

Efficient Global Minimization Methods for Image Segmentation Models with Four Regions

Egil Bae and Xue-Cheng Tai

the date of receipt and acceptance should be inserted later

Abstract We propose an exact global minimization framework for variational image segmentation models, such as the Chan-Vese model, involving four regions. A global solution is guaranteed if the data term satisfies a mild condition. It is shown theoretically and experimentally that such a condition holds in practice for the most commonly used type of data terms, such as the Chan-Vese model and Mumford Shah model. If the condition is violated, convex and submodular relaxations are proposed which are not guaranteed to produce exact solutions, but tends to do so in practice. We also build up a convex relaxation for Pott's model with four regions, which is at least as tight as the tightest existing relaxation, but significantly simpler. Algorithms are proposed which are very efficient due to the simple formulations.

1 Introduction

Image segmentation is one of the core problems in image processing and computer vision. Based on the intensity values of the input image, one would like to partition the image domain into several regions, each representing an object. Energy minimization formulations of such problems have demonstrated to be especially powerful, and have been developed independently in the discrete and variational community. The Mumford-Shah model [34] and Chan-Vese model [15, 39] have been established as some of the most fundamental variational image segmentation models, whereas Pott's model has become one of the most important discrete optimization models. Pott's model and the Mumford-Shah model are closely related in the limit as the number of pixels goes to infinity.

Minimization of the energy in these models poses a fundamental challenge from a computational point of view. In a discrete setting Pott's model is NP-hard, therefore an algorithm for minimizing the energy exactly with reasonable efficiency is unlikely to exist. Numerical approaches for the variational models, such as the level set method, involves the minimization of non-convex energy functionals. Therefore algorithms for minimizing the energy may easily get stuck in poor local minima close to the initialization.

One notable exception where efficient global minimization methods are available is segmentation problems with 2 regions. Potts model restricted to two regions is computationally tractable in the discrete setting and can be minimized by established algorithms such as max-flow/min-cut (graph cuts) [19, 9]. Convex reformulations of variational segmentation models with 2 regions have been proposed in [19], which can be used to design algorithms for computing global minima.

If the number of regions is larger than two, there are no available algorithms that can produce global minima. A level set formulation of the Chan-Vese model with multiple regions appeared in [39], which has since become very popular. It was proposed to solve the resulting gradient descent equations numerically, leading to a local minima of the non-convex energy functional close to the initialization. In a discrete setting, alpha-expansion and alpha-beta swap [9, 10] are the most popular algorithms for approximately minimizing the energy in Pott's model. More recently, attempts to derive convex relaxations in a continuous variational setting have been proposed [42, 28, 35, 6, 12]. Instead of solving the original non-convex problem to a local minimum, a convex "relaxation" of the problem is solved globally. These approaches cannot in general produce global solutions of the original problems, but lead to good ap-

proximations in many practical situations. If the regularization term promotes a linear inclusion property of the regions the optimization problem becomes easier, and the energy can be minimized exactly by graph cuts [24], but this assumption is usually not valid in image segmentation applications.

The main advantages of variational models over discrete optimization models is the rotational invariance and ability to accurately represent geometrical entities such as curve length and surface area without grid bias. The discrete models are 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. On the other hand, discrete optimization problems are much easier to analyze and characterize by the established field of computational complexity. It is also easier to design global minimization algorithms in the discrete setting, by applying established combinatorial optimization algorithms. This includes in particular algorithms for max-flow and min-cut, which may also be very efficient under certain implementations [9]. However, a disadvantage is that these algorithms do not parallelize as easily, in contrast to continuous optimization algorithms which are suitable for massive parallel implementation on graphics processing units (GPUs).

1.1 Contributions

This paper proposes an exact global minimization framework for segmentation problems with 4 regions in the level set framework of Vese and Chan [39], both in a discrete setting and in a convex variational setting. Because of a slight simplification of the regularization term in this model, global minimization is not NP-hard. First, in Section 3, it is shown that a discrete version of the model can be minimized globally by computing the minimum cut a novel graph, under a (mild) submodularity condition on the data term. In Section 5, a reformulation of the Chan-Vese model is proposed in the variational setting, which is convex under the same condition on the data term that made the discrete version submodular. A thresholding scheme, similar to the one that appeared in [16] for two region problems, is proposed for converting solutions of the convex relaxed problem into global solutions of the original problem.

It is shown theoretically (Section 4) and experimentally (Section 9.1) that the condition on the data term, which guarantees a global minimum, is expected to hold for commonly used data terms, such as those in the Mumford-Shah model and Chan-Vese model. If the condition does not hold, simpler submodular and convex relaxations are proposed in Section 6. The relaxations are not guaranteed to produce a global minimum of the

original problems, but conditions on the computed solution are derived for when they do. Experiments demonstrate that global solutions can be obtained in practice also in these cases.

In section 7, a new convex relaxation for variational partitioning problems with Potts regularization is proposed, by building on the previous results. The relaxation is at least as tight as the tightest existing relaxation [35], but significantly simpler. In consequence, it is much easier to handle computationally. We also believe the proposed relaxation is the strictly tightest that exists, but a proof of this will be quite involved and is postponed to a future paper. Instead some arguments are given to support the claim.

Efficient algorithms for the new convex formulations and relaxations are proposed in Section 8, inspired by previous work on continuous max-flow [40, 3].

We focus fully on problems with 4 regions in this work. Problems with 4 regions are important. By the 4 colour theorem, 4 regions suffice to describe any partition of a 2D image. It is an open problem how to incorporate this property into the segmentation models in general. The method applies directly in some settings like [23], where the four colour theorem was used to segment 2D images in a level set framework with four regions, in case rough a priori information of the object locations was provided in advance. Furthermore, in many applications one would like to divide the image into 4 regions, such as brain MRI segmentation where one wants to separate cerebrospinal fluid, gray matter, white matter and background. The formulations can also be extended to problems with more than four regions. However, it is complicated to derive a general framework that supports an arbitrary number of regions. Instead, each special case (e.g. 8, 16 regions etc.) should be investigated separately. Such generalizations will be the topic of a future paper.

The sections on discrete optimization: Section 3 and 6.1 also appeared in shorter form in our conference paper [4, 5]

2 Chan-Vese model, Potts model and the piecewise constant Mumford Shah model

The Pott's model originates from statistical mechanics [37], but has become fundamentally important in image processing and computer vision, especially as energy minimization formulation of segmentation and grouping problems. 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 generalized to the continuous setting attempts 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

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

Here $|\partial\Omega_i|$ denotes the size of the boundary of the subdomain Ω_i , *i.e.* the curve length in 2D and surface area in 3D. In this paper, we let N denote the dimension of Ω . The data cost functions f_i should depend in some sense on the input image I , an important example is

$$f_i(x) = |I(x) - c_i|^\beta, \quad i = 1, \dots, n \quad (2)$$

where $c_i \in \mathbb{R}$ are parameters associated with each Ω_i , for instance the mean intensity value inside Ω_i when $\beta = 2$. The piecewise constant Mumford Shah model [34] has the form of (1) with data terms (2) and $\beta = 2$, but minimizes additionally over the parameters $c_i \in \mathbb{R}$

$$\begin{aligned} \min_{\{\Omega_i\}_{i=1}^n, \{c_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} |I(x) - c_i|^\beta dx + \nu \sum_{i=1}^n |\partial\Omega_i| \quad (3) \\ \text{s.t. } \cup_{i=1}^n \Omega_i = \Omega, \quad \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l, \end{aligned}$$

For fixed c , (3) has the same form as Pott's model. A simple algorithm can be constructed which alternatively minimizes (1) with respect to c_i and Ω_i until convergence. Although a global solution cannot be guaranteed, such a scheme is quite robust as shown in [2]. There has also been attempts to derive convex relaxations for the joint minimization problem (3) in case $n = 2$ [11]. The most challenging problem is to optimize in terms of the regions, and that will be the topic of this paper.

2.1 Representation by level set functions

As a numerical realization, Chan and Vese [15,39] proposed to represent the Mumford-Shah model with level set functions, and solve the resulting gradient descent equations numerically. In case of two region ($n = 2$), the problem can be expressed in terms of a level set function ϕ which satisfies $\phi(x) > 0$ for $x \in \Omega_1$ and $\phi(x) < 0$ for $x \in \Omega_2$ as

$$\min_{\phi} \int_{\Omega} \{H(\phi)f_1 + (1 - H(\phi))f_2\} + \nu |\nabla H(\phi)| dx \quad (4)$$

where $H(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ is the Heaviside function $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$.

Instead of using the non-convex heaviside functions, the problem can also be written directly in terms of a binary function ϕ such that $\phi(x) = 1$ for $x \in \Omega_1$ and

$\phi(x) = 0$ for x in Ω_2 . This was first done in [31,32], coined the piecewise constant level set method (PCLSM), and resulted in the energy functional

$$\min_{\phi \in \mathcal{B}} \int_{\Omega} \{\phi f_1 + (1 - \phi)f_2\} dx + \nu \int_{\Omega} |\nabla \phi| dx. \quad (5)$$

In this paper we use the notation \mathcal{B} for the set of binary functions, *i.e.*

$$\mathcal{B} = \{\phi : \phi(x) \in \{0, 1\}, \forall x \in \Omega\}. \quad (6)$$

In [39], Vese and Chan proposed a multiphase level set framework for the piecewise constant Mumford Shah model. By using $m = \log_2(n)$ level set functions, denoted ϕ^1, \dots, ϕ^m , n regions could be represented in terms of the nonconvex heaviside functions $H(\phi^1), \dots, H(\phi^m)$. An important special case is the representation of 4 regions by two level set functions ϕ^1, ϕ^2 as follows

$$\min_{\phi^1, \phi^2} E(\phi^1, \phi^2) = \nu \int_{\Omega} |\nabla H(\phi^1)| dx + \nu \int_{\Omega} |\nabla H(\phi^2)| dx \quad (7)$$

$$\begin{aligned} + \int_{\Omega} \{H(\phi^1)H(\phi^2)f_2 + H(\phi^1)(1 - H(\phi^2))f_1 \\ + (1 - H(\phi^1))H(\phi^2)f_4 + (1 - H(\phi^1))(1 - H(\phi^2))f_3\} dx. \end{aligned}$$

The above model can also be formulated directly in terms of two binary functions ϕ^1, ϕ^2 , which represents the 4 regions as in Table 1. The resulting energy functional is then

$$\begin{aligned} \min_{\phi^1, \phi^2 \in \mathcal{B}} E(\phi^1, \phi^2) \\ = \nu \int_{\Omega} |\nabla \phi^1| dx + \nu \int_{\Omega} |\nabla \phi^2| dx + E^{data}(\phi^1, \phi^2), \quad (8) \end{aligned}$$

subject to (6), where

$$\begin{aligned} E^{data}(\phi^1, \phi^2) = \int_{\Omega} \{\phi^1 \phi^2 f_2 + \phi^1 (1 - \phi^2) f_1 \\ + (1 - \phi^1) \phi^2 f_4 + (1 - \phi^1) (1 - \phi^2) f_3\} dx. \end{aligned}$$

where f was given by (2) with $\beta = 2$. The problem (8) is nonconvex because the binary constraints (6) are nonconvex and the energy functional $E^{data}(\phi^1, \phi^2)$ is nonconvex in ϕ^1, ϕ^2 . The above level set formulation of the Mumford-Shah model is often referred to as the *Chan-Vese model*.

Note that we have made a permutation in the interpretation of the regions compared to [39], see Table 1. It can be checked that the energy is still exactly the same for all possible ϕ^1, ϕ^2 (if the data functions f_i are permuted accordingly). This permutation is crucial for making the corresponding discrete energy function submodular.

As pointed out in [39], the level set formulation (7), (8) does not correspond exactly to the length term in Pott's model, because two of the six boundaries are counted twice in (8), namely the boundary between Ω_1 and Ω_4 and the boundary between Ω_2 and Ω_3 . The remaining 4 boundaries are counted once.

	Original	Permuted
$x \in \text{region 1}$ iff	$\phi^1(x) = 1, \phi^2(x) = 1$	$\phi^1(x) = 1, \phi^2(x) = 0$
$x \in \text{region 2}$ iff	$\phi^1(x) = 1, \phi^2(x) = 0$	$\phi^1(x) = 1, \phi^2(x) = 1$
$x \in \text{region 3}$ iff	$\phi^1(x) = 0, \phi^2(x) = 1$	$\phi^1(x) = 0, \phi^2(x) = 0$
$x \in \text{region 4}$ iff	$\phi^1(x) = 0, \phi^2(x) = 0$	$\phi^1(x) = 0, \phi^2(x) = 1$

Table 1 Representation of four regions by two binary functions. The interpretation of the regions can be permuted without influencing the energy.

2.1.1 Discrete Approximations

The variational problems in the last section can also be formulated in the discrete setting as combinatorial optimization problems. Let us first mention there are two variants of the total variation term. The isotropic variant, by using 2-norm

$$TV_2(\phi) = \int_{\Omega} |\nabla \phi|_2 dx = \int_{\Omega} \sqrt{|\phi_{x_1}|^2 + \dots + |\phi_{x_N}|^2} dx \quad (9)$$

and the anisotropic variant, by using 1-norm

$$TV_1(\phi) = \int_{\Omega} |\nabla \phi|_1 dx = \int_{\Omega} |\phi_{x_1}| + \dots + |\phi_{x_N}| dx. \quad (10)$$

The anisotropic version is not rotationally invariant and will therefore favor results that are aligned along the coordinate system. The isotropic variant is preferred, but cannot be handled by discrete optimization algorithms (e.g. mapped to the cut on a graph).

Let \mathcal{P} denote the set of grid points, and \mathcal{N}_p^k denote the set of k nearest neighbors of $p \in \mathcal{P}$. In case $N = 2$, $\mathcal{P} = \{(i, j) \in \mathbb{Z}^2\}$ and for each $p = (i, j) \in \mathcal{P}$

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

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

Let ϕ_p^1 and ϕ_p^2 denote the function values of ϕ^1 and ϕ^2 at $p \in \mathcal{P}$ and denote the set of binary functions as

$$\mathcal{B} = \{\phi : \phi_p \in \{0, 1\}, \forall p \in \mathcal{P}\} \quad (11)$$

A discrete approximation of the two region model (5) can be derived as

$$\min_{\phi \in \mathcal{B}} \sum_{p \in \mathcal{P}} \phi_p f_1(p) + (1 - \phi_p) f_2(p) + \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p - \phi_q|, \quad (12)$$

where the usual choice of the data functions f are the discrete version of (2)

$$f_i(p) = |c_i - u_p^0|^\beta. \quad (13)$$

If the weights are set to $w_{pq} = 1$, and the neighborhood system is set to 4 ($k = 4$), the last term corresponds to a forward discretization of the anisotropic total variation of ϕ . The weights w_{pq} can be derived by the Cauchy-Crofton formula of integral geometry as in [8], to approximate the isotropic total variation (8) (euclidian curve length). However, this requires that both the mesh size goes to zero and the number of neighbors in the neighborhood system \mathcal{N}^k goes to infinity, which complicates computation.

In the same manner, a discrete approximation of the model with four regions (8) can be expressed as

$$\begin{aligned} \min_{\phi^1, \phi^2 \in \mathcal{B}} E_d(\phi^1, \phi^2) &= \sum_{p \in \mathcal{P}} E_p^{data}(\phi_p^1, \phi_p^2) \\ &+ \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^1 - \phi_q^1| + \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^2 - \phi_q^2| \end{aligned} \quad (14)$$

where

$$\begin{aligned} E_p^{data}(\phi_p^1, \phi_p^2) &= \{\phi_p^1 \phi_p^2 f_2(p) + \phi_p^1 (1 - \phi_p^2) f_1(p) \\ &+ (1 - \phi_p^1) \phi_p^2 f_4(p) + (1 - \phi_p^1) (1 - \phi_p^2) f_3(p)\}. \end{aligned} \quad (15)$$

2.2 Related work on global optimization

2.2.1 Two regions

In case of two regions, the Chan-Vese/Mumford-Shah/Potts model can be minimized globally. The discrete version can be minimized by computing the minimum cut on a graph [19]. In the continuous setting, a convex relaxation approach can be applied as shown in [16]. The idea is to relax the non-convex binary constraint by the convex constraint $\phi(x) \in [0, 1], \forall x \in \Omega$. It was shown that thresholding the solution of the relaxed problem at almost any threshold in $(0, 1]$ yields an optimal solution of the original problem. Since (5) is convex, this procedure would yield the globally optimal solution.

One might immediately think the same idea could be extended to the multiphase case by iteratively minimizing (8) for ϕ^1 and ϕ^2 in \mathcal{B}' and finally threshold the results. However, since $E^{data}(\phi^1, \phi^2)$ is not convex with respect to ϕ^1 and ϕ^2 , the minimization would not be global.

2.2.2 Multiple linearly ordered labels

In case the number of regions is larger than two, it is possible to solve exactly the simpler problem where the regularization term promotes a linear inclusion property of the regions. Define the "labeling function" u over the continuous image domain Ω or the discrete grid point \mathcal{P} . Any partition $\{\Omega_i\}_{i=1}^n$ can be described in terms of u by the convention $u(x) = i$, $\forall x \in \Omega_i$, $i = 1, \dots, n$. Equivalently, a partition $\{\mathcal{P}_i\}_{i=1}^n$ of the discrete domain \mathcal{P} can be expressed in u as $u_p = i$, $\forall p \in \mathcal{P}_i$, $i = 1, \dots, n$. Consider the following integer constrained problem

$$\min_u \sum_{p \in \mathcal{P}} f_{u_p}(p) + \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |u_p - u_q| \quad (16)$$

subject to $u_p \in \{1, \dots, n\}$, $\forall p \in \mathcal{P}$. A continuous equivalent can be defined as

$$\min_u \int_{\Omega} f_{u(x)}(x) + \nu \int_{\Omega} |\nabla u| dx, \quad (17)$$

subject to $u(x) \in \{1, \dots, n\}$, $\forall x \in \Omega$, which was first done in [30]. In (17) and (16) the data term $f_{u(x)}(x)$ is the data cost of assigning x to region $\Omega_{u(x)}$. However, the regularization term of (17) does not correspond to the length term in the Pott's model (1) because of its dependency size of the discontinuities of u . Instead of penalizing the jump from each region to the next equally, the regularization term overpenalizes the boundary between regions Ω_i and Ω_j where the indices i and j differ by more than one. The overpenalization is much more severe than the representation (8), see Figure 1 for an illustration. Such an overpenalization may cause the boundary between such regions to split.

Ishikawa [24] showed that the discrete model (16) can be minimized globally by computing the minimum cut on a graph. Later, a convex optimization framework was established for the continuous version (17) in [36].

2.2.3 Convex relaxations for Pott's model

Convex relaxation methods have recently been proposed for approximately minimizing the energy in the Pott's model with more than two regions $n > 2$ [42, 28, 35, 13, 6]. However, unlike the relaxation for the model with two regions (5), the model (17) with linearly ordered labels, and the model (8) which will be presented in this work, such relaxation approaches are not in general exact. They may result in exact global solutions in some situations, but will otherwise produce approximate solutions. The relaxation of [35, 13] is tightest, meaning it will lead to the highest energy before binarization (thresholding).

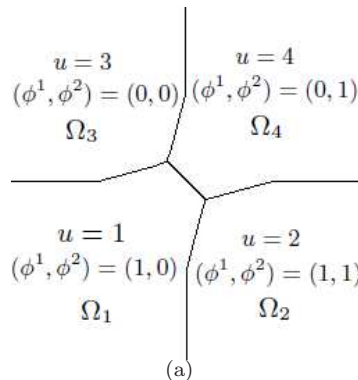


Fig. 1 The model (17) overcounts more severely than (8). In the model (17), the transition $\Omega_1 - \Omega_4$ is penalized 3 times, the transitions $\Omega_1 - \Omega_3$ and $\Omega_2 - \Omega_4$ are penalized 2 times while transitions $\Omega_1 - \Omega_2$ and $\Omega_3 - \Omega_4$ are penalized once. In the model (8), the transition $\Omega_1 - \Omega_4$ is penalized 2 times, while all the other transitions are penalized once.

3 Global minimization of 4-region Chan-Vese model in discrete setting by graph cuts

In this section we show how the discrete approximation of the Chan-Vese model (14) can be minimized exactly by computing the minimum cut on a novel graph. In section 5-8, a different approach is proposed in a variational setting based on convex optimization.

3.1 Brief review of Max-flow and Min-cut

Min-cut and max-flow are optimization problems defined over a graph which are dual to each others. Important energy minimization problems in image processing and computer vision can be represented as min-cut or max-flow problems over certain graphs, and be optimized globally by established efficient max-flow algorithms. Such a min-cut/max-flow approach is often called graph cuts in computer vision [18, 7, 9].

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a set of vertices \mathcal{V} and a set of directed edges \mathcal{E} . We let (v, w) denote the directed edge going from vertex v to vertex w , and let $c(v, w)$ denote the weight on this edge. 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 $(w, v) \in \mathcal{E}$. In the graph cut scenario there are two distinguished vertices in \mathcal{V} , the source $\{s\}$ and the sink $\{t\}$. A cut on \mathcal{G} is a partitioning of the vertices \mathcal{V} into two disjoint connected sets $(\mathcal{V}_s, \mathcal{V}_t)$ such that $s \in \mathcal{V}_s$ and $t \in \mathcal{V}_t$. The cost of the cut is defined as

$$c(\mathcal{V}_s, \mathcal{V}_t) = \sum_{(v,w) \in \mathcal{E} \text{ s.t. } v \in \mathcal{V}_s, w \in \mathcal{V}_t} c(v, w). \quad (18)$$

The minimum cut problem is the problem of finding a cut of minimum cost.

The maximum flow problem can be defined over the same graph. A flow p on \mathcal{G} is a function $p : \mathcal{E} \mapsto \mathbb{R}$. The weights $c(e)$ are upper bounds (capacities) on the flows $p(e)$ for all $e \in \mathcal{E}$, i.e.

$$p(e) \leq c(e), \quad \forall e \in \mathcal{E} \quad (19)$$

For a given flow p , the residual capacities are defined as $R(e) = c(e) - p(e) \forall e \in \mathcal{E}$. In addition, flow conservation is required at each vertex except s and t

$$\sum_{w \in N^+(v)} p(v, w) - \sum_{w \in N^-(v)} p(w, v) = 0, \quad \forall v \in \mathcal{V} \quad (20)$$

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

$$\max_p \sum_{v \in \mathcal{V}} p(s, v) \quad (21)$$

subject to (19) and (20). The theorem of Ford and Fulkerson says this is dual to the problem of finding the cut of minimum cost on \mathcal{G} , the min-cut problem. Therefore, efficient algorithms for finding max-flow, such as the augmented paths method [27] can be used for finding minimum cuts in graphs. An efficient implementation of this algorithm specialized for image processing problems can be found in [7]. This algorithm, which is available on-line has been used in our experiments.

Graph cuts have been used in computer vision for minimizing energy functions of the form

$$\min_{x_i \in \{0,1\}} E^i(x_i) + \sum_{i < j} E^{i,j}(x_i, x_j),$$

where typically E^i is a data term, $E^{i,j}$ is a regularization term, i is the index of each grid point (pixel) and x_i is a binary variable defined for each grid point. In order to be representable as a cut on a graph, it is required that the energy function is submodular [25, 18], i.e. the regularization term must satisfy

$$E^{i,j}(0, 0) + E^{i,j}(1, 1) \leq E^{i,j}(0, 1) + E^{i,j}(1, 0), \quad \forall i < j$$

3.2 Graph construction for energy minimization over multiple regions

Observe that in the discrete energy function (14), not only the regularization term, but also the data term is

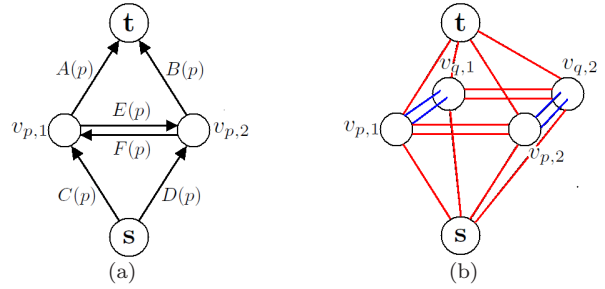


Fig. 2 (a) The graph corresponding to the data term at one grid point p . (b) A sketch of the graph corresponding to the energy function of a 1D signal of two grid points p and q . Data edges are depicted as red edges and regularization edges are depicted as blue arrows.

composed of pairwise interactions between binary variables. In this section we will construct a graph \mathcal{G} such that there is a one-to-one correspondence between cuts on \mathcal{G} and the binary functions ϕ^1 and ϕ^2 , provided the data term is submodular, i.e.

$$E_p^{data}(1, 1) + E_p^{data}(0, 0) \leq E_p^{data}(1, 0) + E_p^{data}(0, 1) \quad (22)$$

for each $p \in \mathcal{P}$. Furthermore, the minimum cost cut will correspond to binary functions ϕ^1 and ϕ^2 that minimize the energy (14)

$$\min_{(\mathcal{V}_s, \mathcal{V}_t)} c(\mathcal{V}_s, \mathcal{V}_t) = \min_{\phi^1, \phi^2 \in \mathcal{B}} E_d(\phi^1, \phi^2) + \sum_{p \in \mathcal{P}} \sigma_p. \quad (23)$$

where $\sigma_p \in \mathbb{R}$ are fixed for each $p \in \mathcal{P}$. In the graph, two vertices are associated with each grid point $p \in \mathcal{P}$. They are denoted $v_{p,1}$ and $v_{p,2}$, and correspond to each of the level set functions ϕ^1 and ϕ^2 . Hence the set of vertices is formally defined as

$$\mathcal{V} = \{v_{p,i} \mid p \in \mathcal{P}, i = 1, 2\} \cup \{s\} \cup \{t\}. \quad (24)$$

The edges are constructed such that the relationship (23) is satisfied. We begin with edges constituting the data term of (14). For each grid point $p \in \mathcal{P}$ they are defined as

$$\mathcal{E}_D(p) = (s, v_{p,1}) \cup (s, v_{p,2})$$

$$\cup (v_{p,1}, t) \cup (v_{p,2}, t) \cup (v_{p,1}, v_{p,2}) \cup (v_{p,2}, v_{p,1}). \quad (25)$$

The set of all data edges are denoted \mathcal{E}_D and defined as $\cup_{p \in \mathcal{P}} \mathcal{E}_D(p)$. The edges corresponding to the regularization term are defined as

$$\mathcal{E}_R = \{(v_{p,1}, v_{q,1}), (v_{p,2}, v_{q,2}) \mid p, q \subset \mathcal{P} \text{ s.t. } q \in \mathcal{N}_p^k\}. \quad (26)$$

For any cut (V_s, V_t) , the corresponding level set functions are defined by

$$\phi_p^1 = \begin{cases} 1 & \text{if } v_{p,1} \in V_s, \\ 0 & \text{if } v_{p,1} \in V_t, \end{cases} \quad \phi_p^2 = \begin{cases} 1 & \text{if } v_{p,2} \in V_s, \\ 0 & \text{if } v_{p,2} \in V_t. \end{cases} \quad (27)$$

Weights are assigned to the edges such that the relationship (23) is satisfied. Weights on the regularization edges are simply given by

$$\begin{aligned} c(v_{p,1}, v_{q,1}) &= c(v_{q,1}, v_{p,1}) = c(v_{p,2}, v_{q,2}) = c(v_{q,2}, v_{p,2}) \\ &= \nu w_{pq}, \quad \forall p \in \mathcal{P}, q \in \mathcal{N}_p^k. \end{aligned} \quad (28)$$

We now concentrate on the weights on data edges \mathcal{E}_D . For grid point $p \in \mathcal{P}$, let

$$c(v_{p,1}, t) = A(p), \quad c(v_{p,2}, t) = B(p), \quad c(s, v_{p,1}) = C(p), \quad (29)$$

$$c(s, v_{p,2}) = D(p), \quad c(v_{p,1}, v_{p,2}) = E(p), \quad c(v_{p,2}, v_{p,1}) = F(p).$$

In Figure 2(a) the graph corresponding to an image of one pixel p is shown. It is clear that these weights must satisfy

$$\begin{cases} A(p) + B(p) &= f_2(p) + \sigma_p \\ C(p) + D(p) &= f_3(p) + \sigma_p \\ A(p) + E(p) + D(p) &= f_1(p) + \sigma_p \\ B(p) + F(p) + C(p) &= f_4(p) + \sigma_p \end{cases} \quad (30)$$

This is a non-singular linear system for the weights $A(p), B(p), C(p), D(p), E(p), F(p)$. Negative weights are not allowed. By choosing σ_p large enough there will exist a solution with $A(p), B(p), C(p), D(p) \geq 0$. However, the requirement $E(p), F(p) \geq 0$ implies that

$$\begin{aligned} f_1(p) + f_4(p) &= A(p) + B(p) + C(p) + D(p) + E(p) + F(p) - 2\sigma_p \\ &\geq A(p) + B(p) + C(p) + D(p) - 2\sigma_p = f_2(p) + f_3(p) \end{aligned}$$

for all $p \in \mathcal{P}$. By inserting the data term (15) for f , we obtain the following requirement

$$|c_2 - u_p^0|^\beta + |c_3 - u_p^0|^\beta \leq |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta, \quad (31)$$

for all $p \in \mathcal{P}$. Or, by assuming the image contains all gray values $I \in [0, L]$, a sufficient condition on the distribution of the constant values c_1, \dots, c_4

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta, \quad \forall I \in [0, L]. \quad (32)$$

If the distribution of the constant values satisfies (32), there exists a solution to (30) with $E(p), F(p) \geq 0$ for all $p \in \mathcal{P}$. Hence the problem can be solved globally by computing the minimum cut on the graph.

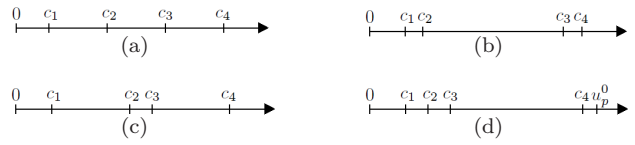


Fig. 3 (a), (b) and (c) distributions of \mathbf{c} which makes energy function submodular for all β . (d) distribution of \mathbf{c} which may make energy function nonsubmodular for small β .

4 Analysis of submodular/convexity condition

We give a detailed analysis of the sufficient submodular condition (32) in case of data terms of the form (2). The condition is also sufficient for making the new continuous variational formulation of (8) convex. The condition says something about how evenly $\{c_i\}_{i=1}^4$ are distributed. First we characterize situations for which (32) is always satisfied.

Proposition 1 *Let $0 \leq c_1 < c_2 < c_3 < c_4$. (32) is satisfied for all $I \in [\frac{c_2 - c_1}{2}, \frac{c_4 - c_3}{2}]$ for any $\beta \geq 1$.*

Proposition 2 *Let $0 \leq c_1 < c_2 < c_3 < c_4$. (32) is satisfied for any $\beta \geq 1$ if $c_2 - c_1 = c_4 - c_3$.*

The proof of Prop 1, 2, 3 and 4 can be found in the appendix. Further, it can be observed that (32) becomes less strict as β increases, as the next two results show.

Proposition 3 *Let $0 \leq c_1 < c_2 < c_3 < c_4$. If (32) is satisfied for some $\beta_0 \geq 1$, then (32) is satisfied for all $\beta \geq \beta_0$.*

Proposition 4 *Let $0 \leq c_1 < c_2 < c_3 < c_4$. There exists a $C \in \mathbb{N}$ such that (32) is satisfied for any $\beta \geq C$.*

In fact we have observed that for $\beta = 2$, (32) is always satisfied in practice for optimal constant values.

Examples where the condition is satisfied and may fail are depicted in Figure 3. Prop. 3 is illustrated in Figure 3 (b) and (c). Figure 3 (d) shows the only possibility in which (32) may be violated, i.e. c_1, c_2, c_3 are clustered compared to c_4 (the symmetrical version would also be a problem). However, the model (3) will disfavor solutions where the constants are very clustered. Numerous experiments on images from the database [1] demonstrate that under L^2 data fidelity, (32) is always satisfied for optimal values of \mathbf{c} . Under L^1 data fidelity, it may be more easily violated.

In case (31) does not hold at some $p \in \mathcal{P}$, the energy function is non-submodular. Not only does this mean it cannot in general be minimized by graph cut. It also implies the minimization problem is NP-hard, hence no algorithm exists that is guaranteed to solve the problem in polynomial time (unless $P = NP$). In Section 6 relaxations are proposed which can produce global solutions for most practical problems.

5 Exact convex formulation of 4-region Chan-Vese model in the continuous setting

Recall that the energy functional of the optimization problem (8) is non-convex in ϕ^1, ϕ^2 . In this section we derive a formulation of (8) which is convex under the same condition (32) that made the discrete energy function (14) submodular, and hence allows for the computation of global minimizers.

For each $x \in \Omega$, let $A(x), B(x), C(x), D(x), E(x), F(x)$ be a solution to the linear system

$$\begin{cases} A(x) + B(x) &= f_2(x) + \sigma(x) \\ C(x) + D(x) &= f_3(x) + \sigma(x) \\ A(x) + E(x) + D(x) &= f_1(x) + \sigma(x) \\ B(x) + F(x) + C(x) &= f_4(x) + \sigma(x) \end{cases} \quad (33)$$

where $\sigma(x)$ can be an arbitrary number. This is the same linear system as (30). Define the weights

$$\begin{aligned} C_s^1(x) &= C(x), & C_s^2(x) &= D(x), \\ C_t^1(x) &= A(x), & C_t^2(x) &= B(x), \\ C^{12}(x) &= E(x), & C^{21}(x) &= F(x). \end{aligned} \quad (34)$$

It can be checked that the following problem is equivalent to the original problem (8)

$$\begin{aligned} \min_{\phi^1, \phi^2} E^P(\phi^1, \phi^2) &= \int_{\Omega} (1 - \phi^1(x))C_s^1(x) + (1 - \phi^2(x))C_s^2(x) dx \\ &+ \int_{\Omega} \phi^1(x)C_t^1(x) + \phi^2(x)C_t^2(x) dx \\ &+ \int_{\Omega} \max\{\phi^1(x) - \phi^2(x), 0\}C^{12}(x) dx \\ &- \int_{\Omega} \min\{\phi^1(x) - \phi^2(x), 0\}C^{21}(x) dx \\ &+ \nu \int_{\Omega} |\nabla \phi^1| dx + \nu \int_{\Omega} |\nabla \phi^2| dx. \end{aligned} \quad (35)$$

such that $\phi^1, \phi^2 \in \mathcal{B}$, where \mathcal{B} is the set of binary functions defined in (6).

Proposition 5 *Let ϕ^1, ϕ^2 be a minimizer of (35), then ϕ^1, ϕ^2 is a minimizer of (8).*

Proof For any $\phi^1, \phi^2 \in \mathcal{B}$, $E(\phi^1, \phi^2) = E^P(\phi^1, \phi^2) + \int_{\Omega} \sigma(x) dx$. Therefore ϕ^1, ϕ^2 is a minimizer of (35) if and only if ϕ^1, ϕ^2 is a minimizer of (8).

The energy functional of (35) is convex if and only if $C^{12}(x), C^{21}(x) \geq 0$ for all $x \in \Omega$, i.e. iff $E(x), F(x) \geq 0$ for all $x \in \Omega$. The weights $C_s^1(x), C_s^2(x), C_t^1(x)$ and $C_t^2(x)$ can be negative without influencing the convexity of (35). As in Section 3, by comparing the sum of row

1-2 and row 3-4 and requiring $E(x), F(x) \geq 0$ we get the condition

$$\begin{aligned} f_2(x) + f_3(x) &= A(x) + B(x) + C(x) + D(x) \\ &\leq A(x) + B(x) + C(x) + D(x) + F(x) + E(x) \\ &= f_1(x) + f_4(x) \end{aligned} \quad (36)$$

for all $x \in \Omega$. By inserting the usual data term (2), we get the requirement

$$|c_2 - u^0(x)|^\beta + |c_3 - u^0(x)|^\beta \leq |c_1 - u^0(x)|^\beta + |c_4 - u^0(x)|^\beta, \quad (37)$$

for all $x \in \Omega$. By assuming the image u^0 contains all gray values I in the interval $[0, L]$ we get the following sufficient condition that (33) has a solution with $E(x), F(x) \geq 0$, and hence that (35) is convex

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta \quad \forall I \in [0, L]. \quad (38)$$

This is exactly the same sufficient condition (32) that made the discrete energy function submodular. Detailed analysis of this condition was given in Section 4. Analogously to the discrete setting [4], we provide here the explicit expression for one of the solutions of (33)

$$\begin{aligned} A(x) &= \max\{f_2(x) - f_4(x), 0\}, \\ C(x) &= \max\{f_4(x) - f_2(x), 0\} \\ B(x) &= \max\{f_4(x) - f_3(x), 0\}, \\ D(x) &= \max\{f_3(x) - f_4(x), 0\} \\ E(x) &= f_1(x) + f_4(x) - f_2(x) - f_3(x), \quad F(x) = 0. \end{aligned}$$

where $f_i(x)$, $i = 1, \dots, 4$ are normally given by (2).

In the following sections, it is shown that the binary non-convex constraints \mathcal{B} can be relaxed and replaced by the convex constraints $\phi^1(x), \phi^2(x) \in [0, 1] \quad \forall x \in \Omega$, such that the overall problem is convex and can be solved globally. In this paper we use the following notation for the relaxed convex constraints

$$\mathcal{B}' = \{\phi : \phi(x) \in [0, 1], \forall x \in \Omega\} \quad (39)$$

We start by deriving a dual formulation of (35) in case $C^{12}(x), C^{21}(x) \geq 0$. The dual model is interpreted as a maximum flow problem over a continuous generalization of the graph in Section 3.2. The dual model eliminates the problem of nondifferentiability of the data term in (35), and allows to build up very efficient algorithms which are presented in Section 8. The dual model is inspired by our previous work [40, 41], where continuous max-flow and min-cut models corresponding to binary (two region) problems were derived.

5.1 Dual model

In the following sections, it is shown that the following maximization problem is a strong dual problem to (35) with either constraints $\phi^1, \phi^2 \in \mathcal{B}$ or $\phi^1, \phi^2 \in \mathcal{B}'$

$$\sup_{p_s, p_t, p^{12}, q} E^D(p_s, p_t, p^{12}, q) = \int_{\Omega} p_s^1(x) + p_s^2(x) dx \quad (40)$$

subject to

$$\begin{aligned} p_s^1(x) &\leq C_s^1(x), \quad p_s^2(x) \leq C_s^2(x), \\ p_t^1(x) &\leq C_t^1(x), \quad p_t^2(x) \leq C_t^2(x), \end{aligned} \quad (41)$$

$$-C^{21}(x) \leq p^{12}(x) \leq C^{12}(x), \quad (42)$$

$$|q^1(x)|_2 \leq \alpha, \quad |q^2(x)|_2 \leq \alpha, \quad \forall x \in \Omega. \quad (43)$$

$$\operatorname{div} q^1(x) - p_s^1(x) + p_t^1(x) + p^{12}(x) = 0, \quad \text{a.e. } x \in \Omega \quad (44)$$

$$\operatorname{div} q^2(x) - p_s^2(x) + p_t^2(x) - p^{12}(x) = 0, \quad \text{a.e. } x \in \Omega. \quad (45)$$

where $p_s^i, p_t^i, p^{12} \in L^1(\Omega)$ and $q^i \in (C^\infty(\Omega))^N$ for $i = 1, 2$. The dual problem (40) can be interpreted as a maximum flow problem over a continuous generalization of the graph proposed in Section 3.2. The dual variables $p_s^i, p_t^i, p^{12}, q^i$ $i = 1, 2$ are interpreted as flow functions on the edges.

5.1.1 Connection to discrete max-flow problem

Recall that for each $p \in \mathcal{P}$ the data term was represented by the 6 edges (25). The flow on each of these edges are constrained by the capacities defined in (29) and (28), i.e.

$$\begin{aligned} P_s^1(p) &\leq c(s, v_{p,1}) \quad \text{flow} \leq \text{capacity on } (s, v_{p,1}) \\ P_s^2(p) &\leq c(s, v_{p,2}) \quad \text{flow} \leq \text{capacity on } (s, v_{p,2}) \\ P_t^1(p) &\leq c(v_{p,1}, t) \quad \text{flow} \leq \text{capacity on } (v_{p,1}, t) \\ P_t^2(p) &\leq c(v_{p,2}, t) \quad \text{flow} \leq \text{capacity on } (v_{p,2}, t) \\ \tilde{P}^{12}(p) &\leq c(v_{p,1}, v_{p,2}) \quad \text{flow} \leq \text{capacity on } (v_{p,1}, v_{p,2}) \\ \tilde{P}^{21}(p) &\leq c(v_{p,2}, v_{p,1}) \quad \text{flow} \leq \text{capacity on } (v_{p,2}, v_{p,1}) \end{aligned}$$

For notational convenience, we use capital letters P, Q to denote flow in discrete edges. The dual variables $p_s^1, p_s^2, p_t^1, p_t^2, \tilde{p}^{12}, \tilde{p}^{21}$ over the continuous domain Ω are inspired by these discrete flow functions. They are constrained by the capacities

$C_s^1(x), C_s^2(x), C_t^1(x), C_t^2(x), C^{12}(x)$ and $C^{21}(x)$ respectively. Note that \tilde{p}^{12} and \tilde{p}^{21} should satisfy

$$0 \leq \tilde{p}^{12}(x) \leq C^{21}(x), \quad 0 \leq \tilde{p}^{21}(x) \leq C^{12}(x)$$

for all $x \in \Omega$, but can be merged in $p^{12} = \tilde{p}^{12} - \tilde{p}^{21}$. The above two constraints then transform into (42).

For each pair of neighboring pixels $(p, q) \in \mathcal{N}$, two edges were constructed in the discrete graph \mathcal{G} : $(v_{p,1}, v_{q,1})$

and $(v_{p,2}, v_{q,2})$. Let Q^1 denote the flow function on $(v_{p,1}, v_{q,1})$ and Q^2 the flow function on $(v_{p,2}, v_{q,2})$. These flows are constrained by

$$0 \leq Q^1(v_{p,1}, v_{q,1}) \leq w_{pq}, \quad 0 \leq Q^2(v_{p,2}, v_{q,2}) \leq w_{pq},$$

for all $(p, q) \in \mathcal{N}$. In the same vein, the two spatial flow fields $q^i \in (C^\infty(\Omega))^N$ $i = 1, 2$ are defined in the continuous setting, and should satisfy (43). Observe that the continuous counterpart allows to measure the magnitude of q^1, q^2 with 2-norm, which in turn produces rotationally invariant results.

The flow conservation constraint (20) should hold at each $v_{p,1}$ and $v_{p,2}$, i.e.

$$\sum_{q \in \mathcal{N}_p^k} Q^1(v_{p,1}, v_{q,1}) - P_s^1(p) + P_t^1(p) - \tilde{P}^{12}(p) + \tilde{P}^{21}(p) = 0,$$

$$\sum_{q \in \mathcal{N}_p^k} Q^2(v_{p,2}, v_{q,2}) - P_s^2(p) + P_t^2(p) + \tilde{P}^{12}(p) - \tilde{P}^{21}(p) = 0,$$

$\forall p \in \mathcal{P}$. The final two constraints (44) and (45) are generalizations of discrete flow conservation to the continuous setting. The total amount of discrete flow in the graph (21) is in this case

$$\sum_{p \in \mathcal{P}} P_s^1(p) + P_s^2(p)$$

The objective functional (40) is a generalization of the above, and measures the total amount of flow that originates from the source.

5.2 Primal-dual model

Introduce the lagrange multipliers, ϕ^1 and ϕ^2 , to the linear equality constraints (44) and (45). The problem (40) can then be reformulated as the primal-dual problem

$$\sup_{p_s, p_t, p^{12}, q} \inf_{\phi^1, \phi^2} \int_{\Omega} p_s^1(x) + p_s^2(x) dx \quad (46)$$

$$+ \int_{\Omega} \phi^1(x) (\operatorname{div} q^1(x) - p_s^1(x) + p_t^1(x) + p^{12}(x)) dx$$

$$+ \int_{\Omega} \phi^2(x) (\operatorname{div} q^2(x) - p_s^2(x) + p_t^2(x) - p^{12}(x)) dx.$$

subject to (41), (42) and (43). By rearranging we get

$$\sup_{p_s, p_t, p^{12}, q} \inf_{\phi^1, \phi^2} \int_{\Omega} (1 - \phi^1(x)) p_s^1(x) + (1 - \phi^2(x)) p_s^2(x) dx \quad (47)$$

$$+ \int_{\Omega} \phi^1(x) p_t^1(x) + \phi^2(x) p_t^2(x) dx + \int_{\Omega} (\phi^1(x) - \phi^2(x)) p^{12}(x) dx$$

$$+ \int_{\Omega} \phi^1(x) \operatorname{div} q^1(x) dx + \int_{\Omega} \phi^2(x) \operatorname{div} q^2(x) dx.$$

subject to (41), (42) and (43).

5.3 Primal model

The inf and sup operators of the primal-dual model (47) can be interchanged, because (47) satisfies all the conditions of the mini-max theorem [17]. It also follows that (47) has at least one saddle point. Observe that all the dual variables $p_s^i, p_t^i, p^{12}, q^i, i = 1, 2$ in (47) are independent, and can be optimized separately. The first 4 terms of (47) can be rewritten for $i = 1, 2$ as

$$\sup_{p_s^i(x) \leq C_s^i(x)} (1 - \phi^i(x)) p_s^i(x) = \begin{cases} (1 - \phi^i(x)) C_s^i(x) & \text{if } \phi^i \leq 1 \\ \infty & \text{if } \phi^i > 1 \end{cases} \quad (48)$$

$$\sup_{p_t^i(x) \leq C_t^i(x)} \phi^i(x) p_t^i(x) = \begin{cases} \phi^i(x) C_t^i(x) & \text{if } \phi^i \geq 0 \\ \infty & \text{if } \phi^i < 0 \end{cases} \quad (49)$$

From (48) and (49) it follows that optimal variables ϕ^1, ϕ^2 must satisfy the constraints

$$\phi^1(x) \in [0, 1], \quad \phi^2(x) \in [0, 1], \quad \forall x \in \Omega, \quad (50)$$

i.e. $\phi^1, \phi^2 \in \mathcal{B}'$ defined in (39). If this was not the case, the primal-dual energy (47) would be infinite, contradicting the existence of at least one saddle point.

The 5th term can also be optimized for p^{12} pointwise as follows

$$\begin{aligned} & \sup_{-C^{21}(x) \leq p^{12}(x) \leq C^{12}(x)} (\phi^1(x) - \phi^2(x)) p^{12}(x) \\ &= \max\{\phi^1(x) - \phi^2(x), 0\} C^{12}(x) \\ & \quad - \min\{\phi^1(x) - \phi^2(x), 0\} C^{21}(x). \end{aligned} \quad (51)$$

The dual representation of total variation (see e.g. [33]) can be used to rewrite the two last terms of (47)

$$\sup_{q \in (C^\infty(\Omega))^N, \text{ s.t. } |q| \leq \alpha} \int_{\Omega} \phi \operatorname{div} q \, dx = \alpha \int_{\Omega} |\nabla \phi| \, dx. \quad (52)$$

Therefore, combining (48), (49), (51) and (52), by maximizing the primal dual model (47) for $p_s^1, p_s^2, p_t^1, p_t^2, p^{12}, q^1, q^2$, we obtain the primal model (35) subject to the convex constraints (50), i.e. $\phi^1, \phi^2 \in \mathcal{B}'$.

5.4 Global solutions by thresholding

We now show there exists binary solutions (i.e. $\phi^1, \phi^2 \in \mathcal{B}$) to the relaxed problem (35) subject to $\phi^1, \phi^2 \in \mathcal{B}'$, which can be obtained by simple thresholding.

Theorem 1 *Assume the data term of (8) satisfies the condition (36), such that $C^{12}(x), C^{21}(x) \geq 0$ for all $x \in \Omega$ in (35). Let ϕ^{1*}, ϕ^{2*} be any solution of (35) subject to $\phi^1(x), \phi^2(x) \in [0, 1] \forall x \in \Omega$. Denote by ϕ_i^ℓ the binary function*

$$\phi_i^\ell(x) = \begin{cases} 1, & \phi^{i*}(x) \geq \ell \\ 0, & \phi^{i*}(x) < \ell \end{cases}. \quad (53)$$

Then for any $\ell \in (0, 1]$, $(\phi_1^\ell, \phi_2^\ell)$ is a solution of (35) and (8) subject to $\phi^1(x), \phi^2(x) \in \{0, 1\} \forall x \in \Omega$.

This theorem shows two important things. First, the convex relaxation (35) subject to (50) is exact, the minimum energy of the relaxed problem always equals the minimum energy of the original problem. Secondly, the thresholding technique allows to convert solutions of the relaxed problem into binary solutions of the original problem, in case of non-uniqueness.

Proof Let $p_s^{i*}, p_t^{i*}, p^{12*}, q^{i*}, \phi^{i*}$ be optimal to the primal-dual model (47). Define the level sets

$$S_i^\ell = \{x : \phi^{i*}(x) \geq \ell\} \quad (54)$$

for $i = 1, 2$. Then, by (48), for any point $x \in \Omega \setminus S_i^\ell$

$$p_s^{i*}(x) = C_s^i(x) \quad (55)$$

for $i = 1, 2$. This, together with the fact that for a.e. $x \in S_1^\ell$

$$p_s^{1*}(x) = \operatorname{div} q^{1*}(x) + p_t^{1*}(x) + p^{12*}(x). \quad (56)$$

and for a.e. $x \in S_2^\ell$

$$p_s^{2*}(x) = \operatorname{div} q^{2*}(x) + p_t^{2*}(x) - p^{12*}(x). \quad (57)$$

implies that the total max-flow can be written

$$\begin{aligned} & \int_{\Omega \setminus S_1^\ell} C_s^1(x) \, dx + \int_{\Omega \setminus S_2^\ell} C_s^2(x) \, dx \\ & + \int_{S_1^\ell} \operatorname{div} q^{1*}(x) + p_t^{1*}(x) + p^{12*}(x) \, dx \\ & + \int_{S_2^\ell} \operatorname{div} q^{2*}(x) + p_t^{2*}(x) - p^{12*}(x) \, dx. \end{aligned}$$

By (49), for any $x \in S_i^\ell$

$$p_t^{i*}(x) = C_t^i(x) \quad (58)$$

for $i = 1, 2$. The terms involving $q^{i*}, i = 1, 2$, are just

$$\int_{S_i^\ell} \operatorname{div} q^{i*}(x) = L_{S_i^\ell} \quad (59)$$

where $L_{S_i^\ell}$ denotes the perimeter of the set S_i^ℓ , see e.g. Prop 4 of [6]. The terms involving p^{12*} can be rewritten as follows

$$\begin{aligned} & \int_{S_1^\ell} p^{12*}(x) \, dx - \int_{S_2^\ell} p^{12*}(x) \, dx = \int_{S_1^\ell \setminus (S_1^\ell \cap S_2^\ell)} p^{12*}(x) \, dx \\ & + \int_{S_1^\ell \cap S_2^\ell} p^{12*}(x) - p^{12*}(x) \, dx - \int_{S_2^\ell \setminus (S_1^\ell \cap S_2^\ell)} p^{12*}(x) \, dx \\ & = \int_{S_1^\ell \setminus (S_1^\ell \cap S_2^\ell)} C^{12}(x) \, dx + \int_{S_2^\ell \setminus (S_1^\ell \cap S_2^\ell)} C^{21}(x) \, dx \end{aligned}$$

where the last equality follows by (51) since p^{12*} is optimal. By combining the above observations, the total max-flow is

$$\begin{aligned} & \int_{\Omega \setminus S_1^\ell} C_s^1(x) dx + \int_{\Omega \setminus S_2^\ell} C_s^2(x) dx \\ & + \int_{S_1^\ell} C_t^1(x) dx + \int_{S_2^\ell} C_t^2(x) dx \\ & + \int_{S_1^\ell \setminus S_1^\ell \cap S_2^\ell} C^{12}(x) dx + \int_{S_2^\ell \setminus S_1^\ell \cap S_2^\ell} C^{21}(x) dx \\ & + \alpha L_{S_1^\ell} + \alpha L_{S_2^\ell}. \end{aligned}$$

By writing this expression in terms of the characteristic functions ϕ_1^ℓ and ϕ_2^ℓ of the sets S_1^ℓ and S_2^ℓ , we get the primal model energy (35) of ϕ_1^ℓ and ϕ_2^ℓ . Hence ϕ_1^ℓ and ϕ_2^ℓ must be optimal to the primal model (35).

6 Nonsubmodular/non-convex data terms

Assume now that the submodular/convexity condition (32) is violated. Although the problem can be represented with negative $E(\cdot), F(\cdot)$, it is no longer computationally tractable. In a discrete setting, this requires to find the minimal cut on a graph which contains negative edges, which is NP-hard. In a variational setting, if $E(x)$ or $F(x)$ are negative for some $x \in \Omega$, the variational problem (35) becomes non-convex, hence the minimization algorithm may get stuck in a local minima. In the next two subsections, we present simpler problems (relaxations), which are submodular in the discrete setting and convex in the continuous setting. Conditions are derived for when the computed solutions of the simpler problems are also solutions to the original problems. It is our observation that these holds for most practical problems. If they should not hold, the relaxations will provide good approximate solutions. We start with the discrete setting in Section 6.1 and then present similar results in a convex variational setting in Section 6.2.

6.1 Submodular relaxation in discrete setting

Minimization of non-submodular energy functions in computer vision has been the subject of previous research, see [26] for a review. One of the most successful approaches is the Quadratic Pseudo Boolean Optimization (QBPO) [20, 26]. In [38] it was shown that removing negative edges, often called truncation, can be effective in minimizing non-submodular functions. We show that such a computationally simple submodular relaxation can produce global solution in practice for our

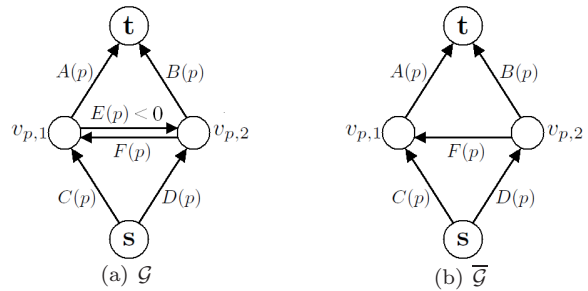


Fig. 4 Illustration of graph $\bar{\mathcal{G}}$ in case $E(p) < 0$.

problem. Furthermore, a criterion is derived for when the minimum cut on the graph with removed edges of negative weight is also a minimum cut on the original graph with negative edge weights. This relaxation is also easily transferable to the continuous setting.

For all $p \in \mathcal{P}$ where (32) is violated, let $A(p), B(p), C(p), D(p), E(p), F(p)$ be a solution to the linear system with $E(p) < 0$ or $F(p) < 0$. Observe that a solution always exists where either $E(p) = 0$ or $F(p) = 0$. The explicit expression for one such solution is given in Section 6.3.

Let $\bar{\mathcal{G}}$ be the graph identically to \mathcal{G} except that all edges of negative weight are removed. That is, for each $p \in \mathcal{P}$, the weights on the data edges in $\bar{\mathcal{G}}$ are constructed as

$$\begin{aligned} c(v_{p,1}, t) &= A(p), \quad c(v_{p,2}, t) = B(p), \\ c(s, v_{p,1}) &= C(p), \quad c(s, v_{p,2}) = D(p), \\ c(v_{p,1}, v_{p,2}) &= \max(E(x), 0), \quad c(v_{p,2}, v_{p,1}) = \max(F(x), 0), \end{aligned} \quad (60)$$

while the regularization edges are given as before by (28). The minimum cut on $\bar{\mathcal{G}}$ can easily be computed by max-flow. As discussed in the previous section, the condition (32) may only be violated if c_1, c_2, c_3 are close to each other compared to c_4 and u_p^0 at $p \in \mathcal{P}$ is close to c_4 . Measured by the data term, the worst assignment of p is to phase 1, which has the cost $|c_1 - u_p^0|^\beta$. By removing the edge with negative weight $E(p) < 0$, the cost of this assignment becomes even higher $|c_1 - u_p^0|^\beta - E(p)$. Alternatively, if c_2, c_3, c_4 are close to each other compared to c_1 and u_p^0 is close to c_1 then $F(p) < 0$. By removing the edge with negative weight, the cost of the worst assignment of u_p^0 becomes higher $|c_4 - u_p^0|^\beta - F(p)$. We therefore expect minimum cuts on $\bar{\mathcal{G}}$ to be almost identical to minimum cuts on \mathcal{G} . Define the sets

$$\mathcal{P}^1 = \{p \in \mathcal{P} \mid E(p) < 0, F(p) \geq 0\},$$

$$\mathcal{P}^2 = \{p \in \mathcal{P} \mid F(p) < 0, E(p) \geq 0\},$$

consisting of all $p \in \mathcal{P}$ for which either $E(p) < 0$ or $F(p) < 0$.

Assume the maximum flow has been computed on $\overline{\mathcal{G}}$, let $R_A(p), R_B(p), R_C(p), R_D(p)$ denote the residual capacities on the edges $(v_{p,1}, t), (v_{p,2}, t), (s, v_{p,1}), (s, v_{p,2})$ respectively. The following theorem gives a criterion for when the minimum cut on $\overline{\mathcal{G}}$ yields the optimal solution of the original problem.

Theorem 2 *Let \mathcal{G} be a graph as defined in (24)-(26) and (28), with weights $A(p), B(p), C(p), D(p), E(p), F(p)$ satisfying (30). Let $\overline{\mathcal{G}}$ be the graph with weights as in \mathcal{G} , with the exception $c(v_{p,1}, v_{p,2}) = 0 \forall p \in \mathcal{P}^1$ and $c(v_{p,2}, v_{p,1}) = 0 \forall p \in \mathcal{P}^2$.*

Assume the maximum flow has been computed on the graph $\overline{\mathcal{G}}$. If

$$\begin{aligned} R_A(p) + R_D(p) &\geq -E(p), \quad \forall p \in \mathcal{P}^1 \\ \text{and } R_B(p) + R_C(p) &\geq -F(p), \quad \forall p \in \mathcal{P}^2, \end{aligned} \quad (61)$$

then $\text{min-cut}(\mathcal{G}) = \text{min-cut}(\overline{\mathcal{G}})$.

Proof We will create a graph $\underline{\mathcal{G}}$ of only positive edge weights, such that the minimum cut problem on $\underline{\mathcal{G}}$ is a relaxation of the minimum cut problem on \mathcal{G} . The graph $\underline{\mathcal{G}}$ is constructed with weights as in $\overline{\mathcal{G}}$ with the following exceptions

$$\begin{aligned} c(v_{p,1}, t) &= A(p) - R_A(p), \quad \forall p \in \mathcal{P}^1, \\ c(s, v_{p,2}) &= D(p) - R_D(p), \quad \forall p \in \mathcal{P}^1 \\ c(v_{p,2}, t) &= B(p) - R_B(p), \quad \forall p \in \mathcal{P}^2, \\ c(s, v_{p,1}) &= C(p) - R_C(p), \quad \forall p \in \mathcal{P}^2. \end{aligned}$$

We first show $\text{min-cut}(\underline{\mathcal{G}}) \leq \text{min-cut}(\mathcal{G}) \leq \text{min-cut}(\overline{\mathcal{G}})$. The right inequality follows because all the edges in the graph $\overline{\mathcal{G}}$ have greater or equal weight than the edges in the graph \mathcal{G} . To prove the left inequality, observe that only data edges for $p \in \mathcal{P}_1 \cup \mathcal{P}_2$ differ between $\underline{\mathcal{G}}$ and \mathcal{G} . For each $p \in \mathcal{P}_1$ there are 4 possibilities for the cut (V_s, V_t) . Since $R_A(p), R_B(p), R_C(p), R_D(p) \geq 0$, the cost of all the 3 cuts $v_{p,1}, v_{p,2} \in V_s, v_{p,1}, v_{p,2} \in V_t$ and $v_{p,1} \in V_t, v_{p,2} \in V_s$ are lower in $\underline{\mathcal{G}}$ than in \mathcal{G} . The last cut $v_{p,1} \in V_s, v_{p,2} \in V_t$ has the cost $A(p) + B(p) - E(p)$ in the \mathcal{G} and the cost $A(p) + D(p) - (R_A(p) + R_D(p)) \leq A(p) + D(p) + E(p)$ in the graph $\underline{\mathcal{G}}$. The same argument shows that all possible cuts have a lower or equal cost in $\underline{\mathcal{G}}$ than in \mathcal{G} for $p \in \mathcal{P}_2$.

Both $\underline{\mathcal{G}}$ and $\overline{\mathcal{G}}$ have only positive edge weights. Since all the edges have greater or equal weight in $\overline{\mathcal{G}}$ than in $\underline{\mathcal{G}}$ it follows that

$$\text{max-flow}(\underline{\mathcal{G}}) \leq \text{max-flow}(\overline{\mathcal{G}}).$$

Hence, since the max flow on $\overline{\mathcal{G}}$ is feasible on $\underline{\mathcal{G}}$ it is also optimal on $\underline{\mathcal{G}}$. Therefore, by duality $\text{min-cut}(\underline{\mathcal{G}}) = \text{min-cut}(\overline{\mathcal{G}})$ which implies $\text{min-cut}(\mathcal{G}) = \text{min-cut}(\overline{\mathcal{G}})$.

Therefore, by computing the max flow on $\overline{\mathcal{G}}$ and examining the residual capacities for criterion (61), it can be checked whether the solution is optimal on \mathcal{G} . It is also possible to apply non-submodular minimization algorithms such as

6.2 Convex relaxation in continuous variational setting

For all points $x \in \Omega$ where (32) is violated, let $A(x), B(x), C(x), D(x), E(x), F(x)$ be a solution to the linear system with $E(x) < 0$ or $F(x) < 0$. See Section 6.3 for the explicit expression of one such solution. Let Ω^1 be the set of points where $E(x) < 0$ and Ω^2 the set of points where $F(x) < 0$, i.e.

$$\Omega^1 = \{x \in \Omega : E(x) < 0\}, \quad \Omega^2 = \{x \in \Omega : F(x) < 0\}. \quad (62)$$

Let P denote the original primal problem (35) with weights set to

$$\begin{aligned} C_s^1(x) &= C(x), \quad C_s^2(x) = D(x), \\ C_t^1(x) &= A(x), \quad C_t^2(x) = B(x), \\ C^{12}(x) &= E(x), \quad C^{21}(x) = F(x), \quad \forall x \in \Omega \end{aligned} \quad (63)$$

as before. Since $C^{12}(x)$ or $C^{21}(x)$ are assumed negative for some $x \in \Omega$, the minimization problem (35) is non-convex. A dual formulation of this problem cannot be derived as in Section 5.1 because the arguments in this section are not valid if $C^{12}(x)$ or $C^{21}(x)$ are negative. In the same manner, max-flow does not make sense in the discrete setting when edge capacities are negative.

A problem is now defined where all the negative terms are removed. Primal-dual and dual formulations of this problem can be derived as in Section 5.1. We denote the primal problem as \overline{P} , primal-dual problem as \overline{PD} and dual problem as \overline{D} and define them as the optimization of (35), (47) and (40) respectively, with weights constructed as

$$\begin{aligned} C_s^1(x) &= C(x), \quad C_s^2(x) = D(x), \\ C_t^1(x) &= A(x), \quad C_t^2(x) = B(x), \\ C^{12}(x) &= \max(E(x), 0), \quad C^{21}(x) = \max(F(x), 0), \quad \forall x \in \Omega \end{aligned} \quad (64)$$

Since all the weights are non-negative, the problem \overline{P} is convex and can be minimized globally. We are interested to know when a solution to the convex problem \overline{P} is also optimal to the problem P . This can be answered by investigating the solution of the dual problem \overline{D} .

Theorem 3 *Let $\overline{\phi}^i; \overline{p}_s^i, \overline{p}_t^i, \overline{p}^{12}, \overline{q}^i; i = 1, 2$ be a solution of \overline{PD} , i.e. $\overline{\phi}^1, \overline{\phi}^2$ is a solution of \overline{P} and $\overline{p}_s^i, \overline{p}_t^i, \overline{p}^{12}, \overline{q}^i; i = 1, 2$ a solution of \overline{D} . Define the residual capacities as*

$$R_s^1(x) = C_s^1(x) - \overline{p}_s^1(x), \quad R_s^2(x) = C_s^2(x) - \overline{p}_s^2(x), \quad (65)$$

$$R_t^1(x) = C_t^1(x) - \bar{p}_t^1(x), \quad R_t^2(x) = C_t^2(x) - \bar{p}_t^2(x).$$

If

$$R_t^1(x) + R_s^2(x) \geq -E(x), \quad \forall x \in \Omega^1 \quad (66)$$

and

$$R_t^2(x) + R_s^1(x) \geq -F(x), \quad \forall x \in \Omega^2, \quad (67)$$

then $(\bar{\phi}^1, \bar{\phi}^2)$ is optimal to the original problem P .

Proof Let E^P denote the energy functional of the primal problem (35) with original weights (63), and let \bar{E}^P denote the energy functional of (35) with weights (64). Then for any functions ϕ^1, ϕ^2 such that $\phi^1(x), \phi^2(x) \in [0, 1]$ for all $x \in \Omega$

$$\bar{E}^P(\phi^1, \phi^2) \geq E^P(\phi^1, \phi^2). \quad (68)$$

Define a new problem \underline{P} as the minimization of (35) with weights (64) for all $x \in \Omega \setminus (\Omega^1 \cup \Omega^2)$ and

$$\begin{aligned} C_s^1(x) &= C(x), & C_s^2(x) &= \bar{p}_s^2(x), \\ C_t^1(x) &= \bar{p}_t^1(x), & C_t^2(x) &= B(x), \quad \forall x \in \Omega^1, \end{aligned} \quad (69)$$

$$\begin{aligned} C_s^1(x) &= \bar{p}_s^1(x), & C_s^2(x) &= D(x), \\ C_t^1(x) &= A(x), & C_t^2(x) &= \bar{p}_t^2(x), \quad \forall x \in \Omega^2, \end{aligned} \quad (70)$$

$$C^{12}(x) = 0, \quad C^{21}(x) = 0, \quad \forall x \in \Omega^1 \cap \Omega^2. \quad (71)$$

Let \underline{E}^P denote the energy functional of (35) with the above defined weights. We will show that

$$\underline{E}^P(\phi^1, \phi^2) \leq E^P(\phi^1, \phi^2) \leq \bar{E}^P(\phi^1, \phi^2), \quad \forall \phi^1, \phi^2 \in \mathcal{B}'. \quad (72)$$

The right inequality is just a repetition of (68). To show the left inequality, observe that for each $x \in \Omega^1$

$$\begin{aligned} \underline{E}^P(\phi^1(x), \phi^2(x)) &= (1 - \phi^1(x))C(x) + (1 - \phi^2(x))\bar{p}_s^2(x) \\ &+ \phi^1(x)\bar{p}_t^1(x) + \phi^2(x)B(x) \\ &= (1 - \phi^1(x))C(x) + (1 - \phi^2(x))(D(x) - R_s^2(x)) \\ &+ \phi^1(x)(A(x) - R_t^1(x)) + \phi^2(x)B(x) \\ &= (1 - \phi^1(x))C(x) + (1 - \phi^2(x))D(x) + \phi^1(x)A(x) \\ &+ \phi^2(x)B(x) + \phi^1(x)(-R_t^1(x)) + (1 - \phi^2(x))(-R_s^2(x)) \end{aligned}$$

The last two terms can be bounded by

$$\begin{aligned} &\phi^1(x)(-R_t^1(x)) + (1 - \phi^2(x))(-R_s^2(x)) \\ &\leq \phi^1(x)(1 - \phi^2(x))(-R_t^1(x)) + \phi^1(x)(1 - \phi^2(x))(-R_s^2(x)) \\ &= \phi^1(x)(1 - \phi^2(x))(-R_t^1(x) - R_s^2(x)) \\ &\leq \max(\phi^1(x) - \phi^2(x), 0)(-R_t^1(x) - R_s^2(x)) \\ &\leq \max(\phi^1(x) - \phi^2(x), 0)E(x) \end{aligned}$$

Therefore

$$\underline{E}^P(\phi^1(x), \phi^2(x)) \leq E^P(\phi^1(x), \phi^2(x)). \quad (73)$$

The same argument shows that (73) also holds for all $x \in \Omega^2$. For $x \in \Omega \setminus (\Omega^1 \cup \Omega^2)$, $\underline{E}^P(\phi^1(x), \phi^2(x)) = E^P(\phi^1(x), \phi^2(x))$. Observe that since the maximum flow $\bar{p}_t^i, \bar{p}_s^{12}, \bar{q}^i; i = 1, 2$ in problem \bar{D} is feasible in \underline{D} , it is by (72) also optimal on \underline{D} . It follows that $\underline{E}^P(\bar{\phi}^1, \bar{\phi}^2) = \bar{E}^P(\bar{\phi}^1, \bar{\phi}^2)$, which by (72) implies $E^P(\bar{\phi}^1, \bar{\phi}^2) = \bar{E}^P(\bar{\phi}^1, \bar{\phi}^2)$.

6.3 Explicit expression for one solution

The explicit expression for one of the solutions of the linear system (33) and (30) in case (32) is violated is presented. We stick to the discrete notation where p denotes spatial coordinates, the continuous analogue where x denotes spatial coordinates is obvious (just replace u_p with $u(x)$). Consider two cases: first, if $u_p^0 > c_3$, then

$$E(p) = f_1(p) + f_4(p) - f_2(p) - f_3(p), \quad F(p) = 0 \quad (74)$$

$$A(p) = \max\{f_2(p) - f_4(p), 0\} - E(p), \quad (75)$$

$$C(p) = \max\{f_4(p) - f_2(p), 0\} - E(p), \quad (76)$$

$$B(p) = \max\{f_4(p) - f_3(p), 0\} - E(p), \quad (77)$$

$$D(p) = \max\{f_3(p) - f_4(p), 0\} - E(p), \quad (78)$$

in which case $E(p) < 0$. If $u_p^0 < c_2$, then

$$F(p) = f_1(p) + f_4(p) - f_2(p) - f_3(p), \quad E(p) = 0 \quad (79)$$

$$A(p) = \max\{f_1(p) - f_3(p), 0\} - F(p), \quad (80)$$

$$C(p) = \max\{f_3(p) - f_1(p), 0\} - F(p), \quad (81)$$

$$B(p) = \max\{f_2(p) - f_1(p), 0\} - F(p), \quad (82)$$

$$D(p) = \max\{f_1(p) - f_2(p), 0\} - F(p), \quad (83)$$

in which case $F(p) < 0$. By Prop 1, the condition holds whenever $u_p^0 \in [c_2, c_3]$.

7 A tight convex relaxation of the Potts model

In this section we show that the convex reformulation (35) can be used as the basis of a new convex relaxation of Potts model (1) with four regions. The convex relaxation cannot in general result in global solutions of (1), but is at least as tight as the tightest existing relaxation [35]. It is significantly simpler than [35], since the number of side constraints and variables are much lower. As a direct consequence, it is easier to handle computationally. The purpose of this paper is to propose the relaxation, we leave detailed analysis to a future work. The new relaxation is possibly also the strictly tightest

that exist, but a proof of this will be quite involved and is left as an open problem.

In [35] it was observed that multiple countings of boundaries in the model (17) with linearly ordered labels, could be suppressed by introducing additional constraints on the dual variables of the total variation term of (17). The number of constraints grow quadratically in the number of regions. In particular, 3 dual variables and 7 dual constraints are necessary to represent 4 regions. The resulting convex relaxation is not guaranteed produce a global binary solution, but as shown in [35], such an approach may produce good approximations after thresholding. Furthermore, as shown in [35], Prop 4, the relaxation is strictly tighter than other recently proposed approaches [42, 29, 6].

In section 5 it was shown that the model (35) could be optimized globally by relaxing the binary constraints of ϕ^1 and ϕ^2 . As discussed in Section 2.2.2 this model approximates Pott's model much more closely than the model (17). Two of the boundaries are measured twice in (35), while all the remaining 4 boundaries are measured once. In contrast, the model (17) overcounts the boundaries much more severely, see e.g. Figure 1. In consequence, it suffices to introduce one additional constraint on the dual variables q^1 and q^2 in the model (35) in order to measure each boundary exactly once.

Consider first the convex model (35), expressed in terms of the dual formulation of total variation (52). We then obtain

$$\begin{aligned} \min_{\phi^1, \phi^2} \sup_{q^1, q^2} E^P(\phi^1, \phi^2) &= \int_{\Omega} (1 - \phi^1(x)) C_s^1(x) \\ &+ (1 - \phi^2(x)) C_s^2(x) + \phi^1(x) C_t^1(x) + \phi^2(x) C_t^2(x) dx \\ &+ \int_{\Omega} \max\{\phi^1(x) - \phi^2(x), 0\} C^{12}(x) dx \\ &- \int_{\Omega} \min\{\phi^1(x) - \phi^2(x), 0\} C^{21}(x) dx \\ &+ \nu \int_{\Omega} \phi^1 \operatorname{div} q^1 dx + \nu \int_{\Omega} \phi^2 \operatorname{div} q^2 dx. \end{aligned} \quad (84)$$

subject to

$$\phi^1(x) \in \{0, 1\} \quad \phi^2(x) \in \{0, 1\}, \quad \forall x \in \Omega, \quad (85)$$

$$|q^1(x)| \leq 1, \quad |q^2(x)| \leq 1, \quad \forall x \in \Omega. \quad (86)$$

Let x^0 be a point on the boundary between region Ω_3 and Ω_2 . Then $\phi_-^1(x^0) = \phi_-^2(x^0) = 0$ and $\phi_+^1(x^0) = \phi_+^2(x^0) = 1$, therefore both the discontinuity of ϕ^1 and ϕ^2 contributes to the energy of (84). In effect, such a transition has twice the cost of other transitions. In order to reduce the cost of this transition, the following constraint is added to the dual variables

$$|q^1(x) + q^2(x)| \leq 1, \quad \forall x \in \Omega. \quad (87)$$

Table 2 Number of variables and constraints

	Primal variables	Primal constraints	Dual variables	Dual constraints
Pock et. al.	3	4	3	6
Proposed	2	2	2	3

Since ϕ^1 and ϕ^2 are equal in a small neighborhood $B(x^0)$ around x^0 it follows that

$$\begin{aligned} \nu \int_{B(x^0)} \phi^1 \operatorname{div} q^1 dx + \nu \int_{B(x^0)} \phi^2 \operatorname{div} q^2 dx \\ = \nu \int_{B(x^0)} \phi^1 \operatorname{div}(q^1 + q^2) dx. \end{aligned}$$

Hence, taking the supremum w.r.t. q^1, q^2 over (86) and (87), we obtain $\nu \int_{B(x^0)} |\nabla \phi^1| dx$ as desired. For x^0 at the boundary between Ω_1 and Ω_4 , $\phi_-^1(x^0) = 1, \phi_-^2(x^0) = 0$ and $\phi_+^1(x^0) = 0, \phi_+^2(x^0) = 1$. Therefore, this transition also costs twice as much. Exactly the same argument shows that the additional constraint (87) also suppresses overcounting of this boundary.

A convex relaxation is now formulated where the non-convex binary constraints (85) are replaced by

$$\phi^1(x) \in [0, 1] \quad \phi^2(x) \in [0, 1], \quad \forall x \in \Omega, \quad (88)$$

In contrast to the model (35), the thresholding Theorem 1 is not generally valid if the additional constraint (87) is introduced. However, if the computed solution ϕ^1, ϕ^2 is binary everywhere, it is also a global solution to the nonconvex Potts model (1). Otherwise, thresholding of ϕ^1, ϕ^2 will result in good approximate solutions. This relaxation is at least as tight as [35]. It is also possibly strictly tighter, but a proof will be quite involved and is postponed to a future work. The special case of a minimal cone in 3D, where the boundaries of 4 regions intersect, is one example we believe can be used to show that our relaxation has a strictly higher minimal energy than [35]. The number of primal and dual variables and constraints are summarized in Table 2. Our approach involves significantly less variables and constraints than [35] and is consequently much easier to handle computationally.

If the data term does not satisfy the condition (37), the relaxation from section 6.2 and theorem 3 can straightforwardly be generalized to the convex relaxation of Pott's model in Section 7, in case the data term is not submodular: Let $\bar{\phi}^i; \bar{p}_t^i, \bar{p}^{12}, \bar{q}^i, i = 1, 2$ be a solution to $\overline{\text{PD}}$ with constraints (86), (87) and (85) and weights set to (64). If $R_s^1(x), R_s^2(x), R_t^1(x), R_t^2(x)$ defined in (65) satisfies (66) and (67), then $\bar{\phi}^i; \bar{p}_t^i, \bar{p}^{12}, \bar{q}^i, i = 1, 2$ is also optimal to the same problem with original weights (63). The proof is identical to the proof of Theorem 3.

Algorithm 1 Multiplier-Based Algorithm

Choose some initialization for p_s^i, p_t^i, q^i and ϕ^i , let $k = 1$ and start k -th iteration, which includes the following steps, until convergence:

- Optimize $p_s^i, i = 1, 2$ by fixing other variables

$$\begin{aligned} p_s^{i,k+1} &:= \arg \max_{p_s^i(x) \leq C_s^i(x)} L_c(p_s^i, p_t^i, q^i, \phi^k) \\ &:= \arg \max_{p_s^i(x) \leq C_s^i(x)} \int_{\Omega} p_s^i dx \\ &\quad - \frac{c}{2} \left\| p_s^i - (p_t^i + (-1)^{i+1} p^{12k} + \operatorname{div} q^i) + \phi^i / c \right\|^2. \end{aligned}$$

The optimal p_s^i can be easily computed pointwise at each $x \in \Omega$;

- Optimize p^{12} by fixing other variables

$$\begin{aligned} p^{12k+1} &:= \arg \max_{-C^{12}(x) \leq p^{21}(x) \leq C^{12}(x)} L_c(p_s^{i,k+1}, p_t^i, q^i, \phi^k) \\ &:= \arg \max_{-C^{12}(x) \leq p^{21}(x) \leq C^{12}(x)} \\ &\quad - \sum_{i=1,2} \frac{c}{2} \left\| p_s^{i,k+1} + (p_t^i + (-1)^{i+1} p^{12} + \operatorname{div} q^i) + \phi^i / c \right\|^2. \end{aligned}$$

- Optimize $q^i, i = 1, 2$, by fixing other variables

$$\begin{aligned} q^{i,k+1} &:= \arg \max_{q \in K_i} L_c(p_s^{i,k+1}, p_t^i, q, \phi^k). \\ &:= \arg \max_{q \in K_i} - \frac{c}{2} \left\| \operatorname{div} q + p_s^{i,k+1} - p_t^i + (-1)^{i+1} p^{12k+1} - \phi^i \right\|^2, \end{aligned} \quad (89)$$

for $i = 1, 2$. The above problem can either be solved iteratively by Chambolle's projection algorithm [14], or approximately in one step through

$$q_i^{k+1} = \Pi_{K_i} \left(q_i^k + c \nabla (\operatorname{div} q_i^k + p_s^{i,k+1} - p_t^i + (-1)^{i+1} p^{12k+1} - \phi^i) \right), \quad (90)$$

where Π_{K_i} is the projection operator onto K_i .

- Optimize $p_t^i, i = 1, 2$ by fixing other variables

$$\begin{aligned} p_t^{i,k+1} &:= \arg \max_{p_t^i(x) \leq C_t^i(x)} L_c(p_s^{i,k+1}, p_t^i, p^{12k+1}, q^i, \phi^k) \\ &:= \arg \max_{p_t^i(x) \leq C_t^i(x)} \\ &\quad - \frac{c}{2} \left\| p_t^i + (p_s^{i,k+1} + (-1)^{i+1} p^{12k+1} + \operatorname{div} q^{i,k+1}) + \phi^i / c \right\|^2. \end{aligned}$$

- Update multipliers $\phi^i, i = 1, 2$, by

$$\phi^{i,k+1} = \phi^i - c (\operatorname{div} q^{i,k+1} - (-p_s^{i,k+1} + p_t^i + (-1)^{i+1} p^{12k+1}));$$

- Let $k = k + 1$ go to the $k + 1$ -th iteration until converge.
-

8 Algorithms

Algorithms for the convex formulations and relaxations in Section 5-7 are presented based on the augmented la-

grangian method. In [3, 40, 41] the augmented lagrangian method was applied on continuous max-flow formulations of minimization problems with binary and linearly ordered labels. The algorithms were shown to be very efficient and outperform alternative approaches. We derive similar algorithms based on the max-flow formulations of the problems (35) and (84). Observe that the lagrange multipliers ϕ^1, ϕ^2 are unconstrained in (46). However, by construction optimal ϕ^1, ϕ^2 will satisfy the relaxed binary constraints (50). In this section it is assumed that Ω , the unknowns $p_s^i, p_t^i, p^{12}, q^i, \phi^i; i = 1, 2$ and the differential and integration operators are discretized, but we stick with the continuous notation for simplicity. The augmented lagrangian functional can be formulated as

$$\begin{aligned} L(p_s, p_t, p^{12}, q, \phi) &= \int_{\Omega} p_s^1(x) + p_s^2(x) dx \\ &\quad + \sum_{i=1}^2 \int_{\Omega} \phi^i(x) (\operatorname{div} q^i(x) - p_s^i(x) + p_t^i(x) + (-1)^{i+1} p^{12}(x)) dx \\ &\quad - \frac{c}{2} \sum_{i=1}^2 \left\| \operatorname{div} q^i(x) - p_s^i(x) + p_t^i(x) + (-1)^{i+1} p^{12}(x) \right\|^2 \end{aligned} \quad (91)$$

An algorithm for optimizing (46) is constructed based on the alternating direction method of multipliers [21], see Alg. 7. A similar algorithm for two region problems [41] was shown to converge very quickly, and outperform earlier approaches. The sets K_1 and K_2 in (89) are the unit disks

$$K_1 = K_2 = \{q : \Omega \mapsto \mathbb{R}^N : \|q\|_{\infty} \leq \nu, q_n|_{\partial\Omega} = 0\}.$$

Here $\|q\|_{\infty} = \max_{x \in \Omega} |q(x)|_2$.

8.1 Algorithm for Convex Relaxed Pott's model

The algorithm for the convex relaxed Pott's model does not change, except the variable q^1 in step (89) is optimized over the set

$$\begin{aligned} q^1 \in K_1 &= \{q : \Omega \mapsto \mathbb{R}^N : \\ &\quad \|q\|_{\infty} \leq \nu, \|q + q^{2k}\|_{\infty} \leq \nu, q_n|_{\partial\Omega} = 0\} \end{aligned} \quad (92)$$

and q^2 is optimized over the set

$$\begin{aligned} q^2 \in K_2 &= \{q : \Omega \mapsto \mathbb{R}^N : \\ &\quad \|q\|_{\infty} \leq \nu, \|q^{1k+1} + q\|_{\infty} \leq \nu, q_n|_{\partial\Omega} = 0\}. \end{aligned} \quad (93)$$

The sets (92) and (93) consist of an intersection of two spheres. The euclidian projection onto such an intersection can be computed analytically, the details were

given in appendix A of [3]. Therefore, the simple gradient projection step (90) generalizes directly to this setting. Alternatively, during each iteration, q^1 and q^2 can be simultaneously projected onto the set (87), (86) by Dykstra's iterative algorithm.

9 Experiments

Numerical examples are shown in Figure 5-11, where the power $\beta = 2$ has been used in the data term. In order to estimate the optimal constant values $\{c_i\}_{i=1}^4$, an alternating minimization algorithm is applied as follows:

Find initialization $\{c_i^0\}_{i=1}^4$ and solve for $k = 1, \dots$ until convergence

1. $\{\phi_i^k\}_{i=1}^4 = \arg \min_{\{\phi_i\}_{i=1}^4} E(\{\phi_i\}_{i=1}^4, \{c_i^{k-1}\}_{i=1}^4)$,
2. $\{c_i^k\}_{i=1}^4 = \arg \min_{\{c_i\}_{i=1}^4} E(\{\phi_i^k\}_{i=1}^4, \{c_i\}_{i=1}^4)$.

Step 1. is solved by the algorithms developed in this paper. The optimization problem in step 2 is simple and has a closed form solution: c_i is the mean intensity value within region Ω_i^k when $\beta = 2$ and median intensity when $\beta = 1$. Convergence means that the partition does not change from one iteration to the next, and will usually occur in around 10 iterations. The constant values can be initialized efficiently by the isodata algorithm [22]. During each iteration of the above algorithm, the energy minimization problem was submodular.

On the relatively simple image in Figure 5, the level set method finds a good local minima. If the initialization is bad, the level set method gets stuck in an inferior local minima also for this simple image as shown in Figure 6. White points indicate the zero level set of ϕ^1 and dark points indicate the zero level set of ϕ^2 .

More difficult images are presented in Figure 7 - 11. The L^2 data fidelity term has been used ($\beta = 2$) and the different methods are compared by keeping the same optimal constant values $\{c_i^*\}_{i=1}^4$ and regularization parameter ν fixed, while minimizing in terms of the regions. One can clearly see the advantages of the global approach over earlier approaches.

In subfigure 8, 9 and 10 (b) the computed global minimum of the discrete energy is shown. In subfigures 8, 9 and 10 (c) the computed global minimum of the convex reformulation of the Chan-Vese model are shown, which takes values in $[0, 1]$, but is binary at most points. The binary result after thresholding ϕ^{1*}, ϕ^{2*} at the level $\ell = \frac{1}{2}$ are shown in subfigures (d), which is a global minimum of the original problem according to Theorem 1. Observe that the continuous version is

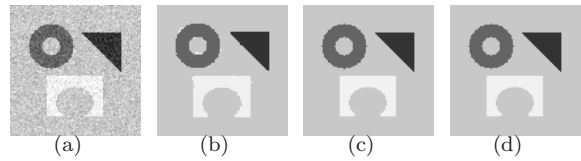


Fig. 5 L^2 data fidelity: (a) input, (b) level set method gradient descent, (c) alpha expansion/alpha beta swap, (d) our approach.

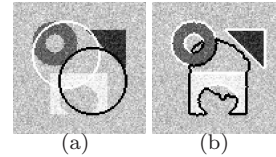


Fig. 6 Level set method: (a) bad initialization, (b) result.

rotationally invariant and, in contrast to the discrete approach, produces results that are not biased by the discrete grid.

9.1 Experiments on L_2 data fitting term: submodularity

In Section 4 we gave theoretical insights on how submodularity of the energy function was related to the distribution of the values c_i , $i = 1, \dots, 4$. It was shown that the condition becomes less strict as β increases. In this section we demonstrate that for L_2 data fitting term ($\beta = 2$ in (2)), the energy function is submodular in practice. The L_2 norm tolerates rather uneven distributions of c_i , $i = 1, \dots, 4$. In addition, the parameters c_i $i = 1, \dots, 4$ that minimize the energy function will not get very clustered. To verify this, we have run the alternating optimization algorithm described at the beginning of this section for optimizing the parameters c_i , $i = 1, \dots, 4$ on all images from the data base [1]. In all experiments, the submodularity condition was satisfied during each iteration of the algorithm.

9.2 Non-submodular data terms

The purpose of this section is to demonstrate the relaxation approaches from Section 6 for minimization the energy in case the data term is not submodular/convex. For that reason, we have used L_1 data term and fixed the constant values c in such a way that the submodularity condition is violated.

One such example is shown in Figure 12, which is a modified version of the example in Figure 5, where the average intensities values of 3 of the objects are close

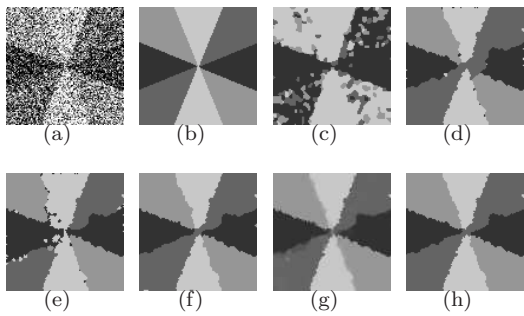


Fig. 7 (a) Input image, (b) ground truth, (c) level set method gradient descent, (d) alpha expansion, (e) alpha-beta swap, (f) global minimum computed by new graph cut approach in discrete setting, (g) New convex optimization approach in continuous setting before threshold, (h) convex minimization approach after threshold (global optimum).

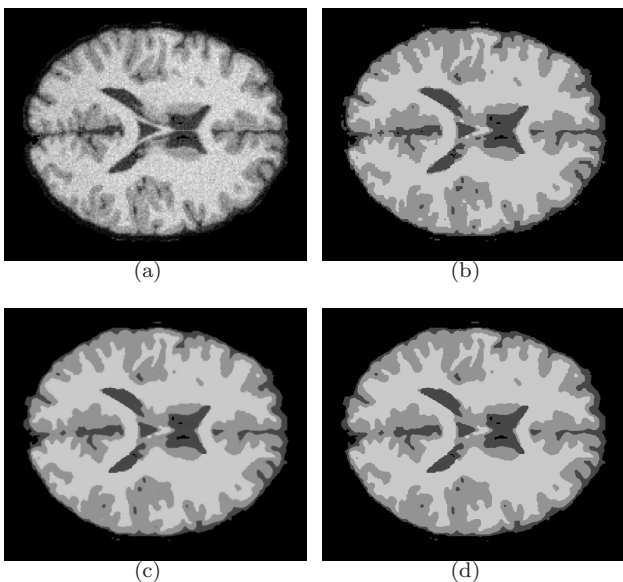


Fig. 8 L_2 data fidelity: (a) Input, (b) global minimum discrete Chan-Vese model 4 neighbors, (c) convex formulation before threshold, (d) convex formulation after threshold (global minimum).

compared to the 4th object. Some more natural examples are shown in Figure 13. Subfigures (b) show the set of pixels $p \in \mathcal{P}_1 \cup \mathcal{P}_2$, where the submodular condition was violated. In all experiments $\beta = 1$. Subfigures (c) show the set of pixels where the residual capacity conditions (61) ((66) and (67) in continuous setting) were violated, which is the empty set in all cases. Therefore, the solutions obtained by the cut on the graphs $\overline{\mathcal{G}}$ and the solutions obtained from the convex relaxations, are also global solutions to the original problems. These examples are typical: if the regularization parameter ν is not set extremely high, the residual capacity condition tends to be satisfied. To save space, only a subset of

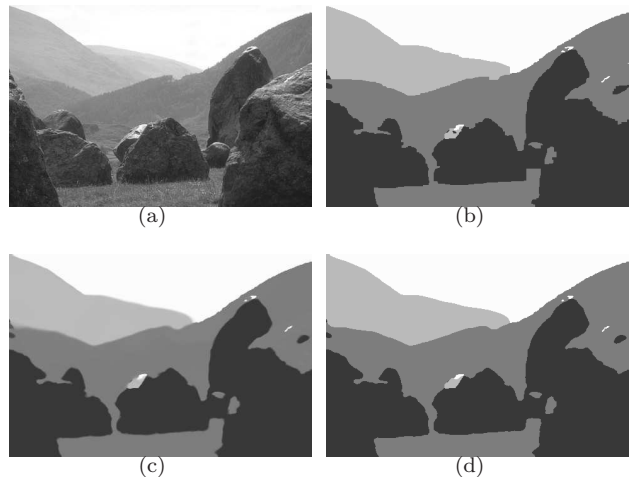


Fig. 9 Segmentation with L^2 data term: (a) Input, (b) graph cut 4 neighbors (c) convex formulation before threshold, (d) convex formulation after threshold (global minimum).

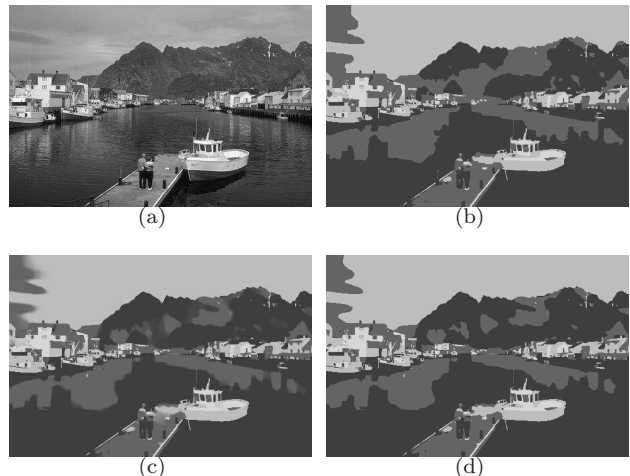


Fig. 10 Segmentation with L^2 data term: (a) Input, (b) result graph cut 8 neighbors in discrete setting (c) result convex formulation before threshold, (d) result convex formulation after threshold (global optimum).

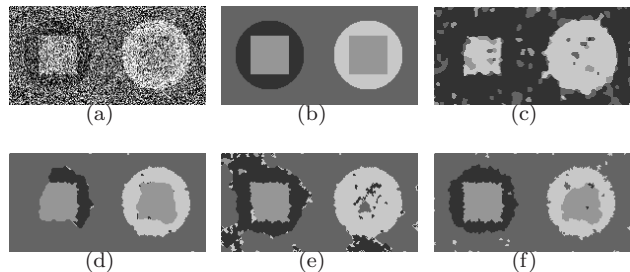


Fig. 11 (a) Input image, (b) ground truth, (c) gradient descent, (d) alpha expansion, (e) alpha-beta swap, (f) our approach.

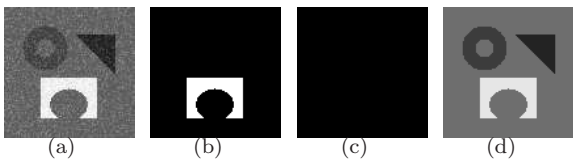


Fig. 12 L^1 data fidelity. Note that the constant values of c_1, c_2, c_3 are close to each other compared to c_4 . (a) Input image, (b) set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was violated, (c) set of pixels where the residual capacity criterion (61) is not satisfied (empty set), (d) output image.

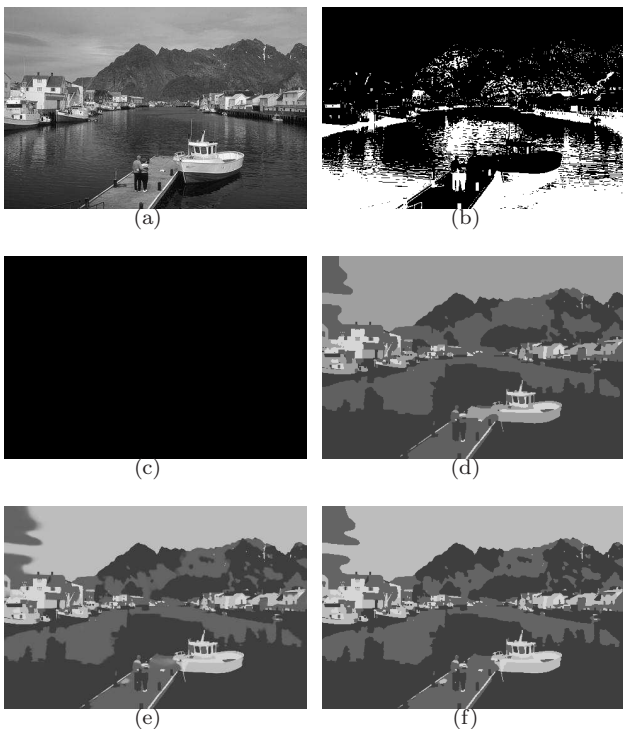


Fig. 13 Segmentation with L^1 data term: (a) Input, (b) set of points where submodular condition (32) was violated, (c) set of points where the residual capacity conditions (66) or (67) were not satisfied (empty set), (d) result before threshold, (e) result after threshold (global optimum).

our experiments are shown. If ν is set extremely high, the residual capacity condition may be violated at a small set of pixels. In this case, it cannot be concluded whether the solution of the relaxed problems are also solutions of the original problems, but will in any case be good approximations.

Table 3 Number of iterations and number of flops per iteration to reach energy precision 10^{-3} on Figure 8 (a).

Pock et al.		Proposed	
iterations	flops pr. it.	iterations	flops pr. it.
290	$4.7 * 10^8$	170	$4.1 * 10^7$

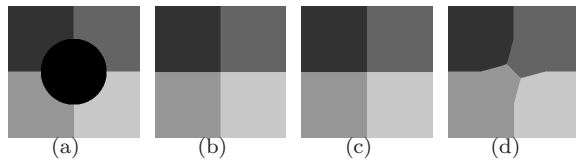


Fig. 14 (a) Input, all data terms are set to zero in the dark region, (b) result of relaxation [42,28] and [6], (c) result of Chan-Vese model (8) (d) result new relaxation from Section 7.

9.3 Convex relaxed Pott's model

The convex relaxation of Pott's model from Section 7 is demonstrated in this section. As seen in Figure 14, the relaxation is tight enough to reconstruct the global solution on the quadruple junction example, Fig. 14 (due to slight asymmetry the solution is unique), whereas the relaxations [42,28] and [6] fail. In [13,35] it was shown that [35] could also produce a nearly binary solution on a similar example. The results after thresholding using the schemes suggested in the different approaches are shown.

In [35] a primal-dual gradient projection algorithm was proposed for the relaxation based on (17). The algorithm needs to project the 3 dual variables q^1, q^2, q^3 onto a convex set described by 6 inequalities each iteration. Since there is no closed form solution for computing such a projection, an iterative algorithm must be applied (Dijkstra's algorithm was suggested). This slows down the convergence. Our relaxation involves only 2 dual variables q^1, q^2 and 3 inequalities (87) and (86), which is simpler to handle computationally. As seen in Table 3, our algorithm converges in less number of iterations than [35] for the image in Figure 8 (a). More importantly, the computational cost per iteration is much lower, as seen by the number of flops per iteration.

10 Conclusions

We have presented an exact global optimization framework for image segmentation models with four regions, both in a discrete setting and a variational setting, which includes the Chan-Vese model. If a condition on the data term was satisfied, a global minimum was guaranteed. It was shown theoretically and experimentally that the condition holds for the most commonly used data terms. If the condition was violated, which could happen for the L_1 data fitting term, relaxations were proposed which could produce global solutions in practice. Conditions on the "residual capacities" of the computed solution could be checked to verify whether a global minima of the original problem had been ob-

tained. A new convex relaxation of Pott's model with four regions was also proposed, which was at least as tight as the tightest existing relaxation, but significantly simpler to handle computationally. In the future, it should be investigated whether the dominance of the new relaxation over [35] is also strict.

In this work, we have restricted our attention to four (or less) regions. The results can be generalized to 2^m regions by using m level set functions. In case of $m = 3$ and 8 regions, the linear system which determines the data term contains 12 unknowns (edges) and 8 equations. In general, the conditions which guarantees submodularity will be more strict, therefore it will be valuable to investigate relaxations as in Section 6. On the other hand, four regions suffice in theory to segment any 2D image by the four color theorem, therefore it would be interesting to attempt formulating the overall problem in term of 4 disconnected regions, where different data cost functions are assigned to each disconnected component.

A Proofs of Prop 1-4

A.1 Proof of Prop 1

Proof Let $\frac{c_1-c_2}{2} \leq I \leq \frac{c_4-c_3}{2}$. Then

$$|c_2 - I|^\beta \leq |c_1 - I|^\beta \quad \text{and} \quad |c_3 - I|^\beta \leq |c_4 - I|^\beta,$$

for any $\beta \geq 1$. Therefore, adding these two inequalities

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta.$$

A.2 Proof of Prop 3

When $\frac{c_1-c_2}{2} \leq I \leq \frac{c_4-c_3}{2}$, the result follows from Prop (2). Consider $I < \frac{c_2-c_1}{2}$, then

$$|I - c_1|^{\beta_0} \leq |I - c_2|^{\beta_0} \leq |I - c_3|^{\beta_0} \leq |I - c_4|^{\beta_0}.$$

Together with (32), this implies

$$0 < |I - c_2|^{\beta_0} - |I - c_1|^{\beta_0} \leq |I - c_4|^{\beta_0} - |I - c_3|^{\beta_0}.$$

Therefore, there exists $\theta_1 \geq \theta_2 > 1$ such that

$$|I - c_4|^{\beta_0} = \theta_1 |I - c_3|^{\beta_0} \quad \text{and} \quad |I - c_2|^{\beta_0} = \theta_2 |I - c_1|^{\beta_0}.$$

For $\beta \geq \beta_0$

$$|I - c_4|^\beta = \theta_1^{\beta-\beta_0} |I - c_3|^\beta \quad \text{and} \quad |I - c_2|^\beta = \theta_2^{\beta-\beta_0} |I - c_1|^\beta,$$

where $\theta_1^{\beta-\beta_0} \geq \theta_2^{\beta-\beta_0} > 1$, hence

$$\begin{aligned} |I - c_2|^\beta + |I - c_3|^\beta &= \theta_2^{\beta-\beta_0} |I - c_1|^\beta + \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta \\ &\leq \theta_1^{\beta-\beta_0} |I - c_1|^\beta + \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta \leq \theta_1^{\beta-\beta_0} |I - c_1|^\beta + \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta, \end{aligned}$$

where the last inequality follows from $|I - c_1|^\beta \leq |I - c_4|^\beta$. Exactly the same argument can be used to show Prop 3 when $I > \frac{c_4-c_3}{2}$.

A.3 Proof of Prop 4

Proof Assume first $I > c_3$, then

$$|c_1 - I| > |c_2 - I| > |c_3 - I|$$

Therefore, there exists a $\theta > 1$ s.t.

$$|I - c_1| = \theta |c_2 - I|.$$

Pick $\mathcal{C}_I^1 \in \mathbb{N}$ s.t.

$$\theta^\beta \geq 2, \quad \forall \beta \geq \mathcal{C}_I^1.$$

Then

$$|c_1 - I|^\beta + |c_4 - I|^\beta \geq |c_1 - I|^\beta \geq 2|c_2 - I|^\beta > |c_2 - I|^\beta + |c_3 - I|^\beta. \quad \forall \beta \geq \mathcal{C}_I^1$$

If $I < c_2$, then

$$|c_4 - I| > |c_3 - I| > |c_2 - I|$$

and thus the same argument can be used to show there exists $\mathcal{C}_I^2 \in \mathbb{N}$ such that

$$|c_4 - I|^\beta + |c_1 - I|^\beta > |c_2 - I|^\beta + |c_3 - I|^\beta, \quad \forall \beta \geq \mathcal{C}_I^2.$$

In case $c_2 \leq I \leq c_3$, the existence of such a \mathcal{C} was proved in Prop 1, e.g. $\mathcal{C} = 1$.

Therefore the condition (32) is satisfied for any $I \in [0, L]$ by choosing $\beta \geq \mathcal{C} = \max_{I \in [0, L]} \max\{\mathcal{C}_I^1, \mathcal{C}_I^2\}$.

A.4 Proof of Prop 2

We will show the condition holds for $\beta = 1$, which by Prop 3 implies it holds for all $\beta \geq 1$. Observe that if c_1, c_2 and c_3, c_4 are equidistant it follows that $c_1 + c_4 = c_2 + c_3$. For $I < c_2$

$$|I - c_2| + |I - c_3| = (c_2 - I) + (c_3 - I) = -2I + (c_2 + c_3)$$

$$= -2I + (c_1 + c_4) = (c_1 - I) + (c_4 - I) \leq |I - c_1| + |I - c_4|.$$

For $I \geq c_3$

$$|I - c_2| + |I - c_3| = (I - c_2) + (I - c_3) = 2I - (c_2 + c_3)$$

$$= 2I - (c_1 + c_4) = (I - c_1) + (I - c_4) \leq |I - c_1| + |I - c_4|.$$

References

1. <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>.
2. E. Bae and X.C. Tai. Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In *Scale Space and Variational Methods in Computer Vision*, pages 1–13. Springer, 2009.
3. E. Bae, J. Yuan, X.C. Tai, and Y. Boykov. A fast continuous max-flow approach to non-convex multilabeling problems. Technical report CAM10-62, UCLA, CAM, 2010.
4. Egil Bae and Xue-Cheng Tai. Efficient global minimization for the multiphase Chan-Vese model of image segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 28–41, 2009.
5. Egil Bae and Xue-Cheng Tai. Efficient global optimization for the multiphase chan-ve-se model of image segmentation by graph cuts. *UCLA cam-report 09-53*, 2009.

6. Egil Bae, Jing Yuan, and Xue-Cheng Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.
7. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 359–374, 2001.
8. Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 26, Washington, DC, USA, 2003. IEEE Computer Society.
9. 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.
10. 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.
11. Ethan S. Brown, Tony F. Chan, and Xavier Bresson. Completely convex formulation of the Chan-Vese image segmentation model. *International Journal of Computer Vision*, 2011. doi: 10.1007/s11263-011-0499-y.
12. 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.
13. A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany, November 2008.
14. Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
15. T. Chan and L.A. Vese. *Active contours without edges*. IEEE Image Proc., 10, pp. 266–277, 2001.
16. Tony F. Chan, Selim Esedoğlu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648 (electronic), 2006.
17. Ivar Ekeland and Roger Téman. *Convex analysis and variational problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.
18. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in uncertain reasoning*, pages 452–472. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.
19. 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.
20. P.L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 01 optimization. *Mathematical Programming*, 28:121–155, 1984.
21. M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.
22. E. Hodneland. *Segmentation of digital images*. Cand. Scient Thesis, Department of Mathematics, University of Bergen. Available online at www.mi.uib.no/~tai, 2003.
23. 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.
24. Hiroshi Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1333–1336, 2003.
25. V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.
26. Vladimir Kolmogorov and Carsten Rother. Minimizing non-submodular functions with graph cuts—a review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1274–1279, 2007.
27. D. Fulkerson L. Ford. *Flows in networks*. Princeton University Press, 1962.
28. J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, HCI, IWR, Uni. Heidelberg, IWR, Uni. Heidelberg, November 2008.
29. 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. Mörken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision (SSVM 2009)*, volume 5567 of *LNCIS*, pages 150–162. Springer, 2009.
30. 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, 2006.
31. 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.
32. 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.
33. Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.
34. D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42, 42:577–685, 1989.
35. 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.
36. 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.
37. Renfrey B. Potts. Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society, Vol. 48*, pages 106–109, 1952.
38. Carsten Rother, Sanjiv Kumar, Vladimir Kolmogorov, and Andrew Blake. Digital tapestry. In *IEEE Proc. Computer Vision and Pattern Recognition*, 2005.
39. 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.
40. J. Yuan, E. Bae, and X.C. Tai. A study on continuous max-flow and min-cut approaches. In *CVPR, USA*, San Francisco, 2010.
41. J. Yuan, E. Bae, X.C. Tai, and Y. Boykov. A study on continuous max-flow and min-cut approaches. Technical report CAM10-61, UCLA, CAM, September 2010.
42. 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.