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

Egil Bae[1] and Xue-Cheng Tai[1,2]

[1] Department of Mathematics, University of Bergen, Norway
Egil.Bae@math.uib.no
[2] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
tai@mi.uib.no

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, exactly the same model was also formulated by using binary level set functions. 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 the multiphase Chan-Vese model based on graph cuts. If the average intensity values of the different phases are sufficiently evenly distributed, the discretized energy function becomes submodular. Otherwise, a novel method for minimizing nonsubmodular functions is proposed with particular emphasis on this energy function.

1 Introduction

Multiphase image segmentation is a fundamental problem in image processing. Variational models such as Mumford-Shah [1] are powerful for this task, but efficient numerical computation of the global minimum is a big challenge. The level set method [2,3] is a powerful tool which can used for numerical realization. It was first proposed for the Mumford-Shah model in [4] for two phases and [5] 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 [6,7,8,9,10,11] 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 [12,13,14,15]. 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 [16,17], 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

D. Cremers et al. (Eds.): EMMCVPR 2009, LNCS 5681, pp. 28–41, 2009.

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 [18,19] 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 [5] by using binary level set functions as in [20]. 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 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 [21]. Such a generalization is not obvious for alpha expansion.

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.

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 [1] 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

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

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 [4,5] proposed to represent the above functional with level set functions, and solve the resulting gradient descent equations numerically. 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 function can then be written

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

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

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

Note that the length term in (1) is slightly approximated, since some of the boundaries are counted twice. Note also that we have made a small permutation in the interpretation of the phases compared to [5]. The energy is still exactly identical for all feasible solutions. This permutation is crucial for making the corresponding discrete energy function submodular.

The functional in this variational problem is highly non-convex 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 [20], the same multiphase model was formulated using binary level set functions $\phi^1, \phi^2 \in D = \{\phi |\ \phi : \Omega \mapsto \{0, 1\}\}$, representing the phases as in Table 1. This resulted in the energy functional

$$\min_{\phi^1,\phi^2 \in D,c_1,...,c_4} E(\phi^1, \phi^2, c_1, ..., c_4) = \nu \int_\Omega |\nabla \phi^1| dx + \nu \int_\Omega |\nabla \phi^2| dx + E^{data}(\phi^1, \phi^2), \tag{3}$$

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 constraint D was represented by a polynomials in ϕ^1 and ϕ^2. Minimization was carried out by the augmented lagrangian method. Since both the constraint D and the energy functional is non-convex, global minimization could not be

Table 1. Representation of four phases by traditional and binary level set functions

	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	$\phi^1(x) < 0, \phi^2(x) > 0$	$\phi^1(x) = 0, \phi^2(x) = 1$

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 [22].

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. The idea, first presented in [23] 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. The same idea could also be used to minimize (3). The problem is that (3) is not convex, and hence the algorithm may converge to a local minimum.

In general, discrete graph cuts 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 very easily be generalized to minimize non-local measurements of the curve lengths as was done for two phases in [21], by using regularization term

$$\nu \int_\Omega |\nabla_{NL}\phi^1| dx + \nu \int_\Omega |\nabla_{NL}\phi^2| dx.$$

However, that is not the focus of this paper. We will propose a method which globally minimizes (3) 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 (3) 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 (3).

2.1 Brief Overview of Graph Cuts in Computer Vision

Graph cuts were first introduced as a computer vision tool by Greig et. al. [8] in connection with markov random fields [6]

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. A theorem of Ford and Fulkerson [24] 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 [24] can be used for finding minimum cuts in graphs. An efficient implementation of this algorithm specialized for image processing problems can be found in [9]. 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 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) [10,6], i.e.

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

2.2 Discretization of Energy Functional

Instead of discretizing the Euler-Lagrange equations, we will discretize the variational problem (3). 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 $TV_2(\phi) = \int_\Omega |\nabla\phi|_2\, dx = \int_\Omega \sqrt{|\phi_{x_1}|^2 + |\phi_{x_2}|^2}\, dx$, and the anisotropic variant, by using 1-norm $TV_1(\phi) = \int_\Omega |\nabla\phi|_1\, dx = \int_\Omega |\phi_{x_1}| + |\phi_{x_2}|\ dx$. The anisotropic variant is graph representable and will be considered here. More isotropic variants can be derived by splitting the calculation of $TV_1(\phi)$ between several rotated coordinate systems, see [25].

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

$$\min_{\phi^1,\phi^2\in D,c_1,\ldots,c_4} E_d(\phi^1,\phi^2,c_1,...,c_4) = \nu\sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{N}_p^k} w_{pq}|\phi_p^1-\phi_q^1| + \nu\sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{N}_p^k} w_{pq}|\phi_p^2-\phi_q^2| \tag{4}$$

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

where

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

The weights w_{pq} are used to approximate the curve lengths. They can be derived from the continuous functional as in full version [25], or from the Cauchy-Crofton formula as in [12].

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 (4).

$$\min_{(\mathcal{V}_s,\mathcal{V}_t)} c(\mathcal{V}_s,\mathcal{V}_t) = \min_{\phi^1,\phi^2} E_d(\phi^1,\phi^2,c_1,...,c_4) + \sum_{p\in\mathcal{P}}\sigma_p. \tag{5}$$

where $\sigma_p \in \mathbb{R}$ 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

$$\mathcal{V} = \{v_{p,i} \mid\ p\in\mathcal{P},\ \ i=1,2\}\cup\{s\}\cup\{t\}. \tag{6}$$

The edges are constructed such that the relationship (5) is satisfied. We begin with the edges constituting the data term of (4). For each grid point $p \in \mathcal{P}$ they are defined as

$$\mathcal{E}_D(p) = (s,v_{p,1})\cup(s,v_{p,2})\cup(v_{p,1},t)\cup(v_{p,2},t)\cup(v_{p,1},v_{p,2})\cup(v_{p,2},v_{p,1}). \tag{7}$$

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

$$\mathcal{E}_R = \{(v_{p,1},v_{q,1}),(v_{p,2},v_{q,2})\ \forall p,q\subset\mathcal{P}\text{ s.t.}q\in\mathcal{N}_p^k\}. \tag{8}$$

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

$$\phi_p^1 = \begin{cases}1 & \text{if } v_{p,1}\in V_s,\\ 0 & \text{if } v_{p,1}\in V_t,\end{cases} \qquad \phi_p^2 = \begin{cases}1 & \text{if } v_{p,2}\in V_s,\\ 0 & \text{if } v_{p,2}\in V_t.\end{cases} \tag{9}$$

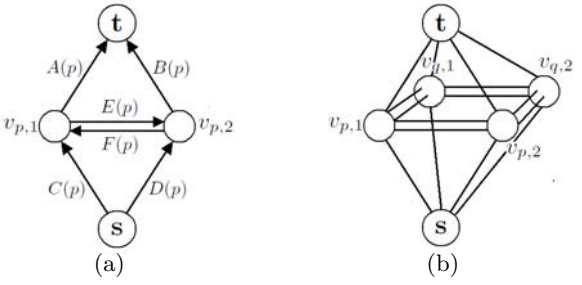


Fig. 1. (a) The graph corresponding to the data term at one grid point p. (b) A sketch of the graph corresponding to the energy function of a 1D signal of two grid points p and q.

Weights are assigned to the edges such that the relationship (5) is satisfied. Weights on the regularization edges are simply given by

$$c\left(v_{p,1}, v_{q,1}\right) = c\left(v_{q,1}, v_{p,1}\right) = c\left(v_{p,2}, v_{q,2}\right) = c\left(v_{q,2}, v_{p,2}\right) = \nu w_{pq}, \qquad \forall (p,q) \in \mathcal{N}. \tag{10}$$

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

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

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

It is clear that these weights must satisfy

$$\begin{cases} A(p) + B(p) & = |c_2 - u_p^0|^\beta + \sigma_p \\ C(p) + D(p) & = |c_3 - u_p^0|^\beta + \sigma_p \\ A(p) + E(p) + D(p) & = |c_1 - u_p^0|^\beta + \sigma_p \\ B(p) + F(p) + C(p) & = |c_4 - u_p^0|^\beta + \sigma_p \end{cases} \tag{11}$$

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

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

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

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

$$|c_2 - I|^\beta + |c_3 - I|^\beta \leq |c_1 - I|^\beta + |c_4 - I|^\beta, \quad \forall\ I \in [0, L], \tag{12}$$

where L is the maximum intensity value. This condition can be seen in the light of submodular energy functions [10,6]. 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 (12) is satisfied.

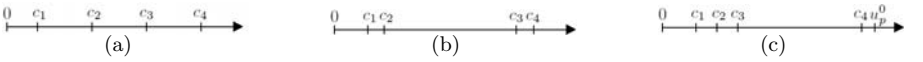


Fig. 2. (a) and (b) distributions of $\mathbf{c}$ which makes energy function submodular for all β. (c) distribution of $\mathbf{c}$ which may make energy function nonsubmodular for small β.

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 is a first observation, the proof of this and the following lemmas can be found in the full version of this work [25]

Lemma 1. *Let $0 \leq c_1 < c_2 < c_3 < c_4$. There exists a $\mathcal{B} \in \mathbb{N}$ such that (12) is satisfied for any $\beta \geq \mathcal{B}$.*

So (12) becomes less strict for larger β. In fact we have observed that for $\beta = 2$, (12) is realistic for most practical images. Here is another observation

Lemma 2. *Let $0 \leq c_1 < c_2 < c_3 < c_4$. (12) is satisfied for all $I \in [c_2, c_3]$.*

The possibility that (12) is not satisfied may happen in two situations: If c_1, c_2, c_3 are very close compared to c_4 and intensity I is close to c_4, or if c_2, c_3, c_4 are very close compared to c_1 and I is close to c_1.

Let us go back to the linear system (11), with restriction $E(p), F(p) \geq 0$. Assuming (12) holds, this has infinitely many solutions.

It was shown in [10] 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 exact construction of the solution can be found in the full version [25]. Hence, at most three edges are required to represent the data term at each grid point. Therefore, by analyzing the complexity of our method in the augmenting 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.

2.4 Minimization of Non-submodular Energy Functions

In the last section, we have observed that the energy function (4) is submodular if $c_1, ..., c_4$ satisfies (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 nonsubmodular functions via graph cuts has been investigated previously, see [26] for a review. The usual idea is to develop a method for determining most of the variables, while leaving some of the variables undetermined. In our approach, we instead aim to determine all the variables. Even when (12) does not hold, the energy function is "almost submodular", which may explain 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 (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. It can be easily seen that if $E(p) < 0$, there exists a solution to the linear system with $F(p) = 0$. Vice versa, if $F(p) < 0$ there exists a solution with $E(p) = 0$. See [25] for the exact construction.

It is difficult interpret physically what is 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 (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 (12) not being satisfied. In this case c_1, c_2, c_3 are close to each other compared to c_4 and I_p 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 $|c_1 - u_p^0|^2 - E(p)$. We therefore expect the minimum cut on $\overline{\mathcal{G}}$ to be almost identical to the minimum cut on $\mathcal{G}$. For easy of notation, we define the sets

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

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 1. *Let $\mathcal{G}$ be a graph as defined in (6)-(8) and (10), with weights $A(p), B(p), C(p), D(p), E(p), F(p)$ satisfying (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

$$R_A(p) + R_D(p) \geq -E(p), \quad \forall p \in \mathcal{P}^1 \quad \text{and} \quad R_B(p) + R_C(p) \geq -F(p), \quad \forall p \in \mathcal{P}^2, \tag{13}$$

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

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

$$c(v_{p,1}, t) = A(p) - R_A(p) \quad \text{and} \quad c(s, v_{p,2}) = D(p) - R_D(p), \qquad \forall p \in \mathcal{P}^1$$

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

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

We have observed that it is often possible to stop at this stage, since (13) is very often satisfied. If not, 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 $\mathcal{G}_0$
while($\mathcal{G}_i \neq \mathcal{G}_{i-1}$ or $i = 0$){
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 $v_{p,2} \in V_t$): set $c(v_{p,1}, t) = A(p) + E(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_s$): set $c(s, v_{p,2}) = D(p) + E(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_t$): set $c(s, v_{p,1}) = A(p) + E(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_s$): set $c(s, v_{p,1}) = D(p) + E(p)$ in $\mathcal{G}_{i+1}$

for all $p \in \mathcal{P}_0^2$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_t$): set $c(v_{p,2}, t) = B(p) + F(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_s$): set $c(s, v_{p,1}) = C(p) + F(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_s$ and $v_{p,2} \in V_t$): set $c(s, v_{p,2}) = B(p) + F(p)$ in $\mathcal{G}_{i+1}$
if($v_{p,1} \in V_t$ and $v_{p,2} \in V_s$): set $c(s, v_{p,2}) = C(p) + F(p)$ in $\mathcal{G}_{i+1}$

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 graph $\mathcal{G}_{i+1}$
4. $i \leftarrow i + 1$
}

Theorem 2. *Let $\mathcal{G}_n$ be the graph at termination of Algorithm 1. Then min-cut($\mathcal{G}_n$) = min-cut($\mathcal{G}$).*

Proof. The proof follows some of the same ideas as the proof of theorem 1. 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 $\overline{\mathcal{G}_n} = \mathcal{G}_{n-1}$, 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 $\overline{\mathcal{G}_n}$)

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

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

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 in $s - t$ paths passing through vertices $v_{p,1}, v_{p,2}$ 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.

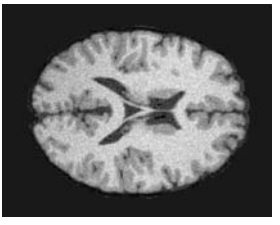
We are trying to develop a convergence theory for this algorithm. Numerical experiments indicate that convergence is fast and no oscillations occur. We have so far investigated convergence experimentally by applying the algorithm to all images from the segmentation database [27]. We have used both the 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. However, by setting ν unnaturally high, pathological cases could be created. In order to verify the convergence of the algorithm, we have also successfully tried these extreme choices of ν.

2.5 Local Minimization Algorithm for Estimating c

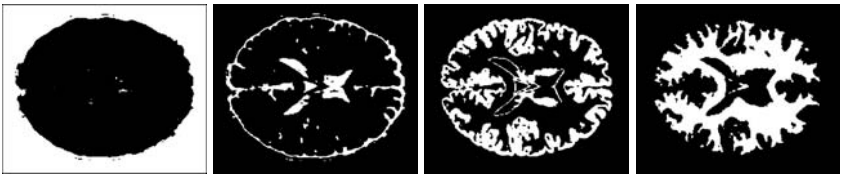
In order to minimize with respect to both ϕ^1, ϕ^2 and $\mathbf{c}$, we alternate between optimization of ϕ^1, ϕ^2 for fixed $\mathbf{c}$ and optimizing $\mathbf{c}$ for fixed ϕ^1, ϕ^2, as explained in more detail in [25]. This algorithm is shown to be robust and typically only require a few iterations, but can of course not be proven to find a global minimum.

3 Numerical Results

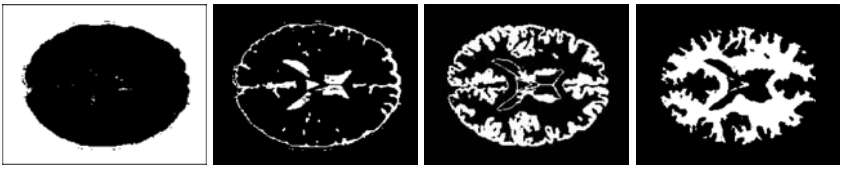
Numerical experiments are made to demonstrate the new minimization methods. We also make comparisons between the PDE approach and combinatorial approach for minimizing (2). In all results, the phases are depicted as bright regions. The values $\mathbf{c}$ used in all experiments are generated from the algorithm in Section 2.5.



(a) Input image



(b) graph cut



(c) gradient descent

Fig. 3. Experiment 2: From left to right: phase 1 - phase 4

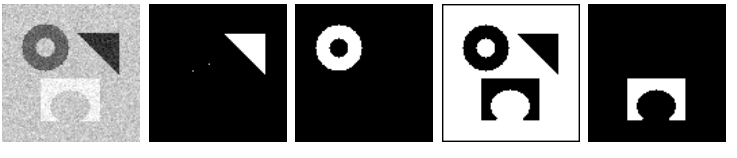


Fig. 4. Experiment 1: L^2 data fidelity

In experiment 1 and 2, Figure (4) and (3), the L^2 norm is used in the data term. The constant values $\{c_i\}_{i=1}^4$ satisfy condition (12) initially and in all iterations until convergence. We next try to use L^1 data fidelity on these images. In this case, condition (12) was not satisfied for all pixels. However, after finding the max flow on $\overline{\mathcal{G}}$ and examining the residual capacities, the criterion (13) was satisfied, and hence the global minimum had been obtained. See Table 2 for computation times.

For the next image, Figure (5), the L^1 norm was used, and for some grid points neither condition (12) nor the criterion (13) was satisfied. Therefore, Algorithm 1 had to be used. For each combination of $\{c_i\}_{i=1}^4$ generated by the algorithm in Section 2.5, it converged in 5-8 iterations. As already mentioned, we have also

Fig. 5. Experiment 3: from left to right: input image, phase 1 - phase 4. L^1 norm

Table 2. Computation times in seconds for gradient descent vs graph cut optimization with $\beta = 2$

	Size	Phases	Gradient descent	Graph Cut
Experiment1	100x100	4	25.3	0.10
Brain	933x736	4	3077	19.4

tested the convergence of Algorithm 1 experimentally by applying it to all images from the database [27]. This includes pathological cases with ν set very high. The different constant values in these experiments were generated by the algorithm in Section 2.5. More experiments can be found in [25].

References

1. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. 42, 577–685 (1989)
2. Dervieux, A., Thomasset, F.: 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). Lecture Notes in Math., vol. 771, pp. 145–158. Springer, Berlin (1980)
3. Osher, S., Sethian, J.: Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. J. Comput. Phys. 79(1), 12–49 (1988)
4. Chan, T., Vese, L.: Active contours without edges. IEEE Image Proc. 10, 266–277 (2001)
5. Vese, L.A., Chan, T.F.: A new multiphase level set framework for image segmentation via the mumford and shah model. International Journal of Computer Vision 50, 271–293 (2002)
6. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In: Readings in uncertain reasoning, pp. 452–472. Morgan Kaufmann Publishers Inc., San Francisco (1990)
7. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. In: ICCV (1), pp. 377–384 (1999)
8. Greig, D.M., Porteous, B.T., Seheult, A.H.: Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, Series B, 271–279 (1989)
9. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 359–374 (2001)

10. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 26(2), 147–159 (2004)
11. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic mrfs. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR 2007, June 17-22, pp. 1–8 (2007)
12. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: ICCV 2003: Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, p. 26. IEEE Computer Society Press, Los Alamitos (2003)
13. Boykov, Y., Kolmogorov, V., Cremers, D., Delong, A.: An integral solution to surface evolution pdes via geo-cuts. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 409–422. Springer, Heidelberg (2006)
14. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part i: Fast and exact optimization. J. Math. Imaging Vis. 26(3), 261–276 (2006)
15. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part ii: Levelable functions, convex priors and non-convex cases. J. Math. Imaging Vis. 26(3), 277–291 (2006)
16. Darbon, J.: A note on the discrete binary mumford-shah model. In: Gagalowicz, A., Philips, W. (eds.) MIRAGE 2007. LNCS, vol. 4418, pp. 283–294. Springer, Heidelberg (2007)
17. Zehiry, N.E., Xu, S., Sahoo, P., Elmaghraby, A.: Graph cut optimization for the mumford-shah model. In: Proceedings of the Seventh IASTED International Conference visualization, imaging and image processing, pp. 182–187. Springer, Heidelberg (2007)
18. Pock, T., Chambolle, A., Bischof, H., Cremers, D.: A convex relaxation approach for computing minimal partitions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, Florida (to appear, 2009)
19. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnorr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: SSVM 2009, pp. 150–162 (2009)
20. Lie, J., Lysaker, M., Tai, X.: A binary level set model and some applications to mumford-shah image segmentation. IEEE Transactions on Image Processing 15(5), 1171–1181 (2006)
21. Bresson, X., Chan, T.: Non-local unsupervised variational image segmentation models (2008)
22. Lie, J., Lysaker, M., Tai, X.: A variant of the level set method and applications to image segmentation. Math. Comp. 75(255), 1155–1174 (2006) (electronic)
23. Nikolova, M., Esedoglu, S., Chan, T.F.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics 66(5), 1632–1648 (2006)
24. Ford, L., Fulkerson, D.: Flows in networks. Princeton University Press, Princeton (1962)
25. Bae, E., Tai, X.C.: Efficient global optimization for the multiphase chan-vese model of image segmentation by graph cuts. UCLA, Applied Mathematics, CAM-report-09-53 (June 2009)
26. Kolmogorov, V., Rother, C.: Minimizing nonsubmodular functions with graph cuts-a review. IEEE Trans. Pattern Anal. Mach. Intell. 29(7), 1274–1279 (2007)
27. `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`