EFFICIENT GLOBAL OPTIMIZATION FOR THE MULTIPHASE CHAN-VESE MODEL OF IMAGE SEGMENTATION BY GRAPH CUTS *

EGIL BAE† AND 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 for this model by representing the phases by several overlapping level set functions. Recently, binary level set functions were proposed as a variant representation of the Chan-Vese model. 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 the binary level set representation of multiphase Chan-Vese model based on graph cuts. If the average intensity values of the different phases are sufficiently evenly distributed, the energy function becomes submodular. Otherwise, a novel method for minimizing nonsubmodular functions is proposed with particular emphasis on this energy function. We also show that non-local extensions of the multiphase Chan-Vese model can be globally minimized by our method.

1. Introduction. Multiphase image segmentation is a fundamental problem in image processing. Variational models such as Mumford-Shah [31] are powerful for this task, but efficient numerical computation of the global minimum is a big challenge. The level set method [15, 33] is a powerful tool which can used for numerical realization. It was first proposed for the Mumford-Shah model in [11] for two phases and [36] for multiple phases. This approach still has the disadvantage of slow convergence and potential of getting stuck in a local minima.

Graph cuts from combinatorial optimization [16, 7, 19, 4, 24, 25, 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, 13, 14]. 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 [12, 37], 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 α-expansion algorithms [7]. Recently, also convex formulations of the continuous multiphase problem have been made in [34, 27] 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 to globally and efficiently minimize the Mumford-Shah model in the multiphase level set framework of Vese and Chan [36] by using binary level set functions as in [28]. Since the length term is slightly approximated in this framework, global minimization is no longer NP hard. We will construct a graph such that the discrete variational problem can be minimized exactly by finding the minimum cut on the graph. However, the energy function may not be submodular

*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).

if the average intensity values of the phases are distributed very unevenly. To handle these cases, we have developed a method for minimizing non-submodular functions with particular emphasis on our energy function. The minimization is global if these values are fixed. A local minimization approach for determining these values is also proposed.

Note that in contrast to α-expansion, the approximation is done in the model rather than in the minimization method. Experimental comparison with alpha expansion is out of the scope of this paper. What can be said is that our method is certainly a lot faster. It is also straight forwardly generalizable to non-local measurements of the curve lengths as was done for two phases in [8].

In this work we focus on the case of 4 or less phases, but aim to generalize the results 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 by assigning different constant values to each disconnect component of the phases.

While image segmentation is the focus of this paper, the method we propose can potentially also be applied for other variational multiphase problems outside of image processing.

The paper is organized as follows: Section 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. The submodular case is presented in Section 2.2, while the non-submodular case is presented in Section 2.3. Section 3 extends the model and minimization method to the non-local setting. Section 4 presents a local minimization approach for determining the average intensity values of the phases, while numerical experiments are presented in Section 5.

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 [31] 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

$$\min_{\{c_i\},\{\Omega_i\}} E_{MS}(\{c_i\},\{\Omega_i\}) = \sum_{i=1}^{n} \int_{\Omega_i} |c_i - u^0|^\beta dx + \sum_{i=1}^{n} \nu \int_{\partial\Omega_i} ds, \tag{1.1}$$

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 [11, 36] 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

$$\min_{\phi,c_1,c_2} \int_\Omega (|\nabla H(\phi)| + \lambda\{H(\phi)|c_1 - u^0|^\beta + (1 - H(\phi))|c_2 - u^0|^\beta\}dx, \tag{1.2}$$

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 1.1. The energy functional could then be written

$$\min_{\phi^1,\phi^2,c_1,...,c_4} E_{CV}(\phi^1,\phi^2,c_1,...,c_4) = \nu \int_\Omega |\nabla H(\phi^1)| + \nu \int_\Omega (|\nabla H(\phi^2)| \tag{1.3}$$

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

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

Note that the length term in (1.1) is slightly approximated, since some of the boundaries are counted twice. Note also that we have made a permutation in the interpretation of the phases compared to [36]. The energy is still exactly the same for all possible solutions. This permutation is crucial for making the corresponding discrete energy function submodular.

The functional in this variational problem 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 [30, 28], 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 1.1. This resulted in the energy functional

$$\min_{\phi^1,\phi^2\in D,c_1,...,c_4} E_{CV}(\phi^1,\phi^2,c_1,...,c_4) = \nu \int_\Omega |\nabla \phi^1| dx + \nu \int_\Omega |\nabla \phi^2| dx + E^{data}(\phi^1,\phi^2), \tag{1.4}$$

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 [30, 28], 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 [29].

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

$$\min_{\phi\in D,c_1,c_2} \nu \int_\Omega |\nabla \phi| dx + \{\phi|c_1 - u^0|^\beta + (1-\phi)|c_2 - u^0|^\beta\}dx. \tag{1.5}$$

The idea, presented in [32] 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

	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 1.1

Representation of four phases by traditional and binary level set functions (note: a little permutation compared to the original paper [36]).

threshold in $[0, 1]$ yields the optimal solution within D. Since (1.5) is convex, this procedure would yield the globally optimal solution.

One could immediately think the same idea could be extended to the multiphase case by iteratively minimizing (1.4) 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 cuts has the disadvantage of some metrification 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]. This will be discussed in Section 2.4 by using a different regularization term. In the next section we will propose a method which globally minimizes (1.4) 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.

2. Graph cut minimization. We will discretize the problem (1.4) and show that this discrete problem can be minimized globally by finding the minimum cut on a specially designed graph. This is possible when the constant values $c_1, ..., c_4$ are sufficiently evenly distributed. We show that such a distribution makes the discrete energy function sub-modular. The evenness criterion will soon be defined more clearly. We have observed that this criterion makes sense for most practical images. Nevertheless, we later develop an algorithm for minimizing non-submodular functions with particular emphasize on functions of the form (1.4).

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 [26], there are several fast algorithms for this problem. It was introduced as a computer vision tool by Greig et. al. [19] in connection with markov random fields [16] and has later been studied by Kolmogorov et. al. [4, 24]. Its applications range from stereo vision [23], segmentation [3, 21, 37, 12] to noise removal [13, 14, 10].

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 [26] 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) [24, 16], 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. Discretization of energy functional. Instead of discretizing the Euler-Lagrange equations, we will discretize the variational problem (1.4). In the next section we show how to minimize the resulting discrete energy function exactly. Let us first mention there are two variants of the total variation term. The isotropic variant, by using 2-norm

$$TV_2(\phi) = \int_\Omega |\nabla\phi|_2 \, dx = \int_\Omega \sqrt{|\phi_{x_1}|^2 + |\phi_{x_2}|^2} \, dx \tag{2.1}$$

and the anisotropic variant, by using 1-norm

$$TV_1(\phi) = \int_\Omega |\nabla\phi|_1 \, dx = \int_\Omega |\phi_{x_1}| + |\phi_{x_2}| \, dx. \tag{2.2}$$

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

$$TV_{1,\frac{\pi}{4}}(\phi) = \frac{1}{2}\int_\Omega \left\{|\nabla\phi(x)|_1 + |R_{\frac{\pi}{4}}\nabla\phi(x)|_1\right\} \, dx, \tag{2.3}$$

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

$$\min_{\phi^1,\phi^2\in D, c_1,\ldots,c_4} E_d(\phi^1, \phi^2, c_1, ..., c_4) = \nu \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\phi_p^1 - \phi_q^1| + \nu \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p^k} w_{pq}|\phi_p^2 - \phi_q^2| \tag{2.4}$$

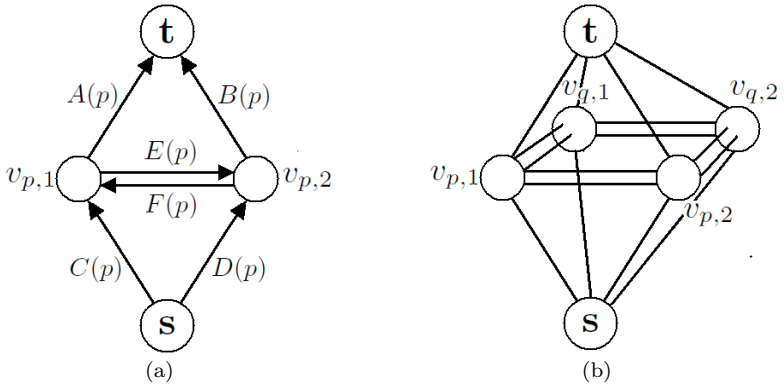


FIGURE 2.1. *(a) The graph corresponding to the data term at one grid point p. (b) A sketch of the graph corresponding to the energy function of a 1D signal of two grid points p and q*

$$+\sum_{p\in\mathcal{P}} E_p^{data}(\phi_p^1,\phi_p^2),$$

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].

2.3. Graph construction. 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 (2.4).

$$\min_{(\mathcal{V}_s,\mathcal{V}_t)} c(\mathcal{V}_s,\mathcal{V}_t) = \min_{\phi^1,\phi^2} E_d(\phi^1,\phi^2,c_1,...,c_4) + \sum_{p\in\mathcal{P}}\sigma_p. \tag{2.5}$$

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 corresponds to each of the level set functions ϕ^1 and ϕ^2. Hence the set of vertices is formally defined as

$$\mathcal{V} = \{v_{p,i} \mid \ p\in\mathcal{P},\ \ i=1,2\}\cup\{s\}\cup\{t\}. \tag{2.6}$$

The edges are constructed such that the relationship (2.5) is satisfied. We begin with the edges constituting the data term of (2.4). For each grid point $p\in\mathcal{P}$ they are defined as

$$\mathcal{E}_D(p) = (s,v_{p,1})\cup(s,v_{p,2})\cup(v_{p,1},t)\cup(v_{p,2},t)\cup(v_{p,1},v_{p,2})\cup(v_{p,2},v_{p,1}). \tag{2.7}$$

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

$$\mathcal{E}_R = \{(v_{p,1},v_{q,1}),(v_{p,2},v_{q,2})\ \forall p,q\subset\mathcal{P}\text{ s.t.}q\in\mathcal{N}_p^k\}. \tag{2.8}$$

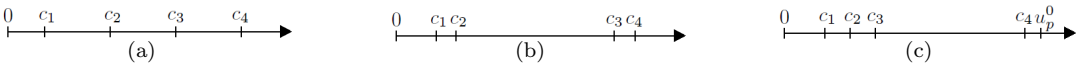


FIGURE 2.2. *(a) and (b) distributions of* $\mathbf{c}$ *which makes energy function submodular for all* β*. (c) distribution of* $\mathbf{c}$ *which may make energy function nonsubmodular for small* β

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

$$\phi_p^1 = \begin{cases} 1 & \text{if } v_{p,1} \in V_s, \\ 0 & \text{if } v_{p,1} \in V_t, \end{cases} \qquad \phi_p^2 = \begin{cases} 1 & \text{if } v_{p,2} \in V_s, \\ 0 & \text{if } v_{p,2} \in V_t. \end{cases} \tag{2.9}$$

Weights are assigned to the edges such that the relationship (2.5) is satisfied. Weights on the regularization edges are simply given by

$$c\left(v_{p,1}, v_{q,1}\right) = c\left(v_{q,1}, v_{p,1}\right) = c\left(v_{p,2}, v_{q,2}\right) = c\left(v_{q,2}, v_{p,2}\right) = \nu w_{pq}, \qquad \forall p \in \mathcal{P},\ q \in \mathcal{N}_p^k. \tag{2.10}$$

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

$$\begin{cases} A(p) + B(p) & = |c_2 - u_p^0|^\beta + \sigma_p \\ C(p) + D(p) & = |c_3 - u_p^0|^\beta + \sigma_p \\ A(p) + E(p) + D(p) & = |c_1 - u_p^0|^\beta + \sigma_p \\ B(p) + F(p) + C(p) & = |c_4 - u_p^0|^\beta + \sigma_p \end{cases} \tag{2.11}$$

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 $p \in \mathcal{P}$. Hence, the following condition on the constant values $c_1, ..., c_4$ must be satisfied

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta, \qquad \forall\ I \in [0, L], \tag{2.12}$$

where L is the maximum intensity value. This condition can be seen in the light of submodular energy functions [24, 16]. In fact, the pairwise term $\sum_{p \in \mathcal{P}} E_p^{data}(\phi_p^1, \phi_p^2)$ is submodular if and only if the condition (2.12) is satisfied.

Let us analyze this condition further. We assume the constant values are ordered increasingly $0 \leq c_1 < c_2 < c_3 < c_4$. The condition says something about how evenly $\{c_i\}_{i=1}^4$ are distributed. Here are some observations.

LEMMA 2.1. *Let $0 \leq c_1 < c_2 < c_3 < c_4$. (2.12) is satisfied for all $I \in [c_2, c_3]$ for any $\beta \in \mathbb{N}$ (natural numbers).*

Proof. Let $c_2 \leq I \leq c_3$. Then, clearly

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

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

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

□

LEMMA 2.2. *Let $0 \leq c_1 < c_2 < c_3 < c_4$. There exists a $\mathcal{B} \in \mathbb{N}$ such that (2.12) is satisfied for any $\beta \geq \mathcal{B}$.*

Proof. Assume first $I > c_3$, then

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

Therefore, there exists a const > 1 s.t.

$$|I - c_1| = \text{const}\, |c_2 - I|.$$

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

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

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 Lemma 2.1, e.g. $\mathcal{B} = 1$.

Therefore the condition (2.12) is satisfied for any $I \in [0, L]$ by choosing $\beta \geq \mathcal{B} = \max_{I \in [0,L]} \max\{\mathcal{B}_I^1, \mathcal{B}_I^2\}$.□

So (2.12) becomes less strict for larger β. In fact we have observed that for $\beta = 2$, (2.12) is realistic for most practical images.

The possibility that (2.12) is not satisfied may happen in two situations: If c_1, c_2, c_3 are close to each other compared to c_4 and intensity u_p^0 is close to c_4, or 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, see Figure 2.2.

Let us go back to the linear system (2.11), with restriction $E(p), F(p) \geq 0$. Assuming (2.12) holds, this has infinitely many solutions. It was shown in [24] 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 $\mathcal{E}_D(p)$ becomes zero for each $p \in \mathcal{P}$. The construction of the solution is as follows

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

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

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

The value σ_p is given implicitly by this solution.

Therefore, by analyzing the complexity of our method in the augmented paths framework, it is easily seen that the cost of our method 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 $F(p) = \infty$.

Remark 1. *In this work, we concentrate on multiphase image segmentations problems. We want to emphasis that the graph constructed here can be used for other multiphase problems besides image segmentation problem.*

2.4. Minimization of non-submodular energy functions. In the last section, we have observed that the energy function (2.4) is submodular if $c_1, ..., c_4$ satisfies (2.12). Although this is realistic for most images, we will develop a method for minimizing nonsubmodular functions with particular emphasis on nonsubmodular terms of the kind encountered here. Minimization of non-submodular functions via graph cuts has been investigated previously, see [22] for a review. The usual idea is to develop a method for determining most of the variables in the energy function, while leaving some of the variables undetermined. In our approach, we instead aim to determine all the variables. Even when (2.12) does not hold, the energy function is "almost submodular", which we believe explains why the following very efficient algorithms works so well in practice.

Consider now the situation

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

for some $p \in \mathcal{P}$. In this case the linear system (2.11) 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

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

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

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

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

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

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

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

in which case $F(p) < 0$. Remember that by Lemma 2.1, the condition holds whenever $u_p^0 \in [c_2, c_3]$.

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. In the extreme case of negative weight on all edges, this becomes equivalent to the max-cut on a graph with negated edge weights. The first step of our procedure finds a good feasible solution, and therefore also a good upper bound on the objective function (2.4). Very often this feasible solution is in fact the optimal solution. All edges of negative weight will be removed, resulting in a new graph $\overline{\mathcal{G}}$. The motivation is as follows. The previous section discussed the possibility of condition (2.12) not being satisfied. In this case 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 negative edge with $E(p) < 0$, 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 this 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 the minimum cut on $\overline{\mathcal{G}}$ to be almost identical to the minimum cut on $\mathcal{G}$. For ease of notation, we define the sets

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

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

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

THEOREM 2.3. *Let $\mathcal{G}$ be a graph as defined in (2.6)-(2.8) and (2.10), with weights $A(p), B(p), C(p), D(p), E(p), F(p)$ satisfying (2.11). 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

$$R_A(p) + R_D(p) \geq -E(p), \quad \forall p \in \mathcal{P}^1 \quad \textit{and} \quad R_B(p) + R_C(p) \geq -F(p), \quad \forall p \in \mathcal{P}^2, \tag{2.13}$$

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

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

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

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

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

Then min-cut($\underline{G}$) $\leq$ min-cut($\mathcal{G}$) $\leq$ min-cut($\overline{\mathcal{G}}$). The max flow on $\overline{\mathcal{G}}$ is feasible on $\underline{\mathcal{G}}$ and therefore also optimal. 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 (2.13), we can check whether the solution is optimal on $\mathcal{G}$. We have observed that it is often possible to stop at this stage. However, if (2.13) does not hold for all pixels one could either accept the solution as suboptimal, or make use of the following algorithm, which is designed to handle such cases. The idea is to create a succession of graphs $\{\mathcal{G}_i\}_{i=1}^n$ with only positive edge weights, such that min-cut($\mathcal{G}_i$) $\leq$ min-cut($\overline{\mathcal{G}}$) for all i, min-cut($\mathcal{G}_0$) = min-cut($\overline{\mathcal{G}}$) and min-cut($\mathcal{G}_n$) = min-cut($\mathcal{G}$). For a given flow we define two new sets $\mathcal{P}_0^1 \subseteq \mathcal{P}^1$ and $\mathcal{P}_0^2 \subseteq \mathcal{P}^2$

$$\mathcal{P}_0^1 = \{p \in \mathcal{P}^1 \mid R_A(p) + R_D(p) < -E(p)\},$$

$$\mathcal{P}_0^2 = \{p \in \mathcal{P}^2 \mid R_B(p) + R_C(p) < -F(p)\}.$$

The graphs $\mathcal{G}_i$ are constructed such that the minimum cut problems on $\mathcal{G}_i$ are relaxations of the minimum cut problem on $\overline{\mathcal{G}}$. Particularly, for each $p \in \mathcal{P}_0^1$ and each $p \in \mathcal{P}_0^2$, the cost of one of the 4 possible phase assignments is reduced, while the rest of the assignment costs are correct (including the one that was set too high in $\overline{\mathcal{G}}$). The cut on $\mathcal{G}_i$ is feasible if no $p \in \mathcal{P}_0^1 \cup \mathcal{P}_0^2$ is assigned to a phase of reduced cost. The algorithm is iterated until the cut on $\mathcal{G}_i$ becomes feasible.

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 $p \in \mathcal{P}_0^1$
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 $p \in \mathcal{P}_0^2$
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 $\mathcal{P}_0^1$ and $\mathcal{P}_0^2$ by examining residual capacities in $\mathcal{G}_{i+1}$
4. $i \leftarrow i + 1$
}

THEOREM 2.4. *Let $\mathcal{G}_n$ be the graph at termination of Algorithm 1. Then min-cut($\mathcal{G}_n$) = min-cut($\mathcal{G}$).*

Proof. If the algorithm terminates with $\mathcal{G}_0$, optimality was proved in theorem 2.3. Assume therefore $n \geq 1$. The proof follows some of the same ideas as the proof of theorem 2.3. 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 (residuals R obtained from flow on $\mathcal{G}_n$)

$$c(v_{p,1}, t) = A(p) - R_A(p), \quad \forall p \in \mathcal{P}^1 \backslash \mathcal{P}_0^1,$$

$$c(s, v_{p,2}) = D(p) - R_D(p), \quad \forall p \in \mathcal{P}^1 \backslash \mathcal{P}_0^1$$

$$c(v_{p,2}, t) = B(p) - R_B(p), \quad \forall p \in \mathcal{P}^2 \backslash \mathcal{P}_0^2,$$

$$c(s, v_{p,1}) = C(p) - R_C(p), \quad \forall p \in \mathcal{P}^2 \backslash \mathcal{P}_0^2.$$

Then min-cut$(\underline{\mathcal{G}}) \leq$ min-cut$(\mathcal{G}) \leq$ min-cut$(\mathcal{G}_n)$. 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$(\mathcal{G}_n)$.□

Observe that 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.

Note also that it is possible to prove that potential edges of negative weight can only go in one direction. That is, if $E(p) < 0$ for some $p \in \mathcal{P}$, then $F(p) \geq 0$ for all $p \in \mathcal{P}$. If $F(p) < 0$ for some $p \in \mathcal{P}$, then $E(p) \geq 0$ for all $p \in \mathcal{P}$.

We are trying to develop a convergence theory for this algorithm. Numerical experiments indicate that convergence is fast and no oscillations occur. We have so far investigated convergence experimentally by applying the algorithm to all images from the segmentation database [1]. We have used both the L^1 and L^2 data fidelity term, and different values on the regularization parameter ν, always resulting in convergence in an average of 3-4 iterations. Let us point out that Algorithm 1 was very rarely needed. When we used L^2 norm in data fidelity ($\beta = 2$), $\mathcal{P}_0^1$ or $\mathcal{P}_0^2$ was usually empty, but occasionally contained a very small subset of $\mathcal{P}^1$ or $\mathcal{P}^2$. In these cases the algorithm converged in 1-3 iterations. When using L^1 norm ($\beta = 1$) the subsets $\mathcal{P}_0^1$ or $\mathcal{P}_0^2$ were comparatively a little larger.

We observed that pathological cases could be created by setting ν unnaturally high, since then the sets $\mathcal{P}_0^1$ or $\mathcal{P}_0^2$ became larger. In order to verify the convergence of the algorithm, we have therefore also tested it on these pathological cases by using L^1 norm and setting ν very high. After going through all images from the database [1], we observed that it always converged. The average number of iterations was 9 (when it was needed), and the maximum number of iterations was 30.

3. Extension to non-local operators and edge detectors. We have so far only considered euclidian measurements of the length term in the Chan-Vese model. Recently, powerful extensions of this model with two phases have been made by using non-local operators [17, 8] or active contours with edge detectors [9]. Both of these extensions use a priori information of the image in the length measure. These concepts can now elegantly be generalized to multiple phases by using our global minimization approach. The multiphase Chan-Vese model with edge detector can be written

$$\min_{\phi^1,\phi^2\in D,c_1,...,c_4} E(\phi^1,\phi^2,c_1,...,c_4) = \nu\int_\Omega g(|\nabla u^0(x)|)|\nabla\phi^1|dx+\nu\int_\Omega g(|\nabla u^0(x)|)\,|\nabla\phi^2|dx+ \\ E^{data}(\phi^1,\phi^2), \tag{3.1}$$

where $g(s)$ is a decreasing function, e.g. $g(s) = \frac{1}{s^2+\varepsilon}$, and $E^{data}(\phi^1,\phi^2)$ is given as before. It forces the boundaries of the phases to locations of high gradients, the edges in the image. Another interesting extension is the use of non-local differential operators in the above functional. Non-local filters and operators have recently gained huge popularity in image denoising for their ability to simultaneously denoise smooth and textured regions [2, 18, 20]. Recently its power has also been demonstrated for image segmentation [17, 8]. In [8], the two phase Chan-Vese model was extended the non-local setting, and was shown to better segment fine and small structures. Both the data term and regularization term could be extended. Here we focus on the regularization term. As in [8] we define a non-local version of the gradient $\nabla_{NL}\phi : \Omega\times\Omega \mapsto \mathbb{R}$ of ϕ between any pairs of points $(x,y)\in\Omega\times\Omega$

$$\nabla_{NL}\phi(x,y) = (\phi(x)-\phi(y))\sqrt{w(x,y)}, \tag{3.2}$$

where w is some weight function. The magnitude of the gradient is defined as its inner product with itself

$$|\nabla_{NL}\phi(x)| = \sqrt{\int_\Omega(\phi(x)-\phi(y))^2 w(x,y)dy}. \tag{3.3}$$

Hence the non-local total variation can then be defined as

$$\int_\Omega|\nabla_{NL}\phi|dx = \int_\Omega\sqrt{\int_\Omega(\phi(x)-\phi(y))^2 w(x,y)dy}\,dx \tag{3.4}$$

The work of [8] proposed a non-local extension of the two phase Chan-Vese model with this operator

$$\min_{\phi\in D,c_1,c_2}\nu\int_\Omega|\nabla_{NL}\phi|dx+\{\phi|c_1-u^0|^\beta+(1-\phi)|c_2-u^0|^\beta\}dx. \tag{3.5}$$

They showed that this functional could be globally minimized (for fixed c_1,c_2) by using continuous or discrete graph cut. We are now able to generalize this result to multiple phases. The multiphase non-local Chan-Vese functional can be written

$$E_{NL-CV}(\phi^1,\phi^2,c_1,...,c_4) = \nu\int_\Omega|\nabla_{NL}\phi^1|dx+\nu\int_\Omega|\nabla_{NL}\phi^2|dx+E^{data}(\phi^1,\phi^2). \tag{3.6}$$

We will show that a discretized version of this functional can be minimized globally via graph cuts. Remember that since $E^{data}(\phi^1,\phi^2)$ is not convex, continuous graph cuts cannot be used.

3.1. Discretization and graph representation. The discrete versions of (3.1) and (3.6) have the same form as (2.4), which is repeated here for convenience

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

$$+\sum_{p\in\mathcal{P}} E_p^{data}(\phi_p^1,\phi_p^2),$$

but the weights w_{pq} differs. In case of edge detector (3.1) they are given as

$$w_{pq} = \nu\frac{2\delta^2(|\nabla_d u^0(p)| + |\nabla_d u^0(q)|)}{k||p-q||_2} \quad \forall p\in\mathcal{P}, q\in\mathcal{N}_p^k,$$

where ∇_d is a discrete gradient. The edges are defined as before in Section 2.3. In case of the non-local functional, the neighborhood system $\mathcal{N}_p^k$ is not local in the spatial sense, but may potentially contain all points in $\mathcal{P}$.

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

$$+\sum_{p\in\mathcal{P}} E_p^{data}(\phi_p^1,\phi_p^2),$$

The weights can be chosen in the discrete setting as in [2, 8]

$$w_{pq} = \exp(-\frac{||U_p^0 - U_q^0||_{L^2}}{h}), \qquad \forall (p,q)\in\mathcal{P}\times\mathcal{P},$$

where U_p^0 and U_q^0 are patches of the image u^0 centered at pixels p and q respectively, and h is a scale parameter. The points p and q are interpreted as close if their image patches are similar. The regularization edges and weights are now given as

$$c\left(v_{p,1},v_{q,1}\right) = c\left(v_{q,1},v_{p,1}\right) = c\left(v_{p,2},v_{q,2}\right) = c\left(v_{q,2},v_{p,2}\right) = \nu w_{pq}, \qquad \forall (p,q)\in\mathcal{P}\times\mathcal{P}. \tag{3.7}$$

It is not necessary to use all these edges in practice, since that would slow down the computational cost considerably. Only a selected few connections are needed for each vertex. In the two phase segmentation case or image denoising case [2, 8] discussed details on how to select the most relevant edges.

4. 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

$$c_i = \frac{\int_{\Omega_i} u^0\,dx}{\int_{\Omega_i} dx},\ i=1,...,n \tag{4.1}$$

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

	Size	Phases	Gradient descent	Graph Cut
Experiment1	100x100	4	25.3	0.10
Brain	933x736	4	3077	19.4
Experiment 3	100x100	4	–	0.14
Experiment 4	320x480	4	–	10.2
Experiment 5	320x480	4	–	9.9

TABLE 5.1

Computation times in seconds for gradient descent vs graph cut optimization (the computation times when using L^1 data term are not available for gradient descent)

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

$$\phi = arg\, min_{\tilde{\phi}^1, \tilde{\phi}^2} E_d(\mathbf{c}^l, \tilde{\phi}^1, \tilde{\phi}^2). \tag{4.2}$$

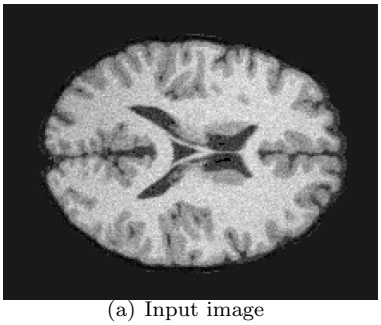
Update $\mathbf{c}^{l+1}$ according to equation (4.1).
Update $l \leftarrow l + 1$

Note that no initialization of the level set function is required. Only the values $\mathbf{c}^0$ need to be initialized, which can be achieved very efficiently by the isodata algorithm [35]. In all our experiments, convergence was reached in 4-12 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.

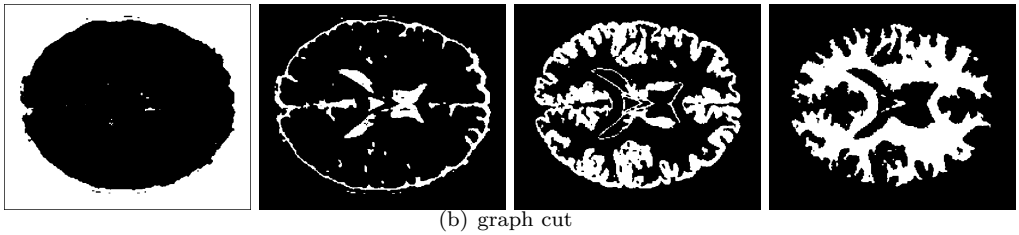
5. Numerical results. Numerical experiments are made to demonstrate the new minimization methods. We also make comparisons between the PDE approach and combinatorial approach for minimizing (1.3). In all results, the phases are depicted as bright regions. In experiment 1 and 2, Figure (5.2) and (5.1), the L^2 norm is used in the data term. The constant values $\{c_i\}_{i=1}^4$ satisfy condition (2.12) initially and in all iterations of Algorithm 2 until convergence.

We next try to use L^1 data fidelity on the image in Figure 5.2 (b). For this modified image, we observe that in the optimal solution, c_1, c_2, c_3 are close to each other compared to c_4. Therefore, as expected, condition (2.12) was not satisfied for all pixels. However, after finding the max flow on $\overline{\mathcal{G}}$ and examining the residual capacities, the criterion 2.13 was satisfied, and hence the global minimum had been obtained. If the constant values had been even closer to each other, the model would merge some of the regions together. See Table 5 for computation times.

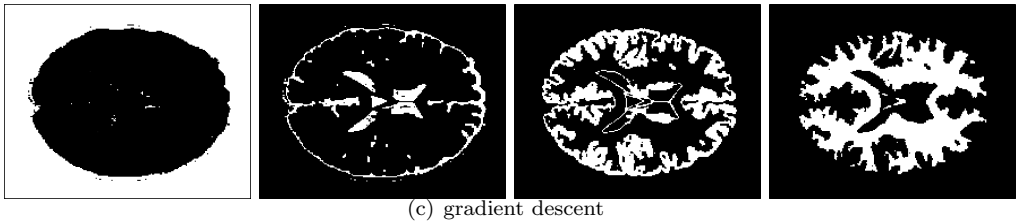
For the next two images, c.f. Figure 5.3(a) and (b), the L^1 norm was used, and for some grid points neither condition (2.12) or the criterion (2.13) was satisfied. Therefore, Algorithm 1 had to be used. For each iteration of Algorithm 2, it converged in 5-8 iterations. See Table 5 for computation times (the cost of Algorithm 1 has been neglected). Because ν is set so high, the results does not look very good. The purpose of these examples is to show that Algorithm 1 can be used to find the global minimum of the energy function in case criterion (2.13) is not satisfied. As already mentioned, we have also tested the convergence of Algorithm 1 experimentally by applying it to all images from the database [1]. This includes pathological with ν set very high. The different constant values in these experiments were given automatically from each iteration of Algorithm 2.



(a) Input image

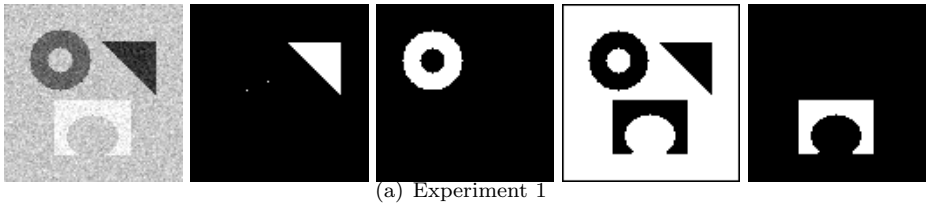


(b) graph cut

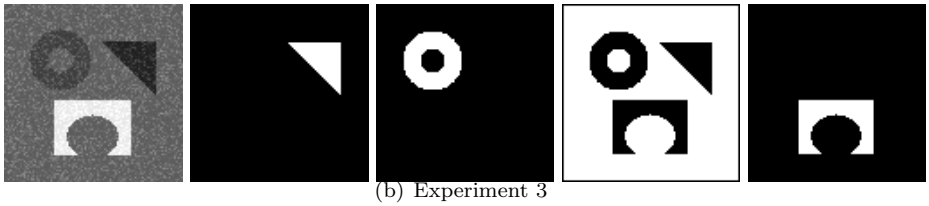


(c) gradient descent

FIGURE 5.1. *Experiment 2: From left to right: phase 1 - phase 4.*

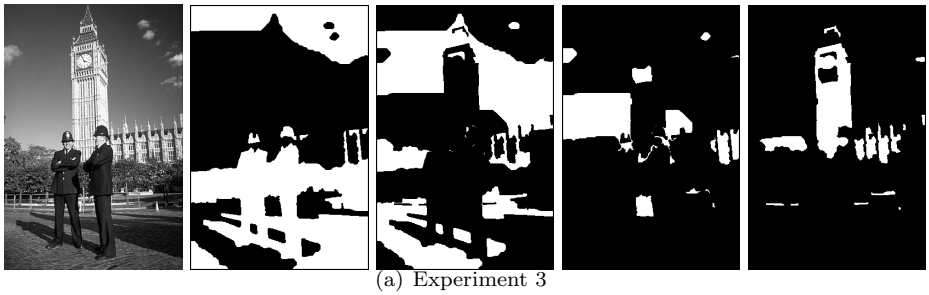


(a) Experiment 1

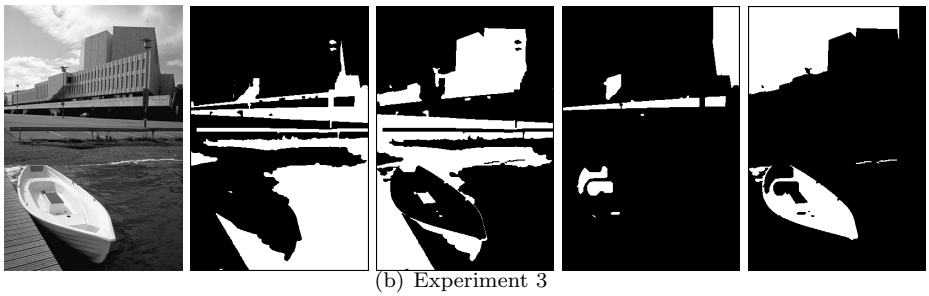


(b) Experiment 3

FIGURE 5.2. *(a) Experiment 1: L^2 data fidelity. (b) Experiment 3: L^1 data fidelity. Note that the constant values of phase 1-3 are very close to each other. From left to right: input image, phase 1, phase 2, ..., phase 4.*



(a) Experiment 3



(b) Experiment 3

FIGURE 5.3. *(a)Experiment 4. (b) Experiment 5. From left to right: input image, phase 1 - phase 4. L*1 *norm*

6. 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. If the average intensity values (constant values) of the phases were sufficiently evenly distributed, the energy function became submodular. A method for minimizing non-submodular was developed in order to handle all possible distributions of the constant values. This method was specially designed for energy functions of our kind. We also showed that non-local extensions of the multiphase Chan-Vese model could be globally minimized by our method.

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 the four color theorem. Therefore, algorithms can alternatively be designed to take advantage of this, which makes extensions to more than four phases unnecessary.

REFERENCES

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

[2] J.-M. M. A. BUADES, B. COLL, *On image denoising methods*, SIAM Multiscale Modeling and Simulation (MMS), 4 (2005), pp. 490–530.

[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 algo-*

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

[10] 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.

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

[12] 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.

[13] 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.

[14] ———, *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.

[15] 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.

[16] 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.

[17] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, SIAM Multiscale Modeling and Simulation (MMS), 6 (2007), pp. 595–630.

[18] ———, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028.

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

[20] S. O. T. C. Guy Gilboa, Jerome Darbon, *Nonlocal convex functionals for image regularization*, UCLA, Applied Mathematics, CAM-report-06-57, (2006).

[21] 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.

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

[23] 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.

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

[25] 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.

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

[27] 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.

[28] 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.

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

[30] 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).
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] F. R. D. Velasco, *Thresholding using the ISODATA clustering algorithm*, IEEE Trans. Systems Man Cybernet., 10 (1980), pp. 771–774.
[36] 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.
[37] 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.