V_0 \subset V_1 \dots \subset L^2(\mathbb{R})$.
2. $\bigcap_j V_j = \{0\}$, and $\bigcup_j V_j$ is dense in $L^2(\mathbb{R})$.
3. $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$.
4. $f(t) \in V_0 \Rightarrow f(t - k) \in V_0$.
5. There exists a *scaling function* $\varphi \in V_0$, such that $\{\varphi(t - k)\}$ is an orthonormal basis of V_0 .

The first condition specifies that the subspace V_j is contained in all the higher subspaces, since information at resolution level j is necessarily included in the information at a higher resolution. The second condition states that the only function that can be approximated at an arbitrarily coarse scale is the zero function ($\lim_{j \rightarrow -\infty} V_j = \{0\}$), and if the resolution instead goes to infinity, the entire $L^2(\mathbb{R})$ space should be recovered ($\lim_{j \rightarrow +\infty} V_j \rightarrow L^2(\mathbb{R})$). The third condition defines the difference between two successive resolutions. This requirement of scale or dilation invariance means that V_{j+1} approximates functions at a twice finer scale than V_j . It is the key condition that determines dyadic refinement. The fourth condition is the requirement of translation or shift invariance of the space V_j .

The final condition requires the approximation spaces $\{V_j\}$ to be spanned by scaling functions. Using the scale and translation invariance, the normalized scaling functions $\{\varphi_{jk}(t)\}$, where $\varphi_{jk}(t) = 2^{j/2} \varphi(2^j t - k)$, form an orthonormal basis for the space V_j . Since $V_0 \subset V_1$, the scaling function $\varphi \in V_0$ can be expanded in terms of the basis function of V_1 , $\{\varphi_{1k}(t)\}$:

$$\varphi(t) = \sqrt{2} \sum_k h_k \varphi(2t - k), \quad (3.2a)$$

$$h_k = \sqrt{2} \int_{-\infty}^{+\infty} \varphi(t) \varphi(2t - k) dt. \quad (3.2b)$$

Equation (3.2a) is known as the *dilation equation*, which relates the scaling functions at two consecutive scales.

For an MRA, the detail space W_j between two consecutive spaces V_j and V_{j+1} is the complementary space of V_j , i.e. $V_{j+1} = V_j \oplus W_j$. Similar to the fact that the scaling function $\varphi_{jk}(t)$ defines a family of bases for the approximation space V_j , it would be useful to have a compact representation of its complementary space W_j . In this context, the *wavelet function* is introduced. The space W_j has an orthonormal wavelet basis $\{\psi_{jk}(t)\}$, where $\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k)$ is the scaled and translated wavelet. The function $\psi(t)$ that generates all the basis functions of the W space is referred to as the basic wavelet or the mother wavelet.

Since $W_0 \subset V_1$, the wavelet function $\psi \in W_0$ can also be expanded in terms of the basis functions of V_1 , $\{\varphi_{1k}(t)\}$:

$$\psi(t) = \sqrt{2} \sum_k g_k \varphi(2t - k) , \quad (3.3a)$$

$$g_k = \sqrt{2} \int_{-\infty}^{+\infty} \psi(t) \varphi(2t - k) dt . \quad (3.3b)$$

Equation (3.3a) is called the *wavelet equation*, which relates the basic wavelet to the scaling function at the next finer scale.

Moreover, Mallat (1989) proved that the coefficients $\{g_k\}$ in the wavelet equation (equation (3.3a)) are related to the coefficients $\{h_k\}$ in the dilation equation (equation (3.2a)) through the expression

$$g_k = (-1)^k h_{1-k} . \quad (3.4)$$

We have seen that wavelets constitute a special class of basis functions for $L^2(\mathbb{R})$, which are usually enough for the multiresolution representation. However, in many applications, it is more convenient to work with the spaces $\{V_j\}$ explicitly, since the scaling function has analytic and operational properties. Once the scaling function φ is chosen, the coefficients $\{h_k\}$ are obtained from equation (3.2b). In turn, the coefficients of $\{g_k\}$ are known from equation (3.4). Finally, using the wavelet equation (equation (3.3a)), we have the wavelet function ψ .

Subsequently, a multiresolution representation can be obtained via the so-called *wavelet transform*:

$$f_{j+1}(t) = f_j(t) + d_j(t) = \sum_k s_{jk} \varphi_{jk}(t) + \sum_k w_{jk} \psi_{jk}(t) , \quad (3.5)$$

where the scaling coefficients s_{jk} and the wavelet coefficients w_{jk} can be computed from the scaling coefficients at the next finer level $s_{(j+1)k}$, and vice versa. In the orthogonal case, the wavelet transform takes the form:

$$s_{jk} = \sum_l h_{l-2k} s_{(j+1)l} , \quad (3.6a)$$

$$w_{jk} = \sum_l g_{l-2k} s_{(j+1)l} . \quad (3.6b)$$

The corresponding relation that is used in the *inverse wavelet transform* is

$$s_{(j+1)l} = \sum_k (h_{l-2k} s_{jk} + g_{l-2k} w_{jk}) . \quad (3.7)$$

The direction from finer to coarser in equation (3.6) is called the *decomposition process* or the *analysis process*; the direction from coarser to finer in equation (3.7) is called the *reconstruction process* or the *synthesis process*.

3.1.2 Haar Multiresolution Analysis

The Haar wavelet is the simplest form of possible wavelets. Let V_0 be the space of functions $f(t) \in L^2(\mathbb{R})$ that are constant on the unit intervals $[k, k+1)$ for integer k , and let V_j be the space of functions that are constant on the intervals $[k/2^j, (k+1)/2^j)$. We can easily verify that this definition satisfies all the conditions on the subspaces of an MRA. The scaling function $\varphi(t)$ is chosen to be the characteristic function on the interval $[0, 1)$:

$$\varphi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1, \\ 0 & \text{else,} \end{cases} \quad (3.8)$$

and the basic wavelet function of Haar wavelet, $\psi(t)$, is given by

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2, \\ -1 & \text{if } 1/2 \leq t < 1, \\ 0 & \text{else.} \end{cases} \quad (3.9)$$

In this thesis, the Haar wavelet transform (Mallat, 1989, 1999) is used to generate a multiresolution version of the original image at different scales (levels) (see Figure 3.1). As already noted, in our problem, the resolution ‘level’ l is different from that in the general definition in subsection 3.1.1. The data at level $l = 0$ corresponds to the original image; and the data at level l are the scaling coefficients after performing the wavelet transform l times.

3.2 Model Definition at Multiple Resolutions

Many tasks in computer vision defined in term of energy minimization suffer from a general difficulty, that is, the algorithm can easily become stuck in irrelevant sub-optimal local minima. The reason for this drawback is that even if well-defined, these applications typically require finding the minimum of a complicated (*i.e.* non-convex) underlying cost functional in a very high-dimensional space. For instance, in our case, the problem of image segmentation is defined as distinguishing roads from their background. The discrete version of this problem can be viewed as labeling each pixel as either foreground or background. Thus, the space of possible segmentations is 2^N , where N is the number of the pixels in the image. Searching over all such solutions is intractable even for small images. Smoothing the original image (by reducing the resolution) is equivalent to smoothing the energy functional itself, which has the consequence of diminishing the number of local minima. However, unfortunately, as the images at higher resolutions get finer, more complex, have higher dimensionality, and more complicated constructions, more local minima appear due to the increase in complexity of the energy functional. The geometric prior knowledge included in the primary model is insufficient to overcome this effect thoroughly, which compels us to seek another, better solution.

Therefore, to tackle the complexity of information contained in VHR images, in this section, we first propose a multiresolution statistical data model integrating the data information from multiple resolutions. We then consider a two-step robust multiresolution

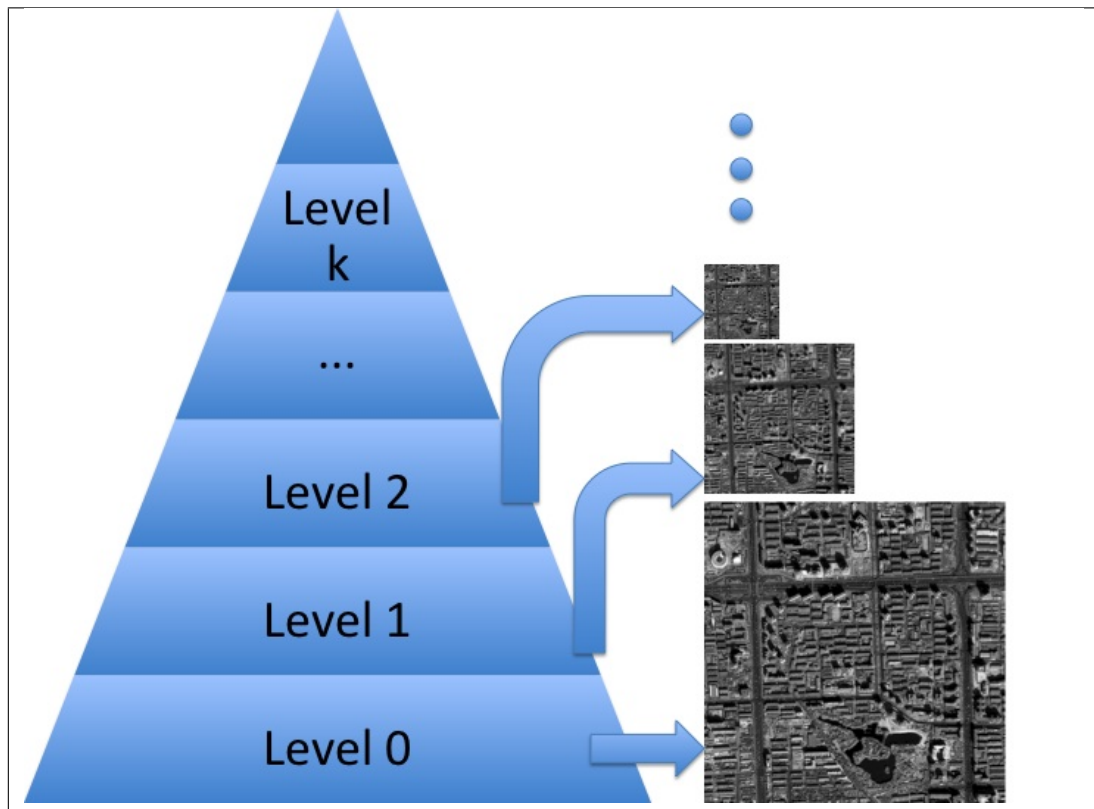


Figure 3.1. The multiresolution representation of the original data via a wavelet transform.

framework to overcome the remaining problems. In this framework, an approximate detection is first obtained at a coarse resolution; at the original resolution, this preliminary result is incorporated into the model to serve as a flexible constraint.

3.2.1 Multiresolution Data Model

First, let us recall the energy functional of the primary model E_{primary} at a single resolution, which was introduced in section 2.4:

$$E_{\text{primary}} = \theta(E_0 + E_S) + E_D, \quad (3.10)$$

where

$$E_0(\phi) = \int_{\Omega} \left\{ \frac{1}{2} \nabla \phi(x) \cdot \nabla \phi(x) + \lambda \left(\frac{1}{4} \phi^4(x) - \frac{1}{2} \phi^2(x) \right) + \alpha \left(\phi(x) - \frac{1}{3} \phi^3(x) \right) \right\} dx, \quad (3.11a)$$

$$E_S(\phi) = -\frac{\beta}{2} \iint_{\Omega^2} \nabla \phi(x) \cdot \nabla \phi(x') \Psi\left(\frac{|x-x'|}{d}\right) dx dx', \quad (3.11b)$$

$$E_D(I, \phi) = - \int_{\Omega} \left\{ [\ln P_+(I(x)) + \theta_v \ln Q_+(V(x))] \phi_+(x) + [\ln P_-(I(x)) + \theta_v \ln Q_-(V(x))] \phi_-(x) \right\} dx. \quad (3.11c)$$

Our first attempt to improve the data model is to propose a multiresolution statistical data model with fusion of information from multiple resolutions. As already mentioned, at coarser resolutions, the extraction task is easier. On the other hand, image pixels on the edge of the road region are the average values of road pixels and background pixels in the full resolution image; consequently, the accuracy of the result is limited by the reduced resolution of the image. Conversely, at full resolution, while greater precision is in principle attainable, a great deal of ‘noise’ appears too, rendering the extraction task very difficult. With the aim of combining the advantages and balancing the disadvantages of coarser and finer resolution data, we introduce a multiresolution data energy, $E_{D,\text{MUL}}$, defined as the sum of energies computed at several different levels:

$$E_{D,\text{MUL}}(I, \phi) = \sum_l E_{D,l}(I_l, \phi), \quad (3.12)$$

where $I_l, l \in \{0, 1, \dots, k\}$, are the scaling coefficients at level l of the Haar wavelet transform. $E_{D,l}$ is the data energy recalled in equation (3.11c). For the scaling coefficients at each considered resolution l , the statistics of intensity and variance are computed, for each of the two classes (main road, background). Therefore, at each level, the parameters of two-component Gaussian mixture distributions $P_{\pm,l}$ and of Gamma distributions $Q_{\pm,l}$, $l \in \{0, 1, \dots, k\}$ are determined. One can interpret the energy from a maximum entropy point of view, in which case it amounts to assuming that the mean energy at each level is fixed. In practice, since the size of the image varies with a factor of 2^2 from level l to level $l+1$, we up-sample all I_l

to the finest resolution. The evolution equation for multiple resolutions involves replacing the data part in equation (2.24) by

$$\sum_{l=0}^k \left\{ \frac{1}{2} \ln \frac{P_{+,l}}{P_{-,l}} + \frac{\theta_v}{2} \ln \frac{Q_{+,l}}{Q_{-,l}} \right\}. \quad (3.13)$$

We refer to the energy $E_{\text{primary,mul}} = \theta(E_0 + E_S) + E_{D,\text{MUL}}$ (equations (3.11a), (3.11b) and (3.12)) as the multiresolution model.

3.2.2 Multiresolution Framework

The multiresolution framework presented in this subsection looks at the coarse scale nature of a problem before considering the fine scale nature. In fact, such an idea has been introduced into many active contour models, to improve noise resistance and speed up processes. For example, in (Leroy et al., 1996), the image is down-sampled to a coarse scale, and a ‘snake’ is evolved until convergence. The resulting active contour is up-sampled to a finer scale of the image, and used as its initialization. The process is continued on successively finer scale representations of the image until the active contour is evolved in the image itself. Since in the classical active contour, the initial contour cannot be too far away from the desired boundary, a big problem of multiscale techniques like (Leroy et al., 1996) is that the result at the previous coarser level gives a very strong constraint on the result at the next finer level, and finally on the result in the original image. In other words, the final contours at all the higher resolutions just move slightly around the final contour at the coarsest resolution. This is problematic in many applications: because errors that occur at the beginning due to the reduced resolution may not be corrected during the evolutions at the higher resolutions; in addition, many small objects, which can be discerned at full resolution, are probably invisible at a reduced resolution.

In contrast, in our proposed framework, prior to the segmentation of the original image, we extract, by a pre-segmentation of a reduced resolution image, an effective zone where the geometric prior knowledge carried by the standard higher-order active contour (HOAC) term E_S will be included. Although the final segmentation at full resolution is still somehow forced to lie close to the rough pre-segmentation, this constraint in our framework is quite flexible so as to permit reasonable changes everywhere in the image domain at full resolution. This property will be clearly demonstrated by the experimental results.

To clarify the above heuristic explanation, the detailed algorithm scheme is as follows:

1. The energy $E_{\text{primary}} = \theta(E_0 + E_S) + E_D$ (equation (3.10)) is applied to the data at a coarse resolution. At convergence, a rough pre-segmentation is obtained.
2. This first result is up-sampled to the full resolution through nearest neighbor interpolation. Let R_0 be the interpolated result, whose phase field function is ϕ_{R_0} . Then, we define a weight factor for each pixel by $\phi_{R_{0+}}(x) = (1 + \phi_{R_0}(x))/2$. As a result, $\phi_{R_{0+}} \simeq 1$ for $x \in R_0$; and $\phi_{R_{0+}} \simeq 0$ for $x \notin R_0$.

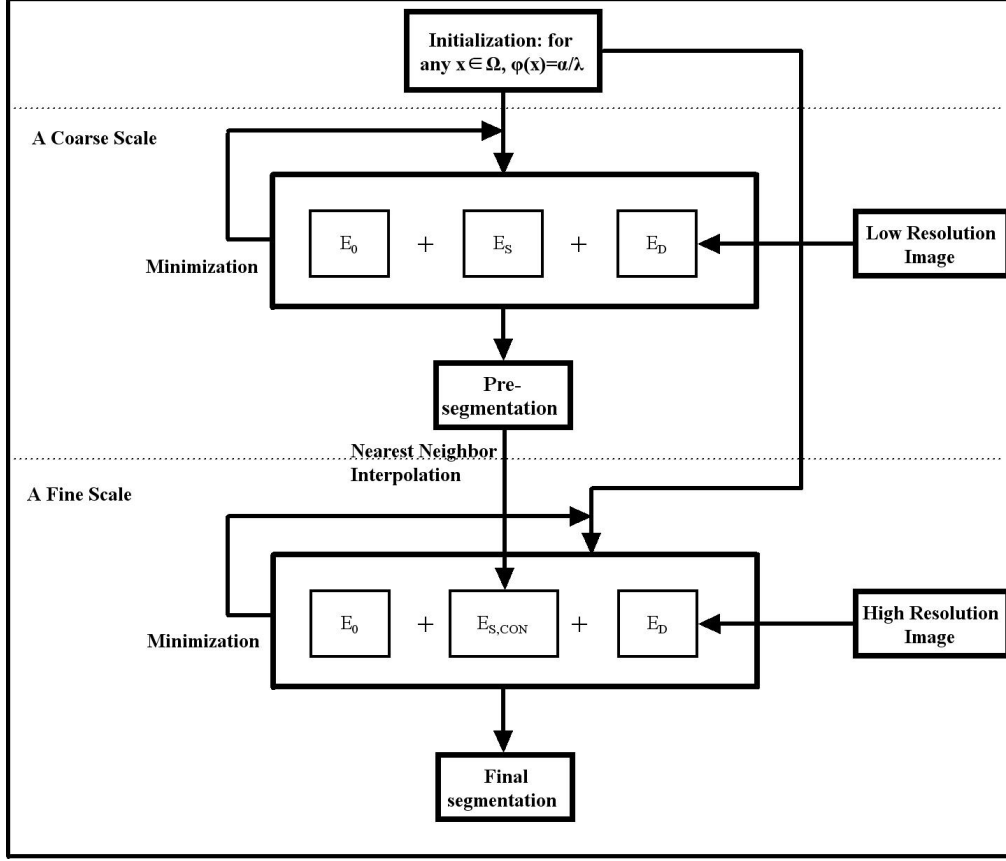


Figure 3.2. The flowchart of the proposed multiresolution framework.

3. At full resolution, a weight constrained quadratic term is formulated as

$$E_{S,CON}(\phi) = -\frac{\beta}{2} \iint_{\Omega^2} \nabla \phi(x) \cdot \nabla \phi(x') \Psi\left(\frac{|x - x'|}{d}\right) \phi_{R0+}(x) dx dx' . \quad (3.14)$$

The energy $E_{HR} = \theta(E_0 + E_{S,CON}) + E_D$ is applied to the original image. The geometric prior knowledge encoded by the quadratic term $E_{S,CON}$ thus takes effect only near the pre-segmentation result. This can significantly diminish false detections in the background, such as buildings, shadows, etc., which have similar geometric properties to the object. On the other hand, thanks to the effects of other terms (mainly the data term), components of the object can still be extracted in the exterior of the pre-segmentation region.

Note that E_0, E_S, E_D are recalled in equations (3.11). The flowchart is illustrated in Figure 3.2.

3.3 Experimental Results and Comparisons

The input data are several pieces of a QuickBird panchromatic image of dense urban areas. To extract the main road networks, we carry out a series of experiments using the primary model at a single resolution, the multiresolution model, and the two-step multiresolution framework. The advantages of the proposed approaches are highlighted in these experimental results and the comparisons with other methods in the literature.

3.3.1 Results Using the Single-Resolution Model

First we apply the single-resolution primary model E_{primary} (equation (3.10)) to the scaling coefficients of the original image in Figure 1(a) at different levels of the wavelet decomposition (Mallat, 1999). The implementation details, described in subsection 2.4.2, are followed. We start at level 3. Figure 3.3 shows the thresholded phase field function at iterations 1 and 400 of gradient descent, and at the final iteration 18,000. The parameters of the model are set to $(\theta, \theta_v, \lambda, \alpha, \beta, d) = (200, 0.02, 3, 0.0905, 0.02, 10)$. The computational time is about 15 minutes². Thanks to the incorporation of strong geometric prior knowledge, the segmentation of the main road network appears very successful, but the road region is actually not very accurate. This observation is also supported by quantitative measures. Compare the first row of Table 3.1 to the first row of Table 3.2. Accuracy is limited both directly, by the low resolution of the phase field, and indirectly, because each scaling coefficient in the data at level 3 is the average of 8×8 pixels at full resolution. Coefficients near the road border therefore include both road and background contributions, and the road width is thereby distorted.

The level 3 image is already quite complex, and we observe experimentally that if we try to apply the same single-resolution primary model E_{primary} at finer resolutions, using the images at levels 2, 1, or 0, the details of the scene in the image make road extraction more difficult (see Figure 3.4). These experiments show, as expected, that this primary model is not sufficient to overcome the difficulties of VHR images at finer resolutions. The erroneous detections in the background result from regions of poor contrast between the roads and the buildings or areas of vegetation, which have statistical properties similar to the roads. On the other hand, the shadows of high buildings, cars, road markings and bridges lead to jagged borders or gaps along the roads. The former indicates a lack in the single level data model, while the latter is due to a weakness in the prior model, which therefore needs to be improved in order to enforce the road geometry more effectively.

This set of experiments shows the primary model is promising, suggesting that at least at lower resolutions, for the main road network, the previous (standard) prior model could be kept. It is clear, though, that to move to higher resolution, and thereby profit from the

²In fact, the computational time depends on the number of iterations and the number of operations at each iteration. The number of iterations depends greatly on the complexity of the considered image scene. If radiometric properties of the road are very similar to those of the background, it takes longer time to make false detections in the background disappear. Second, the set of parameters which leads to the convergence of the solution is not unique: it is possible to find other parameter settings which may obtain the similar result but converge faster. Third, the computational efficiency is also related to the computer's CPU. Results presented in this thesis were obtained with several different computers.

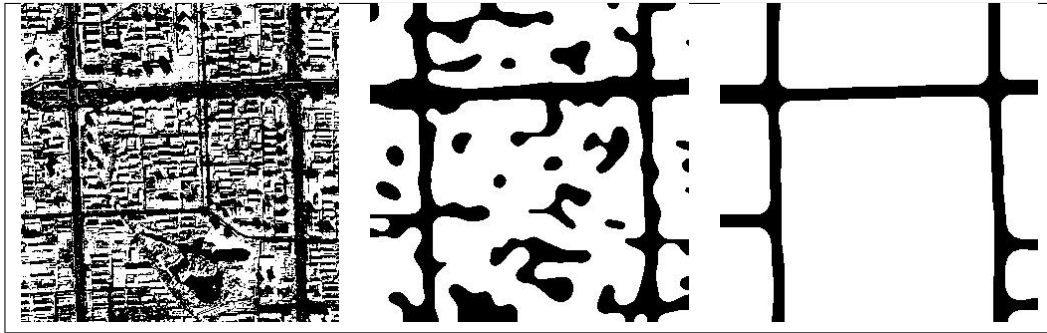


Figure 3.3. Experiment on Figure 1(a) at level 3 (size: 320×320 , road width ≈ 12 pixels), using the single-resolution primary model E_{primary} . Left to right: the thresholded phase field function at iterations 1 and 400, and at the final iteration 18,000.

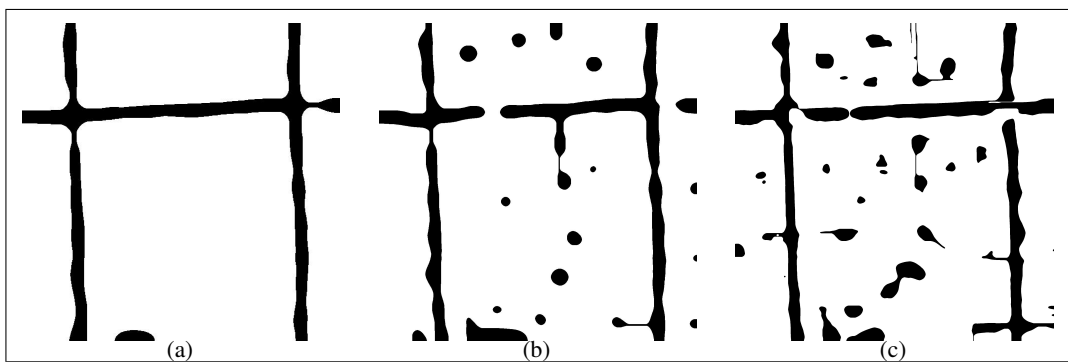


Figure 3.4. Experiments on Figure 1(a) at finer resolutions, using the single-resolution primary model E_{primary} . (a): result at level 2 (size: 640×640 , road width ≈ 24 pixels); (b): result at level 1 (size: 1280×1280 , road width ≈ 48 pixels); (c): result at full resolution, level 0 (size: 2560×2560 , road width ≈ 96 pixels).

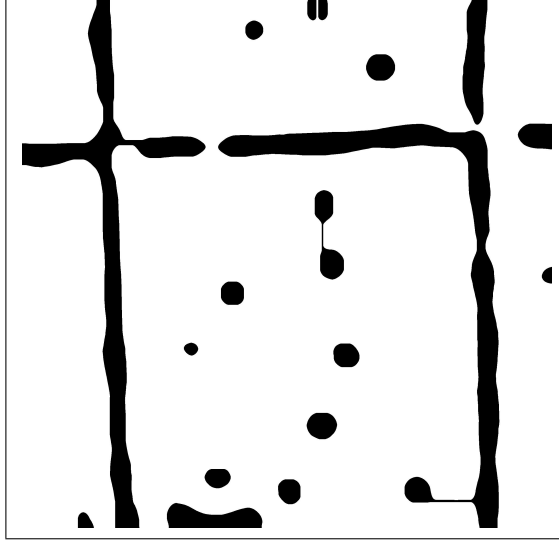


Figure 3.5. Experiment on Figure 1(a) at full resolution (size: 2560×2560 , road width ≈ 96 pixels), using the multiresolution model $E_{\text{primary,mul}}$.

precision offered by VHR data, both the data model and the prior model of the road network have to be changed.

3.3.2 Results Using the Multiresolution Model

In an attempt to overcome the problems at fine resolutions, we use the multiresolution model $E_{\text{primary,mul}}$ and apply it to the image in Figure 1(a). We start with information from the level at which the model energy with a single level works well, *i.e.* level 3. Thus, information from levels 3, 2, 1, and 0 is used.

The result is shown in Figure 3.5. It is not perfect, but it is an improvement over the result obtained at full resolution using the single-resolution model E_{primary} (see Figure 3.4(c)). It shows that MRA is an encouraging avenue. However, there are still some false detections in the background and the road borders are rather inaccurate due to geometric noise along the boundaries of the road. The result indicates that a simple sum of data energies at several different scales, while helpful, is not sufficient to solve the problem completely. We propose thereafter an alternative more efficient way to integrate the information from multiple resolutions.

3.3.3 Results Using the Multiresolution Framework and Comparisons

In this subsection, we first apply the single-resolution primary model E_{primary} to the data at a reduced resolution, and then integrate the segmentation result into the high-resolution model E_{HR} for the original data. To evaluate the performance of our model, the results at both resolutions are compared to those obtained with four other methods: maximum likelihood

Method \ Measure	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
E_{primary} (Figure 3.6(d))	0.9517	0.7823	0.7525
MLE (Figure 3.8(a))	0.9409	0.1607	0.1591
$\theta E_0 + E_D$ (Figure 3.8(b))	0.8468	0.6007	0.5418
Wang (Figure 3.8(c))	0.9028	0.5790	0.5450
Yu (Figure 3.8(d))	0.9776	0.7077	0.6964

Table 3.1. Quality measures for the different methods tested on Figure 3.6(a) at 1/4 resolution (T = true, F = false, P = positive, N = negative). The completeness is the percentage of ground truth road network that is extracted; the correctness is the percentage of extracted road network that is correct; and the quality is the most important measure of the “goodness” of the result, because it takes into account the completeness and the correctness.

estimation (MLE, *i.e.* $\theta = 0$); a standard active contour formed by dropping the quadratic term E_S or $E_{S,\text{CON}}$ in our model (*i.e.* $\beta = 0$); a classification, tracking, and morphology algorithm by Wang and Zhang (2003); and a fast but rough segmentation technique based on “straight line density” by Yu et al. (2004). The two last approaches are detailed in Appendix E.

The first example of extraction is shown in Figure 3.6(a). The parameters $(\theta, \theta_v, \lambda, \alpha, \beta, d)$ are (200, 0, 3, 0.0905, 0.03, 20) at 1/4 resolution, and correspondingly (2000, 0, 3, 0.0905, 0.03, 80) at full resolution. The computational time is 3 minutes and 281 minutes respectively at the two resolutions. Figures 3.6(b)-3.6(d) show the thresholded phase field function at iterations 1 and 100 of gradient descent, and at convergence at iteration 2, 200, using the single-resolution primary model E_{primary} on the 1/4 resolution data. The main road networks are mostly retrieved, but since the information has been simplified at the reduced resolution, there are still some problems in this result: parts of the parking lot are mistakenly included; the position of the bottom road somewhat shifts. When applying the single-resolution primary model E_{primary} to the original image, due to the appearance of too much detail and noise, the extraction result is not satisfactory (see Figure 3.7(a)). When we use the high-resolution model E_{HR} instead, Figure 3.7(b) shows the final result in the original image. Although the computational time is longer at full resolution, the segmentation is very successful: the whole of the road network has been extracted completely and accurately.

By comparison, Figures 3.8 and 3.9 show the results obtained using the other four me-

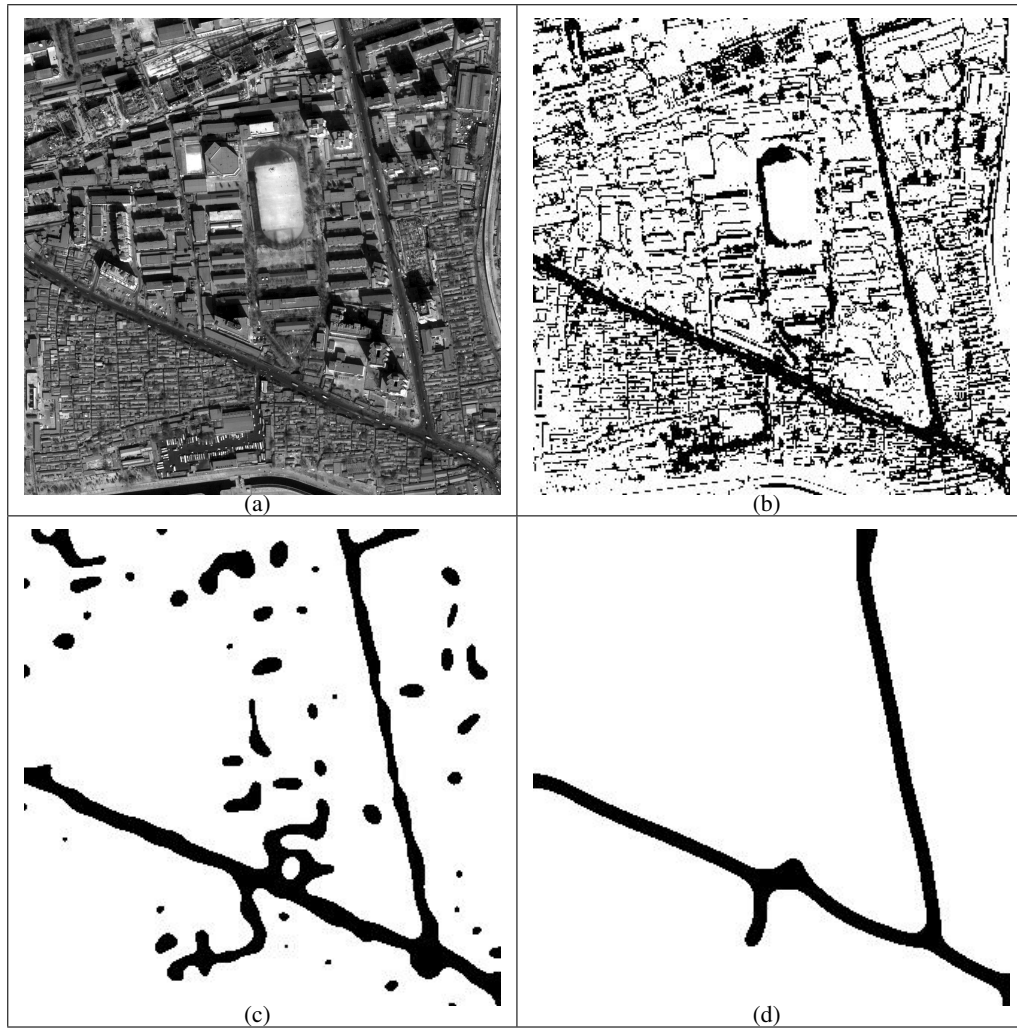


Figure 3.6. Data, and experiment at $1/4$ resolution. (a): a QuickBird image (size: 1280×1280); (b)-(d): the thresholded phase field function at iterations 1, 100, and at convergence at iteration 2,200, using the single-resolution primary energy E_{primary} .

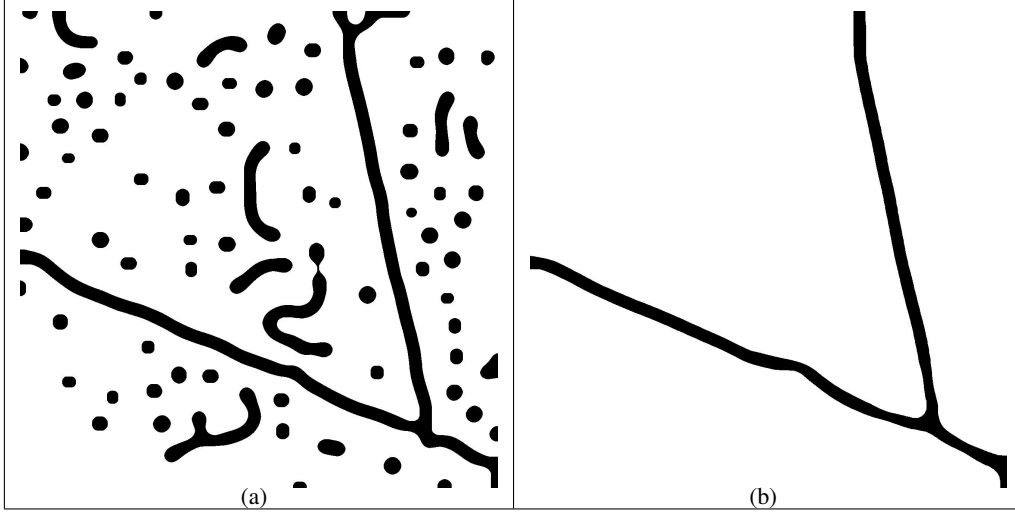


Figure 3.7. Experiments on Figure 3.6(a) at full resolution. (a)-(b): results obtained respectively with the single-resolution primary energy E_{primary} and the high-resolution energy E_{HR} .

thods mentioned earlier, at 1/4 resolution and at full resolution, while Tables 3.1 and 3.2 present some quantitative evaluation measures (Heipke et al., 1997) of all the results at both resolutions. The ground truth used to calculate quantitative criteria is segmented by hand (see Figures 3.8(e) and 3.9(e)). The comparisons show that our model at both resolutions achieves the highest values of the ‘quality’ of the result. MLE is obviously far from enough to distinguish the roads from the background. The results obtained without the quadratic term, *i.e.* $\theta E_0 + E_D$, exemplify the importance of the geometric knowledge. The results using the method of Wang and Zhang (2003) show a great deal of noise, and the accuracy obtained in the delineation of the road boundary is poor. According to the quantitative measures, the method of Yu et al. (2004) is the second best after our model, but this approach is highly restricted to straight roads (for instance, the curved road at the bottom right corner in Figure 3.6(a) is missing).

Figure 3.10 presents another experiment on Figure 1(a) using the multiresolution framework, which shows significant improvement both on the result in Figure 3.4(c) and on the result in Figure 3.5. The pre-segmentation of the reduced resolution image is obtained at level 3 (Figure 3.3). The parameters $(\theta, \theta_v, \lambda, \alpha, \beta, d)$ used at full resolution are set up as (2000, 0.02, 3, 0.0905, 0.02, 80). They meet the requirement of model stability (section 2.3). With respect to the parameters for level 3, except the change of d which controls the road width, and the change of the overall weight θ of the prior energy (which results from the appearance of a great deal of noise at full resolution), the other parameters remain the same between the two resolutions. We will give evaluations of all the experiments on this image later when comparing to the new model proposed in the next chapter.

Figure 3.11 shows segmentation results on another piece of QuickBird image. The parameters $(\theta, \theta_v, \lambda, \alpha, \beta, d)$ at 1/4 and full resolution are respectively (200, 0, 3, 0.0905, 0.02, 20)

Method \ Measure	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
E_{HR} (Figure 3.7(b))	0.9329	0.9199	0.8628
$E_{primary}$ (Figure 3.7(a))	0.9766	0.3439	0.3411
MLE (Figure 3.9(a))	0.9340	0.1612	0.1594
$\theta E_0 + E_D$ (Figure 3.9(b))	0.8120	0.5855	0.5156
Wang (Figure 3.9(c))	0.9307	0.4707	0.4547
Yu (Figure 3.9(d))	0.9651	0.8092	0.7862

Table 3.2. Quality measures for the different methods tested on Figure 3.6(a) at full resolution (T = true, F = false, P = positive, N = negative). See Table 3.1 for an explanation of completeness, correctness and quality.

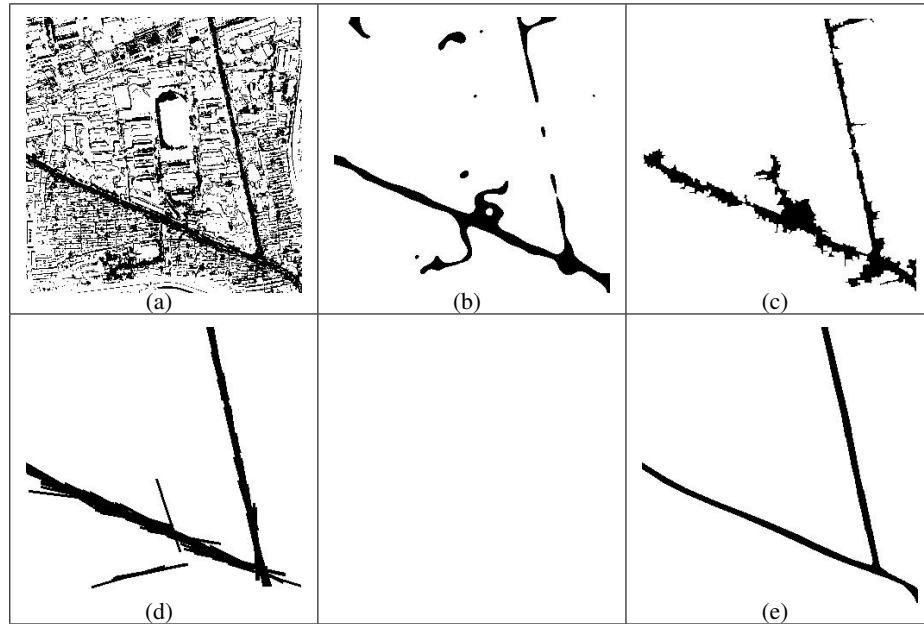


Figure 3.8. Comparisons on Figure 3.6(a) at 1/4 resolution. (a)-(d): results obtained with the MLE (*i.e.* $\theta = 0$), a standard active contour by dropping the quadratic term (*i.e.* $\beta = 0$), the work of Wang and Zhang (2003) and of Yu et al. (2004). (e): the ground truth.

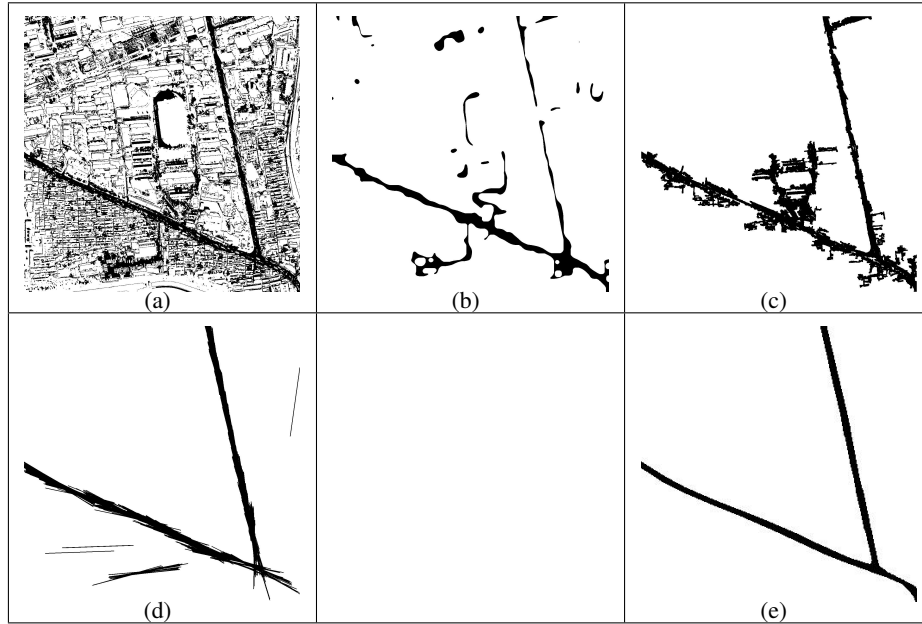


Figure 3.9. Comparisons on Figure 3.6(a) at full resolution. (a)-(d): results obtained with the MLE (*i.e.* $\theta = 0$), a standard active contour by dropping the quadratic term (*i.e.* $\beta = 0$), the work of Wang and Zhang (2003) and of Yu et al. (2004). (e): the ground truth.

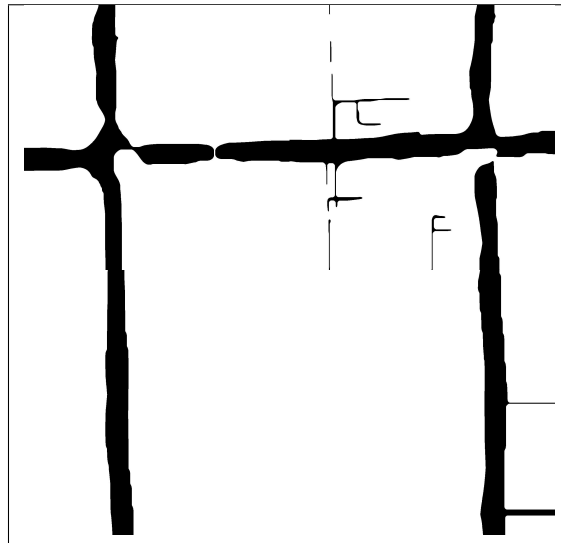


Figure 3.10. Experiment on Figure 1(a) at full resolution (size: 2560×2560 , road width ≈ 96 pixels), using the multiresolution framework.

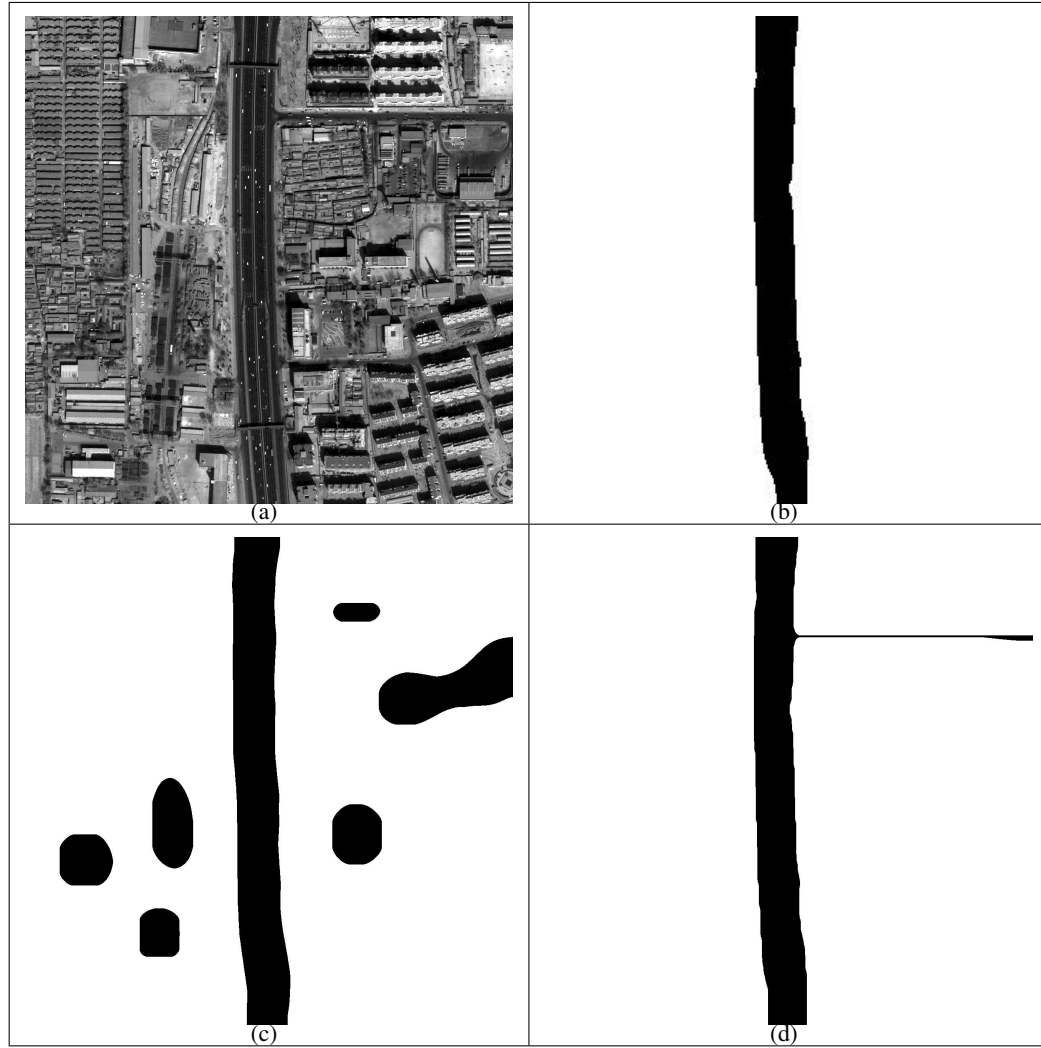


Figure 3.11. More experiments on the QuickBird image. (a): image data (size: 1280×1280); (b): result at 1/4 resolution, using the single-resolution primary model E_{primary} ; (c)-(d): results at full resolution, using the single-resolution primary model E_{primary} and the high-resolution model E_{HR} respectively.

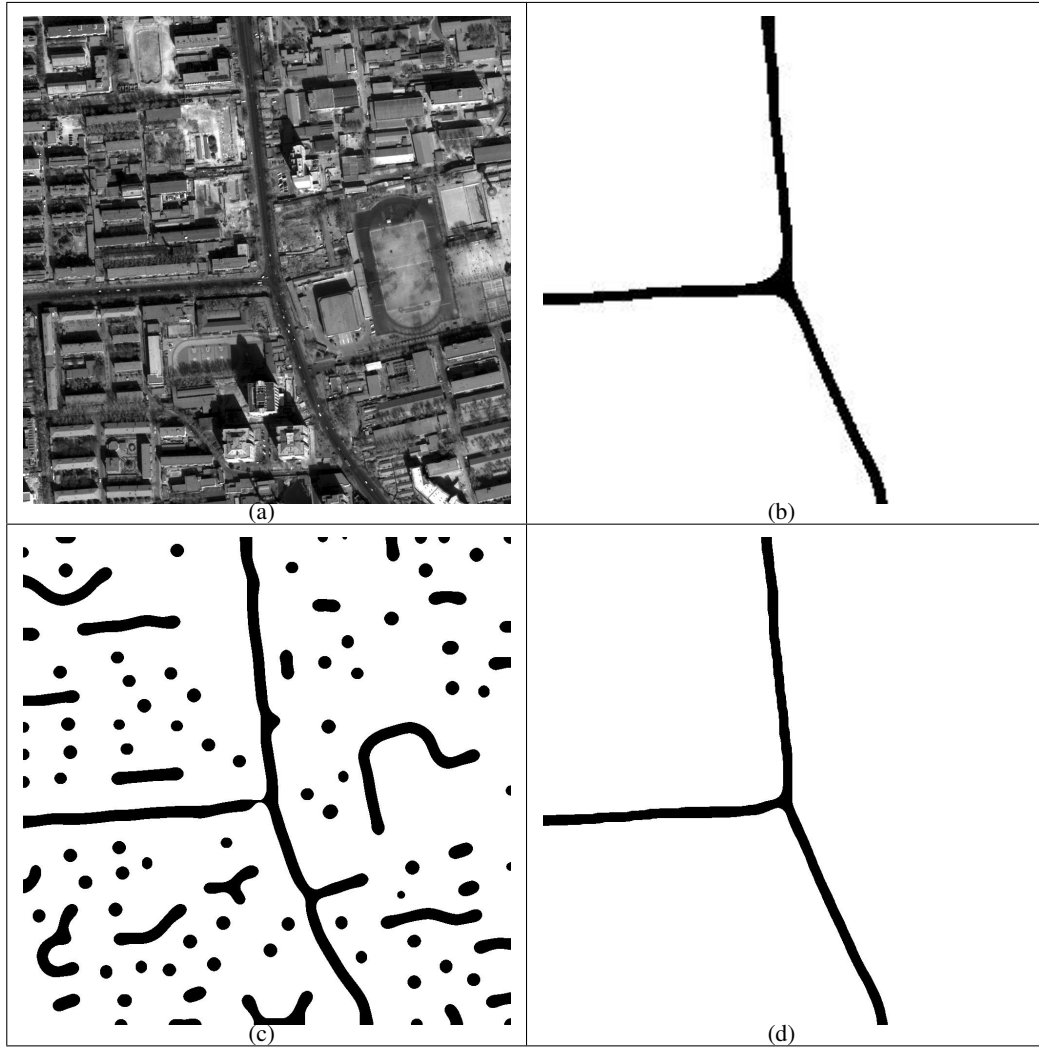


Figure 3.12. More experiments on the QuickBird image. (a): image data (size: 1000×1000); (b): result at 1/4 resolution, using the single-resolution primary model E_{primary} ; (c)-(d): results at full resolution, using the single-resolution primary model E_{primary} and the high-resolution model E_{HR} respectively.

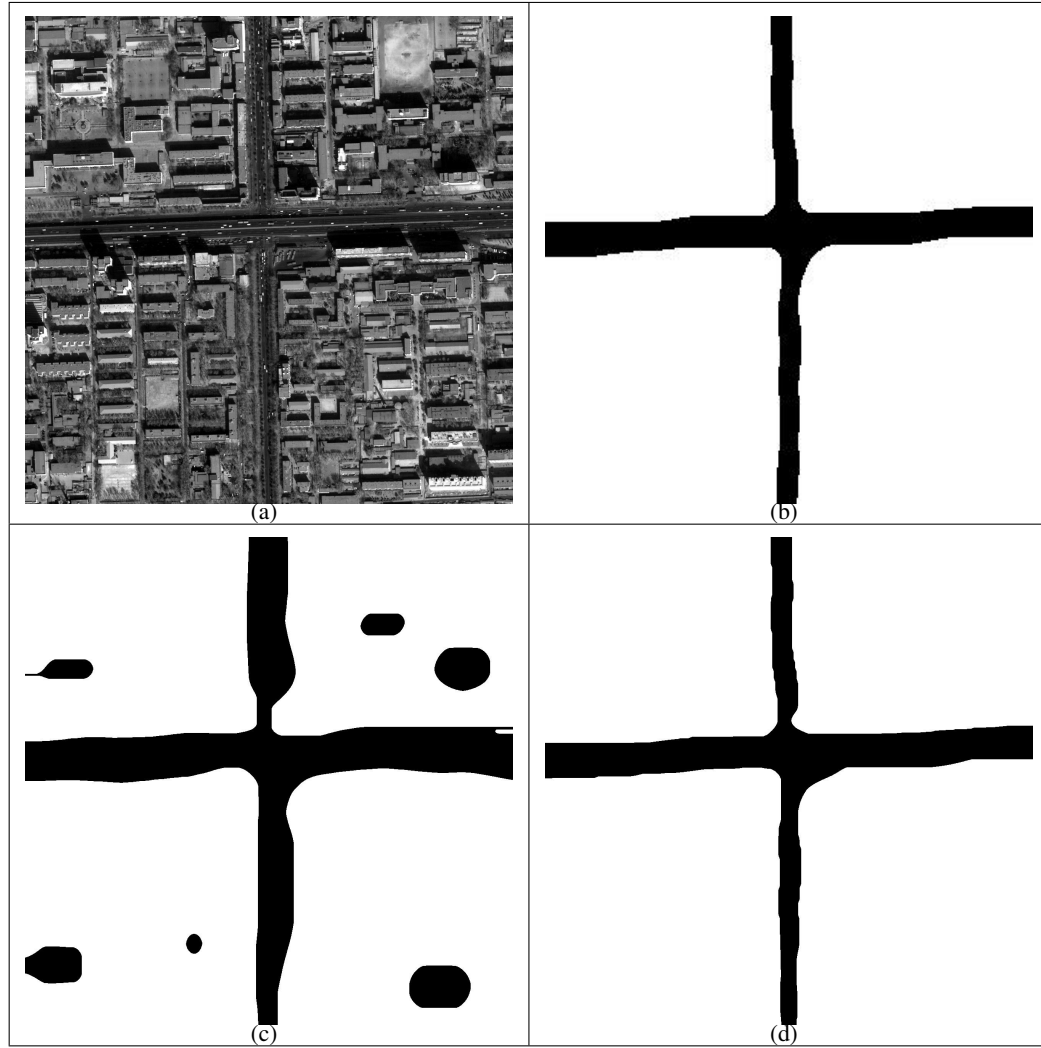


Figure 3.13. More experiments on the QuickBird image. (a): image data (size: 1280×1280); (b): result at $1/4$ resolution, using the single-resolution primary model E_{primary} ; (c)-(d): results at full resolution, using the single-resolution primary model E_{primary} and the high-resolution model E_{HR} respectively.

and $(2000, 0, 3, 0.0905, 0.03, 80)$. In the result at 1/4 resolution obtained with the primary model E_{primary} (Figure 3.11(b)), the small road on the top right of the image is not included, and the borders of the vertical main road are not very accurate. If we still use the single-resolution primary model E_{primary} but on the full resolution image, a large amount of false detections appears (see Figure 3.11(c)). Figure 3.11(d) shows the final segmentation at full resolution using the proposed multiresolution framework (or E_{HR}). Note that at full resolution, the small road in the top right corner has been retrieved, and moreover, the position and the width of the vertical main road are more precise. For example, at the lowest end of the vertical road, due to white road markings, the result in Figure 3.11(d) is closer to the actual image than that in Figure 3.11(b). More experiments are illustrated in Figures 3.12 and 3.13. For the former, the parameters $(\theta, \theta_v, \lambda, \alpha, \beta, d)$ at 1/4 and full resolution are respectively $(200, 0, 3, 0.0905, 0.02, 4)$ and $(2000, 0, 3, 0.0905, 0.03, 16)$; and for the latter, the parameters $(\theta, \theta_v, \lambda, \alpha, \beta, d)$ at 1/4 and full resolution are respectively $(200, 0, 3, 0.0905, 0.03, 20)$ and $(2000, 0, 3, 0.0905, 0.03, 80)$. Notice the better performance of the high-resolution model at the road junction.

3.4 Conclusion

In this chapter, we have presented two models for the extraction of the main road network from a multiscale representation of the image. After considering the merits and demerits of the primary model, we have proposed a multiresolution data energy integrating the data information from multiple resolutions. Although the result at full resolutions are better than those obtained with the single-resolution primary model, the multiresolution approach needed further improvements in order to eliminate false detections and improve the accuracy of road border delineation. Consequently, we have introduced a two-step robust multiresolution framework, with aims of reducing the solution space of the energy functional. The algorithm is robust, and experimental results are accurate. They indicate that, when working at full resolution, the combination of segmentation results at different resolution levels is imperative, due to the great complexity of VHR images. Our model gives better results than several other methods in the literature. In the next chapter, we will take Geographical Information System (GIS) information into account for the particular task of map updating.

Chapter 4

GIS Specific Prior for Map Updating

Keeping the road network information contained in Geographical Information Systems (GIS) up to date is crucial for many applications, for example urban planning, vehicle navigation, and environmental monitoring. The high rate of urban growth, especially in many developing countries, means that this has become an increasingly important research topic in remote sensing. In this chapter, we focus on the issue of main road network *updating* from very high resolution (VHR) images in dense urban areas. Specifically, we show how to make use of an outdated GIS digital map and a recently acquired QuickBird image to generate an up-to-date road network of the observed region. The proposed GIS-based higher-order active contour (HOAC) model includes three different types of prior geometric knowledge characterized by their level of generality. From the most *generic* to the most *specific*, they are (i) generic boundary smoothness constraints, equivalent to a standard active contour prior; (ii) knowledge of the geometric properties of road networks (*i.e.* that they occupy regions composed of long, low-curvature segments joined at junctions), equivalent to a HOAC prior; and (iii) knowledge of the road network at an earlier date derived from GIS data. All three types of prior knowledge prove important for overcoming the complexity of ‘geometric noise’ in VHR images. In section 4.1, we give a brief introduction to map updating, and show that object extraction can benefit from the specific prior knowledge encoded by outdated GIS maps. In section 4.2, we give the definition of a new *specific* prior term, and incorporate it into the phase field framework. In section 4.3, we present experimental results on QuickBird panchromatic images, and validate and compare them with several other techniques from the literature.

4.1 Introduction

Generally, *map updating* consists in three main tasks. The first one is to correct the positions of partly changed objects, and to improve the spatial accuracy of other unchanged objects if necessary. The second one is to extract newly appeared objects, and to incorporate them into the new map to be generated from more recent imagery. The third one is to remove

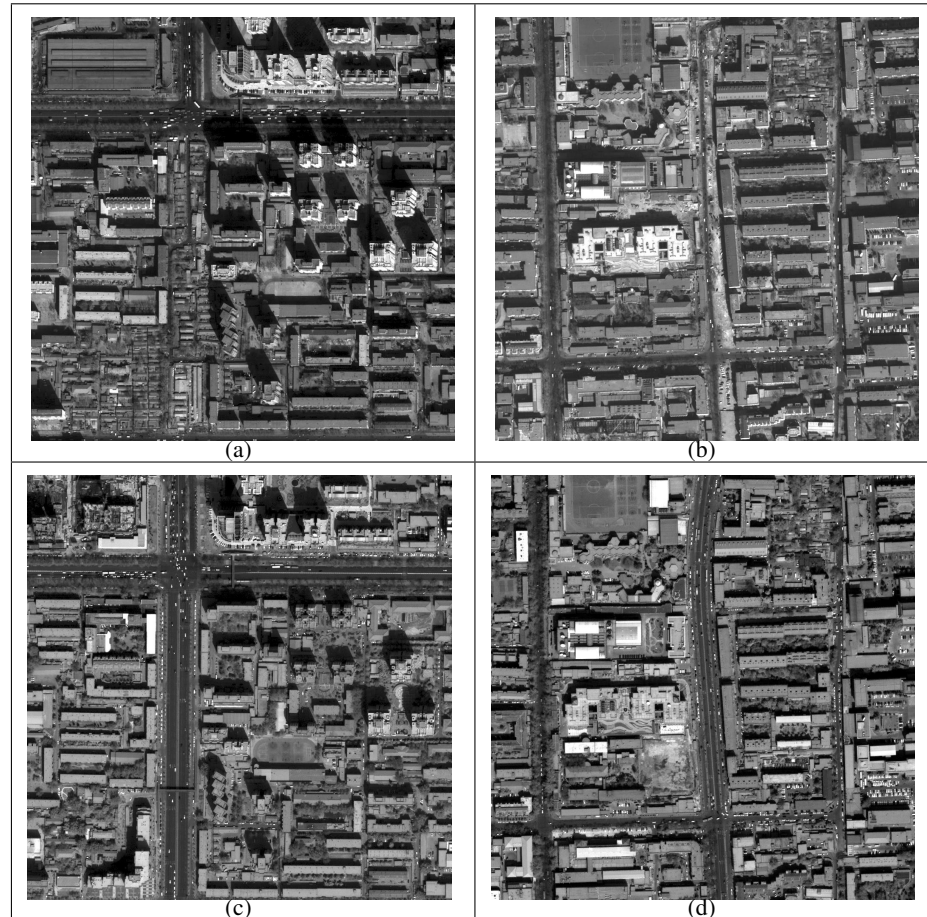


Figure 4.1. Two pairs of QuickBird panchromatic images $\sim 0.61\text{m/pixel}$ (both size: 1000×1000) of Beijing. Top: year 2002; bottom: year 2006.

from the map any object which no longer exists in the image. Compared to the first, the two last tasks are more difficult, because, for a component in one piece of image, we cannot find its counterpart in the other piece of image. Figure 4.1 shows two pairs of QuickBird panchromatic images of Beijing, retrieved respectively in the year 2002 and the year 2006. Due to the great changes in the past few years, the two image pairs give good examples in which map updating is necessary.

For an object extraction process, the GIS map at an earlier date is of great help, because it can serve as an exogenous source of information to the data, and provide valuable (though not completely correct) prior information. The incorporation of an existing map into the algorithm can yield useful semantic and spatial information, and thus increase the robustness of the extraction. More concretely, in our problem, the abundance of ‘artifacts’ of VHR images in dense urban environments, such as shadows, occlusions, cars, etc., may be effectively overcome by adding the older digital map. Since the prior knowledge fur-

nished by existing digital maps concerns the particular scene under consideration, we refer to it as *specific* prior knowledge. This type of knowledge says that the region sought must be ‘close’ to an exemplar region ¹.

In contrast, the most *generic* prior knowledge concerns the regularity properties of the boundary ∂R of the region of interest R . These properties apply to almost any entity, not only road networks. As a consequence, this prior knowledge is included in almost all region models, *e.g.* the Ising model (Ising, 1925), and most active contour models (Kass et al., 1988). It suffices to include a term penalizing the length of ∂R . Between the most *generic* and the most *specific* is prior knowledge derived from a common understanding of the object of interest. In our case, one example of such knowledge of any road network is the statement ‘roads have parallel borders’.

Note that there is an important assumption behind all GIS-guided techniques, *i.e.* the new scene in the image data and the old map should have some regions more or less similar to each other. Otherwise, if the scene were completely different from the GIS map, the map would not be useful, and would probably hinder obtaining the correct result. Of course, this is true of all prior information: if it is incorrect, then it will not help and will probably hinder.

In the rest of this chapter, based on the standard HOAC prior model $E_0 + E_S$, we introduce an additional specific prior term E_{GIS} , derived from an outdated GIS digital map, to solve the above three aspects of road map updating.

4.2 Specific Prior Energy

Suppose that we are given a region R_0 representing the road network at an earlier date than the image data. R_0 can also be described by its minimum energy phase field function ϕ_{R_0} . We propose a specific prior energy, E_{GIS} , that incorporates knowledge of the earlier road network, R_0 , and has effect in the whole of the image domain. We define

$$E_{GIS}(\phi, \phi_{R_0}) = \int_{\Omega} [\omega_+ \phi_{R_0+}(x) + \omega_- \phi_{R_0-}(x)] [\phi(x) - \phi_{R_0}(x)]^2 dx, \quad (4.1)$$

where $\phi_{R_0\pm} = (1 \pm \phi_{R_0})/2$ denote the characteristic functions for the regions R_0 and \bar{R}_0 . The two terms of E_{GIS} correspond to the two components of the symmetric area difference between the segmenting region R and the prior shape R_0 : $x \in R \cap \bar{R}_0$ and $x \in \bar{R} \cap R_0$. These are separated so that they can be weighted differently by the parameters ω_+ and ω_- . The values of ω_{\pm} will be fixed based on factors such as the time lapse between the GIS map and the image. For example, if the time lapse is large, then ω_{\pm} should be small, indicating a weak link between the GIS map and the current road network. If the time lapse is unknown, a mixture model over ω_{\pm} is conceivable. Here we do not discuss these possibilities further, and choose ω_{\pm} manually. E_{GIS} expresses the fact that the evolving phase field ϕ should be ‘close’ to ϕ_{R_0} , which is equivalent to R being ‘close’ to R_0 . At each iterative step, ϕ is modified so as to resemble the given reference shape ϕ_{R_0} in the image domain. Since

¹In other applications, the reference shape may be set as the mean of a set of training shapes with a small variation, as seen in subsection 1.1.3.

this term takes into account the exterior of R_0 (*i.e.* \bar{R}_0), it counteracts the background noise appearing in the data.

It is worth noting that the above formulation is similar to the shape prior term proposed by Chan and Zhu (2003) (equation (1.26)). However, our energy E_{GIS} presents three main differences. First, the outdated GIS data has been globally registered to the image, so pose estimation between the evolving shape and the specific prior shape can be neglected. Second, because of the lack of any hard constraints on ϕ in the phase field modeling, we can avoid the Heaviside function, and thus avoid the numerical error which may be introduced due to approximate estimation of the Heaviside function. Third, due to the updating application, the specific prior ϕ_{R_0} may include some errors.

After augmenting our primary model E_{primary} with the additional specific prior term E_{GIS} , we have the overall GIS-based HOAC model $\theta(E_0 + E_S + E_{GIS}) + E_D$, where E_0 , E_S and E_D have been defined in equations (2.6), (2.8), and (2.21) respectively. In this case of updating, the parameters of the Gaussian mixture and Gamma distributions in the data term are learned from the image data, using the known region R_0 to create samples of road and non-road. Note that the samples may contain errors, since R_0 does not correspond exactly to the road network in the image (*e.g.* we use Figure 4.3(b) as R_0 for Figure 1(a)). The evolution equation for the specific prior is

$$\frac{\partial \phi(x)}{\partial t} = -\frac{\delta E_{GIS}(\phi)}{\delta \phi(x)} = -2(\phi - \phi_{R_0})[\omega_+ \phi_{R_0+} + \omega_- \phi_{R_0-}] . \quad (4.2)$$

This term is added into the rest of the evolution equation for E_{primary} (equation (2.24)).

4.3 Experimental Results and Comparisons

In this section, we test and evaluate the GIS-based HOAC model on QuickBird panchromatic images, and present comparisons with other approaches at full resolution.

4.3.1 Results Using the GIS-Based HOAC Model

Figure 4.2 illustrates two first experiments, using the images in Figure 4.1 acquired in 2006 as the input data. The top row of Figure 4.2 shows GIS maps of the main road networks for the same zones from before the year 2002. Significant changes exist between the maps and the satellite images. The results obtained with the GIS-based HOAC model using the GIS maps as R_0 are shown in the bottom row of Figure 4.2. For these two experiments, the parameter values $(\theta, \theta_v, \lambda, \alpha, \beta, d, \omega_+, \omega_-)$ are $(300, 0.02, 6, 0.1, 0.013, 50, 4.7 \times 10^{-4}, 9 \times 10^{-4})$ and $(300, 0, 5, 0.1, 0.016, 26, 0.0027, 0.0013)$; the computational time is 172 minutes and 137 minutes. Both old maps are successfully updated.

Since a scene that involves all three kinds of changes (*i.e.* some roads missing, some roads added, and some roads narrowed or broadened) seldom happens in the real data, a further test is done on Figure 1(a), using the GIS-based HOAC model with a ‘damaged’ GIS map. Then we can compare to results on the same image obtained with the models proposed in chapter 3. The reason for not using the real older GIS map for this example is

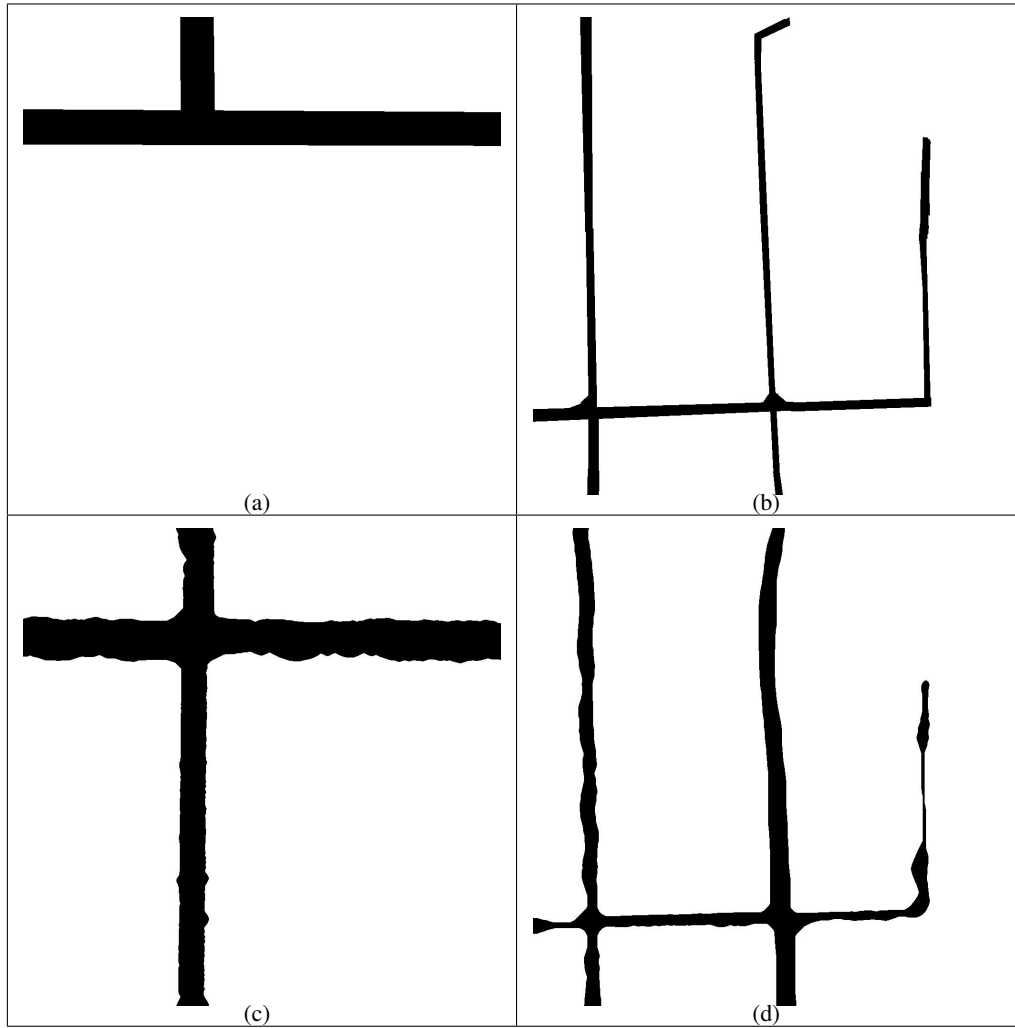


Figure 4.2. Experiments on the later pair of images in Figure 4.1 at full resolution using the GIS-based HOAC model. Top: GIS maps of the main road networks from before the year 2002, used as R_0 ; bottom: the main road networks updated using the QuickBird images from 2006.

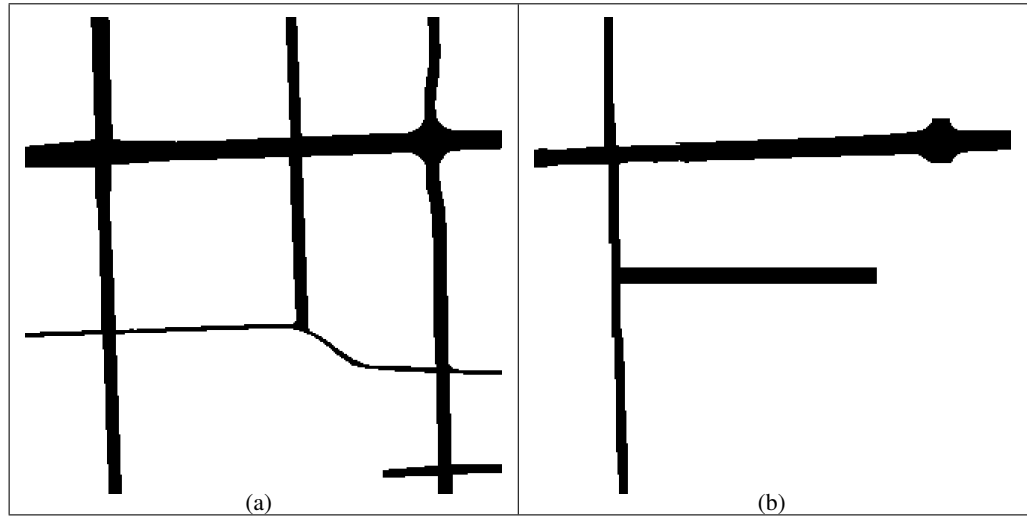


Figure 4.3. Ground truth and simulated GIS map. (a): ground truth of Figure 1(a); (b): deliberately ‘damaged’ ground truth of the same figure, to simulate an earlier GIS map.

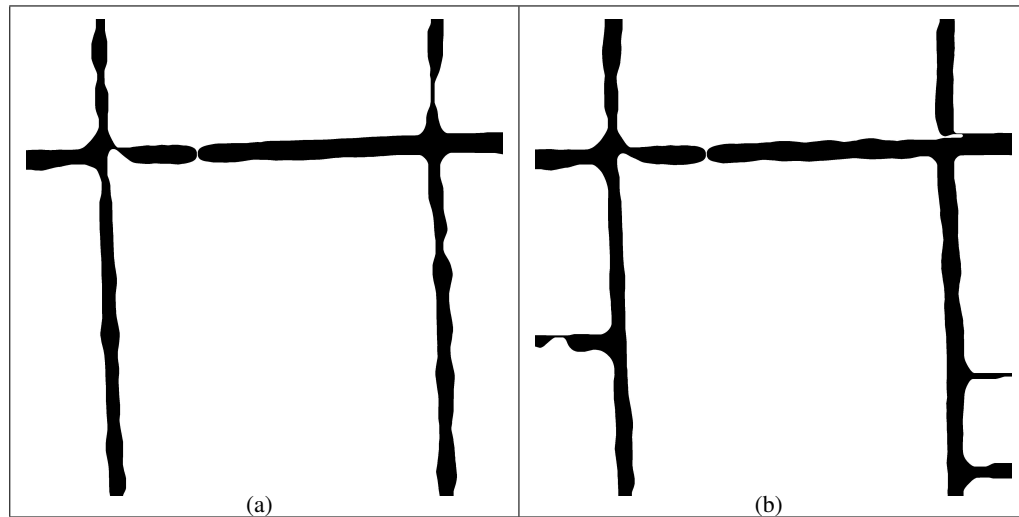


Figure 4.4. Experiments on Figure 1(a) at full resolution (size: 2560×2560 , road width ≈ 96 pixels) using the GIS-based HOAC model. (a): result obtained with ‘damaged’ ground truth (Figure 4.3(b)); (b): result obtained using the result obtained at level 3 (Figure 3.3) as a replacement for the GIS information.

that we want to introduce all kinds of changes into R_0 . Therefore, the associated GIS map, obtained a few year earlier than the satellite image in the zone, is used in two ways: first, to create ground truth via a small manual correction (see Figure 4.3(a)); and to create an inaccurate road network region to serve as R_0 (see Figure 4.3(b)). Note that this ‘damaged’

Measure Method	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
GIS-based model (<i>e.g.</i> Figure 4.4(a))	0.7920	0.8914	0.7198
E_{primary} at level 3, but up-sample to full resolution (<i>e.g.</i> Figure 3.3)	0.7177	0.7571	0.5834
E_{primary} (<i>e.g.</i> Figure 3.4(c))	0.8159	0.6758	0.5893
$E_{\text{primary,mul}}$ (<i>e.g.</i> Figure 3.5)	0.7859	0.7301	0.6141
Multiresolution framework (<i>e.g.</i> Figure 3.10)	0.7098	0.8867	0.6566

Table 4.1. Average quality measures for our different models (T = true, F = false, P = positive, N = negative). The completeness is the percentage of ground truth road network that is extracted; the correctness is the percentage of extracted road network that is correct; and the quality is the most important measure of the “goodness” of the result, because it takes into account the completeness and the correctness.

GIS map is very different from the ground truth. Secondary roads have been kept in the ground truth; this is to allow comparison with other methods, which attempt to find all roads, not just the main road network.

The result is illustrated in Figure 4.4(a). The parameters are set as $(\theta, \theta_v, \lambda, \alpha, \beta, d, \omega_+, \omega_-) = (300, 0.02, 3, 0.0905, 0.02, 80, 3.3 \times 10^{-4}, 6 \times 10^{-4})$. The addition of E_{GIS} provides another great improvement, when compared to the results in Figures 3.4(c) and 3.5. Its main effect is to eliminate false positives in the background, while preserving the correct segmentation of the roads themselves. Figure 4.4(b) shows the result we obtain when we use as R_0 , not the GIS map, but the result obtained at reduced resolution, *i.e.* at level 3 (see Figure 3.3). Thus, in the case of extracting the road network at full resolution with GIS data unavailable, we can replace GIS information in the specific prior term by a low resolution result. We can free ourselves from the need to have a GIS map available. Moreover, it includes multiresolution information in another way. In the multiresolution framework (*i.e.* E_{HR}) proposed in Chapter 3, a low resolution result is used in the weight of the standard HOAC term E_S so as to define an effective zone where the geometric prior knowledge is included; while here, a low resolution result acts as an independent energy term in the model to indicate privileged road regions. These experiments show that for the main roads, *at full resolution*, the GIS-based HOAC model is able to keep unchanged roads, to correct mistakes, and to extract new roads.

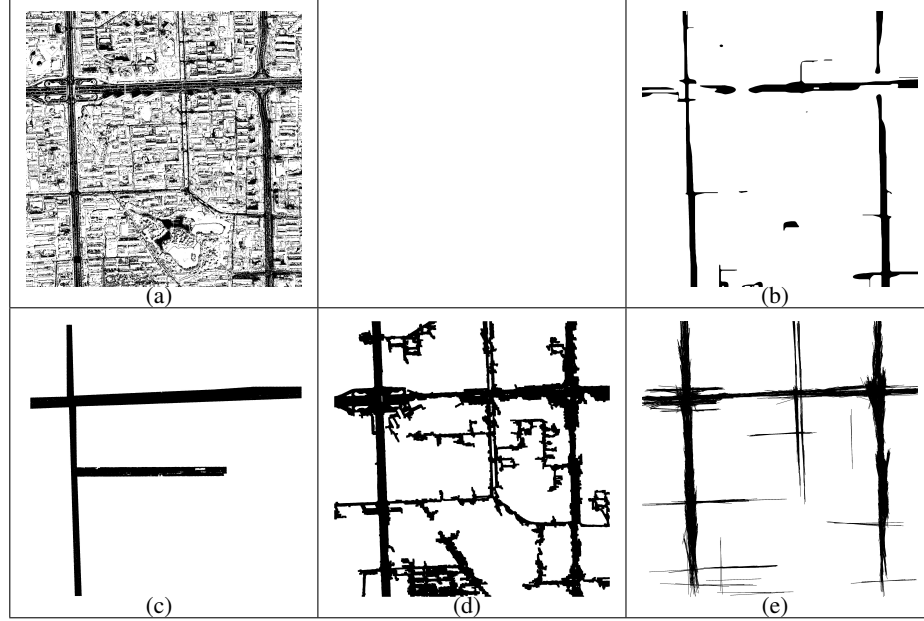


Figure 4.5. Road extraction results from Figure 1(a) at full resolution. (a)-(e): results obtained using MLE (*i.e.* $\theta = 0$), a standard active contour (*i.e.* $\beta = 0$), the work of Bailloeul (2005), Wang and Zhang (2003), and Yu et al. (2004).

4.3.2 Evaluation and Comparison

To evaluate the performance of the various models we have proposed in the previous chapter and in this chapter, we compare our results with five other methods. In the comparisons in chapter 3, we used four of them: maximum likelihood estimation (MLE, *i.e.* $\theta = 0$), a standard active contour (*i.e.* $\beta = 0$), the work of Wang and Zhang (2003), and the work of Yu et al. (2004). The last one is a level set approach incorporating a global shape constraint by Bailloeul (2005); an outdated map is needed to initialize the contour. (For more details, see Appendix E.) Except for those involving the multiresolution model, all results are obtained from full resolution images.

Figure 4.3(a) shows the ground truth used to calculate quantitative criteria. Figures 4.5(a)-4.5(e) illustrate the results obtained using the five methods taken from the literature mentioned above. We have already analyzed the disadvantages of the four methods used in the previous chapter, which we do not repeat here. Besides, the ‘flexible active contour’ method of Bailloeul (2005) (initially dedicated to building extraction) fails because it is not able to eliminate road sections that exist in the map but not in the image, and to extract new road sections that exist in the image but not in the map. Figures 4.6 and 4.7 show the comparisons based on the original images shown in Figures 4.1(c) and 4.1(d). Some quantitative evaluation measures (Heipke et al., 1997) of our proposed models and other different methods are shown respectively in Tables 4.1 and 4.2. For each method and each measure, the average value from three experiments (using the images in Figures 1(a), 4.1(c) and 4.1(d))

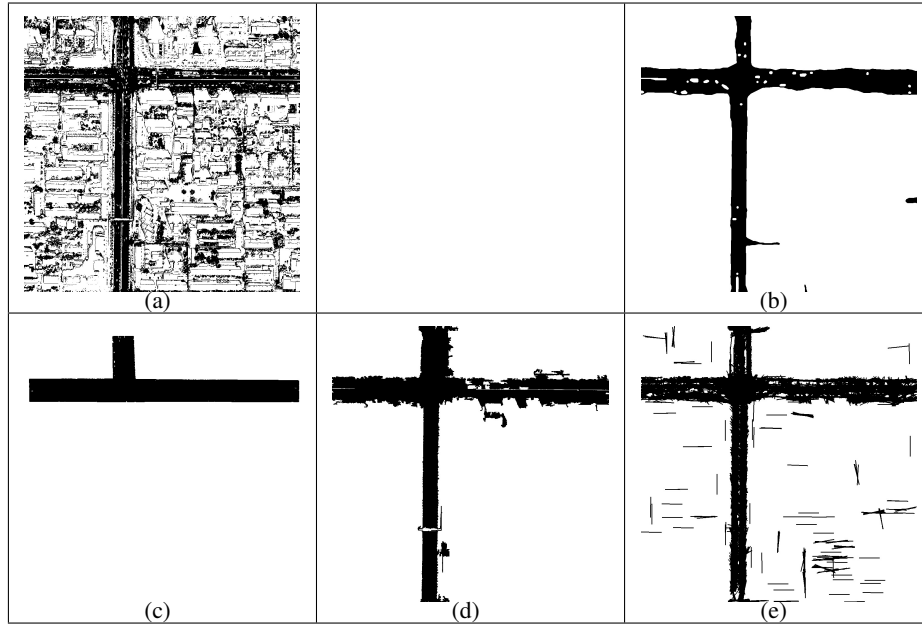


Figure 4.6. Road extraction results from Figure 4.1(c) at full resolution. (a)-(e): results obtained using MLE (*i.e.* $\theta = 0$), a standard active contour (*i.e.* $\beta = 0$), the work of Bailloeul (2005), Wang and Zhang (2003), and Yu et al. (2004).

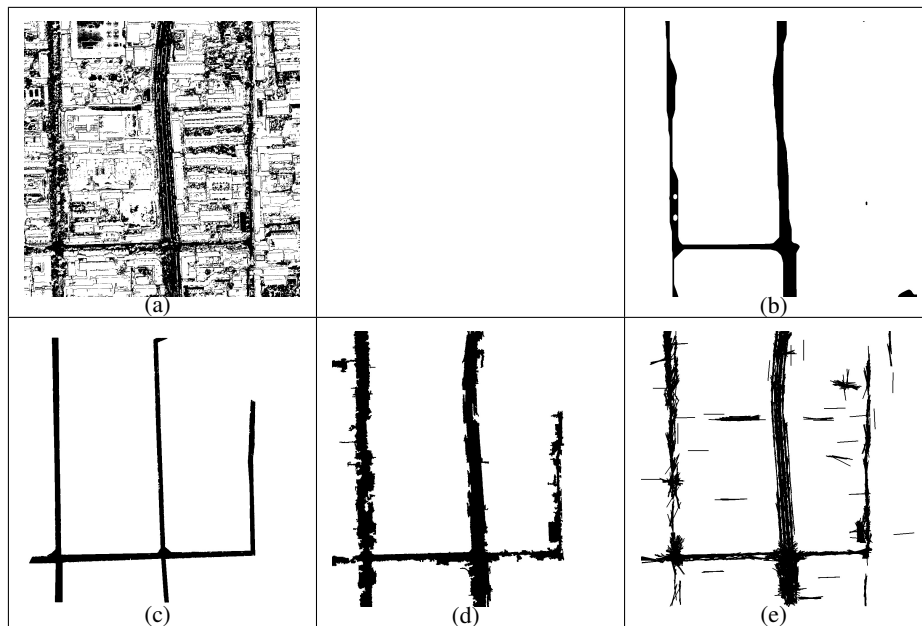


Figure 4.7. Road extraction results from Figure 4.1(d) at full resolution. (a)-(e): the results obtained using MLE (*i.e.* $\theta = 0$), a standard active contour (*i.e.* $\beta = 0$), the work of Bailloeul (2005), Wang and Zhang (2003), and Yu et al. (2004).

Measure Method	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
GIS-based model (<i>e.g.</i> Figure 4.4(a))	0.7920	0.8914	0.7198
MLE (<i>e.g.</i> Figure 4.5(a))	0.7567	0.2775	0.2545
$\theta E_0 + E_D$ (<i>e.g.</i> Figure 4.5(b))	0.6358	0.8743	0.5810
Bailloeul (<i>e.g.</i> Figure 4.5(c))	0.5529	0.8318	0.4990
Wang (<i>e.g.</i> Figure 4.5(d))	0.8918	0.6180	0.5717
Yu (<i>e.g.</i> Figure 4.5(e))	0.7743	0.7196	0.5893

Table 4.2. Average quality measures for the different methods (T = true, F = false, P = positive, N = negative). See Table 4.1 for an explanation of completeness, correctness and quality.

is shown.

4.4 Conclusion

We have proposed a GIS-based HOAC model for the updating of road maps in dense urban areas by extracting the main road network from VHR satellite images. The GIS-based HOAC approach incorporates three different types of prior geometric knowledge: generic knowledge about smoothness; knowledge of the geometry of road networks in general; and knowledge of the specific road network at a different date, supplied as GIS data. Our results indicate that, in order to work at full resolution, the combination of all three types of prior knowledge is essential, to overcome the geometric noise of VHR images. However, in the case of extracting the road network at full resolution with GIS data unavailable, we can use instead a result obtained at lower resolution, where such knowledge appears not to be necessary provided the other two types are present, to replace the GIS information in the specific prior term. Hence, one can free oneself from the need for GIS data. This also leads to a second type of multiresolution analysis. Our model gives better results than several other methods published in the literature, even when smaller roads, which our model is not designed to detect, are included in the ground truth. In the next chapter, we will consider secondary road extraction. Due to the particular difficulties of secondary roads, we will develop new types of HOAC prior.

Chapter 5

Modeling Shape for Network Extraction

In previous chapters, we solved the problem of the extraction and the updating of main road networks. Compared to the main roads, the secondary roads are much more difficult to deal with, because, on the one hand, of the low discriminative power of the grey-level distributions of road regions and the background, and on the other hand, of the greater effect of occlusions and other noise on narrower roads. Moreover, the previously developed higher-order active contour (HOAC) prior energy suffers from a serious limitation, when it is used to model networks: the interactions between points on the same side of a network branch have the same range and strength as the interactions between points on opposite sides. To tackle the above mentioned problems, in this chapter, we show how to separate the control of branch straightness and branch width, thus allowing better prolongation of the network for a given road width. This chapter is organized as follows. In section 5.1, we discuss the difficulties of secondary road extraction. In sections 5.2 and 5.3, we propose a new *nonlinear nonlocal* HOAC prior energy and a new *linear nonlocal* HOAC prior energy, and reformulate them as nonlocal phase field energies. We perform the stability analysis of these two prior models, and thereby establish parametrical constraints. In section 5.4, we illustrate the application of both models to road extraction from very high resolution (VHR) images, and compare the results to other methods in the literature. The results demonstrate the superiority of our models.

5.1 Introduction

The difficulty with secondary roads lies in the following aspects. As shown in Figure 5.1, first, the radiometric properties of narrow roads are similar to those of the background; second, narrow roads are more often obscured by shadows and trees, which could cause gaps in the extracted network. For both reasons, data driven/bottom-up models fail to retrieve the roads correctly: strong geometric prior information is needed.



Figure 5.1. An example of small roads, to demonstrate the difficulty of extracting them compared to main roads.

Although the network model of the standard HOAC prior energy $E_{C,S}$ (equation (2.4))

$$E_{C,S}(R) = \lambda_C L(\partial R) + \alpha_C A(R) - \frac{\beta_C}{2} \iint_{(\partial R)^2} \dot{\gamma}(t) \cdot \dot{\gamma}(t') \Psi\left(\frac{|\gamma(t) - \gamma(t')|}{d}\right) dt dt' , \quad (5.1)$$

where

$$\Psi(x) = \begin{cases} \frac{1}{2} \left(2 - |x| + \frac{1}{\pi} \sin(\pi|x|) \right) & \text{if } |x| < 2 , \\ 0 & \text{else ,} \end{cases} \quad (5.2)$$

or equivalently, the network model of the standard phase field HOAC prior energy $E_0 + E_S$ in equations (2.6) and (2.8), contains such prior knowledge, it suffers from a significant limitation when it comes to modeling networks. Apart from a sign change, the interaction between two points on the same side of a network branch ($\dot{\gamma}(t) \cdot \dot{\gamma}(t') > 0$) is the same, and in particular has the same strength and range as the interaction between two points on opposite sides of a network branch ($\dot{\gamma}(t) \cdot \dot{\gamma}(t') < 0$). The former interaction controls the curvature of network branches by trying to align tangent vectors, while the latter controls branch width by creating a repulsive force. The effect is that for a stable network branch, a typical maximum curvature of a branch κ is connected to the width of that branch W approximately by $\kappa \sim 1/W$. In other words, the length/range along which the network branch is expected to be straight is the same as the width of the branch itself. The standard HOAC prior energy $E_{C,S}$ (or its phase field counterpart $E_0 + E_S$) thereby provides a poor model of networks with straight narrow branches or highly curved, wide branches.

However, in our case of road extraction in cities, road width gives only an (approximate) upper bound on the radius of curvature of the road: most roads are straighter than they are wide, *i.e.* $\kappa \ll 1/W$. For narrow roads, this is particularly problematic, since the road region is relatively unconstrained due to the small road width. In particular, using $E_{C,S}$ (or equivalently $E_0 + E_S$), the road prolongation will be of short range, while to overcome the effects of geometric noise mentioned above, we want it to be long-range.

To solve these problems, we need to be able to model longer-range or stronger ¹ in-

¹To augment the magnitude of the interaction function or to increase its range can approximately be considered as equivalent.

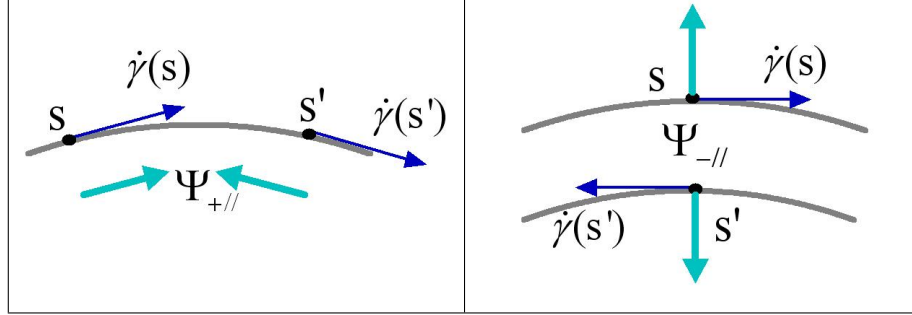


Figure 5.2. The interaction functions $\Psi_{+//}$ and $\Psi_{-//}$ (thin blue arrow: tangent vector; thick cyan arrow: interaction force; black dot: interacting point).

interactions along the road, without changing the interactions across the road. This means that we have to separate the two interaction functions, and hence allow separate control of branch straightness and width, as shown in Figure 5.2. To achieve this goal, we will construct new *nonlocal* HOAC prior energies, which act in a complementary way to the standard HOAC term. The first new energy term is a *nonlinear, nonlocal* HOAC energy E_{NL} , which increases the magnitude of the interaction along one side of a network branch. The second one is a *linear, nonlocal* HOAC energy E_L , which provides a longer-range interaction along one side of a network branch; through the stability analysis of this new model, we further demonstrate that the *linear nonlocal* term permits the modeling of two widths simultaneously.

It is often convenient to formulate an active contour model in terms of the parametric curve, and then reformulate it as a phase field model for implementation. In the rest of this chapter, we follow this procedure by considering each new prior energy using first a parametric contour representation, and then a phase field representation. We also calculate the conditions for which each model has stable behaviors.

5.2 Nonlinear Nonlocal HOAC Prior Energy

We introduce a *nonlinear nonlocal* HOAC prior energy E_{NL} that differentiates between interactions along the road and across the road, and thus enables more sophisticated prior knowledge to be included.

5.2.1 Contour Model Definition

Since roads are elongated structures, the interaction between points on the same side of a road must have longer range (or be stronger) than the interaction between points on opposite sides of a road. In order to separate the two interactions, the interaction function must depend on the tangent/normal vectors at the pairs of points that are its argument. Although the length scale in the interaction function of equation (5.1), d , could be made to depend on the inner product between the tangent/normal vectors at the two pixels, it would lead to

complicated functional derivatives. Alternatively, we prefer to perform a linear interpolation between two interaction functions. In the contour formulation, our new HOAC prior energy E_{HO} takes the form:

$$E_{HO}(\gamma) = - \iint_{S^1 \times S^1} \left\{ f_{+||}(\dot{\gamma}(s) \cdot \dot{\gamma}(s')) \Psi_{+||} - f_{-||}(\dot{\gamma}(s) \cdot \dot{\gamma}(s')) \Psi_{-||} \right\} ds ds' , \quad (5.3)$$

where $\gamma : S^1 \rightarrow \Omega$, is an arc length parameterization of the region boundary ∂R ; $\dot{\gamma}(s)$ is the tangent vector to the boundary at s (thus $\dot{\gamma}(s) \cdot \dot{\gamma}(s') \in [-1, 1]$); ‘+ ||’ denotes parallel vectors and ‘- ||’ denotes antiparallel vectors. We define $f_{+||}(x), f_{-||}(x) : [-1, 1] \rightarrow [0, 1]$ as linear functions:

$$f_{+||}(x) = (1 + x)/2 , \quad (5.4a)$$

$$f_{-||}(x) = (1 - x)/2 . \quad (5.4b)$$

$\Psi_{+||}$ and $\Psi_{-||}$ are interaction functions similar to that in equation (5.2), but have different range or magnitude. They compete with each other: when $\dot{\gamma}(s) \cdot \dot{\gamma}(s') \in [0, 1]$, *i.e.* the two interacting tangent vectors are more parallel, $\Psi_{+||}$ is dominant; while when $\dot{\gamma}(s) \cdot \dot{\gamma}(s') \in [-1, 0]$, *i.e.* the two interacting tangent vectors are more antiparallel, $\Psi_{-||}$ is dominant. We decide to adjust only the magnitude of the interaction (although this effectively changes its range also). Thus, we assume that the magnitude of the interaction of parallel vectors is stronger than that of antiparallel vectors, *i.e.* $\Psi_{+||} = a\Psi_{-||}$, where $a > 1$, is a constant. Then, equation (5.3) becomes

$$E_{HO}(\gamma) = -\frac{1}{2} \iint_{S^1 \times S^1} [(a - 1) + (a + 1)(\dot{\gamma}(s) \cdot \dot{\gamma}(s'))] \Psi_{-||} ds ds' . \quad (5.5)$$

5.2.2 Phase Field Model Definition

In order to implement $E_{HO}(\gamma)$ in the phase field framework, it needs to be reformulated as a function of the phase field ϕ , instead of the arc length parameterization γ used in equation (5.5). Since the constant length of $\dot{\gamma}(s)$ corresponds to the fixed change in ϕ across the interface², we replace tangent vectors by normal vectors, and then normal vectors by $\nabla\phi$. Subsequently, the range of interactions is extended from the region boundary ∂R to the whole of the image domain Ω . Due to the fact that $\nabla\phi(x)$ is approximately equal to zero everywhere outside the narrow interface R_C in Ω , the boundary indicator function

$$S(\phi) = (\nabla\phi(x) \cdot \nabla\phi(x))(\nabla\phi(x') \cdot \nabla\phi(x')) \simeq \begin{cases} 1 & \forall x, x' \in R_C , \\ 0 & \text{otherwise} , \end{cases} \quad (5.6)$$

is inserted into the first term of equation (5.5). Thus we have

$$E_{HO}(\phi) = -\frac{1}{2} \iint_{\Omega^2} [(a - 1)S(\phi) + (a + 1)(\nabla\phi(x) \cdot \nabla\phi(x'))] \Psi\left(\frac{|x - x'|}{d}\right) dx dx' . \quad (5.7)$$

²In an arbitrary parameterization, the terms that do not depend on tangent vectors acquire factors of $|\dot{\gamma}||\dot{\gamma}'|$ arising from $ds ds'$.

When $a = 1$, this reduces to the standard phase field HOAC term E_S (up to a factor of $\beta/2$). Therefore, we define our new additional energy term E_{NL} by

$$\begin{aligned} E_{NL}(\phi) &= -\frac{\beta_2}{4} \iint_{\Omega^2} S(\phi) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x)) (\nabla\phi(x') \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx'. \end{aligned} \quad (5.8)$$

(Note that d is the same as that in E_S .) Thus in this case, we have a term that is quartic in $\nabla\phi$.

The functional derivative of E_{NL} is (see details in Appendix B)

$$\begin{aligned} \frac{\delta E_{NL}(\phi)}{\delta\phi(x)} &= \beta_2 \int_{\Omega} \left\{ (\nabla\phi(x') \cdot \nabla\phi(x')) \nabla^2\phi(x) \Psi\left(\frac{|x-x'|}{d}\right) \right. \\ &\quad \left. + (\nabla\phi(x') \cdot \nabla\phi(x')) \left(\nabla\phi(x) \cdot \nabla\Psi\left(\frac{|x-x'|}{d}\right) \right) \right\} dx'. \end{aligned} \quad (5.9)$$

Since this functional derivative of E_{NL} , unlike that of E_S , contains a term nonlinear in $\nabla\phi$, due to $S(\phi)$ being $O(\phi^4)$, we refer to it as the *nonlinear nonlocal* term³.

Note that whether the two tangent/normal vectors at a pair of interacting points are parallel or antiparallel, the effect of E_{NL} is always to encourage two points inside the range of the interaction to attract each other. Thus E_{NL} weakens the repulsive effect of E_S between opposite sides, so that E_S along a network branch can be strengthened without changing the width. Consequently, the interaction between pairs of points on the same side of a network branch is stronger than that between pairs of points on opposite sides of a network branch.

We now add this new *nonlinear* term E_{NL} into our primary model E_{primary} (section 2.4), and define our new nonlinear overall model as $E = \theta(E_0 + E_S + E_{NL}) + E_D$. Since the derivatives $\delta E_S/\delta\phi$ and $\delta E_{NL}/\delta\phi$ are nonlocal, to avoid performing explicit convolutions, they are, as usual, calculated in the Fourier domain. The intermediate steps of this calculation are detailed in Appendix B. The resulting evolution equation for E_{NL} is

$$\begin{aligned} -\frac{\delta E_{NL}(\phi)}{\delta\phi(x)} &= -\beta_2 \nabla^2\phi(x) \mathcal{F}^{-1} \left\{ d\hat{\Psi}(kd) \mathcal{F} \{ \nabla\phi(x) \cdot \nabla\phi(x) \} \right\} \\ &\quad - \beta_2 \nabla\phi(x) \cdot \nabla \left\{ \mathcal{F}^{-1} \left\{ d\hat{\Psi}(kd) \mathcal{F} \{ \nabla\phi(x) \cdot \nabla\phi(x) \} \right\} \right\}. \end{aligned} \quad (5.10)$$

This term is added into the evolution equation for our primary model (equation (2.24)). We follow all the other implementation details as described in subsection 2.4.2.

5.2.3 Stability Analysis of the Nonlinear Nonlocal HOAC Total Prior Model

The *nonlinear nonlocal* HOAC total prior model takes the form $E_{P,NL} = E_0 + E_S + E_{NL}$. Due to the incorporation of E_{NL} , an additional parameter β_2 has been introduced. So now, there are five parameters $(\alpha, \lambda, \beta, \beta_2, d)$ under consideration. These parameters need to be

³We worked also on another *nonlinear nonlocal* HOAC prior term. We finally do not use it because of the difficulty of its implementation. We describe it in Appendix C.

chosen carefully, because the gradient descent equation of E_{NL} is a highly nonlinear non-local partial differential equation. To study the behavior of the model with respect to these five parameters, we use an *ansatz* model of a bar, as defined in Figure 2.5, and conduct a stability analysis. In this subsection, we first implement this calculation, and then give the stability conditions for the *nonlinear nonlocal* HOAC total prior model.

Nonlinear Nonlocal Phase Field HOAC Term of a Bar

Following a similar calculation for the standard phase field HOAC term, we integrate equation (5.8) over the surface defined by the bar *ansatz*. The nonlinear nonlocal HOAC term of a bar is

$$E_{NL} = -\frac{16\beta_2 L}{w^2} \left\{ \int_0^\infty \Psi\left(\frac{z}{d}\right) dz + d \int_{\hat{W}}^\infty \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \Psi(\eta) d\eta \right\}, \quad (5.11)$$

where $\hat{W} = W/d$ is the scaled width. Based on the definition of the interaction function Ψ (equation (5.2)), we obtain the nonlinear nonlocal phase field HOAC energy, per unit length of bar, e_{NL} :

$$\begin{aligned} e_{NL}(w, \hat{W}) &= -\frac{16\beta_2 d}{w^2} - \frac{8\beta_2 d}{w^2} \int_{\hat{W}}^2 \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \left(2 - \eta + \frac{1}{\pi} \sin(\pi\eta)\right) d\eta \\ &= -\frac{16\beta_2 d}{w^2} - \frac{8\beta_2 d}{w^2} \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta. \end{aligned} \quad (5.12)$$

Nonlinear Nonlocal HOAC Total Prior Model of a Bar

Combining the nonlinear nonlocal phase field HOAC term per unit length of bar e_{NL} (equation (5.12)), with the basic phase field term per unit length of bar e_0 (equation (2.13)), and the standard phase field HOAC term per unit length of bar e_S (equation (2.15)), we can express the nonlinear nonlocal HOAC total prior model, per unit length of bar, $e_{P,NL}$ as (after removing all the constants not related to the variables):

$$\begin{aligned} e_{P,NL}(w, \hat{W}) &= e_0 + e_S + e_{NL} \\ &= \frac{4}{3} \alpha \hat{W} d + \frac{4}{15} \lambda w + \frac{4}{w} - \frac{16\beta_2 d}{w^2} + 4d\left(\beta - \frac{2\beta_2}{w^2}\right) \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta. \end{aligned} \quad (5.13)$$

To minimize the energy $e_{P,NL}$, we need to calculate its first derivatives and its second derivatives with respect to w and \hat{W} . The first derivatives are set to zero:

$$\frac{\partial e_{P,NL}}{\partial \hat{W}} = \frac{4}{3} \alpha d - 4\hat{W}d\left(\beta - \frac{2\beta_2}{w^2}\right) \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta = 0, \quad (5.14a)$$

$$\frac{\partial e_{P,NL}}{\partial w} = \frac{4}{15} \lambda - \frac{4}{w^2} + \frac{16\beta_2 d}{w^3} \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta = 0. \quad (5.14b)$$

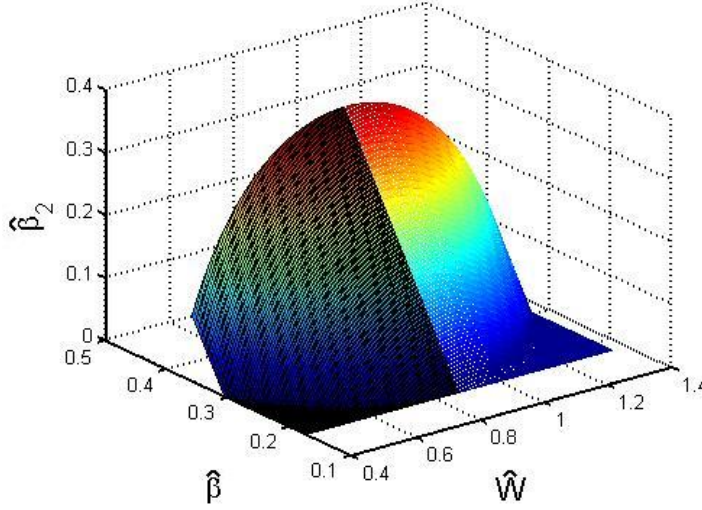


Figure 5.3. Diagram of the scaled control parameters $\hat{\beta}$ and $\hat{\beta}_2$, and the scaled width \hat{W} , when $w = 2$. The left part of the surface represents the parameter space for which $e_{P,NL}$ has one local maximum; in the right part of the surface, $e_{P,NL}$ has one local minimum; and the borderline between them, where $\hat{W} = 0.8798$, represents the situation when two extrema become an inflection point (i.e. $\partial e_{P,NL}/\partial \hat{W} = 0$ and $\partial^2 e_{P,NL}/\partial \hat{W}^2 = 0$).

This is a complicated system of two-variable equations. To simplify the problem, we fix arbitrarily the reasonable value of w beforehand ($w = 2 \sim 4$), and then solve equation (5.14a) to get a sub-optimal solution of \hat{W} . In this way, there is no need to calculate the second derivative of $e_{P,NL}$ with respect to w , but we still need to ensure the derivative of order two of $e_{P,NL}$ with respect to \hat{W} non-negative. It is given by (see details in Appendix A.2.2)

$$\frac{\partial^2 e_{P,NL}}{\partial \hat{W}^2} = -4d(\beta - \frac{2\beta_2}{w^2}) \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right) + 4d(\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta \geq 0. \quad (5.15)$$

The above analysis shows that the stability of the nonlinear nonlocal HOAC total prior model is related to the scaled control parameters $\hat{\beta} = \beta/\alpha$ and $\hat{\beta}_2 = \beta_2/\alpha$, and to the scaled

width $\hat{W} = W/d$. Therefore, we obtain the parameter constraints:

$$\hat{\beta} - \frac{2\hat{\beta}_2}{w^2} = \frac{1}{3\hat{W}I_1(\hat{W})}, \quad (5.16a)$$

$$I_2(\hat{W}) \geq \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right), \quad (5.16b)$$

where

$$I_1(\hat{W}) = \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta, \quad (5.17a)$$

$$I_2(\hat{W}) = \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta. \quad (5.17b)$$

Now we can draw the diagram of the relationship among the scaled control parameters $\hat{\beta}$ and $\hat{\beta}_2$, and the scaled width \hat{W} . We obtain a surface, according to the stability conditions. When w is set to 2, an example of such a surface is illustrated in Figure 5.3. Note that the curve on the surface of $\hat{\beta}_2 = 0$ is identical to that we have presented in Figure 2.6. The left part of the surface represents the extremum as a local maximum, *i.e.* the second derivative is negative; the right part of the surface represents the extremum as a local minimum, *i.e.* the second derivative is positive; and the borderline between these two surfaces, where $\hat{W} = 0.8798$, indicates the situation when two extrema of $e_{P,NL}$ become an inflection point (*i.e.* $\partial e_{P,NL}/\partial \hat{W} = 0$ and $\partial^2 e_{P,NL}/\partial \hat{W}^2 = 0$). Clearly, once \hat{W} is given, $\hat{\beta}$ and $\hat{\beta}_2$ have to be chosen in the right part of this surface. Due to the introduction of the additional parameter $\hat{\beta}_2$, for each given width \hat{W} , there is a curve to show all the possible groups of scaled parameters $\hat{\beta}$ and $\hat{\beta}_2$. As before, one of the parameters, α , has been eliminated.

5.3 Linear Nonlocal HOAC Prior Energy

In the previous section, we proposed a *nonlinear nonlocal* HOAC prior term E_{NL} to overcome the limitation of the standard HOAC prior term. Since the effect of E_{NL} is to reinforce the interaction along the bar branch and to weaken the interaction across the bar branch, the interactions in these two directions are still somehow dependent, and hence the two effects cannot be tuned in a completely independent way. Furthermore, since the computational cost of this nonlinear term is high, as we will show, we cannot perform road extraction on images at full resolution with this nonlinear model. To overcome these drawbacks, in this section, we construct a new, Euclidean invariant *linear nonlocal* HOAC prior term E_L as a replacement for E_{NL} . We analyze the stability of the new model, and find that in addition to solving the above limitation by separating the interactions between points on the same and opposite sides of a network branch, the new model permits a broader range of widths to be modeled simultaneously, and can even model two disjoint width ranges.

5.3.1 Contour Model Definition

As we have seen in chapter 2, one general class of quadratic HOAC terms can be written as

$$E_{HO}(R) = - \iint_{(\partial R)^2} \dot{\gamma}(t) \cdot \mathbf{G}_C(\gamma(t), \gamma(t')) \cdot \dot{\gamma}(t') dt dt' , \quad (5.18)$$

where \mathbf{G}_C is a map from Ω^2 to 2×2 matrices. Imposing Euclidean invariance on this term leads to two possibilities: $G_C(\gamma(t), \gamma(t')) = \Psi(|\Delta\gamma|/d) \delta$, where δ is the unit matrix, and $\Delta\gamma = \gamma(t) - \gamma(t')$; or $G_C(\gamma(t), \gamma(t')) = \Psi(|\Delta\gamma|/d) \Delta\gamma \Delta\gamma^T$. The first possibility leads to the standard HOAC term in $E_{C,S}$ (equation (5.1)). The second possibility leads to our new *linear nonlocal* HOAC prior energy, E_L :

$$E_L(R) = - \iint_{(\partial R)^2} (\dot{\gamma}(t) \cdot \Delta\gamma(t, t')) (\dot{\gamma}(t') \cdot \Delta\gamma(t, t')) \Psi\left(\frac{|\gamma(t) - \gamma(t')|}{d_2}\right) dt dt' , \quad (5.19)$$

where we use the same Ψ (equation (5.2)) as in $E_{C,S}$, but with a different range d_2 .

E_L compares each tangent vector to the vector $\Delta\gamma(t, t')$ joining the two interacting points. When two points have tangent vectors that are both nearly aligned or anti-aligned with $\Delta\gamma$, the product of the dot products is positive. The energy E_L can decrease further by further aligning these tangent vectors with $\Delta\gamma$ and hence with each other. This situation corresponds to two points on the same side of a network branch, as shown in Figure 5.4(a). The energy thus favors straight lines, within a range controlled by d_2 . On the other hand, when at least one of the two tangent vectors is nearly orthogonal to $\Delta\gamma$, the product of the dot products is small. In this configuration, changing the distance between the two points in the argument to Ψ does not change the energy much, and thus the force between two such points is small. This situation corresponds to two points on opposite sides of a network branch, as shown in Figure 5.4(b).

As a result, when E_L is added to $E_{C,S}$, the width of the network branches is controlled largely by the parameter d of $E_{C,S}$, while the distance over which the branch will be straight is controlled largely by d_2 , if $d_2 > d$. For thin, straight bars, we will indeed fix $d_2 > d$. The exception to this rule is again shown in Figure 5.4(b). From the above, $\gamma(t')$ exerts no force on $\gamma(t)$, but for both $\gamma_L(t')$ and $\gamma_R(t')$, the product of the dot products is negative. The energy E_L can decrease when the value of Ψ becomes less positive, *i.e.* both $\gamma_L(t')$ and $\gamma_R(t')$ repel $\gamma(t)$, as shown by the force arrows F_L and F_R in the figure. The tangential parts of F_L and F_R cancel, and there is an overall normal repulsion F . If the weight of E_L in the model is too large, this repulsion may begin to dominate the bar width.

We can notice an interesting fact by changing the notation of the *linear nonlocal* HOAC term E_L in equation (5.19). We make the following change of variables:

$$\dot{\gamma}(t) = (\cos \theta, \sin \theta), \quad \dot{\gamma}(t') = (\cos \theta', \sin \theta'), \quad \Delta\gamma(t, t') = (\cos \eta, \sin \eta).$$

Then, by developing the product of equation (5.19), E_L can be rewritten as (up to a multiplicative)

$$E_L(R) = - \iint_{(\partial R)^2} (\cos(\theta - \theta') + \cos(\theta + \theta' - 2\eta)) \Psi\left(\frac{|\gamma(t) - \gamma(t')|}{d_2}\right) dt dt' . \quad (5.20)$$

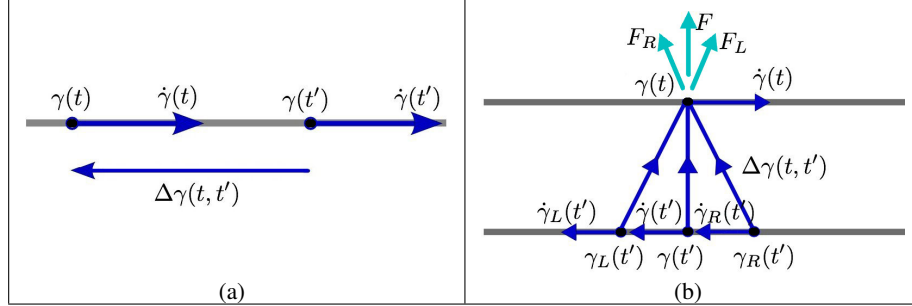


Figure 5.4. The effects of E_L in equation (5.19) (blue arrow: vector; cyan arrow: interaction force; black dot: interacting point). (a): when two tangent vectors are nearly aligned or anti-aligned with $\Delta\gamma$, the energy E_L favors their alignment; (b): when at least one of the two tangent vectors is nearly orthogonal to $\Delta\gamma$, there is only a very small force between the two points, but contributions from many points can add up to a significant repulsion, as denoted by F .

E_L is not only a function of $(\theta - \theta')$, but also a function of $(\theta + \theta')$. If $\Delta\gamma(t, t')$ acts as a canonical axis, $(\theta - \eta)$ and $(\theta' - \eta)$ are now the “fundamental” angles. Note that the first cosine function in equation (5.20) is just the inner product $\dot{\gamma}(t) \cdot \dot{\gamma}(t')$; thus the first part of equation (5.20) has the same form as the standard HOAC term (except that $d_2 \neq d$). The second cosine function is the inner product of $\dot{\gamma}(t)$ with the reflection of $\dot{\gamma}(t')$ in $\Delta\gamma(t, t')$. So the second part is a new term that results in the new effect of E_L , but we will still make use of E_L in equation (5.19), because it is more convenient to analyse and implement this form.

5.3.2 Phase Field Model Definition

We now reformulate $E_L(R)$ in the phase field framework. We rotate tangent vectors to normal vectors, and replace the latter by $\nabla\phi$. Since $\nabla\phi$ is very small outside R_C , the domains of integration can be extended from ∂R to Ω without significantly changing the energy, except for a multiplicative factor. By introducing a weight parameter β_2 , we define the *linear nonlocal* HOAC phase field term $E_L(\phi)$ as

$$E_L(\phi) = -\frac{\beta_2}{2} \iint_{\Omega^2} (\nabla\phi(x) \times (x - x')) (\nabla\phi(x') \times (x - x')) \Psi\left(\frac{|x - x'|}{d_2}\right) dx dx', \quad (5.21)$$

where \times is the 2D vectorial antisymmetric product. The functional derivative of E_L is

$$\frac{\delta E_L(\phi)}{\delta \phi(x)} = \beta_2 \int_{\Omega} \nabla \cdot (\epsilon(x - x')(x - x')^T \epsilon^T) \cdot \nabla\phi(x') \Psi\left(\frac{|x - x'|}{d_2}\right) dx', \quad (5.22)$$

where ϵ rotates the tangent vectors to the inward normal vectors. Since this functional derivative is linear in ϕ , we call E_L as the *linear nonlocal* term.

After incorporating the additional *linear nonlocal* term E_L into our primary model E_{primary} (section 2.4), we define the total energy as $E = \theta(E_0 + E_S + E_L) + E_D$. To minimize E ,

we perform, as usual, gradient descent with the neutral initialization. The functional derivative of E_L , $\delta E_L/\delta\phi$, like all the other nonlocal terms, is calculated in the Fourier domain. So the evolution equation of E_L is given by

$$-\frac{\delta E_L(\phi)}{\delta\phi(x)} = \beta_2 \mathcal{F}^{-1} \left\{ \left(k_1^2 \mathcal{F} \left\{ r^2 \sin^2 \theta \Psi(r/d_2) \right\} - 2k_1 k_2 \mathcal{F} \left\{ r^2 \sin \theta \cos \theta \Psi(r/d_2) \right\} + k_2^2 \mathcal{F} \left\{ r^2 \cos^2 \theta \Psi(r/d_2) \right\} \right) \hat{\phi}(k) \right\}, \quad (5.23)$$

where $k = (k_1, k_2)^T$ is frequency in the Fourier domain, and x is represented by its polar coordinate (r, θ) . For more details, see Appendix B. By adding this term into equation (2.24), we obtain the overall evolution equation of our new linear model.

5.3.3 Stability Analysis of the Linear Nonlocal HOAC Total Prior Model

The sum of the three energies, $E_{P,L} = E_0 + E_S + E_L$ constitutes the *linear nonlocal* HOAC total prior model. Compared to the standard prior model, two more parameters β_2 and d_2 have been introduced, so the situation is much more complicated. The behavior of $E_{P,L}$ depends on the six parameters $(\alpha, \lambda, \beta, \beta_2, d, d_2)$, and can vary significantly. Therefore, in this subsection, we analyze the stability conditions of the model for a network branch.

Linear Nonlocal Phase Field HOAC Term of a Bar

We evaluate the linear nonlocal HOAC term E_L (equation (5.21)) on the *ansatz* of a long straight bar described in equation (2.9). We obtain

$$E_L = -8\beta_2 L \left\{ \int_0^\infty z^2 \Psi\left(\frac{z}{d_2}\right) dz - d^3 \int_{\hat{W}}^\infty \eta \sqrt{\eta^2 - \hat{W}^2} \Psi\left(\frac{\eta}{\hat{d}_2}\right) d\eta \right\}, \quad (5.24)$$

where $\hat{d}_2 = d_2/d$ is the ratio between two interaction ranges. Replacing Ψ by equation (5.2), we have the linear nonlocal phase field HOAC energy, per unit length of bar, e_L :

$$\begin{aligned} e_L(w, \hat{W}) &= -\frac{16\beta_2 d_2^3}{3} \left(1 - \frac{3}{\pi^3}\right) + 4\beta_2 d^3 \int_{\hat{W}}^\infty \eta \sqrt{\eta^2 - \hat{W}^2} \Psi\left(\frac{\eta}{\hat{d}_2}\right) d\eta \\ &= 4\beta_2 d^3 \int_{\hat{W}}^{2\hat{d}_2} \eta \sqrt{\eta^2 - \hat{W}^2} \left(2 - \frac{\eta}{\hat{d}_2} + \frac{1}{\pi} \sin\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta + C_3, \end{aligned} \quad (5.25)$$

where $C_3 = -16\beta_2 d_2^3(1 - 3/\pi^3)/3$ is a constant independent of w and \hat{W} .

Linear Nonlocal HOAC Total Prior Model of a Bar

Now we sum up all the energies per unit length, *i.e.* e_0 (equation (2.13)), e_S (equation (2.15)) and e_L , and remove the constant terms. The linear nonlocal HOAC total prior model per unit

length of bar, $e_{P,L}$, is

$$\begin{aligned}
 e_{P,L}(w, \hat{W}) &= e_0 + e_S + e_L \\
 &= \frac{4}{3}\alpha\hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + 4\beta d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta \\
 &\quad + 4\beta_2 d^3 \int_{\hat{W}}^{2\hat{d}_2} \eta \sqrt{\eta^2 - \hat{W}^2} \left(2 - \frac{\eta}{\hat{d}_2} + \frac{1}{\pi} \sin\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta. \tag{5.26}
 \end{aligned}$$

As for $e_{P,NL}$, we find the minimum of $e_{P,L}$ by setting its first derivatives, with respect to w and \hat{W} , to zero, while ensuring that its second derivative, with respect to \hat{W} , is non-negative. This leads to lengthy calculations that we detail in Appendix A.2.3. Finally we find

$$\begin{aligned}
 \frac{\partial e_{P,L}}{\partial \hat{W}} &= \frac{4}{3}\alpha d - 4\beta\hat{W}d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta \\
 &\quad - \frac{4\beta_2\hat{W}d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\frac{\pi\eta}{\hat{d}_2})) d\eta, \tag{5.27a}
 \end{aligned}$$

$$\frac{\partial e_{P,L}}{\partial w} = \frac{4}{15}\lambda - \frac{4}{w^2}, \tag{5.27b}$$

$$\begin{aligned}
 \frac{\partial^2 e_{P,L}}{\partial \hat{W}^2} &= -4\beta d \ln\left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}}\right) + 4\beta d \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta \\
 &\quad - \frac{4\beta_2 d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \frac{\eta^2 - 2\hat{W}^2}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\frac{\pi\eta}{\hat{d}_2})) d\eta. \tag{5.27c}
 \end{aligned}$$

It follows that stability is related to the three scaled control parameters $\hat{\beta} = \beta/\alpha$, $\hat{\beta}_2 = \beta_2 d^2/\alpha$ and $\hat{d}_2 = d_2/d$, and also to the scaled width $\hat{W} = W/d$. The parameter constraints are then as follows:

$$1 - 3\hat{\beta}\hat{W}I_1(\hat{W}) - \frac{3\hat{\beta}_2\hat{W}}{\hat{d}_2}I_3(\hat{W}) = 0, \tag{5.28a}$$

$$\hat{\beta}I_2(\hat{W}) - \frac{\hat{\beta}_2}{\hat{d}_2}I_4(\hat{W}) \geq \hat{\beta} \ln\left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}}\right), \tag{5.28b}$$

$$\lambda = \frac{15}{w^2}, \tag{5.28c}$$

where $I_1(\hat{W})$ and $I_2(\hat{W})$ have been defined in equation (5.17); and $I_3(\hat{W})$ and $I_4(\hat{W})$ take the

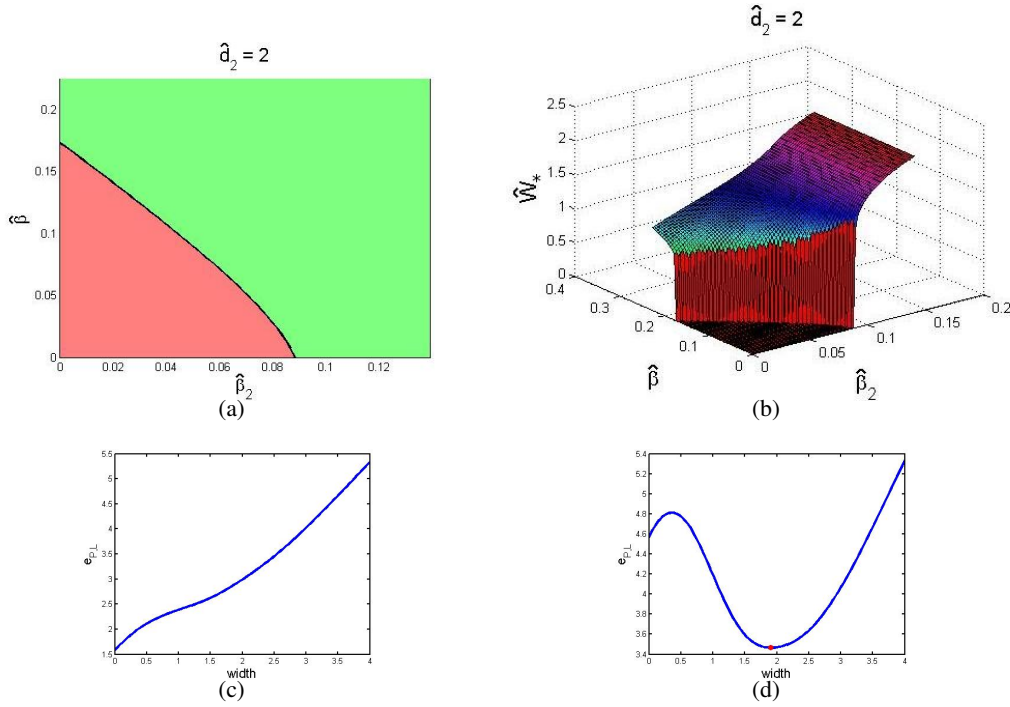


Figure 5.5. $\hat{d}_2 = 2 < \hat{D}_2$. (a): different regions in the $\hat{\beta}_2 - \hat{\beta}$ plane. $e_{P,L}$ has either no local minimum (red) or one local minimum (green). (b): the associated stable bar width \hat{W}_* . (c): $e_{P,L}$ with no local minimum ($\hat{\beta} = 0.05, \hat{\beta}_2 = 0.04$). (d): $e_{P,L}$ with one local minimum ($\hat{\beta} = 0.2, \hat{\beta}_2 = 0.1$).

form:

$$I_3(\hat{W}) = \int_{\hat{W}}^{2\hat{d}_2} \sqrt{\eta^2 - \hat{W}^2} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta, \quad (5.29a)$$

$$I_4(\hat{W}) = \int_{\hat{W}}^{2\hat{d}_2} \frac{\eta^2 - 2\hat{W}^2}{\sqrt{\eta^2 - \hat{W}^2}} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta. \quad (5.29b)$$

For w , the constraint is trivial, leading to $\lambda = 15/w^2$. The stable width \hat{W} , *i.e.* the value of \hat{W} where a local minimum of $e_{P,L}$ is found, depends on the parameters $\hat{\beta}$, $\hat{\beta}_2$ and \hat{d}_2 . Let us compare the first constraint in equation (5.28a) to the constraint of the standard HOAC total prior model in equation (2.19a). In the latter model, the part $\hat{W}I_1(\hat{W})$ is a simple curve with one maximum inside the interval $[0, 2)$, so for a given $\hat{\beta}$, the number of solutions of this equation is at most two: one corresponds to the local minimum of the energy; and the other the local maximum. However, due to the incorporation of the new linear prior energy E_L , we have an additional part in the above first constraint. This additional part $\hat{W}I_3(\hat{W})/\hat{d}_2$ is also a simple curve with one maximum but in the interval $[0, 2\hat{d}_2)$. As a result, in the case

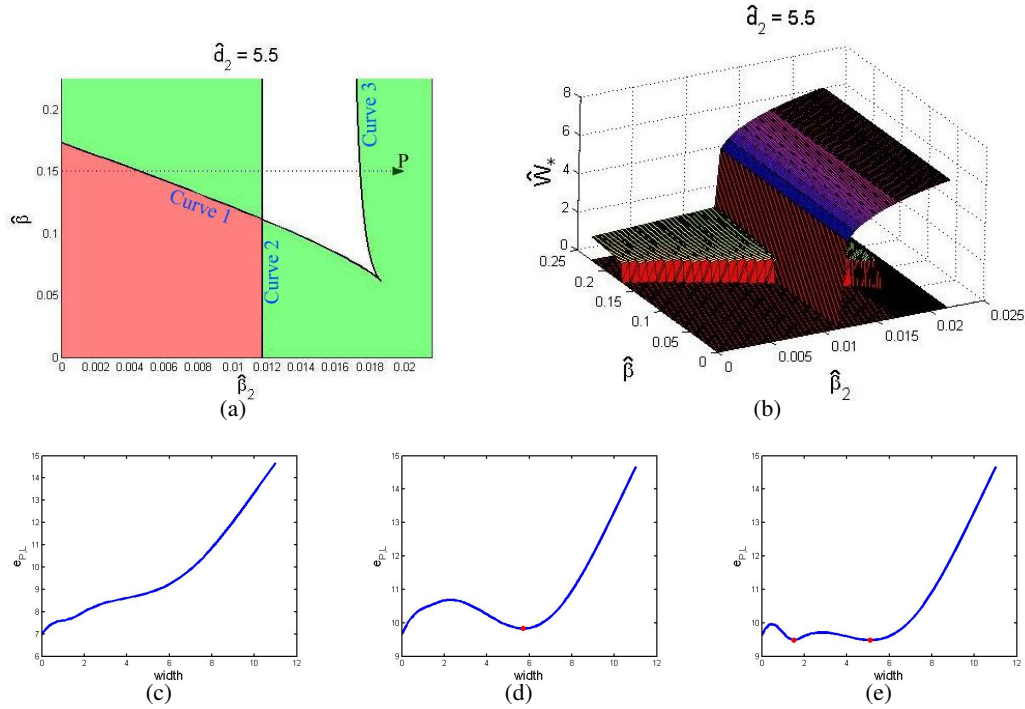


Figure 5.6. $\hat{d}_2 = 5.5 > \hat{D}_2$. (a): different regions in the $\hat{\beta}_2 - \hat{\beta}$ plane. $e_{P,L}$ has either no local minimum (red), one local minimum (green), or two local minima (white). (b): the associated stable bar width(s) \hat{W}_* . (c): $e_{P,L}$ with no local minimum ($\hat{\beta} = 0.1, \hat{\beta}_2 = 0.01$). (d): $e_{P,L}$ with one local minimum ($\hat{\beta} = 0.05, \hat{\beta}_2 = 0.015$). (e): $e_{P,L}$ with two local minima ($\hat{\beta} = 0.2, \hat{\beta}_2 = 0.013$).

where \hat{d}_2 is small, the effective intervals of the two curves are mostly overlapped, and the first constraint possesses the same property as the previous constraint in equation (2.17a): the possible number of solutions is zero or two; while in the case that \hat{d}_2 is large enough so that one curve has a longer effective interval than the other, the property of the first constraint changes significantly: the possible number of solutions is zero, two, or four.

Consequently, there is a singular point of \hat{d}_2 , where the maximum number of solutions jumps from two to four, and accordingly the stable width(s) that the prior energy can model jump(s) from one to two. However, since equation (5.28a) involves complicated integrals and the piecewise interaction functions, it is non-trivial to obtain analytically the number of its solutions. Therefore, we carry out numerical experiments to find an approximate value for this singular point. We find that such a singular point of \hat{d}_2 (named \hat{D}_2) indeed exists, with $\hat{D}_2 \approx 2.7$. If \hat{d}_2 is less than \hat{D}_2 , at most one local minimum can be found. If $\hat{d}_2 > \hat{D}_2$, there are three cases, depending on the values of $\hat{\beta}$, $\hat{\beta}_2$ and \hat{d}_2 : $e_{P,L}$ has no local minimum; $e_{P,L}$ has one local minimum, with either $\hat{W} \approx 1$ (i.e. $W \approx d$) or $\hat{W} \approx \hat{d}_2$ (i.e. $W \approx d_2$); or $e_{P,L}$ has two local minima, at $\hat{W} \approx 1$ and $\hat{W} \approx \hat{d}_2$.

The two regimes are illustrated in Figures 5.5 and 5.6. For both regimes, the associated

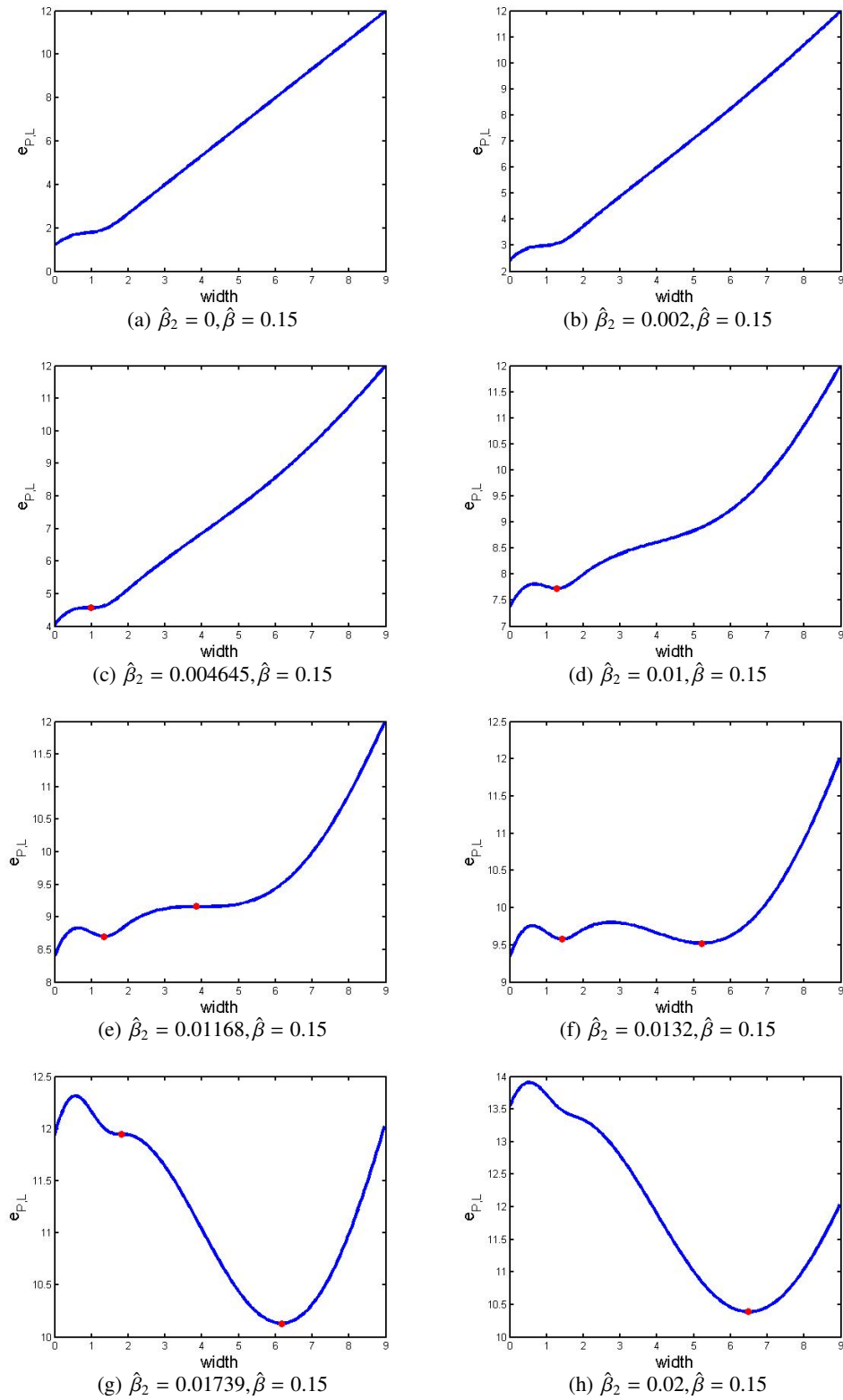


Figure 5.7. Changes in the energy graphs when the parameter setting is moved from left to right along the dotted path P shown in Figure 5.6 ($\hat{d}_2 = 5.5$). Red dot: local minimum.

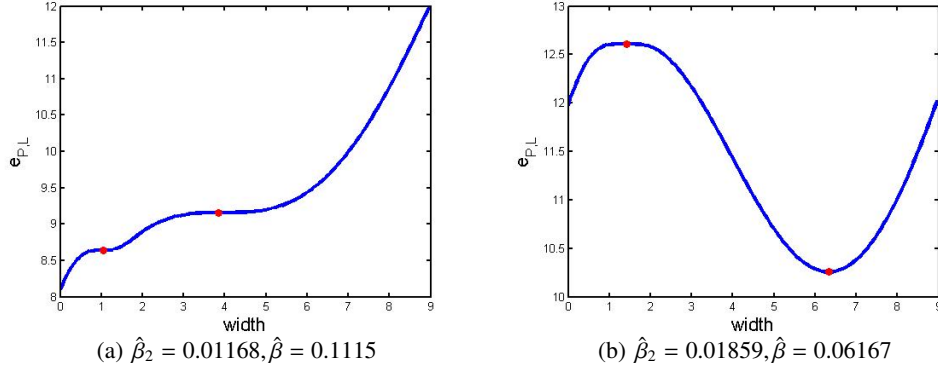


Figure 5.8. Energy graphs when the parameter setting is chosen in the two intersection points of curves shown in Figure 5.6 ($\hat{d}_2 = 5.5$). Red dot: local minimum.

stable bar width, where the local minimum of the prior energy $e_{P,L}$ is found, as well as the possible behaviors of the energy, are shown in the same figure. Figure 5.5 gives an example for $\hat{d}_2 = 2 < \hat{D}_2$. In the $\hat{\beta}_2 - \hat{\beta}$ plane, the red region and the green region correspond to the situations when $e_{P,L}$ has no local minimum, and one local minimum respectively. The sole separation curve between the two regions corresponds to the situation when the two extrema of $e_{P,L}$ become an inflection point. This curve demonstrates a qualitative change in the behavior of $e_{P,L}$.

The case for $\hat{d}_2 > \hat{D}_2$ (here $\hat{d}_2 = 5.5$) is shown in Figure 5.6. To make things clearer, we label the three curves by numbers; and we name the widths where the two possible local minima are found \hat{W}_{MIN_1} and \hat{W}_{MIN_2} , and the widths where the two possible local maxima are found \hat{W}_{MAX_1} and \hat{W}_{MAX_2} , by the ascending order of their values, so $\hat{W}_{\text{MAX}_1} \leq \hat{W}_{\text{MIN}_1} \leq \hat{W}_{\text{MAX}_2} \leq \hat{W}_{\text{MIN}_2}$. (In the case of equality, the extrema of the energy merge as the inflection point(s).) We examine the situation of each curve and each intersection point between two curves:

- Curve 1: \hat{W}_{MAX_1} and \hat{W}_{MIN_1} merge together and become an inflection point. See Figure 5.7(c).
- Curve 2: \hat{W}_{MAX_2} and \hat{W}_{MIN_2} merge together and become an inflection point. See Figure 5.7(e).
- Curve 3: \hat{W}_{MIN_1} and \hat{W}_{MAX_2} merge together and become an inflection point. See Figure 5.7(g).
- Intersection point between curve 1 and curve 2: \hat{W}_{MAX_1} merge with \hat{W}_{MIN_1} and also \hat{W}_{MAX_2} merge with \hat{W}_{MIN_2} . Four extrema of the energy merge as two inflection points. See Figure 5.8(a).
- Corner point between curve 1 and curve 3: \hat{W}_{MAX_1} , \hat{W}_{MIN_1} and \hat{W}_{MAX_2} merge together. The energy has a very wide maximum (the inflection point of the three ex-

trema) and a local minimum. See Figure 5.8(b).

- Corner point between curve 2 and curve 3: it actually goes off to infinity, which arises from the finite range of the interaction function Ψ we use here. It corresponds to the situation where \hat{W}_{MIN_1} , \hat{W}_{MAX_2} and \hat{W}_{MIN_2} merge together. The energy should have a local maximum and a very wide local minimum (the inflection point of the three extrema). This is exactly what we need, for modeling multiple widths of network branches inside a somewhat wide range. Unfortunately, we cannot really use this ‘critical point’. In Appendix D, we discuss how to pull this corner point back to a finite value, by changing to another type of interaction function.

In a word, each curve is a set of points where two extrema of the energy merge as an inflection point, and where a qualitative change in the energy behavior happens; each intersection point combines the properties of the two relative curves.

Let us now observe how the local minimum/minima of the energy $e_{P,L}$ evolve(s) according to the three different states, *i.e.* no local minimum, one local minimum and two local minima, if the parameter setting is moved from left to right along the dotted path P shown in Figure 5.6. The sequence of energies $e_{P,L}$ is plotted in Figure 5.7. The state starts with no local minimum (Figures 5.7(a)-5.7(b)). When the path P meets curve 1, a first local minimum appears, and the state jumps from no local minimum to one local minimum. Initially, the first local minimum is an inflection point (Figure 5.7(c)). The first local minimum gradually becomes deeper, but the state stays in one local minimum, held there stably by the dynamic (Figure 5.7(d)). Eventually, when the path P meets curve 2, a second local minimum, *i.e.* an inflection point, appears (Figure 5.7(e)), and the state jumps from one local minimum to two local minima (Figure 5.7(f)). When the parameter setting continues being moved to the right and meets curve 3, the first local minimum and the second local maximum merge together, and become an inflection point (Figure 5.7(g)). In the end, the first local minimum disappears, and the state jumps back to one local minimum again (Figure 5.7(h)). The red dots in Figure 5.7 denote the local minima. They are shown in Figure 5.6(b).

According to catastrophe theory (Zeeman, 1977), in our problem, the parameters related to energy stability constitute an M -dimensional parameter space C , the width of the bar is a 1-dimensional state space X , and the energy e_P is a smooth function on X parameterized by C . As noted in subsection 2.3.4, the relationship between the parameters of the standard HOAC prior model and the model behavior, shown in Figure 2.6 is an example of a *fold catastrophe*, because in that case, the dimension M of the parameter space is equal to 1 (there is only one parameter $\hat{\beta}$ in space C), and the possible states of the 1-dimensional state space are respectively no local minimum and one local minimum of the energy. Whereas, here for the linear nonlocal HOAC prior model, two additional parameters \hat{d}_2 and $\hat{\beta}_2$ are introduced. Then, $M = 3$, and besides the two previously existing states, a third new behavior in the 1-dimensional state space corresponds to two local minima of the energy. This behavior is an example of a *swallowtail catastrophe*. The analytical study illustrated by the diagrams in Figures 5.5 and 5.6 confirms this statement.

The variety of behaviors is important for applications. As well as being able to model networks with branches of more or less fixed width, but with greater ‘stiffness’ than provided by the model in (Rochery et al., 2006), the new energy can model two widths at the

Measure Method	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
New model E (with E_{NL}) (Figure 5.9(d))	0.9524	0.8591	0.8237
E_{primary} ($\beta_2 = 0$) (Figure 5.10(a))	0.8832	0.8659	0.7769
$\theta E_0 + E_D$ (Figure 5.10(c))	0.4282	0.8314	0.3940
MLE ($\theta = 0$) (Figure 5.10(b))	0.9734	0.1831	0.1822

Table 5.1. Quantitative evaluation criteria for the different methods tested on Figure 5.9(a) at 1/4 resolution (T = True, F = False, P = Positive, N = Negative). The completeness is the percentage of ground truth road network that is extracted; the correctness is the percentage of extracted road network that is correct; and the quality is the most important measure of the “goodness” of the result, because it takes into account the completeness and the correctness.

same time. At certain ‘critical points’ in parameter space, essentially where pairs of minima merge, it could also model a large range of widths, all of which are approximately stable.

5.4 Experimental Results and Comparisons

As input data I , we use a number of images, with average size 1200×1200 pixels, extracted from a QuickBird optical panchromatic image of Beijing. The scenes are characteristic of dense urban regions. Our aim is to extract, completely and accurately, the road network from an image. In order to evaluate the performance of our new *nonlinear* and *linear* models, we compare them quantitatively to ground truth and to other methods from the literature. Ground truth for the evaluations is segmented by hand. We also analyze the effect of the different terms in our energies.

5.4.1 Nonlinear Nonlocal Overall Model

In this subsection, we demonstrate the behavior of our new nonlinear model containing the *nonlinear, nonlocal* HOAC prior term E_{NL} , *i.e.* $E = \theta(E_0 + E_S + E_{NL}) + E_D$ via experiments on real QuickBird panchromatic images in dense urban areas, at reduced resolutions.

Figure 5.9(a) shows one of the input images at 1/4 resolution. The parameters $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d)$ are $(100, 0, 0.12, 3.8, 0.0375, 0.0338, 4)$. The results obtained using the energy with the new nonlinear nonlocal term E_{NL} at iterations 1, 1, 500 and 27, 000 are illustrated in Figures 5.9(b)-5.9(d). The result obtained using our primary model E_{primary} (without E_{NL}) is shown in Figure 5.10(a). We see that adding E_{NL} enables the recovery of the main and

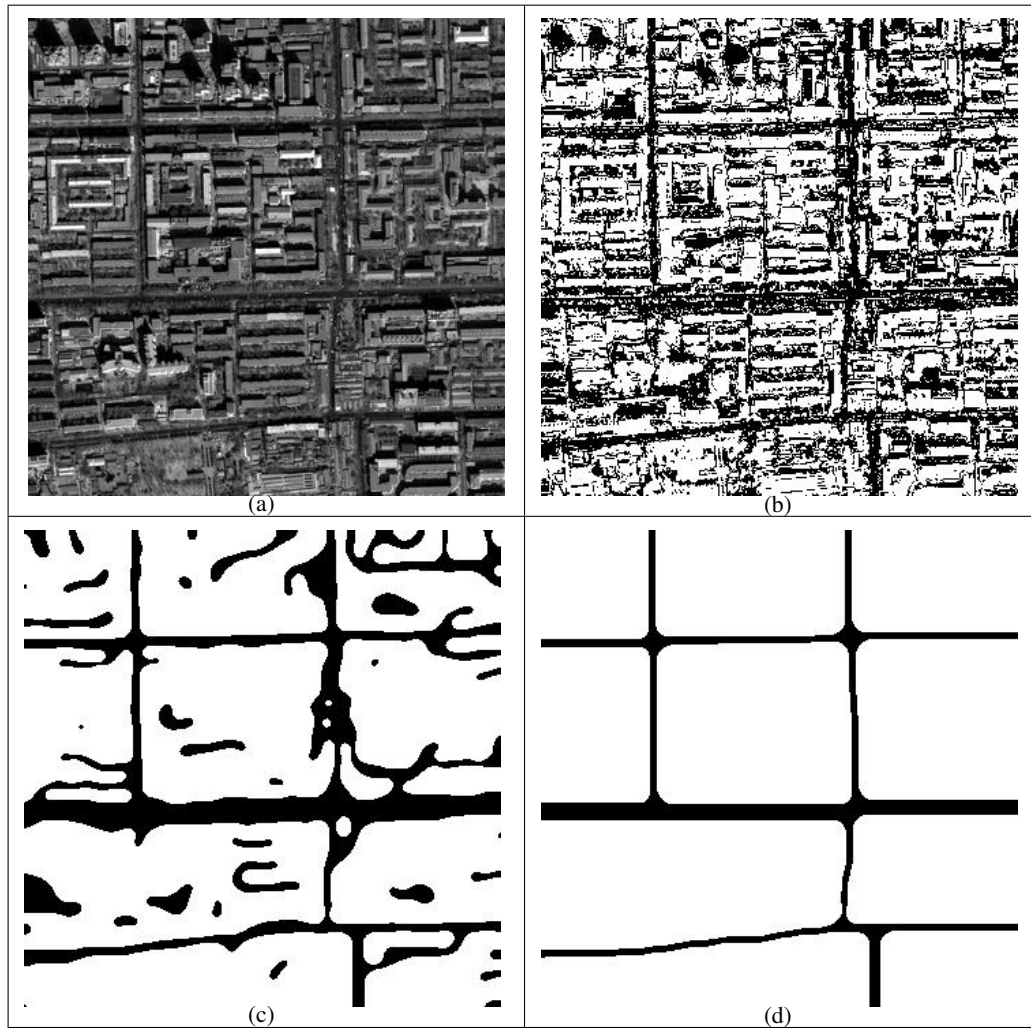


Figure 5.9. Data, and experiment at $1/4$ resolution. (a): image data (size = 350×350 , road width = $3 \sim 5$ pixels). (b)-(d): results obtained using the new nonlinear model (with E_{NL}) at iterations 1, 1,500 and 27,000.

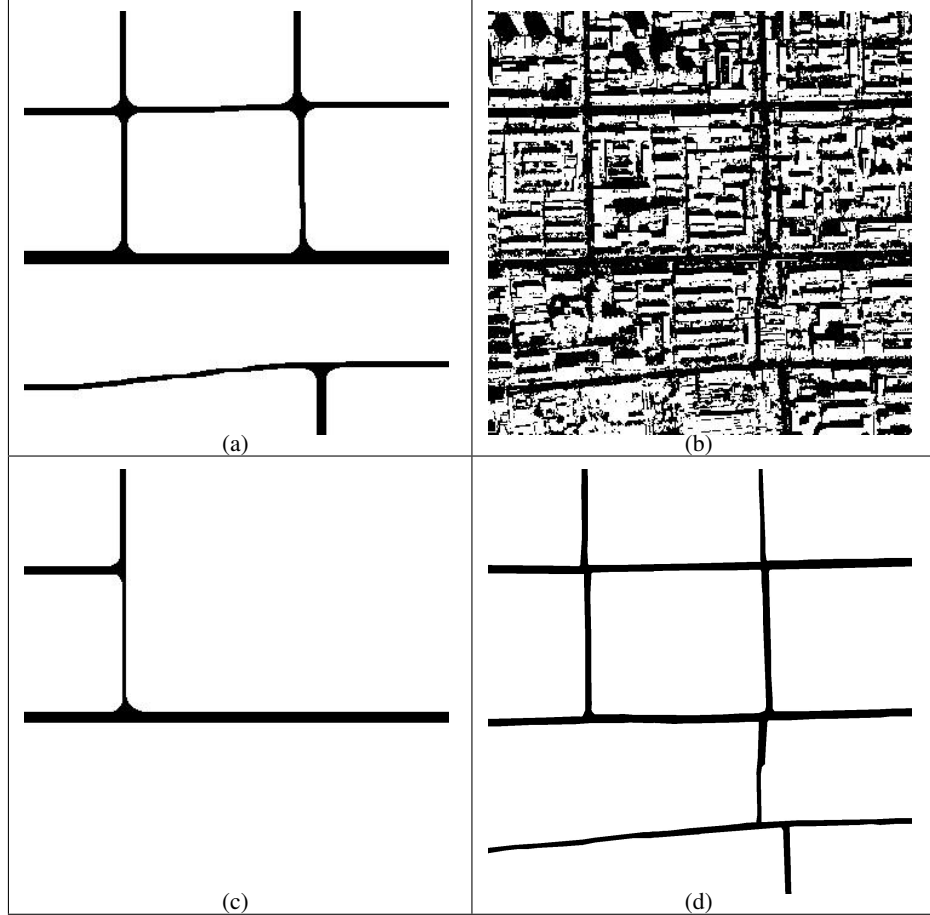


Figure 5.10. Experiments at 1/4 resolution and ground truth. (a)-(c): results obtained respectively using the energy E_{primary} (without E_{NL}), MLE, and a standard, non-higher-order active contour model (with neither E_S nor E_{NL}). (d): ground truth segmented manually.

secondary road network, whereas the model without E_{NL} misses a secondary road. In order to illustrate the effects of other terms in the model, we compute results using maximum likelihood estimation (MLE, *i.e.* $\theta = 0$) and a standard, non-higher-order active contour, *i.e.* $\beta = \beta_2 = 0$ (see Figures 5.10(b) and 5.10(c)). The ground truth is presented in Figure 5.10(d). The MLE result shows that local image information alone is not sufficient to distinguish the roads from the background, while the standard active contour result shows the importance of the geometric knowledge introduced by HOACs. Quantitative evaluations based on standard criteria (Heipke et al., 1997) are shown in Table 5.1. On the other hand, the computation time for the result in Figure 5.9 is around 80 minutes, which is considerably slower than the next nearest time, that obtained with the model E_{primary} (Figure 5.10(a)).

Figure 5.11 presents more results at reduced resolutions. From top to bottom, the parameters $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d)$ are respectively $(200, 0, 0.12, 4.5, 0.035, 0.035, 3)$, $(200, 0, 0.0905,$

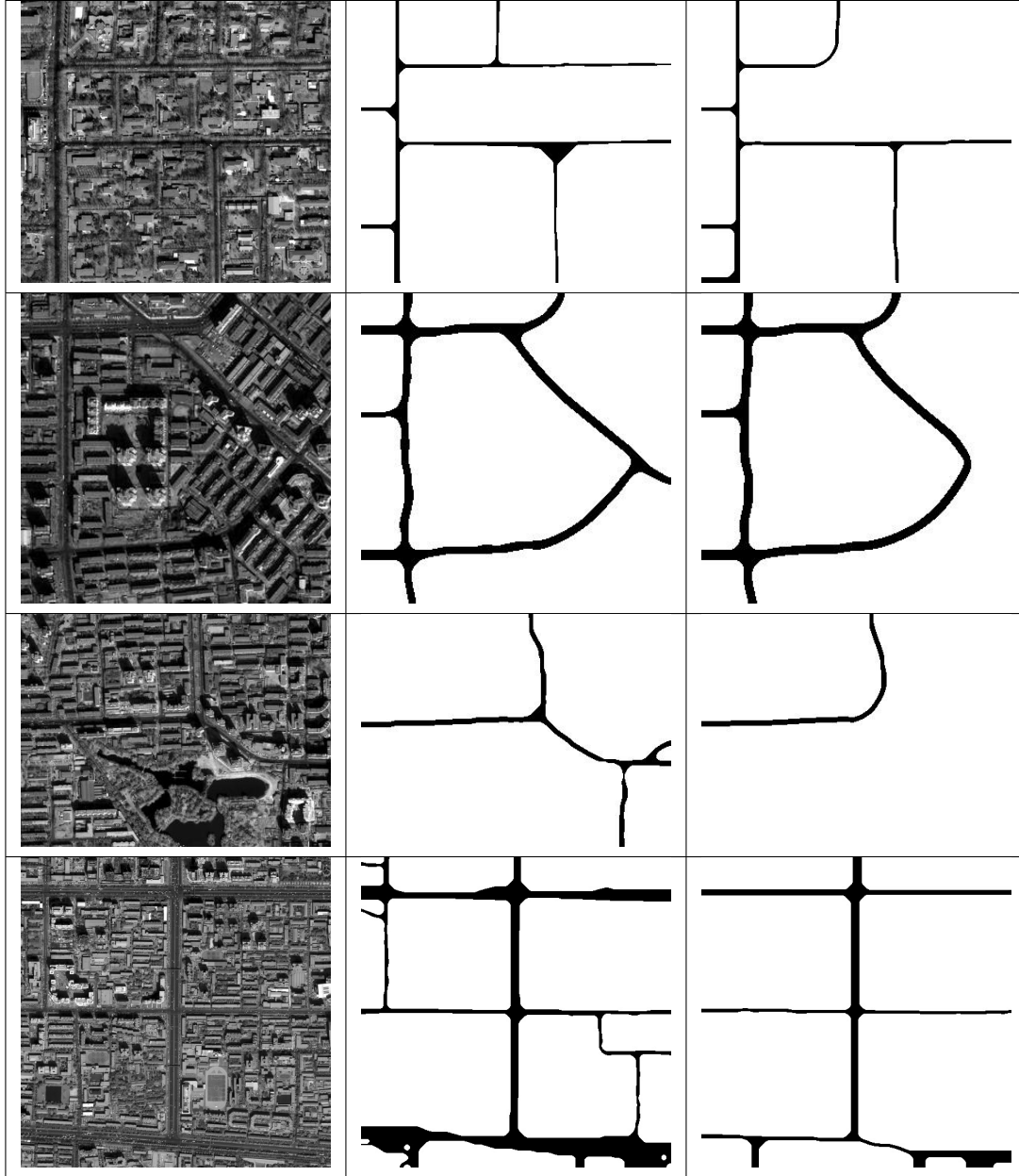


Figure 5.11. More experiments at reduced resolutions. First column: input images, first row: 1/2 resolution, size = 400×440 , road width = 2 ~ 4 pixels; second row: 1/4 resolution, size = 300×300 , road width = 3 ~ 5 pixels; third row: 1/4 resolution, size = 300×400 , road width = 3 ~ 6 pixels; last row: 1/4 resolution, size = 512×512 , road width = 3 ~ 15 pixels. Two rightmost columns: corresponding results obtained using the new nonlinear model (with E_{NL}) and the primary model E_{primary} (without E_{NL}).

4.5, 0.0375, 0.025, 5), (200, 0, 0.0905, 5, 0.0375, 0.025, 4) and (100, 0, 0.1, 4.5, 0.0375, 0.03, 4). The first column shows the input image data, which is either at 1/4 or 1/2 resolution. The two columns on the right show the corresponding results obtained with and without the new nonlinear, nonlocal term E_{NL} . The importance of E_{NL} is clear: it facilitates greatly the retrieval of secondary roads. However, the nonlinear nonlocal prior term is computationally expensive, which is why the method is applied only at reduced resolutions. To solve this problem, which is due to the nonlinear nature of E_{NL} , we show the effect of the *linear nonlocal* prior term in the next subsection.

5.4.2 Linear Nonlocal Overall Model

In this subsection, we focus on two particular cases of road extraction: extraction of a network consisting of roads of roughly the same width; and extraction of networks containing roads of two different widths. In the former case, we choose the parameters so that $e_{P,L}$ has one local minimum. The resulting model can extract roads whose widths' are close to the minimizing value. In the latter case, we choose the parameters so that $e_{P,L}$ has two local minima. Again a small range of widths around each minimum is possible.

Extraction of Roads of Similar Width

We apply our new linear model $E = \theta(E_0 + E_S + E_L) + E_D$ to both the full-resolution and reduced resolution images. We fix the parameters as described in subsection 5.3.3.

The results obtained using the new linear model E (with E_L), at 1/4 resolution and at full resolution, are shown in Figures 5.12(b) and 5.12(c). For this experiment, the parameters $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d, d_2)$ are (200, 0, 0.15, 4, 0.02, 2×10^{-4} , 4, 12) and (200, 0, 0.15, 4, 0.02, 1.25×10^{-5} , 16, 48) at 1/4 and full resolution respectively. Note that apart from the obvious scaling of d and d_2 , and a change in β_2 , the other parameters are the same for the two resolutions. The computational speed of the linear model is equivalent to that of the primary model; theoretically, they are supposed to be two times as fast as the nonlinear model⁴. In practice, the computational time of the linear model for this experiment is 60 minutes and 936 minutes at 1/4 and full resolution. The complete road network is retrieved successfully, at both resolutions. Although the segmentation at 1/4 resolution appears geometrically smoother and converges faster, the extraction result is actually more accurate at full resolution. Accuracy at 1/4 resolution is limited both directly, by the low resolution of the phase field, and indirectly, because each scaling coefficient in the data at level 2 is the average of 16 pixels at full resolution: coefficients near the road border therefore include both road and background contributions, and the road boundary is thereby blurred.

To evaluate the performance of the new model, we now compare our result with other methods at full resolution. As before, we compare results obtained with MLE (*i.e.* $\theta = 0$); with a standard, non-higher-order active contour (*i.e.* $\beta = \beta_2 = 0$); and with the primary

⁴As already mentioned, we compute the functional derivative of the energy in the Fourier domain. At each iterative step, for the nonlinear model, we need to compute twice a Fourier transfer and twice an inverse Fourier transfer; while for the linear model or the primary model, only one Fourier transfer and one inverse Fourier transfer are required.

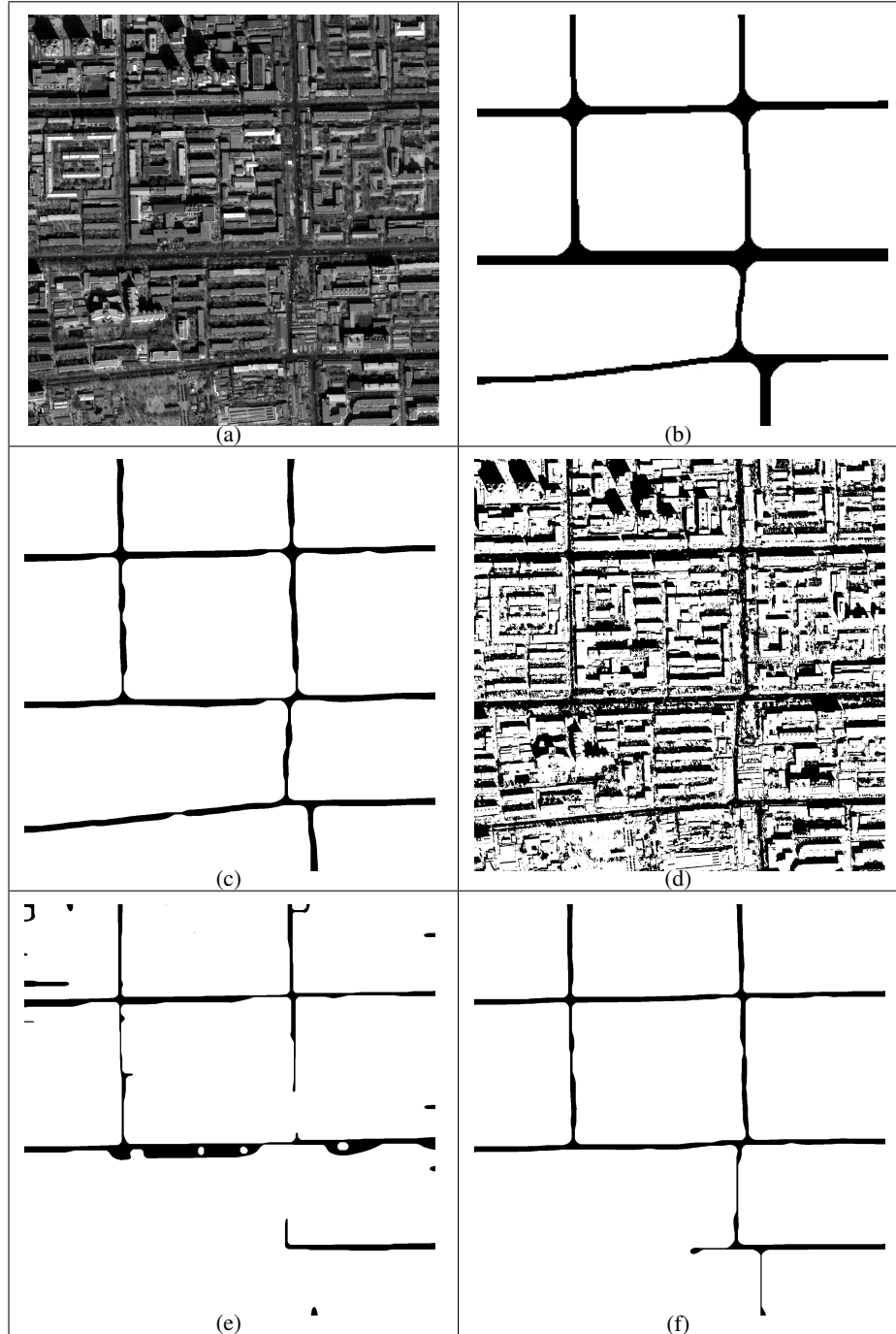


Figure 5.12. Experiments using the new linear model (with E_L), and analysis of the effect of the different terms in the energy. (a): image data (size: 1400×1400); (b): result obtained using the new linear model (with E_L), at $1/4$ resolution; (c): result obtained using the new linear model (with E_L), at full resolution; and results obtained at full resolution using (d): MLE; (e): the model with $\beta = \beta_2 = 0$ (equivalent to a standard active contour); (f): the primary model E_{primary} (without E_L , i.e. $\beta_2 = 0$).

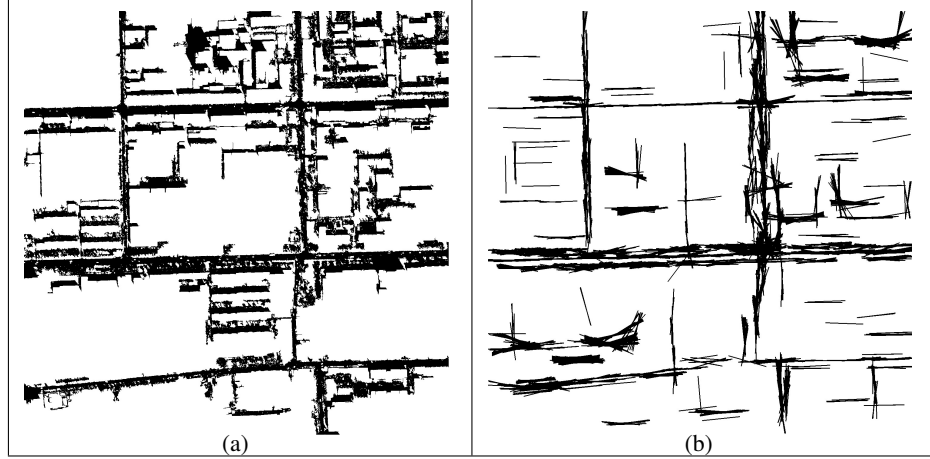


Figure 5.13. Comparisons at full resolution. (a)-(b): results obtained with (Wang and Zhang, 2003) and (Yu et al., 2004).

model E_{primary} (without E_L , *i.e.* $\beta_2 = 0$). They are shown in Figures 5.12(d)-5.12(f). MLE is clearly incapable of distinguishing the roads from the background, while the models without E_L and/or E_S are not able to recover the complete road network (although that with E_S does better than the standard active contour, with only local prior knowledge). In addition, we apply two other methods, proposed in (Wang and Zhang, 2003) and (Yu et al., 2004), and compare them to ours (see Figure 5.13). Without much prior geometric knowledge, they extract many incorrect areas that happen to have statistical properties similar to roads. Moreover, the accuracy of the delineation of the road boundary is poor. Some quantitative evaluations based on standard criteria (Heipke et al., 1997), are shown in Table 5.2. The ‘quality’ is the most important measure because it considers both completeness and correctness. The linear model and the nonlinear model at 1/4 resolution achieve a similar performance; and the linear model at full resolution outperforms all others.

Figure 5.14 presents more results using the new linear model (with E_L), at a reduced resolution and at full resolution. From top to bottom, at a reduced resolution, the parameters $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d, d_2)$ are respectively $(200, 0, 0.15, 5, 0.025, 2 \times 10^{-4}, 3, 9)$, $(200, 0, 0.15, 5, 0.02, 3 \times 10^{-4}, 4, 12)$, $(200, 0, 0.0905, 5, 0.03, 1.5 \times 10^{-4}, 4, 12)$, and $(200, 0, 0.12, 5, 0.02, 2 \times 10^{-4}, 4, 12)$; and at full resolution, the parameters $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d, d_2)$ are respectively $(200, 0, 0.15, 4, 0.025, 5 \times 10^{-5}, 6, 18)$, $(200, 0, 0.15, 5, 0.02, 1.25 \times 10^{-5}, 16, 48)$, $(200, 0, 0.15, 5, 0.022, 1.375 \times 10^{-5}, 16, 48)$, and $(200, 0, 0.15, 5, 0.02, 1.25 \times 10^{-5}, 16, 48)$. In practice, the results are not very sensitive to the precise choice of parameter values, provided they lie in the correct subset of the $\hat{\beta} - \hat{\beta}_2 - \hat{d}_2$ diagram. We also apply the same model to river extraction, as shown in Figure 5.17. Figures 5.15 and 5.16 show respectively a large image (size: 1100×3300) and the road extraction result obtained using our models, at full resolution.

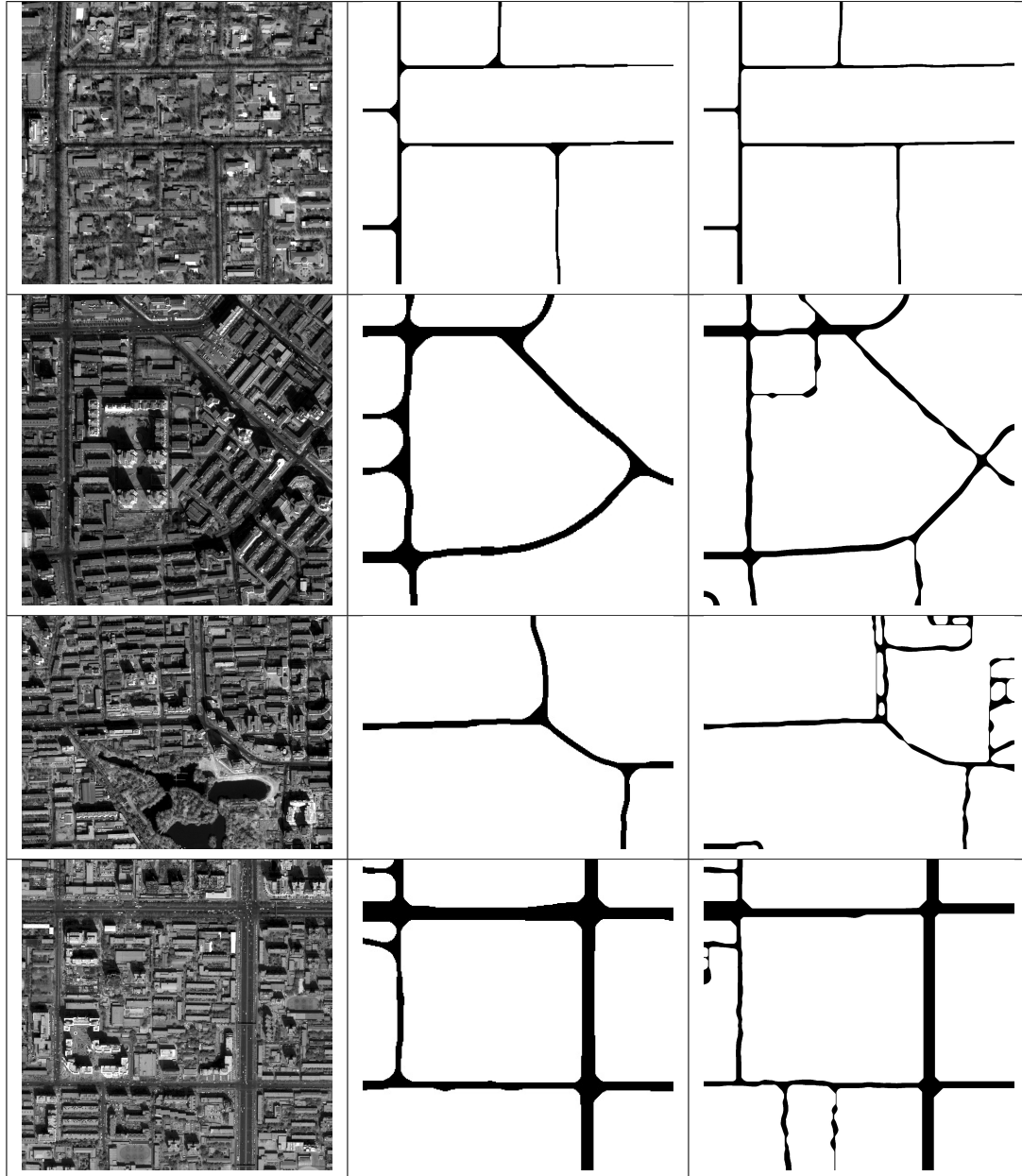


Figure 5.14. More results using the new linear model (with E_L) on pieces of a QuickBird image at a reduced resolution and at full resolution. Image size: first row: 800×880 ; second row: 1200×1200 ; third row: 1200×1600 ; last row: 1400×1400 .



Figure 5.15. A large image (size: 1100 × 3300).

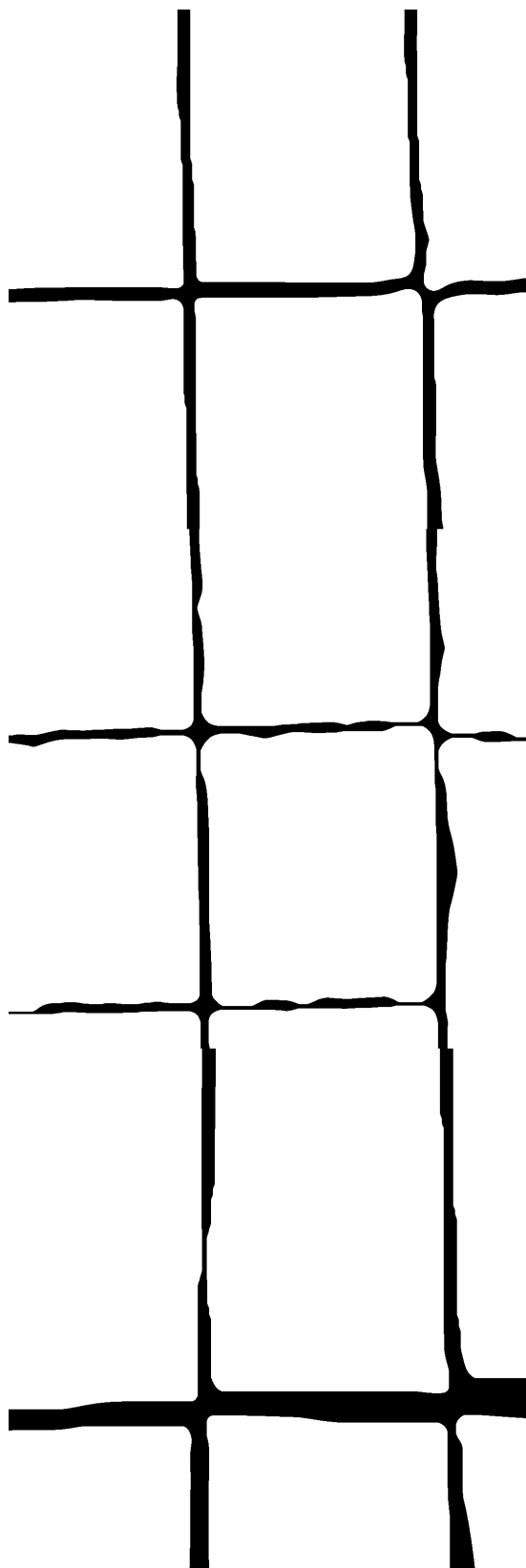


Figure 5.16. Road extraction result on this large image obtained using our models, at full resolution.

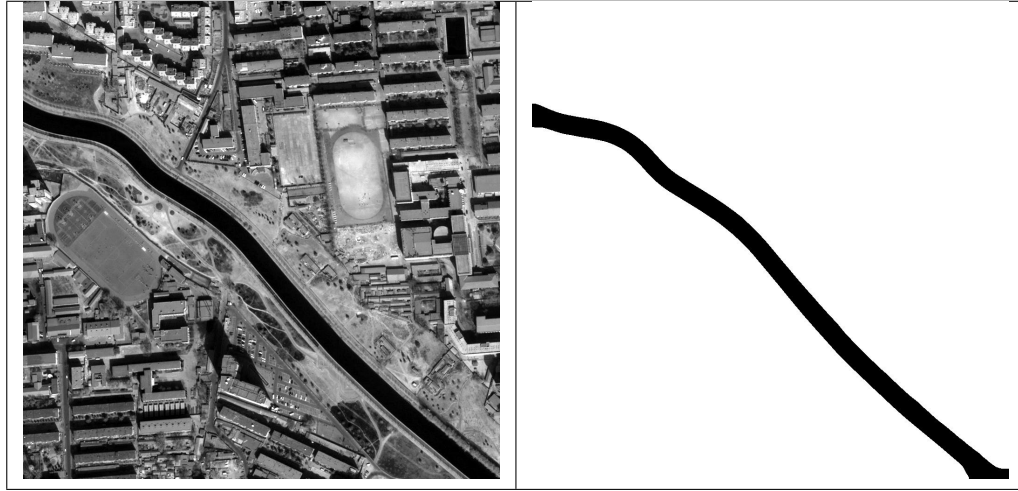


Figure 5.17. Result of river extraction on a QuickBird panchromatic image (size: 1024×1024).

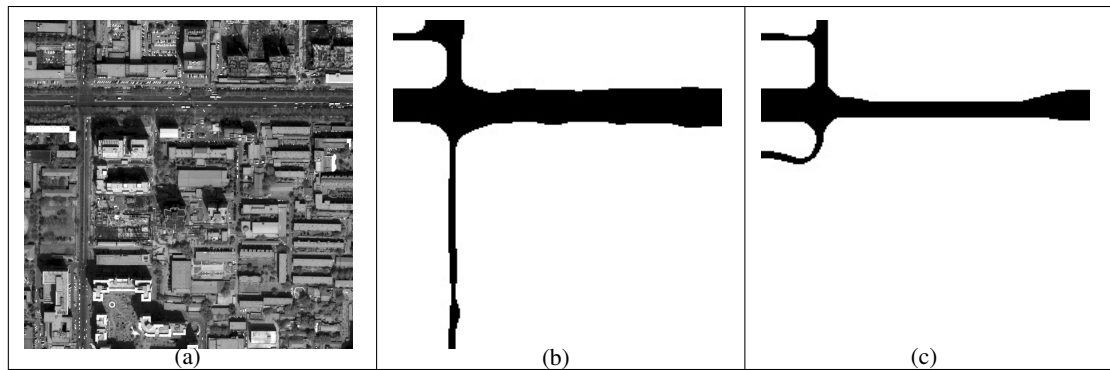


Figure 5.18. Extraction of a road network containing two different widths, at $1/4$ resolution. (a)-(c): image data; results using: the new linear model E ; the primary model E_{primary} ($\beta_2 = 0$).

Measure Method	Completeness TP/(TP+FN)	Correctness TP/(TP+FP)	Quality TP/(TP+FP+FN)
New model E (with E_L) at 1/4 resolution (Figure 5.12(b))	0.9688	0.8519	0.8292
New model E (with E_{NL}) at 1/4 resolution (Figure 5.9(d))	0.9524	0.8591	0.8237
New model E (with E_L) at full resolution (Figure 5.12(c))	0.8756	0.9693	0.8520
MLE (Figure 5.12(d))	0.9356	0.2073	0.2044
$\theta E_0 + E_D$ (Figure 5.12(e))	0.6047	0.8249	0.5359
$\theta(E_0 + E_S) + E_D$ (Figure 5.12(f))	0.6946	0.9889	0.6892
Wang (Wang and Zhang, 2003) (Figure 5.13(a))	0.9350	0.3463	0.3381
Yu (Yu et al., 2004) (Figure 5.13(b))	0.6050	0.3695	0.2977

Table 5.2. Quantitative criteria tested on Figure 5.12(a) at full resolution (except first row) (T = True, F = False, P = Positive, N = Negative). See Table 5.1 for an explanation of completeness, correctness and quality.

Extraction of Roads of Different Widths

Images containing roads of different widths are processed after choosing parameter values for which e_{PL} has two local minima. Figure 5.18(a) shows an input image containing two roads: their widths are approximately 20 pixels and 80 pixels. The results obtained using the new linear model (with E_L) and the primary model E_{primary} (without E_L), at 1/4 resolution, are illustrated in Figures 5.18(b) and 5.18(c) respectively. The parameter values $(\theta, \theta_v, \alpha, \lambda, \beta, \beta_2, d, d_2)$ used in this experiment are $(25, 0, 0.15, 5, 0.02, 1.228 \times 10^{-4}, 4, 22)$. The estimated stable widths for these parameter values are 5.28 and 20.68, corresponding to the road widths at 1/4 resolution, *i.e.* 5 pixels and 20 pixels. This comparison shows clearly that adding E_L enables the detection of roads with both widths, while the primary model without E_L finds only an incomplete network.

5.5 Conclusion

Narrow secondary roads in VHR images are very difficult to extract, because of occlusion effects and the similar radiometric properties of the road region and background. In particular, the standard HOAC prior term has a severe limitation: network branch width is constrained to be similar to maximum network branch radius of curvature, thereby providing a poor model of networks with straight narrow branches or highly curved, wide branches. To deal with this problem, the incorporation of strong geometric prior knowledge of road networks is essential. Building upon our primary model, we have first presented, in this chapter, a novel *nonlinear nonlocal* phase field term. This novel term causes pairs of points inside the range of the interaction to attract each other. In conjunction with the original HOAC geometric term, it allows the interaction between points on the same side of a network branch to be stronger than the interaction between points on opposite sides of a network branch. Therefore, the incorporation of the term enables the generation of longer arm-like branches and better prolongation. Subsequently, we have proposed another novel *linear nonlocal* HOAC term for modeling bar shape and embedded it in the phase field framework. Thus, the prolongation of the network branch is controlled by this new linear term, which includes a longer-range interaction. Based on a stability analysis of a bar with a desired width, we established constraints linking the parameters of the energy function. We explored the possible behaviors of the resulting prior energy as a function of the parameter settings, and showed that as well as separating the interactions between points on the same and opposite sides of a network branch, the new linear model permits the modeling of two widths simultaneously. The analysis also fixes some of the model parameters in terms of network width(s). Moreover, due to the linearity, the linear nonlocal term is more efficient from a computational point of view, and can be applied to images at full resolution. For this reason, in the experiments on road network extraction from VHR satellite images, some thinner networks can also be extracted, and in general, the extraction accuracy has been improved, with the linear nonlocal term at full resolution.

Conclusion

In this chapter, we summarize the goal and the proposed models of this thesis. We also point out some possible ways to improve our models in future work.

Summary

The goal of this thesis was to develop new variational models for the segmentation from an image of entities that have the form of a ‘network’, *i.e.* branches joining together at junctions. More particularly, we focused on incorporating different types of geometric prior on network regions, while taking advantage of a multiresolution analysis of the image. We applied our different models to segment road networks from very high resolution (VHR) QuickBird panchromatic images ($\sim 0.61\text{m/pixel}$) in dense urban areas. This is a hard problem due to the complexity existing in the data, and the complexity of modeling network regions with arbitrary topology.

We first started with a primary model in a phase field formulation. It was composed of a standard higher-order active contour (HOAC) prior model and a region-based likelihood model. Thanks to long-range interactions between pixels of the HOAC, the model enabled the inclusion of sophisticated prior knowledge of region geometry. Phase fields also provided many advantages over other conventional methods for region modeling. To ease the difficulties stemming from the complexity of the image scene at high resolution, we introduced a multiresolution statistical data model and a multiresolution constraint prior model. These two models allowed the effective integration of image information from different resolutions. Multiresolution analysis can greatly increase the robustness of the algorithm. Subsequently, in the context of map updating, we included a specific shape prior for the region, derived from a Geographical Information System (GIS) map, and combined it with the other more generic priors. GIS information, though outdated, can provide partly correct information about the considered scene, and eliminate false detections in the background. We showed that our model was able to improve the accuracy of unchanged objects, to extract newly appeared objects, and to remove objects that had disappeared from the map.

To facilitate the extraction of thinner elongated structures, we proposed two new HOAC prior models. Both of them achieved a similar effect on shape modeling, that is, to control independently the straightness and the width of the road, in conjunction with the standard

HOAC term. Therefore, the incorporation of these two terms enabled the generation of longer arm-like branches and better prolongation. Moreover, the linear term had several advantages over the other nonlinear one: it was more efficient from a computational standpoint, and it was able to model multiple widths simultaneously. In each HOAC total prior model, there were a number of control parameters. It was of great importance to establish the internal constraints among them, so as to fix some in terms of the others. For these prior models, within the phase field representation, we calculated stability conditions, and indicated the related constraints. We showed that the relationships between the control parameters and energy behavior followed some special cases of catastrophe theory.

We tested and evaluated the proposed models on QuickBird images of Beijing, and compared them to several other techniques in the literature. The experimental results and comparisons demonstrate the superiority of our models.

Perspectives

Of course, there are still many possibilities to improve the work in this thesis. Here we list some perspectives:

1. From the stability analysis of the linear nonlocal HOAC total prior model, we showed that the model's behavior depends on the three scaled control parameters $\hat{\beta}$, $\hat{\beta}_2$ and \hat{d}_2 . This is an example of a swallowtail catastrophe. Indeed, with the parameter settings in one of its swallowtail corners, a wide flat minimum can be generated, which seems to be an efficient way to extract branches of multiple widths. However, with the current interaction function with finite support, this effective swallowtail corner has an infinite value, therefore impossible to use in practice. Alternatively, in Appendix D, we started to investigate the feasibility of bringing it back from infinity to a finite value, by changing to another type of interaction function with infinite support. Furthermore, the condition of being at that right corner enables $\hat{\beta}$ and $\hat{\beta}_2$ to be fixed. Only \hat{d}_2 is left as an adjustable parameter. Besides the exponential function used in Appendix D, we may exploit other types of interaction function, if it is monotonically decreasing to zero at infinity.
2. In our last linear nonlocal HOAC prior model, the linear nonlocal term E_L aims to control the longer-range interactions along the network branch, (and of course, when modeling two different widths, E_L is responsible for the larger width,) while the standard term E_S still controls the interactions (with a shorter range) in both longitudinal and transverse directions. A possible replacement of E_S , E_T , can be restricted to have an effect solely in the transverse direction. Following a similar idea to E_L , but changing the cross product to the dot product, we can write E_T as

$$E_T(\phi) = -\frac{\beta}{2} \iint_{\Omega^2} (\nabla\phi(x) \cdot (x - x')) (\nabla\phi(x') \cdot (x - x')) \Psi\left(\frac{|x - x'|}{d}\right) dx dx'.$$

So the prior model writes $E_{P,L} = E_0 + E_T + E_L$, which provides a complete separation of interactions.

3. In our current data model, we only use a simple region-based likelihood data energy based on the image intensity and/or the variance. Gradient information and Haar wavelet coefficients are too noisy to be useful for segmentation in dense urban environments at this resolution. In the future, we may integrate other cues, such as color, texture, and wavelet coefficients with other wavelet kernels into the data model. This would be useful for the extraction of large roads. Another possibility is a more sophisticated data model including interactions between pixels.
4. Due to the high computational cost, it is very important to improve the computational efficiency. We could optimize the program in several different ways. Some sparse matrix operators could be introduced to avoid the calculation of integrals. Another potentially powerful numerical technique to solve the phase field equations is to use the adaptive grid method (Provatas et al., 1998).
5. We may express the phase field prior model in a suitable wavelet basis, and develop the multiscale structure of the prior model. On the one hand, the relationship among the same parameters at different resolutions may be established. On the other hand, the computational time may be reduced, in terms of a fast convergence.
6. Last but not least, we can think about cooperation among the different models we have proposed. Introducing other appropriate prior knowledge is a promising direction to increase robustness of the algorithm. In the future, we could apply our models to VHR data of other cities or to other applications.

Appendix A

Stability Calculations

In this appendix, we detail the calculation of the stability analysis of a long straight bar, of length L and width $W \ll L \rightarrow \infty$ (see Figure 2.5). The *ansatz* for ϕ_R for such a bar is defined as follows: the phase field is given by $\phi(x) = 1$ for $x \in R \setminus R_C$; $\phi(x) = -1$ for $x \in \bar{R} \setminus R_C$, while in R_C , ϕ changes linearly from 1 to -1 . Its expression using Cartesian coordinates is presented in equation (2.9).

A.1 Energy Terms of a Bar

A.1.1 Basic Phase Field Term

For the *basic* phase field term E_0 (see equation (2.6)):

$$E_0(\phi) = \int_{\Omega} \left\{ \frac{1}{2} \nabla \phi(x) \cdot \nabla \phi(x) + \lambda \left(\frac{1}{4} \phi^4(x) - \frac{1}{2} \phi^2(x) \right) + \alpha \left(\phi(x) - \frac{1}{3} \phi^3(x) \right) \right\} dx, \quad (\text{A.1})$$

the image domain Ω decomposes into the region R , the region \bar{R} and the interface R_C . The integration gives

$$E_{0,R} = \left(-\frac{1}{4} \lambda + \frac{2}{3} \alpha \right) (WL - wL), \quad (\text{A.2a})$$

$$E_{0,\bar{R}} = \left(-\frac{1}{4} \lambda - \frac{2}{3} \alpha \right) (A_0 - WL - wL), \quad (\text{A.2b})$$

$$\begin{aligned} E_{0,R_C} &= 2 \int_{-\infty}^{\infty} \left\{ \int_0^w \left\{ \frac{1}{2} \frac{4}{w^2} + \lambda \left(\frac{1}{4} \left(\frac{2}{w} v - 1 \right)^4 - \frac{1}{2} \left(\frac{2}{w} v - 1 \right)^2 \right) \right. \right. \\ &\quad \left. \left. + \alpha \left(\left(\frac{2}{w} v - 1 \right) - \frac{1}{3} \left(\frac{2}{w} v - 1 \right)^3 \right) \right\} dv \right\} du \\ &= 2 \int_{-\infty}^{\infty} \left(\frac{2}{w} - \frac{7}{60} \lambda w \right) du \\ &= \frac{4L}{w} - \frac{7}{30} \lambda w L, \end{aligned} \quad (\text{A.2c})$$

where A_0 is the area of Ω .

A.1.2 Standard Phase Field HOAC Term

Since $\nabla\phi$ is zero in the regions R and \bar{R} , the double integral of the *standard* phase field HOAC term E_S (see equation (2.8)):

$$E_S(\phi) = -\frac{\beta}{2} \iint_{\Omega^2} \nabla\phi(x) \cdot \nabla\phi(x') \Psi\left(\frac{|x-x'|}{d}\right) dx dx', \quad (\text{A.3})$$

needs to be calculated only across the interface R_C . If we assume that Ψ is roughly constant over distances w , each integral simply contributes a factor w , which cancels w in the corresponding gradient $-(2/w)\hat{n}$ (\hat{n} is the outward unit normal vector to the boundary). The energy becomes

$$\begin{aligned} E_S &= -\frac{\beta}{2} \iint_{\Omega^2} \frac{4}{w^2} \hat{n}(x) \cdot \hat{n}(x') \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= -\frac{4\beta}{w^2} \left\{ \iint_{R_{CB} \times R_{CB}} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' - \iint_{R_{CB} \times R_{CT}} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \right\} \\ &= -4\beta \left\{ \iint_{-\infty}^{\infty} \Psi\left(\frac{|u-u'|}{d}\right) du du' - \iint_{-\infty}^{\infty} \Psi\left(\frac{\sqrt{W^2 + (u-u')^2}}{d}\right) du du' \right\}. \end{aligned} \quad (\text{A.4})$$

Putting $z = u - u'$, we then have

$$E_S = -4\beta L \left\{ \int_{-\infty}^{\infty} \Psi\left(\frac{|z|}{d}\right) dz - \int_{-\infty}^{\infty} \Psi\left(\frac{\sqrt{W^2 + z^2}}{d}\right) dz \right\}. \quad (\text{A.5})$$

Now the integrals are symmetric in z , so we can reduce the range of the integrals to positive values. This enables us to remove the absolute value in the first integral. Having done this, we can introduce the variable $\eta^2 = (z^2 + W^2)/d^2$ for the second integral, and again we have no need of an absolute value. Noting that $dz = d\eta \frac{\eta d}{\sqrt{\eta^2 - \hat{W}^2}}$, we have the expression in equation (2.14):

$$E_S = -8\beta L \left\{ \int_0^{\infty} \Psi\left(\frac{z}{d}\right) dz - d \int_{\hat{W}}^{\infty} \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \Psi(\eta) d\eta \right\}. \quad (\text{A.6})$$

A.1.3 Nonlinear Nonlocal Phase Field Term

The *nonlinear nonlocal* phase field term E_{NL} is (see equation (5.8))

$$E_{NL}(\phi) = -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x)) (\nabla\phi(x') \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx'. \quad (\text{A.7})$$

Following a calculation similar to that for E_S , we have

$$\begin{aligned}
E_{NL}(\phi) &= -\frac{\beta_2}{4} \iint_{\Omega^2} \left(\frac{4}{w^2} \hat{\mathbf{n}}(x) \cdot \hat{\mathbf{n}}(x) \right) \left(\frac{4}{w^2} \hat{\mathbf{n}}(x') \cdot \hat{\mathbf{n}}(x') \right) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\
&= -\frac{8\beta_2}{w^4} \left\{ \iint_{R_{CB} \times R_{CB}} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' + \iint_{R_{CB} \times R_{CT}} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \right\} \\
&= -\frac{8\beta_2}{w^2} \left\{ \iint_{-\infty}^{\infty} \Psi\left(\frac{|u-u'|}{d}\right) du du' + \iint_{-\infty}^{\infty} \Psi\left(\frac{\sqrt{W^2 + (u-u')^2}}{d}\right) du du' \right\} \\
&= -\frac{8\beta_2 L}{w^2} \left\{ \int_{-\infty}^{\infty} \Psi\left(\frac{|z|}{d}\right) dz + \int_{-\infty}^{\infty} \Psi\left(\frac{\sqrt{W^2 + z^2}}{d}\right) dz \right\} \\
&= -\frac{16\beta_2 L}{w^2} \left\{ \int_0^{\infty} \Psi\left(\frac{z}{d}\right) dz + d \int_{\hat{W}}^{\infty} \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \Psi(\eta) d\eta \right\}. \tag{A.8}
\end{aligned}$$

This is equation (5.11).

A.1.4 Linear Nonlocal Phase Field Term

The *linear nonlocal* phase field term E_L is (see equation (5.21))

$$E_L(\phi) = -\frac{\beta_2}{2} \iint_{\Omega^2} (\nabla\phi(x) \times (x-x')) (\nabla\phi(x') \times (x-x')) \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx'. \tag{A.9}$$

where \times is the 2D vectorial antisymmetric product. Evaluating it on the *ansatz* for a long straight bar, we have ($\hat{\mathbf{t}}$ is the unit tangent vector)

$$\begin{aligned}
E_L(\phi) &= -\frac{\beta_2}{2} \iint_{\Omega^2} \left(\frac{2}{w} (x-x') \hat{\mathbf{t}}(x) \right) \left(\frac{2}{w} (x-x') \hat{\mathbf{t}}(x') \right) \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx' \\
&= -\frac{4\beta_2}{w^2} \left\{ \iint_{R_{CB} \times R_{CB}} \left(\frac{2}{w} (x-x') \hat{\mathbf{t}}(x) \right)^2 \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx' \right. \\
&\quad \left. - \iint_{R_{CB} \times R_{CT}} \left(\frac{2}{w} (x-x') \hat{\mathbf{t}}(x) \right)^2 \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx' \right\} \\
&= -4\beta_2 \left\{ \iint_{-\infty}^{\infty} (u-u')^2 \Psi\left(\frac{|u-u'|}{d_2}\right) du du' - \iint_{-\infty}^{\infty} (u-u')^2 \Psi\left(\frac{\sqrt{W^2 + (u-u')^2}}{d_2}\right) du du' \right\} \\
&= -4\beta_2 L \left\{ \int_{-\infty}^{\infty} z^2 \Psi\left(\frac{|z|}{d_2}\right) dz - \int_{-\infty}^{\infty} z^2 \Psi\left(\frac{\sqrt{W^2 + z^2}}{d_2}\right) dz \right\}. \tag{A.10}
\end{aligned}$$

To simplify the integral, the variable $\eta^2 = (z^2 + W^2)/d^2$ is introduced. Here the denominator still includes d , instead of d_2 . Noting that $\hat{d}_2 = d_2/d$ is the ratio between two interaction ranges, we obtain the result in equation (5.24):

$$E_L(\phi) = -8\beta_2 L \left\{ \int_0^{\infty} z^2 \Psi\left(\frac{z}{d_2}\right) dz - d^3 \int_{\hat{W}}^{\infty} \eta \sqrt{\eta^2 - \hat{W}^2} \Psi\left(\frac{\eta}{\hat{d}_2}\right) d\eta \right\}. \tag{A.11}$$

A.2 Model Energy per Unit Length

A.2.1 Standard HOAC Total Prior Model

The *standard* HOAC total prior energy, per unit length of bar, $e_{P,S}$, is (see equation (2.16))

$$\begin{aligned} e_{P,S}(w, \hat{W}) &= e_0 + e_S \\ &= \frac{4}{3}\alpha\hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + 4\beta d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta. \end{aligned} \quad (\text{A.12})$$

The first derivatives are

$$\frac{\partial e_{P,S}}{\partial \hat{W}} = \frac{4}{3}\alpha d - 4\beta\hat{W}d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta, \quad (\text{A.13a})$$

$$\frac{\partial e_{P,S}}{\partial w} = \frac{4}{15}\lambda - \frac{4}{w^2}. \quad (\text{A.13b})$$

The second derivative with respect to w is always positive. Noting

$$I_1(\hat{W}) = \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta, \quad (\text{A.14})$$

to obtain the second derivative with respect to \hat{W} , we have to calculate $I_1'(\hat{W})$ (Here ' denotes $\partial I_1 / \partial \hat{W}$):

$$\begin{aligned} I_1'(\hat{W}) &= \int_{\hat{W}}^2 \frac{\hat{W}}{\sqrt{(\eta^2 - \hat{W}^2)^3}} (1 - \cos(\pi\eta)) d\eta - \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) \Big|_{\eta=\hat{W}} \\ &= - \frac{\eta}{\hat{W} \sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) \Big|_{\eta=\hat{W}}^2 + \int_{\hat{W}}^2 \frac{\pi\eta}{\hat{W} \sqrt{\eta^2 - \hat{W}^2}} \sin(\pi\eta) d\eta \\ &\quad - \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) \Big|_{\eta=\hat{W}} \\ &= \frac{\pi}{\hat{W}} \int_{\hat{W}}^2 \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \sin(\pi\eta) d\eta \\ &= - \frac{\pi^2}{\hat{W}} \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} \cos(\pi\eta) d\eta. \end{aligned} \quad (\text{A.15})$$

Using this result, we have the expression in equation (2.18):

$$\begin{aligned}
\frac{\partial^2 e_{P,S}}{\partial \hat{W}^2} &= -4\beta d I_1(\hat{W}) - 4\beta \hat{W} d I_1'(\hat{W}) \\
&= -4\beta d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta + 4\beta \pi^2 d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} \cos(\pi\eta) d\eta \\
&= -4\beta d \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right) + 4\beta d \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta . \quad (A.16)
\end{aligned}$$

A.2.2 Nonlinear Nonlocal HOAC Total Prior Model

The *nonlinear nonlocal* HOAC total prior energy, per unit length of bar, $e_{P,NL}$, is (see equation (5.13))

$$\begin{aligned}
e_{P,NL}(w, \hat{W}) &= e_0 + e_S + e_{NL} \\
&= \frac{4}{3} \alpha \hat{W} d + \frac{4}{15} \lambda w + \frac{4}{w} - \frac{16\beta_2 d}{w^2} + 4d(\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta . \quad (A.17)
\end{aligned}$$

The first derivatives are

$$\frac{\partial e_{P,NL}}{\partial \hat{W}} = \frac{4}{3} \alpha d - 4\hat{W} d (\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta , \quad (A.18a)$$

$$\frac{\partial e_{P,NL}}{\partial w} = \frac{4}{15} \lambda - \frac{4}{w^2} + \frac{16\beta_2 d}{w^3} \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta . \quad (A.18b)$$

Since we decide to fix the value w beforehand with the aim of simplifying the problem, there is no need to calculate the second derivative with respect to w . On the other hand, the second derivative with respect to \hat{W} has to be checked. Using the results of $I_1(\hat{W})$ and $I_1'(\hat{W})$ in equations (A.14) and (A.15), we obtain the formula in equation (5.15):

$$\begin{aligned}
\frac{\partial^2 e_{P,NL}}{\partial \hat{W}^2} &= -4d(\beta - \frac{2\beta_2}{w^2}) I_1(\hat{W}) - 4\hat{W} d (\beta - \frac{2\beta_2}{w^2}) I_1'(\hat{W}) \\
&= -4d(\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta \\
&\quad + 4\pi^2 d (\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} \cos(\pi\eta) d\eta \\
&= -4d(\beta - \frac{2\beta_2}{w^2}) \ln \left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}} \right) + 4d(\beta - \frac{2\beta_2}{w^2}) \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta . \quad (A.19)
\end{aligned}$$

A.2.3 Linear Nonlocal HOAC Total Prior Model

The *linear nonlocal* HOAC total prior energy, per unit length of bar, $e_{P,L}$, is (see equation (5.26))

$$\begin{aligned}
 e_{P,L}(w, \hat{W}) &= e_0 + e_S + e_L \\
 &= \frac{4}{3}\alpha \hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + 4\beta d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} (1 - \cos(\pi\eta)) d\eta \\
 &\quad + 4\beta_2 d^3 \int_{\hat{W}}^{2\hat{d}_2} \eta \sqrt{\eta^2 - \hat{W}^2} \left(2 - \frac{\eta}{\hat{d}_2} + \frac{1}{\pi} \sin\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta . \tag{A.20}
 \end{aligned}$$

The first derivatives are calculated as follows:

$$\begin{aligned}
 \frac{\partial e_{P,L}}{\partial \hat{W}} &= \frac{4}{3}\alpha d - 4\beta \hat{W}d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta \\
 &\quad - 4\beta_2 \hat{W}d^3 \int_{\hat{W}}^{2\hat{d}_2} \frac{\eta}{\sqrt{\eta^2 - \hat{W}^2}} \left(2 - \frac{\eta}{\hat{d}_2} + \frac{1}{\pi} \sin\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta \\
 &= \frac{4}{3}\alpha d - 4\beta \hat{W}d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta \\
 &\quad - \frac{4\beta_2 \hat{W}d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \sqrt{\eta^2 - \hat{W}^2} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta , \tag{A.21a}
 \end{aligned}$$

$$\frac{\partial e_{P,L}}{\partial w} = \frac{4}{15}\lambda - \frac{4}{w^2} . \tag{A.21b}$$

Now, we calculate the second derivative with respect to \hat{W} . Noting

$$I_3(\hat{W}) = \int_{\hat{W}}^{2\hat{d}_2} \sqrt{\eta^2 - \hat{W}^2} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta , \tag{A.22}$$

we first calculate $I'_3(\hat{W})$, i.e. $\partial I_3 / \partial \hat{W}$:

$$I'_3(\hat{W}) = -\hat{W} \int_{\hat{W}}^{2\hat{d}_2} \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta . \tag{A.23}$$

So we have the result in equation (5.27c):

$$\begin{aligned}
\frac{\partial^2 e_{P,L}}{\partial \hat{W}^2} &= -4\beta d I_1 - 4\beta \hat{W} d I'_1(\hat{W}) - \frac{4\beta_2 d^3}{\hat{d}_2} I_3(\hat{W}) - \frac{4\beta_2 \hat{W} d^3}{\hat{d}_2} I'_3(\hat{W}) \\
&= -4\beta d \int_{\hat{W}}^2 \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} (1 - \cos(\pi\eta)) d\eta + 4\beta \pi^2 d \int_{\hat{W}}^2 \sqrt{\eta^2 - \hat{W}^2} \cos(\pi\eta) d\eta \\
&\quad - \frac{4\beta_2 d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \sqrt{\eta^2 - \hat{W}^2} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta \\
&\quad + \frac{4\beta_2 \hat{W}^2 d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta \\
&= -4\beta d \ln\left(\frac{2 + \sqrt{4 - \hat{W}^2}}{\hat{W}}\right) + 4\beta d \int_{\hat{W}}^2 \frac{1 + \pi^2(\eta^2 - \hat{W}^2)}{\sqrt{\eta^2 - \hat{W}^2}} \cos(\pi\eta) d\eta \\
&\quad - \frac{4\beta_2 d^3}{\hat{d}_2} \int_{\hat{W}}^{2\hat{d}_2} \frac{\eta^2 - 2\hat{W}^2}{\sqrt{\eta^2 - \hat{W}^2}} \left(1 - \cos\left(\frac{\pi\eta}{\hat{d}_2}\right)\right) d\eta .
\end{aligned} \tag{A.24}$$

Appendix B

Evolution Equations of New HOAC Prior Energies

In this appendix, we detail the calculation of the derivatives of the two new *nonlocal* HOAC prior energies proposed in chapter 5. The corresponding evolution equation is the negative of the functional derivative.

B.1 Nonlinear Nonlocal HOAC Prior Energy

The *nonlinear nonlocal* HOAC prior energy we propose in equation (5.8) is

$$E_{NL}(\phi) = -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x)) (\nabla\phi(x') \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx', \quad (\text{B.1})$$

where ϕ is the phase field function, and $x, x' \in \Omega$.

We calculate the energy with a small variation of the phase field $\delta\phi$ (neglecting second order terms in $\delta\phi$):

$$\begin{aligned} E_{NL}(\phi + \delta\phi) &= -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x) + 2\nabla\phi(x) \cdot \nabla\delta\phi(x)) \\ &\quad (\nabla\phi(x') \cdot \nabla\phi(x') + 2\nabla\phi(x') \cdot \nabla\delta\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= E_{NL}(\phi) - \frac{\beta_2}{2} \iint_{\Omega^2} \left\{ (\nabla\phi(x) \cdot \nabla\phi(x)) (\nabla\phi(x') \cdot \nabla\delta\phi(x')) \right. \\ &\quad \left. + (\nabla\phi(x') \cdot \nabla\phi(x')) (\nabla\phi(x) \cdot \nabla\delta\phi(x)) \right\} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= E_{NL}(\phi) - \beta_2 \iint_{\Omega^2} (\nabla\phi(x') \cdot \nabla\phi(x')) (\nabla\phi(x) \cdot \nabla\delta\phi(x)) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= E_{NL}(\phi) + \beta_2 \int_{\Omega} \delta\phi(x) \nabla \cdot \left\{ \nabla\phi(x) \int_{\Omega} (\nabla\phi(x') \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx' \right\} dx, \quad (\text{B.2}) \end{aligned}$$

so we obtain the energy derivative in equation (5.9):

$$\begin{aligned} \frac{\delta E_{NL}(\phi)}{\delta \phi(x)} &= \beta_2 \nabla \cdot \left\{ \nabla \phi(x) \int_{\Omega} (\nabla \phi(x') \cdot \nabla \phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx' \right\} \\ &= \beta_2 \int_{\Omega} \left\{ (\nabla \phi(x') \cdot \nabla \phi(x')) \nabla^2 \phi(x) \Psi\left(\frac{|x-x'|}{d}\right) \right. \\ &\quad \left. + (\nabla \phi(x') \cdot \nabla \phi(x')) \left(\nabla \phi(x) \cdot \nabla \Psi\left(\frac{|x-x'|}{d}\right) \right) \right\} dx'. \end{aligned} \quad (\text{B.3})$$

This formula in the Fourier domain is

$$\begin{aligned} \mathcal{F}\left\{\frac{\delta E_{NL}(\phi)}{\delta \phi(x)}\right\} &= \beta_2 ik \cdot \mathcal{F}\left\{\nabla \phi(x) \left(\Psi\left(\frac{|x|}{d}\right) * (\nabla \phi(x) \cdot \nabla \phi(x))\right)\right\} \\ &= \beta_2 ik \cdot \left\{ ik \hat{\phi}(k) * \mathcal{F}\left\{\Psi\left(\frac{|x|}{d}\right) * (\nabla \phi(x) \cdot \nabla \phi(x))\right\} \right\} \\ &= \beta_2 (-k^2 \hat{\phi}(k)) * \left(d \hat{\Psi}(kd) \mathcal{F}\{\nabla \phi(x) \cdot \nabla \phi(x)\} \right) \\ &\quad + \beta_2 (ik \hat{\phi}(k)) * \left(ikd \hat{\Psi}(kd) \mathcal{F}\{\nabla \phi(x) \cdot \nabla \phi(x)\} \right), \end{aligned} \quad (\text{B.4})$$

where $*$ is convolution. \mathcal{F} and \mathcal{F}^{-1} denote the Fourier and the inverse Fourier transform respectively, and a hat $\hat{\cdot}$ indicates the Fourier transform of a variable. After the inverse Fourier transform, the energy derivative is given by

$$\begin{aligned} \frac{\delta E_{NL}(\phi)}{\delta \phi(x)} &= \beta_2 \nabla^2 \phi(x) \mathcal{F}^{-1}\left\{ d \hat{\Psi}(kd) \mathcal{F}\{\nabla \phi(x) \cdot \nabla \phi(x)\} \right\} \\ &\quad + \beta_2 \nabla \phi \cdot \nabla \left(\mathcal{F}^{-1}\left\{ d \hat{\Psi}(kd) \mathcal{F}\{\nabla \phi(x) \cdot \nabla \phi(x)\} \right\} \right). \end{aligned} \quad (\text{B.5})$$

Finally we have the evolution equation in equation (5.10).

B.2 Linear Nonlocal HOAC Prior Energy

The *linear nonlocal* HOAC prior term E_L proposed in equation (5.21) is

$$E_L(\phi) = -\frac{\beta_2}{2} \iint_{\Omega^2} [\nabla \phi(x) \times (x - x')] [\nabla \phi(x') \times (x - x')] \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx'. \quad (\text{B.6})$$

To get rid of the cross products, we rewrite it as

$$E_L(\phi) = -\frac{\beta_2}{2} \iint_{\Omega^2} \nabla \phi(x) \cdot [\epsilon(x - x')(x - x')^T \epsilon^T] \cdot \nabla \phi(x') \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx', \quad (\text{B.7})$$

where T denotes the transpose, and ϵ rotates the tangent vectors to the inward normal vectors.

We calculate the energy with a small variation of the phase field $\delta\phi$:

$$\begin{aligned}
E_L(\phi + \delta\phi) &= -\frac{\beta_2}{2} \iint_{\Omega^2} (\nabla\phi(x) + \nabla\delta\phi(x)) \cdot [\epsilon(x-x')(x-x')^T \epsilon^T] \cdot (\nabla\phi(x') \\
&\quad + \nabla\delta\phi(x')) \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx' \\
&= E_L(\phi) - \beta_2 \iint_{\Omega^2} \nabla\delta\phi(x) \cdot [\epsilon(x-x')(x-x')^T \epsilon^T] \cdot \nabla\phi(x') \Psi\left(\frac{|x-x'|}{d_2}\right) dx dx' \\
&= E_L(\phi) + \beta_2 \int_{\Omega} \left\{ \delta\phi(x) \int_{\Omega} \nabla \cdot [\epsilon(x-x')(x-x')^T \epsilon^T] \cdot \nabla\phi(x') \Psi\left(\frac{|x-x'|}{d_2}\right) dx' \right\} dx. \quad (B.8)
\end{aligned}$$

The functional derivative of E_L is

$$\frac{\delta E_L(\phi)}{\delta\phi(x)} = \beta_2 \int_{\Omega} \nabla \cdot [\epsilon(x-x')(x-x')^T \epsilon^T] \cdot \nabla\phi(x') \Psi\left(\frac{|x-x'|}{d_2}\right) dx', \quad (B.9)$$

as shown in equation (5.22).

To avoid calculating the convolution, we apply the Fourier transform to the above formula. It becomes

$$\begin{aligned}
\mathcal{F}\left\{\frac{\delta E_L(\phi)}{\delta\phi(x)}\right\} &= \beta_2 \mathcal{F}\left\{\left(\nabla \cdot \left[\Psi\left(\frac{|x|}{d_2}\right) \epsilon x x^T \epsilon^T\right]\right) * \nabla\phi(x)\right\} \\
&= \beta_2 i k \cdot \mathcal{F}\left\{\Psi\left(\frac{|x|}{d_2}\right) \epsilon x x^T \epsilon^T\right\} \cdot i k \hat{\phi}(k). \quad (B.10)
\end{aligned}$$

We replace x in equation (B.10) by its polar coordinates, *i.e.* $x = (r \cos \theta, r \sin \theta)^T$. Thus, $\epsilon x = (r \sin \theta, -r \cos \theta)^T$. Equation (B.10) becomes

$$\begin{aligned}
\mathcal{F}\left\{\frac{\delta E_L(\phi)}{\delta\phi(x)}\right\} &= -\beta_2 k \cdot \mathcal{F}\left\{\begin{pmatrix} r^2 \sin^2 \theta \Psi(r/d_2) & -r^2 \sin \theta \cos \theta \Psi(r/d_2) \\ -r^2 \sin \theta \cos \theta \Psi(r/d_2) & r^2 \cos^2 \theta \Psi(r/d_2) \end{pmatrix}\right\} \cdot k \hat{\phi}(k) \\
&= -\beta_2 \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \cdot \begin{pmatrix} \mathcal{F}\{r^2 \sin^2 \theta \Psi(r/d_2)\} & \mathcal{F}\{-r^2 \sin \theta \cos \theta \Psi(r/d_2)\} \\ \mathcal{F}\{-r^2 \sin \theta \cos \theta \Psi(r/d_2)\} & \mathcal{F}\{r^2 \cos^2 \theta \Psi(r/d_2)\} \end{pmatrix} \cdot \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \hat{\phi}(k) \\
&= -\beta_2 (k_1^2 \mathcal{F}\{r^2 \sin^2 \theta \Psi(r/d_2)\} - 2k_1 k_2 \mathcal{F}\{r^2 \sin \theta \cos \theta \Psi(r/d_2)\} \\
&\quad + k_2^2 \mathcal{F}\{r^2 \cos^2 \theta \Psi(r/d_2)\}) \hat{\phi}(k). \quad (B.11)
\end{aligned}$$

After the inverse Fourier transform, the energy derivative is

$$\begin{aligned}
\frac{\delta E_L(\phi)}{\delta\phi(x)} &= -\beta_2 \mathcal{F}^{-1}\left\{\left(k_1^2 \mathcal{F}\{r^2 \sin^2 \theta \Psi(r/d_2)\} - 2k_1 k_2 \mathcal{F}\{r^2 \sin \theta \cos \theta \Psi(r/d_2)\} \right. \right. \\
&\quad \left. \left. + k_2^2 \mathcal{F}\{r^2 \cos^2 \theta \Psi(r/d_2)\}\right)\right\} \hat{\phi}(k). \quad (B.12)
\end{aligned}$$

So we get the evolution equation in equation (5.23).

Appendix C

Another Nonlinear Nonlocal HOAC Prior Term

In this appendix, we describe another *nonlinear nonlocal* HOAC prior term \tilde{E}_{NL} , an alternative to E_{NL} , which we do not use. We explain the reason after describing the energy.

C.1 Definition of \tilde{E}_{NL}

The general HOAC prior energy E_{HO} in the contour formulation, which involves the interpolation between interaction functions, is given in equation (5.3):

$$E_{HO}(\gamma) = - \iint_{S^1 \times S^1} \left\{ f_{+||}(\dot{\gamma}(s) \cdot \dot{\gamma}(s')) \Psi_{+||} - f_{-||}(\dot{\gamma}(s) \cdot \dot{\gamma}(s')) \Psi_{-||} \right\} ds ds', \quad (C.1)$$

where $\gamma : S^1 \rightarrow \Omega$, is an arc length parameterization of the region boundary ∂R ; $\dot{\gamma}(s)$ is the tangent vector to the boundary at s (thus $\dot{\gamma}(s) \cdot \dot{\gamma}(s') \in [-1, 1]$); ‘+ ||’ denotes parallel vectors and ‘- ||’ denotes antiparallel vectors.

Different from the discussion in subsection 5.2.1, we assume that $\Psi_{+||} = \Psi_{-||}$, and change the definitions of $f_{+||}(x)$ and $f_{-||}(x)$. In principle, $f_{+||}(x) : [-1, 1] \rightarrow [0, 1]$ has to be a non-decreasing function with the boundary conditions of $f_{+||}(-1) = 0$ and $f_{+||}(1) = 1$, and $f_{-||}(x)$ is another function such that $f_{-||}(-x)$ satisfies the same constraints as $f_{+||}(x)$. Here instead of the linear functions, we take both $f_{+||}(x)$ and $f_{-||}(x)$ to be quadratic. After imposing the boundary and monotonicity conditions, and simplifying f by eliminating the constant term from it, we choose the integrand of E_{HO} to have the form:

$$f_{||}(x) = (a + 1)x + (a - 1)x^2. \quad (C.2)$$

Note that $f_{||}(1) = 2a$, while $f_{||}(-1) = -2$. The monotonicity constraint applied directly to $f_{||}$ implies that $(1/3) \leq a \leq 3$, which is the same as the constraint on a arising from the monotonicity constraints on $f_{+||}$ and $f_{-||}$ separately, couple with the vanishing of the constant term. Moreover, to guarantee that $f_{||}(1)$ has the larger magnitude than $f_{||}(-1)$, we have another constraint that $a > 1$. Therefore, the parameter a takes the values on the interval $[1, 3]$.

The contour energy thus becomes

$$\tilde{E}_{HO}(\gamma) = -\frac{1}{2} \iint_{S^1 \times S^1} [(a+1)(\dot{\gamma}(s) \cdot \dot{\gamma}(s')) + (a-1)(\dot{\gamma}(s) \cdot \dot{\gamma}(s'))^2] \Psi_{-||} ds ds' . \quad (C.3)$$

Since each term of \tilde{E}_{HO} includes tangent vector, when changing the contour formulation to the phase field formulation, there is no need to insert the boundary indicator function. Accordingly, the corresponding phase field energy is given by

$$\begin{aligned} \tilde{E}_{HO}(\phi) = & -\frac{1}{2} \iint_{\Omega^2} [(a+1)(\nabla\phi(x) \cdot \nabla\phi(x')) \\ & + (a-1)(\nabla\phi(x) \cdot \nabla\phi(x'))(\nabla\phi(x) \cdot \nabla\phi(x'))] \Psi\left(\frac{|x-x'|}{d}\right) dx dx' . \end{aligned} \quad (C.4)$$

The second term is different from E_S . We thus define another *nonlinear nonlocal* HOAC prior energy \tilde{E}_{NL} as

$$\tilde{E}_{NL}(\phi) = -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x'))(\nabla\phi(x) \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' . \quad (C.5)$$

Obviously, \tilde{E}_{NL} is also quartic in ϕ , but compared to E_{NL} , this term leads to some implementation difficulties when performing the energy minimization. We will explain after first calculating its derivative.

C.2 Derivative of \tilde{E}_{NL}

To calculate the derivative of \tilde{E}_{NL} , we first calculate \tilde{E}_{NL} with a small variation of the phase field $\delta\phi$:

$$\begin{aligned} \tilde{E}_{NL}(\phi + \delta\phi) &= -\frac{\beta_2}{4} \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x') + \nabla\delta\phi(x) \cdot \nabla\phi(x') + \nabla\phi(x) \cdot \nabla\delta\phi(x')) \\ &\quad (\nabla\phi(x) \cdot \nabla\phi(x') + \nabla\delta\phi(x) \cdot \nabla\phi(x') + \nabla\phi(x) \cdot \nabla\delta\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= \tilde{E}_{NL}(\phi) - \frac{\beta_2}{2} \iint_{\Omega^2} \left\{ (\nabla\phi(x) \cdot \nabla\phi(x'))(\nabla\delta\phi(x) \cdot \nabla\phi(x')) \right. \\ &\quad \left. + (\nabla\phi(x) \cdot \nabla\phi(x'))(\nabla\phi(x) \cdot \nabla\delta\phi(x')) \right\} \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= \tilde{E}_{NL}(\phi) - \beta_2 \iint_{\Omega^2} (\nabla\phi(x) \cdot \nabla\phi(x'))(\nabla\phi(x') \cdot \nabla\delta\phi(x)) \Psi\left(\frac{|x-x'|}{d}\right) dx dx' \\ &= \tilde{E}_{NL}(\phi) + \beta_2 \int_{\Omega} \delta\phi(x) \nabla \cdot \left\{ \int_{\Omega} \nabla\phi(x') (\nabla\phi(x) \cdot \nabla\phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx' \right\} dx , \end{aligned} \quad (C.6)$$

so the energy derivative is

$$\begin{aligned}
\frac{\delta \tilde{E}_{NL}(\phi)}{\delta \phi(x)} &= \beta_2 \nabla \cdot \left\{ \int_{\Omega} \nabla \phi(x') (\nabla \phi(x) \cdot \nabla \phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) dx' \right\} \\
&= \beta_2 \int_{\Omega} \left\{ (\nabla \phi(x') \cdot \nabla \nabla \phi(x) \cdot \nabla \phi(x')) \Psi\left(\frac{|x-x'|}{d}\right) \right. \\
&\quad \left. + (\nabla \phi(x) \cdot \nabla \phi(x')) \left(\nabla \phi(x') \cdot \nabla \Psi\left(\frac{|x-x'|}{d}\right) \right) \right\} dx' . \tag{C.7}
\end{aligned}$$

Unlike the gradient in equation (5.9), the above gradient cannot be expressed as any form of convolution. The computation is thus very expensive. In the end, we choose the energy E_{NL} as the final definition of the *nonlinear nonlocal* HOAC prior term.

Appendix D

Another Type of Interaction Function

As we have seen in subsection 5.3.3, due to the finite range of the interaction function Ψ (equation (5.2)), for a given $\hat{d}_2 > \hat{D}_2$, a corner point of the $\hat{\beta}-\hat{\beta}_2$ diagram goes off to infinity, whereas, at this critical point, the total prior energy should have a local maximum and a very wide local minimum (the inflection point of the other two local minima and the other local maximum). This is exactly what we need for modeling multiple widths of network branches over a wide range. In this appendix, we describe how to bring in this swallowtail corner, and thus achieve a wide minimum, by changing to a new type of interaction function with infinite support.

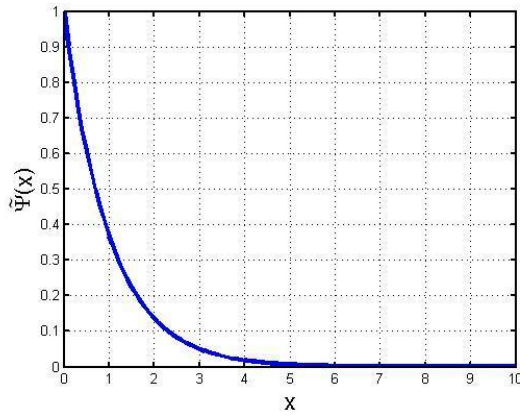


Figure D.1. The new interaction function $\tilde{\Psi}$.

We define the new interaction function $\tilde{\Psi}$ (see Figure D.1) as

$$\tilde{\Psi}(x) = e^{-x} . \quad (\text{D.1})$$

Note that this form is assumed to be the same in each HOAC energy term. We denote the standard HOAC term and the linear nonlocal HOAC term with $\tilde{\Psi}$ respectively as \tilde{E}_S and \tilde{E}_L . Thus, the overall prior model is the sum of the three energies: $\tilde{E}_{P,L} = E_0 + \tilde{E}_S + \tilde{E}_L$. As

before, we obtain the new prior energy calculated using the *ansatz* for a long straight bar, per unit length, $\tilde{e}_{P,L}$ (after removing the constant terms):

$$\begin{aligned} \tilde{e}_{P,L}(w, \hat{W}) = & \frac{4}{3}\alpha\hat{W}d + \frac{4}{15}\lambda w + \frac{4}{w} + 8\beta d \int_{\hat{W}}^{\infty} \sqrt{\eta^2 - \hat{W}^2} e^{-\eta} d\eta \\ & + 8\beta_2 d^3 \int_{\hat{W}}^{\infty} \eta \sqrt{\eta^2 - \hat{W}^2} e^{-\frac{\eta}{\hat{d}_2}} d\eta. \end{aligned} \quad (\text{D.2})$$

After several intermediate steps, we find the first derivatives with respect to w and \hat{W} , and the second derivative with respect to \hat{W} , of the above formulation. They are as follows:

$$\frac{\partial \tilde{e}_{P,L}}{\partial \hat{W}} = \frac{4}{3}\alpha d - 8\beta \hat{W}d \int_{\hat{W}}^{\infty} \frac{1}{\sqrt{\eta^2 - \hat{W}^2}} e^{-\eta} d\eta - \frac{8\beta_2 \hat{W}d^3}{\hat{d}_2} \int_{\hat{W}}^{\infty} \sqrt{\eta^2 - \hat{W}^2} e^{-\frac{\eta}{\hat{d}_2}} d\eta, \quad (\text{D.3a})$$

$$\frac{\partial \tilde{e}_{P,L}}{\partial w} = \frac{4}{15}\lambda - \frac{4}{w^2}, \quad (\text{D.3b})$$

$$\frac{\partial^2 \tilde{e}_{P,L}}{\partial \hat{W}^2} = -8\beta d \int_{\hat{W}}^{\infty} \frac{1 - \eta^2 + \hat{W}^2}{\sqrt{\eta^2 - \hat{W}^2}} e^{-\eta} d\eta - \frac{8\beta_2 d^3}{\hat{d}_2} \int_{\hat{W}}^{\infty} \frac{\eta^2 - 2\hat{W}^2}{\sqrt{\eta^2 - \hat{W}^2}} e^{-\frac{\eta}{\hat{d}_2}} d\eta. \quad (\text{D.3c})$$

Following the same procedure as in subsection 5.3.3, we observe that in this case, the singular point of \hat{d}_2 , *i.e.* \hat{D}_2 , is approximately 3.4. Here we are only interested in a swallowtail corner, so in the following analysis, we consider only the case where $\hat{d}_2 > \hat{D}_2$. For different given values of \hat{d}_2 , their $\hat{\beta} - \hat{\beta}_2$ diagrams, and corresponding $\tilde{e}_{P,L}$ against W with parameters at the upper swallowtail corner, are illustrated in Figure D.2. This shows clearly that this corner point now occurs at a finite value, and that the energy per unit length of bar $\tilde{e}_{P,L}$, with $\hat{\beta}$ and $\hat{\beta}_2$ values at this corner, has a flat minimum. We see from these figures that, with the increase of \hat{d}_2 , the region where $\tilde{e}_{P,L}$ has two local minima becomes larger; and in the energy plot, the peak of the first local maximum becomes higher. Note that the case where \hat{d}_2 is near \hat{D}_2 (see Figure D.2(a)) is interesting. Since the three intersection points between every pair of the three curves are close, all four extrema of $\tilde{e}_{P,L}$ tend to merge together, and thus generate a very wide minimum (see Figure D.2(b)).

Generating a wide flat minimum seems to be an effective way to extract branches of multiple widths. Moreover, the condition of being at that corner enables $\hat{\beta}$ and $\hat{\beta}_2$ to be fixed; thereafter only \hat{d}_2 needs to be chosen. Parameter setting is thereby facilitated, but a new problem arises: in this situation, we cannot control the amplitude of the first local maximum and the width of the flat minimum at the same time. If the amplitude of the first local maximum is too large, many undesired structures in the image may not be able to be eliminated during the evolution; on the other hand, if this amplitude is too small, some thin features might slide down to zero width, and thus vanish. We may need to add another parameter of the right type to the interaction function, in order to achieve separate control of the above two factors, which is a subject for future work.

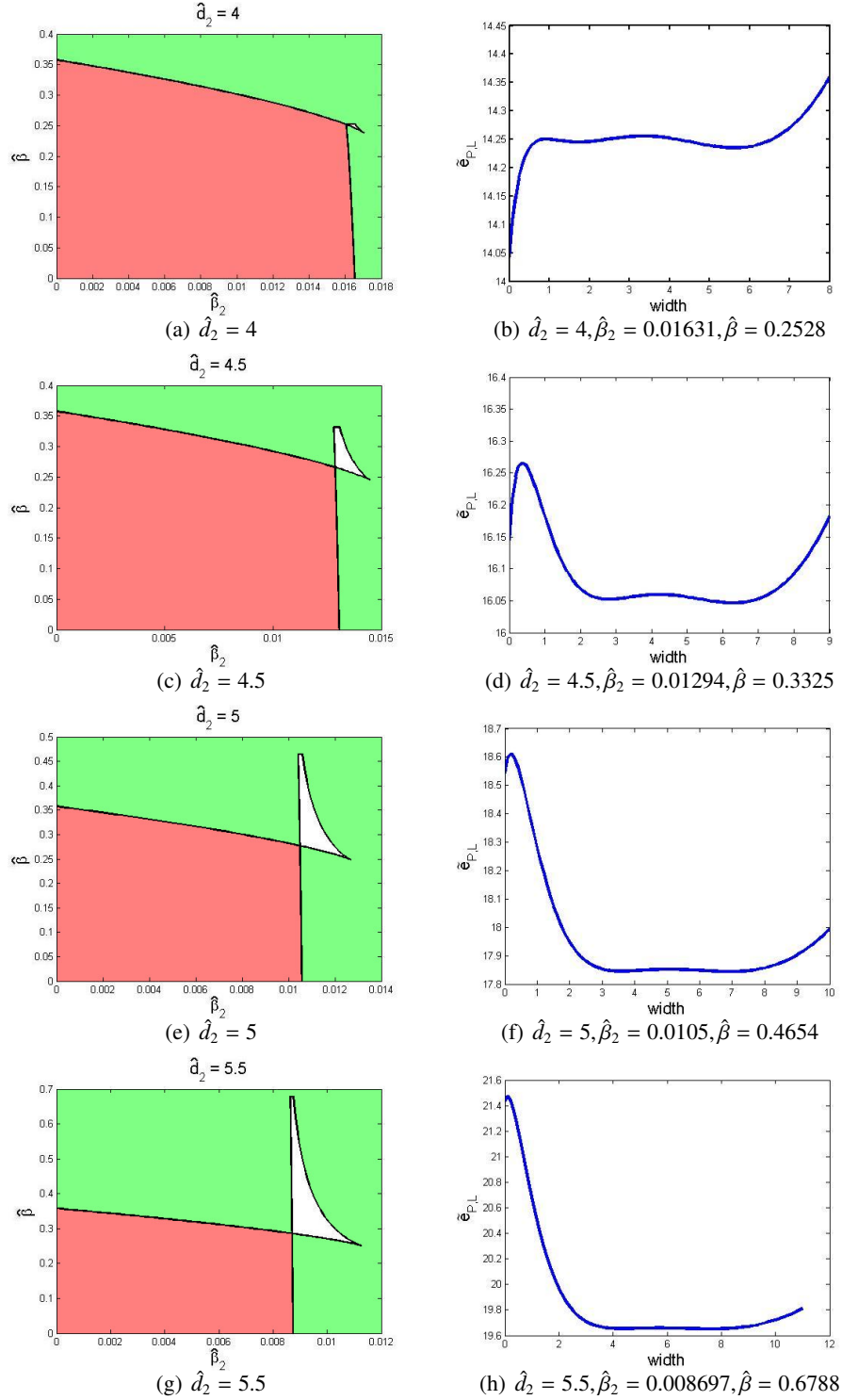


Figure D.2. For different given values of $\hat{d}_2 > \hat{D}_2$, left column: $\tilde{\beta} - \hat{\beta}_2$ diagrams. $\tilde{e}_{P,L}$ has either no local minimum (red), one local minimum (green), or two local minima (white). Right column: the corresponding energy per unit length of bar $\tilde{e}_{P,L}$ against W with parameters at the upper swallowtail corner.

Appendix E

Summary of Other Methods Used in Our Comparisons

In this appendix, we recall the other methods in the literature that have been used in this manuscript for comparison purposes. They are (i) a classification, tracking, and morphology algorithm by Wang and Zhang (2003) (section E.1); (ii) a fast but rough segmentation technique based on “straight line density” by Yu et al. (2004) (section E.2); and (iii) a level set approach with global shape constraint by Bailloeul (2005) (section E.3).

E.1 Method by Wang

Wang and Zhang (2003) attempted to extract urban road networks using pan-sharpened multispectral and panchromatic QuickBird imagery in two stages.

The first stage is a conventional multispectral classification of a pan-sharpened QuickBird multispectral image. An unsupervised fuzzy K-means classifier is used on the red, green and near infrared channels.

In the second stage, the edge image for road extraction is obtained from the corresponding panchromatic image by a Robert edge detector. After performing edge-aided segmentation, the classified road network is combined with the edge information so that some misclassified objects are eliminated. Then a connected component labeling algorithm is employed to register each isolated object. Finally, the roads are refined by a mathematical morphology operation.

E.2 Method by Yu

To register the satellite image to the GIS data, Yu et al. (2004) performed a method based on “straight line density” to extract roughly the road network. They considered only the main road features having a pronounced straight linear pattern, while other road features with high curvatures are omitted. A constraint is made that for any pixel on the main roads, there exists at least one straight segment with given length passing through it, and most of the pixels on the line segment possess road-specific spectral values.

A set of candidate pixels whose gray values are in the road-specific spectral range $[S_1, S_2]$ is extracted. For each candidate pixel, a set of straight segments with a pre-defined length L passing through this candidate pixel is obtained. Then, for each line segment, the ratio of the number of pixels whose gray value belongs to $[S_1, S_2]$ to the number of all pixels on the line is checked. If the ratio is above a given value, *i.e.* 95%, this line segment is considered as one part of the main roads; otherwise, this line is removed from the candidate list.

E.3 Method by Bailloeul

For the purpose of building map updating, Bailloeul (2005) used the Chan-Vese functional (Chan and Vese, 2001a) plus a shape prior (Chan and Zhu, 2003):

$$E_{CZ}(\phi, \psi_0) = E_{CV}(\phi) + \lambda_{\text{shape}} E_{\text{shape}}(\phi, \psi_0), \quad (\text{E.1})$$

where

$$E_{CV}(\phi) = \lambda_{\text{in}} \int_{\Omega} |I(x) - \mu_{\text{in}}|^2 H(\phi(x)) dx + \lambda_{\text{out}} \int_{\Omega} |I(x) - \mu_{\text{out}}|^2 (1 - H(\phi(x))) dx, \quad (\text{E.2a})$$

$$E_{\text{shape}}(\phi, \psi_0) = \int_{\Omega} \{H(\phi(x)) - H(T_{\text{sim}} \circ \psi_0(x))\}^2 dx. \quad (\text{E.2b})$$

In these formulations, μ_{in} and μ_{out} respectively denote the image mean grey level inside and outside the evolving active contour, which is embedded into the level set function ϕ . ψ_0 is the level set function embedding the prior reference shape. H represents the Heaviside function. T_{sim} is a similarity transformation computed at each iteration.

In (Bailloeul, 2005), an old digital map of the building is used to initialize the contour. A gradient descent algorithm is used to minimize E_{CZ} , while a Simplex algorithm is used to estimate the parameters of T_{sim} . At convergence, a refined version of the old map is obtained. However, one limit of this technique is its inability to handle objects newly appeared in the image.

Appendix F

Publications and Scientific Activities of the Author

Journal papers

1. T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. Extended phase field higher-order active contour models for networks. *International Journal of Computer Vision*, under submission.
2. T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. Incorporating generic and specific prior knowledge in a multi-scale phase field model for road extraction from VHR image. *IEEE Trans. Geoscience and Remote Sensing Special Issue: Selected Topics in Applied Earth Observations and Remote Sensing*, 1(2):139-146, 2008.
3. T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. A robust framework based on phase field modeling for road network extraction from VHR satellite image. *Chinese Journal of Computers*, under review.

Conference papers

1. T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. An extended phase field higher-order active contour model for networks and its application to road network extraction from VHR satellite image. In *Proc. European Conference on Computer Vision 2008 (ECCV 2008)*, Marseille, France, October 13-16, 2008.
2. T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. Extraction of main and secondary roads in VHR images using a higher-order phase field model. In *Proc. XXI ISPRS Congress, Commission III, Part A*, Beijing, China, July 3-11, 2008.
3. T. Peng, I. H. Jermyn, V. Prinet, J. Zerubia, and B. Hu. A phase field model incorporating generic and specific prior knowledge applied to road network extraction from VHR satellite images. In *Proc. British Machine Vision Conference 2007 (BMVC 2007)*, Warwick, UK, September 10-13, 2007.

4. T. Peng, I. H. Jermyn, V. Prinet, J. Zerubia, and B. Hu. Urban road extraction from VHR images using a multiscale approach and a phase field model of network geometry. In *Proc. 4th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN 2007)*, Paris, France, April 11-13, 2007.

Talks

1. Road extraction from very high resolution satellite images, Beijing Municipal Committee of Communications, Beijing, China, April 28, 2008.
2. Road extraction from very high resolution satellite images, Beijing Institute of Surveying and Mapping, Beijing, China, April 17, 2008.
3. Phase field models and higher-order active contours for road updating, Alcatel Alenia Space (new name: Thales Alenia Space), Toulouse, France, November 29, 2006.

Bibliography

- P. Agouris, A. Stefanidis, and S. Gyftakis. Differential snakes for change detection in road segments. *Photogrammetric Engineering and Remote Sensing*, 67(12):1391–1399, 2001.
- A. Amini, A. Weymouth, and R. Jain. Using dynamic programming for solving variational problem in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(9): 855–867, 1990.
- M. Amo, F. Martínez, and M. Torre. Road extraction from aerial images using a region competition algorithm. *IEEE Trans. on Image Processing*, 15:1192–1201, 2006.
- K. Appel and W. Haken. Every planar map is four colorable. *Contemporary Math*, 98, 1989.
- G. Aubert and P. Kornprobst. *Mathematical problems in image processing: Partial differential equations and the calculus of variations*. Springer-Verlag, New York, 2002.
- G. Aubert, J. F. Aujol, and L. Blanc-Féraud. Detecting codimension - two objects in an image with Ginzburg-Landau models. *International Journal of Computer Vision*, 65(1-2):29–42, 2005.
- T. Bailloeul. *Active contours and prior knowledge for change analysis: Application to digital urban building map updating from optical high resolution remote sensing images*. PhD thesis, CASIA and INPT, October 2005.
- M. Barzohar and D. B. Cooper. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(7):707–721, 1996.
- A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein. Multi-resolution, semantic objects, and context for road extraction. In *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, Basel, Switzerland, May 1997.
- A. Baumgartner, S. Hinz, and C. Wiedemann. Efficient methods and interfaces for road tracking. *International Archives of Photogrammetry and Remote Sensing*, XXXVI:28–31, 2002.
- M. Beneš, V. Chaloupecký, and K. Mikula. Geometrical image segmentation by the Allen-Cahn equation. *Applied Numerical Mathematics*, 51:187–205, 2004.

- J.M. Benharrosh. *Extraction de thèmes cartographiques dans les images satellites ou aériennes. Application à la génération de quick-looks adaptatifs et à la compression des images*. PhD thesis, Université de Nice-Sophia Antipolis, April 1998.
- R. Bonnefon, P. Dhérété, and J. Desachy. Geographic information system updating using remote sensing image. *Pattern Recognition Letters*, 23:1073–1083, 2002.
- V. Bucha, S. Uchida, and S. Ablameyko. Interactive road extraction with pixel force fields. In *Proc. IEEE International Conference on Pattern Recognition*, Hong Kong, China, August 2006.
- V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, 1993.
- V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
- T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2):266–277, 2001a.
- T. F. Chan and L. A. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. In *Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision*, Vancouver, Canada, July 2001b.
- T. F. Chan and W. Zhu. Level set based shape prior segmentation. Technical Report 03-66, UCLA, August 2003.
- T. F. Chan, B. Y. Sandberg, and L. A. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image representation*, 11:130–141, 2000.
- J. Chanussot and P. Lambert. An application of mathematical morphology to road network extraction on SAR images. In *Proc. International Symposium on Mathematical Morphology*, Amsterdam, The Netherlands, June 1998.
- L. Chen. Phase-field models for microstructure evolution. *Annual Review of Materials Research*, 50:113–140, 2002.
- Y. Chen, S. Thiruvankadam, H. Tagare, F. Huang, D. Wilson, and E. Geiser. On the incorporation of shape priors into geometric active contours. In *Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision*, Vancouver, Canada, July 2001.
- Y. Chen, H. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328, 2002.
- L. D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, 1991.

- L. D. Cohen and I. Cohen. Finite element methods for active contour models and balloons for 2D and 3D images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.
- I. Couloigner and T. Ranchin. Mapping of urban areas: a multiresolution modeling approach for semi-automatic extraction of streets. *Photogrammetric Engineering and Remote Sensing*, 66(7):867–874, 2000.
- D. Cremers. Dynamical statistical shape priors for level set-based tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(8):1262–1273, 2006.
- D. Cremers. Nonlinear dynamical shape priors for level set segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, June 2007.
- D. Cremers and S. Soatto. A pseudo-distance for shape priors in level set segmentation. In *Proc. IEEE Workshop Variational, Geometric, and Level Set Methods in Computer Vision*, Nice, France, October 2003.
- D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. *International Journal of Computer Vision*, 50(3):295–313, 2002.
- D. Cremers, S. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 69(3):335–351, 2006.
- D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: (i)ntegrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.
- A. P. Dal Poz and G. M. do Vale. Dynamic programming approach for semi-automated road extraction from medium- and high-resolution images. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, Germany, September 2003.
- A. P. Dal Poz and M. A. O. Silva. Active testing and edge analysis for road centerline extraction. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Graz, Austria, September 2002.
- M. A. Fischler, J. M. Tenenbaum, and H. C. Wolf. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing*, pages 201–223, 1981.
- M. F. A. Fortier, D. Ziou, C. Armenakis, and S. Wang. Survey of work on road extraction in aerial and satellite images. Technical Report 241, Université de Sherbrooke, Quebec, Canada, 1999.

- M. F. A. Fortier, D. Ziou, C. Armenakis, and S. Wang. Automated correction and updating of road databases from high-resolution imagery. *Canadian Journal of Remote Sensing*, 27(1):76–89, 2001.
- P. Gamba, F. Dell’Acqua, and G. Lisini. Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts. *IEEE Geoscience and Remote Sensing Letters*, 3:387–391, 2006.
- D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(1):1–14, 1996.
- S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- V. L. Ginzburg and L. D. Landau. On the theory of superconductivity. *Journal of Experimental and Theoretical Physics (USSR)*, page 1064, 1950.
- R. González-Cinca, R. Folch, R. Benítez, L. Ramírez-Piscina, J. Casademunt, and A. Hernández-Machado. Phase-field models in interfacial pattern formation out of equilibrium. *Advances in Condensed Matter and Statistical Mechanics*, pages 203–236, 2004.
- H. Grossauer and O. Scherzer. Using the complex Ginzburg-Landau equation for digital inpainting in 2D and 3D. In *Proc. Scale Space Methods in Computer Vision*, Lecture Notes in Computer Science 2695, pages 225–236. Springer, June 2003.
- A. Grote, M. Butenuth, M. Gerke, and C. Heipke. Segmentation based on normalized cuts for the detection of suburban roads in aerial imagery. In *Proc. 4th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Paris, France, April 2007.
- A. Gruen and H. Li. Semi-automatic linear feature extraction by dynamic programming and LSB-snakes. *Photogrammetric Engineering and Remote Sensing*, 63(8):985–995, 1997.
- C. Heipke, H. Mayr, C. Wiedemann, and O. Jamet. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, XXXII:47–56, 1997.
- J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Trans. on Geoscience and Remote Sensing*, 45:4144–4157, 2007.
- X. Hu, C. V. Tao, and Y. Hu. Automatic road extraction from dense urban area by integrated processing of high resolution imagery and Lidar data. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, September 2004.
- E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

- X. Jin and C. H. Davis. Multispectral IKONOS imagery automatic road extraction from high-resolution. In *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Toulouse, France, July 2003.
- M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- A. Katartzis, H. Sahli, V. Pizurica, and J. Cornelis. A model-based approach to the automatic extraction of linear features from airborne images. *IEEE Trans. on Geoscience and Remote Sensing*, 39(9):2073–2079, 2001.
- S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proc. IEEE International Conference on Computer Vision*, Boston, MA, USA, June 1995.
- J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Non-parametric methods for image segmentation using information theory and curve evolution. In *Proc. International Conference on Image Processing*, Rochester, New York, USA, September 2002.
- D. Klang. Automatic detection of changes in road databases using satellite imagery. *International Archives of Photogrammetry and Remote Sensing*, XXXII:293–298, 1998.
- C. Lacoste, X. Descombes, and J. Zerubia. Point processes for unsupervised line network extraction in remote sensing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1568–1579, 2005.
- J. Lai, A. Sowmya, and J. Trinder. Support vector machine experiments for road recognition in high resolution images. In *Machine Learning and Data Mining in Pattern Recognition*, Leipzig, Germany, July 2005.
- I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner. Automatic extraction of roads from aerial images based on scale space and snakes. *Machine Vision and Applications*, 12(1):23–31, 2000.
- B. Leroy, I. Herlin, and L. D. Cohen. Multi-resolution algorithms for active contour models. In *Proc. International Conference on Analysis Optimization*, Stuttgart, Germany, June 1996.
- M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, June 2000.
- H. Long and Z. Zhao. Urban road extraction from high resolution optical satellite images. *International Journal of Remote Sensing*, 26(22):4907–4921, 2005.
- S. G. Mallat. A theory for multiresolution signal decomposition: A wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.
- S. G. Mallat. *A wavelet tour of signal processing*. Academic Press, 2nd edition, 1999.

- F. Mangin, M. Berthod, and J. Zerubia. A cooperative network for contour grouping. In *Proc. IEEE International Conference on Pattern Recognition*, The Hague, Netherlands, August 1992.
- D. Marr. *Vision*. W. H. Freeman and Company, New York, 1982.
- G. Matheron. *Random sets and integral geometry*. Wiley, New York, 1975.
- H. Mayer, I. Laptev, and A. Baumgartner. Multi-scale and snakes for automatic road extraction. In *Proc. European Conference on Computer Vision*, volume 1, Freiburg, Germany, July 1998.
- T. McInerney and D. Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4:73–91, 2000.
- J. B. Mena. State of the art on automatic road extraction for GIS update: A novel classification. *Pattern Recognition Letters*, 24(16):3037–3058, 2003.
- S. Menet, P. Saint-Marc, and G. Medioni. B-snakes: Implementation and application to stereo. In *Proc. Image Understanding Workshop*, Pittsburgh, Pennsylvania, USA, September 1990.
- N. Merlet and J. Zerubia. New prospects in line detection by dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(4):426–431, 1996.
- Y. Meyer. *Wavelets and Operators*. Cambridge University, Cambridge, 1992.
- M. Mokhtarzade and M. J. Valadan Zoej. Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 9(1):32–40, 2007.
- D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, California, USA, June 1985.
- D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.
- M. Negri, P. Gamba, G. Lisini, and F. Tupin. Junction-aware extraction and regularization of urban road networks in high-resolution SAR images. *IEEE Trans. on Geoscience and Remote Sensing*, 44:2962–2971, 2006.
- M. Nielsen, L. Florack, and R. Deriche. Regularization, scale space, and edge detection filters. *Journal on Mathematical Imaging and Vision*, 7(4):291–307, 1997.
- S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

- N. Paragios and R. Deriche. Geodesic active regions: A new framework to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, 2002.
- N. Paragios, O. Mellina-Gottardo, and V. Ramesh. Gradient vector flow fast geometric active contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(3):402–407, 2004.
- T. Peng, I. H. Jermyn, V. Prinet, J. Zerubia, and B. Hu. Urban road extraction from VHR images using a multiscale approach and a phase field model of network geometry. In *Proc. 4th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Paris, France, April 2007a.
- T. Peng, I. H. Jermyn, V. Prinet, J. Zerubia, and B. Hu. A phase field model incorporating generic and specific prior knowledge applied to road network extraction from VHR satellite images. In *Proc. British Machine Vision Conference*, Warwick, England, September 2007b.
- T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. Extraction of main and secondary roads in VHR images using a higher-order phase field model. In *Proc. XXI ISPRS Congress, Commission III, Part A*, Beijing, China, July 2008a.
- T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. Incorporating generic and specific prior knowledge in a multi-scale phase field model for road extraction from VHR image. *IEEE Trans. Geoscience and Remote Sensing Special Issue: Selected Topics in Applied Earth Observations and Remote Sensing*, 1(2):139–146, 2008b.
- T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia. An extended phase field higher-order active contour model for networks and its application to road network extraction from very high resolution satellite image. In *Proc. European Conference on Computer Vision*, Marseille, France, October 2008c.
- P. Pérez, A. Blake, and M. Gangnet. JetStream: Probabilistic contour extraction with particles. In *Proc. IEEE International Conference on Computer Vision*, Vancouver, Canada, July 2001.
- R. Péteri and T. Ranchin. Detection and extraction of road networks from high resolution satellite images. In *Proc. International Conference on Image Processing*, Barcelona, Spain, September 2003.
- V. Prinet, S. Ma, and O. Monga. Scale selection for curvilinear structures detection from remote-sensed images. In *Proc. Asian Conference on Computer Vision*, Taiwan, January 2000.
- N. Provatas, N. Goldenfeld, and J. Dantzig. Efficient computation of dendritic microstructures using adaptive mesh refinement. *Physical Review Letters*, 80:3308–3311, 1998.

- T. Riklin-Raviv, N. Kiryati, and N. Sochen. Unlevel sets: Geometry and prior-based segmentation. In *Proc. European Conference on Computer Vision*, volume 1, Prague, Czech Republic, May 2004.
- T. Riklin-Raviv, N. Kiryati, and N. Sochen. Prior-based segmentation and shape registration in the presence of perspective distortion. *International Journal of Computer Vision*, 72(3):309–328, 2007.
- M. Rochery. *Contours actifs d'ordre supérieur et leur application à la détection de linéiques dans des images de télédétection*. PhD thesis, Université de Nice-Sophia Antipolis, September 2005.
- M. Rochery, I. H. Jermyn, and J. Zerubia. Higher order active contours and their application to the detection of line networks in satellite imagery. In *Proc. 2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, Nice, France, October 2003.
- M. Rochery, I. H. Jermyn, and J. Zerubia. Phase field models and higher-order active contours. In *Proc. IEEE International Conference on Computer Vision*, Beijing, China, October 2005a.
- M. Rochery, I. H. Jermyn, and J. Zerubia. Higher order active contours. Research Report 5656, INRIA, August 2005b.
- M. Rochery, I. H. Jermyn, and J. Zerubia. Higher-order active contours. *International Journal of Computer Vision*, 69(1):27–42, 2006.
- F. Rottensteiner, J. Trinder, and S. Clode. Data acquisition for 3D city models from LIDAR. In *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Seoul, South Korea, July 2005.
- M. Rousson. *Cue integration and front evolution in image segmentation*. PhD thesis, Université de Nice-Sophia Antipolis, December 2004.
- M. Rousson and N. Paragios. Shape priors for level set representations. In *Proc. European Conference on Computer Vision*, volume 2, Copenhagen, Denmark, May 2002.
- M. Rousson and N. Paragios. Prior knowledge, level set representations & visual grouping. *International Journal of Computer Vision*, 76(3):1573–1405, 2007.
- P. Saint-Marc, J. Chen, and G. Medioni. Adaptive smoothing: a general tool for early vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6):514–529, 1991.
- C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A variational model for image classification and restoration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(5):460–472, 2000.
- J. Serra. *Image analysis and mathematical morphology, Volume 1. Theoretical advances, Volume 2*. Academic Press, London, 1982.

- P. Soille and M. Pesaresi. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Trans. on Geoscience and Remote Sensing*, 40(9):2042–2055, 2002.
- M. Song and D. Civco. Road extraction using SVM and image segmentation. *Photogrammetric Engineering and Remote Sensing*, 70(12):1365–1372, 2004.
- R. Stoica, X. Descombes, and J. Zerubia. A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision*, 57(2):121–136, 2004.
- F. Tupin, H. Maitre, J. F. Mangin, J. M. Nicolas, and E. Pechersky. Detection of linear features in SAR images: Application to road network extraction. *IEEE Trans. on Geoscience and Remote Sensing*, 36(2):434–453, 1998.
- M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing: Part i – Theory. *IEEE Trans. on Signal Processing*, 41(2):821–833, 1993a.
- M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing: Part ii – Efficient design and applications. *IEEE Trans. on Signal Processing*, 41(2):834–848, 1993b.
- S. Urago, J. Zerubia, and M. Berthod. A Markovian model for contour grouping. In *Proc. IEEE International Conference on Pattern Recognition*, Jerusalem, Israel, October 1994.
- L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
- G. Vosselman and J. de Knecht. Road tracing by profile matching and kalman filtering. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Suisse, April 1995.
- R. Wang and Y. Zhang. Extraction of urban road network using Quickbird pan-sharpened multispectral and panchromatic imagery by performing edge-aided post-classification. In *Proc. International Society for Photogrammetry and Remote Sensing (ISPRS)*, Quebec City, Canada, October 2003.
- D. J. Williams and M. Shah. A fast algorithm for active contours and curvature estimation. *CVGIP: Image Understanding*, 55(1):14–26, 1992.
- S. Wolfram. *The mathematica[©] book*. Cambridge: Cambridge University Press, 4th edition, 1999.
- C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Processing*, 7:359–369, 1998.
- C. Xu, J. Yezzi, and J. Prince. On the relationship between parametric and geometric active contours. In *Proc. Asilomar Conference Signal, Systems, and Computers*, Pacific Grove, California, USA, October 2000.

- N. Yager and A. Sowmya. Support vector machines for road extraction from remotely sensed images. In *Computer Analysis of Images and Patterns*, Groningen, The Netherlands, August 2003.
- J. Youn and J. S. Bethel. Adaptive snakes for urban road extraction. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, September 2004.
- Z. Yu, V. Prinet, C. Pan, and P. Chen. A novel two-steps strategy for automatic GIS-image registration. In *Proc. International Conference on Image Processing*, Singapore, October 2004.
- E. C. Zeeman. *Catastrophe theory. Selected papers 1972-1977*. Addison-Wesley, 1977.
- C. Zhang, S. Murai, and E. P. Baltsavias. Road network detection by mathematical morphology. In *Proc. of ISPRS Workshop "3D Geospatial Data Production: Meeting Application Requirements"*, Paris, France, April 1999.
- Q. Zhang and I. Couloigner. Automatic road change detection and GIS updating from remotely-sensed image. In *Proc. of Asia GIS 2003 Conference*, Wuhan, Hubei, China, October 2003.
- J. Zhou, W. F. Bischof, and T. Caelli. Robust and efficient road tracking in aerial images. *International Archives of Photogrammetry and Remote Sensing*, XXXVI:35–40, 2005.
- J. Zhou, W. F. Bischof, and T. Caelli. Road tracking in aerial images based on human-computer interaction and bayesian filtering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(2):108–124, 2006.
- P. Zhu, Z. Lu, X. Chen, K. Honda, and A. Eiumnoh. Extraction of city roads through shadow path reconstruction using laser data. *Photogrammetric Engineering and Remote Sensing*, 70(12):1433–1440, 2004.
- S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.
- 马建文. 遥感数据模型与处理方法. 中国科学技术出版社出版, 合肥, 2001.
- 马颂德, 张正友. 计算机视觉. 科学出版社, 北京, 1997.

ABSTRACT

The objective of this thesis is to develop and validate robust approaches for the semi-automatic extraction of road networks in dense urban areas from very high resolution (VHR) optical satellite images. Our models are based on the recently developed higher-order active contour (HOAC) phase field framework. The problem is difficult for two main reasons: VHR images are intrinsically complex and network regions may have arbitrary topology. To tackle the complexity of the information contained in VHR images, we propose a multiresolution statistical data model and a multiresolution constrained prior model. They enable the integration of segmentation results from coarse resolution and fine resolution. Subsequently, for the particular case of road map updating, we present a specific shape prior model derived from an outdated GIS digital map. This specific prior term balances the effect of the generic prior knowledge carried by the HOAC model, which describes the geometric shape of road networks in general. However, the classical HOAC model suffers from a severe limitation: network branch width is constrained to be similar to maximum network branch radius of curvature, thereby providing a poor model of networks with straight narrow branches or highly curved, wide branches. We solve this problem by introducing two new models: one with an additional nonlinear nonlocal HOAC term, and one with an additional linear nonlocal HOAC term. Both terms allow separate control of branch width and branch curvature, and furnish better prolongation for the same width, but the linear term has several advantages: it is more efficient from a computational standpoint, and it is able to model multiple widths simultaneously. To cope with the difficulty of parameter selection of these models, we analyze the stability conditions for a long bar with a given width described by these energies, and hence show how to choose rigorously the parameters of the energy functions. Experiments on VHR satellite images and comparisons with other approaches demonstrate the superiority of our models.

Keywords: higher-order, active contour, phase field, prior, constraint, shape, multiresolution, parameter, road network extraction, map updating, very high resolution (VHR), dense urban area, remote sensing images.

RÉSUMÉ

L'objectif de cette thèse est de développer et de valider des approches robustes d'extraction semi-automatique de réseaux routiers en zone urbaine dense à partir d'images satellitaires optiques à très haute résolution (THR). Nos modèles sont fondés sur une modélisation par champs de phase des contours actifs d'ordre supérieur (CAOS). Le problème est difficile pour deux raisons principales : les images THR sont intrinsèquement complexes, et certaines zones des réseaux peuvent prendre une topologie arbitraire. Pour remédier à la complexité de l'information contenue dans les images THR, nous proposons une modélisation statistique multi-résolution des données ainsi qu'un modèle multi-résolution contraint a priori. Ces derniers permettent l'intégration des résultats de segmentation de résolution brute et de résolution fine. De plus, dans le cadre particulier de la mise à jour de réseaux routiers, nous présentons un modèle de forme a priori spécifique, dérivé d'une ancienne carte numérique issue d'un SIG. Ce terme spécifique a priori équilibre l'effet de la connaissance a priori générique apportée par le modèle de CAOS, qui décrit la forme géométrique générale des réseaux routiers. Cependant, le modèle classique de CAOS souffre d'une limitation importante : la largeur des branches du réseau est contrainte à d'être similaire au maximum du rayon de courbure des branches du réseau, fournissant ainsi un modèle non satisfaisant dans le cas de réseaux aux branches droites et étroites ou aux branches fortement incurvées et larges. Nous résolvons ce problème en proposant deux nouveaux modèles : l'un contenant un terme additionnel, non-local, non-linéaire de CAOS, et l'autre contenant un terme additionnel, nonlocal, linéaire de CAOS. Ces deux termes permettent le contrôle séparé de la largeur et de la courbure des branches, et fournissent une meilleure prolongation pour une même largeur. Le terme linéaire a plusieurs avantages : d'une part il se calcule plus efficacement, d'autre part il peut modéliser plusieurs largeurs de branche simultanément. Afin de remédier à la difficulté du choix des paramètres de ces modèles, nous analysons les conditions de stabilité pour une longue barre d'une largeur donnée décrite par ces énergies, et montrons ainsi comment choisir rigoureusement les paramètres des fonctions d'énergie. Des expériences sur des images satellitaires THR et la comparaison avec d'autres modèles démontrent la supériorité de nos modèles.

Mots clefs : ordre supérieur, contour actif, champ de phase, a priori, contrainte, forme, multi-résolution, paramètre, extraction de réseaux routiers, mise à jour de carte, très haute résolution (THR), urbain dense, images de télédétection.