Efficient Global Minimization for the Multiphase Chan-Vese Model of Image Segmentation *

Egil Bae† Xue-Cheng Tai‡

Abstract

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

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

1 Introduction

Multiphase image segmentation is a fundamental problem in image processing. Variational models such as the Mumford-Shah model [27] are powerful for this task, but efficient numerical computation of a global minimum is a big challenge. The level set method [14, 29] is a powerful tool which can used for numerical realization. It was first proposed for the Mumford-Shah model in [10] for two phases and [33] for multiple phases. Both approaches have the disadvantage of slow convergence and potential of getting stuck in a local minima.

*Support from the Norwegian Research Council (eVita project 166075), National Science Foundation of Singapore (NRF2007IDM-IDM002-010) and Ministry of Education of Singapore (Moe Tier 2 T207B2202) are gratefully acknowledged.

†Department of Mathematics, University of Bergen, Norway (Egil.Bae@math.uib.no).

‡Department of Mathematics, University of Bergen, Norway and Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. (tai@mi.uib.no).

Graph cuts from combinatorial optimization [15, 7, 16, 4, 20, 21, 3] is another technique which can perform image segmentation by minimizing certain discrete energy functions. In the recent years, the relationship between graph cuts and continuous variational problems have been much explored [5, 6, 12, 13]. It turns out graph cuts are very similar to the level set method, and can be used for many variational problems with the advantage of a much higher efficiency and ability to find global minima. It can be applied to the 2-phase Mumford-Shah model [11, 35], but for multiple phases it is probably not possible to find the exact, global minimum in polynomial time as this is an NP-hard problem. The usual approach to minimization problems with several regions is some heuristic method for finding an approximate, local minimum. Most popular in computer vision are the alpha-expansion and alpha-beta swap algorithms [7]. Recently, also convex formulations of the continuous multiphase problem have been made in [30, 23, 34] by relaxing the integrality constraint. A suboptimal solution is found by converting the real valued relaxed solution to an integral one (e.g. by thresholding).

In this paper we propose a method for globally and efficiently minimizing the energy in the Mumford-Shah model in the multiphase level set framework of Vese and Chan [33] by using binary level set functions as in [24]. Since the length term is slightly approximated in this framework, global minimization is not NP hard.

We will construct a graph such that the discrete variational problem can be minimized exactly by finding a minimum cut on the graph, provided the energy function is submodular. A sufficient condition on the data term is derived for when the energy is submodular. It is shown theoretically and experimentally that the condition is expected to hold for the most commonly used data term.

The submodularity of the energy function depends on how evenly the average intensity values of each region are distributed and may sometimes be violated under an L_1 data fitting term. To handle these cases, we have developed a method for minimizing non-submodular functions with particular emphasis on this energy function. It is shown in experiments that the algorithm can efficiently compute global solutions in practice, but we cannot prove it will always do so (which would conflict with NP-hardness of the problem).

The global optimization framework applies if the average intensity values of each region are fixed. One can also simultaneously minimize with respect to the average intensities values, by an alternating algorithm as in [2]. Although a global minimum with respect to the average intensities is not guaranteed, such an algorithm is much more robust to initialization.

Note that in contrast to alpha-expansion and alpha-beta swap, the approximation is done in the model rather than in the minimization method. Our approach only requires to find a minimum cut on a graph once, while alpha expansion and alpha-beta swap need to iteratively solve a sequence of minimum cut problems. By analyzing the complexity it can be easily seen that our method is more efficient.

In this work we will focus fully on the case of 4 or less phases, but aim to generalize to more phases later. Nevertheless, these are important cases since

by the four colour theorem, four phases in theory suffices to segment any 2D image. This can potentially be exploited in an algorithm in the future, by assigning different constant values to each disconnect component of the phases.

The paper is organized as follows: Section 1.1 reviews the Mumford-Shah model, the Chan-Vese model and the different level set representations. Section 2 presents the new global minimization approach for the multiphase Chan-Vese model. In Section 3 some theoretical analysis of the submodularity condition is presented. Section 4 presents algorithms for minimizing non-submodular energy functions.

Section 5 presents a local minimization approach for determining the average intensity values of the phases, while numerical experiments are presented in Section 6.

1.1 The Mumford-Shah model and its level set representation

Image segmentation is the task of partitioning the image domain Ω into a set of n meaningful disjoint regions $\{\Omega_i\}_{i=1}^n$.The Mumford-Shah model [27] is an established image segmentation model with a wide range of applications. An energy functional to be minimized is defined over the regions $\{\Omega_i\}_{i=1}^n$, and an approximation image u of the input image u_0. In an especially popular form, u is assumed to be constant within each region Ω_i, in which case the model reads

$$\min_{\{c_i\},\{\Omega_i\}} E_{MS}(\{c_i\},\{\Omega_i\}) = \sum_{i=1}^{n} \int_{\Omega_i} |c_i - u^0|^\beta dx + \sum_{i=1}^{n} \nu \int_{\partial\Omega_i} ds, \qquad (1)$$

where $\partial\Omega_i$ is the boundary of Ω_i. The power β is usually chosen as $\beta = 2$. As a numerical realization, Chan and Vese [10, 33] proposed to represent the above functional with level set functions, and solve the resulting gradient descent equations numerically. For two phases ($n = 2$) the level set representation yielded the variational problem

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

where $H(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ is the Heaviside function $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$. The multiphase case was handled by introducing more level set functions. By using $m = \log_2(n)$ level set functions, denoted $\phi^1, ..., \phi^m$, n phases could be represented. An important special case is the representation of 4 phases by two level set functions ϕ^1,ϕ^2, as in Table 1. The energy functional could then be written

$$\min_{\phi^1,\phi^2,c_1,...,c_4} E_{CV}(\phi^1, \phi^2, c_1, ..., c_4) = \nu \int_\Omega |\nabla H(\phi^1)| + \nu \int_\Omega |\nabla H(\phi^2)| \qquad (3)$$

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

	Traditional level set functions	Binary level set functions
$x \in$ phase 1 iff	$\phi^1(x) > 0, \phi^2(x) < 0$	$\phi^1(x) = 1, \phi^2(x) = 0$
$x \in$ phase 2 iff	$\phi^1(x) > 0, \phi^2(x) > 0$	$\phi^1(x) = 1, \phi^2(x) = 1$
$x \in$ phase 3 iff	$\phi^1(x) < 0, \phi^2(x) < 0$	$\phi^1(x) = 0, \phi^2(x) = 0$
$x \in$ phase 4 iff	$\phi^1(x) < 0, \phi^2(x) > 0$	$\phi^1(x) = 0, \phi^2(x) = 1$

Table 1: Representation of four phases by traditional and binary level set functions (note: a little permutation compared to the original paper [33]).

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

Note also that we have made a permutation in the interpretation of the phases compared to [33]. The energy is still exactly the same for all possible solutions. This permutation is crucial for making the corresponding discrete energy function submodular. Note also that the length term in (1) is slightly approximated, since some of the boundaries are counted twice. In fact there are a total of 6 types of boundaries between two different regions. In this model only two of them are counted twice, while the remaining 4 are counted once. This is very close to the ideal situation where each boundary is counted once.

The functional in the variational problem (3) is highly non-convex, even for fixed constant values $c_1, ..., c_4$. The traditional minimization approach of solving the gradient descent equations can therefore easily get stuck in a local minima. Furthermore, the numerical solution of the gradient descent PDEs is expensive computationally.

In [26, 24], the same multiphase model was formulated using binary level set functions $\phi^1, \phi^2 \in D = \{\phi \mid \phi : \Omega \mapsto \{0, 1\}\}$, representing the phases as in Table 1. This resulted in the energy functional

$$\min_{\phi^1,\phi^2 \in D, c_1,...,c_4} E_{CV}(\phi^1, \phi^2, c_1, ..., c_4) = \nu \int_\Omega |\nabla \phi^1| dx + \nu \int_\Omega |\nabla \phi^2| dx + E^{data}(\phi^1, \phi^2), \tag{4}$$

where

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

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

The functional was written in a slightly different form in [26, 24], but is exactly equal to the above in case ϕ^1 and ϕ^2 are binary functions. The constraint D was represented by the polynomials $K(\phi^1) = 0$ and $K(\phi^2) = 0$, where $K(\phi) = \phi(1 - \phi)$. Minimization of this constrained problem was carried out by the augmented lagrangian method. Since both the side constraints were non-convex, global minimization could not be guaranteed. Also, convergence was slow just as in the traditional level set approach. A similar approach could also be used for finding a local minimum with exact curve lengths [25].

Let us mention that a method often referred to as continuous graph cut can be used to globally minimize the Mumford Shah model in case of two phases. By letting $\phi \in D$, this model can be written

$$\min_{\phi \in D, c_1, c_2} \nu \int_\Omega |\nabla \phi| dx + \{\phi |c_1 - u^0|^\beta + (1-\phi)|c_2 - u^0|^\beta\} dx. \tag{5}$$

The idea, presented in [28] is to relax the constraint D by the convex constraint $D' = \{\phi \mid \phi : \Omega \mapsto [0,1]\}$. It was shown that thresholding this solution at almost any threshold in $(0,1]$ yields the optimal solution within D. Since (5) is convex, this procedure would yield the globally optimal solution.

One might immediately think the same idea could be extended to the multiphase case by iteratively minimizing (4) for ϕ^1 and ϕ^2 in D' and finally threshold the results. However, since $E^{data}(\phi^1, \phi^2)$ is not convex with respect to ϕ^1 and ϕ^2, the minimization would not be global.

In general, discrete graph cut has the disadvantage of some metrication artifacts over continuous graph cuts. However, discrete graph cuts is faster and can elegantly be used for minimization problems with non-local operators. The method we propose can easily be generalized to minimize non-local measurements of the curve lengths as was done for two phases in [8].

In the next section we will propose a method which globally minimizes (4) for fixed constant values $c_1, ..., c_4$. This new approach is also shown to be very superior in terms of efficiency compared to gradient descent. We start by deriving a discretizatation of (4).

1.2 Discretization of energy functional

Instead of discretizing the Euler-Lagrange equations, we will discretize the variational problem (4). In the next section we show how to minimize the resulting discrete energy function exactly by graph cuts. Let us first mention there are two variants of the total variation term. The isotropic variant, by using 2-norm

$$TV_2(\phi) = \int_\Omega |\nabla \phi|_2 \, dx = \int_\Omega \sqrt{|\phi_{x_1}|^2 + |\phi_{x_2}|^2} \, dx \tag{6}$$

and the anisotropic variant, by using 1-norm

$$TV_1(\phi) = \int_\Omega |\nabla \phi|_1 \, dx = \int_\Omega |\phi_{x_1}| + |\phi_{x_2}| \, dx. \tag{7}$$

Only the anisotropic variant is graph representable and will be considered here. A more isotropic graph representable version can be obtained by splitting TV_1 using the original gradient operator, and one rotated counterclockwise $\pi/4$ radians

$$TV_{1,\frac{\pi}{4}}(\phi) = \frac{1}{2} \int_\Omega \left\{ |\nabla \phi(x)|_1 + |R_{\frac{\pi}{4}} \nabla \phi(x)|_1 \right\} dx, \tag{8}$$

where $R_{\frac{\pi}{4}} \nabla$ is the gradient in the rotated coordinate system. It is also possible to create even more isotropic versions by considering more such rotations.

Let $\mathcal{P} = \{(i,j) \subset \mathbb{Z}^2\}$ denote the set of grid points. For each $p = (i,j) \in \mathcal{P}$, the neighborhood system $\mathcal{N}_p^k \subset \mathcal{P}$ is defined as

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

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

The discrete energy function can be written

$$\min_{\phi^1,\phi^2 \in D, c_1,...,c_4} E_d(\phi^1, \phi^2, c_1, ..., c_4) = \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^1 - \phi_q^1| + \nu \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p^k} w_{pq} |\phi_p^2 - \phi_q^2| + \sum_{p \in \mathcal{P}} E_p^{data}(\phi_p^1, \phi_p^2), \tag{9}$$

where

$$E_p^{data}(\phi_p^1, \phi_p^2) = \{\phi_p^1 \phi_p^2 |c_2 - u_p^0|^\beta + \phi_p^1(1 - \phi_p^2)|c_1 - u_p^0|^\beta) + (1 - \phi_p^1)\phi_p^2 |c_4 - u_p^0|^\beta + (1 - \phi_p^1)(1 - \phi_p^2)|c_3 - u_p^0|^\beta\},$$

and $k = 4$ for TV_1 and $k = 8$ for $TV_{1,\frac{\pi}{4}}$. The weights w_{pq} are then given by $w_{pq} = \frac{4\delta^2}{k||p-q||_2}$. Similar weights can also be derived from the Cauchy-Crofton formula of integral geometry as was done for two phases in [5]. It can furthermore be proved that as the mesh size decreases and the size of the neighborhood system increases, the minimizers of the discrete energy function converges to minimizers of the continuous energy functional.

2 Graph cut minimization

We will show that the discrete problem (9) can be minimized globally by finding the minimum cut on a specially designed graph. This is possible provided the constant values $c_1, ..., c_4$ are sufficiently evenly distributed. We show that such a distribution makes the discrete energy function sub-modular. Some analysis of the condition is given in Section 3, where it is argued it will be satisfied in practice for the most commonly used data terms. In Section 4, an algorithm is developed for minimizing non-submodular functions with particular emphasize on functions of the form (4).

2.1 Brief overview of graph cuts in computer vision

Graph cut is a well known optimization problem. Due to a duality theorem by Ford and Fulkerson [22], there are several fast algorithms for this problem. It was introduced as a computer vision tool by Greig et. al. [16] in connection with markov random fields [15] and has later been studied by Kolmogorov et. al. [4, 20]. Its applications range from stereo vision [19], segmentation [3, 17, 35, 11] to noise removal [12, 13, 9].

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

$$c(\mathcal{V}_s, \mathcal{V}_t) = \sum_{(i,j)\in\mathcal{E} \text{ s.t. } i\in\mathcal{V}_s, j\in\mathcal{V}_t} c(i, j).$$

A flow f on $\mathcal{G}$ is a function $f : \mathcal{E} \mapsto \mathbb{R}$. For a given flow, the residual capacities are defined as $R(e) = c(e) - f(e)\ \forall e \in \mathcal{E}$. The max flow problem is to find maximum amount of flow that can be pushed from $\{s\}$ to $\{t\}$, under flow conservation constraint at each vertex. The theorem of Ford and Fulkerson says this is the dual to the problem of finding the cut of minimum cost on $\mathcal{G}$, the min-cut problem. Therefore, efficient algorithms for finding max-flow, such as the augmented paths method [22] can be used for finding minimum cuts in graphs. An efficient implementation of this algorithm specialized for image processing problems can be found in [4]. This algorithm, which is available on-line has been used in our experiments.

In computer vision this has been exploited for minimizing energy functions of the form

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

Typically, $i = 1, ..., m$ denotes the set of grid points and x contains one binary variable for each grid point. In order to be representable as a cut on a graph, it is required that the energy function is submodular (or regular) [20, 15], i.e.

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

2.2 Graph construction

Observe that the energy function E is composed of pairwise interaction terms between binary variables. Such energy functions can be minimized exactly via graph cuts, provided the pairwise interaction potentials are submodular [20, 15]. In particular, this requires the data term to satisfy

$$E_p^{data}(1,1) + E_p^{data}(0,0) \leq E_p^{data}(1,0) + E_p^{data}(0,1) \tag{10}$$

We will construct a graph $\mathcal{G}$ such that there is a one-to-one correspondence between cuts on $\mathcal{G}$ and the level set functions ϕ^1 and ϕ^2. Furthermore, the minimum cost cut will correspond to the level set functions ϕ^1 and ϕ^2 minimizing the energy (9).

$$\min_{(\mathcal{V}_s,\mathcal{V}_t)} c(\mathcal{V}_s, \mathcal{V}_t) = \min_{\phi^1,\phi^2} E_d(\phi^1, \phi^2, c_1, ..., c_4) + \sum_{p\in\mathcal{P}} \sigma_p. \tag{11}$$

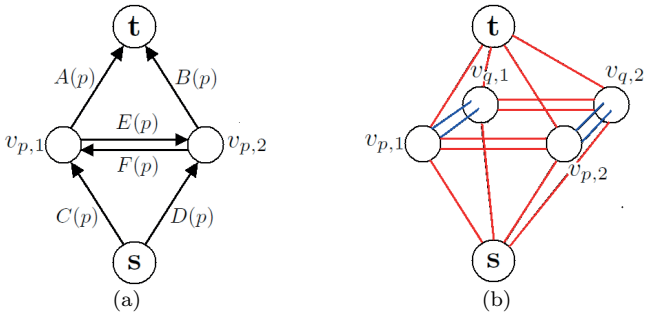


Figure 1: (a) The graph corresponding to the data term at one grid point p. (b) A sketch of the graph corresponding to the energy function of a 1D signal of two grid points p and q, red: data edges, blue: regularization edges.

where $\sigma_p \in \mathbb{R}$ are fixed for each $p \in \mathcal{P}$. In the graph, two vertices are associated to 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

$$\mathcal{V} = \{v_{p,i} \mid \ p \in \mathcal{P}, \ \ i = 1, 2\} \cup \{s\} \cup \{t\}. \tag{12}$$

The edges are constructed such that the relationship (11) is satisfied. We begin with the edges constituting the data term of (9). For each grid point $p \in \mathcal{P}$ they are defined as

$$\mathcal{E}_D(p) = (s, v_{p,1}) \cup (s, v_{p,2}) \cup (v_{p,1}, t) \cup (v_{p,2}, t) \cup (v_{p,1}, v_{p,2}) \cup (v_{p,2}, v_{p,1}). \tag{13}$$

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

$$\mathcal{E}_R = \{(v_{p,1}, v_{q,1}), (v_{p,2}, v_{q,2}) \ \forall p, q \subset \mathcal{P} \text{ s.t.} q \in \mathcal{N}_p^k\}. \tag{14}$$

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

$$\phi_p^1 = \begin{cases} 1 & \text{if } v_{p,1} \in V_s, \\ 0 & \text{if } v_{p,1} \in V_t, \end{cases} \qquad \phi_p^2 = \begin{cases} 1 & \text{if } v_{p,2} \in V_s, \\ 0 & \text{if } v_{p,2} \in V_t. \end{cases} \tag{15}$$

Weights are assigned to the edges such that the relationship (11) is satisfied. Weights on the regularization edges are simply given by

$$c\left(v_{p,1}, v_{q,1}\right) = c\left(v_{q,1}, v_{p,1}\right) = c\left(v_{p,2}, v_{q,2}\right) = c\left(v_{q,2}, v_{p,2}\right) = \nu w_{pq}, \qquad \forall p \in \mathcal{P}, \ q \in \mathcal{N}_p^k. \tag{16}$$

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

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

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

In Figure 1(a) the graph corresponding to an image of one pixel p is shown. It is clear that these weights must satisfy

$$\begin{cases} A(p) + B(p) & = |c_2 - u_p^0|^\beta + \sigma_p \\ C(p) + D(p) & = |c_3 - u_p^0|^\beta + \sigma_p \\ A(p) + E(p) + D(p) & = |c_1 - u_p^0|^\beta + \sigma_p \\ B(p) + F(p) + C(p) & = |c_4 - u_p^0|^\beta + \sigma_p \end{cases} \tag{17}$$

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

$$|c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta = A(p) + B(p) + C(p) + D(p) + E(p) + F(p) - 2\sigma_p$$

$$\geq A(p) + B(p) + C(p) + D(p) - 2\sigma_p = |c_2 - u_p^0|^\beta + |c_3 - u_p^0|^\beta.$$

This condition must hold for all grid points $p \in \mathcal{P}$, which is exactly the submodular condition (10). Hence, the following condition on the constant values $c_1, ..., c_4$ must be satisfied

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta, \quad \forall\, I \in [0, L], \tag{18}$$

where L is the maximum intensity value.

A detailed analysis of the condition (18) is given in Section 3.

Assuming (18) holds, the linear system (17) has infinitely many solutions. It was shown in [20] that at most three edges are required for representing a general submodular term of two binary variables. Therefore, it is possible to pick a solution such that at least three of the weights $A(p), B(p), C(p), D(p), E(p), F(p)$ in $\mathcal{E}_D(p)$ become zero for each $p \in \mathcal{P}$. The construction of the solution is as follows

$$A(p) = \max\{|c_2 - u_p^0|^\beta - |c_4 - u_p^0|^\beta, 0\}, \tag{19}$$

$$C(p) = \max\{|c_4 - u_p^0|^\beta - |u_p^0 - c_2|^\beta, 0\} \tag{20}$$

$$B(p) = \max\{|c_4 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, 0\}, \tag{21}$$

$$D(p) = \max\{|c_3 - u_p^0|^\beta - |c_4 - u_p^0|^\beta, 0\}, \tag{22}$$

$$E(p) = |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta - |c_2 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, \quad F(p) = 0. \tag{23}$$

The value σ_p is given implicitly by this solution.

Therefore, by analyzing the complexity of the method in the augmented paths framework, it is easily seen that the computational cost is equal to the cost of one single iteration of the alpha expansion method.

Note finally that three phase segmentation is a special case that can be handled by putting infinite cost to one of the four possible solutions, i.e. $E(p) = \infty$ or $F(p) = \infty$.

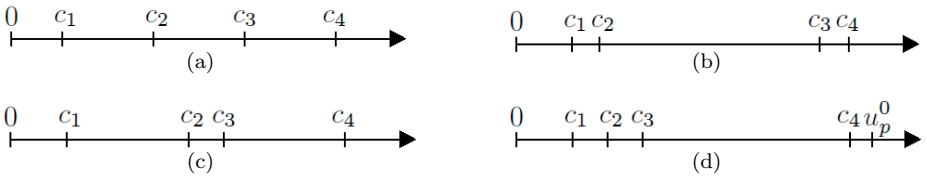


Figure 2: (a), (b) and (c) distributions of $\mathbf{c}$ which makes energy function submodular for all β. (d) distribution of $\mathbf{c}$ which may make energy function non-submodular for small β

3 Analysis of submodular condition

The condition (18) says something about how evenly $\{c_i\}_{i=1}^4$ are distributed. First we characterize situations for which (10) is always satisfied.

Proposition 1. Let $0 \leq c_1 < c_2 < c_3 < c_4$. (18) is satisfied for all $I \in [\frac{c_2-c_1}{2}, \frac{c_4-c_3}{2}]$ for any $\beta \geq 1$.

Proposition 2. Let $0 \leq c_1 < c_2 < c_3 < c_4$. (18) is satisfied for any $\beta \geq 1$ if $c_2 - c_1 = c_4 - c_3$.

Further, it can be observed that (18) becomes less strict as β increases, as the next two results show.

Proposition 3. Let $0 \leq c_1 < c_2 < c_3 < c_4$. If (18) is satisfied for some $\beta \geq 1$, then (10) is satisfied for all powers $\alpha \geq \beta$.

Proposition 4. Let $0 \leq c_1 < c_2 < c_3 < c_4$. There exists a $\mathcal{B} \in \mathbb{N}$ such that (18) is satisfied for any $\beta \geq \mathcal{B}$.

In fact we have observed that for $\beta = 2$, (18) is always satisfied in practice for optimal constant values.

Figure 2 shows examples where the condition is satisfied and may fail. Prop. 3 is illustrated in Figure 2 (b) and (c). Figure 2 (d) shows the only possibility in which (18) may be violated, i.e. c_1, c_2, c_3 are clustered compared to c_4 (the symmetrical version would also be a problem). However, the model (4) will disfavor solutions where the constants are clustered. Experiments will demonstrate that under L^2 data fidelity (18) is always satisfied for optimal values of $\mathbf{c}$. Under L^1 data fidelity, it may be more easily violated.

In case (18) does not hold at some $p \in \mathcal{P}$, the energy function is non-submodular. Not only does this mean it cannot in general be minimized by graph cut. It also implies the minimization problem is NP-hard, hence no algorithm exist that can solve the problem in polynomial time (unless $P = NP$).

4 Minimization of non-submodular energy functions

Assume that the submodularity condition (18) is not satisfied. We will develop a method for minimizing non-submodular energy functions with particular emphasize on energy functions of the form (9). The algorithm cannot be guaranteed to always find a global solution, but works well in practice. Minimization of non-submodular functions via graph cuts has been investigated previously, see [18] for a review. The usual idea is to develop a method for determining most of the variables, while leaving some of the variables undetermined. We instead aim to determine all the variables. Even when (18) does not hold, the energy function is "almost submodular", which we believe explains why the following very efficient algorithms work so well in practice.

Assume that

$$|c_2 - u_p^0|^\beta + |c_3 - u_p^0|^\beta > |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta,$$

for some $p \in \mathcal{P}$. In this case the linear system (17) has a solution only if either $E(p) < 0$ or $F(p) < 0$, in which case one of the edges, $(v_{p,1}, v_{p,2})$ or $(v_{p,2}, v_{p,1})$, will have negative weight. In order to construct the solution we consider two cases. If $u_p^0 > c_3$, then

$$E(p) = |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta - |c_2 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, \quad F(p) = 0 \tag{24}$$

$$A(p) = \max\{|c_2 - u_p^0|^\beta - |c_4 - u_p^0|^\beta, 0\} - E(p), \tag{25}$$

$$C(p) = \max\{|c_4 - u_p^0|^\beta - |u_p^0 - c_2|^\beta, 0\} - E(p), \tag{26}$$

$$B(p) = \max\{|c_4 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, 0\} - E(p), \tag{27}$$

$$D(p) = \max\{|c_3 - u_p^0|^\beta - |c_4 - u_p^0|^\beta, 0\} - E(p), \tag{28}$$

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

$$F(p) = |c_1 - u_p^0|^\beta + |c_4 - u_p^0|^\beta - |c_2 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, \quad E(p) = 0 \tag{29}$$

$$A(p) = \max\{|c_1 - u_p^0|^\beta - |c_3 - u_p^0|^\beta, 0\} - F(p), \tag{30}$$

$$C(p) = \max\{|c_3 - u_p^0|^\beta - |u_p^0 - c_1|^\beta, 0\} - F(p), \tag{31}$$

$$B(p) = \max\{|c_2 - u_p^0|^\beta - |c_1 - u_p^0|^\beta, 0\} - F(p), \tag{32}$$

$$D(p) = \max\{|c_1 - u_p^0|^\beta - |c_2 - u_p^0|^\beta, 0\} - F(p), \tag{33}$$

in which case $F(p) < 0$. By Prop 1, the condition holds whenever $u_p^0 \in [c_2, c_3]$. In this section we let $\mathcal{G}$ denote the graph with data edges set according to (24)-(33) (some of which may be negative) and regularization edges set to (16).

4.1 Truncation of non-submodular terms

It is difficult to interpret what is physically meant by max flow on a graph with negative edge weights. The concept of min-cut, on the other hand, certainly have a meaning even if some of the edges have negative weight. If all the edges have negative weight, the min-cut problem becomes equivalent to the max-cut problem on a graph with negated edge weights. The first step of the method finds a good feasible solution, and therefore also a good upper bound on the objective function (9). It seems that most often this feasible solution is in fact the optimal solution. All edges with negative weight will be removed, resulting in a new graph $\overline{\mathcal{G}}$. It has been observed in [31] that removing negative edges, often called truncation, can be effective in minimizing non-submodular functions. We will see that this applies especially well to our energy function. Furthermore, we will derive a criterion for when the minimum cut on the graph with removed edges of negative weight is also a minimum cut on the original graph with negative edge weights.

Let $\overline{\mathcal{G}}$ be the graph identically to $\mathcal{G}$ except that all edges of negative weight are removed. The minimum cut on $\overline{\mathcal{G}}$ can be easily computed by max-flow. As discussed in the previous section, the condition (18) 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

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

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

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

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

Theorem 5. Let $\mathcal{G}$ be a graph as defined in (12)-(14) and (16), with weights $A(p), B(p), C(p), D(p), E(p), F(p)$ satisfying (17). 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

$$R_A(p) + R_D(p) \geq -E(p), \quad \forall p \in \mathcal{P}^1 \quad \text{and} \quad R_B(p) + R_C(p) \geq -F(p), \quad \forall p \in \mathcal{P}^2, \tag{34}$$

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

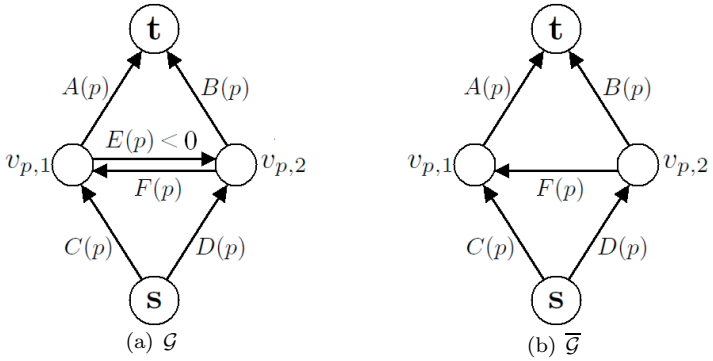


Figure 3: Illustration of graph $\overline{\mathcal{G}}$ in case $E(p) < 0$.

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

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

We will 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 $\underline{\mathcal{G}}$ and $\mathcal{G}$. For each $p \in \mathcal{P}_1$ there are 4 possibilities for the cut (V_s, V_t). Since $R_A(p), R_B(p), R_C(p), R_D(p) \geq 0$, the cost of all the 3 cuts $v_{p,1}, v_{p,2} \in V_s$, $v_{p,1}, v_{p,2} \in V_t$ and $v_{p,1} \in V_t, v_{p,2} \in V_s$ are lower in $\underline{\mathcal{G}}$ than in $\mathcal{G}$. The last cut $v_{p,1} \in V_s, v_{p,2} \in V_t$ has the cost $A(p) + B(p) - E(p)$ in the $\mathcal{G}$ and the cost $A(p) + D(p) - (R_A(p) + R_D(p)) \leq A(p) + D(p) + E(p)$ in the graph $\underline{\mathcal{G}}$. The same argument shows that all possible cuts have a lower or equal cost in $\underline{\mathcal{G}}$ than in $\mathcal{G}$ for $p \in \mathcal{P}_2$.

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

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

Hence, since the max flow on $\overline{\mathcal{G}}$ is feasible on $\underline{\mathcal{G}}$ it is also optimal on $\underline{\mathcal{G}}$. Therefore, by duality 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 (34), we can check whether the solution is optimal on $\mathcal{G}$. Most often it is possible to stop at this stage, since the residual capacity conditions are satisfied everywhere.

4.2 Refinement of truncated solution

If the regularization parameter is set extremely high, the residual criterion (34) may be violated at a small set of the pixels. In that case one could accept the computed solution as suboptimal. We have also developed an algorithm which iterates upon the cut on $\overline{\mathcal{G}}$ until it finds a cut on $\mathcal{G}$. We cannot prove the algorithm always converges (which would conflict with the NP-hardness of the problem), but it works well in practice.

If the residual capacity conditions are violated, there is a possibility the optimal assignment is $\phi^1(p) = 0, \phi^2(p) = 1$ if $p \in \mathcal{P}_1$ and $\phi^1(p) = 1, \phi^2(p) = 0$ if $p \in \mathcal{P}_2$, even if a different assignment was produced by the cut on the graph $\overline{\mathcal{G}}$. In the first step of the algorithm, the weights on the data edges at p are modified such that the cost of the above assignments are correct, at the expense of one of the remaining 3 assignments. There are possibilities for several variants. In the variant below, the weights are modified such that the cost of either assignment $\phi^1(p), \phi^2(p) = 0$ or $\phi^1(p), \phi^2(p) = 1$ are reduced, while the remaining assignments are correct. Which assignment to be modified, is determined by examining the cut on the graph $\overline{\mathcal{G}}$. The cut on the modified graph is expected to be closer to a cut on $\mathcal{G}$. If the cut on the new graph may causes either: (1) the residual capacity condition to be violated at some new $p \in \mathcal{P}_1 \cup \mathcal{P}_2$, (2) the assignment of $\phi^1(p), \phi^2(p)$ to a region of reduced cost, one cannot guarantee the obtained cut is also optimal on $\mathcal{G}$. In that case, the same procedure can be reiterated until neither (1) or (2) are violated.

The algorithm creates a sequence of graphs $\{\mathcal{G}_i\}_{i=1}^n$, with $\mathcal{G}_0 = \overline{G}$, of only positive edge weights, such that the min-cut on $\mathcal{G}_i$ converges to a min-cut on $\mathcal{G}$. The graphs satisfy min-cut$(\mathcal{G}_0)$ = min-cut$(\overline{\mathcal{G}})$, min-cut$(\mathcal{G}_i) \leq$ min-cut$(\overline{\mathcal{G}})$ for all i, and min-cut$(\mathcal{G}_n)$ = min-cut$(\mathcal{G})$. For a given flow on $\mathcal{G}_i$ we define two new sets $\mathcal{P}_i^1 \subseteq \mathcal{P}^1$ and $\mathcal{P}_i^2 \subseteq \mathcal{P}^2$

$$\mathcal{P}_i^1 = \{p \in \mathcal{P}^1 \mid R_A^i(p) + R_D^i(p) < -E(p)\},$$

$$\mathcal{P}_i^2 = \{p \in \mathcal{P}^2 \mid R_B^i(p) + R_C^i(p) < -F(p)\},$$

where $R_A^i(p), R_B^i(p), R_C^i(p)$ and $R_D^i(p)$ are the residual capacities in graph $\mathcal{G}_i$ on edges $(v_{p,1}, t), (v_{p,2}, t), (s, v_{p,1}), (s, v_{p,2})$ respectively.

By construction, a sufficient stopping criterion that ensures both (1) and (2) above is to require $\mathcal{G}_i = \mathcal{G}_{i-1}$, that is, the weights on all edges of $\mathcal{G}_i$ and $\mathcal{G}_{i-1}$ are equal. The algorithm is as follows.

Algorithm 1:
$\mathcal{G}_0 = \overline{\mathcal{G}}, \mathcal{G}_{-1} = \emptyset, i = 0$
Find max flow on $\mathcal{G}_0$, update $\mathcal{P}_0^1$ and $\mathcal{P}_0^2$

if($\mathcal{P}_0^1$ and $\mathcal{P}_0^2$ are empty)
stop, optimal solution found
else:
while($\mathcal{G}_i \neq \mathcal{G}_{i-1}$){
1. Construct $\mathcal{G}_{i+1}$ as in $\overline{\mathcal{G}}$ except for the following weights

for all $p \in \mathcal{P}_i^1$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_t$ in $\mathcal{G}_i$): set $c(v_{p,1}, t) = A(p) + E(p)$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_s$ in $\mathcal{G}_i$): set $c(s, v_{p,2}) = D(p) + E(p)$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_t$ in $\mathcal{G}_i$): set $c(s, v_{p,1}) = A(p) + E(p)$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_s$ in $\mathcal{G}_i$): set $c(s, v_{p,1}) = D(p) + E(p)$

for all $p \in \mathcal{P}_i^2$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_t$ in $\mathcal{G}_i$): set $c(v_{p,2}, t) = B(p) + F(p)$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_s$ in $\mathcal{G}_i$): set $c(s, v_{p,1}) = C(p) + F(p)$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_t$ in $\mathcal{G}_i$): set $c(s, v_{p,2}) = B(p) + F(p)$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_s$ in $\mathcal{G}_i$): set $c(s, v_{p,2}) = C(p) + F(p)$

2. Find max-flow on $\mathcal{G}_{i+1}$
3. Update $\mathcal{P}_{i+1}^1$ and $\mathcal{P}_{i+1}^2$ by examining residual capacities in $\mathcal{G}_{i+1}$ 4. $i \leftarrow i + 1$
}

Theorem 6. Let $\mathcal{G}_n$ be the graph at termination of Algorithm 1 and let $(\mathcal{V}_s, \mathcal{V}_t)$ be the minimum cut on $\mathcal{G}_n$ computed at the last iteration. Then $(\mathcal{V}_s, \mathcal{V}_t)$ is a minimum cut on $\mathcal{G}$.

Proof. If the algorithm terminates with $\mathcal{G}_0$, optimality was proved in theorem 5. Assume therefore $n \geq 1$. The proof follows some of the same ideas as the proof of theorem 5. We will use $\mathcal{G}_n$ to construct a graph $\underline{\mathcal{G}}$ such that the minimum cut problem on $\underline{\mathcal{G}}$ is a relaxation of the minimum cut problem on $\mathcal{G}$. Observe first that since $\mathcal{G}_n = \mathcal{G}_{n-1}$, the minimum cut on $\mathcal{G}_n$ is feasible, no edges in the cut have a reduced cost. Therefore, min-cut$(\mathcal{G}_n) \geq$ min-cut$(\mathcal{G})$

The graph $\underline{\mathcal{G}}$ is constructed with weights as in $\mathcal{G}_n$ except

$$c(v_{p,1}, t) = A(p) - R_A^n(p), \quad \forall p \in \mathcal{P}^1 \backslash \mathcal{P}_n^1,$$

$$c(s, v_{p,2}) = D(p) - R_D^n(p), \quad \forall p \in \mathcal{P}^1 \backslash \mathcal{P}_n^1$$

$$c(v_{p,2}, t) = B(p) - R_B^n(p), \quad \forall p \in \mathcal{P}^2 \backslash \mathcal{P}_n^2,$$

$$c(s, v_{p,1}) = C(p) - R_C^n(p), \quad \forall p \in \mathcal{P}^2 \backslash \mathcal{P}_n^2.$$

Then min-cut$(\underline{\mathcal{G}}) \leq$ min-cut$(\mathcal{G}) \leq$ min-cut$(\mathcal{G}_n)$. These inequalities can be shown by the same argument as in the proof of Theorem 5. By construction, the max flow on $\mathcal{G}_n$ is feasible on $\underline{\mathcal{G}}$, and therefore also optimal on $\underline{\mathcal{G}}$. Hence, by duality min-cut$(\underline{\mathcal{G}}) =$ min-cut$(\mathcal{G}_n)$ which implies min-cut$(\mathcal{G}) =$ min-cut$(\mathcal{G}_n)$. □

There is a lot of redundancy in this algorithm. It is not necessary to compute the max-flow from scratch in each iteration, especially in the augmenting paths framework. Rather, starting with the max flow in $\mathcal{G}_i$, flow can be pulled back along $s-t$ paths passing through vertices $v_{p,1}$ or $v_{p,2}$ for $p \in \mathcal{P}_0^1 \cup \mathcal{P}_0^2$ until it becomes feasible in graph $\mathcal{G}_{i+1}$. With such an initial flow, only a few augmenting paths are required to find the max flow on $\mathcal{G}_{i+1}$. Since $\mathcal{P}^1$ and $\mathcal{P}^2$ are small subsets of $\mathcal{P}$, and $\mathcal{P}_0^1 \cup \mathcal{P}_0^2$ are small subsets of $\mathcal{P}^1 \cup \mathcal{P}^2$, the cost of this algorithm is negligible.

5 Unknown constant values, algorithm

The algorithm presented in the last sections minimizes $E_d(\mathbf{c}, \phi^1, \phi^2)$ with respect to ϕ^1, ϕ^2 for a fixed $\mathbf{c}$. Vice versa, for a fixed ϕ^1, ϕ^2 the values $\mathbf{c}$ minimizing $E_d(\mathbf{c}, \phi^1, \phi^2)$ are given by the average intensity in each region

$$c_i = \frac{\int_{\Omega_i} u^0 \, dx}{\int_{\Omega_i} dx}, \; i = 1, ..., n \tag{35}$$

We want an algorithm to minimize both with respect to ϕ and $\mathbf{c}$. As in [2], this is achieved by combining the two above results in the following iterative descent algorithm

Algorithm 2:
Estimate initial values $\mathbf{c}^0$, set $l = 0$.
while($||\mathbf{c}^l - \mathbf{c}^{l-1}|| > \text{tol}$)
Use graph cuts to estimate ϕ from

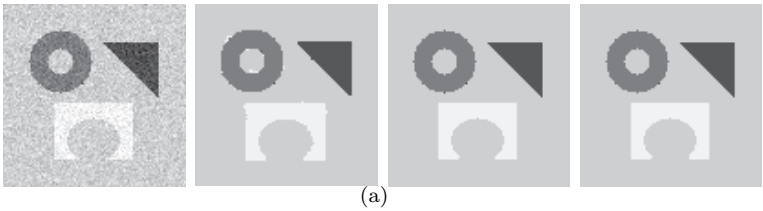
$$\phi = arg\, min_{\tilde{\phi}^1, \tilde{\phi}^2} E_d(\mathbf{c}^l, \tilde{\phi}^1, \tilde{\phi}^2). \tag{36}$$

Update $\mathbf{c}^{l+1}$ according to equation (35).
Update $l \leftarrow l + 1$

Note that no initialization of the level set functions are required. Only the values $\mathbf{c}^0$ need to be initialized, which can be achieved very efficiently by the isodata algorithm [32]. In our experiments, convergence was reached in around 5-15 iterations. It must be noted that this algorithm is no longer guaranteed to find the global minima. Theoretically, it may get trapped in a local minima close to the initial values $\mathbf{c}^0$. However, in practice it is usually rather insensitive to initialization.

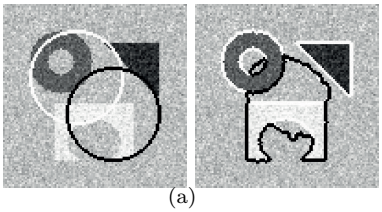
6 Numerical results

Numerical examples are shown in Figure 4 - 7, where we have used the power $\beta = 2$ in the data term. The constant values $\{c_i\}_{i=1}^4$ are estimated by running Algorithm 2 until convergence (6-10 iterations). During each iteration, the energy minimization problem was submodular.



(a)

Figure 4: L^2 data fidelity. From left to right: input, level set method gradient descent, our approach, alpha expansion/alpha beta swap.



(a)

Figure 5: Level set method: (a) bad initialization, (b) result.

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

More difficult images are presented in Figure 6 - 7. The L^2 data fidelity term has been used ($\beta = 2$) and the different methods are compared by keeping the same constant values $\{c_i^*\}_{i=1}^4$ and regularization parameter ν fixed, while minimizing in terms of the regions. One can clearly see the advantages of the global approach over earlier approaches.

6.1 Experiments on L_2 data fitting term: submodularity

In Section 3 we gave theoretical insights on how submodularity of the energy function was related to the distribution of the values c_i, $i = 1, ..., 4$. It was shown that the condition becomes less strict as β increases. In this section we demonstrate that for L_2 data fitting term ($\beta = 2$ in (3)), the energy function is submodular in practice. The L_2 norm tolerates rather uneven distributions of c_i, $i = 1, ..., 4$. In addition, the parameters c_i $i = 1, ..., 4$ minimizing the energy function are not expected to get too clustered. To verify this, we have run Algorithm 2 for optimizing the parameters c_i, $i = 1, ..., 4$ on all images from the data base [1]. For all the experiments, the submodularity condition was

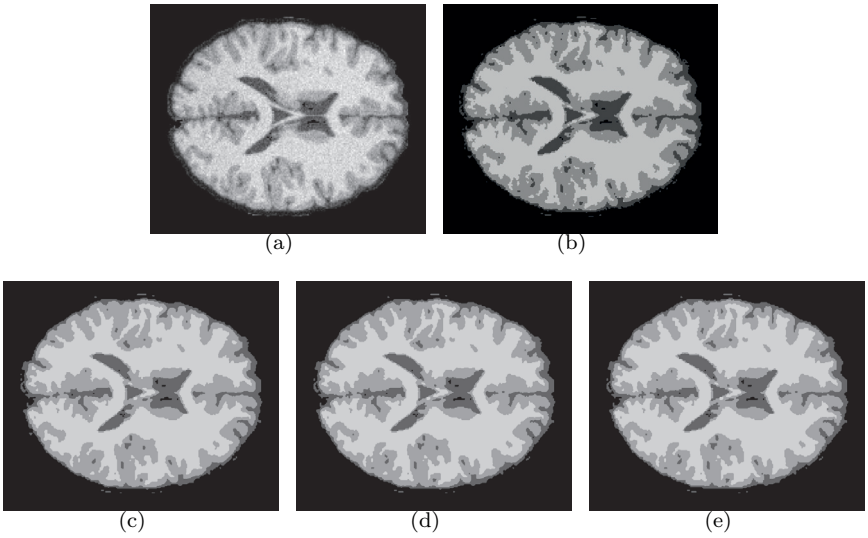


Figure 6: Experiment 2: (a) Input, (b) gradient descent, (c) our approach, (d) alpha expansion, (e) alpha-beta swap.

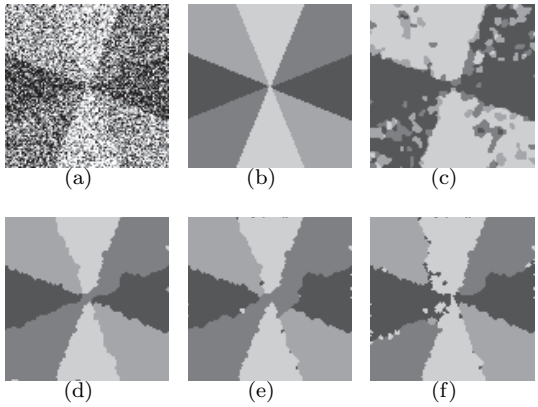
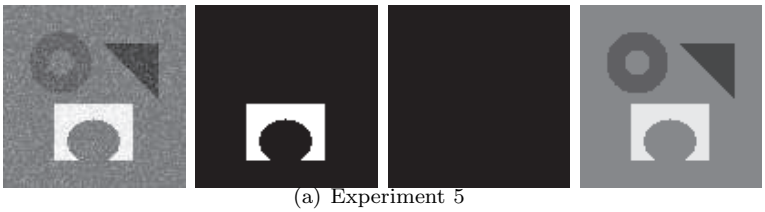


Figure 7: Experiment 3: (a) Input image, (b) ground truth, (c) gradient descent, (d) our approach, (e) alpha expansion, (f) alpha-beta swap.



(a) Experiment 5

Figure 8: L^1 data fidelity. Note that the constant values of c_1, c_2, c_3 are very close to each other compared to c_4. From left to right: input image, set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was not satisfied, set of pixels where residual criterion (34) is not satisfied (empty set), output image.

satisfied during each iteration of the algorithm.

6.2 Experiments on non-submodular energy minimization

The purpose of this section is to demonstrate the methods for minimizing non-submodular energy functions. For that reason, we will use the L_1 data fidelity term and fix the values c_i, $i = 1, ..., 4$ in such a way that the submodular condition (18) is violated. Figure 8 shows such an example, which is a modified version of the example in Figure 4, where the average intensities values of 3 of the objects are close compared to the 4th object. Some more natural examples are shown in Figure 9 and 10. Subfigures (b) show the set of pixels $p \in \mathcal{P}_1 \cup \mathcal{P}_2$, where the submodular condition was violated. Subfigures (c) show the set of pixels where the residual capacity conditions (34) were violated, which is the empty set in all cases. Therefore, the solutions obtained by the cut on the graphs $\overline{\mathcal{G}}$ are also global solutions to the original problems.

If the regularization parameter ν is set extremely high, the residual capacity condition (34) may also be violated at a small set of the pixels. Two such examples are shown in Figure 11 and 12. As we see from subfigures (c), the residual capacity condition is only violated at a small set of the pixels. By applying the refinement algorithm from Section 4.2, the exact global solution can be obtained in two iterations, as shown in subfigures (c)-(e).

7 Conclusions

We have developed a global minimization method for the multiphase Chan-Vese model of image segmentation based on graph cuts. Numerical experiments also demonstrated superior efficiency of the new approach over gradient descent. It was shown that the energy function was submodular provided the average intensity values of each region was sufficiently evenly distributed. The strictness on of the condition depended on the data term. For L_p data term with $p \geq 2$, the condition was satisfied in all our experiments. For L_1 data term the condition

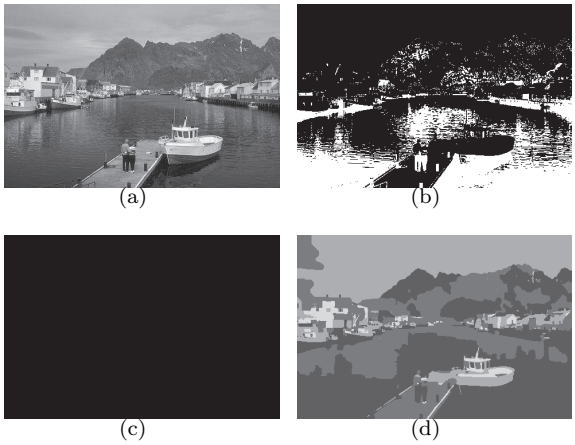


Figure 9: Segmentation with L^1 data fidelity ($\beta = 1$): (a) Input; (b) Set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was not satisfied; (c) Set of pixels where residual capacity criterion (34) was not satisfied (empty set); (d) Output.

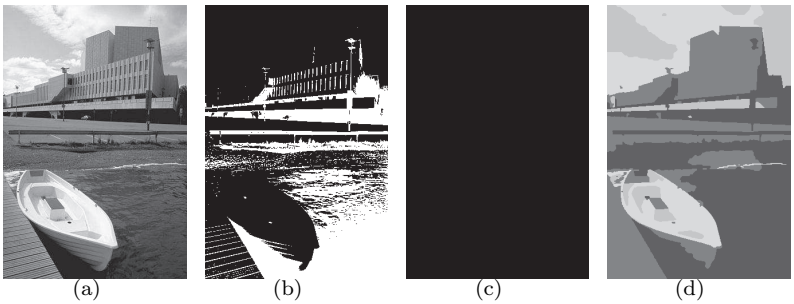


Figure 10: Segmentation with L^1 data fidelity ($\beta = 1$): (a) Input; (b) Set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was not satisfied; (c) Set of pixels where residual capacity criterion (34) was not satisfied (empty set); (d) Output.

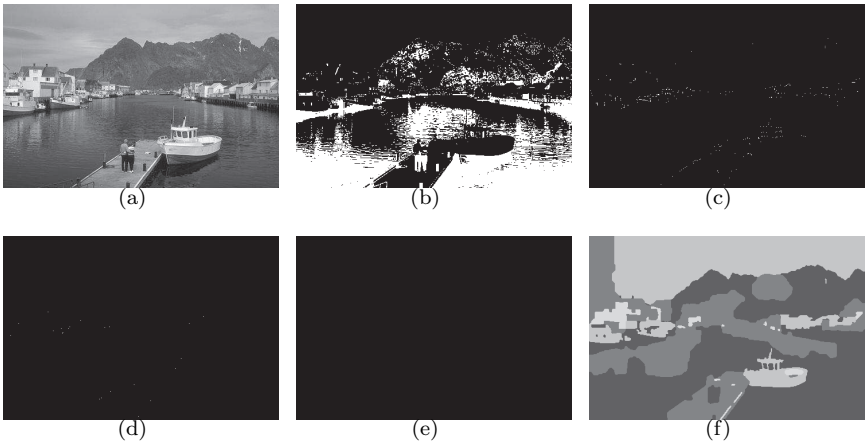


Figure 11: Segmentation with L^1 data fidelity ($\beta = 1$) and very high regularization ν: (a) Input image; (b) Set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was not satisfied; (c) Set of pixels where the residual capacity condition (34) was violated; (c) -(e) Set of pixels where the weights on graph $\mathcal{G}_i$ differ from the weights on the graph $\mathcal{G}_{i-1}$, $i = 1, ..., 3$; (f) Output (global solution).

was more easily violated. Algorithms for minimizing non-submodular functions were developed, which most often computed global solution, but could not be proved to always do so.

In this work, we have restricted our attention to four (or less) phases. The results can be generalized to more phases by using more level set functions. For m level set functions, m vertices in the graph will be associated to each grid point. Since the data term then would involve interactions between m binary variables, we expect submodularity to be more restrictive. We plan to investigate how submodularity is related to the constant values in these cases, and extend the non-submodular algorithm to this setting. On the other hand, four phases suffices in theory to segment any 2D image by the four color theorem. Therefore, algorithms can alternatively be designed to take advantage of this, which makes extensions to more than four phases unnecessary.

A Proofs of Prop 1-4

A.1 Proof of Prop 1

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

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

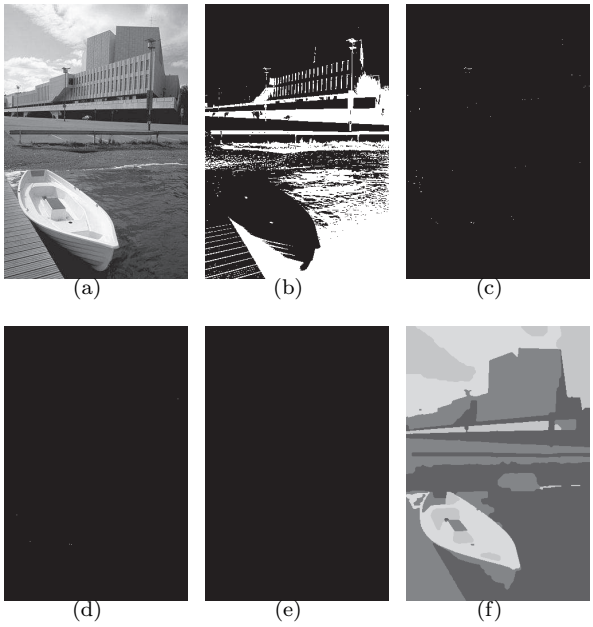


Figure 12: Segmentation with L^1 data fidelity ($\beta = 1$) and very high regularization ν: (a) Input image; (b) Set of pixels $\mathcal{P}^1 \cup \mathcal{P}^2$ where the submodular condition was not satisfied; (c) Set of pixels where the residual capacity condition (34) was violated; (c) -(e) Set of pixels where the weights on graph $\mathcal{G}_i$ differ from the weights on the graph $\mathcal{G}_{i-1}$, $i = 1, ..., 3$; (f) Output (global solution).

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

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

□

A.2 Proof of Prop 3

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

$$|I - c_1|^\beta \leq |I - c_2|^\beta \leq |I - c_3|^\beta \leq |I - c_4|^\beta. \tag{37}$$

Together with (10), this implies

$$0 < |I - c_2|^\beta - |I - c_1|^\beta \leq |I - c_4|^\beta - |I - c_3|^\beta. \tag{38}$$

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

$$|I - c_4|^\beta = \theta_1 |I - c_3|^\beta \quad \text{and} \quad |I - c_2|^\beta = \theta_2 |I - c_1|^\beta. \tag{39}$$

For $\alpha \geq \beta$

$$|I - c_4|^\alpha = \theta_1^{\alpha-\beta} |I - c_3|^\alpha \quad \text{and} \quad |I - c_2|^\alpha = \theta_2^{\alpha-\beta} |I - c_1|^\alpha, \tag{40}$$

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

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

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

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

A.3 Proof of Prop 4

Proof. Assume first $I > c_3$, then

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

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

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

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

$$\theta^\beta \geq 2, \qquad \forall \beta \geq \mathcal{B}.$$

Then

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

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 $\mathcal{B}_I^2 \in \mathbb{N}$ such that

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

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

Therefore the condition (10) is satisfied for any $I \in [0, L]$ by choosing $\beta \geq \mathcal{B} = \max_{I \in [0,L]} \max\{\mathcal{B}_I^1, \mathcal{B}_I^2\}$. □

A.4 Proof of Prop 2

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

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

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

For $I \geq c_3$

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

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

References

[1] *http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/.*

[2] E. Bae and X. Tai, *Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation*, in Scale Space and Variational Methods in Computer Vision, Springer, 2009, pp. 1–13.

[3] Y. Boykov and G. Funka-Lea, *Graph cuts and efficient n-d image segmentation*, Int. J. Comput. Vision, 70 (2006), pp. 109–131.

[4] Y. Boykov and V. Kolmogorov, *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, 2001, pp. 359–374.

[5] ——, *Computing geodesics and minimal surfaces via graph cuts*, in ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, 2003, IEEE Computer Society, p. 26.

[6] Y. Boykov, V. Kolmogorov, D. Cremers, and A. Delong, *An integral solution to surface evolution pdes via geo-cuts*, ECCV06, (2006), pp. 409–422.

[7] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*, in ICCV (1), 1999, pp. 377–384.

[8] X. Bresson and T. Chan, *Non-local unsupervised variational image segmentation models*, UCLA, Applied Mathematics, CAM-report-08-67, (2008).

[9] A. Chambolle, *Total variation minimization and a class of binary mrf models*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer Berlin / Heidelberg, 2005, pp. 136–152.

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

[11] J. Darbon, *A note on the discrete binary mumford-shah model*, Proceedings of Computer Vision/Computer Graphics Collaboration Techniques, (MIRAGE 2007), LNCS Series vol. 44182, (March 2007), pp. 283–294.

[12] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation part i: Fast and exact optimization*, J. Math. Imaging Vis., 26 (2006), pp. 261–276.

[13] ———, *Image restoration with discrete constrained total variation part ii: Levelable functions, convex priors and non-convex cases*, J. Math. Imaging Vis., 26 (2006), pp. 277–291.

[14] A. Dervieux and F. Thomasset, *A finite element method for the simulation of a Rayleigh-Taylor instability*, in Approximation methods for Navier-Stokes problems (Proc. Sympos., Univ. Paderborn, Paderborn, 1979), vol. 771 of Lecture Notes in Math., Springer, Berlin, 1980, pp. 145–158.

[15] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, in Readings in uncertain reasoning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990, pp. 452–472.

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

[17] H. Ishikawa and D. Geiger, *Segmentation by grouping junctions*, in CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 1998, IEEE Computer Society, pp. 125–131.

[18] V. Kolmogorov and C. Rother, *Minimizing nonsubmodular functions with graph cuts-a review*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 1274–1279.

[19] V. Kolmogorov and R. Zabih, *Multi-camera scene reconstruction via graph cuts*, in ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part III, London, UK, 2002, Springer-Verlag, pp. 82–96.

[20] ——, *What energy functions can be minimized via graph cuts?*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26 (2004), pp. 147–159.

[21] N. Komodakis, G. Tziritas, and N. Paragios, *Fast, approximately optimal solutions for single and dynamic mrfs*, Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, (17-22 June 2007), pp. 1–8.

[22] D. F. L. Ford, *Flows in networks*, Princeton University Press, 1962.

[23] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnorr, *Convex multi-class image labeling by simplex-constrained total variation*, in SSVM09, 2009, pp. 150–162.

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

[25] ——, *A variant of the level set method and applications to image segmentation*, Math. Comp., 75 (2006), pp. 1155–1174 (electronic).

[26] J. Lie, M. Lysaker, and X.-C. Tai, *A variant of the level set method and applications to image segmentation*, UCLA, Applied Mathematics, CAM-report-03-50, (2003).

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

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

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

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

[31] C. Rother, S. Kumar, V. Kolmogorov, and A. Blake, *Digital tapestry*, in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1, Washington, DC, USA, 2005, pp. 589–596.

[32] F. R. D. Velasco, *Thresholding using the ISODATA clustering algorithm*, IEEE Trans. Systems Man Cybernet., 10 (1980), pp. 771–774.

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

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

[35] N. E. Zehiry, S. Xu, P. Sahoo, and A. Elmaghraby, *Graph cut optimization for the mumford-shah model*, in Proceedings of the Seventh IASTED International Conference visualization, imaging and image processing, Springer Verlag, 2007, pp. 182–187.