Efficient Global Minimization Methods for Variational Problems in Imaging and Vision

Phd Thesis

Egil Bae

Department of Mathematics
University of Bergen

April, 2011

Abstract

Energy minimization has become one of the most important paradigms for formulating image processing and computer vision problems in a mathematical language. Energy minimization models have been developed in both the variational and discrete optimization community during the last 20-30 years. Some models have established themselves as fundamentally important and arise over a wide range of applications.

One fundamental challenge is the optimization aspect. The most desirable models are often the most difficult to handle from an optimization perspective. Continuous optimization problems may be non-convex and contain many inferior local minima. Discrete optimization problems may be NP-hard, which means algorithms are unlikely to exist which can always compute exact solutions without an unreasonable amount of effort.

This thesis contributes with efficient optimization methods which can compute global or close to global solutions to important energy minimization models in imaging and vision. New insights are given in both continuous and combinatorial optimization, as well as a strengthening of the relationships between these fields.

One problem that is extensively studied is minimal perimeter partitioning problems with several regions, which arise naturally in e.g. image segmentation applications and is NP-hard in the discrete context. New methods are developed that can often compute global solutions and otherwise very close approximations to global solutions. Experiments show the new methods perform significantly better than earlier variational approaches, like the level set method, and earlier combinatorial optimization approaches. The new algorithms are significantly faster than previous continuous optimization approaches.

In the discrete community, max-flow and min-cut (graph cuts) have gained huge popularity because they can efficiently compute global solutions to certain energy minimization models. It is shown that new types of problems can be solved exactly by max-flow and min-cut. Furthermore, variational generalizations of max-flow and min-cut are proposed which bring the global optimization property to the continuous setting, while avoiding grid bias and metrication errors which are major disadvantages of the discrete models. Convex optimization algorithms are derived from the variational max-flow models, which are very efficient and are more parallel friendly than traditional combinatorial algorithms.

Preface

This thesis is submitted as a partial fulfilment of the requirements for the degree Doctor of Philosophy (PhD) at the Department of Mathematics, the University of Bergen. The thesis is divided into two parts, where part I provides motivations, theoretical background, overview of the field, a summary of the scientific contributions and conclusions. Part II is composed of a series of research papers, which are listed on the next page. Most of the papers also involve co-workers. In all papers, the authors are ordered according to their contributions, meaning the first author contributed most, the second author contributed second most etc., some details can be found in the beginning of Section 6: Summary of papers.

Acknowledgements

This work has been funded by the Norwegian Research Council through the eVita project 166075.

During my time as a phd student, I feel privileged to have worked with many interesting, skilled and motivated people.

First, I want to thank my supervisor Xue-Cheng Tai. I appreciate his enthusiasm and constant encouragement both before and during my phd period, while at the same time leaving me a lot of freedom. I am also grateful for being introduced to several researchers and research groups, which has resulted in fruitful collaborations, and for hosting me for two semesters at Nanyang Technological University in Singapore.

Thanks goes to my main collaborators Jing Yuan, Juan Shi, Wan Min and Yu Wang. It has been great to work with you during the last two years, especially the collaboration with Jing Yuan has been fruitful.

Thanks to Stanley Osher and Luminita Vese who have been great hosts during my stay at the department of mathematics at University of California, Los Angeles for one semester.

Finally, I would like to thank my parents for always being supportive.

Egil Bae
Bergen, April 3rd, 2011

List of Publications

Publications included in this thesis

A **Egil Bae** and Xue-Cheng Tai,
Graph Cut Optimization for the Piecewise Constant Level Set Method Applied to Multiphase Image Segmentation. In *Proc. Second International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2009),* pages 1-13. Lecture Notes in Computer Science, Springer, Berlin, 2009.

B **Egil Bae** and Xue-Cheng Tai,
Efficient Global Minimization for the Multiphase Chan-Vese Model of Image Segmentation. In *Proc. Seventh International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2009),* pg.28-41, Lecture Notes in Computer Science, Springer, Berlin, 2009. (an extended journal version is included)

C **Egil Bae**, Jing Yuan and Xue-Cheng Tai,
Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach. *International Journal of Computer Vision*, 2010, DOI: 10.1007/s11263-010-0406-y

D Jing Yuan, **Egil Bae** and Xue-Cheng Tai,
A Study on Continuous Max-Flow and Min-Cut Approaches. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2010. (an extended journal version is included)

E **Egil Bae**, Jing Yuan, Xue-Cheng Tai and Yuri Boykov
A Fast Continuous Max-Flow Approach to Non-Convex Multilabeling Problems. Submitted for journal publication.

F **Egil Bae** and Xue-Cheng Tai,
Exact Convex Formulation of the Chan-Vese Model and a Tight Convex Relaxation of Pott's Model with 4 Regions. (To be submitted to a journal)

G **Egil Bae**, Juan Shi and Xue-Cheng Tai,
Graph Cuts for Curvature based Image Denoising. *IEEE Transactions on Image Processing, 2010, DOI: 10.1109/TIP.2010.2090533*,

H Jing Yuan, **Egil Bae**, Xue-Cheng Tai and Yuri Boykov,
A Continuous Max-Flow Approach to Potts Model. In *European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, Volume 6316/2010, pg. 379-392*, 2010

I Jing Yuan, **Egil Bae**, Yuri Boykov and Xue-Cheng Tai
A Continuous Max-Flow Approach to Minimal Partitions with Label Cost Prior. To appear In *Proc. Third International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)* 2011.

J Min Wan, Yu Wang, **Egil Bae**, Xue-Cheng Tai, Desheng Wang, Reconstructing Open Surfaces via Graph-Cuts. Submitted for journal publication.

Publications not included in this thesis

K **Egil Bae** and Joachim Weickert, Partial differential equations for interpolation and compression of surfaces. In *Proc. Seventh International Conference on Mathematical Methods for Curves and Surfaces (MMCS 2008),* pg. 1-14, Lecture Notes in Computer Science, Springer, Berlin, 2010. (based on some of the results in my master thesis)

Contents

Part I

Background

Chapter 1

Introduction

Image processing and computer vision are, broadly speaking, about inferring information about the physical world from visual observations of the physical world. Mathematically, it is the problem of finding order and pattern in high dimensional data, such as two or three dimensional digital images. One important example is the abstraction of objects from digital images, often called image segmentation. Other examples include reconstruction of images that are damaged or distorted in some way, such as removal of noise, or obtaining 3 dimensional knowledge from 2 dimensional data (2D-3D reconstruction).

There has been a lot of research on this subject during the last decades, and the field is expanding. The impact on society is constantly increasing. Medical image processing allows medical doctors to diagnose deceases from image modalities such MRI and CT. Driving assistants in automobiles help to prevent accidents by analyzing visual and sensory data from the surroundings of the car. In the industry, computer vision and image processing are used for visual inspection and manufacturing and has lead to significant reduction in manual labor. Efficient compression algorithms have made possible the wide distribution and accessibility of images and video on the internet.

Energy minimization has become an important paradigm for formulating image processing and computer vision problems in a mathematical language. Various problems can elegantly be formulated as the minimization of some predefined energy function or functional. Of special importance are variational and optimization problems composed of a data fitting term and a regularization term. The data fitting term forces the unknown to fit some observed data. However, many solutions can usually be explained by the observed data, therefore additional information is necessary to distinguish the best candidates and make the problem well posed. One basic assumption which can be made in image processing and computer vision, is that the data one wish to reconstruct is regular. That is, the values at different data points are correlated. The physical world contains a lot of order, as opposed to randomness, therefore information extracted from the physical world will also contain order and patterns. Correlated or regular data

will equivalently be sparse in some basis (such as in the fourier or wavelet basis), therefore sparsity is another word that is frequently used with the same meaning. The regularization term is designed such that regular solutions are favored. What is meant by regular is rather vague, will depend on the problem, and is still an open research question. Past research has shown that a very powerful criterion is to favor solutions that are spatially varying smoothly in some sense, and penalize random oscillatory patterns.

Models consisting of a data term and a regularization term have been developed both in the "variational community" and the "discrete optimization community" during the past 20-30 years. Variational models seek a function defined over a continuous domain (typically the image domain) which takes values in a continuous set and minimizes the energy. Such models are therefore in the form of continuous optimization problems. On the contrary, discrete models regard the image domain as discrete and seeks a variable at each discrete pixel taking values in a predefined finite set which minimizes the energy. Therefore, discrete models have the form of combinatorial optimization problems. In the discrete optimization community such models are often called markov random fields (mrfs).

Some models have established themselves as fundamentally important and occur over a wide range of applications. One such type of models is the minimal perimeter partitioning models, which divides the image into several regions according to some data fitting measure, while minimizing the lengths of the region boundaries (in 2D) as spatial regularity. The most important application is image segmentation into several regions. Such models are often called the "piecewise constant Mumford-Shah model" or "Chan-Vese model" in the continuous setting and Potts model in the discrete setting. Other important models include total variation regularized models, and some models which involve higher order derivatives.

One fundamental challenge is the optimization aspects. The models can be formulated in a relatively intuitively simple way, yet to compute the solutions can be very difficult. The most desirable models are often the most difficult to handle from an optimization perspective. Continuous optimization problems may be non-convex and contain many inferior local minima. Discrete optimization problems may be NP-hard, which means an algorithm is unlikely to exist which can always compute an exact solution without an unreasonable amount of effort.

In the discrete optimization community, combinatorial optimization algorithms have been developed and applied for solving the problems. One of the most successful and effective class of methods are based on algorithms for the graph optimization problems max-flow and min-cut. Some of the discrete optimization problems encountered in image processing and computer vision can be formulated as finding the minimum cut on a graph. The min-cut problem and max-flow problems are dual to each other, therefore one can solve the much easier max-flow problem to find minimum cuts. There exists very efficient algorithms

for computing max-flow when implemented directly on CPU. These algorithms can therefore be applied for efficiently solving the optimization problems exactly. Such algorithms are popularly called graph cuts. The most important class of problems that can be solved by graph cuts is partitioning problems, where the number of regions is fixed to two. However, a large class of the discrete optimization problems are NP-hard. Therefore, one cannot expect to develop algorithms which can solve these problems exactly on all instances, unless a practically infinite amount of computing time is allowed. Instead, algorithms have been developed for computing approximate solutions.

In the variational community, most existing optimization algorithms are based on solving the Euler Lagrange partial differential equations associated with the variational models. However, the Euler Lagrange equations do not distinguish between local and global solutions. Therefore, these approaches may easily result in local solutions, which could deviate far from the desired global solutions. The only exception is if the variational model is convex, in which case every local solution is also a global solution. For partitioning problems and geometrical problems which involve curve lengths and surface area, the level set method has become the most popular approach. It is attractive especially for its ability to deal with unknown topologies. However, the level set method results in non-convex formulations and may therefore easily get stuck in inferior local minima. This is both a theoretical and practical problem, as such local minima may deviate far from what one is trying to express with the model. For partitioning problems with two regions, convex formulations have recently been developed, which makes it possible to compute global solutions.

Discrete energy minimization models have the advantage over variational models, that existing combinatorial optimization algorithms can compute global solutions in some special cases. This is particularly the case for problems that can be stated as the minimum cut problem on a graph. On the other hand, variational models have many advantages over discrete models. Most importantly is the rotational invariance and ability to accurately represent geometrical entities such as curve lengths and surface areas. The discrete models are always biased by the discrete grid, and will favor curves or surfaces that are oriented along with the grid, e.g. in horizontal, vertical or diagonal directions. In the variational framework, discretization is also necessary to compute solutions numerically. Instead of discretizing the variational model directly, the Euler-Lagrange equations are discretized and solved numerically. This allows for much more accurate representations of geometrical entities, without grid bias and metrication artifacts. Furthermore, continuous optimization algorithms have a much more parallel friendly nature. The algorithms most often reduce to large scale matrix and vector operations, which are very suitable for implementation on graphic processing units (GPUs) or general purpose GPUs. In the future, improvement in processing technology is expected to be largely of the parallel aspect. Moore's law is reaching fundamental limits when it comes to the exponential improvements

of single core CPUs, and we have already started to see a massive investment in parallel hardware. Combinatorial algorithms can also be parallelized, but cannot be expected to scale nearly as well as continuous optimization algorithms.

The main contributions of this thesis are efficient optimization methods that can compute global or close to global solutions to energy minimization models that arise in image processing and computer vision. They include both combinatorial optimization and continuous optimization via variational models, as well as stronger unifications between these fields. Before stating the contributions in more detail, some more background information is provided.

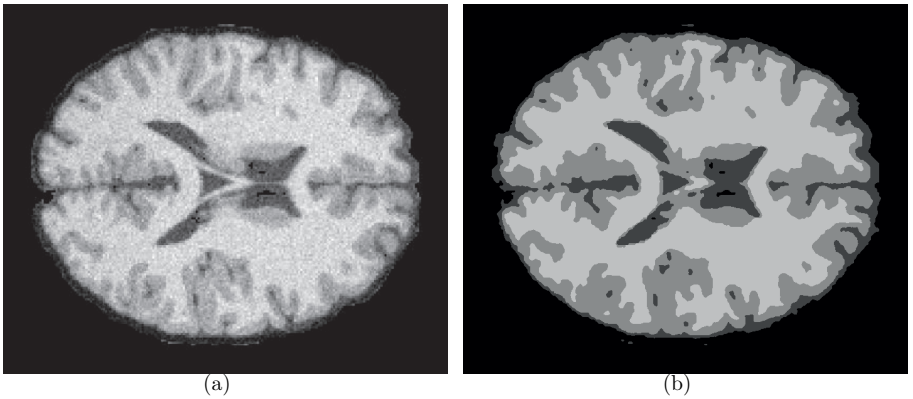
1.1 Preliminary definitions, fundamental problems in imaging and vision and important energy minimization models

In this section, we will give a brief overview of some of the most important problems in image processing and computer vision and introduce some fundamental energy minimization models. We start with some preliminary definitions.

An image is a function $I : \Omega \mapsto \mathcal{L}$ defined over the image domain $\Omega \subset \mathbb{R}^N$. The image domain is usually a rectangular 2-dimensional domain, that is $N = 2$, but 3-dimensional images also occur frequently, e.g. in medical imaging and seismology. Ideally, Ω can be a continuous domain of infinitely many points. In practice, Ω consists of a finite set of pixels. The image function I takes values in the set $\mathcal{L}$. For gray scale images, $\mathcal{L} \subset \mathbb{R}$ and I is a scalar function. For color images, $\mathcal{L} \subset \mathbb{R}^3$ consists of three color channels and I is a vector function. The set $\mathcal{L}$ can either be discrete or continuous. If $\mathcal{L}$ is discrete, the image is said to be quantized, typically $\mathcal{L} = \{0, ..., L\}$ consists of a set of integer gray values. If $\mathcal{L}$ is continuous, it is typically assumed the image is scaled such that $\mathcal{L} = [0, 1]$.

Some of the most important problems in image processing and computer vision are image segmentation, denoising, inpainting and 2D-3D reconstruction. They will all be touched in this thesis. For completeness, we give a very brief statement of these problems.

Image segmentation: Image segmentation is the problem of partitioning the image domain Ω into several meaningful regions, based on the image intensity function I. It is one of the most important problems in computer vision, and perhaps also the most challenging one. The human brain is remarkably good at abstracting objects from visual information, yet an algorithmic formulation of the problem has turned out to be very difficult. In low level vision, one assume no a priori knowledge of the shapes of the unknown objects, therefore the image intensity function I is the only guide for abstracting the regions. The problem can be defined more formally as obtaining a set of regions $\{\Omega_i\}_{i=1}^n$ which covers



(a) (b)

Figure 1.1: Image segmentation is the problem of inferring objects from visual information. (a) Input MRI image. (b) Segmentation into 4 regions: cerebrospinal fluid, gray matter, white matter and background. Each region is visualized with a distinct gray value. The result was computed by the algorithm in paper C.

the whole image domain Ω without any overlap, i.e.

$$\cup_{i=1}^{n}\Omega_i \,=\, \Omega\,, \quad \Omega_k \cap \Omega_l \,=\, \emptyset\,, \forall k \neq l.$$

The number of regions n can either be fixed in advance or be unknown. An example is given in Figure 1.1, where one want to classify the input MRI image into the regions: cerebrospinal fluid, gray matter, white matter and background.

Image denosing: Noise occurs naturally during the acquisition processes of images due to uncertainty in the measurements. It also occurs during analogue transmission of images. Digital image denoising is therefore one of the most important tasks of image processing. Assume the image I is corrupted by noise. We let I_0 denote the true image, which is unknown. Most often it is assumed the noise is additive, meaning the image function I can be written

$$I = I_0 + \eta, \tag{1.1}$$

where η is the unknown noise function defined over Ω. Image denoising is the problem of reconstructing an image which as close as possible resembles I_0 from I

2D-3D reconstruction: The goal of 2D-3D reconstruction is to infer three dimensional information about the physical world from one of several two dimensional observations (i.e. 2D images) of the physical world. This is one of the oldest problems in computer vision and has received considerable amounts of attention. Shape from shading applies physical laws of light reflection to construct

three dimensional objects from a two dimensional image. Multiview 2D-3D reconstruction searches for correspondences between several 2D images of a scene to create a three dimensional representation of the scene. Shape from shading and multiview reconstruction can also be combined.

Interpolation and inpainting: Other problems that can be formulated in an energy minimization framework and will be touched in this thesis are interpolation and inpainting. Both these problems are related, in that their goal is to construct new data from existing data. Typically, some data is missing for various reasons and should be reconstructed by using the information which is available. In interpolation, the available data is scattered, while in inpainting the data is missing in larger regions and data is available at the boundaries of the regions. Both these problems can be formulated as the minimization of an energy function/functional composed of a data term and regularization term. The regularization term play the role of filling in missing information, while the data term forces the solution to coincide with, or approximate, the available data at the known locations.

All the above problems can be formulated as the minimization of an energy composed of a data fitting term and a regularization term. Some models arise naturally in many of the above problems, and can be used over a wide range of applications. The energy minimization models developed for image denoising, inpainting and interpolation are closely related and can be used interchangeably with small modifications. The same goes for models developed for problems such as image segmentation, surface interpolation and 2D-3D reconstruction. As stated in the last section, such optimization formulations have been developed in both the variational community and discrete optimization community. The discrete models and variational models are closely related. Some models have established themselves as fundamentally important.

One important class are energy minimization formulations of segmentation and grouping problems, which seek partitions of the image domain into several regions. The data fitting term sets a cost for the assignment of each individual image point (pixel) to each region based on its intensity value. A simple, but powerful, spatial regularization term can be added which favors regions of small boundary length (in 2D) or surface area (in 3D). In the special case of an L_2 data fitting term, such models are often called piecewise constant Mumford-Shah model in the variational setting. Equivalent discrete representations are often referred to as Pott's model in the discrete optimization community. Generalizations can also be made by introducing spatially varying strength of the regularization term, to align the region boundaries to edges in the image, or using non-local operators which distinguish texture from noise. The most important application of these models are image segmentation, but they can also be applied in other problems, like 3D reconstruction from 2D images.

Another important class of models, are the total variation regularized models. Total variation has become one of the most popular and powerful regularization

terms for inverse problems. Its attractiveness can be explained by its combined ability to preserve edges and convexity, which makes computation trackable. Total variation is perhaps best known in connection with the Rudin-Osher-Fatemi model of image denoising, which uses the L_2 norm between the noisy image and the reconstructed image as data fidelity term and total variation as regularization term. It also arise naturally in many other applications, like 2D-3D stereo reconstruction, where the data term can possibly be non-convex. Total variation has also been studied in the discrete setting as a special case of first order markov random fields.

Other notable models have also been developed for image denoising and inpainting. Except for total variation, the most successful first order image denoising model is the general Mumford-Shah model, where the regularization term favors solution images which are smooth everywhere except for a smooth discontinuity set. The Mumford-Shah model has advantages over total variation, but is much more difficult to handle computationally. It can also be cast as a special case of first order markov random fields, where its computational complexity also becomes apparent. Another class of denoising and inpainting models use higher order regularization terms. In the recent years, there has been increasing activity in this area. One of the first and most sucessful is the Euler's elastica model which uses a regularization term that favors solutions with level lines of small curvature. It is one of the few higher order models, that are also discontinuity preserving.

1.2 Contributions of this thesis

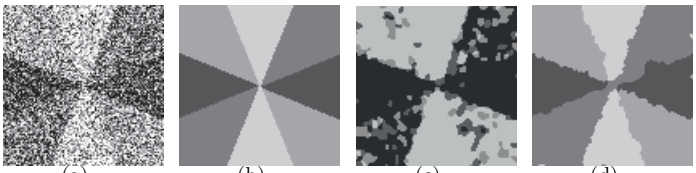
The main contribution of this thesis is the development of efficient methods that can compute global or close to global solutions to important energy minimization models in imaging and vision. The thesis contributes with new developments in continuous and combinatorial optimization, as well as a strengthening of the relationships between these fields.

Algorithms for computing global solutions are generally further developed in the field of combinatorial and integer optimization than continuous optimization. One example is the max-flow and min-cut (graph cuts) algorithms, which have been used to compute global solutions to certain discrete energy minimization models in imaging and vision for a long time. In fact, any first order markov random field which is not NP-hard, can be minimized via graph cuts. Although much has been said on this topic already, more insights will be given. In paper B it is shown that a new type of multi-region problems, which has become very popular in the variational community, can efficiently be solved exactly in the discrete setting, by computing the minimum cut on a special graph, under some conditions on the data term. If the conditions are not met, algorithms are proposed which are not guaranteed to find the exact global optimum, but tends to do so in

practice.

One goal of this thesis is a stronger unification of variational and discrete optimization models, and continuous and combinatorial optimization methods. Some initial relationships between the level set method and graph cuts are discussed in paper A. In paper D and E, continuous generalizations of max-flow and min-cut are proposed. As pointed out, variational models have many advantages over discrete models: they can much more accurately represent geometric entities like curve lengths, and hence avoid the metrication errors; continuous and convex optimization algorithms are much more easily parallelizable than combinatorial algorithms. Continuous formulations of max-flow and min-cut combine the advantages of both discrete and variational, they make possible the efficient computation of global solutions to certain non-convex variational models. Paper D deals with partitioning problem with two regions and Paper E deals with total variation regularized problems, where the data term is arbitrary and the unknown can take values in a predefined finite or continuous set. The discrete counterparts of such problems can be solved globally via max-flow and min-cut. Continuous generalizations of the discrete max-flow and min-cut models are proposed which are shown form strong primal-dual pairs, and provide convex optimization frameworks for solving the original non-convex variational problems globally. Convex formulations of partitioning problems in the continuous setting with two regions have also been presented in [19], and recently [70] presented a convex formulation of total variation regularized models with arbitrary data term. The max-flow and min-cut primal-dual treatment presented in paper D and E offers many algorithmic and analytical advantages.

For many problems, the computation of an exact solution is NP-hard in the discrete setting. This is particularly the case for minimal perimeter partitioning problems with several regions, which can be applied in e.g. image segmentation models. In the variational setting, the level set method has been the most widely used approach for solving such problems. As pointed out, the level set method converges rather slowly and may get stuck in local minima which deviates far from global solutions. In the discrete setting, the most popular optimization approaches are the alpha expansion and alpha beta swap algorithms, which solve a sequence of two region problems by graph cuts until convergence to an approximate solution of the original multi-region problem. A significant part of this thesis is devoted to new and better methods for solving partitioning problems with several regions, mainly in a variational setting. Instead of aiming to solve the problems exactly in all cases, approximate convex models (convex relaxations) are studied and developed, which can often compute global solutions of the original models, and otherwise good approximations. During the last two years, there has also been activity in this area from several research groups [92, 51, 67, 9]. In paper C, a relatively simple convex model [92, 51] is analyzed from a dual perspective. The dual model gives insight into the exactness of the relaxation. It is shown that global solutions of the original model can often be obtained from



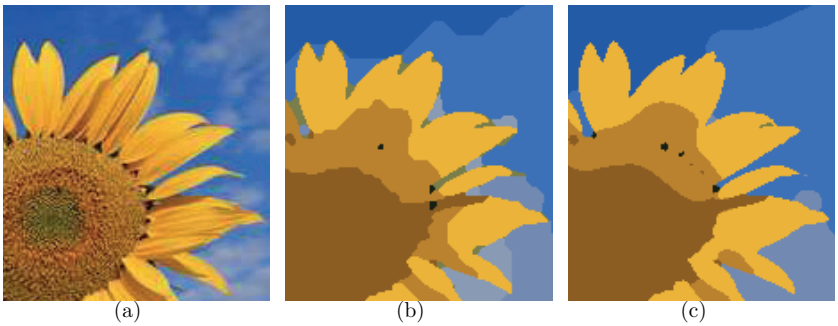
(a) (b) (c) (d)

Figure 1.2: (a) Input image, (b) ground truth, (c) the level set method, (d) new global approach. Each region is visualized with a distinct gray value.

a dual solution of the convex model, and otherwise close approximations. The convex models can approximate the original model more tightly by adding many additional constraints, as was done in [67], however this significantly increases the computational complexity. In paper F, a relaxation is proposed which is both the simplest and tightest of all the approaches, but its application is so far limited to problems with four regions. A formal proof that the dominance over [67] is also strict needs to be worked out, but arguments are presented to support that this is expected. In paper F it is also shown that the Chan-Vese model, a very popular level set formulation of partitioning problems, can be solved globally under the same assumptions, by generalizing the results of Paper B to the continuous setting. An illustration is given in Figure 1.2, where one seeks a partition of the leftmost image into 4 regions. As seen in Figure 1.2 (c), the level set method gets stuck in an inferior local minima, whereas the new approach from paper B and F computes a true global solution, Figure 1.2 (d). Another example is given in Figure 1.3. Alpha expansion (middle) produces an approximate solution with noticeable errors, such as the green misclassifications between the flower and the sky, and a bias towards the discrete grid. A result produced by the method in paper C is shown on the right, which is closer to a global solution and avoids the grid bias.

Another contribution of this thesis is the design of efficient algorithms, both in the discrete and continuous setting.

There has been much work on fast optimization algorithms for total variation regularized models with L_2 (ROF) or L_1 data fidelity terms, such as the "split bregman"/"augmented Lagrangian" [32, 88, 26] methods and algorithms based on the dual formulation [14]. Total variation also arise naturally in many other settings, like geometrical problems involving curve lengths or surface areas. For example, it occurs naturally in the formulation of minimal perimeter partitioning problems. For problems where the unknown is constrained to a finite set and general data term, for instance L_0 norm, one is also interested in fast algorithms. Some of the algorithms are applications of more general frameworks, like the projected gradient algorithm or the alternating direction method of multipliers.



(a) (b) (c)

Figure 1.3: Segmentation into 10 regions by using Potts model. (a) Input, (b) result graph cut-based alpha-expansion [7] with "8-neighborhood system", (c) result of the optimization approach in paper C. Each region is visualized with a distinct color. The full images and experiment are presented in Paper C.

The main novelties in these cases lie in reformulations of the problems in ways that make realization of these algorithms possible. In particular, it is shown that dual formulations of the problems offer many algorithmic advantages, such as the dual formulation of partitioning problems in paper C and the variational max-flow models. The dual problems deal with all the constraints of the primal variables implicitly and has a simpler structure, which makes computation easier. It is shown in numerical experiments, that the new algorithms are very efficient, and outperform alternative approaches.

In paper G, a fast algorithm is constructed for solving energy minimization problems with Euler's elastica as regularization term (curvature of all level lines) on a discrete grid with L quantized gray values. By solving a sequence of simpler subproblems, each of which can be solved efficiently in O(log(L)) by graph cuts, a local minimum of the Euler's elastica model can be obtained. Alternative approaches instead solve the Euler Lagrange equation, which is a 4th order PDE. Due to the small time step restrictions, such approaches are very slowly convergent. The proposed iterative splitting algorithm is also new from a general algorithmic point of view, and can be applied for other problems in the future.

A method for constructing open and non-orientable surfaces from point cloud data is proposed in paper J. By specifying appropriate boundary condition on a crust around the point cloud, it is shown that graph cuts can also be applied for such problems.

1.3 Outline of this thesis

This thesis is divided into Part I and Part II. Part I gives an overview of the field and describes related work in detail. Part II is composed of 10 research papers.

In part I, we will give an introduction to some of the most important energy minimization models in image processing and computer vision, some of their properties and some of the most effective methods for computing solutions. Both discrete and variational models will be covered and some relationships will be revealed.

Part I is organized as follows: In Chapter 2 some background material on optimization is provided. Section 2.1 deals with continuous optimization and Section 2.2 deals with integer and combinatorial optimization. Chapter 3 introduces fundamental variational models in image processing and computer vision. Chapter 4 gives an overview of existing continuous optimization methods. In Section 5.1 of Chapter 5, discrete energy minimization models, also called markov random fields, are introduced. In Section 5.2 we give an overview of the most effective combinatorial optimization algorithms.

Chapter 2

Background on Optimization

This section introduces some background material on optimization, which will be useful in this thesis. The introduction given here is very brief. For a detailed treatment of the subject of convex optimization, we refer the reader to [25, 74], nonlinear optimization to [64] and combinatorial optimization to [86].

2.1 Continuous optimization

A continuous optimization problem is in general the problem of finding the smallest value of an objective function $f : X \mapsto \mathbb{R}$ over a given subset $S \subset X$

$$\inf_{x \in S} f(x). \tag{2.1}$$

Usually X is a finite dimensional subset of $\mathbb{R}^M$. More generally, X can be any Banach space with norm $||.||_X$. If X is infinite dimensional, the problem (2.1) is called a variational problem. In this case X is usually some function space and the objective $f(x)$ in (2.1) is an energy functional which assigns a real value to every $x \in X$. For ease of expression, we will also refer to the objective f as a function in this chapter.

We say that a point x^* is a global minimizer (or minimum) of (2.1) if $f(x^*) \leq f(x)$ for all $x \in S$. A point $x^* \in X$ is said to be a local minimizer of (2.1) if for some $\epsilon > 0$, $f(x) \geq f(x^*)$ for all $x \in S$ such that $||x - x^*||_X < \epsilon$.

Often S is described in terms of a set of inequalities and equalities, in which case the optimization problem can be written

$$\inf_{x \in X} f(x), \quad \text{subject to} \begin{cases} g_i(x) & \geq 0, \quad i = 1, ..., m_1 \\ h_i(x) & = 0, \quad i = 1, ..., m_2 \end{cases} \tag{2.2}$$

If g, h and f are linear, (2.2) is often called a linear program. Linear programs can be solved by e.g. algorithms such as the simplex method and interior point methods when X is finite dimensional. In general, to find a global minimizer x^* of (2.2) is very difficult. Convexity is a crucial notion for characterizing problems that can be solved with reasonable computational effort.

2.1.1 Convexity

Definition 2.1 *A set $S \subset X$ is called convex if for any $x, y \in S$ all points on the line segment between x and y are contained in S, i.e. if $tx + (1-t)y \in S$ for every $t \in [0,1]$.*

A function f is convex if and only if the set of points above the graph of f is a convex set. Such a set is defined more formally as the epigraph of f.

Definition 2.2 *The epigraph of $f : X \mapsto [-\infty, \infty]$ is defined as*

$$Epi(f) = \{(x,t) \in X \times \mathbb{R},\ s.t.\ t \geq f(x)\} \tag{2.3}$$

Convex functions can also be defined as follows.

Definition 2.3 *We say a function $f : X \mapsto [-\infty, \infty]$ is convex if for any $x, y \in X$*

$$f(tx + (1-t)y) \leq t\, f(x) + (1-t)\, f(y) \tag{2.4}$$

for any $t \in [0,1]$.

A function is concave if and only if the set of points below the graph of f is a convex set, or equivalently, if $-f$ is a convex function.

We say that an optimization problem of the form (2.1) is convex if both the objective function f is a convex function, and the constraint set S is a convex set. If the inequality in (2.4) is strict, we say the problem is strictly convex. For strictly convex problems the global minimizer is unique. In general, there can exist several global minimizers. A fundamental property of convex optimization problems is that any local minimizer is also a global minimizer. Therefore, to solve convex problems it suffices to search for local minimizers. In the development of convex optimization theory, it will be useful to assume the objective function satisfies the following mild conditions.

Definition 2.4 *A function f is said to be proper if f is not identically $-\infty$ or ∞.*

Definition 2.5 *A function f is lower semi-continuous (l.s.c.) if and only if for any $x \in X$ and any sequence $\{x^n\}_{n=1}^{\infty}$ converging to x*

$$\liminf_{n\to\infty} f(x^n) = f(x) \tag{2.5}$$

and upper semi-continuous (l.s.c.) if and only if for any such sequence

$$\limsup_{n\to\infty} f(x^n) = f(x) \tag{2.6}$$

The Legendre-Fenchel conjugate is an important concept in the duality theory of convex optimization. We start by defining the dual space X^* of X.

Definition 2.6 *The dual space X^* of X is defined as all linear functionals on X, i.e.*

$$X^* = \{\phi \,:\, X \mapsto \mathbb{R},\ s.t.\ \phi\ is\ linear\} \tag{2.7}$$

For any $x \in X$ and $\phi \in X^*$, we define the bilinear mapping $\langle ., . \rangle \,:\, X \times X^* \mapsto \mathbb{R}$ as

$$\langle x, \phi \rangle = \phi(x) \tag{2.8}$$

A Hilbert space is a Banach space with an inner product $\langle ., . \rangle$. If X is a Hilbert space, the Riesz representation theorem (see e.g. [90]) says that any element in X^* can uniquely be "represented" by an element of $y \in X$. That is, for any $\phi \in X^*$, there exists a unique $y \in X$ such that

$$\phi(x) = \langle x, y \rangle, \quad \forall x \in X \tag{2.9}$$

It follows that if X is a Hilbert space, then X^* is isomorphic to X. Essentially, this means a Hilbert space is its own dual. Examples of Hilbert spaces include subspaces of $\mathbb{R}^M$ and L^2 (space of square integrable functions, the inner product between $f, g \in L^2$ is given by $\int fg$).

We can now define the Legendre-Fenchel conjugate of a functional on X.

Definition 2.7 *The Legendre-Fenchel conjugate f^* of a function f is defined for each $y \in X^*$ as*

$$f^*(y) = sup_{x \in X} \langle x, y \rangle - f(x) \tag{2.10}$$

We denote $f^{**} = f^{**}$. A fundamental result of convex duality is the following

Theorem 2.8 *Let f be a convex, proper and l.s.c. function on X, then $f^{**} = f$.*

The subgradient is a generalization of the gradient to non-differentiable convex functions. It is a crucial tool for stating optimality conditions of non-convex functions.

Definition 2.9 *The subgradient ∂f of a convex function $f \,:\, X \mapsto \mathbb{R}$, at $x \in X$ is defined as*

$$\partial f(x) = \{v \in X^* \,:\, f(y) \geq f(x) + \langle v, y - x \rangle \ \forall y \in X\} \tag{2.11}$$

Observe that if f is differentiable, the subgradient reduces to $\partial f = \nabla f$ if X is finite dimensional and the Gateaux differential [25] if X is infinite dimensional.

Minimizers of convex functions can be characterized in terms of the subgradient as follows.

Proposition 2.10 *Let $f \,:\, X \mapsto \mathbb{R}$ be convex, then $x \in \arg\min_{x \in X} f(x)$ if and only if*

$$0 \in \partial f(x) \tag{2.12}$$

This can be easily seen by inserting in (2.11): $f(y) \geq f(x) + \langle 0, y - x \rangle \ \forall y \in X$.

The following, the Legendre-Fenchel identity, is another useful result.

Proposition 2.11 *If $f : X \mapsto \mathbb{R}$ is convex, proper and l.s.c. then $y \in \partial f(x)$ if and only if $x \in \partial f^*(y)$*

see e.g. [75, 15].

Observe that any constrained minimization problem of the form (2.1) can be written as the unconstrained problem

$$\inf_{x \in X} f(x) + \mathbb{I}_S(x) \tag{2.13}$$

where $\mathbb{I}_S$ is the indicator function of the set S, which is given by

$$\mathbb{I}_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases} . \tag{2.14}$$

If S is a convex set and f is a convex function, one can easily check that $f + \mathbb{I}_S$ is a convex function. Therefore it suffices to state optimality conditions for unconstrained problems.

2.1.2 Saddle point problems

Another important class of optimization problems aim to maximize the objective function with respect to some variables while minimizing with respect to other variables. Such problems are often called saddle point problems, because the optimizers are saddle points of the objective function. Let X, Y be two Banach spaces and let $S \subset X$ and $C \subset Y$. Consider the problem

$$\sup_{x \in S} \inf_{y \in C} E(x, y), \tag{2.15}$$

where $E : X \times Y \mapsto \mathbb{R}$. Convex saddle point problems are important special cases. Under some more (mild) assumptions, the following important result can be shown for such problems

Theorem 2.12 *Assume S and C are convex, $E(., y)$ is convex and l.s.c. for any fixed $y \in C$ and $E(x, .)$ is concave u.s.c. for any fixed $x \in S$. Then the problem* (2.15) *can equivalently be defined by interchanging the* min *and* max *operators of* (2.15)*, i.e.*

$$\sup_{x \in S} \inf_{y \in C} E(x, y) = \inf_{y \in C} \sup_{x \in S} E(x, y) \tag{2.16}$$

see e.g. [25] for a proof. Saddle point problems arise naturally in connection with duality. Consider the primal problem

$$\inf_{x \in S} E^P(x), \tag{2.17}$$

and assume that for each $x \in S$, $E^P(x)$ can be written as

$$E^P(x) = \sup_{y \in C} E(x, y). \tag{2.18}$$

for some E. This is in particular the case for convex, proper and l.s.c. functions. Observe that under those assumptions $E^{P**} = E^P$. Therefore, by applying the formula for the convex conjugate twice, we obtain

$$E^P(x) = E^{P**}(x) = \sup_{y \in X^*} \langle x, y \rangle - E^{P*}(y) \tag{2.19}$$

which is exactly in the form (2.18).

The minimization problem (2.17) can therefore equivalently be formulated as the saddle point problem (2.15). The associated dual problem of (2.17) is defined as

$$\sup_{y \in C} E^D(y) = \sup_{y \in C} \{ \inf_{x \in S} E(x, y) \}. \tag{2.20}$$

If the conditions of theorem 2.12 are satisfied it follows immediately that

$$\inf_{x \in S} E^P(x) = \inf_{x \in S} \sup_{y \in C} E(x, y) = \sup_{y \in C} \inf_{x \in S} E(x, y) = \sup_{y \in C} E^D(y). \tag{2.21}$$

We then say that $E^P(.)$ and $E^D(.)$ form a strong primal-dual pair. For such a pair we have in general that

$$E^P(x) \leq E(x, y) \leq E^D(y), \quad \forall x \in S,\ y \in C. \tag{2.22}$$

If for some $x^* \in S$, $y^* \in C$

$$E^P(x^*) = E(x^*, y^*) = E^D(y^*), \tag{2.23}$$

we say that (x^*, y^*) is a saddle point. It follows that x^* is optimal to the primal problem (2.17) and y^* is optimal to the dual problem (2.20).

2.1.3 Lagrange duality

Saddle point problems also arise naturally in connection with lagrange duality. Consider the problem

$$\inf_{x \in S} f(x), \quad \text{s.t. } g_i(x) = 0,\ i = 1, ..., m_1, \forall x \in S \tag{2.24}$$

where $g_i : X \mapsto Y$, $i = 1, ..., m_1$ for some space Y,

The side constraints can be moved directly to the objective functional, by introducing Lagrange multipliers $\lambda_i \in Y^*$, $i = 1, ..., m_1$. Define the Lagrangian functional as

$$L(x; \lambda_1, ..., \lambda_{m_1}) = f(x) + \sum_{i=1}^{m_1} \langle \lambda_i, g_i \rangle \tag{2.25}$$

The primal minimization problem (2.24) can be expressed as a saddle point problem in L, i.e. as

$$\inf_{x \in S} \sup_{\lambda_1, ..., \lambda_{m_1}} L(x; \lambda_1, ..., \lambda_{m_1}) \tag{2.26}$$

Clearly, at an optimum (x^*, λ^*) of (2.26)

$$\partial_{\lambda_i} L = g_i = 0. \quad i = 1, ..., m_1$$

We also introduce the augmented Lagrangian functional

$$L_c(x; \lambda_1, ..., \lambda_{m_1}) = f(x) + \sum_{i=1}^{m_1} \langle \lambda_i, g_i \rangle + \frac{1}{2} c \sum_{i=1}^{m_1} ||g_i||^2 \tag{2.27}$$

An algorithm, known as the "augmented Lagrangian method", or "alternating direction method of multipliers" can be constructed for optimizing (2.26) when X is finite dimensional. It is especially powerful for problems where the unknown x can be separated into two or more groups, e.g. $x = (x_1, ..., x_k)$, and there exists subsets $S_i \subset X$ such that $x \in S$ if and only if $x_i \in S_i$ for all $i = 1, ..., k$. The algorithm consists of optimizing (2.27) for each x_i independently and computing ascent steps in λ_i alternatively as follows: For $n = 1, ...$

$$x_i^{n+1} = \underset{x_i \in S_i}{\arg\min}\, L_c(x_1^{n+1}, ..., x_{i_1}^{n+1}, x_i, x_{i+1}^n, ..., x_k^n, \lambda_1^n, ..., \lambda_{m_1}^n), \qquad i = 1, ..., k \tag{2.28}$$

$$\lambda_i^{n+1} = \lambda_i^n + c g_i, \qquad i = 1, ..., k, \tag{2.29}$$

see e.g. [36, 73]

2.1.4 Proximal operators

We introduce the proximal operator, which will be very useful in the design of algorithms for constrained optimization problems. Let P be a convex, proper, l.s.c. function. For any $x \in X$ and $\delta > 0$, the proximal operator $\text{prox}_{\delta G}$ is defined as

$$\text{prox}_{\delta G}(x) = \underset{y \in X}{\arg\min} \frac{1}{2} ||x - y||^2 + \delta G(y). \tag{2.30}$$

The above minimization problem has a unique solution, hence the proximal operator is well defined. By the first order optimality conditions, $\text{prox}_{\delta G}(x)$ is the unique point $y \in X$ which satisfies

$$0 \in \delta \partial G(y) + y - x \tag{2.31}$$

Therefore, y is given by

$$y = (I + \delta \partial G)^{-1}(x) \tag{2.32}$$

As one important special case, G is the indicator function of some set $S \in X$, i.e.

$$G(x) = \mathbb{I}_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}. \tag{2.33}$$

In this case $\text{prox}_{\mathbb{I}_S}$ is the orthogonal projection operator onto the set S, which maps x to the point of S of smallest euclidian length to x. The projection operator $\text{prox}_{\mathbb{I}_S}$ will also be denoted Π_S in this work.

The proximal operator can be applied in an algorithm for solving optimization problems composed as

$$\inf_{x \in X} f(x) + G(x), \tag{2.34}$$

where it is assumed prox_G is easily computed. For instance, if $G(x)$ is the indicator function of some set $S \subset X$, the above minimization problem is equivalent to

$$\inf_{x \in S} f(x).$$

The algorithm consists of alternatively computing one explicit step of gradient descent in f and one implicit gradient descent step in G as follows

$$x^{n+1} = (I + \delta G)^{-1}(x^n - \delta \nabla f(x^n)), \tag{2.35}$$

for $n = 1, ...$ Such algorithms are often called forward-backward splitting algorithms, see e.g. [57].

When G is the characteristic function $\mathbb{I}_S$ of some set $S \in X$, algorithm 2.35 reduces to the projected gradient algorithm [31, 53]

$$x^{n+1} = \Pi_S(x^n - \delta \nabla f(x^n)). \tag{2.36}$$

2.1.5 Relaxations

Relaxations are important tools for reformulating difficult problems in a simpler manner. Sometimes, it is possible to reformulate non-convex problems as convex problems.

Definition 2.13 *A problem*

$$\inf_{x \in \tilde{S}} \tilde{f}(x) \tag{2.37}$$

is a relaxation of the problem

$$\inf_{x \in S} f(x) \tag{2.38}$$

if $S \subseteq \tilde{S}$ and $\tilde{f}(x) \geq f(x)$ for all $x \in S$.

If f and S in the original problem are nonconvex, one can instead attempt to solve the relaxed problem, where $\tilde{f}$ and $\tilde{S}$ are convex. The question arises as to how minimizers of the relaxed problem (2.37) are related to minimizers of the original problem (2.38). Let x^* be optimal to the relaxed problem (2.37), then x^* is optimal to the original problem (2.38) if and only if $x^* \in S$ and $f(x^*) = \tilde{f}(x^*)$.

A relaxation is said to be exact if there exists an $x \in S$ such that

$$f(x) = \tilde{f}(x) = \inf_{x \in \tilde{S}} \tilde{f}(x) \tag{2.39}$$

If one can prove there exists $x \in S$ such that (2.39) holds for every input data f of the problem class, the relaxed problem (2.37) is said to be a convex formulation of the original problem (2.38).

2.2 Combinatorial and integer optimization

Integer optimization problems have the same form as (2.2) (most often f, g and h are linear) with the additional constraint that the unknown variables should take integer values, i.e.

$$x_i \in \mathbb{Z}, \quad i = 1, ..., M \tag{2.40}$$

In this case there are either finitely or countably many feasible solutions.

Closely related are combinatorial optimization problems. Given a finite set $\Omega = \{1, ..., M\}$ and a set of feasible subsets of Ω, denoted $\mathcal{F}$. A combinatorial optimization problem can in general be formulated as

$$\min_{S \subseteq \Omega : S \in \mathcal{F}} f(S) = \sum_{i \in S} c_i \tag{2.41}$$

Observe that there are also finitely many feasible solutions for combinatorial optimization problems. By defining binary (integer) variables x_i such that $x_i = 1$ iff $i \in S$ and $x_i = 0$ iff $i \notin S$, any combinatorial optimization problem can equivalently be formulated as a (linear) integer optimization problem.

In this thesis, we are also interested in infinite dimensional combinatorial optimization problem. That is, there are infinitely many variables, but each variable is constrained to take values in a finite set.

In the last section we saw that convexity is a crucial notion to describe whether a problem can be solved easily. However, convexity is not a necessary condition. Sometimes non-convex problems can be transformed into convex problems and therefore be solved globally, which we will see many examples of in this thesis. Integer and combinatorial optimization problems are obviously non-convex because of the non-convex integer constraints. Yet, efficient algorithms exist which can compute global solutions to some of these problems. Instead of convexity, a much stronger classification theory exists for combinatorial and integer optimization problems. This is the theory of computational complexity, which can also be used to classify other combinatorial problems besides optimization problems.

2.2.1 Computational complexity theory: a brief overview

We will give a brief overview of computational complexity theory without going into details, in order to give the reader a sense of the challenges associated with some of the problems encountered in this thesis. By the size of a problem, we shall mean the number of bits required for a "standard" representation of the input data. In image processing, the input data is typically an image. The size of a typical image processing problem is therefore the number of bits required to represent the input image (without any compression).

We say that a problem is polynomially solveable if there exists an algorithm which can compute an exact solution in a number of iterations which is bounded by a polynomial in the size of the problem, for any input data. The class of all polynomially solveable problems is denoted P.

Obviously, if a problem is not polynomially solveable, exact solution becomes very difficult in general. For certain input data, this means the best possible algorithm for solving the problem will need a number iterations which grows at least exponentially in the size of the problem. In image processing applications there are millions of pixels and the sizes of the problems are huge. In practice, this means infinite computation time is required to calculate an exact solution, even with hypothetical future computers of superior efficiency.

A large class of problems, denoted NP, are believed to not be polynomially solveable. All the problems in NP have in common, that if a polynomial algorithms is discovered for one of them, the same algorithm can be used to solve all the other problems in NP. If this was the case, one could immidiately conclude that $P = NP$. However, it is widely believed that $P \neq NP$, although neither has yet been proved. A consequence of $P \neq NP$ is that problems which belongs to NP are very difficult to solve exactly in general, and impossible for large problem sizes. Problems in NP are often called NP-hard.

2.2.2 Submodular objective functions

One class of polynomially solvable combinatorial optimization problems, is problems with submodular objective function. The function f in (2.41) assigns a real value to each subset $S \subseteq \Omega$. f is said to be submodular if and only if for any two subsets $A, B \subseteq \Omega$

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B). \tag{2.42}$$

If the objective function f satisfies the submodular condition (2.42), the optimization problems of the form (2.41) can be solved polynomially by a "greedy" algorithm.

2.2.3 Network flow problems

An important class of polynomially solveable combinatorial optimization problems is the network flow problems. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of vertices $\mathcal{V}$ and a set of directed edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The directed edge from vertex $v \in \mathcal{V}$ to $w \in \mathcal{V}$ is denoted (v, w). For each $v \in \mathcal{V}$ the neighborhood system $N^+(v)$ is defined as all $w \in \mathcal{V}$ such that $(v, w) \in \mathcal{E}$ and $N^-(v)$ is defined as all $w \in \mathcal{V}$ such that $(w, v) \in \mathcal{E}$.

Optimization problems can be defined over the graph $\mathcal{G}$. Of central importance in this thesis are the max-flow and min-cut problems. Assume the set $\mathcal{V}$ contains two distinguished vertices, the source $\{s\}$ and the sink $\{t\}$. For each edge $(v, w) \in \mathcal{E}$, let $c(v, w) \in \mathbb{R}$ be a predefined cost on (v, w).

Definition 2.14 *The min-cut problem is to find a partition of the vertices into two sets (V_s, V_t) which minimizes*

$$\min_{(V_s,V_t)} \sum_{(v,w)\in\mathcal{E}\,:\,v\in V_s,\ w\in V_t} c(v, w) \tag{2.43}$$

subject to

$$V_s \cap V_t = \emptyset, \quad V_s \cup V_t = \mathcal{V}, \quad s \in V_s, \quad t \in V_t \tag{2.44}$$

Closely related to the min-cut problem is the max-flow problem. For each $(v, w) \in \mathcal{E}$ define the "flow" variable $p(v, w) \in \mathbb{R}$. The weights $c(v, w)$ are upper bounds (capacities) on the the flows $p(v, w)$ for all $(v, w) \in \mathcal{E}$, i.e.

$$p(v, w) \leq c(v, w), \quad \forall (v, w) \in \mathcal{E} \tag{2.45}$$

In addition, flow conservation is required at each vertex except s and t

$$\sum_{w\in N^+(v)} p(v, w) - \sum_{w\in N^-(v)} p(w, v) = 0, \quad \forall v \in \mathcal{V} \tag{2.46}$$

The max-flow problem is to find the maximum amount of flow that can be pushed from $\{s\}$ to $\{t\}$ under the above flow constraints. The total amount of flow in the graph is at any time equal to the total amount of outgoing flow on the source edges. The max-flow problem can therefore be formulated as

Definition 2.15 *The max-flow problem is to find a flow function p which maximizes*

$$\max_{p} \sum_{v\in\mathcal{V}} p(s, v) \tag{2.47}$$

subject to (2.45) *and* (2.46).

The following is the most important theorem of network flow problems.

Theorem 2.16 *The maximum amount of flow* (2.47) = *the cost of the minimum cut* (2.43).

For proofs, see e.g. [29]. See also the paper D,E, where proofs are given in the infinite dimensional setting in some special cases. This result says that the max-flow problem and min-cut problem are strong dual problems. Therefore, if one of the problems is solved, a solution also becomes available to the other problem. The max-flow problem (2.47) has a structure which makes it very easily solveable. One algorithm for doing so, is the Ford-Fulkerson algorithm [29], which is essentially a greedy algorithm that iteratively pushes flow along paths from s to t until no such s-t path with at least one saturated edge exists. An efficient implementation specialized for problems in imaging and vision can be found in [6].

The variables of the max-flow problem (primal problem) are related to the variables in the min-cut problem (dual problem) as follows. Given a maximal flow function $p^* : \mathcal{E} \mapsto \mathbb{R}$. Define the graph $\mathcal{G}_r$ where each of the edges are assigned the residual capacity function $f - p^* : \mathcal{E} \mapsto \mathbb{R}$. A vertex $v \in \mathcal{V}$ is contained in the set V_s if there exists a path of edges of non-zero capacities from s to v in the graph $\mathcal{G}_r$. Similarly, a vertex $v \in \mathcal{V}$ is contained in the set V_t if there exists a path of edges of non-zero capacities from v to t in the graph $\mathcal{G}_r$. Vertices that satisfy neither of these criteria, can belong to either V_s or V_t.

Chapter 3

Fundamental Variational Models in Imaging and Vision

In this chapter we introduce some variational models in image processing and computer vision, which have become fundamentally important and occur over a wide range of applications. They all have in common a data fitting term and a regularization term. Section 3.1 - 3.3 deal with first order models, meaning the energy functional only involves the first order derivatives of the unknowns. Some higher order models are introduced in Section 3.4.

3.1 Pott's model and the piecewise constant Mumford-Shah model

Image segmentation with several regions is one of the core problems in image processing and computer vision. One would like to classify the image into regions, each representing an object. Each object should in some sense share the same intensity profile. As a prior assumption, the objects should possesses a certain regularity. They are not just random noise scattered around the image domain. In a variational framework, a simple, but very powerful, integration of such a prior is to seek objects with small boundary lengths (in 2D), or surface area (in 3D). Such formulations were first proposed [17] in a variational framework. The discrete equivalent of such models are called Pott's model. The Pott's model originates from statistical mechanics [72], but has become fundamentally important in image processing and computer vision, especially as energy minimization formulation of segmentation and grouping problems. Its first appearance in this connection dates back to [30], but in the discrete setting and a little different form, we refer the reader to Section 5.1.2 for more details. In this thesis we also refer to Pott's model as its the continuous counterpart.

If the image domain Ω is continuous, one seeks a partition $\{\Omega_i\}_{i=1}^n$ of Ω. Let $f_i(x)$ be the cost of assigning x to Ω_i, the Pott's model tries to assign each x

to the region Ω_i with smallest cost $f_i(x)$, while minimizing the lengths of the boundaries of the partitions Ω_i as spatial regularity.

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x)\,dx + \nu \sum_{i=1}^n |\partial\Omega_i| \qquad (3.1)$$
$$\text{s.t. } \cup_{i=1}^n \Omega_i = \Omega\,, \quad \Omega_k \cap \Omega_l = \emptyset\,, \forall k \neq l\,,$$

Here $|\partial\Omega_i|$ denotes the perimeter of the subdomain Ω_i (e.g. the length of the boundary $\partial\Omega$ in 2D). The data cost functions f_i should depend in some sense on the input image I. An important example is the Mumford-Shah data term

$$f_i(x) = |I(x) - c_i|^\beta, \quad i = 1, ..., n \qquad (3.2)$$

where $c_i \in \mathbb{R}$ are parameters associated with each Ω_i. The model (3.1) with data term (3.2) and $\beta = 2$ can minimized over the parameters c_i as proposed in [17]

$$\min_{\{\Omega_i\}_{i=1}^n, \{c_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} |I(x) - c_i|^\beta\,dx + \nu \sum_{i=1}^n |\partial\Omega_i| \qquad (3.3)$$
$$\text{s.t. } \cup_{i=1}^n \Omega_i = \Omega\,, \quad \Omega_k \cap \Omega_l = \emptyset\,, \forall k \neq l.,$$

The optimal values of $\{c_i\}_{i=1}^n$ will naturally be the mean intensity values inside Ω_i when $\beta = 2$ (usually $\beta = 2$) as observed in [17]. Often (3.3) is referred to as the piecewise constant Mumford-Shah model, although this is slightly incorrect, see Section 3.3.

Segmentation models of the form (3.1) are often called region based models, because the image is divided according to characteristics of each region. Another type of segmentation models distinguish the objects by using edge information of the image, i.e. they try to locate the object boundaries. In a variational framework, such models are often called geodesic active contour models [45, 11]. In [8, 82], it was shown that combination of region based and edge based models have many advantages. They only studied problems with two regions, but their approach can straight forwardly be extended to problems with several regions.

Let $C_i(s_i)$, $s_i \in [0, 1)$ be a parametrization of the curve $\partial\Omega_i$. A more general variant of (3.1) can be written as

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x)\,dx + \nu \sum_{i=1}^n \int_{s_i \in [0,1)} g(C_i(s_i))\,ds_i \qquad (3.4)$$
$$\text{s.t. } \cup_{i=1}^n \Omega_i = \Omega\,, \quad \Omega_k \cap \Omega_l = \emptyset\,, \forall k \neq l\,,$$

If the function $g(C_i(s_i)) = \nu$ for all $s_i \in [0, 1)$, then (3.4) reduces to (3.1). $g(s)$ be can also be an edge detector function, for instance

$$g(s_i) = \frac{1}{c|\nabla I_\sigma(C_i(s_i)|^2 + 1}\,, \qquad (3.5)$$

where I_σ is a smoothed version of the input image I. It forces the boundaries of the regions to be aligned along the edges of the image. The more general model (3.4) also fits directly into the optimization frameworks of this thesis.

Energy functionals of the form (3.1) also appear naturally in other problems, like surface reconstruction from point cloud data and 2D-3D surface reconstruction. More details about these two applications can be found in Paper J and Paper E respectively.

The Pott's model is natural for problems like image segmentation, because the regularization term does not favour any particular inclusion of the regions. On the other hand, computation of a solution raises some fundamental problems. The discrete counterpart of (3.1) can be be seen as a multiway cut problem, which is known to be NP-hard for $n > 2$ [21].

3.2 Total variation

Total variation is an important concept in inverse problems and numerical analysis. In image processing it is the only regularizer which is both convex and edge preserving. It also occurs naturally in implicit representations of geometrical problems involving curve lengths and surface areas, like the Pott's model.

The total variation (TV) of a function $u : \Omega \mapsto \mathbb{R}$ is defined as

$$\mathrm{TV}(u) = \int_\Omega |\nabla u| \, dx, \tag{3.6}$$

where $|\nabla u|$ is meant in the distributional sense. A more strict definition can be given by duality

Definition 3.1 *The total variation of a function $u \in L^1(\Omega)$, $\Omega \subset \mathbb{R}^N$, is defined as*

$$TV(u) = \sup_p \{ \int_\Omega u \operatorname{div} p \, dx \ : \ p \in C_c^\infty(\Omega, \mathbb{R}^N), \ |p(x)| \leq 1, \forall x \in \Omega \}. \tag{3.7}$$

Here $C_c^\infty(\Omega, \mathbb{R}^N)$ is the set of smooth vector fields on Ω with compact support. The norm $|.|$ in (3.6) and (3.7) is usually the 2-norm, $|p(x)|_2 = \sqrt{p_1^2(x) + ... + p_N^2(x)}$, in which case (3.6) and (3.7) are called isotropic total variation. If the 1-norm is used, $|p(x)|_1 = |p_1(x)| + ... + |p_N(x)|$, then (3.6) and (3.7) are called anisotropic total variation. If u is smooth, the equivalence between (3.6) and (3.7) follows directly by integration by parts. Functions with bounded total variation are said to belong to the space of bounded variation.

Definition 3.2 *The space of functions of bounded variation (BV) is defined as*

$$BV(\Omega) = \{ u \in L^1(\Omega) \ s.t. \ TV(u) < \infty \}. \tag{3.8}$$

Minimization problems involving total variation are examples of variational problems of the form (2.1), where X is typically $\mathrm{BV}(\Omega)$. Note that $\mathrm{BV}(\Omega)$ is a Banach space, but not a Hilbert space (inner product space).

3.2.1 Important properties

A crucial property of the BV space is that it allows for discontinuous functions, that is, the total variation of a discontinuous function is finite. For instance, it can be shown

Theorem 3.3 *Let u be the binary characteristic function of the set $S \subset \Omega \subset \mathbb{R}^N$, i.e.*

$$u(x) := \begin{cases} 1\,, & \text{if} \quad x \in S \\ 0 & \text{else} \end{cases}, \tag{3.9}$$

then

$$\int_\Omega |\nabla u|\, dx = |\partial S| \tag{3.10}$$

where $|\partial S| = \mathcal{H}^{N-1}(\partial S)$ is the $N-1$ - dimensional Hausdorff measure of ∂S, i.e. the curve length of ∂S for $N = 2$ and the surface area of ∂S for $N = 3$.

For a proof see e.g. [2, 27, 61]. Therefore, total variation will be one of the keys for representing geometrical minimization problems, like the Pott's model, in terms of functionals instead of subsets. Another important property of total variation is the coarea formula, which will be used extensively in this thesis.

Theorem 3.4 *Let $u \in BV(\Omega)$, denote by S^ℓ the ℓ- upper level set of u*

$$S^\ell = \{x \in \Omega \,:\, u(x) \geq \ell\} \tag{3.11}$$

and let u^ℓ denote its characteristic function

$$u^\ell(x) := \begin{cases} 1\,, & \text{if} \quad u(x) \geq \ell \\ 0 & \text{else} \end{cases}. \tag{3.12}$$

Then

$$TV(u) = \int_{-\infty}^{\infty} |\partial S^\ell|\, d\ell = \int_{-\infty}^{\infty} \int_\Omega |\nabla u^\ell|\, dx\, d\ell. \tag{3.13}$$

The equality (3.13) *is called the coarea formula.*

Proofs of can found in [28]. We also list some other important properties.

Theorem 3.5 *$TV(u)$ is*

- *convex*

- *one-homogeneous, that is, for any* $t > 0$

$$TV(tu) = t\,TV(u) \tag{3.14}$$

- *lower semi-continuous*

For proofs, see e.g [15]. The second property can be seen from

$$t\mathrm{TV}(u) = t \sup_p \{ \int_\Omega u \operatorname{div} p \, dx \ : \ p \in C_c^\infty(\Omega),\ |p(x)| \leq 1, \forall x \in \Omega \}$$

$$= \sup_p \{ \int_\Omega u \operatorname{div}(tp) \, dx \ : \ p \in C_c^\infty(\Omega),\ |p(x)| \leq 1, \forall x \in \Omega \}$$

$$= \sup_p \{ \int_\Omega u \operatorname{div} p \, dx \ : \ p \in C_c^\infty(\Omega),\ |p(x)| \leq t, \forall x \in \Omega \}$$

This expression can alternatively be written

$$t\mathrm{TV}(u) = \sup_{p \in C_t} \int_\Omega u \operatorname{div} p \, dx, \tag{3.15}$$

where the set C_t is defined as

$$C_t = \{ p \in C_c^\infty(\Omega, \mathbb{R}^N),\ \text{s.t. } |p(x)| \leq t, \forall x \in \Omega \}. \tag{3.16}$$

3.2.2 The ROF model and total variation regularized models

Discontinuity preservation is one of the motivations for the Rudin-Osher-Fatemi (ROF) model of image denoising [49], which uses total variation as a regularization term for reconstructing an image u from a noisy image I by minimizing

$$\min_u \frac{1}{2} \int_\Omega |u(x) - I(x)|^2 \, dx + \nu \int_\Omega |\nabla u| \, dx. \tag{3.17}$$

More general models can be formulated with data term

$$\min_u \int_\Omega \rho(u(x), x) \, dx + \nu \int_\Omega g(x) |\nabla u| \, dx. \tag{3.18}$$

For instance

$$\rho(u(x), x) = |u(x) - I(x)|^\beta \tag{3.19}$$

where $\beta = 2$ corresponds to the ROF model and $\beta = 1$ corresponds to the $TV-L1$ model. Both these models are convex, l.s.c. and coercive, therefore minimizers exist. The ROF model, being strictly convex, possesses a unique solution, which is in contrast to the $TV - L1$ model. Let us mention that models can also be

formulated that does not have the exact form of (3.18), such as the deblurring model

$$\min_u \frac{1}{2}\int_\Omega |Au - I|^2\,dx + \nu\int_\Omega |\nabla u|\,dx. \tag{3.20}$$

where A is a blurring operator.

There has been a lot of attention paid to numerical method for the ROF model and recently the $TV - L1$ model, some of the most successful algorithms for ROF are based on its equivalent dual formulation.

3.2.3 Dual formulation of the ROF model

The ROF model (3.17) has the form of a saddle point problem

$$\min_u \sup_{p\in C_\nu} \frac{1}{2}\int_\Omega |u(x) - I(x)|^2\,dx + \int_\Omega u \operatorname{div} p\,dx. \tag{3.21}$$

A dual formulation can be derived which only depends on the variable p, see e.g. [10, 20, 14]. Simpler and faster algorithms can be derived based on the dual formulation. Observe that the min and sup operators can be interchanged by the minimax theorem (see e.g., [25] Chapter 6, Proposition 2.4), resulting in

$$\sup_{p\in C_\nu} \min_u \frac{1}{2}\int_\Omega |u(x) - I(x)|^2\,dx + \int_\Omega u \operatorname{div} p\,dx. \tag{3.22}$$

For any p, the optimal u is given by the formula

$$u = I - \operatorname{div} p. \tag{3.23}$$

By inserting in (3.22), we obtain the dual formulation of the ROF model

$$\sup_{p\in C_\nu} \frac{1}{2}\int_\Omega |I(x)|^2 - (I - \operatorname{div} p)^2\,dx,$$

which, by ignoring the constant term, is equivalent to

$$\sup_{p\in C_\nu} -\frac{1}{2}\int_\Omega (I - \operatorname{div} p)^2\,dx. \tag{3.24}$$

In this work, we also derive dual formulations of other important models, where total variation is involved, like total variation with general data term (3.18).

3.2.4 Discretization

In order to compute a solution numerically, discretization is necessary. There are several ways to discretize energy functionals involving total variation. We state here the most common approach based on finite differences, which will be useful

in the review of related work. In the papers, discretizations based on the mimetic finite difference method are also applied. More details about this scheme can be found in paper C and [41, 40]. We restrict ourselves to 2-D grids for simplicity, generalizations to N-D should be straight forward. Define the discrete grid Ω_d as

$$\Omega_d = \{(ih, jh) \,:\, i = 1, ..., N_1,\ j = 1, ..., N_2\} \tag{3.25}$$

where h is the grid size. Let $u_{i,j}$ denote the discrete approximation of u at the point (ih, jh). The space of all discrete images is defined as $X = \mathbb{R}^{N_1 N_2}$, which is finite dimensional and therefore a Hilbert space. A forward discretization of the gradient $\nabla \,:\, X \mapsto X \times X$ can be derived as

$$(\nabla u)_{i,j} = \big((D^x u)_{i,j},\ (D^y u)_{i,j}\big)^T \tag{3.26}$$

where

$$(D^x u)_{i,j} = u_{i+1,j} - u_{i,j}/h, \qquad i = 1, ..., N_1 - 1,\ j = 1, ..., N_2, \tag{3.27}$$
$$(D^x u)_{i,j} = 0 \qquad i = N_1,\ j = 1, ..., N_2 \tag{3.28}$$
$$(D^y u)_{i,j} = u_{i,j+1} - u_{i,j}/h, \qquad i = 1, ..., N_1,\ j = 1, ..., N_2 - 1, \tag{3.29}$$
$$(D^y u)_{i,j} = 0, \qquad i = 1, ..., N_1,\ j = N_2. \tag{3.30}$$

A discrete approximation of the isotropic total variation can then be expressed as

$$\mathrm{TV}_d(u) = h^2 \sum_{i,j} \sqrt{\Big|\frac{u_{i+1,j} - u_{i,j}}{h}\Big|^2 + \Big|\frac{u_{i,j+1} - u_{i,j}}{h}\Big|^2} \tag{3.31}$$

and the anisotropic total variation can be written

$$\mathrm{TV}_d(u) = h^2 \sum_{i,j} \Big|\frac{u_{i+1,j} - u_{i,j}}{h}\Big| + \Big|\frac{u_{i,j+1} - u_{i,j}}{h}\Big| \tag{3.32}$$

The divergence div is defined as the adjoint ∇^* of the operator ∇ by

$$\langle \nabla u, p\rangle = \sum_{i,j} \nabla u_{i,j}^T p_{i,j} = \sum_{i,j} u_{i,j} (\nabla^* p)_{i,j} = \sum_{i,j} u_{i,j} (\operatorname{div} p)_{i,j} = \langle u, \nabla^* p\rangle \tag{3.33}$$

for all $u \in X$ and $p = (p^x, p^y) \in X \times X$. The expression for the adjoint applied to p is given by

$$(\operatorname{div} p)_{i,j} = (\nabla^* p)_{i,j} = \frac{p^x_{i,j} - p^x_{i-1,j} + p^y_{i,j} - p^y_{i,j-1}}{h} \tag{3.34}$$

where $p^x_{0,j} = 0,\ j = 1, ..., N_2$ and $p^y_{i,0} = 0,\ i = 1, ..., N_1$. For ease of notation we will often use the continuous operators $\int, \nabla, \operatorname{div}$ as a replacement of the corresponding discrete operators throughout this thesis.

3.3 The general Mumford-Shah model

The Mumford-Shah model [63] is another discontinuity preserving model, besides the ROF model, which can be used for image restoration, primarily image denoising. The Mumford-Shah model seeks an approximation image u which is smooth everywhere except for a discontinuity set Γ. Regularity is also imposed on the discontinuity set Γ itself.

$$\min_{\Gamma,u} E_\alpha^{MS}(\Gamma, u) = \int_\Omega |u(x) - I(x)|^\beta dx + \alpha \int_{\Omega\setminus\Gamma} |\nabla u|^2 \, dx + \nu \int_\Gamma ds. \tag{3.35}$$

The Mumford-Shah model is in some ways more ideal than the ROF model. The energy functional does not depend on the sizes of the discontinuities. The staircasing artifacts are also avoided. However, from a computational perspective it is much harder. The energy functional of (3.36) is obviously non-convex. Even worse, a combinatorial variant of (3.36) is NP-hard, see Section 5.1.4.

In the limit as the smoothness parameter α goes to infinity, (3.36) becomes

$$\min_{\Gamma,u} \lim_{\alpha\to\infty} E_\alpha^{MS}(\Gamma, u). \tag{3.36}$$

which is often called the piecewise constant Mumford-Shah model. As α goes to infinity, the solution u will naturally be constant everywhere except for the discontinuity set Γ. In consequence, the discontinuity set Γ will be a set of closed curves and will separate the image domain Ω into several regions $\{\Omega_i\}_{i=1}^N$. Within each Ω_i, it can be easily seen that u will take the mean value of I

$$u|_{\Omega_i} = \frac{\int_{\Omega_i} I(x)\, dx}{\int_{\Omega_i} dx}, \quad i = 1, ..., N. \tag{3.37}$$

The Mumford-Shah model has a close connection to the Chan-Vese model (3.3), in which case the optimal parameters c_i will be also be mean values within each Ω_i, $i = 1, ..., n$. One difference is that the number of regions is fixed in advance in (3.3), while the number of regions are unknown in the piecewise constant Mumford-Shah model (3.36).

3.4 Higher order models

Higher order models have recently been developed, mainly for the application of noise removal and inpainting. The main motivation is to yield smoother results and avoid artifacts, like staircasing, which is often associated with first order variational models. The first such higher order models were based on Euler's elastica [62, 60]. You and Kaveh [89] studied 4th order diffusion equations, where the diffusion coefficient depends on the laplacian of the unknown. Lysaker et al.

[58] proposed a higher order extension of total variation, the corresponding Euler-Lagrange equation is also a fourth order PDE. In [59] a method was proposed where the total curvature of the image acted as a regularization term, by using total variation to smooth the normal vectors. See also [16, 33, 91, 95] for related works. Higher order models that minimize some functionals of the curvature of the image [18], [76], [83] and [95] are especially important. The main disadvantage of these models is the high computational cost of minimizing the energies. We will concentrate especially on the Euler's elastica model [62, 60]. In Paper G a new algorithm is proposed for minimizing the energy in this model, which is much more efficient than alternative PDE approaches [18].

3.4.1 Euler's elastica

Euler's elastica as an image processing model was first introduced in [62]. It was studied further in [62, 60]. PDE based optimization methods were developed in [18]. In image processing, the model can be formulated as the minimization of Euler's elastica of all level curves of the image. The Euler's elastica of a curve Γ is given by the energy

$$E(\Gamma) = \int_{\Gamma} (a + b \cdot |k|^{\beta}(s)) ds, \tag{3.38}$$

where a and b are two parameters and k is the curvature of Γ at position s. By setting $b = 0$, $E(\Gamma)$ measures the total length of the curve. If $a = 0$, $E(\Gamma)$ measures the total curvature of the curve. Therefore, the Euler's elastica of all level curves of an image u can be written as:

$$\int_{\ell=0}^{L} \int_{\gamma_\ell : u=\ell} (a + b \cdot |k|^{\beta}(s)) ds d\ell. \tag{3.39}$$

The power β can be set to $\beta = 1$ as in [60], or $\beta = 2$ as in [18]. The choice of $\beta = 1$ allows cracks (sharp corners) of the level curves. By setting $\beta = 2$ differentiation is much easier, such that the Euler-Lagrange equations do not get too complicated.

Note that the curvature of the level curve can be expressed as a function of u by

$$k(u) = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right). \tag{3.40}$$

Using this fact, (3.39) can be expressed more simply as

$$\int_{\Omega} \left(a + b \left| \nabla \cdot \frac{\nabla u}{|\nabla u|} \right|^{\beta} \right) |\nabla u| \, dx. \tag{3.41}$$

For the image denoising application, Euler's elastica can be used as a regularization term to approximate the noisy image I with an image u by minimizing

$$\min_u E^{\mathrm{EL}}(u) = \int_\Omega |u - I|^2 \, d\mathbf{x} + \int_\Omega \left(a + b \left| \nabla \cdot \frac{\nabla u}{|\nabla u|} \right|^\beta \right) |\nabla u| dx. \tag{3.42}$$

Chapter 4

Continuous Optimization Approaches for Variational Models in Imaging and Vision

In this chapter, we give an overview of existing optimization approaches for the variational problems introduced in Chapter 3.

4.1 Level set methods for partitioning problems

The first attempts to solve segmentation and partitioning problems in a variational framework was via the level set method [65]. The level set method was originally developed as a computational tool for solving Hamilton-Jacobi equations involving evolving fronts and shocks. It has later been applied with great success in image processing, computer vision and computer graphics. Given a subset $S \subset \Omega \subset \mathbb{R}^N$, where Ω is a bounded domain in $\mathbb{R}^N$. The boundary of S is a closed orientable manifold in $\mathbb{R}^N$ denoted ∂S. If $N = 2$, then ∂S is a curve, if $N = 3$ then ∂S is a two dimensional surface. The boundary ∂S can be represented as the zero level set of a "level set function" ϕ defined in $\mathbb{R}^N$ as follows

$$\begin{cases} \phi(x) > 0, & \text{for } x \in S, \\ \phi(x) < 0, & \text{for } x \in \Omega \backslash S, \\ \phi(x) = 0, & \text{for } x \in \partial S, \end{cases} \tag{4.1}$$

For numerical stability reasons, ϕ should not be too steep or too shallow. A usual choice of ϕ is the signed distance function.

$$\phi(x) = \begin{cases} \text{dist}(x, \partial S), & \text{for } x \in S, \\ -\text{dist}(x, \partial S), & \text{for } x \notin S, \\ 0, & \text{for } x \in \partial S. \end{cases} \tag{4.2}$$

4.1.1 Chan-Vese model two regions

The Chan-Vese model (3.3) restricted to two regions can be written in terms of a level set function ϕ by requiring $\phi(x) > 0$ for $x \in \Omega_1$, $\phi(x) < 0$ for $x \in \Omega_2$ and $\phi(x) = 0$ for $x \in \partial\Omega_1 = \partial\Omega_2$ as

$$\min_{\phi,c_1,c_2} \int_\Omega \{H(\phi)|c_1 - u^0|^\beta + (1 - H(\phi))|c_2 - u^0|^\beta\} dx + 2\nu|\nabla H(\phi)|\, dx \tag{4.3}$$

where $H(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ is the Heaviside function: $H(s) = 0$ if $x < 0$ and $H(s) = 1$ if $x \geq 0$. $H(\phi)$ is the characteristic function of the region Ω_1, therefore the last term of (4.3) equals $2|\partial\Omega_1|$, by the result (3.10). The functional in (4.3) is highly nonconvex due to the heaviside function H. To solve the minimization problem numerically, [17] proposed to solve the Euler Lagrange equation of (4.3) by gradient descent. The gradient descent equation can be derived as

$$\phi_t = \delta(\phi)\Big\{ -\{(c_1 - u^0)^2 - (c_2 - u^0)^2\} + \nu\, \nabla \cdot \left(\frac{\nabla\phi}{|\nabla\phi|}\right)\Big\}, \tag{4.4}$$

The last term of (4.4) is the mean curvature of ∂S. For this reason, the evolution (4.4) is often called the mean curvature flow if the data terms are ignored.

In addition, (4.3) can be optimized for c_1, c_2 by updating the means

$$c_1 = \frac{\int_\Omega H(\phi(x;t)\, I(x)\, dx}{\int_\Omega H(\phi(x;t)\, dx}, \quad c_2 = \frac{\int_\Omega (1 - H(\phi(x;t))\, I(x)\, dx}{\int_\Omega (1 - H(\phi(x;t))\, dx}, \tag{4.5}$$

after each time step of (4.4).

4.1.2 Multiphase level set method of Zhao et al.

There has been several attempts to extend the level set approach to problems with multiple regions. Zhao et al. [37] did not study minimization problems of the form (3.1), but developed a level set method for motion of multiple junctions under mean curvature flow. However, their approach can also be extended to minimization problems such as (3.1).

Any partition Ω_i, $i = 1, ..., n$ of the domain Ω satisfying the constraints in (3.1), can be described in terms of n level set functions ϕ_i, $i = 1, ..., n$ as follows

$$\begin{cases} \phi_i(x) > 0, & \text{for } x \in \Omega_i, \\ \phi_i(x) < 0, & \text{for } x \in \Omega \backslash \Omega_i, \\ \phi_i(x) = 0, & \text{for } x \in \partial\Omega_i, \end{cases} \tag{4.6}$$

for $i = 1, ..., n$. The vacuum and overlap constraints (3.1) of the subsets Ω_i, $i = 1, ..., n$, can be imposed by placing the following constraint on ϕ_i, $i = 1, ..., n$

$$\sum_{i=1}^{n} H(\phi_i(x)) = 1, \quad \forall x \in \Omega. \tag{4.7}$$

The Potts model (3.1) can then be written in terms of ϕ_i, $i = 1, ..., n$ as

$$\min_{\phi} \ \nu \sum_{i=1}^{n} \int_{\Omega} |\nabla H(\phi_i)| + H(\phi_i(x)) f_i(x) \, dx \tag{4.8}$$

subject to (4.7). Gradient descent equations can be derived by introducing a lagrange multiplier for the constraint (4.7), the details are omitted here. Since both the objective function and constraint is non-convex, this approach can of course not compute global solutions. This representation is especially relevant for the material in Paper C.

4.1.3 Multiphase level set method of Vese and Chan

In [84] L. Vese and T. Chan proposed another level set approach for representing multiple regions, which has become very popular. By using $m = \log_2(n)$ level set functions, denoted $\phi^1, ..., \phi^m$, n region could be represented. An important special case is the representation of 4 regions by two level set functions ϕ^1,ϕ^2, as in Table 4.1. The energy functional could then be written

$$\min_{\phi^1,\phi^2,c_1,...,c_4} E_{CV}(\phi^1, \phi^2, c_1, ..., c_4) = \nu \int_{\Omega} |\nabla H(\phi^1)| + \nu \int_{\Omega} |\nabla H(\phi^2)| \tag{4.9}$$

$$+ \int_{\Omega} \{H(\phi^1)H(\phi^2)|c_1 - u^0|^\beta + H(\phi^1)(1 - H(\phi^2))|c_2 - u^0|^\beta)$$

$$+(1 - H(\phi^1))H(\phi^2)|c_3 - u^0|^\beta + (1 - H(\phi^1))(1 - H(\phi^2))|c_4 - u^0|^\beta\} dx.$$

(4.9) is often referred to as the multiphase Chan-Vese model. Observe that (4.9) does not correspond exactly to the Potts model (3.1), because the regularization term in (4.9) "overcounts" two of the boundaries. The length of the boundaries between Ω_1, Ω_4 and between Ω_2, Ω_3 are counted twice, while the remaining 4 different boundaries are counted twice. This small approximation usually does not have a large impact on the results and the model (4.9) has become very popular for its simplicity.

The energy functional of (4.9) is non-convex due to the heaviside functions and multiplications between ϕ^1 and ϕ^2. The traditional minimization approach of solving the gradient descent equations can therefore easily get stuck in a local minima. Furthermore, the numerical solution of the gradient descent PDEs is expensive computationally.

4.1.4 Piecewise Constant Level set Method

In [56, 55, 54], the piecewise constant level set method was proposed, and applied to the Mumford-Shah model. The use of heaviside functions can be avoided by instead requiring the level set function itself to take integer values. Let $\{\Omega_i\}_{i=1}^n$

$x \in$ phase 1 iff	$\phi^1(x) > 0, \phi^2(x) > 0$
$x \in$ phase 2 iff	$\phi^1(x) > 0, \phi^2(x) < 0$
$x \in$ phase 3 iff	$\phi^1(x) < 0, \phi^2(x) > 0$
$x \in$ phase 4 iff	$\phi^1(x) < 0, \phi^2(x) < 0$

Table 4.1: Representation of four phases by two level set functions ϕ^1 and ϕ^2.

be a partition of the Ω into n regions. Any such partition can be described by a piecewise constant level set function ϕ as follows

$$\phi = i \quad \text{in} \quad \Omega_i \quad \text{for} \ \ i = 1, 2, ..., n. \tag{4.10}$$

Note that all interphases are represented by discontinuities in ϕ. The Mumford-Shah functional can then be written in terms of ϕ as

$$E(\mathbf{c}, \phi) = \int_\Omega (u - u^0)^2 dx + \nu \sum_{i=1}^{n} \int_\Omega |\nabla \psi_i| dx, \tag{4.11}$$

where $u = \sum_{i=1}^n c_i \psi_i$, and ψ_i is the characteristic function of Ω_i. It can be expressed as a polynomial in ϕ as

$$\psi_i = \frac{1}{\alpha_i} \prod_{j=1 j \neq i} (\phi - j) \ \text{ with } \ \alpha_i = \prod_{k=1 k \neq i} (i - k). \tag{4.12}$$

Due to the high nonlinearity and nonconvexity of ψ_i, (4.13) is often approximated as

$$E(\mathbf{c}, \phi) = \int_\Omega (u - u^0)^2 dx + \nu \int_\Omega |\nabla \phi| dx, \tag{4.13}$$

In order to force a solution taking only integer values, the following constraint was imposed

$$K(\phi) = \prod_{i=1}^{n} (\phi - i) \tag{4.14}$$

The constrained optimization problem (4.13)-(4.14) can be solved by the augmented Lagrangian method as in [56, 55, 54], by introducing a lagrange multiplier for the constraint (4.14). Some attempts to speed up the computation can be found in [79].

4.2 Convex formulation of partitioning problems with two regions

The level set methods cannot in general find global solutions due to the non-convexity of their formulations. In practice they may easily get trapped in local

minima, unless the initialization is very good, even for simple images. For problems with two regions, a convex formulation was given in [19, 78] which made it possible to compute global solutions. Observe that by requiring $\phi_1(x) = 1$ for $x \in \Omega_1$ and $\phi_2(x) = 0$ for $x \in \Omega_2$, like in the piecewise constant level set method, the Potts model can be written

$$\min_{\phi} \nu \int_{\Omega} |\nabla \phi| dx + \{\phi f_1 + (1 - \phi) f_2\} dx. \tag{4.15}$$

such that

$$\phi(x) \in \{0, 1\}, \quad \forall x \in \Omega \tag{4.16}$$

This problem is non-convex since the binary set (4.16) is non-convex. A convex relaxation can be formulated by instead minimizing (4.15) over the larger convex set

$$\phi(x) \in [0, 1], \quad \forall x \in \Omega. \tag{4.17}$$

In [19, 78] it was shown that such a relaxation is always exact. That is, there exists binary solutions of the relaxed problem (4.15) subject to (4.17), which are also globally optimal to the original problem (4.15) subject to (4.16). Such binary solutions can be obtained by thresholding the solutions of the relaxed problem as the next result shows.

Theorem 4.1 *Let ϕ be globally optimal to the relaxed problem* (4.15) *subject to* (4.17). *For any $t \in (0, 1]$ define ϕ^t by*

$$\phi^t(x) = \begin{cases} 1, & \text{if } \phi(x) \geq t \\ 0, & \text{else.} \end{cases} . \tag{4.18}$$

Then, for almost every $t \in (0, 1]$, ϕ^t is a global solution of (4.15) *subject to* (4.16).

The proof can be found in [19].

4.3 Convex relaxation approaches for partitioning problems with more than two regions

Partitioning problems with more than two regions are NP-hard in general, as mentioned in section 3.1, and which will be discussed in more detail in Section 5.1.2. In the last two years, and simultaneously with much of the work in this thesis, attempts to convexify (3.1) have been proposed [92, 51, 67, 9]. One first thing one might try, is to apply the same binary relaxation approach to the multiphase level set formulation of Chan and Vese from Section 4.1.3. The model (4.9) can be written with in terms of $\phi^1, \phi^2 \in D = \{\phi \mid \phi : \Omega \mapsto \{0, 1\}\}$ as

$$\min_{\phi^1, \phi^2 \in D, c_1, ..., c_4} E_{CV}(\phi^1, \phi^2, c_1, ..., c_4) = \nu \int_{\Omega} |\nabla \phi^1| dx + \nu \int_{\Omega} |\nabla \phi^2| dx + E^{data}(\phi^1, \phi^2), \tag{4.19}$$

where

$$E^{data}(\phi^1,\phi^2) = \int_\Omega \{\phi^1\phi^2|c_1-u^0|^\beta + \phi^1(1-\phi^2)|c_2-u^0|^\beta)$$
$$+(1-\phi^1)\phi^2|c_3-u^0|^\beta + (1-\phi^1)(1-\phi^2)|c_4-u^0|^\beta\}dx.$$

The connections between ϕ^1, ϕ^2 and the regions are given in Table 4.2. In fact, such a formulation appeared in [55], as a variant of the piecewise constant level set method (PCLSM). In order to solve the equations, the constraints were represented with non-convex polynomials and lagrange multipliers in [55]. One could attempt to replace the non-convex constraint $\phi^1, \phi^2 \in \{\phi \mid \phi : \Omega \mapsto \{0,1\}\}$ with the convex constraints $\phi^1, \phi^2 \in \{\phi \mid \phi : \Omega \mapsto [0,1]\}$ and then solve the problem. However, the objective function of (4.19) is not convex in ϕ^1, ϕ^2, therefore such an approach will not produce a global solution.

As suggested in [92, 51], one could instead attempt the same binary relaxation for the multiphase representation of Zhao et al. [37] from Section 4.1.2, i.e. minimize (4.8) subject to (4.7). By writing (4.8) subject to (4.7) with binary functions ϕ_i, $i = 1, ..., n$ instead of level set functions, one obtains

$$\min_\phi \ \nu \sum_{i=1}^n \int_\Omega |\nabla\phi_i| + \phi_i(x) f_i(x)\, dx, \tag{4.20}$$

subject to

$$\sum_{i=1}^n \phi_i(x) = 1, \quad \forall x \in \Omega, \tag{4.21}$$

and the binary constraints

$$\phi^1, ..., \phi^n \in \{\phi \mid \phi : \Omega \mapsto \{0,1\}\}. \tag{4.22}$$

If one were to relax the binary constraints (4.22) with the convex constraint

$$\phi^1, ..., \phi^n \in \{\phi \mid \phi : \Omega \mapsto [0,1]\}. \tag{4.23}$$

the overall problem would be convex. The question which now arises, is how minimizers of the relaxed problem (4.20) subject to (4.21) and (4.23) are related to the original problem. There are no simple thresholding scheme, like theorem 4.1 for problems with two regions, to obtain binary solutions. In [92, 51], it was proposed to use the indicator function of the largest component ϕ_i as the final ϕ, i.e. the thresholded solution $\tilde{\phi}$ was selected as

$$\tilde{\phi}_k(x) \ = \ \begin{cases} 1 & \text{if } k = \arg\max_{i=1,...,n} u_k^* \\ 0 & \text{otherwise} \end{cases} \quad . \tag{4.24}$$

Such a binary $\tilde{\phi}$ may not be a global optimum of the original problem, but can be accepted as an approximate solution. In paper C, the problem is analyzed from a

dual perspective. A thresholding scheme is derived for producing solutions of the original problem from a dual solution of the relaxed problem. The thresholded solution is global under some conditions and close to global otherwise. An efficient algorithm is also proposed based on the dual formulation, which is shown to produce final binary solutions significantly faster than other alternatives. In paper H, the dual model is reformulated as continuous max-flow problem, and a continuous max-flow algorithm is developed.

In [67] another attempt to convexify the problem (3.1) was made. This approach can in fact be shown to be tighter than (4.20) subject to (4.21) and (4.23), but is much more difficult to handle computationally. Especially since the number of side constraints grow quadratically in n. The proposed optimization algorithm requires to project the variables onto the feasible set for every iteration of the main algorithm. Since no closed form solution exists for such a complex projection, it must be computed by an iterative algorithm. As part of Paper E, a faster algorithm is proposed which avoid to compute projections by an iterative algorithm, but is still much slower than the other approaches due to the quadratic complexity of the problem. The details can be found in paper E.

In paper F, a convex formulation of the level set representation of the Chan and Vese (4.9) model with 4 regions is proposed. This approach is guaranteed to produce a global solution provided the data term satisfies some mild conditions. Since there is a slight simplification of the length term in this model, the optimization problem is not NP-hard. Furthermore, a new convex relaxation of problems with Potts regularization (i.e. without such a simplification) is proposed. This relaxation is both the tightest and most simple that exists, but is so far limited to problems with 4 regions.

It should be mentioned that some other work on fast algorithms for the above relaxations (except the one in paper F), have appeared after and simultaneously with the papers in this thesis. A direct comparison have therefore not yet been made with these. In [50, 52] algorithms for 4.20 and the relaxation [67] were proposed based on the same idea as the augmented lagrangian / split-bregman method for the ROF model (to be discussed in Section 4.4.2). It was shown the expensive iterative algorithm for computing projections as in [67] could be avoided by instead introducing more lagrange multipliers. Although no direct comparison is made, in paper D it was shown that the new algorithms based on continuous max-flow converged faster than the split-bregman algorithm applied to convex formulations of partitioning problems with two regions.

$x \in$ phase 1 iff	$\phi^1(x) = 1, \phi^2(x) = 1$
$x \in$ phase 2 iff	$\phi^1(x) = 1, \phi^2(x) = 0$
$x \in$ phase 3 iff	$\phi^1(x) = 0, \phi^2(x) = 1$
$x \in$ phase 4 iff	$\phi^1(x) = 0, \phi^2(x) = 0$

Table 4.2: Representation of four phases by binary level set functions.

4.4 Numerical methods for total variation minimization

There has recently been much efforts on designing efficient algorithms for minimizing the energy in the ROF model and TV-L1 model. Some of these approaches have been further applied in algorithms for solving more complex problems with more constraints on the unknowns or more general data terms. We will briefly discuss the most important approaches. In this section, we assume that the functions involved are discretized. The operators $\int, \nabla, \operatorname{div}$ are used as replacements for discrete integrals, gradients and sums for ease of notation. The algorithms allow for many discretization schemes of these operators, for instance the discretization schemes discussed in Section 3.2.4 can be applied.

4.4.1 Algorithms for ROF based on the Dual formulation

An algorithm for the ROF model can be constructed based on its equivalent dual formulation by the projected gradient method. Observe that the dual formulation of the ROF model can be written

$$\max_{p \in C_\nu} -\frac{1}{2}|| \operatorname{div} p - I||^2 + G(p), \tag{4.25}$$

where

$$G(p) = \mathbb{I}_{C_\nu} = \begin{cases} 0 & \text{if } |p| \leq \nu \\ \infty & \text{else} \end{cases} . \tag{4.26}$$

The forward-backward splitting algorithm (2.35) then reduces to: for $k = 1, ...$

$$p^{k+1} = \Pi_{C_\nu}(p^k + \delta \nabla(\operatorname{div} p^k + I)) \tag{4.27}$$

Such an algorithm can alternatively be written

$$p^{k+1} = \frac{p^k + \delta \nabla(\operatorname{div} p^k + I)}{\max(\nu, |p^k + \delta \nabla(\operatorname{div} p^k + I)|_2)} \tag{4.28}$$

Chambolle [14] derived a little different algorithm based on the dual formulation as follows

$$p^{k+1} = \frac{p^k + \delta \nabla(\operatorname{div} p^k + I)}{\nu + |p^k + \delta \nabla(\operatorname{div} p^k + I)|_2} \tag{4.29}$$

4.4.2 Augmented lagrangian approaches

In order to deal with non-differentiability of the primal formulation of the ROF model (3.17) another variable can be introduced as $q = \nabla u$. Then (3.17) can equivalently be formulated as

$$\min_{u,q} \int_\Omega |q| + (u - I)^2 \, dx \tag{4.30}$$

such that

$$q = \nabla u \tag{4.31}$$

By introducing a lagrange multiplier λ for the constraint (4.31), the augmented Lagrangian functional of (4.30) can be written

$$\min_{u,q} \sup_{\lambda} L(u, q, \lambda) = \int_\Omega |q| + (u - I)^2 + \lambda(\nabla u - q) + \frac{c}{2}(\nabla u - q)^2 \, dx \tag{4.32}$$

The augmented Lagrangian method (2.28) outlined in Section 2.1.3 reduces in this case to

$$u^{k+1} = \arg\min_u L(q^k, u, \lambda^k) \tag{4.33}$$

$$q^{k+1} = \arg\min_u L(q, u^{k+1}, \lambda^k) \tag{4.34}$$

$$\lambda^{k+1} = \lambda^k + c(q^{k+1} - \nabla u^{k+1}) \tag{4.35}$$

$$\tag{4.36}$$

This algorithm is equivalent to the split-bregman algorithm introduced in [32], which has been pointed out in several recent works [81, 26, 77]. The dual variables λ also have a strong connection to the dual variables p in (4.25), see e.g. [81].

4.4.3 Primal-dual Arrow-Hurwicz type approaches

A third class of algorithms is the primal-dual methods, which alternates gradient descent steps in the primal variables and gradient ascent steps in the dual variables. Consider the general problem

$$\min_{u \in S} \max_{p \in C} \langle p, Au \rangle - h(u), \tag{4.37}$$

where A is some linear operator. We assume the function $h(u)$ is differentiable. If $S = BV(\Omega)$, $C = C_\nu$, $h(u) = ||u - I||^2$, $A = -\nabla$ and $C = C_\nu$ then (4.37) corresponds to the ROF model. In [44] an algorithm for optimizing (4.37) was proposed. Choose two time steps σ, τ and solve for $k = 1, ...$

$$p^{k+1} = \Pi_C(p^k + \sigma A u^k) \tag{4.38}$$

$$u^{k+1} = \Pi_S(u^k + \tau(A^* p^{k+1} + \frac{\partial h}{\partial u}(u)). \tag{4.39}$$

In the context of the ROF model, such a scheme was first proposed in [93]. The scheme (4.38), (4.39) also applies directly if u constrained to some set S, like for instance the set $[0, 1]$ in the convex formulation of Pott's model with two regions (4.15) subject to (4.17). An acceleration of such schemes was later proposed by Popov in 1980 [71]. This scheme has recently been applied in recent works where the unknowns are constrained to various sets [68, 69, 67].

4.5 Algorithms for the Mumford-Shah model and higher order models

We will just briefly mention some numerical approaches for the piecewise smooth Mumford-Shah model and higher order models, without going into details. The piecewise smooth Mumford-Shah model [63] can be addressed numerically by phase field approaches [1]. It can also be solved approximately as the steady state of the Perona-Malik nonlinear diffusion equations [66], which can be shown converges to a local minimum of the piecewise smooth Mumford-Shah model. Recently, a convex relaxation of the piecewise smooth Mumford-Shah model appeared in [68], using a higher dimensional formulation of the problem. The higher order models have typically been addressed by solving the Euler-Lagrange PDEs directly, a process which is very time consuming due to the time strict step restrictions. Very recently, some faster optimization methods have been developed, as applications of more general convex optimization algorithms, the Euler's elastica model in [80], and the LLT model in [87]. A more extensive review can be found in the paper G.

Chapter 5

Fundamental Discrete Energy Models and Optimization Methods

Variational models in imaging are formulated under the assumption of a continuous image domain and are approached numerically by continuous optimization techniques. Energy minimization models can also be formulated over the discrete image domain directly. Such models have been independently developed in the discrete and combinatorial optimization community. We will see that the discrete models have a lot in common with the variational models discussed in Chapter 3 and are often equivalent in the continuum limit. The discrete optimization algorithms have many advantages over the continuous algorithms. In some cases they can find global solutions and often have a high efficiency. One crucial disadvantage of discrete models is the grid bias and lack of subgrid accuracy. The algorithms do not parallelize as well as continuous optimization algorithms, which can be easily implemented on GPU.

5.1 Discrete energy models and Markov random fields

Optimization problems are defined over a set of grid points. For example, in 2-D, such a set of pixels can be enumerated by

$$\mathcal{P} = \{(i, j), i = 1, ..., N_1, j = 1, ..., N_2\}. \tag{5.1}$$

If a uniform mesh, with mesh size h, is assumed, $\mathcal{P}$ enumerates the grid points of the discrete grid

$$\{(ih, jh), i = 1, ..., N_1, j = 1, ..., N_2\}.$$

Let u be a function defined over the discrete grid. For $p = (i, j) \in \mathcal{P}$, the notation $u_p = u_{i,j} = u(ih, jh)$ is used to refer to the function value of u at location (ih, jh). Usually the mesh size h is assumed to be 1. For each $p \in \mathcal{P}$ a neighborhood system $\mathcal{N}_p$ is defined, consisting of elements $q \in \mathcal{P}$ that are "close" to p according to some measure. For the 2D grid (5.1) "close" could mean (and usually means) grid points with the smallest euclidian distance to p, two examples would be the 4 neighborhood system and 8 neighborhood systems.

$$\mathcal{N}_p^4 = \{(i \pm 1, j), (i, j \pm 1)\} \cap \mathcal{P}$$

$$\mathcal{N}_p^8 = \{(i \pm 1, j), (i, j \pm 1), (i \pm 1, j \pm 1)\} \cap \mathcal{P}.$$

Extensions to 16, 32, 64 etc. neighbors should be obvious. For simplicity of notation, we also define the set of neighboring grid points as

$$\mathcal{N} = \{(p, q) \in \mathcal{P} \times \mathcal{P} \,:\, q \in \mathcal{N}(p) \text{ for some } p \in \mathcal{P}\}. \tag{5.2}$$

An energy function can be constructed which contains one variable u_p for each grid point $p \in \mathcal{P}$. Such a function is often called a "labeling function" and is constrained to take values from a predefined finite set of real numbers $\mathcal{L} = \{\ell_1, ..., \ell_n\}$. In quantized gray scale image restoration and processing, each label may correspond to a gray value, for instance $\mathcal{L} = \{1, ..., 256\}$. In image segmentation and partitioning problems, each label may correspond to a region, i.e. $\mathcal{L} = \{1, ..., n\}$ and $u_p = i$ if and only if p is assigned to region i.

First order markov random fields (mrfs) contain interaction terms between at most two of the variables. Since their introduction in imaging and vision by Geman and Geman [30], they have become among the most popular and widely used energy minimization models due to their simplicity and expressive power. 1st order mrfs can in general be written in the form

$$\min_u \sum_{p \in \mathcal{P}} f_p(u_p) + \sum_{(p,q) \in \mathcal{N}} V^{pq}(u_p, u_q) \tag{5.3}$$

such that

$$u_p \in \{\ell_1, ..., \ell_n\}, \, \forall p \in \mathcal{P} \tag{5.4}$$

The most useful first order mrfs in imaging and vision have the form of

$$\min_u \sum_{p \in \mathcal{P}} f_p(u_p) + \sum_{(p,q) \in \mathcal{N}} g(u_p - u_q) \tag{5.5}$$

such that

$$u_p \in \{\ell_1, ..., \ell_n\}, \, \forall p \in \mathcal{P} \tag{5.6}$$

If the function g is convex, problems of the form (5.5) can be optimized exactly by graph cuts, as shown in [42]. The function $f_p(.)$ can be arbitrary, possibly

non-convex. In [23] it was shown that, more generally, if the function V^{pq} is submodular, (5.7) can be optimized exactly by graph cuts. Although polynomially solveable, the algorithms are extremely computationally demanding in the general case, especially since the number of edges in the graph grows quadratically in n. In some special cases, faster algorithms are available. More details on this will be provided in Section 5.2. If V^{pq} is not submodular, or g is not convex, the problems (5.7) and (5.5) are NP-hard.

5.1.1 1st order binary markov random fields

When the set of feasible labels is binary, the problem (5.7) becomes

$$\min_u \sum_{p\in\mathcal{P}} f_p(u_p) + \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}^k(p)} V^{pq}(u_p, u_q) \tag{5.7}$$

such that

$$u_p \in \{0, 1\},\ \forall p \in \mathcal{P} \tag{5.8}$$

One can assume, without loss of generality, that $\ell_1 = 0$ and $\ell_2 = 1$. Submodularity reduces in this case to

$$V^{pq}(1,1) + V^{pq}(0,0) \leq V^{pq}(1,0) + V^{pq}(0,1), \forall (p,q) \in \mathcal{N} \tag{5.9}$$

We will in particular consider

$$V^{pq}(u_p, u_q) = w_{pq}|u_p - u_q| \tag{5.10}$$

If the 4-neighborhood system $\mathcal{N}_p^4$ is chosen and $w_{pq} = h\nu$ for all $(p,q) \in \mathcal{N}$, (5.10) reduces to

$$\sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}^4(p)} h\nu|u_p - u_q| = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} h^2\nu(|\frac{u_{i+1,j} - u_{i,j}}{h}| + |\frac{u_{i,j+1} - u_{i,j}}{h}|) \tag{5.11}$$

which is exactly the forward discretization of the anisotropic total variation from Section 3.2.4. As discussed in Section 3.2.1, the total variation of a binary characteristic function of a set S in 2-D, is equal to the euclidian length of the boundary of the set S. However, this result only holds for the isotropic variant of total variation (3.7) with the 2-norm. It is also possible to approximate the euclidian length with arbitrary precision with $V^{pq}(.,.)$ chosen in the form (5.10). By using a result from integral geometry called the Cauchy-Crofton formula, the weights w_{pq} can be derived as [5]

$$w_{pq} = \frac{\nu\, h^2\, \pi}{k|ph - qh|_2} \tag{5.12}$$

It can be shown that as the mesh size goes to zero and the number of neighbors k in the neighborhood system $\mathcal{N}^k(p)$ goes to infinity, the last term of (5.7) converges

to the curve length of the boundary of the set described by S. More precicely, let $S \subset \Omega$. For each $p \in \mathcal{P}$ define $u_p = 1$ if $p \in S$ and $u_p = 0$ else. Choose the weights w_{pq} by (5.12), then as $\delta \to 0$ and $k \to \infty$

$$\sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}^k(p)} w_{pq} |u_p - u_q| \to |\partial S|_E, \tag{5.13}$$

where $|\partial S|_E$ is the euclidian length of the boundary of S. More details can be found in [5]. This result shows that the grid bias in the discrete models can be reduced, but doing so requires progressively larger neighborhood systems, which in turn complicates the computation. In contrast, the continuous models can approximate the euclidian curve length with arbitrary precision by only decreasing the mesh size, while keeping the neighborhood system fixed to 4. In 3-D and N-D, this property of the continuous models becomes even more important.

5.1.2 Potts model and the Potts interaction term

Pott's model originates from statistical mechanics and describes the interaction between spins of particles. By regarding each u_p as a particle it can be written as a special case of the interaction term in the model (5.5) with

$$g(u_p - u_q) = w_{pq}\big(1 - \delta(u_p - u_q)\big), \tag{5.14}$$

where

$$\delta(s) = \begin{cases} 1\,, & \text{if } s = 0 \\ 0\,, & \text{else} \end{cases} . \tag{5.15}$$

By inserting in (5.5) we obtain the model

$$\min_u \sum_{p \in \mathcal{P}} f_p(u_p) + \sum_{(p,q) \in \mathcal{N}} w_{pq}\big(1 - \delta(u_p - u_q)\big) \tag{5.16}$$

such that

$$u_p \in \{\ell_1, ..., \ell_n\},\ \forall p \in \mathcal{P} \tag{5.17}$$

The interaction term (5.14) is valued w_{pq} if $u_p \neq u_q$ and 0 else. All changes in the labeling function are penalized equally.

Let $\mathcal{L} = \{1, ..., n\}$ and assume the labeling function u describes a partition of $\mathcal{P}$ into n regions $\mathcal{P}_i$, $i = 1, ..., n$ such that $u_p = i$ for all $p \in \mathcal{P}_i$. Then, clearly

$$\cap_{i=1}^n \mathcal{P}_i = \emptyset, \quad \cup_{i=1}^n \mathcal{P}_i = \mathcal{P}.$$

The Pott's interaction term penalizes each jump in the labeling function equally, independent of its actual values. In other words, the energy function penalizes the transition between any two regions $\mathcal{P}_i$ and $\mathcal{P}_j$, $i, j \in \{1, ..., n\}$ equally. By choosing the weights w_{pq} by the Cauchy-Crofton formula (5.12), we obtain, in

the continuous limit as the mesh size goes to zero and the number of neighbors k goes to infinity, the continuous Pott's model (3.1) from Section 3.1. The discrete Pott's model (5.16) can be seen as a multiway cut problem, which is known to be NP-hard [21].

As proposed in [94], the Pott's model can also be extended by adding a term that places a penalty on the number of labels that appear in the solution

$$\min_u \sum_{p\in\mathcal{P}} f_p(u_p) + \sum_{(p,q)\in\mathcal{N}} w_{pq}\big(1-\delta(u_p-u_q)\big) + \gamma\#\{\,1\le i\le n \,|\, u_p = i \text{ for some } p\in\mathcal{P}\}. \tag{5.18}$$

This model may select the best "models" (data cost functions corresponding to each label) out of possibly many, in order to explain the image.

5.1.3 Linear interaction potential

Another important special case of the interaction term in (5.5) is

$$g(u_p - u_q) = w_{pq}|u_p - u_q| \tag{5.19}$$

As in Section 5.1.1, by choosing $w_{pq} = h\nu$, the interaction potential then reduces to, in case of a 4 neighborhood system

$$\sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{N}^4(p)} h\nu|u_p - u_q| = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} h^2\nu\left(\frac{|u_{i+1,j} - u_{i,j}|}{h} + \frac{|u_{i,j+1} - u_{i,j}|}{h}\right) \tag{5.20}$$

which is the discrete anisotropic total variation. The complete model including the data term is repeated here for convenience.

$$\min_u \sum_{p\in\mathcal{P}} f_p(u_p) + \sum_{(p,q)\in\mathcal{N}} w_{pq}|u_p - u_q| \tag{5.21}$$

such that

$$u_p \in \{\ell_1, ..., \ell_n\},\ \forall p\in\mathcal{P} \tag{5.22}$$

Being a special case of (5.5) with convex g, (5.24) can be optimized by graph cuts as in [42]. When $f_p(.)$ is a convex function for all $p\in\mathcal{P}$, (5.24) can also be optimized by an algorithm which solves a sequence of binary mrf problems. More details will be given in Section 5.2.3.

5.1.4 Truncated quadratic interaction potential / Piecewise smooth Mumford Shah model

Another possible interaction potential is the truncated quadratic

$$g(u_p - u_q) = w_{pq}\min(|u_p - u_q|^2, r) \tag{5.23}$$

In this case the complete energy function to be minimized is

$$\min_u \sum_{p\in\mathcal{P}} f_p(u_p) + \sum_{(p,q)\in\mathcal{N}} w_{pq} \min(|u_p - u_q|^2, r) \tag{5.24}$$

In [13] a very similar problem to (5.24), was studied where $f_p(u_p) = |u_p - I_p|^2$, $k = 4$, $w_{pq} = \alpha$ and $r = \frac{\nu}{\alpha}$. The only difference was that the unknown u takes values in the continuous set $[0, L-1]$ as opposed to the discrete set $\{0, ..., L-1\}$. It was shown that such a model converges to the general Mumford-Shah model in the continuum limit. Therefore the model (5.23) can be seen as a discrete equivalent of the general piecewise smooth Mumford-Shah model.

Therefore, we can conclude that the model (5.24) is also NP-hard to optimize. This gives some insight on the difficulty of optimizing the general Mumford-Shah model.

5.1.5 Higher order Markov random fields

Higher order markov random fields, which contain interactions between more than 2 variables, have advantages over 1st order mrfs. They have been largely ignored in the past because of the difficult computational aspect. In all, except for some very restricted special cases, such mrfs are NP-hard. A review of recent literature can be found in the introduction of paper G. Let $\mathcal{N}^3$ denote some set of triple wise neighbors, i.e.

$$\mathcal{N}^3 \subseteq \mathcal{P} \times \mathcal{P} \times \mathcal{P}$$

2nd order mrfs can in general be written

$$\min_u \sum_{p\in\mathcal{P}} f_p(u_p) + \sum_{(p,q,r)\in\mathcal{N}^3} V^{pqr}(u_p, u_q, u_r) \tag{5.25}$$

such that

$$u_p \in \{\ell_1, ..., \ell_n\},\ \forall p \in \mathcal{P}, \tag{5.26}$$

where $V^{p,q,r}(.,.,.)\ \mathcal{L} \times \mathcal{L} \times \mathcal{L} \mapsto \mathbb{R}$. The generalization to n-order mrfs is straight forward.

5.2 Optimization methods for discrete energy models

We will give a brief overview of optimization methods for solving discrete problems of the form (5.7) and (5.5).

5.2.1 Binary and submodular 1st order mrfs

Picard and Ratliffe [43] were the first to observe that optimization of binary energies interactions between at most two variables could be minimized by computing the minimum cut on a graph. Greig et al. [34] showed essentially the same result for problems which can be written in the form (5.7), under the submodularity condition (5.9). Kolmogorov et al. [47] gave sufficient conditions on which energies with binary variables could be minimized by computing the minimum cut on a graph. Details about the graph constructions can for instance be found in Paper D and A. The details are therefore omitted here. Efficient implementations specialized for image processing applications have been proposed in [6].

5.2.2 Convex g or submodular V^{pq}

We consider now general problems of the form (5.7) and (5.5).

When g is convex, or V^{pq} is submodular, (5.5) or (5.7) can be optimized exactly by graph cuts for any data term $f_p(.)$ (possible non-convex). The details are given in [42] for the former and [23] for the latter and are omitted here. The idea is to construct a large graph where n vertices are associated with each pixel. The min-cut strategy can be used to compute a global minimizer with a max-flow algorithm of pseudo-polynomial complexity. Strictly, a polynomial algorithm should grow as $O(\log(n))$. Since the number of edges in the graph grows quadratically in n, the complexity of such an an algorithm grows as $O(n^2)$. This is too slow to be practical if n is relatively large, for instance if n is the number of gray values in image restoration.

In the special case that g is linear, a smaller graph can be constructed, where the number of edges instead grows linearly in n as was shown in [42]. This makes computation more tractable, but still rather slow if n is large. Details about the constructions are given in Paper E and A and are therefore omitted here.

Yet another special case occurs when the data term f is convex and g is linear. In this case, very efficient algorithms are available of polynomial, as opposed to pseudo polynomial complexity. This subject will be treated in the next section.

5.2.3 Convex f and linear g

Assume that the data term f is convex and the regularization term g is linear. As was shown in [38], [12] and [22] the problem (5.5) can then be reduced to that of solving a set of binary problems for each label. Furthermore, since there is a lot of redundancy, one can get away with solving $\log_2(n) - 1$ binary problems.

Since u is constrained to take values in the set $\{1, ..., n\}$, there are $n - 1$ distinct functions of upper level sets $\ell = 0, ..., n - 1$, which are point-wise given

by

$$\theta_p^\ell = \begin{cases} 1 & \text{if } u_p \geq \ell, \\ 0 & \text{else.} \end{cases} \tag{5.27}$$

In the discrete setting, the coarea formula (3.13) can be written

$$\sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|u_p - u_q| = \sum_{\ell=0}^{n-1} \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\theta_p^\ell - \theta_q^\ell|. \tag{5.28}$$

To approximate the euclidian curve length, the weights w_{pq} can be derived by the Cauchy Crofton formula as (5.12).

$$\int_\Omega |\nabla\theta^\ell(x)|\, dx\, d\ell \approx \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\theta_p^\ell - \theta_q^\ell|. \tag{5.29}$$

If u is a function, it must be single valued at each point $p \in \mathcal{P}$. In that case the family $\{\theta^\ell\}_{\ell=0}^{n-2}$ is monotonically non-increasing: $\theta^{\ell+1} \leq \theta^\ell,\ \forall\ \ell = 0, ..., n-2$. For any such family of monotonically decreasing binary functions, the function u can be recovered by the formula $u_p = \max\{\ell :\ \theta_p^\ell = 1\}$.

Note also that for each $p \in \mathcal{P}$ the data term of (5.5) can be written in terms of $\{\theta_p^\ell\}_\ell$ as

$$\sum_{p\in\mathcal{P}} f_p(u_p) = \sum_{\ell=0}^{n-1} \sum_{p\in\mathcal{P}} (\theta_p^\ell - \theta_p^{\ell+1}) f_p(\ell). \tag{5.30}$$

Therefore, combining (5.28) and (5.30), (5.5) with $g(.) = |.|$ can be written

$$\sum_{\ell=0}^{n-2} \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\theta_p^\ell - \theta_q^\ell| + \sum_{\ell=0}^{n-2} \sum_{p\in\mathcal{P}} (\theta_p^\ell - \theta_p^{\ell+1}) f_p(\ell) = \sum_{\ell=0}^{n-2} E^\ell(\theta^\ell), \tag{5.31}$$

subject to

$$\theta_p^{\ell+1} \leq \theta_p^\ell, \quad \forall p \in \mathcal{P},\ \ell = 0, ..., n-2, \tag{5.32}$$

where

$$E^\ell(\theta^\ell) = \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\theta_p^\ell - \theta_q^\ell| + (\theta_p^\ell - \theta_p^{\ell+1}) f_p(\ell) \tag{5.33}$$

It was shown in [12] that by minimizing each $E^\ell(\theta^\ell)$ independently, the constraints (5.32) were automatically satisfied. Furthermore, it was observed that each $E^\ell(\theta^\ell)$ has the form of a binary MRF, and could be efficiently minimized via graph cuts as was shown by Greig et. al. [35]. This leads to several algorithms. Since there is a lot of redundancy between the problems $E^\ell(\theta^\ell)$, it is not necessary to solve the problems sequentially. That would yield an algorithm of linear complexity in the number of gray values. Instead, a dyadic algorithm was proposed in [38], [12], [22], which has a logarithmic complexity in the number of gray values, i.e. at most $\log_2(n) - 1$ binary MRFs had to be solved.

5.2.4 Approximate optimization for nonconvex g or non-submodular V^{pq}

If g is non convex, or V^{pq} is non-submodular, (5.5) or (5.7) are NP-hard. In particular this is the case for the Pott's interaction potential (5.14) and the truncated quadratic potential (5.23). There are several methods which can compute approximate solutions for such problems, like iterated conditional modes [4], simulated annealing [30], message passing [85, 46], linear programming [48]. Most popular are the graph cut based alpha expansion and alpha-beta swap [7], which are widely considered state of the art for minimizing such energy functions. This thesis proposes new methods for solving (5.16) and contains comparisons with alpha expansion and alpha beta swap. We therefore give a brief review of these approaches.

Alpha expansion and alpha-beta swap applies under some assumptions on V^{pq}. The interaction potential $V^{pq}(.,.)$ is said to be a semi-metric if for any $\alpha, \beta \in \mathcal{L}$ and all $(p,q) \in \mathcal{N}$, $V^{pq}(\alpha,\beta) = V^{pq}(\beta,\alpha) \geq$ and $V^{pq}(\alpha,\beta) = 0 \Leftrightarrow \alpha = \beta$. If $V^{pq}(.,.)$ also satisfies the triangle inequality, i.e. for any $\alpha, \beta, \gamma \in \mathcal{L}$, $V^{pq}(\alpha,\beta) + V^{pq}(\beta,\gamma) \leq V^{pq}(\alpha,\gamma)$, then $V^{pq}(.,.)$ is said to be a metric.

The core of alpha expansion and alpha-beta swap is optimization of binary energy functions of the form (5.7) by graph cuts. For any labeling function $u : \mathcal{P} \mapsto \mathcal{L}$, define the set $\mathcal{P}_{\ell_i}(u) = \{p \in \mathcal{P} \quad u_p = \ell_i\}$, $i = 1, ..., n$. Given two label values $\alpha, \beta \in \mathcal{L}$, a move from a labeling function u to a labeling function v is called an alpha-beta swap if $\mathcal{P}_{\ell_i}(u) = \mathcal{P}_{\ell_i}(v)$ for all $\ell_i \in \mathcal{L} \backslash (\alpha \cup \beta)$. For a label value $\alpha \in \mathcal{L}$ a move from u to v is called an alpha expansion if $\mathcal{P}_\alpha(u) \subseteq \mathcal{P}_\alpha(v)$ and $\mathcal{P}_{\ell_i}(v) \subseteq \mathcal{P}_{\ell_i}(u)$ for all $\ell_i \in \mathcal{L} \backslash \alpha$.

For any $\alpha, \beta \in \mathcal{L}$ and any u, the optimal labeling function v which is within one $\alpha - \beta$ swap from u and which minimizes the energy (5.7) can be computed by binary graph cuts, provided $V^{pq}(.,.)$ is a semi-metric. In the same vain, for any $\alpha \in \mathcal{P}$ and any labeling function u, the optimal labeling function v which is within one α expansion from u and which minimizes the energy (5.7) can be computed by binary graph cuts, provided $V^{pq}(.,.)$ is a metric. The $\alpha - \beta$ swap and α expansion algorithms starts with an initial labeling function u and iteratively computes $\alpha - \beta$ swap moves and α expansion moves respectively, until the energy does not change. The final labeling function is accepted as an approximate solution to the minimization problem (5.7). The details are provided in Algorithm 5.2.4 and Algorithm 5.2.4.

Observe that for each iteration, alpha expansion needs to solve n binary optimization problems and alpha swap swap needs to solve n^2 binary problems. This iterative process is repeated until convergence.

The alpha expansion algorithm was also recently generalized in [24] to approximately minimize energies with label cost prior of the form (5.18).

Algorithm 1 $\alpha - \beta$ swap algorithm

- Select initial labeling function u
- Repeat until $w_p = v_p \; \forall p \in \mathcal{P}$
  - For each pair $\alpha, \beta \in \mathcal{L}$
    - $$u \leftarrow \arg\min_u E(u), \quad \text{s.t. } u \text{ is one } \alpha - \beta \text{ swap move from } v \tag{5.34}$$
    - set $v \leftarrow u$
  - set $w \leftarrow v$
- Output labeling function w.

Algorithm 2 α expansion algorithm

- Select initial labeling function u
- Repeat until $w_p = v_p \; \forall p \in \mathcal{P}$
  - For each $\alpha \in \mathcal{L}$
    - $$u \leftarrow \arg\min_u E(u), \quad \text{s.t. } u \text{ is one } \alpha \text{ expansion move from } v \tag{5.35}$$
    - set $v \leftarrow u$
  - set $w \leftarrow v$
- Output labeling function w.

5.2.5 Higher order mrfs

Very little work has appeared on optimization methods for mrfs of order 2 or higher, due to their difficult computational nature. Recently, some methods have appeared for computing approximate solutions. An algorithm for solving minimization problems where Euler's elastica acts as a regularization term is presented in paper G. A more detailed review of some recent work can be found in the introduction of this paper, and is therefore omitted here.

Chapter 6

Summary of Papers

There are 10 articles included in this thesis. All of them are based on global and efficient optimization, and build to varying degree on each other. Each paper stand on its own, but I also hope to give the reader a sense of progression and unity. With some exceptions, most of them are ordered in chronological order, according to when they were produced, as this is also the most reasonable order in presenting them.

Both conference and journal papers/preprints are included. Two journal preprints, B and D, are direct extensions of earlier conference papers. In those cases the shorter conference versions are omitted from the thesis. Therefore, each included paper presents its own scientific contribution. However, it should be noted that there is some overlap in the review section, appendix and section containing analysis of submodular condition between paper F and the extended version of paper B.

As the papers also involve co-authors, a short clarification of the contributions are given below:

- Paper A,B and F: The main idea, theoretical findings, implementations, experiments and writing of the papers were done by me. The last author was supervisor and gave useful advices. In paper A he had the idea of the proof of Theorem 1.

- Paper C,D,E,H,I: The main ideas and theoretical findings were developed by the two first authors, Jing Yuan and me. The first author of the respective papers did the majority of writing, implementation and experiments. In paper H and I, I also helped with some experiments. The remaining authors were supervisors and gave useful advices.

- Paper G: The initial idea and methodology were developed by me. I wrote the majority of the paper, while the second author Shi Juan performed the majority of experiments. The methods were implemented by both of us. The last author was supervisor and gave useful advices.
- Paper J: I had the initial idea of computing the medial axis of the crust in order to specify boundary conditions in such a way that graph cuts could be applied to solve the problems. The first two authors did implementations, experiments and wrote the majority of the paper. I helped writing the paper and gave suggestions of revisions.

6.1 Paper A: Graph Cut Optimization for the Piecewise Constant Level Set Method Applied to Multiphase Image Segmentation

Egil Bae and Xue-Cheng Tai
In Scale space and variational methods in computer vision. Lecture Notes in Computer Science 5567, pp. 1-13, Springer 2009.

Abstract: The piecewise constant level set method (PCLSM) has recently emerged as a variant of the level set method for variational interphase problems. Traditionally, the Euler-Lagrange equations are solved by some iterative numerical method for PDEs. Normally the speed is slow. In this work, we focus on the piecewise constant level set method (PCLSM) applied to the multiphase Mumford-Shah model for image segmentation. Instead of solving the Euler-Lagrange equations of the resulting minimization problem, we propose an efficient combinatorial optimization technique, based on graph cuts. Because of a simplification of the length term in the energy induced by the PCLSM, the minimization problem is not NP hard. Numerical experiments on image segmentation demonstrate that the new approach is very superior in terms of efficiency, while maintaining the same quality.

Main results:

- A relationship between the piecewise constant level set method and a class of discrete optimization methods is revealed
- A simplification of the graph of Ishikawa [42] is proposed, which requires $n-1$ layers of vertices instead of n, and avoids edges with infinite capacities. A proof is given that optimal cuts correspond to optimal level set functions in a one-to-one manner.
- An alternating algorithm for minimizing the energy jointly with respect to the regions and mean intensities of each region.

6.2 Paper B: Efficient Global Minimization for the Multiphase Chan-Vese Model of Image Segmentation

Egil Bae and Xue-Cheng Tai
In Energy minimization methods in computer vision and pattern recognition (EMMCVPR). Lecture notes in Computer Science 5681, pp 28-41, Springer 2009 (an extended journal version included).

Abstract: The Mumford-Shah model is an important variational image segmentation model. A popular multiphase level set approach, the Chan-Vese model, was developed as a numerical realization by representing the phases by several overlapping level set functions. Recently, a variant representation of the Chan-Vese model with binary level set functions was proposed. In both approaches, the gradient descent equations had to be solved numerically, a procedure which is slow and has the potential of getting stuck in a local minima.

In this work, we develop an efficient and global minimization method for a discrete version of the level set representation of the Chan-Vese model with 4 regions, based on graph cuts. If the average intensity values of the different phases are sufficiently evenly distributed, the energy function is submodular. It is shown theoretically and experimentally that the condition is expected to hold for the most commonly used data terms. We have also developed a method for minimizing nonsubmodular functions, that can produce global solutions in practice should the condition not be satisfied, which may happen for the L^1 data term.

Main results:

- It is shown that a new type of energy functions involving multiple regions can be solved exactly, by computing the minimum cut on a graph.
- A discrete version of the level set representation of the Chan Vese model (4.9) with 4 region fits into this framework, provided the data term satisfies a condition.
- It is shown theoretically and experimentally that the condition is expected to hold for the most commonly used data terms.
- An algorithm is proposed for minimizing non-submodular energy functions, with particular emphasis on this energy function, should the condition not be satisfied. The algorithm cannot be proven to always produce a global solution, but tends to do so in practice.

6.3 Paper C: Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach

Egil Bae, Jing Yuan and Xue-Cheng Tai
International Journal on Computer Vision, 2010, DOI: 10.1007/s11263-010-0406-y

Abstract: This paper is devoted to the optimization problem of continuous multi-partitioning, or multi-labeling, which is based on a convex relaxation of the continuous Potts model. In contrast to previous efforts, which are tackling the optimal labeling problem in a direct manner, we first propose a novel dual model and then build up a corresponding duality-based approach. By analyzing the dual formulation, sufficient conditions are derived which show that the relaxation is often exact, i.e. there exists optimal solutions that are also globally optimal to the original nonconvex Potts model. In order to deal with the nonsmooth dual problem, we develop a smoothing method based on the log-sum exponential function and indicate that such a smoothing approach leads to a novel smoothed primal-dual model and suggests labelings with maximum entropy. Such a smoothing method for the dual model also yields a new thresholding scheme to obtain approximate solutions. An expectation maximization like algorithm is proposed based on the smoothed formulation which is shown to be superior in efficiency compared to earlier approaches from continuous optimization. Numerical experiments also show that our method outperforms several competitive approaches in various aspects, such as lower energies and better visual quality.

Main results:

- The relatively simple convex relaxation of Pott's model (4.20) subject to (4.21), (4.23) is formulated and analyzed from a dual perspective.
- Conditions are derived for when global solutions of Pott's model can be computed from a dual solution of the relaxed problem.
- A rounding scheme is proposed for obtaining obtaining global solutions of Pott's model when the conditions are satisfied, and approximate solutions if they are not.
- An efficient algorithm based on the dual formulation is proposed.

6.4 Paper D: A Study on Continuous Max-Flow and Min-Cut Approaches

Jing Yuan, Egil Bae and Xue-Cheng Tai
In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010 (an extended journal version is included (with Yuri Boykov))

Abstract: We propose and investigate novel max-flow models in the spatially continuous setting, with or without supervised constraints, under a comparative study of graph based max-flow / min-cut. We show that the continuous max-flow models correspond to their respective continuous min-cut models as primal and dual problems, and the continuous min-cut formulation without supervision constraints regards the well-known Chan-Esedoglu-Nikolova model [19] as a special case. In this respect, basic conceptions and terminologies applied by discrete max-flow / mincut are revisited under a new variational perspective. We prove that the associated nonconvex partitioning problems, unsupervised or supervised, can be solved globally and exactly via the proposed convex continuous max-flow and min-cut models. Moreover, we derive novel fast max-flow based algorithms whose convergence can be guaranteed by standard optimization theories. Experiments on image segmentation, both unsupervised and supervised, show that our continuous max-flow based algorithms outperform previous approaches in terms of efficiency and accuracy.

Main results:

- A continuous and convex formulation of max-flow and min-cut models for solving binary partitioning problems.
- The max-flow and min-cut models are shown to form strong dual pairs. Further, it is shown the max-flow model is a strong dual to the convex model of Chan-Esedoglu-Nikolova [19] introduced in Section 4.2.
- The continuous max-flow model is extended to supervised partitioning problems, the same duality correspondence is shown for such problems.
- An algorithm is proposed based on the max-flow formulations of the problems. The algorithm is an application of the augmented Lagrangian method and is shown to be very efficient.

6.5 Paper E: A Fast Continuous Max-flow Approach to Non-convex Multilabeling Problems

Egil Bae, Jing Yuan, Xue-Cheng Tai and Yuri Boykov
Submitted for journal publication

Abstract: This work addresses a class of multilabeling problems over a spatially continuous image domain, where the data fidelity term can be any bounded function, not necessarily convex. Two total variation based regularization terms are considered, the first favoring a linear relationship between the labels and the second independent of the label values (Pott's model). In the spatially discrete setting, Ishikawa [42] showed that the first of these labeling problems can be solved exactly by standard max-flow and min-cut algorithms over specially designed graphs. We will propose a continuous analogue of Ishikawa's graph construction [42] by formulating continuous max-flow and min-cut models over a specially designed domain. These max-flow and min-cut models are equivalent under a primal-dual perspective. They can be seen as exact convex relaxations of the original problem and can be used to compute global solutions. Fast continuous max-flow based algorithms are proposed based on the max-flow models whose efficiency and reliability can be validated by both standard optimization theories and experiments. In comparison to previous work [70, 69] on continuous generalization of Ishikawa's construction, our approach differs in the max-flow dual treatment which leads to the following main advantages: A new theoretical framework which embeds the label order constraints implicitly and naturally results in optimal labeling functions taking values in any predefined finite label set; A more general thresholding theorem which, under some conditions, allows to produce a larger set of non-unique solutions to the original problem; Numerical experiments show the new max-flow algorithms converge faster than the fast primal-dual algorithm of [70, 69]. The speedup factor is especially significant at high precisions. In the end, our dual formulation and algorithms are extended to a recently proposed convex relaxation of Pott's model [67], thereby avoiding expensive iterative computations of projections without closed form solution.

Main results:

- A continuous and convex max-flow formulation of variational problems with total variation regularization where the unknown is constrainted to a finite set and general non-convex data term.
- A thresholding scheme which, under some conditions, can produce a larger

set of solutions to the original problem than earlier works [70, 69].

- A new algorithm based on the max-flow formulation, which is significantly more efficient than the primal-dual algorithm (4.38) with Popov's acceleration [71] from Section 4.4.3.
- A new algorithm for a recent convex relaxation of Pott's model [68] which avoids the bottleneck of iterative computations of projections of the dual variables without closed form solution.

6.6 Paper F: Exact Convex Formulation of the Chan-Vese Model and a Tight Convex Relaxation of Pott's Model with 4 Regions

Egil Bae and Xue-Cheng Tai
In finalization, to be submitted to a journal

Abstract: We propose an exact convex formulation of the Chan-Vese model with 4 regions. A global solution is guaranteed if the data term satisfies a (mild) submodularity condition. Theoretical and experimental arguments are given that such a condition will hold in practice for the most commonly used type of data terms. Otherwise, a convex truncation scheme is proposed which tends to produce global solutions in practice, should this not be the case.

Secondly, we build up a convex relaxation for Pott's model with 4 regions. While several convex relaxations for Pott's model have been proposed recently, ours is both the most simple and tightest that exists for such problems. Algorithms are proposed which we believe are more efficient than most previous work due to the simple formulations.

Main results:

- A convex and global optimization framework for the Chan-Vese model (4.9) with 4 regions in the continuous setting, under the same condition that made the discrete counterpart in paper B graph representable.
- As in paper B, it is shown that the condition is expected to hold for the most commonly used data terms.
- A truncation scheme to convexify the problem is proposed in case the condition is not satisfied. A condition is derived for when a solution of the truncated convex problem is also a solution of the original problem.
- A convex relaxation of Pott's model (3.1) with 4 regions, which is both the simplest and tightest that exists so far. That the dominance over [67] is strict remains to be proved, but arguments are given to support that this is expected.

6.7 Paper G: Graph Cuts for Curvature based Image Denoising

Egil Bae, Shi Juan and Xue-Cheng Tai
IEEE Transactions on Image processing, 2010, DOI: 10.1109/TIP.2010.2090533

Abstract: Minimization of total variation (TV) is a well known method for image denoising. Recently, the relationship between TV minimization problems and binary MRF models has been much explored. This has resulted in some very efficient combinatorial optimization algorithms for the TV minimization problem in the discrete setting via graph cuts. To overcome limitations, such as staircasing effects, of the relatively simple TV model, variational models based on higher order derivatives have been proposed. The Eulers elastica model is one such higher order model of central importance, which minimizes the curvature of all level lines in the image. Traditional numerical methods for minimizing the energy in such higher order models are complicated and computationally complex. In this work we will present an efficient minimization algorithm based on graph cuts for minimizing the energy in the Eulers elastica model, by simplifying the problem to that of solving a sequence of easy graph representable problems. This sequence has connections to the gradient flow of the energy function, and converges to a minimum point. The numerical experiments show that our new approach is more effective in maintaining smooth visual results while preserving sharp features better than TV models.

Main results:

- A fast algorithm for minimizing the energy in the Euler's elastica model (3.42) is proposed, where the gray values are assumed quantized in L levels.
- The problem is decomposed into separate subproblems, each of which can be solved by graph cuts with a complexity which is logarithmic in L, using the algorithm described in Section 5.1.3.
- Numerical experiments demonstrate the efficiency and properties of the algorithm.

6.8 Paper H: A Continuous Max-Flow Approach to Potts Model

Jing Yuan, Egil Bae, Xue-Cheng Tai and Yuri Boykov
In ECCV. Lecture Notes in Computer Science, 2010, Volume 6316/2010, 379-392,

Abstract: We address the continuous problem of assigning multiple (unordered) labels with the minimal perimeter. The corresponding discrete Potts model is typically addressed with a-expansion which can generate metrication artifacts. Existing convex continuous formulations of the Potts model use TV-based functionals directly encoding perimeter costs. Such formulations are analogous to min-cut problems on graphs. We propose a novel convex formulation with a continous max-flow functional. This approach is dual to the standard TV-based formulations of the Potts model. Our continous max-flow approach has significant numerical advantages; it avoids extra computational load in enforcing the simplex constraints and naturally allows parallel computations over different labels. Numerical experiments show competitive performance in terms of quality and significantly reduced number of iterations compared to the previous state of the art convex methods for the continuous Potts model.

Main results:

- The dual problem of Paper C is reformulated and given a new max-flow interpretation.
- Another algorithm is constructed for solving the reformulated dual problem. The algorithm is based on the augmented Lagrangian method.
- Numerical experiments demonstrate the efficiency of the algorithm.

6.9 Paper I: A Continuous Max-Flow Approach to Minimal Partitions with Label Cost Prior

Jing Yuan, Egil Bae, Yuri Boykov and Xue-Cheng Tai
In Third International Conference on Scale Space and Variational Methods in Computer Vision, 2011

Abstract: This paper investigates a convex relaxation approach for minimum description length (MDL) based image partitioning or labeling, which proposes an energy functional regularized by the spatial smoothness prior joint with a penalty for the total number of appearences or labels, the so-called label cost prior. As common in recent studies of convex relaxation approaches, the total-variation term is applied to encode the spatial regularity of partition boundaries and the auxiliary label cost term is penalized by the sum of convex infinity norms of the labeling functions. We study the proposed such convex MDL based image partition model under a novel continuous flow maximization perspective, where we show that the label cost prior amounts to a relaxation of the flow conservation condition which is crucial to study the classical duality of max-flow and min-cut! To the best of our knowledge, it is new to demonstrate such connections between the relaxation of flow conservation and the penalty of the total number of active appearences. In addition, we show that the proposed continuous max-flow formulation also leads to a fast and reliable max-flow based algorithm to address the challenging convex optimization problem, which significantly outperforms the previous approach by direct convex programming, in terms of speed, computation load and handling large-scale images. Its numerical scheme can by easily implemented and accelerated by the advanced computation framework, e.g. GPU.

Main results:

- An extension of the max-flow model in the previous paper to Potts model with label cost prior (5.18), where the number of appearing regions are penalized linearly. It is shown that such problems can be formulated by relaxing the flow conservation condition in the max-flow model of the previous paper.
- An efficient algorithm for solving the problem is proposed in a convex framework, again by exploiting the dual formulation of the problem.
- Numerical experiments demonstrate a significantly faster convergence rate than direct second order cone programming. The problem can be solved in a few minutes instead of several hours.

6.10 Paper J: Reconstructing Open Surfaces via Graph-Cuts

Min Wan, Yu Wang, Egil Bae, Xue-Cheng Tai, Desheng Wang
Submitted for journal publication

Abstract: A novel graph-cuts-based method is proposed for reconstructing open surfaces from unordered point sets. Through a boolean operation on the crust around the data set, the open surface problem is translated to a watertight surface problem within a restricted region. Integrating the variational model, Delaunay-based tetrahedra mesh framework and multi-phase technique, the proposed method can reconstruct open surfaces robustly and effectively. Furthermore, a surface reconstruction method based on domain decomposition is presented, which is based on the new open surface reconstruction method. This method can also handle more general surfaces, like non-orientable surfaces. The algorithm is designed in a parallel-friendly way, such that the surface patch in each subdomain can be approached independently. Necessary measures are taken to eliminate the cracks at the interface between the subdomains. Numerical examples are included to demonstrate the robustness and effectiveness of the proposed method on open surfaces, non-orientable surfaces and combinations of such.

Main results:

- An algorithm for reconstructing open and non-orientable surfaces from point cloud data is proposed
- The algorithm consists of two steps. In the first step, a crust is constructed around the point cloud. By calculating the medial axis of the crust, boundary conditions can be specified in such a way that graph cuts can be used to reconstruct the open surface in the second step.
- Non-orientable surfaces are regarded as the union of open surfaces, and are approached by domain decomposition, where the subproblem in each domain involves the reconstruction of an open surface.

Chapter 7

Conclusions

7.1 Summary

This thesis has focused on the development of efficient global optimization methods for important variational and discrete energy minimization problems that arise in image processing and computer vision. Powerful image processing and computer vision models have been developed during the last decades, both in the variational and discrete optimization community. They all have in common an objective function consisting of a data fitting term and a regularization term. However, the design of methods for computing solutions to the models have been very challenging, and the available methods have many limitations. Variational optimization approaches, such as the level set method may easily get stuck in inferior local minima and converges relatively slowly. Combinatorial optimization algorithms can compute global solutions for certain problems, but suffer from metrication errors and grid bias and does not parallelize as easily. Many important discrete optimization problems are NP-hard, like minimal perimeter partitioning problems. For such problems the available combinatorial algorithms compute approximate solutions which may lead to noticeable errors.

The thesis has contributed to both combinatorial optimization and continuous optimization and has aimed to bring these fields closer together. The contributions can roughly be divided in two parts. First is the development of algorithms which are guaranteed to compute global solutions to their respective problems. Second is the development of algorithms for NP-hard problems, that cannot be guaranteed to compute a global solution for every input data, but often do so and otherwise provide good approximations. In particular, minimal perimeter partitioning problems have been studied extensively. In both cases, it has been a central aim to develop methods that are also efficient, or have a nature which makes them easily computable (e.g. on parallel processors).

Max-flow and min-cut (graph cuts) have been central in the achievement of the first goal. Max-flow and min-cut can be used to solve many polynomially

solvable combinatorial optimization problems that arise in image processing and computer vision, but their application is somewhat limited, reflecting the fact that most combinatorial problems are NP-hard. In particular, segmentation problems with multiple regions have been problematic, unless one assume a linear inclusion property of the regions. In paper B it was shown that graph cuts can be used to globally solve a new class of problems with multiple regions: segmentation problems with 4 regions in the popular overlapping level set framework of Chan and Vese. In paper G, problems with higher order regularization terms involving curvature were solved by graph cuts, by decomposing the problem into a set of easier 1st order subproblems. The approach cannot be guaranteed to compute a global solution, but performs well in experiments. In paper J, graph cuts were generalized to solve problems involving open or non-orientable surfaces from point cloud data.

Global minimization methods are generally less developed in the variational setting. In paper D and E, continuous generalizations of max-flow and min-cut were given, which made it possible to compute global solutions to certain variational problems, while avoiding the metrication artifacts and grid bias of combinatorial max-flow and min-cut. Convex formulations of such problems have also recently been proposed in [19, 70]. The max-flow and min-cut generalizations resulted in the same convex relaxation of two phase problems as in [19] and a comparable convex relaxation of the total variation regularized problems in [70]. The max-flow and min-cut models had a structure which allowed for very efficient computation and were useful in analyzing the connections between the convex relaxed problems and original problems. In paper F, a continuous generalization of the max-flow and min-cut formulation of the Chan-Vese model presented in paper B was given, which resulted in a convex optimization framework for solving the problem.

Minimal perimeter partitioning problems, which are NP-hard in the discrete context, have been studied extensively in this thesis. Several methods have been developed for solving variational formulations of the problem in a convex manner. During the last two years, there has also been a lot of activity in the area from several research groups [92, 51, 67, 9]. They all have advantages and disadvantages compared to each other which will be elaborated here. A relatively simple binary relaxation approach that was suggested in [92, 51], was analyzed from a dual perspective in Paper C. It was shown that by instead solving the dual problem, a global solution of the original problem could be obtained via a new thresholding scheme under some conditions on the dual variables. A smooth approximation of the dual problem was proposed, which tended to force the conditions to be satisfied (but on the contrary introduced an approximation error) and led to an efficient algorithm which was significantly faster than other alternatives. Another efficient algorithm based on the dual formulation without smoothing was proposed in paper H, where some connections between the dual model and the max-flow models of paper E and D were also pointed out. The

dual formulation and algorithm were also be extended to problems which penalizes the number of regions in paper I, with some significant structural advantages over other alternatives. The relaxation of [67] is tighter than those in paper C,I and [92, 51], but is much more computationally complex for two reasons: (1) the number of side constraints grow quadratically in the number of regions, (2) iterative computation of projections without closed form solution of the variables are required every iteration of the main algorithm. In paper E, an algorithm was proposed which avoided to compute projections by an iterative algorithm, but was still significantly slower than those in paper C, I, due to the quadratic complexity of the problem. In paper F a relaxation was proposed, which was so far limited to applications with 4 regions, but was both the simplest and tightest that exists. A formal proof that the dominance over [67] is also strict remains to be worked out, but arguments were given to support that this is expected.

7.2 Future research

In this section some open challenges and new research opportunities are discussed, some are rather concrete, while others are more abstract.

In paper G, an algorithm was constructed for solving minimizing problems with Euler's elastica as regularization term (curvature of all level lines) on a discrete grid with L quantized gray values. The problem was decomposed into a set of simpler subproblems, each of the form (5.24), which could be solved efficiently by graph cuts in $O(log(L))$ as discussed in Section 5.2.3. The algorithm can also be applied for other problems, one of the most interesting extensions would be first order mrfs with general convex data term and convex regularization term. Existing algorithms for such problems are much more computationally complex and grow at least linearly in L as discussed in Chapter 5. Instead of a penalty parameter, a lagrange multiplier can be introduced to force the solution between successive iterations to be close, and be updated every iteration. I believe it is possible to prove convergence to an exact solution in a finite number of steps when applied to such problems. The efficiency of the algorithm can be improved even further by reusing flow from each iteration to the next, instead of solving the subproblems from scratch each iteration. Another extension of the algorithm is to reconstruct surfaces of minimal curvature. It could for instance be applied to reconstruct surfaces from point cloud data of minimal curvature, instead of minimal surface area as in paper J. The curvature of the distance function to the surface at iteration n could be used as weights during iteration $n+1$. Regularizing with curvature has many advantages, such an approach would present an efficient algorithm for solving the minimization problem.

Minimal perimeter partitioning problems (Potts model) have been a central topic of this thesis. Image segmentation is one of the most fundamental problems in image processing and computer vision, and the Pott's model has a structure

which is the most intuitive for energy minimization formulation of such problems: it does not favor any particular inclusions of the unknown regions. Different convex relaxations have been studied in this work. The relaxation proposed in paper F is both the tightest and most simple of them all, meaning it is the least computationally demanding and will most often produce global solutions, but it is so far only applicable to problems with four or less regions. A natural extension of this work will be problems with more than four regions, but the submodularity condition will become more strict for such problems. An alternative is to formulate the problems in such a way, that four phases suffices to find the optimal partitions. By the four colour theorem, any partition can be described with four phases. An open problem would be to assign different region parameters to each connected component of the 4 phases. The method would apply directly in some cases when rough a priori information about the object locations are provided in advance, and thereby a four colour map, in the same way as the four colour theorem was exploited in a 4 phase level set framework in [39] for such problems. In the same manner, some of the less tight relaxations could be used to provide more rough partitioning, and thereby a 4 color map, in the general case. For instance, the method in paper C can produce a hard partition whereever $f(x) + \operatorname{div} p^*(x)$ has two or less minimizers, which would provide an excellent such 4 colour map.

As we have seen, the convex models for partitioning problems perform well in practice, an open challenge is to derive error bounds in worst case scenarios. In paper C, it would be interesting to investigate further the case where the dual solution vector $f + \operatorname{div} p^*$ has 3 or more non-unique smallest components.

The algorithm in paper F avoided iterative projections of the dual variables as in earlier works. With some modifications on the constraint sets, the same algorithm can be applied to a recently proposed convex relaxation of the general Mumford-Shah model [68], which is also very slow to optimize by existing algorithms due to expensive iterative computations of projections.

Image segmentation is by no means a solved problem. The results produced by todays most advanced algorithms cannot compete with the human brain in recognizing and classifying objects. I believe many future research challenges from a modeling perspective will involve incorporation of a priori knowledge in the image segmentation models. There has already been devoted work in this direction, where the a priori information is incorporated via new parameters (such as rotations and translations of a shape or family of shapes), see e.g [3] for a review. Good optimization algorithms for these problems are challenging and still lacking. Another open challenge is better algorithms for optimizing jointly with respect to the regions and the parameters (such as mean intensity values) describing each region.

The Pott's model will continue to be a core energy minimization formulation of segmentation problems in the future. I believe many of the future modeling developments will involve either modifications of the data term in Pott's model or the addition of new terms. The regularization term itself can also be made more

sophisticated. The most natural extensions would be applications of non-local operators, or to seek partitions with small curvature. This of course introduces new computational challenges. For the former, exactly the same algorithms can be applied with small modifications. For the latter, deeper structural modifications will be necessary.

We have seen the importance of optimization in image processing and computer vision. Many problems can be formulated and modeled in a mathematical language as the minimization of some energy function or functional. Equally important, and perhaps even more challenging, is the development of methods for solving the minimization problems. For computational complexity reasons, it is often not possible to compute exact solutions, except for some simpler special cases. We have seen that it is still possible to solve the problems very closely and often exactly. In the future, energy minimization models will only get more complicated, and the development of good optimization methods will become even more important. We now have a rather good knowledge of optimization problems in imaging and vision that can be solved exactly and their respective algorithms. Most new models that are developed can be expected to be NP-hard. A key characteristic of image processing and computer vision problems is their large scale nature. One assume a theoretically continuum of variables. Algorithms that have a higher than polynomial complexity in the number of pixels become very impractical. The large scale nature of the problems also raises new questions and research opportunities. Can one develop simpler (e.g. convex) models that approximates the NP-hard models tightly? Is it possible that the approximation error decreases as the resolution increases? Can conditions be derived for when the simpler model provide a solution to the original model, even if it is NP-hard? Optimization is a powerful and expressive modeling language in image/signal processing and computer vision. At the same time, the application of optimization in these fields raises many interesting challenges and opportunities for new mathematical developments of high practical value.

Bibliography

[1] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Comm. Pure Appl. Math.*, 43:9991036, 1990.

[2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. The Clarendon Press, Oxford University Press, New York, 2000.

[3] M. Bergtholdt, D. Cremers, and C. Schnörr. Variational segmentation with shape priors. In O. Faugeras N. Paragios, Y. Chen, editor, *Handbook of Mathematical Models in Computer Vision*. Springer, 2005.

[4] J. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B*, 48(3):259302, 1986.

[5] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Ninth IEEE International Conference on Computer Vision, 2003. Proceedings*, pages 26–33, 2003.

[6] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:359–374, 2001.

[7] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222 – 1239, 2001.

[8] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.

[9] Ethan S. Brown, Tony F. Chan, and Xavier Bresson. A convex relaxation method for a class of vector-valued minimization problems with applications to mumford-shah segmentation. *UCLA, Applied Mathematics, CAM-report-10-43*, July, 2010.

[10] Jamylle Laurice Carter. Dual methods for total variation-based image restoration (ph.d. thesis). *UCLA, Applied Mathematics, CAM-report-02-13*, April, 2002.

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

[12] A. Chambolle. Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition: 5th International Workshop*, volume 3757, pages 136–152. Springer, 2005.

[13] Antonin Chambolle. Image segmentation by variational methods: Mumford and shah functional and the discrete approximations. *SIAM J. Appl. Math.*, 55:827–863, June 1995.

[14] Antonin Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, January 2004.

[15] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical Foundations and Numerical Methods for Sparse Recovery*, page 263340, 2010.

[16] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM J. SCI. COMPUT*, 22(2):503–516, 2000.

[17] T. Chan and L.A. Vese. *Active contours without edges.* IEEE Image Proc., 10, pp. 266-277, 2001.

[18] T.F. Chan, S.H. Kang, and J. Shen. Euler's elastica and curvature-based inpainting. *SIAM Journal on Applied Mathematics*, pages 564–592, 2002.

[19] Tony F. Chan, Selim Esedoḡlu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648 (electronic), 2006.

[20] Tony F. Chan, Gene H. Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20:1964–1977, May 1999.

[21] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiway cuts (extended abstract). In *STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 241–251, New York, NY, USA, 1992. ACM.

[22] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, 2006.

[23] Jérôme Darbon. Global optimization for first order markov random fields with submodular priors. *Discrete Appl. Math.*, 157(16):3412–3423, 2009.

[24] Andrew Delong, Anton Osokin, Hossam Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. In *CVPR*, 2010.

[25] Ivar Ekeland and Roger Téman. *Convex analysis and variational problems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[26] Ernie Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *UCLA cam-report 09-31*, April 2009.

[27] K. J. Falconer. *The geometry of fractal sets*. Cambridge University Press, New York, NY, USA, 1986.

[28] Hebert Federer. *Geometric measure theory*. Springer-Verlag New York Inc., New York, 1969, 2000.

[29] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[30] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 6, 1984.

[31] A. A. Goldstein. Convex programming in hilbert space. *Bull. Amer. Math. Soc*, 70(5):709–710, 1964.

[32] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009.

[33] J. B. Greer and A. L. Bertozzi. Traveling wave solutions of fourth order pdes for image processing. *SIAM J. Math. Anal*, 36:38–68, 2004.

[34] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, pages 271–279, 1989.

[35] D.M. Greig, B.T. Porteous, and A.H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989.

[36] M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303320, 1969.

[37] B. Merriman S.Osher H.K. Zhao, T.F. Chan. A variational level set approach to multiphase motion. *J. Comput. Phys.*, 127:179–195, 1996.

[38] D. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM*, 48(2):686–701, 2001.

[39] Erlend Hodneland, Xue-Cheng Tai, and Hans-Hermann Gerdes. Four-color theorem and level set methods for watershed segmentation. *International Journal of Computer Vision*, 82(3):264–283, 2009.

[40] J. M. Hyman and M. J. Shashkov. Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids. *Appl. Numer. Math.*, 25(4):413–442, 1997.

[41] J. M. Hyman and M. J. Shashkov. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.*, 33(4):81–104, 1997.

[42] Hiroshi Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1333–1336, 2003.

[43] H. Ratlif J. Picard. Minimum cuts and related problem. *Networks*, 5:357–370, 1975.

[44] L. Hurwicz K. J. Arrow and H. Uzawa. *Studies in linear and non-linear programming*. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford University Press, Stanford, Calif., 1958.

[45] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[46] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006.

[47] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:65–81, 2004.

[48] Nikos Komodakis and Georgios Tziritas. Approximate labeling via graph-cuts based on linear programming. In *In Pattern Analysis and Machine Intelligence*, page 2007, 2007.

[49] E. Fatemi L. Rudin, S. Osher. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[50] J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal-dual iterations for variational problems in computer vision. *European Conference on Computer Vision (ECCV)*, 2010.

[51] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In X.-C. Tai, K. Mórken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision (SSVM 2009)*, volume 5567 of *LNCS*, pages 150–162. Springer, 2009.

[52] J. Lellmann and C. Schnörr. Continuous multiclass labeling approaches and algorithms. *CoRR*, abs/1102.5448, 2011.

[53] E. S. Levitin and B. T. Polyak. Constrained minimization problems. *USSR. Comput. Math. Math. Phys.*, 6:1–50, 1966.

[54] J. Lie, M. Lysaker, and X.C. Tai. Piecewise constant level set methods and image segmentation. In *Scale Space and PDE Methods in Computer Vision*, pages 573–584. Springer, 2005.

[55] J. Lie, M. Lysaker, and X.C. Tai. A binary level set model and some applications to mumford-shah image segmentation. *IEEE Transactions on Image Processing*, 15(5):1171–1181, 2006.

[56] J. Lie, M. Lysaker, and X.C. Tai. A variant of the level set method and applications to image segmentation. *Math. Comp.*, 75(255):1155–1174 (electronic), 2006.

[57] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[58] M. Lysaker, A. Lundervold, and X.C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, 2003.

[59] M. Lysaker, S. Osher, and X.C. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Transactions on Image Processing*, 13(10):1345–1357, 2004.

[60] Simon Masnou and Jean-Michel Morel. Level lines based disocclusion. In *ICIP (3)*, pages 259–263, 1998.

[61] Jean Michel Morel and Sergio Solimini. *Variational methods in image segmentation*. Birkhauser Boston Inc., Cambridge, MA, USA, 1995.

[62] D. Mumford. Elastica and computer vision. In *Algebraic Geometry and its Applications*, 1994.

[63] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math, 42*, 42:577–685, 1989.

[64] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, 2006.

[65] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[66] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[67] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.

[68] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

[69] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. Global solutions of variational models with convex regularization. Technical report, Institute for Computer Graphics and Vision, Graz University of Technology, 2010.

[70] Thomas Pock, Thomas Schoenemann, Gottfried Graber, Horst Bischof, and Daniel Cremers. A convex formulation of continuous multi-label problems. In *ECCV '08*, pages 792–805, 2008.

[71] L. Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical Notes*, 28(5):845848, 1980.

[72] Renfrey B. Potts. Some generalized order-disorder transformations. *In Proceedings of the Cambridge Philosophical Society, Vol. 48*, pages 106–109, 1952.

[73] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization ed. by R. Fletcher, Academic Press, New York, NY*, 1969.

[74] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Methematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[75] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[76] T. Schoenemann and D. Cremers. Introducing curvature into globally optimimal image segmentation: Minimum ratio cycles on product graphs. In *IEEE International Conference on Computer Vision (ICCV). Rio de Janeiro, Brazil*, 2007.

[77] Simon Setzer. Split bregman algorithm, douglas-rachford splitting and frame shrinkage. In *Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, SSVM '09, pages 464–476, Berlin, Heidelberg, 2009. Springer-Verlag.

[78] Gil. Strang. Maximal flow through a domain. *Mathematical Programming*, 26:123–143, 1983.

[79] X.C. Tai, O. Christiansen, P. Lin, and I. Skjaelaaen. Image segmentation using some piecewise constant level set methods with mbo type of project. *International Journal of Computer Vision*, 73:61–76, 2007.

[80] Xue-Cheng Tai, Jooyoung Hahn, and Ginmo Jason Chung. A fast algorithm for euler's elastica model using augmented lagrangian method. Technical report CAM10-47, UCLA, CAM, July 2010.

[81] Xue-Cheng Tai and Chunlin Wu. Augmented lagrangian method, dual methods and split bregman iteration for rof model. In *Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, SSVM '09, pages 502–513, Berlin, Heidelberg, 2009. Springer-Verlag.

[82] W. Tao and X. Tai. Multiple Piecewise Constant Active Contours for Image Segmentation using Graph Cuts Optimization. *UCLA cam-report 09-13*, 2009.

[83] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher. Geometric surface smoothing via anisotropic diffusion of normals. *IEEE Visualization*, pages 125–132, 2002.

[84] L. A. Vese and T. F. Chan. A new multiphase level set framework for image segmentation via the mumford and shah model. *International Journal of Computer Vision*, 50:271–293, 2002.

[85] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, 51:3697–3717, 2002.

[86] Laurence A. Wolsey. *Integer programming*. Wiley-Interscience Series in Discrete Mathematics and Optimization, 1998.

[87] Chunlin Wu and Xue-Cheng Tai. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. Technical report CAM09-76, UCLA, CAM, September 2009.

[88] Chunlin Wu and Xue-Cheng Tai. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. *SIAM J. Imaging Sci.*, 3:300–339, 2010.

[89] Yu-Li You and Mostafa Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, 2000.

[90] Nicholas Young. *An Introduction to Hilbert Space*. Cambridge University Press, 1988.

[91] Jing Yuan, Christoph Schnorr, and Gabriele Steidl. Total-Variation Based Piecewise Affine Regularization. In *Second International Conference, SSVM. Voss, Norway*, pages 552–564. Springer, 2009.

[92] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*, 2008.

[93] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical report CAM08-34, UCLA, CAM, May 2008.

[94] Song Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *PAMI*, 18:884–900, 1996.

[95] Wei. Zhu and Tony. Chan. Image denoising using mean curvature, Available in: http://www.cims.nyu.edu/ wzhu/meancurvature06Nov30.pdf. 2009.