Efficient Global Minimization Methods for Image Segmentation Models with Four Regions

Egil Bae · Xue-Cheng Tai

Abstract We propose an exact global minimization framework for certain variational image segmentation models, such as the Chan–Vese model, involving four regions. A global solution is guaranteed if the data term satisfies a certain condition. We give a theoretical analysis of the condition for L^p type of data terms, such as in the Chan–Vese model and Mumford Shah model for $p = 2$. We show experimentally that the condition tends to hold in practice for $p \geq 2$ and also holds in many cases for more advanced data terms. If the condition is violated, convex and submodular relaxations are proposed which are not guaranteed to produce exact solutions, but tend to do so in practice. We also build up simple convex relaxations for some other four region segmentation models, including Potts' model. Algorithms are proposed which are very efficient compared to related work due to the simple and compact formulations.

Keywords Image segmentation · Variational models · Global optimization · Convex optimization · Combinatorial optimization

1 Introduction

Image segmentation is one of the core problems in image processing and computer vision. In mathematical terms, one wishes to find a partition $\{\Omega_i\}_{i=1}^n$ of the closed image domain

E. Bae (✉)
Department of Mathematics, University of California,
Los Angeles, CA, USA
e-mail: ebae@math.ucla.edu

X.-C. Tai
Department of Mathematics, University of Bergen, Bergen, Norway
e-mail: tai@math.uib.no

$\Omega \subset \mathbb{R}^N$. This paper deals with unsupervised image segmentation, where the intensity values of the input image I defined over Ω are the data available for making decisions about the partition. 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 [1] and Chan–Vese model [2,3] have been established as some of the most fundamental variational image segmentation models, whereas Potts' model is one of the most important discrete optimization models. Potts' 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 Potts' 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, involve 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 two 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) [4,5]. Convex reformulations of variational segmentation models with two regions have been proposed in [5], 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 guarantee convergence to a global minimum with reasonable efficiency (in polynomial time). A level set formulation of the Chan–Vese model with

multiple regions appeared in [3], which has since become very popular. It was proposed to solve the resulting gradient descent equations numerically, leading to a local minimum of the non-convex energy functional close to the initialization. In a discrete setting, alpha-expansion and alpha-beta swap [4,6] are the most popular algorithms for approximately minimizing the energy in Potts' model. More recently, attempts to derive convex relaxations in a continuous variational setting have been proposed for image partition problems [7–12] and vector valued labeling problems [8,13,14]. 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 instead lead to good approximations 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 [15], but this assumption is usually not valid in image segmentation applications.

The main advantages of variational models over discrete optimization models are 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 [4]. 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 four regions in the level set framework of Vese and Chan [3], 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. It is shown that a discrete version of the model can be minimized globally by computing the minimum cut on a graph, under a (mild) submodularity condition on the data term. Furthermore, 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.

We give an analysis of a sufficient condition under which L^p type of data terms are submodular for $p \geq 1$, which includes the Chan–Vese /Mumford–Shah data terms for $p = 2$. Arguments are given that the condition is expected to hold in practice for $p \geq 2$, which is also confirmed by experiments on a large image segmentation data base. If the condition does not hold, simpler submodular and convex relaxations are proposed. The relaxations are not guaranteed to produce global minima 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. We also propose relaxations for some other regularization terms including the continuous Potts' regularizer, which are compared experimentally with other recently proposed relaxations.

A dual formulation of the proposed convex problems is derived and related to the maximum flow problem over the graph used for combinatorial optimization of the discrete problems, inspired by the previous work [17–19]. The derivation of the dual problem is presented informally for variational problems and we leave rigorous proofs of existence of Lagrange multipliers as an open problem. After the eventual spatial discretization, rigorous proofs are provided by applying known duality theory. Efficient algorithms are derived based on the discretized dual problem, which are shown experimentally to converge faster than algorithms for related convex relaxations.

We focus fully on problems with four regions in this work. Problems with four regions are important: By the 4-colour theorem, four regions suffice to describe any partition of a 2D image. Some work has appeared on incorporating this property in image segmentation models recently [20,21], although a completely convex framework is still open. Our methods apply directly in some settings like [22], 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 important applications one would like to divide the image into four 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, but the conditions which guarantee global minima will be more strict, in which case one may need to settle for approximate solutions. Such generalizations are out of the scope of this paper and is the subject of a future work.

1.2 Organization

We start with a brief review of the segmentation models and previous optimization approaches in Sect. 2. Section 3 presents the new discrete approach for computing global minima by graph cuts, in case of submodular and non-submodular data terms. In Sect. 4 we present an exact convex reformulation of the segmentation models in a variational setting, in case the data term is submodular. Section 5 gives a detailed analysis of a condition that guarantees submodularity/convexity of the energy. In Sect. 6, we present convex relaxations in a variational setting for problems with non-submodular data term and some other regularization terms. Algorithms for the new convex models are derived in Sect. 7 and numerical experiments are presented Sect. 8.

The reader who is only interested in continuous optimization may optionally skip Sect. 3 on discrete optimization. However, Sect. 3 may be useful to understand the motivation behind some of the models and theorems, in particular the primal and dual formulations which are closely linked to min-cut and max-flow duality in the discrete setting.

Parts of this article are based on our preliminary conference paper [23]. More specifically, Sect. 3 on discrete optimization is an extended and elaborated version of [23]. Prop 2 and 5 without proofs also appeared in [23]. The convex formulation of the Chan Vese model in Sect. 4 and 6.1 appeared initially as part of the first author's Ph.D thesis [24].

2 Chan–Vese Model, the Mumford Shah Model and Potts' Model

2.1 Overview of the Models

We are interested in variational models which can generally be formulated as

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x)\,dx + \frac{\nu}{2} R(\{\partial\Omega_i\}_{i=1}^n) \tag{1}$$

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

The first term of the energy functional (1) is a data fitting term: $f_i(x)$ is the cost of assigning x to region Ω_i for each $x \in \Omega$ and each $i = 1, ..., n$. In general, it assumed the functions $f_i : \Omega \mapsto \mathbb{R}$, $i = 1, \ldots, n$ are uniformly bounded. A prominent example of the data term, which was first proposed in the seminal papers of Mumford-Shah [1] and Chan-Vese [2,3], is as follows

$$f_i(x) = |I(x) - c_i|^\beta, \quad \forall x \in \Omega, \ i = 1, \ldots, n. \tag{2}$$

Here $c_i \in \mathbb{R}$ are parameters associated with each Ω_i, for instance the mean intensity value inside Ω_i when $\beta = 2$.

The Chan–Vese model and the piecewise constant Mumford Shah model [1] have the form of (1) with data terms (2) and $\beta = 2$, but minimizes additionally over the parameters $c_i \in \mathbb{R}$

$$\min_{\{\Omega_i\}_{i=1}^n, \{c_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} |I(x) - c_i|^\beta\,dx + \nu R(\{\partial\Omega_i\}_{i=1}^n) \tag{3}$$

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

For fixed c, (3) has the same form as (1). A simple algorithm can be constructed which alternatingly 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 [25]. There has also been attempts to derive convex relaxations for the joint minimization problem (3) in case $n = 2$ [26]. The most challenging problem is to optimize in terms of the regions, and that will be the topic of this paper.

The last term $R(\{\partial\Omega_i\}_{i=1}^n)$ is a regularization term, weighted by the parameter $\nu \in \mathbb{R}$. Its purpose is to enforce regular region boundaries, typically as a function of the boundary lengths $|\partial\Omega_i|$. A simple and intuitive example is the total length of the region boundaries, i.e.

$$R(\{\partial\Omega_i\}_{i=1}^n) = \frac{1}{2}\sum_{i=1}^n |\partial\Omega_i| = \frac{1}{2}\sum_{i=1}^n \int_{\partial\Omega_i} d\mathcal{H}^{N-1}. \tag{4}$$

In a variational setting, such a boundary regularization was proposed by Mumford and Shah [1]. If the image domain is discrete, the equivalent discrete optimization problem with the above regularization term is often called Potts' model in the discrete optimization community. We will also refer to the continuous version (4) as Potts' model in this paper.

2.2 Representation by Level Set Functions and Binary Constrained Optimization

As a numerical realization, Chan and Vese [2,3] 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

$$\min_\phi \int_\Omega \{H(\phi)f_1 + (1 - H(\phi))f_2\} + \nu|DH(\phi)|, \tag{5}$$

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 [27,28], coined the

piecewise constant level set method (PCLSM), and resulted in the energy functional

$$\min_{\phi \in \mathcal{B}} \int_{\Omega} \{\phi f_1 + (1 - \phi) f_2\} dx + \nu \int_{\Omega} |D\phi|. \tag{6}$$

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

$$\mathcal{B} = \{\phi \in BV(\Omega) : \phi(x) \in \{0, 1\} \text{ a.e } x \in \Omega\}, \tag{7}$$

where the space of functions of bounded variations is defined as

$$BV(\Omega) = \{\phi \in L^1(\Omega) : \int_{\Omega} |D\phi| < \infty\}. \tag{8}$$

The total variation is defined in a distributional sense as (see e.g. [29])

$$\nu \int_{\Omega} |D\phi| := \sup_{q \in C_\nu} \int_{\Omega} \phi \operatorname{div} q, \, dx \tag{9}$$

where

$$C_\nu = \{q \in (C_c^\infty(\Omega))^N : \sup_{x \in \Omega} |q(x)|_2 \le \nu, \}.$$

if $\Omega = \mathbb{R}^N$. In this work, we assume the image domain Ω is closed and bounded, in which case the set C_ν is

$$C_\nu = \{q \in (C^\infty(\Omega))^N : \sup_{x \in \Omega} |q(x)|_2 \le \nu, \, q \cdot n = 0 \text{ at } \partial\Omega\}. \tag{10}$$

The problem (6) is non-convex since the side constraint $\phi \in \mathcal{B}$ is a non-convex set. In the seminal papers [16,30] it was realized that the problem can be made convex by instead minimizing of the convex set $\phi \in \mathcal{B}'$, where

$$\mathcal{B}' = \{\phi \in BV(\Omega) : \phi(x) \in [0, 1] \text{ a.e } x \in \Omega\}. \tag{11}$$

By first solving the convex relaxed problem (6) with the above constraints, global minimizers of the binary constrained problem can be obtained by thresholding the solution of the relaxed problem at any threshold in the interval (0, 1].

In [3], 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 $\phi^1, \ldots, \phi^m$, n regions could be represented in terms of the nonconvex heaviside functions $H(\phi^1), \ldots, H(\phi^m)$. An important special case is the representation of four regions by two level set functions ϕ^1, ϕ^2 as follows

$$\min_{\phi^1, \phi^2} E(\phi^1, \phi^2) = \nu \int_{\Omega} |DH(\phi^1)| + \nu \int_{\Omega} |DH(\phi^2)|$$

$$+ \int_{\Omega} \{H(\phi^1)H(\phi^2)f_2 + H(\phi^1)(1 - H(\phi^2))f_1$$

$$+ (1 - H(\phi^1))H(\phi^2)f_4$$

$$+ (1 - H(\phi^1))(1 - H(\phi^2))f_3\} dx. \tag{12}$$

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

$$\min_{\phi^1, \phi^2 \in \mathcal{B}} E(\phi^1, \phi^2)$$

$$= \nu \int_{\Omega} |D\phi^1| + \nu \int_{\Omega} |D\phi^2| + E^{data}(\phi^1, \phi^2), \tag{13}$$

subject to (7), where

$$E^{data}(\phi^1, \phi^2) = \int_{\Omega} \{\phi^1\phi^2 f_2 + \phi^1(1 - \phi^2)f_1$$

$$+ (1 - \phi^1)\phi^2 f_4 + (1 - \phi^1)(1 - \phi^2)f_3\} dx, \tag{14}$$

where f was given by (2) with $\beta = 2$. The problem (13) is nonconvex because both the binary constraints (7) are nonconvex and the energy functional $E^{data}(\phi^1, \phi^2)$ is nonconvex in ϕ^1, ϕ^2. Since the energy functional itself it nonconvex, one cannot obtain global minimizers by simply minimizing over $\phi^1, \phi^2 \in \mathcal{B}'$ as in the two region case. 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 [3], 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 [3], the level set formulation (12), (13) does not correspond exactly to the length term in Potts' model, because two of the six boundaries are counted twice in (13), 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.

Table 1 Representation of four regions by two binary functions

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

The interpretation of the regions can be permuted without influencing the energy

2.3 Related Models and Convex Relaxations

In this section, we give a closer examination of two other convex relaxation models that were discussed in the introduction, which are especially related to our work.

In case the number of regions is larger than two, it is possible to solve globally the simpler problem where the regularization term promotes a linear inclusion property of the regions. Define the integer valued "labeling function" u over the continuous image domain Ω. Any partition $\{\Omega_i\}_{i=1}^n$ can be described in terms of u by the convention $u(x) = i$ if $x \in \Omega_i$, $i = 1, \dots, n$. Consider the following integer constrained problem

$$\min_{u} \int_{\Omega} f_{u(x)}(x) + v \int_{\Omega} |Du|, \tag{15}$$

subject to $u(x) \in \{1, \dots, n\}$ a.e. $x \in \Omega$. In (15) the data term $f_{u(x)}(x)$ is the data cost of assigning x to region $\Omega_{u(x)}$. However, the regularization term of (15) does not correspond to the length term in the Potts' model (1) because of its dependency on the 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 (13), see Fig. 1 for an illustration. Such an overpenalization may cause the boundary between such regions to split. Ishikawa [15] showed that a discrete version of the model (15) could be minimized globally by computing the minimum cut on a graph. Later, a convex optimization framework was established for the continuous version (15)
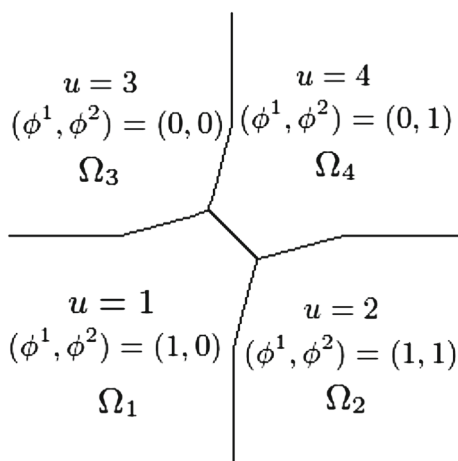

Fig. 1 The model (15) overcounts more severely than (13). In the model (15), the transition $\Omega_1 - \Omega_4$ is penalized three times, the transitions $\Omega_1 - \Omega_3$ and $\Omega_2 - \Omega_4$ are penalized 2 times while transitions $\Omega_1 - \Omega_2$ and $\Omega_3 - \Omega_4$ are penalized once. In the model (13), the transition $\Omega_1 - \Omega_4$ is penalized two times, while all the other transitions are penalized once

in [31]. The convex relaxation of Potts' model [11] minimizes (15) with additional constraints on the dual variables to prevent overcounting of boundaries.

There has recently been attempts to generalize the relaxation of the model (15), to the case that the unknown is a vector function, i.e. to solve

$$\min_{(u_1, \dots, u_K)} \int_{\Omega} f_{(u_1, \dots, u_K)(x)}(x) + v \sum_{i=1}^{K} \int_{\Omega} |Du_i|, \tag{16}$$

The work [14] proposed the tightest convex relaxation of the data term of this problem based on the convex envelope of the data term of (16). This has been studied further in [32], where convex relaxations for partition problems with generalized representations of the regions in terms of several integer or vector valued functions were derived. In case the unknown $(u_1, \dots, u_K)$ is constrained to a discrete set of binary values, the model (16) captures the model (13). By considering the relaxation (11) in [14] for $K = 2$ and binary labels and making a substitution of the simplex constrained variable v such that $v_i^1 = \phi^i$ and $v_i^2 = 1 - \phi^i$, we obtain

$$\min_{\{\phi^i(x) \in [0,1] \text{ a.e. } x \in \Omega\}_{i=1}^m} \sup_{\{p_0^k\}_{k=1}^m, \{p_1^k\}_{k=1}^m}$$

$$\times \int_{\Omega} p_0^1(1 - \phi^1) + p_1^1 \phi^1 + p_0^2(1 - \phi^2) + p_1^2 \phi^2 \, dx$$

$$+ \sum_{k=1}^{2} \int_{\Omega} |D\phi^k| \tag{17}$$

such that

$$p_1^1 + p_0^2 \le f_1, \quad p_1^1 + p_1^2 \le f_2, \quad p_0^1 + p_0^2 \le f_3, \quad p_0^1 + p_1^2 \le f_4.$$

Comparison of this problem with our convex models and relaxations will be given in Sect. 4.3. A crucial tool for proving exactness of the relaxations is the coarea formula. Note that it is not obvious that (17) satisfies the coarea formula, therefore it cannot be determined if the relaxation will produce global minimizers.

3 Global Minimization of 4-Region Chan–Vese Model in Discrete Setting by Graph Cuts

In this section we show how a discrete approximation of the Chan–Vese model (13) can be minimized exactly by computing the minimum cut on a novel graph.

3.1 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

$$TV_2(\phi) = \int_\Omega |D\phi|_2 = \int_\Omega \sqrt{|\phi_{x_1}|^2 + \ldots + |\phi_{x_N}|^2} \qquad (18)$$

and the anisotropic variant, by using 1-norm

$$TV_1(\phi) = \int_\Omega |D\phi|_1 = \int_\Omega |\phi_{x_1}| + \ldots + |\phi_{x_N}|. \qquad (19)$$

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 exactly by discrete optimization algorithms (e.g. mapped to the cut on a graph). It can be approximated to arbitrary precision if the size of the neighborhood system in the graph goes to infinity, a more detailed discussion is provided below.

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 $N = 2$, $\mathcal{P} = \{(i, j) \subset \mathbb{Z}^2\}$ 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

$$\mathcal{B} = \{\phi : \phi_p \in \{0, 1\} \ \forall p \in \mathcal{P}\} \qquad (20)$$

A discrete approximation of the two region model (6) can be derived as

$$\min_{\phi \in \mathcal{B}} \sum_{p \in \mathcal{P}} \phi_p f_1(p) + (1 - \phi_p) f_2(p)$$
$$+ \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p - \phi_q|, \qquad (21)$$

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

$$f_i(p) = |c_i - u_p^0|^\beta. \qquad (22)$$

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 [33], to approximate the isotropic total variation (13) (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 (13) can be expressed as

$$\min_{\phi^1, \phi^2 \in \mathcal{B}} E_d(\phi^1, \phi^2) = \sum_{p \in \mathcal{P}} E_p^{data}(\phi_p^1, \phi_p^2) \qquad (23)$$
$$+ \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^1 - \phi_q^1| + \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^2 - \phi_q^2|,$$

where

$$E_p^{data}(\phi_p^1, \phi_p^2) = \{\phi_p^1 \phi_p^2 f_2(p) + \phi_p^1 (1 - \phi_p^2) f_1(p))$$
$$+ (1 - \phi_p^1)\phi_p^2 f_4(p) + (1 - \phi_p^1)(1 - \phi_p^2) f_3(p)\}. \qquad (24)$$

3.2 Brief Review of Min-cut and Max-flow

Min-cut and max-flow are optimization problems defined over a graph which are dual to each other. 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 [4,34].

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. 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 partition 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

$$c(\mathcal{V}_s, \mathcal{V}_t) = \sum_{(v,w)\in\mathcal{E} \text{ s.t. } v\in\mathcal{V}_s, w\in\mathcal{V}_t} c(v, w). \qquad (25)$$

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.

$$p(e) \leq c(e), \quad \forall e \in \mathcal{E}. \qquad (26)$$

The max-flow problem aims to maximize the amount of from from $\{s\}$ to $\{t\}$ under flow conservation at each vertex. The theorem of Fulk–Fulkerson [35] states that this problem is dual/equivalent to the min-cut problem. An efficient implementation of augmented paths max-flow algorithm [35], specialized for image processing problems can be found online in [4]. This algorithm has been used in our experiments. For a given flow p, it is useful to define the residual edge capacities as $R(e) = c(e) - p(e) \ \forall e \in \mathcal{E}$.

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

$$\min_{x_i \in \{0,1\}} \sum_i E^i(x_i) + \sum_{i<j} E^{i,j}(x_i, x_j), \qquad (27)$$

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 [34,36], i.e. the regularization term must satisfy

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

3.3 Graph Construction for Energy Minimization Over Multiple Regions

Observe that in the discrete energy function (23), not only the regularization term, but also the data term is composed of pairwise interactions between binary variables. In this section we show how to 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.

$$E_p^{data}(1,1) + E_p^{data}(0,0) \leq E_p^{data}(1,0) + E_p^{data}(0,1) \tag{28}$$

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 (23)

$$\min_{(\mathcal{V}_s, \mathcal{V}_t)} c(\mathcal{V}_s, \mathcal{V}_t) = \min_{\phi^1, \phi^2 \in \mathcal{B}} E_d(\phi^1, \phi^2) + \sum_{p \in \mathcal{P}} \sigma_p, \tag{29}$$

where $\sigma_p \in \mathbb{R}$ are fixed for each $p \in \mathcal{P}$. In [34,36] a receipt was provided for minimizing a submodular function of the form (27) by computing the minimum cut on a graph. Problems of the form (23), where both the data term and regularization term consist of interactions between pairwise binary variables have not been directly considered. We provide a different geometric derivation of the graph for the specific problem (23) which does not rely on earlier results.

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

$$\mathcal{V} = \{v_{p,i} \mid p \in \mathcal{P}, \ i = 1, 2\} \cup \{s\} \cup \{t\}. \tag{30}$$

The edges are constructed such that the relationship (29) is satisfied. We begin with edges constituting the data term of (23). For each grid point $p \in \mathcal{P}$ they are defined as

$$\mathcal{E}_D(p) = (s, v_{p,1}) \cup (s, v_{p,2}) \cup (v_{p,1}, t) \cup (v_{p,2}, t)$$
$$\cup (v_{p,1}, v_{p,2}) \cup (v_{p,2}, v_{p,1}). \tag{31}$$

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

$$\mathcal{E}_R = \{(v_{p,1}, v_{q,1}), (v_{p,2}, v_{q,2}) \ \forall p, q \subset \mathcal{P} \ \text{s.t.} q \in \mathcal{N}_p^k\}. \tag{32}$$

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

$$\phi_p^1 = \begin{cases} 1 & \text{if } v_{p,1} \in V_s, \\ 0 & \text{if } v_{p,1} \in V_t, \end{cases} \qquad \phi_p^2 = \begin{cases} 1 & \text{if } v_{p,2} \in V_s, \\ 0 & \text{if } v_{p,2} \in V_t. \end{cases} \tag{33}$$

Weights are assigned to the edges such that the relationship (29) is satisfied. Weights on the regularization edges are simply given by

$$c(v_{p,1}, v_{q,1}) = c(v_{q,1}, v_{p,1}) = c(v_{p,2}, v_{q,2}) = c(v_{q,2}, v_{p,2})$$
$$= v w_{pq}, \qquad \forall p \in \mathcal{P}, \ q \in \mathcal{N}_p^k. \tag{34}$$

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

$$c(v_{p,1}, t) = A(p), \ c(v_{p,2}, t) = B(p), \ c(s, v_{p,1}) = C(p),$$
$$c(s, v_{p,2}) = D(p), c(v_{p,1}, v_{p,2}) = E(p), c(v_{p,2}, v_{p,1}) = F(p). \tag{35}$$

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

$$\begin{cases} A(p) + B(p) & = f_2(p) + \sigma_p \\ C(p) + D(p) & = f_3(p) + \sigma_p \\ A(p) + E(p) + D(p) & = f_1(p) + \sigma_p \\ B(p) + F(p) + C(p) & = f_4(p) + \sigma_p. \end{cases} \tag{36}$$

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

$$f_1(p) + f_4(p) = A(p) + B(p) + C(p)$$
$$+ D(p) + E(p) + F(p) - 2\sigma_p \geq A(p) + B(p)$$
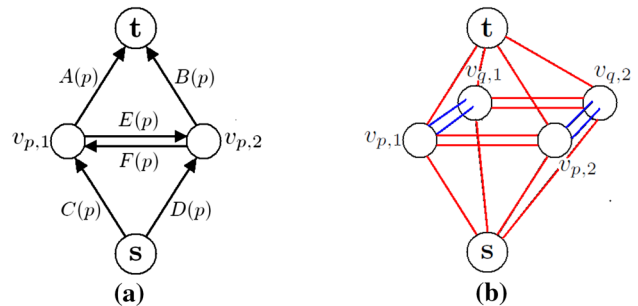$$+ C(p) + D(p) - 2\sigma_p = f_2(p) + f_3(p) \tag{37}$$

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* (Color figure online)

for all $p \in \mathcal{P}$. By inserting the data term (24) for f, we obtain the following requirement

$$|c_2 - u_p^0|^\beta + |c_3 - u_p^0|^\beta \le |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta, \quad (38)$$

for all $p \in \mathcal{P}$. If the distribution of the constant values satisfies (38), there exists a solution to (36) with $E(p), F(p) \ge 0$ for all $p \in \mathcal{P}$. Hence the problem can be solved globally by computing the minimum cut on the graph. In Sect. 5, we derive and analyze a sufficient condition that the data term is submodular, that does not depend on the input image u^0. Together with numerical experiments, it is shown that it is expected to hold in practice in case $\beta \ge 2$.

Note 1 Three region segmentation can be handled as a special case by putting infinite cost on one of the four assignments. For instance, assignment to Ω_4 will have infinite cost by putting $F(p) = \infty$. The cost of the remaining assignments are determined by finding a solution to the 3 first equations of the linear system (36). This system always has a solution with $E(p) \ge 0$, consequently the data term can always be made submodular in case of three regions. One solution is simply $D(x) = f_1(x)$, $E(x) = f_2(x)$, $A(x) = f_3(x)$ and $F(x) = \infty$. It should be noted however, that the resulting graph becomes identical to the graph from [15] with the simplification from [25] (where $n-1$ nodes are assigned to each pixel instead of n). The resulting minimization problem is exactly (15) with $n = 3$ in this case.

3.4 Submodular Relaxation of Non-submodular Data Terms

Consider the situation that the submodularity condition (37) is violated. Although the problem can be represented with negative $E(.)$, $F(.)$, it is not computationally tractable in general, since this amounts to find the minimal cut on a graph which contains negative edges, which is NP-hard in general. However, the problem can be solved exactly for many practical input data by a relatively simple combinatorial relaxation.

Approximate minimization of non-submodular energy functions in computer vision has been the subject of previous research, see [37] for a review. One of the most successful approaches is the Quadratic Pseudo Boolean Optimization (QPBO) [37,38]. In [39] 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 often can produce exact solutions in practice for our problem when the regions are permuted as in Table 1, and derive a condition which can checked after computation to verify whether one has obtained an exact solution. The relaxation can also be extended to the continuous setting, which is the topic of Sect. 6.1. The following relaxation was first published in shorter form in our conference paper [23]. An algorithm for minimizing the discrete energy (23) with QPBO was later proposed in [40].

QPBO can in theory give a better approximation (provided the permutation in Table 1 is made), but we expect it to be slower computationally because: (1) extra auxiliary nodes must be added to the graph before computing the minimum cut, (2) potential "unlabeled" nodes must subsequently be dealt with by an iterative brute force algorithm.

Observe that for all $p \in \mathcal{P}$ where (38) is violated, a solution can always be constructed to the linear system (36) where either $E(p) = 0$ or $F(p) = 0$. To prove this, let

$A(p), B(p), C(p), D(p), E(p), F(p)$ be any solution to the linear system (36) where $E(p) < 0$ or $F(p) < 0$. : Assume without loss of generality that $E(p) < 0$ and $E(p) \le F(p)$. Then another solution to (36) can be constructed with as follows: $A(p) \leftarrow A(p) - F(p)/2$, $B(p) \leftarrow B(p) + F(p)/2$, $C(p) \leftarrow C(p) + F(p)/2$, $D(p) \leftarrow D(p) - F(p)/2$, $E(p) \leftarrow E(p) + F(p)$ and $F(p) \leftarrow 0$. Similarly, if $F(p) < 0$ and $F(p) \le E(p)$, one can construct another solution where $E(p) = 0$.

Let $\overline{\mathcal{G}}$ 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 $\overline{\mathcal{G}}$ are constructed as

$$c(v_{p,1}, t) = A(p), \quad c(v_{p,2}, t) = B(p),$$
$$c(s, v_{p,1}) = C(p), \quad c(s, v_{p,2}) = D(p),$$
$$c(v_{p,1}, v_{p,2}) = \max(E(p), 0), \quad c(v_{p,2}, v_{p,1})$$
$$= \max(F(p), 0), \quad (39)$$

while the regularization edges are given as before by (34). The minimum cut on $\overline{\mathcal{G}}$ can easily be computed by max-flow. As will be discussed in Sect. 5, the condition (64) 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 $\overline{\mathcal{G}}$ to be almost identical to minimum cuts on $\mathcal{G}$. Define the sets

$$\mathcal{P}^1 = \{p \in \mathcal{P} \mid E(p) < 0, F(p) \ge 0\},$$
$$\mathcal{P}^2 = \{p \in \mathcal{P} \mid F(p) < 0, E(p) \ge 0\},$$

consisting of all $p \in \mathcal{P}$ for which either $E(p) < 0$ or $F(p) < 0$ (Fig. 3).

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
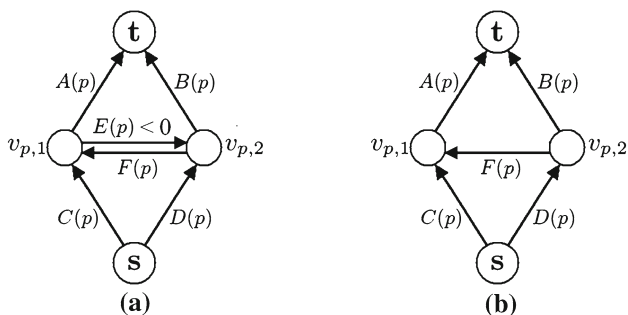
Fig. 3 Illustration of graph $\overline{\mathcal{G}}$ in case $E(p) < 0$. **a** $\mathcal{G}$. **b** $\overline{\mathcal{G}}$

when the minimum cut on $\overline{\mathcal{G}}$ yields the optimal solution of the original problem.

Theorem 1 *Let $\mathcal{G}$ be a graph as defined in (30)-(32) and (34), with weights $A(p), B(p), C(p), D(p), E(p), F(p)$ satisfying (36). 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

$$R_A(p) + R_D(p) \geq -E(p), \quad \forall p \in \mathcal{P}^1$$
$$and \quad R_B(p) + R_C(p) \geq -F(p), \quad \forall p \in \mathcal{P}^2, \qquad (40)$$

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

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

We first show min-cut$(\underline{\mathcal{G}}) \leq$ min-cut$(\mathcal{G}) \leq$ 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 $\mathcal{G}$ and $\underline{\mathcal{G}}$. 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 min-cut$(\underline{\mathcal{G}})$ = min-cut$(\overline{\mathcal{G}})$ which implies min-cut$(\mathcal{G})$ = min-cut$(\overline{\mathcal{G}})$. □

Therefore, by computing the max flow on $\overline{\mathcal{G}}$ and examining the residual capacities for criterion (40), it can be checked whether the solution is optimal on $\mathcal{G}$.

4 Exact Convex Formulation of 4-Region Chan–Vese Model in the Continuous Setting

Recall that the energy functional of the optimization problem (13) is non-convex in ϕ^1, ϕ^2. In this section we derive a formulation of (13) which is convex under the same condition (64) that made the discrete energy function (23) submodular, and hence allows for the computation of global minimizers. The convex formulation allows to deal with the rotationally invariant total variation (18), which cannot be handled by discrete algorithms.

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

$$\begin{cases} A(x) + B(x) & = f_2(x) + \sigma(x) \\ C(x) + D(x) & = f_3(x) + \sigma(x) \\ A(x) + E(x) + D(x) & = f_1(x) + \sigma(x) \\ B(x) + F(x) + C(x) & = f_4(x) + \sigma(x) \end{cases} \qquad (41)$$

where $\sigma(x)$ can be an arbitrary finite number. This is the same linear system as (36). Define the weights

$$C_s^1(x) = C(x), \quad C_s^2(x) = D(x),$$
$$C_t^1(x) = A(x), \quad C_t^2(x) = B(x), \qquad (42)$$
$$C^{12}(x) = E(x), \quad C^{21}(x) = F(x).$$

The above functions are uniformly bounded, since $\{f_i\}_{i=1}^4$ are assumed to be uniformly bounded. It can be checked that the following problem is equivalent to the original problem (13)

$$\min_{\phi^1, \phi^2} E^P(\phi^1, \phi^2) = \int_\Omega \{(1 - \phi^1)C_s^1 + (1 - \phi^2)C_s^2\}(x) \, dx$$
$$+ \int_\Omega \phi^1(x)C_t^1(x) + \phi^2(x)C_t^2(x) \, dx$$
$$+ \int_\Omega \max\{\phi^1(x) - \phi^2(x), 0\}C^{12}(x) \, dx$$

$$-\int_\Omega \min\{\phi^1(x) - \phi^2(x), 0\} C^{21}(x)\, dx$$

$$+\nu \int_\Omega |D\phi^1| + \nu \int_\Omega |D\phi^2|, \qquad (43)$$

such that $\phi^1, \phi^2 \in \mathcal{B}$, where $\mathcal{B}$ is the set of binary functions defined in (7).

Proposition 1 *Let ϕ^1, ϕ^2 be a minimizer of (43), then ϕ^1, ϕ^2 is a minimizer of (13).*

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 (43) if and only if ϕ^1, ϕ^2 is a minimizer of (13). □

The energy functional of (43) 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 (43). As in Sect. 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

$$\begin{aligned}
f_2(x) + f_3(x) &= A(x) + B(x) + C(x) + D(x) \\
&\leq A(x) + B(x) + C(x) + D(x) + F(x) + E(x) \\
&= f_1(x) + f_4(x)
\end{aligned} \qquad (44)$$

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

$$|c_2 - u^0(x)|^\beta + |c_3 - u^0(x)|^\beta \leq |c_1 - u^0(x)|^\beta + |c_4 - u^0(x)|^\beta, \qquad (45)$$

for all $x \in \Omega$. This is exactly the same condition (38) that made the discrete energy function submodular. In Sect. 5 we derive and analyze a sufficient condition that (45) holds for any possible input image. Analogously to the discrete setting, we provide here the explicit expression for one of the solutions of (41)

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

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

As in the discrete case, 3 region segmentation can be handled by putting infinite cost on one of the four assignments. In this case there are no restrictions on the data term and the problem can always be made convex. See Note 1 in Sect. 3.3 for details on how to set the coefficients.

Consider now the convex relaxed problem, where the binary constraints $\phi^1, \phi^2 \in \mathcal{B}$ are replaced by the convex constraints $\phi^1, \phi^2 \in \mathcal{B}'$ defined in (11), i.e.

$$\mathcal{B}' = \{\phi \in \mathrm{BV}(\Omega) : \phi(x) \in [0, 1] \text{ a.e } x \in \Omega\}.$$

A solution to the minimization problem (43) with the above constraints exists in $\mathrm{BV}(\Omega)$ since the energy functional is convex, lower semi-continuous, coercive, bounded below and the constraint set is a closed convex set. The next theorem shows that a global minimizer of the binary constrained non-convex problem can be obtained by thresholding the solution of the convex relaxed problem.

Theorem 2 *Assume the data term of (13) satisfies the condition (44), such that $C^{12}(x), C^{21}(x) \geq 0$ for all $x \in \Omega$ in (43). Let ϕ^{1*}, ϕ^{2*} be any solution of (43) subject to the binary constraints $\phi^1, \phi^2 \in \mathcal{B}'$. Denote by $\phi_i^\ell : \Omega \mapsto \{0, 1\}$ the binary function*

$$\phi_i^\ell(x) = \begin{cases} 1 & \text{if } \phi^{i*}(x) \geq \ell \\ 0, & \text{if } \phi^{i*}(x) < \ell \end{cases} \qquad (46)$$

Then for any $\ell \in (0, 1]$, $(\phi_1^\ell, \phi_2^\ell)$ is a solution of (43) and (13) subject to $\phi^1, \phi^2 \in \mathcal{B}$.

This theorem shows two important things. First, the convex relaxation (43) subject to (11) 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 For any $x \in \Omega$, if $\phi^i(x) \in [0, 1]$ then $\int_0^1 \phi_i^\ell(x)\, d\ell = \phi^i(x)$ and $\int_0^1 1 - \phi_i^\ell(x)\, d\ell = 1 - \phi^i(x)$. Furthermore, if $\phi^1(x) > \phi^2(x)$ then $\phi_1^\ell(x) \geq \phi_2^\ell(x)$ and if $\phi^1(x) \leq \phi^2(x)$ then $\phi_1^\ell(x) \leq \phi_2^\ell(x)$ for all $\ell \in (0, 1]$. Therefore

$$\int_0^1 \max(\phi_1^\ell(x) - \phi_2^\ell(x), 0)\, d\ell = \max(\phi^1(x) - \phi^2(x), 0)$$

and

$$-\int_0^1 \min(\phi_1^\ell(x) - \phi_2^\ell(x), 0)\, d\ell = -\min(\phi^1(x) - \phi^2(x), 0).$$

By the coarea formula

$$\int_0^1 \int_\Omega |D\phi_i^\ell|\, d\ell = \int_\Omega |D\phi^i|.$$

Combining the above properties we obtain that for any $\phi^1, \phi^2 \in \mathcal{B}'$

$$
\int_0^1 E^P(\phi_1^\ell, \phi_2^\ell) \, d\ell
$$

$$
= \int_0^1 \Big\{ \int_\Omega (1 - \phi_1^\ell(x)) C_s^1(x) + (1 - \phi_2^\ell(x)) C_s^2(x)
$$

$$
+ \phi_1^\ell(x) C_t^1(x) + \phi_2^\ell(x) C_t^2(x)
$$

$$
+ \max\{\phi_1^\ell(x) - \phi_2^\ell(x), 0\} C^{12}(x)
$$

$$
- \min\{\phi_1^\ell(x) - \phi_2^\ell(x), 0\} C^{21}(x) \, dx
$$

$$
+ \nu |D\phi_1^\ell| + \nu |D\phi_2^\ell| \Big\} d\ell
$$

$$
= E^P(\phi^1, \phi^2). \tag{47}
$$

For a pair ϕ^1, ϕ^2 that minimizes the energy, clearly $E^P(\phi^1, \phi^2) \leq E^P(\phi_1^\ell, \phi_2^\ell)$ for any $\ell \in (0, 1]$. However, equality (47) can then only be true provided $E^P(\phi^1, \phi^2) = E^P(\phi_1^\ell, \phi_2^\ell)$ for almost every $\ell \in (0, 1]$. In other words, ϕ_1^ℓ, ϕ_2^ℓ also minimizes the energy for almost every $\ell \in (0, 1]$.

It can also be shown that the theorem is valid for every $\ell \in (0, 1]$. The transition from "almost every" to "every" was proved for two regions in [41]. We use the same arguments to prove our theorem involving four regions. For some fixed $\ell \in (0, 1]$, there exists a strictly decreasing sequence $\{\ell_k\}_{k=1}^\infty$ converging to ℓ such that $\phi_1^{\ell_k}, \phi_2^{\ell_k}$ is a minimizer of (43) for every $k = 1, \ldots$ and $\phi_1^{\ell_k}, \phi_2^{\ell_k}$ converges to ϕ_1^ℓ, ϕ_2^ℓ for almost every $x \in \Omega$. By lower semi-continuity of the functional (43), it follows that ϕ_1^ℓ, ϕ_2^ℓ is a minimizer of (43). $\quad\square$

Discretization In order to use the convex relaxation (43) in a numerical algorithm, discretization is necessary. We assume in general that ϕ^1 and ϕ^2 are defined over a discrete domain, also denoted Ω, containing finitely many grid points. For example, a 2D rectangular image domain can be regarded as a regular grid $\Omega = \{(i, j) : i = 1, \ldots, n_x, \ j = 1, \ldots, n_y\}$ where each grid point corresponds to a pixel. The notations $\mathcal{B}$ and $\mathcal{B}'$ denote the set of functions that are binary and contained in $[0, 1]$ respectively, at each grid point in Ω. There are many choices of discretization of the differential and integration operators. In this paper, we avoid dependency on a particular choice of discretization and use the notations $D, \int$ also for general discrete differential and integration operators acting on functions defined over the discrete domain Ω. Using these notations, we can refer to (13) as a discretized model by specifying that all operators are discrete. An important distinction from the discrete models in Sect. 3, is the use of the 2-norm total variation (18), which has the advantage of being rotationally invariant. It should be noted that the coarea formula does not hold exactly for any known discretization of the 2-norm total variation (but does so for the 1-norm),

therefore the thresholding theorem above only holds approximately after discretization. Importantly, the approximation error goes to zero as the discretization becomes finer. The graph representable models (23) require that also the number of neighbors in the neighborhood system goes to infintity in order to converge to the rotationally invariant continuous regularization.

4.1 Primal–Dual Model

In the following parts of this section, we build up primal–dual and dual formulations of the primal model (43) in case $C^{12}(x), C^{21}(x) \geq 0$. The dual formulation eliminates the problem of non-differentiability of the data term in (43), and allows to build up very efficient algorithms which are presented in Sect. 7. Another reason for deriving the dual problem is to reveal connections between the convex relaxation (43) and the discrete min-cut/max-flow approach developed in Sect. 3. The dual model is inspired by our previous work [18,19], where continuous max-flow and min-cut models corresponding to binary (two region) problems were derived.

Remark The dual and primal–dual formulations are presented with continuous notation. We keep the presentation of the spatially continuous formulations somewhat informal. Questions regarding existence of dual and primal–dual solutions in the continuous setting are quite technical and left open to future research. For spatially discretized problems, these properties are proved by existing duality theory.

Define the functional

$$
D(\phi^1, \phi^2)
$$

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

$$
+ \int_\Omega \phi^1(x) p_t^1(x) + \phi^2(x) p_t^2(x)
$$

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

$$
+ \int_\Omega \phi^1(x) \operatorname{div} q^1(x) \, dx + \int_\Omega \phi^2(x) \operatorname{div} q^2(x) \, dx, \tag{48}
$$

subject to

$$
p_s^1(x) \leq C_s^1(x), \quad p_s^2(x) \leq C_s^2(x), \tag{49}
$$

$$
p_t^1(x) \leq C_t^1(x), \quad p_t^2 \leq C_t^2(x), \tag{50}
$$

$$
- C^{21}(x) \leq p^{12}(x) \leq C^{12}(x), \quad \text{a.e. } x \in \Omega \tag{51}
$$

$$
q^1 \in C_\nu, \quad q^2 \in C_\nu, \tag{52}
$$

where in the continuous setting, $p_s^i, p_t^i, p^{12} \in L^1(\Omega)$ and C_ν is defined in (10). Observe that all the dual variables

p_s^i, p_t^i, p^{12}, q^i, $i = 1, 2$ in (48) are independent, and can be optimized separately.

We will show that the minimization problem (43) with the convex constraints $\phi^1, \phi^2 \in \mathcal{B}'$ can be reformulated as the primal–dual model

$$\inf_{\phi^1, \phi^2} D(\phi^1, \phi^2). \qquad (53)$$

For each $x \in \Omega$, the integrand of the first 4 terms of (48) can be rewritten for $i = 1, 2$ as

$$\sup_{p_s^i(x) \leq C_s^i(x)} ((1 - \phi^i) p_s^i)(x)$$
$$= \begin{cases} ((1 - \phi^i) C_s^i)(x) & \text{if } \phi^i(x) \leq 1 \\ \infty & \text{if } \phi^i(x) > 1 \end{cases} \qquad (54)$$

$$\sup_{p_t^i(x) \leq C_t^i(x)} \phi^i(x) p_t^i(x) = \begin{cases} (\phi^i C_t^i)(x) & \text{if } \phi^i(x) \geq 0 \\ \infty & \text{if } \phi^i(x) < 0. \end{cases} \qquad (55)$$

Note that (53) is bounded above. To see this, define the zero function $\emptyset(x) = 0 \ \forall x \in \Omega$. Then $\inf_{\phi^1, \phi^2} D(\phi^1, \phi^2) \leq D(\emptyset, \emptyset) = \int_\Omega C_s^1(x) + C_s^2(x) \, dx$ due to constraints (49). Since C_s^1 and C_s^2 are uniformly bounded, this expression is finite. From (54) and (55) it therefore follows that optimal variables ϕ^1, ϕ^2 must satisfy the constraints

$$\phi^1(x), \phi^2(x) \in [0, 1] \text{ a.e. } x \in \Omega. \qquad (56)$$

If this was not the case, the primal–dual energy (48) would be infinite, contradicting boundedness from above.

The dual representation of total variation (9) can be used to rewrite the two last terms of (48)

$$\sup_{q^i \in C_\nu} \int_\Omega \phi^i \operatorname{div} q^i \, dx = \nu \int_\Omega |D\phi^i|, \quad i = 1, 2. \qquad (57)$$

From (57) it follows that optimal variables ϕ^1, ϕ^2 satisfy $\int_\Omega |D\phi^i| < \infty$, otherwise the energy would be infinite, again contradicting boundedness from above. Together with observation (56), it therefore follows that optimal variables ϕ^1, ϕ^2 are contained in the set $\mathcal{B}'$.

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

$$\sup_{-C^{21}(x) \leq p^{12}(x) \leq C^{12}(x)} (\phi^1(x) - \phi^2(x)) \, p^{12}(x)$$
$$= \max\{\phi^1(x) - \phi^2(x), 0\} C^{12}(x)$$
$$- \min\{\phi^1(x) - \phi^2(x), 0\} C^{21}(x). \qquad (58)$$

Therefore, combining (54), (55), (57) and (58), by maximizing the primal dual model (48) for $p_s^1, p_s^2, p_t^1, p_t^2, p^{12}, q^1, q^2$,

we obtain the primal model (43) subject to the convex constraints $\phi^1, \phi^2 \in \mathcal{B}'$.

Discretization We can also use the above notation for the discretized problem. After discretization, the functions p_s^i, p_t^i $i = 1, 2$ and p^{12} are defined over the same discrete domain Ω as ϕ^1, ϕ^2, while the spatial flow functions q^1 and q^2 are defined between neighboring grid points. For example, if Ω is a 2D regular grid, the flow functions are typically defined on the middle points $\Omega^x = \{(i - 1/2, j) : i = 1, \ldots, n_x + 1, j = 1, \ldots, n_y\}$ and $\Omega^y = \{(i, j - 1/2) : i = 1, \ldots, n_x, j = 1, \ldots, n_y + 1\}$ and $\operatorname{div} : \Omega_d^x \times \Omega_d^y \mapsto \Omega_d$ maps q onto Ω_d. The set C_ν is in this case

$$C_\nu = \{(q^1, q^2) : \Omega_d^x \times \Omega_d^y \mapsto \mathbb{R}^2 :$$
$$\Big(\Big| \frac{q^1(i - \frac{1}{2}, j) + q^1(i + \frac{1}{2}, j)}{2} \Big|^2$$
$$+ \Big| \frac{q^2(i, j - \frac{1}{2}) + q^2(i, j + \frac{1}{2})}{2} \Big|^2 \Big)^{\frac{1}{2}} \leq \nu$$
$$i = 1, \ldots, n_x, \ j = 1, \ldots, n_y \}. \qquad (59)$$

There are several choices for the discrete divergence div, but it must satisfy $\operatorname{div} = -D^*$ for the particular choice of discrete gradient D, such that the defining identity $\int_\Omega \phi \operatorname{div} q \, dx = -\int_\Omega D\phi \cdot q \, dx$ is valid for all discrete ϕ and q. A simple choice of discretization is a forward scheme for $D : \Omega_d \mapsto \Omega_d^x \times \Omega_d^y$ and a backward scheme for div. In our experiments, we use a mimetic discretization scheme [42]. Note that if Ω contains finitely many points, the expression *a.e.* $x \in \Omega$ has the same meaning as $\forall x \in \Omega$.

4.2 Dual model

The terms in the primal–dual functional (48) can be rearranged as follows

$$D(\phi^1, \phi^2) = \{ \sup_{p_s, p_t, p^{12}, q} \int_\Omega p_s^1(x) + p_s^2(x) \, dx \qquad (60)$$

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

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

subject to (49)–(52). Observe that since the variables ϕ^1, ϕ^2 are unconstrained in the primal–dual model (53), they can be interpreted as Lagrange multipliers for the constraints

$$\operatorname{div} q^1(x) - p_s^1(x) + p_t^1(x) + p^{12}(x) = 0, \quad \text{a.e. } x \in \Omega \qquad (61)$$

$$\operatorname{div} q^2(x) - p_s^2(x) + p_t^2(x) - p^{12}(x) = 0, \quad \text{a.e. } x \in \Omega. \qquad (62)$$

see e.g. [43] (0.0.13)-(0.0.14).

Remark In the continuous setting, i.e. for variational problems, it is complicated to establish existence of Lagrange muliipliers unless the constraints satisfy conservative conditions. The nature of the open set C_ν makes it difficult to apply well known theory, see discussion below more detail. We therefore leave strict proofs of existence of solutions as an open problem. The derivations in this section regarding continuous spatial domains must therefore be seen in an informal sense. After discretization, the existence of Lagrange multipliers follows directly from the min-max theorem, more details are given below.

By removing the Lagrange multipliers, we obtain the dual problem

$$\sup_{p_s, p_t, p^{12}, q} E^D(p_s, p_t, p^{12}, q) = \int_\Omega p_s^1(x) + p_s^2(x)\, dx$$
(63)

subject to the constraints (49)–(52), (61) and (62). The dual problem (63) can be interpreted as a maximum flow problem over a continuous generalization of the graph proposed in Sect. 3.3. The dual variables p_s^i, p_t^i, p^{12}, q^i $i = 1, 2$ are interpreted as flow functions on the edges and the constraints (61) and (62) are flow conservation constraints. The interested reader can find the details in Appendix 1.

Discussion on existence of primal–dual solutions In the spatially continuous setting, the above functional spaces are not reflexive, therefore one cannot directly apply the minimax theorem to prove existence of a primal–dual solution. As established under (55), the dual problem is bounded above. This guarantees the existence of a maximizing sequence $\{p_s^{i^k}\}_{k=1}^\infty, \{p_t^{i^k}\}_{k=1}^\infty, \{p^{12^k}\}_{k=1}^\infty, \{q^{i^k}\}_{k=1}^\infty$ inside the constraint set (10) such that $\lim_{k\to\infty} E^D(p^k, q^k)$ is equal to the supremum in problem (63). In the classical definition of existence, there must exist a maximizing subsequence converging to a point within the constraint set. However, this is not true for this problem, because the supremum with respect to the flow variable q may be attained for a discontinuous flow field q, which is in the closure of the constraint set of smooth q (10) and not contained in the set itself. When we speak of a dual solution, we shall mean a solution where q may be discontinuous.

After discretization, the functional spaces become finite dimensional Hilbert spaces. One can therefore directly apply the minimax theorem, Prop. 2.4 of [44] Chapter VI, to conclude that a primal–dual solution exists to the problem (48) and the duality gap between the primal and dual problems is zero.

4.3 Comparison to Product Space relaxation [14]

It is interesting to compare our convex model (43) to the convex relaxation (17) derived from [14]. The data term of the relaxation (17) is based on the convex envelope of the data term of the original problem, which is the tightest convex relaxation of the data term that theoretically exists (meaning it has at least as high value as any other convex relaxation for each feasible ϕ^1, ϕ^2). However, after the regularization term is added, there is no guarantee that an exact solution will be produced. In the last section it was established that our convex model with regularization term exactly represents the original problem: It has the same minimal value and minimizers of the original problem can be obtained by thresholding the solution to the convex problem. This indicates that the relaxation (17) is also exact in case of submodular data terms with regularizer, since the relaxation of the data term must be at least as tight as the proposed relaxation. Therefore (17) should also satisfy the coarea formula and thresholding theorem under these conditions. If the data term is not submodular, one cannot guarantee that (17) will give a global minimizer. If this was true in general, one would have a polynomial algorithm for solving an NP-hard problem, which is unlikely to exist.

In fact, our primal–dual problem (60) has a closer resemblance to (17). An important distinction is that the dual constraints in our problem are all separable, which allows us to prove the coarea formula for our model. This also implies that for a fixed ϕ^1, ϕ^2, our model can be maximized for all dual variables simultaneously, by maximizing in terms of each dual variable separately. This important property allows us to design very efficient algorithms, which are presented in Sect. 7. In contrast, the dual variables are coupled through the dual constraints in (17), which makes projection more difficult. Since no closed form solution exists, a more expensive inexact iterative algorithm for projecting onto the constraint set of $p_0^1, p_0^2, p_1^1, p_1^2$ each iteration would be required.

5 Analysis of Submodularity/Convexity Condition

We give an analysis of the condition (44) for making the optimization problems submodular/convex in case of data terms of the form (2). By assuming the input image contains all gray values $I \in [0, L]$, we get from (38) or (45) the following sufficient condition on the distribution of the constant values $c_1, \ldots, c_4$

$$|c_2 - I|^\beta + |c_3 - I|^\beta \le |c_1 - I|^\beta + |c_4 - I|^\beta, \quad \forall I \in [0, L].$$
(64)

If (64) holds, the problem is submodular for any input image u^0. This condition says something about how evenly $\{c_i\}_{i=1}^4$ are distributed. The following two results show that (64) becomes less strict as β increases

Proposition 2 *Let $0 \le c_1 < c_2 < c_3 < c_4$. If (64) is satisfied for some $\beta_0 \ge 1$, then (64) is satisfied for all $\beta \ge \beta_0$.*
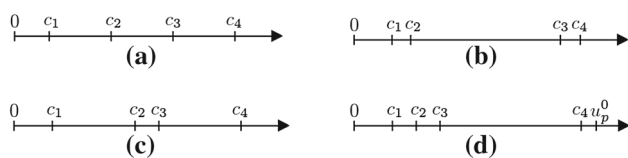
Fig. 4 **a**, **b** and **c** distributions of $\{c_i\}_{i=1}^4$ which makes energy function submodular for all β. **d** distribution of $\{c_i\}_{i=1}^4$ which may make energy function nonsubmodular for small β

Proposition 3 *Let* $0 \leq c_1 < c_2 < c_3 < c_4$. *There exists a* $\mathcal{C} \in \mathbb{N}$ *such that (64) is satisfied for any* $\beta \geq \mathcal{C}$.

The proofs of all propositions in this section can be found in Appendix 1. It can also be observed that the condition can only be violated if I is closer to c_1 or c_4 than the other variables

Proposition 4 *Let* $0 \leq c_1 < c_2 < c_3 < c_4$. *(64) is satisfied for all* $I \in [\frac{c_2+c_1}{2}, \frac{c_4+c_3}{2}]$ *for any* $\beta \geq 1$.

And lastly, if the constants are perfectly symmetrically distributed around their mean, the condition is satisfied for any $\beta \geq 1$

Proposition 5 *Let* $0 \leq c_1 < c_2 < c_3 < c_4$. *(64) is satisfied for any* $\beta \geq 1$ *if* $c_2 - c_1 = c_4 - c_3$.

For $\beta > 1$, greater assymetry is tolerated by Prop 2. These propositions suggest that the condition can only be violated if the assymetry of the distribution of $c_1, \ldots, c_4$ around their mean becomes too "bad".

Examples where the condition is satisfied and may fail are depicted in Fig. 4. Prop. 2 is illustrated in Fig. 4b, c. Figure 4d shows the possibility in which (64) may be violated, i.e. c_1, c_2, c_3 are clustered compared to c_4 (the opposite version where c_2, c_3, c_4 are clustered would also be a problem).

We are mostly interested in the case that $\beta = 2$ and are interested in knowing how often the condition is expected to be satisfied. To get a raw empirical estimate, we have picked the constant values $c_1, \ldots, c_4$ randomly in the interval $[0, 1]$ and investigated how often the condition (64) is violated for $I \in [0, 1]$. The condition was satisfied in 39.6 percent out of 10000 random selections of $c_1, \ldots, c_4$. In cases that the condition was violated, the constants were distributed unevenly around their mean as illustrated in Fig. 4d. However, such distributions of the constants are not expected in practice, because when minimized over the region parameters, the model (13) will favor solutions where the data functions (2) corresponding to each region are dissimilar. In fact, when the model (13) was optimized over the constants $c_1, \ldots, c_4$ in addition to the regions, the condition was satisfied on all 100 images in the Berkeley image segmentation database [45]. More details will be given in Sect. 8.3.

6 Convex Relaxations of NP-Hard Data and Regularization Terms in Continuous Setting

In this section, we develop convex relaxations which can be applied on a broader set of variational problems than the models in Sect. 4. We start by considering data terms where the convexity/submodularity condition (44) is not satisfied in Sect. 6.1. Section 6.2 deals with some other regularization terms such as the Potts' regularizer. In contrary to the models in Sect. 4, it cannot be guaranteed in advance that exact solutions will be produced. Instead, conditions are derived which can be checked after computation whether a global minimum has been attained. Especially for problems with non-submodular data terms, it is our observation that these hold in practice. If they do not hold exactly, the relaxations will provide good approximate solutions. The relaxation (17) could also be used to approximately solve the problem in case of non-submodular data terms. The advantage of the following approach is that the thresholding operation (2) can be applied for producing a global binary minimizer in case the conditions on the computed solution holds. The relaxed problem is also easier to deal with computationally than (17) for the same reasons described in Sect. 4.3.

6.1 Convex Relaxation of Non-convex Data Term

Recall that if $E(x)$ or $F(x)$ are negative for some $x \in \Omega$, the variational problem (43) is non-convex, hence any minimization algorithm based on steepest descent may get stuck in a local minimum. We propose a convex relaxation for this problem, inspired by the submodular relaxation from Sect. 3.4 and derive a condition for determining whether the computed solution is globally optimal.

For all points $x \in \Omega$ where (64) 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 $F(x) < 0$. As discussed in Sect. 3.4, there always exists a solution where either $E(x) = 0$ or $F(x) = 0$.

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.

$$\Omega^1 = \{x \in \Omega : E(x) < 0\}, \quad \Omega^2 = \{x \in \Omega : F(x) < 0\}. \tag{65}$$

Let P denote the original primal problem (43) with weights set to

$$
\begin{aligned}
C_s^1(x) &= C(x), \quad C_s^2(x) = D(x), \\
C_t^1(x) &= A(x), \quad C_t^2(x) = B(x), \\
C^{12}(x) &= E(x), \quad C^{21}(x) = F(x), \quad \forall x \in \Omega
\end{aligned}
\tag{66}
$$

as before. Since $C^{12}(x)$ or $C^{21}(x)$ are assumed negative for some $x \in \Omega$, the minimization problem (43) is nonconvex.

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 Sect. 4.2. 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 (43), (48) and (63) respectively, with weights constructed as

$$C_s^1(x) = C(x), \quad C_s^2(x) = D(x),$$
$$C_t^1(x) = A(x), \quad C_t^2(x) = B(x), \tag{67}$$
$$C^{12}(x) = \max(E(x), 0), \ C^{21}(x) = \max(F(x), 0), \ \forall x \in \Omega.$$

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 $\overline{D}$.

In the following theorem and proof it is assumed for simplicity that the problems are discretized such that ϕ^i, p_s^i, p_t^i, $i = 1, 2$ and p^{12} are defined over a discrete domain Ω containing finitely many grid points. The spatial discretization is denoted D, div $= -D^*$ and can be arbitrary, see Sect. 4 for a more detailed discussion about discretization. In this case, the existence of primal–dual solutions and a zero duality gap follows by the minimax theorem. If one assumes a continuous spatial domain, the following derivations are carried out in an informal way.

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*

$$R_s^1(x) = C_s^1(x) - \overline{p}_s^1(x), \quad R_s^2(x) = C_s^2(x) - \overline{p}_s^2(x),$$
$$R_t^1(x) = C_t^1(x) - \overline{p}_t^1(x), \quad R_t^2(x) = C_t^2(x) - \overline{p}_t^2(x). \tag{68}$$

If

$$R_t^1(x) + R_s^2(x) \geq -E(x), \quad \forall x \in \Omega^1 \tag{69}$$

and

$$R_t^2(x) + R_s^1(x) \geq -F(x), \quad \forall x \in \Omega^2, \tag{70}$$

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

Proof Let E^P denote the energy functional of the primal problem (43) with original weights (66), and let $\overline{E}^P$ denote the energy functional of (43) with weights (67). Then for any functions ϕ^1, ϕ^2 such that $\phi^1(x), \phi^2(x) \in [0, 1]$ for all $x \in \Omega$

$$\overline{E}^P(\phi^1, \phi^2) \geq E^P(\phi^1, \phi^2). \tag{71}$$

Define a new problem $\underline{P}$ as the minimization of (43) with weights (67) for all $x \in \Omega \backslash (\Omega^1 \cup \Omega^2)$ and

$$C_s^1(x) = C(x), \quad C_s^2(x) = \overline{p}_s^2(x),$$
$$C_t^1(x) = \overline{p}_t^1(x), \quad C_t^2(x) = B(x), \quad \forall x \in \Omega^1, \tag{72}$$
$$C_s^1(x) = \overline{p}_s^1(x), \quad C_s^2(x) = D(x),$$
$$C_t^1(x) = A(x), \quad C_t^2(x) = \overline{p}_t^2(x), \quad \forall x \in \Omega^2, \tag{73}$$
$$C^{12}(x) = 0, \quad C^{21}(x) = 0, \quad \forall x \in \Omega^1 \cup \Omega^2. \tag{74}$$

Let $\underline{E}^P$ denote the energy functional of (43) with the above defined weights. We will show that

$$\underline{E}^P(\phi^1, \phi^2) \leq E^P(\phi^1, \phi^2) \leq \overline{E}^P(\phi^1, \phi^2), \quad \forall \phi^1, \phi^2 \in \mathcal{B}'. \tag{75}$$

The right inequality is just a repetition of (71). To show the left inequality, observe that for each $x \in \Omega^1$

$$(1 - \phi^1(x))C(x) + (1 - \phi^2(x))\overline{p}_s^2(x)$$
$$+ \phi^1(x)\overline{p}_t^1(x) + \phi^2(x)B(x)$$
$$= (1 - \phi^1(x))C(x) + (1 - \phi^2(x))(D(x) - R_s^2(x))$$
$$+ \phi^1(x)(A(x) - R_t^1(x)) + \phi^2(x)B(x)$$
$$= (1 - \phi^1(x))C(x) + (1 - \phi^2(x))D(x) + \phi^1(x)A(x)$$
$$+ \phi^2(x)B(x) + \phi^1(x)(-R_t^1(x)) + (1 - \phi^2(x))(-R_s^2(x)).$$

The last two terms can be bounded by

$$\phi^1(x)(-R_t^1(x)) + (1 - \phi^2(x))(-R_s^2(x))$$
$$\leq \phi^1(x)(1 - \phi^2(x))(-R_t^1(x))$$
$$\quad + \phi^1(x)(1 - \phi^2(x))(-R_s^2(x))$$
$$= \phi^1(x)(1 - \phi^2(x))(-R_t^1(x) - R_s^2(x))$$
$$\leq \max(\phi^1(x) - \phi^2(x), 0)(-R_t^1(x) - R_s^2(x))$$
$$\leq \max(\phi^1(x) - \phi^2(x), 0)E(x).$$

Therefore the integrand of the the data term of $\underline{E}^P$ is lower than E^P for any $x \in \Omega^1$. Exactly the same argument can be used to show that this is also true for any $x \in \Omega^2$. For any $x \in \Omega \backslash (\Omega^1 \cup \Omega^2)$, the integrand of the data term is identical in $\underline{E}^P$ and E^P. The regularization term is identical in $\underline{E}^P$ and E^P, hence it follows that $\underline{E}^P(\phi^1, \phi^2) \leq E^P(\phi^1, \phi^2)$ for any $\phi^1, \phi^2 \in \mathcal{B}'$. Observe that since the maximum flow $\overline{p}_t^i, \overline{p}^{12}, \overline{q}^i; i = 1, 2$ in problem $\overline{D}$ is feasible in $\underline{D}$, it is by (75) also optimal on $\underline{D}$. It follows that $\underline{E}^P(\overline{\phi}^1, \overline{\phi}^2) = \overline{E}^P(\overline{\phi}^1, \overline{\phi}^2)$, which by (75) implies $E^P(\overline{\phi}^1, \overline{\phi}^2) = \overline{E}^P(\overline{\phi}^1, \overline{\phi}^2)$. □

6.2 Convex Relaxation of Potts' Regularizer and Some Other Regularizers

In this section we show that the convex reformulation (43) can be used as the basis of a new convex relaxation of Potts' regularization term (4) and some other regularization terms with four regions.

In [11,46] it was observed that multiple countings of boundaries in the model (15) with linearly ordered labels, could be suppressed by introducing additional constraints on the dual variables of the total variation term of (15). The number of constraints grow quadratically in the number of regions. In particular, 3 dual variables and 6 dual constraints are necessary to represent four regions. The resulting convex relaxation is not guaranteed to produce a global binary solution, but as shown in [11,46], such an approach may produce good approximations after thresholding. Furthermore, as shown in [11], Prop 4, the relaxation is strictly tighter than other recently proposed approaches [7,10,12]. In sect. 4 it was shown that the model (43) could be optimized globally by relaxing the binary constraints of ϕ^1 and ϕ^2. As discussed in Sect. 2.3, this model approximates Potts' model more closely than the model (15). Two of the boundaries are measured twice in (43), while all the remaining four boundaries are measured once. In contrast, the model (15) overcounts the boundaries much more severely, see e.g. Fig. 1. This motivates us to use (43) as a starting point for a new convex relaxation of Potts' model.

Consider first the convex model (43), expressed in terms of the dual formulation of total variation (57). We then obtain

$$
\begin{aligned}
\inf_{\phi^1,\phi^2} \sup_{q^1,q^2} E^P(\phi^1,\phi^2) = &\int_\Omega (1-\phi^1(x))C_s^1(x) \\
&+ (1-\phi^2(x))C_s^2(x) + \phi^1(x)C_t^1(x) + \phi^2(x)C_t^2(x)\,dx \\
&+ \int_\Omega \max\{\phi^1(x)-\phi^2(x),0\}C^{12}(x)\,dx \\
&- \int_\Omega \min\{\phi^1(x)-\phi^2(x),0\}C^{21}(x)\,dx \\
&+ \int_\Omega \phi^1\,\mathrm{div}\,q^1\,dx + \int_\Omega \phi^2\,\mathrm{div}\,q^2\,dx,
\end{aligned}
\tag{76}
$$

subject to $\phi^1,\phi^2 \in \mathcal{B}$ and $q^1,q^2 \in C_\nu$, where C_ν is defined in (10).

The distributional derivative $D\phi$ can be decomposed as [46–48] $D\phi = \nabla\phi\,dx + C\phi + (\phi_+ - \phi_-)n_\phi d\mathcal{H}^{N-1} \lrcorner J_\phi$, where $\nabla\phi\,dx$ is the part which is continuous with respect to the Lebesgue measure and $C\phi$ is the Cantor part. Here J_ϕ is the discontinuity set of ϕ and ϕ_+ and ϕ_- are the values of ϕ at the upper and lower side of the discontinuity respectively, such that $\phi_+ > \phi_-$. The vector $n_\phi(x) \in \mathbb{S}^{N-1}$ is the normal vector at point x of the discontinuity J_ϕ pointing in

the direction of the lower side of the discontinuity. A more formal definition of the jump set and normal vector can be found in [46] page 1118. By integration by parts, we have that for a q with compact support or which satisfies the boundary conditions in (10) that

$$
\begin{aligned}
\int_\Omega \phi\,\mathrm{div}\,q\,dx &= -\int_\Omega q \cdot D\phi \\
&= -\int_\Omega q \cdot \nabla\phi\,dx - \int_\Omega q \cdot C u \\
&\quad - \int_{J_\phi} q(x) \cdot n_\phi(x)\,d\mathcal{H}^{N-1}(x)
\end{aligned}
$$

The Cantor part can be shown to vanish for $N-1$ dimensional subsets of finite measure [46]. Assuming that $\phi \in \mathcal{B}$, the two first terms of the above expression therefore disappear. If $\phi^1,\phi^2 \in \mathcal{B}$, the regularization term can therefore be written

$$
\begin{aligned}
\int_\Omega &\phi^1\,\mathrm{div}\,q^1 + \phi^2\,\mathrm{div}\,q^2\,dx \\
&= -\sum_{i=1}^{2}\int_{J_{\phi^i}} q^i(x) \cdot n_{\phi^i}(x)\,d\mathcal{H}^{N-1}(x)
\end{aligned}
$$

which is maximized over $q^1,q^2 \in C_\nu$ by picking $q^i(x) = -\nu n_{\phi^i}(x)$ for $x \in J_{\phi^i}, i = 1, 2$. Observe first that the integral can be decomposed as

$$
\begin{aligned}
- &\int_{J_{\phi^1} \cap J_{\phi^2}} \sum_{i=1}^{2} q^i(x) \cdot n_{\phi^i}(x)\,d\mathcal{H}^{N-1}(x) \\
- &\int_{J_{\phi^1} \setminus J_{\phi^1} \cap J_{\phi^2}} q^1(x) \cdot n_{\phi^1}(x)\,d\mathcal{H}^{N-1}(x) \\
- &\int_{J_{\phi^2} \setminus J_{\phi^1} \cap J_{\phi^2}} q^2(x) \cdot n_{\phi^2}(x)\,d\mathcal{H}^{N-1}(x).
\end{aligned}
$$

The set $J_{\phi^1} \cap J_{\phi^2}$ can be decomposed as $J^+ \cup J^-$, where $J^+ = \{x \in J_{\phi^1} \cap J_{\phi^2} : n_{\phi^1}(x) = n_{\phi^2}(x)\}$ and $J^- = \{x \in J_{\phi^1} \cap J_{\phi^2} : n_{\phi^1}(x) = -n_{\phi^2}(x)\}$. Then the first integral can be written as follows

$$
\begin{aligned}
- &\int_{J_{\phi^1} \cap J_{\phi^2}} \sum_{i=1}^{2} q^i(x) \cdot n_{\phi^i}(x)\,d\mathcal{H}^{N-1}(x) \\
= - &\int_{J^+} (q^1(x) + q^2(x)) \cdot n_{\phi^1}(x)\,d\mathcal{H}^{N-1}(x) \\
- &\int_{J^-} (q^1(x) - q^2(x)) \cdot n_{\phi^1}(x)\,d\mathcal{H}^{N-1}(x).
\end{aligned}
$$

From Table 1 it follows that J^+ is the boundary between Ω_2 and Ω_3. J^- is the boundary between Ω_1 and Ω_4. By inserting the maximizer $q^i(x) = -n_{\phi^i}(x)$, it can be seen that the integrals over J^+ and J^- are valued twice their total length (weighted by ν): $2\nu \int_{J^+} d\mathcal{H}^{N-1}(x)$ and $2\nu \int_{J^-} d\mathcal{H}^{N-1}(x)$. In order to count the length of the boundary between Ω_2 and Ω_3 only once, we add the constraint

$$\sup_{x \in \Omega} |q^1(x) + q^2(x)|_2 \leq \nu. \qquad (77)$$

Similarly, to count the boundary between Ω_1 and Ω_4 only once, add the constraint

$$\sup_{x \in \Omega} |q^1(x) - q^2(x)|_2 \leq \nu. \qquad (78)$$

The integral is maximized over $q^1, q^2 \in C_\nu$, (77) and (78) by picking $q^1 = -n_{\phi^1}$ in $J_{\phi^1} \backslash J_{\phi^1} \cap J_{\phi^2}$, $q^2 = -n_{\phi^2}$ in $J_{\phi^2} \backslash J_{\phi^1} \cap J_{\phi^2}$, $q^1 + q^2 = -n_{\phi^1}$ in J^+ and $q^1 - q^2 = -n_{\phi^1}$ in J^-. If one inserts these values for q^1 and q^2 one obtains

$$= \nu \int_{J^+} d\mathcal{H}^{N-1}(x) + \nu \int_{J^-} d\mathcal{H}^{N-1}(x)$$
$$+ \nu \int_{J_\phi^1 \backslash J_\phi^1 \cap J_\phi^2} d\mathcal{H}^{N-1}(x) + \nu \int_{J_\phi^2 \backslash J_\phi^1 \cap J_\phi^2} d\mathcal{H}^{N-1}(x),$$

which is the total length of each boundary.

Observe that for any $q^1, q^2 \in C_\nu$ which satisfy $|q^1 + q^2| = \nu$, it must hold that $|q^1 - q^2| \leq \nu$, for the following reason

$$|q^1 + q^2|_2^2 = \sum_{i=1}^N |q_i^1 + q_i^2|^2 = \sum_{i=1}^N (q_i^1)^2 + 2q_i^1 q_i^2 + (q_i^2)^2$$
$$|q^1 - q^2|_2^2 = \sum_{i=1}^N |q_i^1 - q_i^2|^2 = \sum_{i=1}^N (q_i^1)^2 - 2q_i^1 q_i^2 + (q_i^2)^2,$$

therefore

$$|q^1 + q^2|_2^2 + |q^1 - q^2|_2^2$$
$$= \sum_{i=1}^N (q_i^1)^2 + (q_i^2)^2 = |q^1|_2^2 + |q^2|_2^2 \leq 2\nu^2.$$

If $|q^1 + q^2|_2^2 = \nu^2$, this can only be true provided $|q^1 - q^2|_2^2 \leq \nu^2$. Vice versa, if $|q^1 - q^2|_2 = \nu$ then $|q^1 + q^2|_2 \leq \nu$. Therefore, the measure of the transition $\Omega_1 - \Omega_4$ is not influenced by adding the constraint (77) and vice versa for $\Omega_2 - \Omega_3$ regarding the constraint (78). By adding both constraints (77) and (78) to the optimization problem, we obtain the length/area of each boundary weighted by ν, i.e. Potts' regularization term (4). It is also possible derive different segmentation models by only adding some of the constraints.

Table 2 Number of variables and constraints in relaxation of Potts' model, N is the dimension of the image domain

	Primal variables	Primal constraints	Dual variables	Dual constraints
Pock et. al.	3	4	$3 * N$ (+4)	6 (+4)
Simplex [12]	4	5	$4 * N$ (+5)	4 (+4)
Proposed	2	2	$2 * N$ (+3)	4 (+3)

Each variable is regarded as a scalar function defined over the image domain. The numbers in parentheses indicate additional dual variables and constraints, in case the data term is represented in primal–dual form (which is used in the algorithms in the experiment section)

This can be an advantage if one knows in advance that certain regions are not supposed to share a common border. Examples will be given in the experiment section to illustrate. Non-standard regularization terms were also investigated in [9], but in the context of the simpler simplex constrained relaxation [10,12].

In order to derive a completely convex problem, the non-convex binary constraints $\phi^1, \phi^2 \in \mathcal{B}$ are replaced by $\phi^1, \phi^2 \in \mathcal{B}'$. In contrast to the model (43), the thresholding Theorem 2 is not generally valid if the additional constraints on the dual variables are introduced. However, if the computed solution ϕ^1, ϕ^2 is binary everywhere, it is also a global solution to the original problem. Otherwise, thresholding of ϕ^1, ϕ^2 will result in good approximate solutions. In the same manner, the relaxation for Potts' model [11] is not generally exact either, but will provide good approximate solutions after thresholding. To analyze theoretical relations between our relaxation and [11] is quite involved and is out of the scope of this paper. Some experimental comparisons are given in Sect. 8, which indicate our relaxation performs as least as well as [11]. Our approach involves significantly less primal and dual variables and constraints than [11] and is consequently easier to handle computationally. The number of primal and dual variables and constraints are summarized in Table 2.

7 Algorithms

Algorithms for the convex formulations and relaxations in Sects. 4 and 6 are presented based on the augmented Lagrangian method. In [17–19] 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 (43) and (76). Observe that the Lagrange multipliers ϕ^1, ϕ^2 are unconstrained in (60). However, by construction optimal ϕ^1, ϕ^2 will satisfy the relaxed binary constraints

Algorithm 1 Multiplier-Based Algorithm

Choose some initialization for p_s^i, p_t^i, q^i and ϕ^i, set $k = 1$ and repeat until convergence:

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

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

- Optimize p^{12} by fixing other variables

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

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

$$
\begin{aligned}
(q^1, q^2)^{k+1} := \arg \max_{(q^1, q^2) \in K} L_c(p_s^{k+1}, p_t^k, p^{12\,k}, q, \phi^k) . \\
:= \arg \max_{(q^1, q^2) \in K} \\
- \frac{c}{2} \sum_{i=1}^2 \left\| \operatorname{div} q^i - p_s^{i\,k+1} + p_t^{i\,k} + (-1)^{i+1} p^{12\,k+1} - \phi^{i\,k}/c \right\|^2,
\end{aligned}
\tag{79}
$$

for $i = 1, 2$. The above problem can either be solved iteratively by Chambolle's projection algorithm [49], or approximately in one step through

$$
\tilde{q}^i = \left(q^{i\,k} + c \nabla (\operatorname{div} q^i - p_s^{i\,k+1} + p_t^{i\,k} + (-1)^{i+1} p^{12\,k+1} - \phi^{i\,k}/c) \right),
$$

$$
(q^1, q^2)^{k+1} = \Pi_{(q^1, q^2) \in K}(\tilde{q}^1, \tilde{q}^2),
\tag{80}
$$

where $\Pi_{(q^1, q^2) \in K}$ is the projection operator onto K.

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

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

- Update multipliers ϕ^i, $i = 1, 2$, by

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

- If not converged, set $k = k + 1$ and repeat.

---

(11). In this section it is assumed that Ω, the unknowns p_s^i, p_t^i, p^{12}, q^i, ϕ^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

$$
\begin{aligned}
L(p_s, p_t, p^{12}, q, \phi) = \int_\Omega p_s^1(x) + p_s^2(x) \, dx \\
+ \sum_{i=1}^2 \int_\Omega \{\phi^i (\operatorname{div} q^i - p_s^i + p_t^i + (-1)^{i+1} p^{12})\}(x) \, dx \\
- \frac{c}{2} \sum_{i=1}^2 || \operatorname{div} q^i - p_s^i + p_t^i + (-1)^{i+1} p^{12}||^2.
\end{aligned}
\tag{81}
$$

An algorithm for optimizing (60) is constructed based on the alternating direction method of multipliers [50], see Alg. 6.2. The set K in (79) consists of the unit disks

$$
\begin{aligned}
K = \{(q^1, q^2) : \ \Omega \mapsto \mathbb{R}^N \times \mathbb{R}^N : \\
||q^i||_\infty \le \nu, \ \ q_n^i|_{\partial\Omega} = 0 \ \ i = 1, 2\}.
\end{aligned}
$$

Here $||q||_\infty = \max_{x \in \Omega} |q(x)|_2$. Note that the optimization problems in (79) and (80) decouple and can be solved separately for q^1 and q^2. The optimization for p_s^i, p^{12} and p_t^i can be easily computed in closed form pointwise at each $x \in \Omega$.

7.1 Algorithm for Convex Relaxed Potts' model

The algorithm for the convex relaxed Potts' model does not change, except the variables q^1 and q^2 in step (79) are optimized over the set

$$
\begin{aligned}
K = \{(q^1, q^2) : \ \Omega \mapsto \mathbb{R}^N \times \mathbb{R}^N : \\
||q^i||_\infty \le \nu, \ \ q_n^i|_{\partial\Omega} = 0, \ \ i = 1, 2, \\
||q^1 + q^2||_\infty \le \nu, \ \ ||q^1 - q^2||_\infty \le \nu\}.
\end{aligned}
$$

There are no closed form solution for such a projection, but it can be computed approximately by a few iterations of Dykstra's algorithm [51] in the same way as [11].

8 Experiments

In most experiments, we choose the data term (2) and set $\beta = 2$ or $\beta = 1$. 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. $\{\phi_i^k\}_{i=1}^4 = \underset{\{\phi^i\}_{i=1}^4}{\arg\min} E(\{\phi^i\}_{i=1}^4, \{c_i^{k-1}\}_{i=1}^4),$

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

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 [52]. In all experiments reported, the condition which guarantee submodularity/convexity was satisfied during each iteration of the algorithm unless otherwise specifed. We have used a mimetic spatial discretization [42] of the differential operators in Algorithm 1.

Abbreviations: We have used the following abbreviations in the comparison experiments:

 GT: ground truth
 CVl: Chan–Vese model (13) solved with level set method
 CVg: Chan–Vese model (13) solved by graph cut
 CVC: convex Chan–Vese model before thresholding
 CVCt: convex Chan–Vese model after thresholding
 AE: Alpha expansion algorithm [6]
 ABS: Alpha–beta swap algorithm [6]
 SR: Simplex relaxation [12]
 SRt: Simplex relaxation [12] after thresholding
 PR: Pock relaxation [11]
 PRt: Pock relaxation [11] after thresholding
 NR: New relaxation of Potts' model
 NRt: New relaxation of Potts' model after thresholding

On the relatively simple image in Fig. 5, the level set method finds a good local minimum. If the initialization is bad, the level set method gets stuck in an inferior local minimum also for this simple image as shown in Fig. 6, courtesy of [3]. White points indicate the zero level set of ϕ^1 and dark points indicate the zero level set of ϕ^2 (Fig. 6).

More difficult images are presented in Figs. 7, 8, 9, 10, 11. 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. We depict the results of the relaxed problems ϕ^1, ϕ^2 before thresholding in one single image I as

$$I = \phi^1 \phi^2 c_2 + \phi^1(1 - \phi^2)c_1 + (1 - \phi^1)\phi^2 c_4$$
$$+ (1 - \phi^1)(1 - \phi^2)c_3$$

In the same way, the results of relaxations [12,53] before threshold are depicted in a single image using the integrand of the data terms of their convex energy functionals. As seen in the figures, the convex Chan–Vese model tends to favor
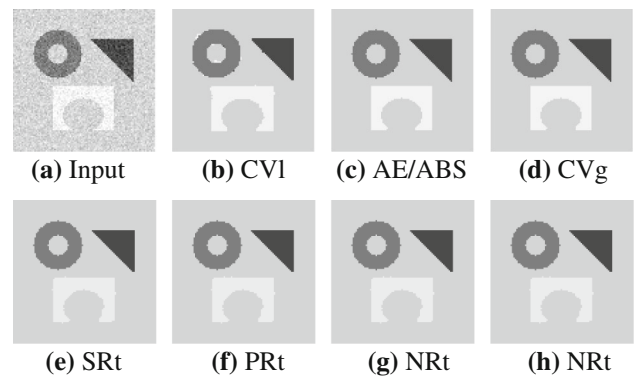

Fig. 5 L^2 data fidelity. **a** Input, **b** level set method gradient descent, **c** alpha expansion/alpha beta swap, **d** Chan–Vese model graph cut eight neighbors (global minimum). *2nd row* Convex relaxations after threshold: **e** simplex, **f** Pock et al. **g** proposed Potts', **h** proposed special
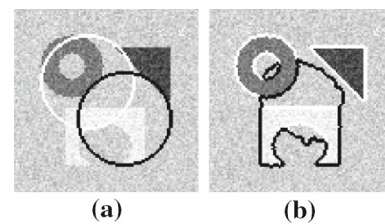

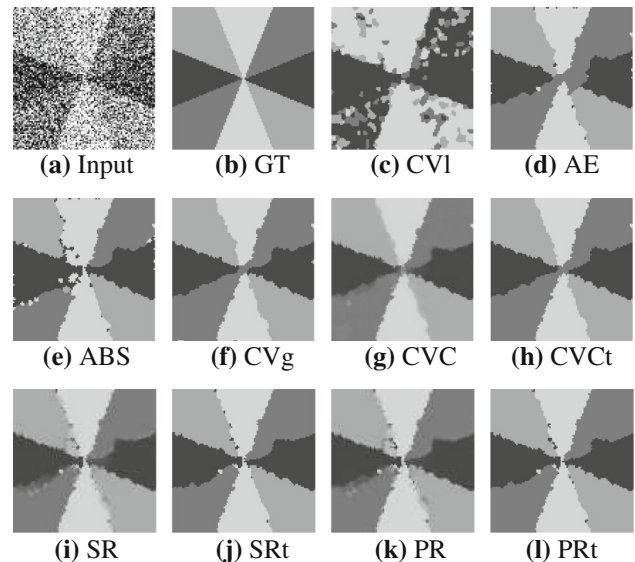**Fig. 6** Level set method: **a** bad initialization, **b** result

Fig. 7 L_1 data term, $\nu = 0.245$. **a** Input image, **b** ground truth, **c** Chan–Vese model level set method, **d** alpha expansion Potts' model, **e** alpha–beta swap alpha expansion Potts' model, **f** graph cut for Chan–Vese model (global minimum), **g** convex Chan–Vese model before threshold, **h** convex Chan–Vese model after threshold, **i** simplex relaxation [12], **j** simplex relaxation [12] after threshold, **k** relaxation [53], **l** relaxation [53] after threshold

solutions that are closer to binary than the other relaxations [12,53] for Pott' model on difficult images with high levels of noise. The percentage of misclassified pixels compared to
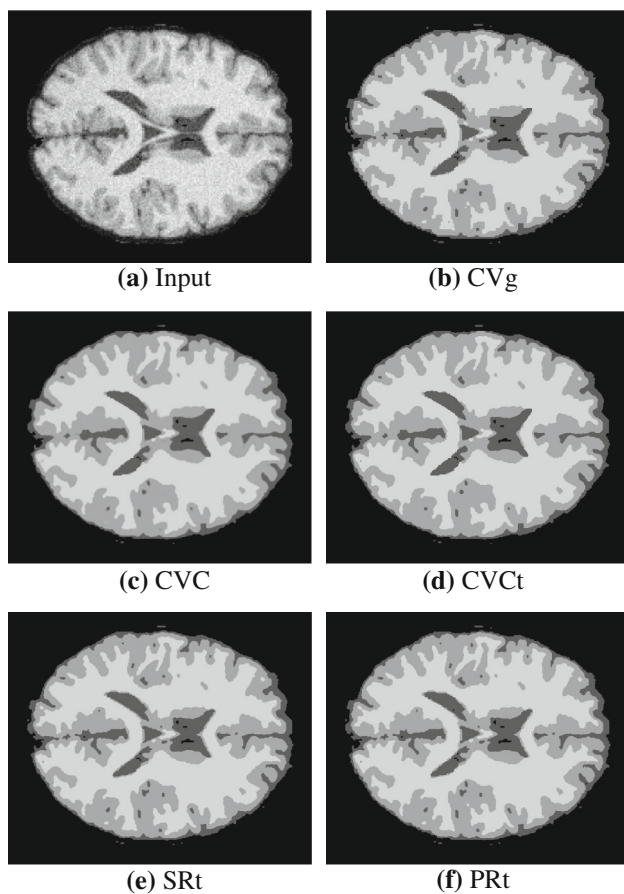
(a) Input   **(b) CVg**

(c) CVC   **(d) CVCt**

(e) SRt   **(f) PRt**

Fig. 8 L_2 data fidelity: **a** Input; **b** graph cut for discrete Chan–Vese model four neighbors (global minimum); **c** convex Chan–Vese model before threshold; **d** convex Chan–Vese model after threshold; **e** special relaxation after threshold; **f** Potts relaxation after threshold

(a)   **(b)**

(c)   **(d)**

Fig. 9 Chan–Vese model with L^2 data term: **a** Input, **b** graph cut approach 4 neighbors, **c** convex formulation before threshold, **d** convex formulation after threshold

(a) Input   **(b) CVg**

(c) CVC   **(d) CVCt**

Fig. 10 Chan–Vese model with L^2 data term. **a** Input, **b** graph cut eight neighbors in discrete setting, **c** convex formulation before threshold, **d** convex formulation after threshold

ground truth, depicted in subfigures b, are shown in Table 3 and indicate that our approach performs favorably. Note that it is not so relevant to compare energies, since the regularizer in the Chan–Vese model is slightly heavier for some boundaries than the Potts' model and therefore has a higher energy value.

In subfigure 8, 9, and 10b the computed global minimum of the discrete energy is shown. In subfigures 8, 9 and 10c the computed global minimizer 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. Observe that the continuous version is rotationally invariant and, in contrast to the discrete approach, produces results that are not biased by the discrete grid.

In Fig. 12, the problem (13) with non-convex data term (14) is solved simultaneously for ϕ^1 and ϕ^2 by the gradient descent method. For simplicity, we have removed the noise and set the regularization parameter to zero. Both initializations lead to incorrect results. The recent inexact convex
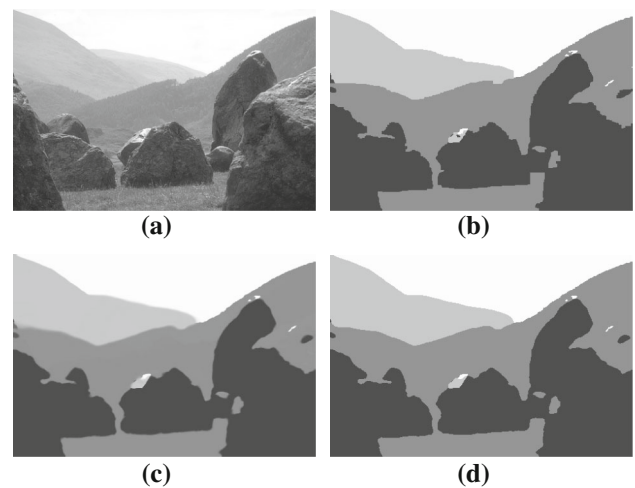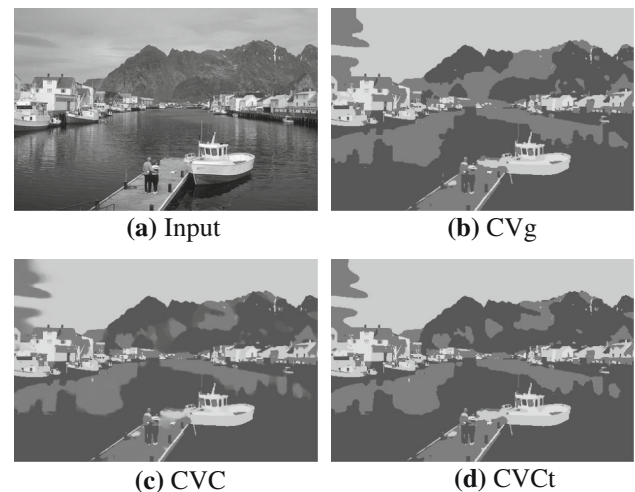
relaxation for the Chan–Vese model [8] was demonstrated to almost find a global minimum on Fig. 5 (except for a few pixels), but differed more substantially on more difficult images like the brain image. Our approach finds exact solutions and is more computationally efficient, since [8] involves more unknown variables in a higher dimensional space.

8.1 Experiments on L_2 data Fitting Term: Submodularity

In Sect. 5 we gave theoretical insights on how submodularity of the energy function was related to the distribution of the values c_i, $i = 1, \ldots, 4$. It was shown that the condition becomes less strict as β increases. In this section, the

(a) Input **(b) GT** **(c) CVl**

(d) AE **(e) ABS** **(f) CVg**

(g) SR **(h) PR** **(i) CVC**
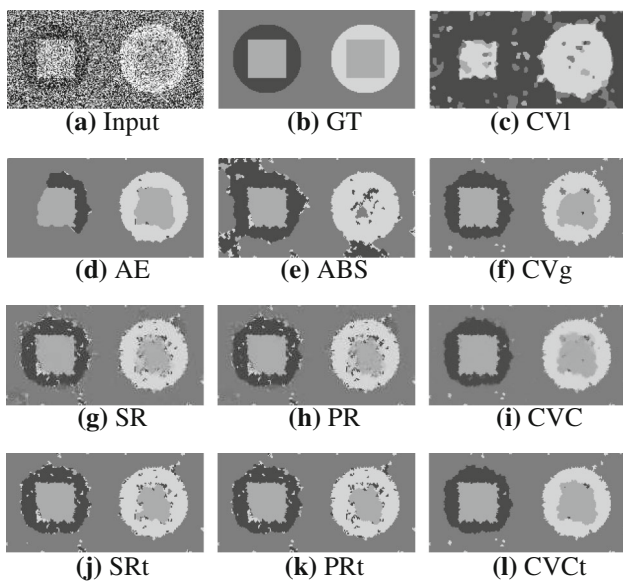
(j) SRt **(k) PRt** **(l) CVCt**

Fig. 11 L_1 data term and $\nu = 0.19$. **a** Input image, **b** ground truth, **c** Chan–Vese model level set method. *2nd row* discrete algorithms eight neighbors, **d** alpha expansion Potts' model, **e** alpha–beta swap Potts' model, **f** discrete Chan–Vese model proposed (global minimum). *3rd row* convex relaxations before threshold: **g** simplex relaxation [12], **h** relaxation [53] **i** convex Chan–Vese model. *4th row* Convex relaxations after threshold: **j** simplex relaxation [12], **k** relaxation [53], **l** convex Chan–Vese model
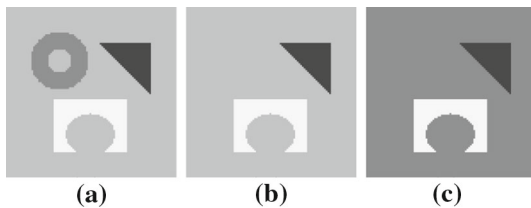

(a) **(b)** **(c)**

Fig. 12 Coupled gradient descent minimization of non-convex energy (13) over constraint $\phi^1(x), \phi^2(x) \in [0, 1] \ \forall x \in \Omega$ and regularization $\nu = 0$: **a** Input image, **b** result initialization $\phi^1, \phi^2 = 0$, **c** result initialization $\phi^1, \phi^2 = 1$

Table 3 Percentage of misclassified pixels compared to ground truth for experiments in Figs. 7 and 11

	α-exp	α-β-swap	Simplex relaxation	Pock [11] relaxation	Convex Chan–Vese
Figure 11	21.7	9.75	8.45	8.04	4.54
Figure 7	5.24	9.30	4.19	4.11	3.57

condition is analyzed empirically and experimentally for the L_2 data fitting term ($\beta = 2$ in (2)).

In Sect. 5 the constant values $c_1, \ldots, c_4$ were sampled randomly in the interval $[0, 1]$. The condition (5) was satisfied for all $I \in [0, 1]$ in 39.6 percent out of the 10000 random selections of $c_1, \ldots, c_4$. In cases that the condition was violated, the constants were clustered asymmetrically around their mean. However, such distributions of the constants are
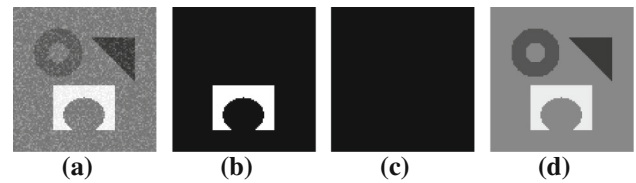

(a) **(b)** **(c)** **(d)**

Fig. 13 Chan Vese model L^1 data fidelity. Note that the constant values c_1, c_2, c_3 are close to each other compared to c_4: $c_1 = 36$, $c_2 = 60$, $c_3 = 110$, $c_4 = 230$. **a** Input image, **b** set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodularity condition was violated, **c** set of pixels where the residual capacity criterion (40) is not satisfied (empty set), **d** output image (global minimum)

not expected in practice, because when minimized over the region parameters, the model (2) will favor solutions where the data functions (2) corresponding to each region are as dissimilar as possible. We have run the alternating optimization algorithm described at the beginning of this section for optimizing the parameters c_i, $i = 1, \ldots, 4$ on all 100 images from the image segmentation data base [45]. In all experiments, the submodularity condition (45) was satisfied during each iteration of the algorithm.

8.2 Non-submodular Data Terms

The purpose of this section is to demonstrate the relaxation approaches from Sect. 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 Fig. 13, which is a modified version of the example in Fig. 5, where the average intensity values of 3 of the objects are close compared to the 4th object. Some more natural examples are shown in Fig. 14. Subfigures (b) show the set of pixels $p \in \mathcal{P}_1 \cup \mathcal{P}_2$, where the submodularity condition was violated. In all experiments $\beta = 1$. Subfigures (c) show the set of pixels where the residual capacity conditions (40) ((69) and (70) 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. If the regularization parameter ν is set to a very large number, the residual capacity condition may be more easily violated. An example is shown in Fig. 14g, h. The set of pixels where it is violated, shown in (g), is small and constitutes 0.2 percent of the pixels in the image. It cannot be concluded whether the solution of the relaxed problem, shown in (h), is also a solution of the original problems, but is in any case a good approximation.

These examples are typical: if the regularization parameter ν is not set extremely high, the residual capacity condition tends to be satisfied. In cases the condition is violated, it only happens for a very small set of the pixels, which means the
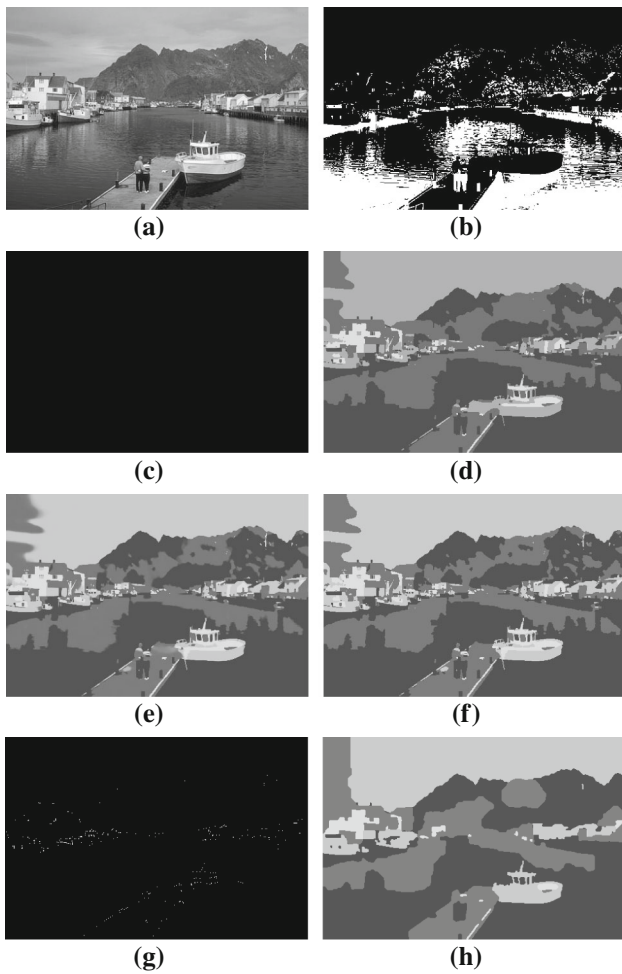
(a) **(b)**

(c) **(d)**

(e) **(f)**

(g) **(h)**

Fig. 14 Chan–Vese model with L^1 data term: **a** input, **b** set of points where submodularity condition (64) was violated, **c** set of points where the residual capacity conditions ((40) for graph cut and (69), (70) for convex model) were not satisfied (empty set), **d** graph cut method eight neighbors (global minimum), **e** convex formulation before threshold, **f** convex formulation after threshold. **g–h** High regularization parameter v: **g** set of points where residual capacity condition (69) was violated (0.2 % of pixels), **h** resulting partition (approximate solution)

results are good approximations. To save space, only a subset of our experiments are shown.

8.3 Experiments with Other Data Terms

In this section, we demonstrate the global minimization methods on some other data terms. The data term in the Chan–Vese model (2) can be derived via the maximum a posteriori estimate

$$\Pi_{i=1}^4 \Pi_{x \in \Omega_i} \frac{1}{2\pi\sigma_i} \exp\left(\frac{(I(x)-c_i)^2}{2\sigma_i}\right) \tag{82}$$

The above expression can be maximized by minimizing its negative logarithm. By adding regularization, one ends up
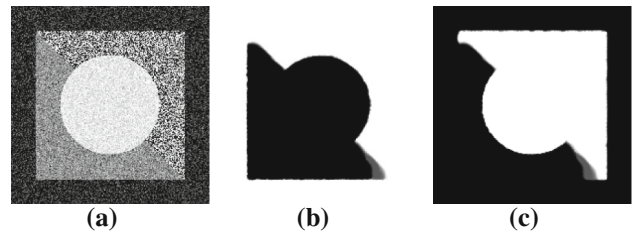

(a) **(b)** **(c)**

Fig. 15 **a** Input image: the *square* in the middle is divided into two regions of the same mean, but different variances. Segmentation with data term involving mean and variance (83). **b** λ_1, **c** λ_2 (white: $\lambda_i = 1$, dark: $\lambda_i = 0$)
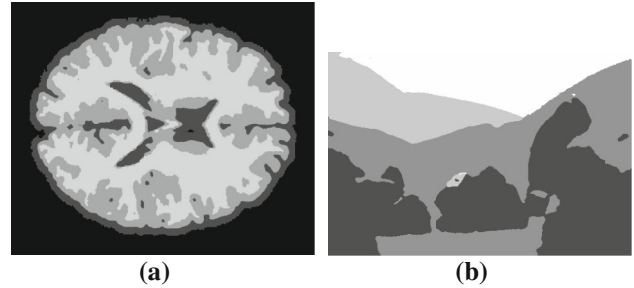

(a) **(b)**

Fig. 16 Segmentation with data term (83). The optimal mean and variance were estimated by the altenating algorithm. **a** Result Fig. 8a, $c = (0.0078, 0.2158, 0.5808, 0.7935)$, $\sigma = (0.0079, 0.1152, 0.0779, 0.0551)$. **b** Result Fig. 9a, $c = (0.2188, 0.4909, 0.7390, 0.9950)$, $\sigma = (0.0563, 0.0783, 0.0578, 0.0066)$

with the the general model (1) with data term

$$f_i(x) = \frac{(I(x)-c_i)^2}{2\sigma_i} - \log(2\pi\sigma_i). \tag{83}$$

The Chan–Vese model involves minimizing (1) with data term (83) with respect to $\{\Omega_i\}_{i=1}^4$ and $\{c_i\}_{i=1}^4$ while keeping the standard deviations $\sigma_i = 1/2$ fixed. In the more advanced model, (83) is minimized with respect to all variables $\{\Omega_i\}_{i=1}^4$, $\{c_i\}_{i=1}^4$ and $\{\sigma_i\}_{i=1}^4$. The minimization problem can be solved numerically by a modification of the alternating algorithm described in the beginning of this section, by updating both c_i and σ_i for the given Ω_i in step 2. each iteration. An illustrative example is given in Figure 15. The square in the middle is divided diagonally into regions of the same mean, but different standard deviations. The more advanced model is able to separate these regions based on the differing standard deviations as shown in (b) and (c). The submodularity condition (44) was satisfied for all points in this example. Results on the other more natural images are shown in Fig. 16. Also in this case the submodularity condition was satisfied during each iteration of the alternating algorithm. The final parameters c and σ are shown in the caption of the figure.

We have repeated the experiments on the 100 images in the segmentation data base [45], by iteratively optimizing in

Table 4 Relaxations for Potts' model, final energy after binarization

	Simplex relaxation	Relaxation [11]	New relaxation
Figure 8	$1.8938 * 10^3$	$1.8897 * 10^3$	$1.8897 * 10^3$
Figure 5	60.9827	60.9686	60.9686

Fig. 17 a Input, all data terms are set to zero in the dark region. **b** Result of relaxation [10,12] and [7] and Chan–Vese model (13). **c–d** Convex relaxation (76) over (10), $\phi^1, \phi^2 \in \mathcal{B}'$ and extra dual constraint (77) (**c**) and (78) (**d**)

terms of regions and updating c, σ. The submodularity condition (44) was satisfied in 65 % of the experiments. In the remaining experiments, we can check whether the condition on the residual capacities (69) and (70) are satisfied. As mentioned, this depends on the strength of the regularization. In our case we used the moderate value of $\nu = 0.12$ and the condition held in all cases. For high values of ν this cannot be expected in general as discussed in Sect. 8.2, in which case the solutions are approximate.

8.4 Convex Relaxation of Other Regularization Terms

The convex relaxation of the Potts' model and other regularization terms from Sect. 6.2 are demonstrated and compared with other recent relaxations. Table 4 depicts the energy of the final binary results after thresholding, i.e. lower energy means better approximation of the global minimum. The curve lengths on the discrete grid have been calculated by the Cauchy–Crofton formula of integral geometry [33] with an 8 neighborhood system. The images are scaled between 0 and 1, $\nu = 0.05$ in Fig. 5 and $\nu = 0.026$ in Fig. 8. As we see, our approach performs on par with [11] and better than the simplex relaxation [12].

In some applications, one know in advance that certain regions are not supposed to share a common border. In brain MRI imaging, for example, the white matter should only border with gray matter and cerebrospinal fluid, but not the background. Therefore, the transition $\Omega_1 - \Omega_4$ can favorably be given a higher penalty in the regularization term. As a new segmentation model for brain imaging, we therefore propose to optimize (76) over $\phi^1, \phi^2 \in \mathcal{B}'$, $q^1, q^2 \in C_\nu$ and the extra dual constraint (77). Finally, a binary (approximate) solution is obtained by thresholding ϕ^1, ϕ^2 as in Theorem 2. The result is depicted in Fig. 8e. In Fig. 5, region 1 and 4 are not supposed to share a common border, therefore the same constraint set also makes sense in this application.

The special regularization term can also be used to enforce uniqueness in applications where the Potts' model suffer from multiple global minima. In the inpainting experiment depicted in Fig. 17, one would like to fill in the middle dark area in (a). The partition problem is solved with data term set to zero in the dark area and the maximum double precision value elsewhere. With Potts' regularizer, this problem has two solutions of equal energy, shown in Fig. 17 c, d respectively. Therefore, a convex relaxation for Potts' model can
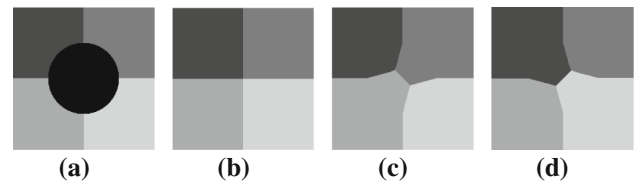
at best obtain a convex combination of these two solutions, which is certainly not binary. By instead using the relaxation from Sect. 6.2 with either constraint (77) or (78), we are able to obtain the correct two solutions of minimal curve lengths. It was reported in [11] that one of the solutions could be obtained on a similar example via their Potts relaxation. However, this is due to asymmetry of their underlying discretized problem, which results in a bias towards one solution over the other.

8.5 Analysis of Runtime and Efficiency

A major advantage of the proposed algorithms is the simple, compact and convex representation of the partition problem which results in a significant speedup over state of the art approaches. Table 5 summarizes the number of iterations and flops (floating point operations) of our algorithms compared to two other convex relaxations. We have calculated the number of flops manually, by direct counting in the code taking into account the number of pixels, in order to avoid operating system and compiler specific uncertainties.

The tight convex relaxation [11] was implemented with the primal–dual algorithm of [11] and the simplex constrained relaxation [12] was implemented with the algorithm of [54], which is the fastest for this problem in our experience. As stopping criteria, we estimate the minimal energy E^* using a huge amount of iterations and determine the iteration k when the relative energy precision $\frac{E^k - E^*}{E^*}$ falls below 10^{-3}.

The convex formulation of the Chan–Vese model is the fastest and significantly outperforms relaxations [11,12,54] in terms of total number of flops and iterations. The relaxation for Potts' model, the special relaxation and [11] are bottle necked by Dykstra's algorithm for projecting dual variables onto the feasible set every iteration. We have used five iterations of Dykstra's algorithm in all these relaxations, as we found that to be the most optimal balance between efficiency and accuracy. This is reflected in a larger number of flops each iteration in the table. Our relaxations involve the least number of variables and dual constraints and are consequently faster than [11]. Overall, our algorithm runs on par with the algorithm [54] for the simpler simplex constrained relaxation [12].

Table 5 Number of iterations k, number of flops per iteration and total number of flops to reach energy precision $\frac{E^k - E^*}{E^*} < 10^{-3}$

	Figure 5			Figure 8		
	Iterations	Flops/iteration	Total flops	Iterations	Flops/iteration	Total flops
Chan–Vese convexification	11	$9.94 * 10^5$	$1.09 * 10^7$	60	$4.20 * 10^6$	$2.52 * 10^8$
Special relaxation	12	$3.92 * 10^6$	$4.70 * 10^7$	60	$1.65 * 10^7$	$9.93 * 10^8$
Potts relaxation	12	$5.07 * 10^6$	$6.09 * 10^7$	60	$2.14 * 10^7$	$1.29 * 10^9$
Simplex relaxation [54]	60	$1.96 * 10^6$	$1.18 * 10^8$	190	$8.21 * 10^6$	$1.56 * 10^9$
Relaxation [11]	295	$1.12 * 10^7$	$3.31 * 10^9$	1020	$4.71 * 10^7$	$4.79 * 10^{10}$

The discrete graph cut method in Sect. 3 is also very efficient, particularly due to the relatively low number of nodes and edges. Comparatively, alpha expansion is an iterative algorithm, which solves a sequence of graph cut problems until convergence. Our algorithm only needs to solve one graph cut problem and consequently converges around 5–7 times faster than alpha expansion in the experiments. Overall, the graph cut method was slightly faster than our matlab implementation of the convex algorithms. However, the convex algorithms consist mainly of floating point matrix and vector operations, which are much better suited for parallel implementation on GPU. We expect a GPU implementation of our algorithms to beat the graph cut method in terms of speed.

9 Conclusions

We have presented an exact global optimization framework for certain image segmentation models involving four regions, such as the Chan–Vese model, both in a discrete setting and a variational setting. If a condition on the data term was satisfied, a global minimum was guaranteed. A theoretical analysis of the condition was given and it was shown experimentally that the condition tends to hold for L^p type of data terms for $p \geq 2$. It also often holds for statistical data terms taking into account the means and variances the regions. If the condition was violated, relaxations were proposed for producing either exact or approximate solutions. Conditions on the "residual capacities" of the computed solution could be checked to verify whether a global minimum of the original problem had been obtained. Experiments showed that these relaxations could produce global minima in practice provided the strength of spatial regularization was not too high. A new convex relaxation of Potts' regularization term and some other regularization terms were also proposed. Experiments demonstrated results on par with [11] for Potts' model energy wise. Algorithms were proposed for the new energy minimization problems. Experiments demonstrated a significant speed up over alternative convex relaxations, which is mainly explained by the com-

pactness and simplicity of our convex minimization problems.

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 binary 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, we expect the conditions which guarantee submodularity to be more strict as m increases, therefore it will be valueable to derive relaxations as in Sect. 6. However, the case of four regions is important and deserves special attention: It is possible to obtain exact binary solutions, both in a discrete and continuous setting; Important practical problems, such as brain MRI segmentation involve four distinct regions; 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 four disconnected regions, where different data cost functions are assigned to each disconnected component. Some developments have already been made in this direction recently [20–22].

Appendix 1: A Proofs of Prop 2–5

Appendix 1.1: Proof of Prop 4

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

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

$\square$

Appendix 1.2: Proof of Prop 2

When $\frac{c_1+c_2}{2} \leq I \leq \frac{c_4+c_3}{2}$, the result follows from Prop 4. 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 (64), 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 $\beta \geq \beta_0$

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

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

$$|I - c_2|^\beta + |I - c_3|^\beta = \theta_2^{\beta-\beta_0} |I - c_1|^\beta + \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta$$

$$\leq \theta_1^{\beta-\beta_0} |I - c_1|^\beta + \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta \leq \theta_1^{\beta-\beta_0} |I - c_1|^\beta$$

$$+ \frac{1}{\theta_1^{\beta-\beta_0}} |I - c_4|^\beta,$$

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 2 when $I > \frac{c_4+c_3}{2}$.

Appendix 1.3: Proof of Prop 3

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 4, e.g. $\mathcal{C} = 1$.

Therefore the condition (64) is satisfied for any $I \in [0, L]$ by choosing $\beta \geq \mathcal{C} = \max_{I \in [0,L]} \max\{\mathcal{C}_I^1, \mathcal{C}_I^2\}$.   □

Appendix 1.4 Proof of Prop 5

We will show the condition holds for $\beta = 1$, which by Prop 2 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 $I \geq c_3$

$$|I - c_2| + |I - c_3| = (I - c_2) + (I - c_3)$$
$$= 2I - (c_2 + c_3)$$
$$= 2I - (c_1 + c_4) = (I - c_1) + (I - c_4)$$
$$\leq |I - c_1| + |I - c_4|.$$

Appendix 2: Relation Between Dual Problem (63) and Discrete Max-flow Problem

Recall that for each $p \in \mathcal{P}$ the data term was represented by the six edges (31). The flow on each of these edges are constrained by the capacities defined in (35) and (34), i.e.

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

For notational convenience, we use capital letters P, Q to denote flow on 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

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

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 (51).

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

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

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 (52). 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.

Flow conservation should hold at each $v_{p,1}$ and $v_{p,2}$, i.e.

$$\sum_{q \in \mathcal{N}_p^k} Q^1(v_{p,1}, v_{q,1}) - P_s^1(p) + P_t^1(p) - \tilde{P}^{12}(p) + \tilde{P}^{21}(p)$$
$$= 0,$$
$$\sum_{q \in \mathcal{N}_p^k} Q^2(v_{p,2}, v_{q,2}) - P_s^2(p) + P_t^2(p) + \tilde{P}^{12}(p) - \tilde{P}^{21}(p)$$
$$= 0,$$

$\forall p \in \mathcal{P}$. Here $N^+(v)$ 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 $(w, v) \in \mathcal{E}$. The final two constraints (61) and (62) are generalizations of discrete flow conservation to the continuous setting. The total amount of discrete flow in the graph is in this case

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

The objective functional (63) is a generalization of the above, and measures the total amount of flow that originates from the source.

References

1. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. **42**, 577–685 (1989)
2. Chan, T., Vese, L.A.: Active contours without edges. IEEE Image Proc. **10**, 266–277 (2001)
3. Vese, L.A., Chan, T.F.: A new multiphase level set framework for image segmentation via the Mumford and Shah model. Int. J. Comput. Vis. **50**, 271–293 (2002)
4. Boykov, Yuri, Kolmogorov, Vladimir: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 359–374 (2001)
5. Greig, D.M., Porteous, B.T., Seheult, A.H.: Exact maximum a posteriori estimation for binary images. J. R. Stat. Soc. Ser. B **51**, 271–279 (1989)
6. Boykov, Yuri, Veksler, Olga, Zabih, Ramin: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1222–1239 (2001)
7. Bae, Egil, Yuan, Jing, Tai, Xue-Cheng: Global minimization for continuous multiphase partitioning problems using a dual approach. Int. J. Comput. Vis. **92**(1), 112–129 (2011)
8. Brown, E.S., Chan, T.F., Bresson, X.: 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
9. Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: IEEE International Conference on Computer Vision (ICCV), pp. 646–653 (2009)
10. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: Tai, X.-C., Mörken, K., Lysaker, M., Lie, K.-A. (eds.) Scale Space and Variational Methods in Computer Vision (SSVM 2009), volume 5567 of LNCS, pp. 150–162. Springer, New York (2009)
11. Pock, T., Chambolle, A., Bischof, H., Cremers, D.: A convex relaxation approach for computing minimal partitions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, Florida
12. Zach, C., Gallup, D., Frahm, J.-M., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: Vision, Modeling and Visualization Workshop (VMV) (2008)
13. Goldluecke, B., Cremers, D.: Convex relaxation for multilabel problems with product label spaces. In: Proceedings of the 11th European conference on Computer vision, ECCV'10, pp. 225–238. Springer, Berlin, Heidelberg (2010).
14. Strekalovskiy, E., Goldluecke, B., Cremers, D.: Tight convex relaxations for vector-valued labeling problems. In: IEEE International Conference on Computer Vision (ICCV) (2011)
15. Ishikawa, Hiroshi: Exact optimization for markov random fields with convex priors. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1333–1336 (2003)
16. Chan, T.F., Esedoğlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math., **66**(5):1632–1648 (electronic), (2006)
17. Bae, E., Yuan, J., Tai, X.C., Boykov, Y.: A fast continuous max-flow approach to non-convex multilabeling problems. Technical report CAM10-62, UCLA, CAM (2010)
18. Yuan, J., Bae, E., Tai, X.C.: A study on continuous max-flow and min-cut approaches. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2217–2224. San Francisco (2010)
19. Yuan, Jing, Bae, Egil, Tai, Xue-Cheng, Boykov, Yuri: A spatially continuous max-flow and min-cut framework for binary labeling problems. Numerische Mathematik **126**(3), 559–587 (2014)
20. Liu, Liman, Tao, Wenbing: Image segmentation by iteratively optimization of multiphase multiple piecewise constant model and four-color relabeling. Pattern Recognit. **44**(12), 2819–2833 (2011)
21. Zhang, R., Bresson, X., Chan, T.F., Tai, X.-C.: Four color theorem and convex relaxation for image segmentation with any number of regions. UCLA, Applied Mathematics, CAM-report-13-16 (2013)
22. Hodneland, Erlend, Tai, Xue-Cheng, Gerdes, Hans-Hermann: Four-color theorem and level set methods for watershed segmentation. Int. J. Comput. Vis. **82**(3), 264–283 (2009)
23. Bae, E., Tai, X.-C.: Efficient global minimization for the multiphase Chan–Vese model of image segmentation. In: Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition. LNCS, vol. 5681, pp. 28–41. Springer, New York (2009)
24. Bae, E.: Efficient global minimization methods for variational problems in imaging and vision. (Phd thesis, University of Bergen, https://bora.uib.no/handle/1956/5017), August 2011
25. Bae, Egil, Tai, Xue-Cheng: Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In: Tai, X.-C., Mörken, K., Lysaker, M., Lie, K.-A. (eds.) Scale Space and Variational Methods in Computer Vision, Volume 5567 of LNCS, pp. 1–13. Springer, New York (2009)
26. Brown, E.S., Chan, T.F., Bresson, X.: Completely convex formulation of the Chan-Vese image segmentation model. Inter-

national Journal of Computer Vision. (2011) doi:10.1007/s11263-011-0499-y

27. Lie, J., Lysaker, M., Tai, X.C.: Piecewise constant level set methods and image segmentation. In: Scale Space and PDE Methods in Computer Vision, pp. 573–584. Springer, New York (2005)

28. Lie, J., Lysaker, M., Tai, X.C.: A binary level set model and some applications to Mumford–Shah image segmentation. IEEE Trans. Image Process. **15**(5), 1171–1181 (2006)

29. Meyer, Y.: 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

30. Strang, Gil: Maximal flow through a domain. Math. Program. **26**, 123–143 (1983)

31. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: ECCV '08, pp. 792–805 (2008)

32. Bae, E., Lellmann, J., Tai, X.-C.: Convex relaxations for a generalized chan-vese model. In: 9th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition. Lecture Notes in Computer Science, vol. 8081, pp. 223–236. Springer, New York (2013)

33. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, p. 26. IEEE Computer Society, Washington, DC (2003).

34. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Readings in Uncertain Reasoning, pp. 452–472. Morgan Kaufmann Inc., San Francisco (1990)

35. Fulkerson, D., Ford, L.: Flows in Networks. Princeton University Press, Princeton (1962)

36. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 147–159 (2004)

37. Kolmogorov, Vladimir, Rother, Carsten: Minimizing nonsubmodular functions with graph cuts: a review. IEEE Trans. Pattern Anal. Mach. Intell. **29**(7), 1274–1279 (2007)

38. Hammer, P.L., Hansen, P., Simeone, B.: Roof duality, complementation and persistency in quadratic 01 optimization. Math. Program. **28**, 121–155 (1984)

39. Rother, C., Kumar, S., Kolmogorov, V., Blake, A.: Digital tapestry. In: IEEE Proceedings of the Computer Vision and Pattern Recognition (2005)

40. El-Zehiry, N.Y., Grady, L.: Combinatorial optimization of the discretized multiphase Mumford-Shah functional. Int. J. Comput. Vis. (2013) doi:10.1007/s11263-013-0617-0

41. Berkels, B.: An unconstrained multiphase thresholding approach for image segmentation. In: SSVM, pp. 26–37 (2009)

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

43. Ito, Kazufumi, Kunisch, Karl: Lagrange Multiplier Approach to Variational Problems and Applications. Society for Industrial and Applied Mathematics, Philadelphia (2008)

44. Ekeland, Ivar, Téman, Roger: Convex Analysis and Variational Problems. Society for Industrial and Applied Mathematics, Philadelphia (1999)

45. http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/

46. Chambolle, Antonin, Cremers, Daniel, Pock, Thomas: A convex approach to minimal partitions. SIAM J. Imaging Sci. **5**(4), 11131158 (2012)

47. Federer, H.: Geometric Measure Theory. Springer, New York (1969)

48. Vol'Pert, A.I.: Spaces BV and quasilinear equations. Mat. Sb. (N.S.) **73**(115), 255302 (1967)

49. Chambolle, Antonin: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**(1), 89–97 (2004)

50. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)

51. Boyle, J.P., Dykstra, R.: A method for finding projections onto the intersection of convex sets in hilbert spaces. In: Advances in Order Restricted Statistical Inference (Iowa City, Iowa, 1985). Lecture Notes in Statistics, vol. 37, p. 2847. Springer, Berlin, 1986 (1985)

52. Hodneland, E.: Segmentation of digital images. Cand. Scient Thesis, Department of Mathematics, University of Bergen. Available online at www.mi.uib.no/tai (2003)

53. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: IEEE International Conference on Computer Vision (ICCV). Kyoto (2009)

54. Yuan, J., Bae, E., Tai, X.-C., Boykov, Yuri: A continuous max-flow approach to potts model. In: European Conference on Computer Vision. LNCS, vol. 6316, pp. 379–392 (2010)
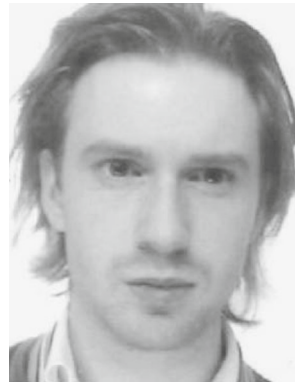
Egil Bae received the B.S. degree in mathematics in 2005 and the M.S. and Ph.D. degrees in applied mathematics in 2007 and 2011 from the University of Bergen in Norway. He is currently a postdoctoral researcher in the mathematics Department at the University of California, Los Angeles. His research interests include optimization, variational, and PDE-based methods in image processing, computer vision, computer graphics and machine learning.

Xue-Cheng Tai received the licenciate degree and the Ph.D. degree in applied mathematics from Jyvaskalya University, Jyvaskalya, Finland in 1989 and 1991, respectively. After holding several research positions in Europe, he became an associate professor in 1994 at the University of Bergen, Bergen, Norway, and a professor since 1997. He was also with Nanyang Technological University, Singapore between 2007–2011. His research interests include numerical PDE for image processing, multigrid, and domain decomposition methods, iterative methods for linear and nonlinear PDE problems, and parallel computing. He has guided numerous masters and Ph.D. students, and published more than 80 scientic papers. He has been a reviewer and an editor for a number of international journals.