J. Vis. Commun. Image R. 25 (2014) 1472–1488

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A new continuous max-flow algorithm for multiphase image segmentation using super-level set functions

Jun Liu [a,*], Xue-cheng Tai [b], Shingyu Leung [c], Haiyang Huang [a]

[a] School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry Education, Beijing 100875, PR China
[b] Department of Mathematics, University of Bergen, Johaness Brunsgate 12, 5007 Bergen, Norway
[c] Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

ARTICLE INFO

Article history:
Received 3 September 2013
Accepted 28 April 2014
Available online 9 May 2014

Keywords:
Image segmentation
Global minimization
Graph cut
Continuous max-flow
Super-level set functions
Convex relaxation
Augmented Lagrangian method
Potts model

ABSTRACT

We propose a graph cut based global minimization method for image segmentation by representing the segmentation label function with a series of nested binary super-level set functions. This representation enables us to use $K-1$ binary functions to partition any images into K phases. Both continuous and discretized formulations will be treated. For the discrete model, we propose a new graph cut algorithm which is faster than the existing graph cut methods to obtain the exact global solution. In the continuous case, we further improve the segmentation accuracy using a number of techniques that are unique to the continuous segmentation models. With the convex relaxation and the dual method, the related continuous dual model is convex and we can mathematically show that the global minimization can be achieved. The corresponding continuous max-flow algorithms are easy and stable. Experimental results show that our model is very competitive to some existing methods.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Image segmentation is a fundamental but important task in computer vision and pattern recognition. It has received much attention by researchers during the past several decades. The objective of image segmentation is to partition an image into several parts according to some similarity measures such as intensity means, histograms, structure tensors and so on. There are many image segmentation methods proposed in the literature, in which the partial differential equation (PDE) based techniques and graph cut based approaches are two of the most popular image segmentation methods.

For PDE methods, the well known level set methods have been proven to be very flexible and quite efficient for image segmentation. The Mumford–Shah segmentation model [1] is an important approach to find a piecewise smooth approximation for a given image. However, the original Mumford–Shah functional is difficult to compute due to its weak mathematical properties such as discontinuity and non-convexity. By using a level set approximation, the discontinuity in the Mumford–Shah model can be easily handled and computed. To get a convex model, Cai, *etc.* [2] proposed a two-stage method for the Mumford–Shah segmentation model. Many methods of the fluid mechanics also can be applied to image segmentation, such as the phase-field method [3] and the Modica–Mortola phase transition method [4].

For the two-phase segmentation, the Chan–Vese model [5] is a very successful simplified version of the Mumford–Shah model. The Chan–Vese model is not convex. This explains why the numerical algorithm may sometimes get stuck at a local minimum close to the initial condition and produce undesirable segmentation results. Later, a binary level set method was proposed in [6] as a variant of the level set method. Meanwhile, the convex relaxation approach developed in [7] shows that one can get global minimizers for the piecewise constant Mumford–Shah functional with the binary approach [6] if we relax the binary constraint. The main idea of the convex relaxation is to relax the binary characteristic function into a continuous interval $[0,1]$ such that the non-convex original problem becomes convex. Solving such a relaxed convex problem can enable one to find a global minimizer, and then the global binary solution of the original problem can be obtained by a threshold process. Combining the convex relaxation and some recently developed TV (total variation) minimization techniques [8,9], Bresson, *etc.* have proposed some fast two-phase global minimization algorithms for image segmentation in [10,11].

For multi-phase segmentation, a generalization of the Chan–Vese model has been proposed in [12] to partition an image into n parts by using $\log_2 n$ level set functions. Similar to the two-phase case, the

* Corresponding author. Fax: +86 010 58808202.
E-mail address: jliu@bnu.edu.cn (J. Liu).

http://dx.doi.org/10.1016/j.jvcir.2014.04.011
1047-3203/© 2014 Elsevier Inc. All rights reserved.

model is non-convex and thus the global minimization cannot be guaranteed. Recently, a convex formulation of 4-phase Chan–Vese model has been proposed in [13,14] provided that the segmentation data term satisfies a convexity condition. Numerical tests show that this condition may be often satisfied in practice. In case this condition is violated, some "truncation" procedures have to be used.

Another multi-phase segmentation method is to use the label function or a PCLSM (piecewise constant level set method [15]) to represent different classes. By using a graph cut implementation, the PCLSM can be globally solved [16]. In the continuous case, functional lifting method [17] can be regarded as a convex formulation of PCLSM. As pointed out in [12,18], the TV of the label function or level set functions in PCLSM and multi-phase Chan–Vese model does not correspond exactly to the length term in Mumford–Shah model. The main drawback of these models is that some parts of the boundary are counted multiple times. Therefore pixels near some of the cluster boundaries will be misclassified (see e.g. [18]).

More recently, some continuous convex relaxation of the Potts model [19] have become popular. Bae, *etc.* proposed a smooth dual model of the Potts model in [20]. Pock, *etc.* [21] developed a tight convex relaxation framework for Potts model. Yuan, *etc.* [22,23] designed a max-flow approach to the Potts model. These continuous methods need to solve K unknown characteristic functions with a partition condition for K-phase clusters.

For the discrete partition problem, graph cut is a powerful tool to optimize the related energy. For example, the discrete Potts model restricted to 2-phase segmentation is computationally tractable by using some graph cut based min-cut/max-flow algorithms [24,25]. It is well known that the discrete Potts model is a NP-hard problem. Namely, if the number of segmentation classes is larger than two, there is no low-complexity algorithm which can find the exact global minimizer of the Potts model (see [26,27]). Instead of exactly solving the Potts model in a discrete setting, some algorithms for approximately minimizing the energy in Potts model have been proposed in [26], which are known as the popularly used alpha-expansion and alpha–beta swap algorithms.

Another approximation for the multi-phase Potts model is Ishikawa's graph cut method [28], in which the regularization term of the Potts model is modified such that it can be solved by a graph cut algorithm, c.f [28,16]. In 2006, Darbon, *etc.*[29] proposed another graph cut method for the multiphase segmentation. However, these graph-based methods generally suffer from metrication errors since the isotropic TV cannot be minimized by discrete max-flow algorithm. This difficulty could cause some zigzag edges in the clusters, which gives unnatural segmentation results. Recently, some continuous max-flow [30,22] algorithms have been developed by analyzing the primal min-cut and the dual max-flow problems with the Lagrangian multiplier method. These algorithms combine the advantages of both the continuous method and discrete model, and thus can provide impressive results.

This paper is devoted to propose a new graph cut method based on the multi-phase segmentation method, and the label function is represented by the binary super-level set. We will show that it is possible to minimize a modified piecewise constant Mumford–Shah segmentation model with the super-level set representation by solving the min-cut problem of a constructed graph.

The contributions of this paper include:

- We reveal the connections between the logical graph cut (discrete method) and the super-level set (continuous method) in the multiphase segmentation, and construct a graph which is related to the PCLSM using the super-level set expression. The proposed method is faster than the Ishikawa's graph cut method [28] due to some special structures of this graph.
- A continuous max-flow algorithm is proposed to further overcome two main drawbacks of the discrete graph cut method: one of the drawbacks of the discrete method is that the isotropic TV cannot be applied, but it can be employed in our method; the other one is that it is impossible to exactly penalize the length of the boundary for the multiphase segmentation in the discrete method, but the proposed method can do this. These two improvements ensure that one can get some better segmentation results from our algorithm. Compared to the Darbon's [29] discrete graph cut method, we proposed a continuous max-flow algorithm to overcome the drawbacks of the discrete method. Compared to some existing continuous methods, the proposed algorithm uses $K-1$ super-level set functions to partition K classes, which reduces the number of unknown variables, so providing a computationally very efficient algorithm. In addition, we use K dual variables to keep the regularization term in the model to be the exact length of the boundary in the continuous dual model, experimental results have shown that this can significantly improve the quality of the segmentation results.
- We give some mathematical analysis on the proposed algorithm, and show that the binary solutions of our algorithm can be obtained by a convex relaxation and a thresholding step, which corresponds to give a binary solution for the Potts model under a certain condition.

The rest of the paper is organized as follows: Section 2 gives some backgrounds on multi-phase segmentation methods; in Section 3, we introduce the proposed method, including the model, the algorithms and related analysis; Section 4 contains some experimental results; finally, some conclusions and discussions are presented in Section 5.

2. Related works

The generic problem of image segmentation is to partition an image domain Ω into K non-overlapping regions Ω_k such that $\Omega = \cup_{k=1}^K \Omega_k$. The well-known Potts model for image segmentation is to minimize the energy

$$\mathcal{E}^{Potts}\left(\{\Omega_k\}_{k=1}^K\right) = \sum_{k=1}^K \int_{\Omega_k} d^k(x)\mathrm{d}x + \mu\sum_{k=1}^K |\partial\Omega_k|, \tag{1}$$

such that $\cup_{k=1}^K \Omega_k = \Omega$ and $\Omega_i \cap \Omega_j = \emptyset$ if $i \neq j$, where $|\partial\Omega_k|$ stands for the perimeter of the boundary of Ω_k and $\mu > 0$ is a parameter. Here the first term is the data term, and each d^k should depend on the input image I. For example, $d^k(x) = |I(x) - c^k|^\lambda,\ \lambda = 1, 2$ represents that the pixels are classified in terms of the intensity means $\{c^k\}_{k=1}^K$. The second term, namely the regularization term, measures the sum of the perimeters of the sets $\Omega_k, k = 1, \ldots, K$. When $\lambda = 2$ and $\{c^k\}_{k=1}^K$ are unknown, (1) coincides with the energy of the piecewise constant Mumford–Shah model [1]:

$$\mathcal{E}^{MS}\left(\{\Omega_k\}_{k=1}^K, \{c^k\}_{k=1}^K\right) = \sum_{k=1}^K \int_{\Omega_k} |I(x) - c^k|^2\mathrm{d}x + \mu\sum_{k=1}^K |\partial\Omega_k|. \tag{2}$$

By introducing a vector-valued characteristic function $\psi(x) = (\psi^1(x), \ldots, \psi^K(x))$ with component functions

$$\psi^k(x) = \begin{cases} 1, & x \in \Omega_k, \\ 0, & x \notin \Omega_k, \end{cases}$$

the Potts model (1) can be reformulated as

$$\mathcal{E}^{Potts}(\psi) = \sum_{k=1}^K \int_\Omega d^k(x)\psi^k(x)\mathrm{d}x + \frac{\mu}{2}\sum_{k=1}^K \int_\Omega |\nabla\psi^k(x)|\mathrm{d}x, \tag{3}$$

in terms of the coarea formula [31]. The partition condition also can be satisfied by the constraint

$$\boldsymbol{\psi} \in \mathbb{B} = \left\{ \boldsymbol{\psi} = (\psi^1, \ldots, \psi^K) : \psi^k \in \{0,1\}, \sum_{k=1}^{K} \psi^k = 1 \right\}. \tag{4}$$

For numerical implementation, the binary function $\boldsymbol{\psi}$ is approximated by a smooth function. If $\boldsymbol{\psi}$ is approximated by the level set functions, this leads to the level set segmentation methods [5,12,6]. While if $\boldsymbol{\psi}$ is approximated by some exponential type functions, which leads to the EM (expectation maximization) algorithm for segmentation [20,32].

In case of two phases ($K = 2$), the last condition in (4) can be easily kept. The energy of the two-phase Chan–Vese model [5] can be written as

$$\mathcal{E}^{CV}(\phi) = \int_\Omega H(\phi) d^1 \mathrm{d}x + \int_\Omega (1 - H(\phi)) d^2 \mathrm{d}x + \mu \int_\Omega |\nabla H(\phi)| \mathrm{d}x,$$

where ϕ is the level set function which satisfies $\phi(x) > 0$ for $x \in \Omega_1$ and $\phi(x) < 0$ for $x \in \Omega_2$, H is a Heaviside function $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geqslant 0$. To address the non-convexity of $\mathbb{B}$ in (4), convex relaxation methods are developed in recent years. In the convex relaxation, $\mathbb{B}$ is usually relaxed as a convex set

$$\boldsymbol{u} \in \mathbb{B}^1 = \left\{ \boldsymbol{u} = (u^1, \ldots, u^K) : u^k \in [0,1], \sum_{k=1}^{K} u^k = 1 \right\}. \tag{5}$$

The energy of the global minimization model for two-phase in [6,7,10] can be expressed by

$$\mathcal{E}^{conv-2}(u) = \int_\Omega u d_1 \mathrm{d}x + \int_\Omega (1-u) d_2 \mathrm{d}x + \mu \int_\Omega |\nabla u| \mathrm{d}x,$$

where $0 \leqslant u \leqslant 1$, which is a special case for condition (5). This continuous model is connected to the discrete 2-phase Potts model, which can be exactly and globally solved by a number of recent developed global minimization methods, for example [25,22,23,10,11].

However, when the number of the phase is more than two, i.e. $K > 2$, the problem becomes complex. For continuous case, a natural choice is to extend the convex relaxation method to multi-phase, and to minimize the following energy

$$\mathcal{E}^{conv-K}(\boldsymbol{u}) = \sum_{k=1}^{K} \int_\Omega u^k d^k \mathrm{d}x + \frac{\mu}{2} \sum_{k=1}^{K} \int_\Omega |\nabla u^k| \mathrm{d}x, \tag{6}$$

such that $\boldsymbol{u} \in \mathbb{B}^1$. This is the model essentially solved in [33,21,34,20]. In this work, we shall proposed a fast max-flow algorithm which just needs to solve $K - 1$ unknown functions.

Another way to solve a multiphase segmentation problem is to use a piecewise constant level set function or the so-called label function to indicate the segmented phases. The label function can be expressed as

$$l(x) = \sum_{k=1}^{K} k \cdot \psi^k(x), \tag{7}$$

in which the different integer values of $l(x)$ stand for the different phases. With this label function, the energy of multi-phase segmentation problem can be modified as the following PCLSM [6]

$$\mathcal{E}^{Lab}(l) = \sum_{k=1}^{K} \int_\Omega \delta_{l,k} d^k \mathrm{d}x + \mu \int_\Omega |\nabla l| \mathrm{d}x, \tag{8}$$

where

$$\delta_{l,k} = \begin{cases} 1, & l(x) = k, \\ 0, & l(x) \neq k. \end{cases}$$

Instead of minimizing the last term, namely vector-valued TV in (6), we only need to optimize the TV of the scalar label function in the last term of PCLSM (8). Similar idea of replacing the vector TV by the typical TV has been recently proposed in [35] in the content of vector-valued image segmentation. More important, the anisotropic TV version of this model can be efficiently solved by Ishikawa's graph cut method [28,16]. However, the regularization term in (8) is slightly different from the original Potts model's. In fact, the last term in (8) is not exactly equal to the length of the cluster boundary. More precisely, some parts of the cluster boundary may be counted multiple times in (8), and it corresponds to that each part of the boundary has different weights to be penalized. This may lead to mis-classification of some pixels near these boundaries. We shall show this in the later experiments.

A third way to do the multiphase segmentation is to express different phases through the super-level set representation of the label function $l(x)$. A γ-super-level set function of any given function $l(x)$ is defined as

$$\phi(x,\gamma) = \begin{cases} 1, & \text{when } l(x) > \gamma, \\ 0, & \text{when } l(x) \leqslant \gamma. \end{cases} \tag{9}$$

By using the generalized coarea formula [31]

$$\int_\Omega |\nabla l| \mathrm{d}x = \int_\Omega \left(\int_{-\infty}^{+\infty} |\nabla \phi(x,\gamma)| \mathrm{d}\gamma \right) \mathrm{d}x, \tag{10}$$

and the layer cake formula [7]

$$l(x) = \int_{-\infty}^{+\infty} \phi(x,\gamma) \mathrm{d}\gamma, \tag{11}$$

a functional lifting method (FLM) has been developed in [17]. Using (10) and (11) we can see that the minimization problem (8) for PCLSM [6] can be written:

$$\mathcal{E}^{FLM}(\phi) = \int_{-\infty}^{+\infty} \left\{ \int_\Omega |\partial_\gamma \phi(x,\gamma)| d^\gamma(x) \mathrm{d}x + \mu \int_\Omega |\nabla \phi(x,\gamma)| \mathrm{d}x \right\} \mathrm{d}\gamma. \tag{12}$$

This is the problem considered in [17] and it shows clearly the relationship between FLM [17] and PCLSM [6]. Recently, a continuous max-flow method has been proposed in [23], which can be regarded as the dual model of (12). Like the models we mentioned earlier, the regularization term in FLM does not give exactly the length of the cluster boundary.

3. The proposed method

Our method is built upon the super-level set function representation. We shall rewrite the label function in PCLSM using the super-level set representation. With this new representation, each super-level set function can indicate a binary classification. Then we will properly design a graph in which the cost of the minimum cut is exactly equal to the minimization of the discrete energy of (12) plus a constant. Thus the global minimization of (12) can be found with the discrete graph cut method.

To overcome the previously mentioned flaw of the regularization term in the discrete setting, we shall propose a continuous dual max-flow model. To get a convex model, we will use the convex relaxation technique. The relaxation we are going to use is different from relaxing the non-convex binary set $\mathbb{B}$ (4) to a convex square area $\mathbb{B}^1$ (5) in most of the existing convex relaxation methods. Instead, we relax $\mathbb{B}$ to a convex triangle constraint, which has an advantage that we need to solve one less unknown function. The condition $\sum_{k=1}^{K} u^k = 1$ in (5) can also be satisfied exactly.

3.1. The super-level set representation

For better formulations, we rewrite the super-level set function $\phi(x,\gamma)$ in (9) as $\phi^\gamma(x)$. For a label function $l(x)$ which takes the

integers from 1 to K, it can be easily calculated that its super-level set function $1 = \phi^0(x) \geqslant \phi^1(x) \geqslant \cdots \geqslant \phi^{K-1} \geqslant \phi^K(x) = 0$ and

$$\phi^{k-1}(x) - \phi^k(x) = \begin{cases} 1, & l(x) = k, \\ 0, & l(x) \neq k, \end{cases} = \delta_{l,k}, \tag{13}$$

for $k = 1, 2, \ldots, K$. Therefore, (8) and (12) can be formulated as

$$\begin{aligned} \mathcal{E}^{Lab-sup}(\boldsymbol{\phi}) &= \sum\nolimits_{k=1}^{K} \int_\Omega (\phi^{k-1} - \phi^k) d^k \mathrm{d}x + \mu \sum_{k=1}^{K-1} \int_\Omega |\nabla \phi^k| \mathrm{d}x \\ &= \int_\Omega d^1 \mathrm{d}x + \sum_{k=1}^{K-1} \int_\Omega (d^{k+1} - d^k) \phi^k \mathrm{d}x + \mu \sum_{k=1}^{K-1} \int_\Omega |\nabla \phi^k| \mathrm{d}x, \end{aligned} \tag{14}$$

where $\boldsymbol{\phi} = (\phi^1, \phi^2, \ldots, \phi^{K-1})$ and $\phi^0 = 1$, $\phi^K = 0$. This particular formulation for data term is equivalent to that in the energy $\mathcal{E}^{Potts}$ if one replaces $\boldsymbol{\psi}$ in (3) by

$$\psi^k = \phi^{k-1} - \phi^k, \quad k = 1, 2, \ldots, K.$$

It is easy to verify that $\boldsymbol{\psi} \in \mathbb{B}$ can be guaranteed by

$$\begin{aligned} \boldsymbol{\phi} \in \widetilde{\mathbb{B}} = \Big\{ \boldsymbol{\phi} = (\phi^1, \phi^2, \ldots, \phi^{K-1}) : \phi^k \in \{0, 1\}, \\ 1 = \phi^0 \geqslant \phi^1 \geqslant \phi^2 \geqslant \cdots \geqslant \phi^{K-1} \geqslant \phi^K = 0 \Big\}. \end{aligned} \tag{15}$$

In the above equation, $\phi^1, \ldots, \phi^{K-1}$ are all binary and nested from 1 to $K-1$. They can indicate the classifications by subtracting arbitrary two neighbors of $\boldsymbol{\phi}$.

We shall treat the discrete problem firstly. Let $\mathbb{P}$ be the set of mesh grid points in Ω, and $\mathbb{N}_p^m$ be the set of m nearest neighbors of $p \in \mathbb{P}$. For $\Omega \subset \mathbb{R}^2$, $\mathbb{P} = \{(i,j) \subset \mathbb{Z}^2\}$ and for each $p = (i,j) \in \mathbb{P}$

$$\mathbb{N}_p^4 = \{(i \pm 1, j), (i, j \pm 1)\} \cap \mathbb{P},$$
$$\mathbb{N}_p^8 = \{(i \pm 1, j), (i, j \pm 1), (i \pm 1, j \pm 1)\} \cap \mathbb{P}.$$

Let ϕ_p^k, d_p^k be the function values of ϕ^k and d^k at $p \in \mathbb{P}$.

Since the minimizer is independent of any constant term in the energy, we can ignore the first constant term in (14) and write the discrete approximation as

$$\mathcal{E}^{Lab-sup-d}(\boldsymbol{\phi}) = \sum_{k=1}^{K-1} \sum_{p \in \mathbb{P}} \left(d_p^{k+1} - d_p^k \right) \phi_p^k + \mu \sum_{k=1}^{K-1} \sum_{p \in \mathbb{P}} \sum_{q \in \mathbb{N}_p^m} |\phi_q^k - \phi_p^k|.$$

Please note that the last regularization term is chosen to be the anisotropic TV

$$TV_1(\phi^k) = \int_\Omega |\nabla \phi^k|_1 \mathrm{d}x = \int_\Omega |\partial_{x_1} \phi^k| + |\partial_{x_2} \phi^k| \mathrm{d}x_1 \mathrm{d}x_2,$$

such that the regularization term can be minimized by graph cut (see [27,16]). Please note that TV_1 does not equal to the length of the boundary of the clusters since it is not rotation invariant.

In the next section, we shall construct a graph cut method to exactly solve the problem

$$\boldsymbol{\phi}^* = \arg \min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}} \mathcal{E}^{Lab-sup-d}(\boldsymbol{\phi}). \tag{16}$$

3.2. Graph cut implementation

3.2.1. Brief introduction of min-cut and max-flow

The min-cut and max-flow are discrete optimization problems defined over a graph. An advantage of min-cut and max-flow optimization is that the global minimization can be found by some max-flow algorithms. In this section, we shall construct a graph such that its min-cut corresponds to the minimizer of problem (16).

The basic principle for minimizing energy using graph cut can be found in [26,27]. A graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ is constituted by a set of vertices $\mathbb{V}$ and a set of directed edges $\mathbb{E}$. For any $v_1, v_2 \in \mathbb{V}$, let $(v_1, v_2) \in \mathbb{E}$ be the directed edge from vertex v_1 to vertex v_2, and denote $C(v_1, v_2)$ the weight on this edge. In the s–t (source–sink) graph, there are two special distinguished vertices in $\mathbb{V}$, the source s and the sink t. A cut on a graph $\mathbb{G}$ is to partition the vertices $\mathbb{V}$ into two disjoint connected set $\mathbb{V}_s$ and $\mathbb{V}_t$ such that $s \in \mathbb{V}_s$ and $t \in \mathbb{V}_t$. The cost of the cut is defined by

$$C(\mathbb{V}_s, \mathbb{V}_t) = \sum_{(v_1, v_2) \in \mathbb{E}, v_1 \in \mathbb{V}_s, v_2 \in \mathbb{V}_t} C(v_1, v_2). \tag{17}$$

The min-cut problem is to find the minimum cost of the cut. As to the max-flow problem, a flow f on the graph $\mathbb{G}$ is a function $f : \mathbb{E} \to \mathbb{R}$. $\forall (v_1, v_2) \in \mathbb{E}$, and the flow f satisfies the following conditions:

1. Maximum capacities

$$f(v_1, v_2) \leqslant C(v_1, v_2). \tag{18}$$

2. Flow conservation

$$\sum_{v_2 \in \mathbb{N}_{v_1}^+} f(v_2, v_1) - \sum_{v_2 \in \mathbb{N}_{v_1}^-} f(v_1, v_2) = 0, \tag{19}$$

where $\mathbb{N}_{v_1}^+, \mathbb{N}_{v_1}^-$ stand for two neighborhood systems of v_1, and

$$\mathbb{N}_{v_1}^+ = \{v_2 \in \mathbb{V} : (v_2, v_1) \in \mathbb{E}\},$$
$$\mathbb{N}_{v_1}^- = \{v_2 \in \mathbb{V} : (v_1, v_2) \in \mathbb{E}\}.$$

The max-flow problem is to find the maximum amount of flow f that can be pushed from source s to sink t under the above two flow constraint conditions (18) and (19). The theorem of Ford and Fulkerson showed that the min-cut and max-flow are dual problems, and the min-cut problem can be solved by some efficient max-flow algorithms such as [25].

3.2.2. Graph Construction

A cut of graph partition vertices into 2 clusters. In order to represent multi-phase segmentation, we must use several binary functions to distinguish the different classes. Generally speaking, a binary vector-valued function $\boldsymbol{\phi} = (\phi^1, \phi^2, \ldots, \phi^{K-1})$ without any constraints can represent 2^{K-1} phases, this is the idea used in the multiphase Chan–Vese model [12]. The representation related to $\boldsymbol{\phi} \in \widetilde{\mathbb{B}}$ in (15) can be used to identify K phases. For example, when $K = 4$, then 4 different phases can be represented with 3 binary functions ϕ_1, ϕ_2 and ϕ_3 with condition $\phi_1 \geqslant \phi_2 \geqslant \phi_3$. At any $x \in \Omega$, $\boldsymbol{\phi}$ can only take the 4 values $(1,1,1)$, $(1,1,0)$, $(1,0,0)$, $(0,0,0)$.

Now, we shall construct a graph $\mathbb{G}$ such that there is a one-to-one correspondence between feasible cuts on $\mathbb{G}$ and the binary vector-valued function $\boldsymbol{\phi} \in \widetilde{\mathbb{B}}$. Furthermore, the minimum cost cut will correspond to the minimizer of problem (16) and

$$\min_{(\mathbb{V}_s, \mathbb{V}_t)} C(\mathbb{V}_s, \mathbb{V}_t) = \min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}} \mathcal{E}^{Lab-sup-d}(\boldsymbol{\phi}) + constant. \tag{20}$$

The vertices, edges, and weights of the graph $\mathbb{G}$ are defined as follows:

1. Vertices: for each mesh grid point $p \in \mathbb{P}$, ϕ_p^k corresponds to a vertex denoted by v_p^k. Hence the set of vertices can be defined as
$$\mathbb{V} = \{v_p^k : p \in \mathbb{P}, k = 1, 2 \ldots, K-1\} \cup \{s\} \cup \{t\}.$$
2. Edges: for each vertex v_p^k, t-links (terminal links) are defined as
$$\mathbb{E}_t(p) = \left\{ (s, v_p^k) : k = 1, 2, \ldots, K-1 \right\} \cup \left\{ (v_p^k, t) : k = 1, 2, \ldots, K-1 \right\}.$$

It means that every vertex is connected to both source s and sink t. The t-links is associated with the data term of energy in (16). There will be two types of n-links (neighbor links) in the graph. The one is given by

$$\mathbb{E}_n(p) = \left\{(v_p^k, v_q^k) : \forall p, q \in \mathbb{P}, q \in \mathbb{N}_p^m, k = 1, 2, \ldots, K-1\right\},$$

which is associated with the regularization term. The other n-links are the links among layers (different k values), the set of edges is defined by

$$\mathbb{E}_l(p) = \left\{(v_p^k, v_p^{k+1}) : k = 1, 2, \ldots, K-2\right\},$$

which is used to satisfy the condition $\phi \in \widetilde{\mathbb{B}}$. In the left figure of Fig. 1, we show the t-links and part of the n-links (layers) in the 1-dimension case. Combining the t-links and n-links of each vertex v_p^k, the set of all the edges in the graph is defined as

$$\mathbb{E} = \bigcup_{p\in\mathbb{P}} \{\mathbb{E}_t(p) \cup \mathbb{E}_n(p) \cup \mathbb{E}_l(p)\}.$$

3. Weights: the weights are assigned as following

$$C(s, v_p^k) = d_p^{k+1}, \quad k = 1, 2, \ldots, K-1.$$
$$C(v_p^k, t) = d_p^k, \quad k = 1, 2, \ldots, K-1.$$
$$C(v_p^k, v_p^{k+1}) = +\infty, \quad k = 1, 2, \ldots, K-2.$$
$$C(v_p^k, v_q^k) = \frac{\mu}{2}.$$

In fact, the weights of the edges (v_p^k, v_p^{k+1}) can be set to be any relatively large values, but they cannot be removed from the graph.

In our graph, some edges have infinity capabilities, and thus the cost of a cut may equal to infinity. Obviously, such a cut would not be the minimum cut. For a cut $(\mathbb{V}_s, \mathbb{V}_t)$, we say $(\mathbb{V}_s, \mathbb{V}_t)$ is a feasible cut when $C(\mathbb{V}_s, \mathbb{V}_t) < +\infty$.

We can show the following result:

Theorem 1. *There is a one-to-one correspondence between the feasible cuts of graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ and the super-level set $\phi \in \widetilde{\mathbb{B}}$, and*

$$\min_{(\mathbb{V}_s, \mathbb{V}_t)} C(\mathbb{V}_s, \mathbb{V}_t) = \min_{\phi\in\widetilde{\mathbb{B}}} \mathcal{E}^{Lab-sup-d}(\phi) + \sum_{k=1}^{K-1}\sum_{p\in\mathbb{P}} d_p^k.$$

Proof. Please find the details in Appendix A. □

Theorem 1 shows that we can exactly solve problem (20) by the discrete max-flow algorithms on the graph $\mathbb{G}$.

3.3. Continuous dual model

Though the discrete model (16) can be globally solved by graph cut, a main drawback is that the regularization term is not exactly equal to the length of the boundary, which causes undesirable results. So, in this section, we apply the continuous method to improve it.

3.3.1. Continuous max-flow model

In this section, we shall build a continuous max-flow model of the constructed graph. The continuous max-flow method for image segmentation was proposed in [30]. Here, we apply a similar idea to build a new dual model with the super-level set representation according to the previous constructed graph.

To simplify the expression, we denote the flow functions as following:

$$f_{p,q}^k = f(v_p^k, v_q^k),$$
$$f_{q,p}^k = f(v_q^k, v_p^k),$$
$$f_{s,p}^k = f(s, v_p^k),$$
$$f_{p,t}^k = f(v_p^k, t),$$
$$f_p^{k,k+1} = f(v_p^k, v_p^{k+1}).$$

In addition, we introduce two convex sets:

$$\widetilde{\mathbb{B}}^1 = \{\boldsymbol{\phi} = (\phi^1, \phi^2, \ldots, \phi^{K-1}) : \phi^k \in [0,1], \phi^1 \geqslant \cdots \geqslant \phi^{K-1}\},$$
$$\mathbb{C} = \{\boldsymbol{g} = (\boldsymbol{g}^1, \ldots, \boldsymbol{g}^K) : \boldsymbol{g}^k = (g_1^k, g_2^k), ||\boldsymbol{g}^k||_\infty = \max_x ||g^k(x)||_2 \leqslant \mu\}.$$

Here $\widetilde{\mathbb{B}}^1$ is a convex relaxation set of binary set $\mathbb{B}$.

A max-flow problem is to find the maximum amount of flow f that tries to stream from source s to sink t under the two flow constraint conditions (18) and (19), i.e.

$$\begin{cases} \max_{f_{p,t}^k, f_{s,p}^k, f_p^{k,k+1}} \sum_{p\in\mathbb{P}}\sum_{k=1}^{K-1} f_{p,t}^k \\ f_{s,p}^k \leqslant d_p^{k+1},\ f_{p,t}^k \leqslant d^k,\ f_p^{k,k+1} = +\infty,\ 0 \leqslant f_{p,q}^k,\ f_{q,p}^k \leqslant \frac{\mu}{2},\ k = 1,\ldots,K-1,\ p \in \mathbb{P}, \\ f_{s,p}^k + f_p^{k-1,k} - f_p^{k,k+1} - f_{p,t}^k + \sum_{q\in\mathbb{N}_p^4} (f_{q,p}^k - f_{p,q}^k) = 0,\ \ k = 1,\ldots,K-1, p \in \mathbb{P}. \end{cases}$$

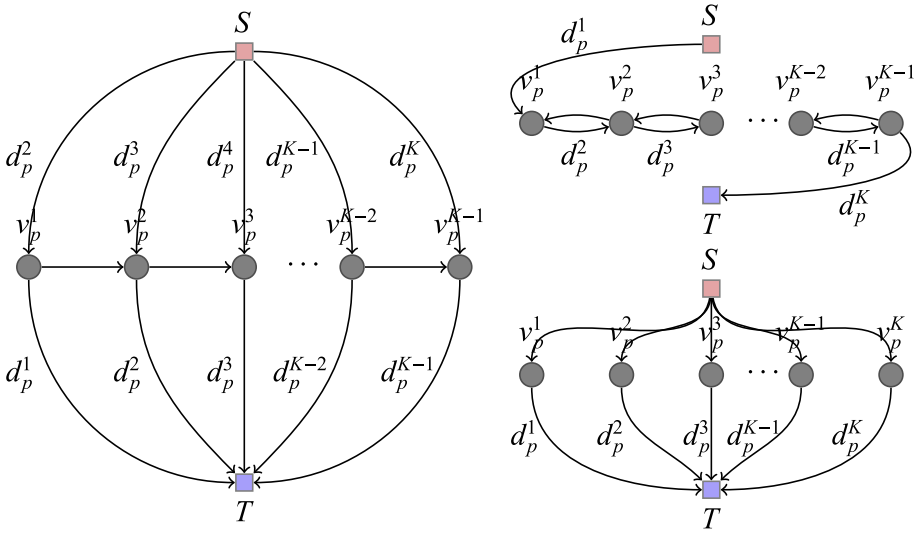


Fig. 1. Comparison of the constructed graphs for different multi-label methods in the 1-D case. Left: the proposed method; top right: Ishikawa's [28]; bottom right: Yuan, *etc.*'s [39].

In the above equation, $f_p^{0,1} = f_p^{K-1,K} = 0$. For the 2-D case, i.e. $p = (i,j) \in \mathbb{R}^2$, $\mathbb{N}_p^4 = \{(i \pm 1, j), (i, j \pm 1)\}$, then

$$\sum_{q \in \mathbb{N}_p^4} (f_{q,p}^k - f_{p,q}^k) = \left(f_{(i-1,j),(ij)}^k - f_{(ij),(i-1,j)}^k\right) - \left(f_{(ij),(i+1j)}^k - f_{(i+1,j),(ij)}^k\right) + \left(f_{(ij-1),(ij)}^k - f_{(ij),(ij-1)}^k\right) - \left(f_{(ij),(ij+1)}^k - f_{(ij+1),(ij)}^k\right).$$

Let

$$g_1^k(i,j) = f_{(ij),(i+1j)}^k - f_{(i+1,j),(ij)}^k,$$
$$g_2^k(i,j) = f_{(ij),(ij+1)}^k - f_{(ij+1),(ij)}^k,$$

we have

$$\sum_{q \in \mathbb{N}_p^4} (f_{q,p}^k - f_{p,q}^k) = (g_1^k(i-1,j) - g_1^k(i,j)) + (g_2^k(i,j-1) - g_2^k(i,j)) = -\partial_{x_1}^- g_1^k - \partial_{x_2}^- g_2^k = -\nabla \cdot \boldsymbol{g}^k,$$

where $\partial_{x_1}^-$, $\partial_{x_2}^-$ are the backward finite difference schemes with respect to x_1, x_2, and $\nabla\cdot$ is the discrete divergence operator. Here $-\nabla \cdot \boldsymbol{g}^k$ has a physical meaning that it indicates the total flow at a vertex. From $0 \leqslant f_{p,q}^k, f_{q,p}^k \leqslant \frac{\mu}{2}$, we have $\|\boldsymbol{g}^k(i,j)\|_1 \leqslant \mu$. It is well known that the L^1 norm in the dual space would lead to an anisotropic TV in the primal problem. To get the isotropic TV in the primal problem, we modified this condition as $\|\boldsymbol{g}^k(x)\|_2 \leqslant \mu$ in the continuous case. As to the condition $f_p^{k,k+1} = +\infty$, it is used to keep the linear inequality constraint $\phi^1 \geqslant \cdots \geqslant \phi^{K-1}$ according to the proof of Theorem 1. Similarly, we can prove that for any large enough constant $\alpha > \max_{p \in \mathbb{P}} 2\sum_{k=1}^K d_p^k$, the condition $f_p^{k,k+1} \geqslant \alpha$ can still ensure that the linear inequality constraint holds. In addition, in order to recover $\phi^0 = 1$ and $\phi^K = 0$, the layer number k takes the values from 0 to K, which means there will be two extra layers vertices v_p^0 and v_p^K in the continuous graph. Accordingly, d^0 and d^{K+1} are both extended to an arbitrary large value. Then the above discrete max-flow problem can be generally reformulated as

$$\begin{cases} \max_{f_t^k, f_s^k, f^{k,k+1}, \boldsymbol{g}^k} \sum_{k=0}^K \int_\Omega f_t^k(x)\mathrm{d}x \\ f_s^k(x) \leqslant d^{k+1}(x),\ f_t^k(x) \leqslant d^k(x),\ f^{k,k+1}(x) \geqslant \alpha,\ \|\boldsymbol{g}^k(x)\|_2 \leqslant \mu,\ k = 0,1\ldots,K,\ x \in \Omega, \\ f_s^k(x) - f_t^k(x) + f^{k-1,k}(x) - f^{k,k+1}(x) - \nabla \cdot \boldsymbol{g}^k(x) = 0,\ k = 0,1,\ldots,K,\ x \in \Omega, \end{cases} \tag{21}$$

where $f^{-1,0}(x) = f^{K,K+1}(x) = 0$.

It can be shown in the following proposition.

Proposition 1. *The max-flow problem (21) is equivalent to a dual problem of*

$$\min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}^1} \left\{\mathcal{E}^{Lab-sup}(\boldsymbol{\phi}) + constant\right\},$$

where $\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$ is defined in (14).

Proof. See Appendix B. □

The model (21) can be directly optimized by the augmented Lagrangian method [36,37]. Since $f^{k,k+1}$ in the graph is used to impose $\boldsymbol{\phi}$ such that $\phi^1 \geqslant \cdots \geqslant \phi^{K-1}$, thus we can remove $f^{k,k+1}$ from the energy and replace it by condition $\phi^1 \geqslant \cdots \geqslant \phi^{K-1}$. Moreover, we do not need to solve ϕ^0 and ϕ^K because $\phi^0 = 1$ and $\phi^K = 0$ are fixed in terms of the proof of Proposition 1. By applying the augmented Lagrangian method, we get the following dual model

$$\min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}^1} \max_{f_t^k \leqslant d^k, f_s^k \leqslant d^{k+1}, \boldsymbol{g} \in \mathbb{C}} \left\{\mathcal{E}^{Dual}(\boldsymbol{f_t}, \boldsymbol{f_s}, \boldsymbol{g}, \boldsymbol{\phi})\right\}, \tag{22}$$

where

$$\mathcal{E}^{Dual}(\boldsymbol{f_t}, \boldsymbol{f_s}, \boldsymbol{g}, \boldsymbol{\phi}) = \sum_{k=1}^{K-1} \int_\Omega f_t^k(x)\mathrm{d}x + \sum_{k=1}^{K-1} \int_\Omega \phi^k(x)(f_s^k(x) - f_t^k(x) - \nabla \cdot \boldsymbol{g}^k(x))\mathrm{d}x - \frac{r}{2}\sum_{k=1}^{K-1} \int_\Omega (f_s^k(x) - f_t^k(x) - \nabla \cdot \boldsymbol{g}^k(x))^2 \mathrm{d}x,$$

and $r > 0$ is a penalty parameter in the augmented Lagrangian method. In the above model, the second term is the Lagrangian multiplier term, while the third term is a concave penalty functional in the maximization problem.

The dual model (22) can be efficiently solved by the following alternating optimization algorithm

$$\begin{cases} (\boldsymbol{f}_t)^{(v+1)} &= \arg\max_{f_t^k \leqslant d^k} \mathcal{E}^{Dual}\left(\boldsymbol{f}_t, (\boldsymbol{f}_s)^{(v)}, (\boldsymbol{g})^{(v)}, (\boldsymbol{\phi})^{(v)}\right), \\ (\boldsymbol{f}_s)^{(v+1)} &= \arg\max_{f_s^k \leqslant d^{k+1}} \mathcal{E}^{Dual}\left((\boldsymbol{f}_t)^{(v+1)}, \boldsymbol{f}_s, (\boldsymbol{g})^{(v)}, (\boldsymbol{\phi})^{(v)}\right), \\ (\boldsymbol{g})^{(v+1)} &= \arg\max_{\boldsymbol{g} \in \mathbb{C}} \mathcal{E}^{Dual}\left((\boldsymbol{f}_t)^{(v+1)}, (\boldsymbol{f}_s)^{(v+1)}, \boldsymbol{g}, (\boldsymbol{\phi})^{(v)}\right), \\ (\boldsymbol{\phi})^{(v+1)} &= \arg\max_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}^1} \mathcal{E}^{Dual}\left((\boldsymbol{f}_t)^{(v+1)}, (\boldsymbol{f}_s)^{(v+1)}, (\boldsymbol{g})^{(v+1)}, \boldsymbol{\phi}\right). \end{cases}$$

Here, we summarize the details of updating formulation of each variable in Algorithm 1.

Algorithm 1. Given the initial values $(\boldsymbol{\phi})^{(0)} = (\boldsymbol{g})^{(0)} = (\boldsymbol{f}_s)^{(0)} = 0$, we choose a time step parameter $\tau_\phi > 0$, and update the following steps until a convergence criterion is reached:

Step 1 for $k = 1,\ldots,K-1$, optimize $(f_t^k)^{(v+1)}$ by

$$(f_t^k)^{v+1} = \min\left\{(f_s^k)^{(v)} - \nabla \cdot (\boldsymbol{g}^k)^{(v)} + \frac{1 - (\phi^k)^{(v)}}{r}, d^k\right\}.$$

Step 2 for $k = 1,\ldots,K-1$, optimize $(f_s^k)^{(v+1)}$ by

$$(f_s^k)^{v+1} = \min\left\{(f_t^k)^{(v)} + \nabla \cdot (\boldsymbol{g}^k)^{(v)} + \frac{(\phi^k)^{(v)}}{r}, d^{k+1}\right\}.$$

Step 3 for $k = 1,\ldots,K-1$, update

$$(\boldsymbol{g}^k)^{(v+1)} = \arg\min_{\boldsymbol{g}^k \in \mathbb{C}} \| - \frac{(\phi^k)^{(v)}}{r} + (f_s^k)^{(v+1)} - (f_t^k)^{(v+1)} - \nabla \cdot \boldsymbol{g}^k\|^2$$

with Chambolle projection algorithm [8].

Step 4 for $k = 1,\ldots,K-1$, optimize $(\phi^k)^{(v+1)}$ by

$$(\phi^k)^{(v+1)} = Proj_{\widetilde{\mathbb{B}}^1}\left((\phi^k)^{(v)} - \tau_\phi\left((f_s^k)^{(v+1)} - (f_t^k)^{(v+1)} - \nabla \cdot (\boldsymbol{g}^k)^{(v+1)}\right)\right).$$

Here the symbol $Proj_{\widetilde{\mathbb{B}}^1}$ is a projection operator on a convex set $\widetilde{\mathbb{B}}^1$.

As to the projection on convex set $\widetilde{\mathbb{B}}^1$, in 2-dimension case (i.e. $K = 2$), for $k = 1, 2$,

$$[Proj_{\widetilde{\mathbb{B}}^1}(\boldsymbol{a})]_k = \begin{cases} \max\{\min\{a_k, 1\}, 0\}, & \text{when } a_1 \geqslant a_2, \\ \max\{\min\{\frac{\sum_{k=1}^2 a_k}{2}, 1\}, 0\}, & \text{when } a_1 < a_2. \end{cases}$$

When the dimension is more than 2, the explicit mathematical formulation of this projection is complicated, and we do not list it here. But it can be easily calculated by a simple Algorithm 2, just as similar as the recursive algorithm in [38].

Algorithm 2. [Projection on $\widetilde{\mathbb{B}}^1$] Let $\boldsymbol{z} = \boldsymbol{a}$, set $k = 1$,

Step 1 if $z_k \geqslant z_{k+1}$, go to the step 4; else let $\mathbb{S} = \{k, k+1\}$, $z_k = z_{k+1} = \frac{\sum_{i' \in \mathbb{S}} a_{i'}}{|\mathbb{S}|}$, where $|\mathbb{S}|$ represents the number of elements in $\mathbb{S}$. let $i = k$.

Step 2 If $i > 1$, then go to the next step; else go to the step 4.

Step 3 if $z_{i-1} \geqslant z_i$, go to the next step; else let $\mathbb{S} = \mathbb{S} \cup \{i-1\}$, $z_j = \frac{\sum_{i' \in \mathbb{S}} a_{i'}}{|\mathbb{S}|}$ for each $j \in \mathbb{S}$, let $i = i - 1$, and go back to step 2.

Step 4 if $k = K-1$, end the algorithm; else $k = k+1$, and go back to step 1.

When the algorithm is done, we have $z_1 \geqslant z_2 \geqslant \cdots \geqslant z_K$, then $[Proj_{\widetilde{\mathbb{B}}^1}(\boldsymbol{a})]_k = \max\{\min\{\boldsymbol{z}_k, 1\}, 0\}$.

The solution produced by Algorithm 1 will be nearly binary in most applications. For obtaining a binary solution, we have the following threshold theorem:

Theorem 2 (*Binary solution by threshold*). *Let* $(\boldsymbol{f}t*, \boldsymbol{f}s*, \boldsymbol{g}^*, \boldsymbol{\phi}^*)$ *be any optimal saddle point of problem* (22)*, for any threshold* $\gamma \in (0,1)$*, define* $\boldsymbol{\psi}^*_\gamma = ((\psi^*_\gamma)^1, \ldots, (\psi^*_\gamma)^{K-1})$ *with the component function*

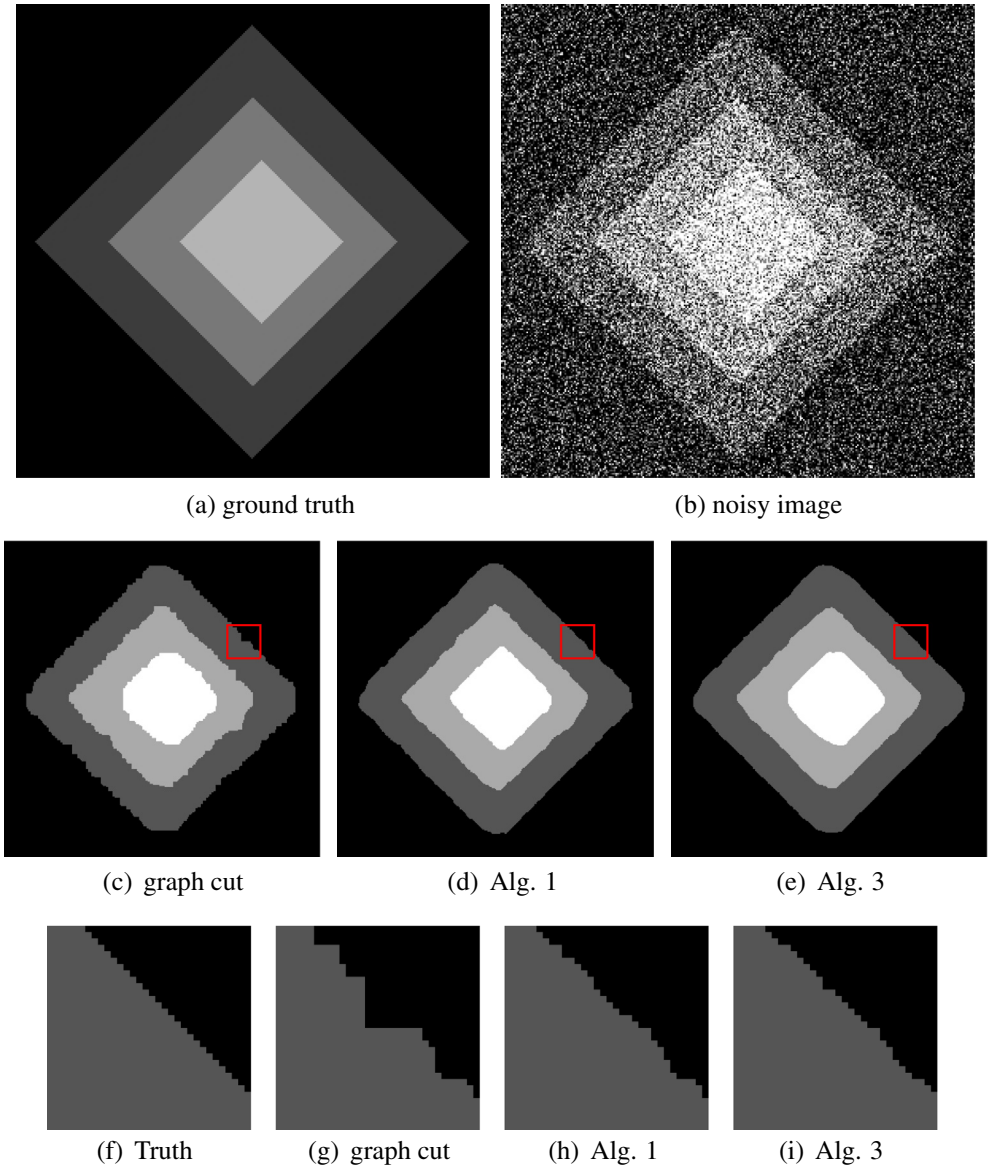
$$(\psi^*_\gamma)^k(x) = \begin{cases} 1, & (\phi^*)^k(x) \geqslant \gamma, \\ 0, & (\phi^*)^k(x) < \gamma, \end{cases}$$

then $\boldsymbol{\psi}^*_\gamma$ *is a global minimizer of the binary problem* $\min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}} \mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$, where $\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$ *is defined in* (14).

Proof. Please find in Appendix C. □

3.3.2. Simplification and Improvements

In the continuous max-flow model, the Lagrangian multiplier functions play the same role of super-level set functions in the primal min-cut problem. It can be seen from the proof of Proposition 1 and Theorem 1 that f_t^k, f_s^k, $f^{k,k+1}$ are used to keep the Lagrangian multiplier functions satisfy $1 = \phi^0 \geqslant \phi^1 \geqslant \cdots \geqslant \phi^{K-1} \geqslant \phi^K = 0$ in the continuous max-flow (21). In fact, we do not need to directly impose constraints to the flow functions and to solve so many unknown variables in the dual space. Based on the analysis of the min-cut, max-flow problems and Eqs. (B.2) and (B.3) in the previous section, we can confirm that the flow functions f_t^k, f_s^k would reach the maximum capabilities d^k and d^{k+1} respectively when the flow is the max-flow. Thus, we can directly set $f_t^k = d^k$, $f_s^k = d^{k+1}$ in max-flow model (21) and impose each Lagrangian function $0 \leqslant \phi^k \leqslant 1$. Let $\phi^0 = 1$, $\phi^K = 0$, $\boldsymbol{\phi} = (\phi^1, \ldots, \phi^{K-1})$. From Proposition 1, we can directly solve



(a) ground truth (b) noisy image

(c) graph cut (d) Alg. 1 (e) Alg. 3

(f) Truth (g) graph cut (h) Alg. 1 (i) Alg. 3

Fig. 2. Segmentation results of the discrete graph cut method, and our dual continuous methods Algorithms 1 and 3.

$$\min_{\phi\in\widetilde{\mathbb{B}}^1}\sum_{k=1}^{K}\int_\Omega(\phi^{k-1}-\phi^k(x))d^k(x)\mathrm{d}x+\mu\sum_{k=1}^{K}|\partial\Omega_k|.$$

Since

$$\mu\sum_{k=1}^{K}|\partial\Omega_k|=\frac{\mu}{2}\sum_{k=1}^{K}\int_\Omega|\nabla\psi^k(x)|\mathrm{d}x=\frac{\mu}{2}\sum_{k=1}^{K}\int_\Omega|\nabla\phi^{k-1}(x)-\nabla\phi^k(x)|\mathrm{d}x$$
$$=\max_{\boldsymbol{g}\in\mathbb{C}}\left\{\sum_{k=1}^{K}\int_\Omega\left(\phi^{k-1}(x)-\phi^k(x)\right)\nabla\cdot\boldsymbol{g}^k(x)\mathrm{d}x\right\},$$

and thus we propose the following convex relaxation dual model

$$\min_{\phi\in\widetilde{\mathbb{B}}^1}\max_{\boldsymbol{g}\in\mathbb{C}}\mathcal{E}^{Conv-sup}(\boldsymbol{\phi},\boldsymbol{g})=\sum_{k=1}^{K-1}\int_\Omega(d^{k+1}-d^k)\phi^k\mathrm{d}x$$
$$+\sum_{k=1}^{K}\int_\Omega(\phi^{k-1}-\phi^k)\nabla\cdot\boldsymbol{g}^k\mathrm{d}x. \tag{23}$$

In the above model, we use $K-1$ super-level set functions to segment the image into K parts, while to exactly penalize the length of boundary, K dual variables are adopted. Please note that the regularization terms in the above $\mathcal{E}^{Conv-sup}$ and $\mathcal{E}^{Lab-sup}$ are slightly different. For ϕ^k, $k=1,\ldots,K-1$ are a series of super-level set functions, the regularization term in $\mathcal{E}^{Conv-sup}$ equals to the length of cluster boundaries exactly but $\mathcal{E}^{Lab-sup}$ usually not.

The saddle point of $\mathcal{E}^{Conv-sup}$ can be found by a simple alternating algorithm

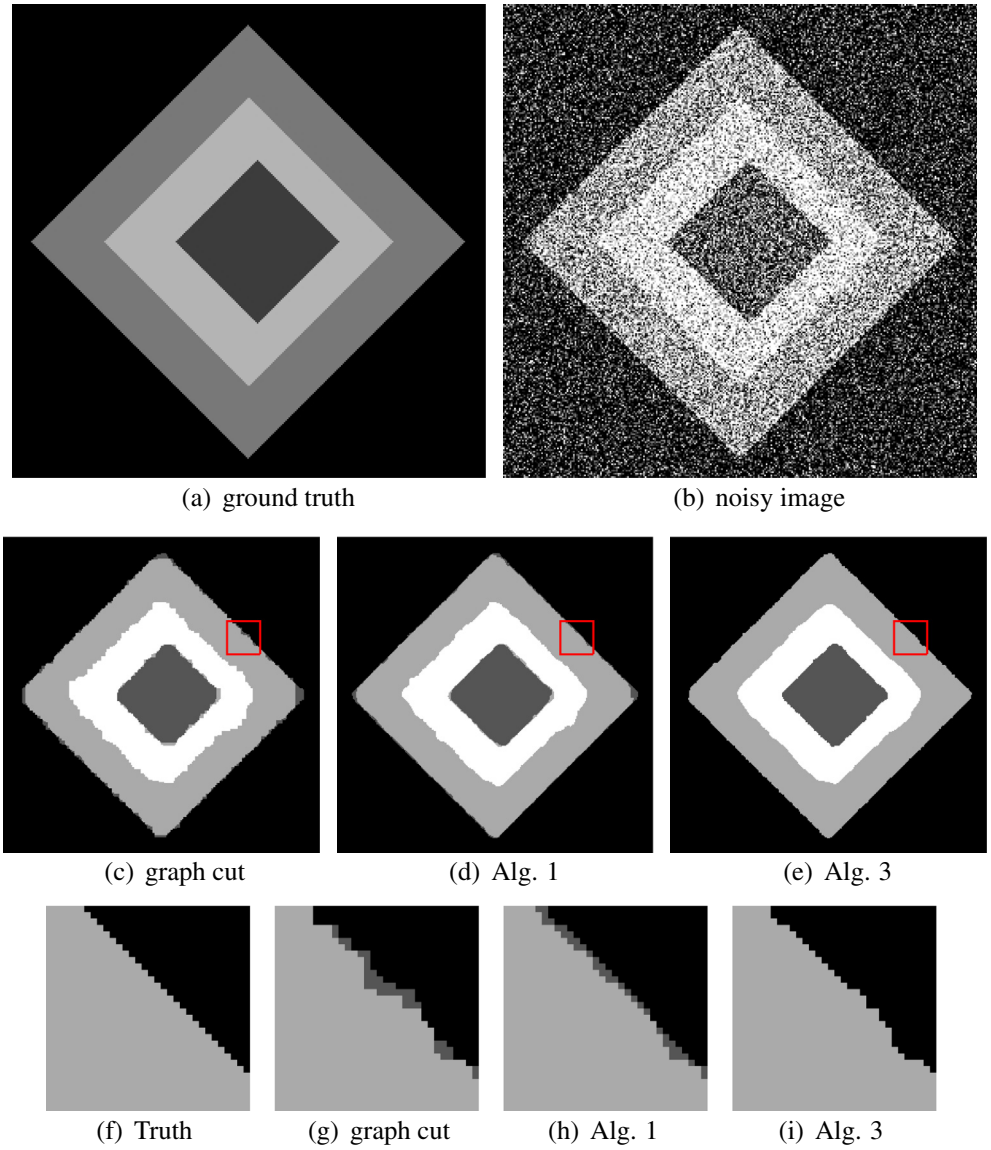
$$\boldsymbol{\phi}^{(v+1)}=\arg\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}^1}\mathcal{E}^{Conv-sup}(\boldsymbol{\phi},\boldsymbol{g}^{(v)}), \tag{24}$$

$$\boldsymbol{g}^{(v+1)}=\arg\min_{\boldsymbol{g}\in\mathbb{C}}\mathcal{E}^{Conv-sup}(\boldsymbol{\phi}^{(v+1)},\boldsymbol{g}), \tag{25}$$

where v is the number of iterations. The subproblems (24) and (25) can both be solved by projection gradient algorithms, which are sumarized in Algorithm 3.

Algorithm 3. Given the initial values $\boldsymbol{\phi}^{(0)},\boldsymbol{g}^{(0)}=\mathbf{0}$, and choosing two time step parameters $\tau_\phi,\tau_{\boldsymbol{g}}>0$, updating the following steps until a convergence criterion is reached:

Step 1, for $k=1,2,\ldots,K-1$,



(a) ground truth (b) noisy image

(c) graph cut (d) Alg. 1 (e) Alg. 3

(f) Truth (g) graph cut (h) Alg. 1 (i) Alg. 3

Fig. 3. Segmentation results of the discrete graph cut method, and our dual continuous methods Algorithms 1 and 3.

$$(\phi^k)^{(v+1)} = (\phi^k)^{(v)} + \tau_\phi\left(d^k - d^{k+1} + \nabla\cdot(\boldsymbol{g}^k)^{(v)} - \nabla\cdot(\boldsymbol{g}^{k+1})^{(v)}\right).$$
$$\boldsymbol{\phi}^{(v+1)} = Proj_{\widetilde{\mathbb{B}}^1}(\boldsymbol{\phi}^{(v+1)}).$$

Step 2, for $k = 1, 2, \ldots, K$,

$$(\boldsymbol{g}^k)^{(v+1)} = (\boldsymbol{g}^k)^{(v)} - \tau_{\boldsymbol{g}}\left((\nabla\phi^{k-1})^{(v+1)} - (\nabla\phi^k)^{(v+1)}\right).$$
$$\boldsymbol{g}^{(v+1)} = Proj_{\mathbb{C}}(\boldsymbol{g}^{(v+1)}).$$

Here the symbol $Proj_{\widetilde{\mathbb{B}}^1}$, $Proj_{\mathbb{C}}$ are projection operators on convex sets $\widetilde{\mathbb{B}}^1$ and $\mathbb{C}$.

Both of the projections onto $\mathbb{C}$ and $\widetilde{\mathbb{B}}^1$ can be easily calculated. For any vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_K) \in \mathbb{R}^K$, the projection of $\boldsymbol{a}$ on convex set $\mathbb{C}$ has a closed-form expression

$$Proj_{\mathbb{C}}(\boldsymbol{a}) = \begin{cases} \boldsymbol{a}, & ||\boldsymbol{a}||_2 \leqslant \mu, \\ \frac{\mu\boldsymbol{a}}{||\boldsymbol{a}||_2}, & ||\boldsymbol{a}||_2 \leqslant \mu. \end{cases}$$

On the other hand, the projection on $\widetilde{\mathbb{B}}^1$ can be solved by Algorithm 2.

Compared to Algorithms 1 and 3, for the K-phase segmentation, it at least requires to solve $3\times(K-1)$ variables. But in Algorithm 3, it only needs to solve $K-1$ super-level set functions, K dual variable and two projections onto convex sets. Though the convergence of the augmented Lagrangian method is faster than that of the projection gradient method, the cost of Algorithm 3 at each iteration is much less than Algorithm 1 since both of the projections can be easily calculated. Generally speaking, Algorithm 3 is slightly faster than Algorithm 1.

It is well known that the existence of the global minimizer of the Potts model is still an open problem, but we can show that the global minimizer of the Potts model can be achieved under a certain condition:

Theorem 3. *Suppose* $(\boldsymbol{\phi}^*, \boldsymbol{g}^*)$ *is a saddle point of the convex relaxation model* (23), $\forall x \in \Omega$, *if there is a unique minimizer for* $\min\{d^1(x) + \nabla\cdot(\boldsymbol{g}^*)^1(x), \ldots, d^K(x) + \nabla\cdot(\boldsymbol{g}^*)^K(x)\}$, *then* $\boldsymbol{\psi}^* = ((\psi^*)^1, \ldots, (\psi^*)^K)$, *where* $(\psi^*)^k = (\phi^*)^{k-1} - (\phi^*)^k$ *is a global minimizer of the Potts model* $\min_{\boldsymbol{\psi}\in\mathbb{B}}\mathcal{E}^{Potts}(\boldsymbol{\psi})$.

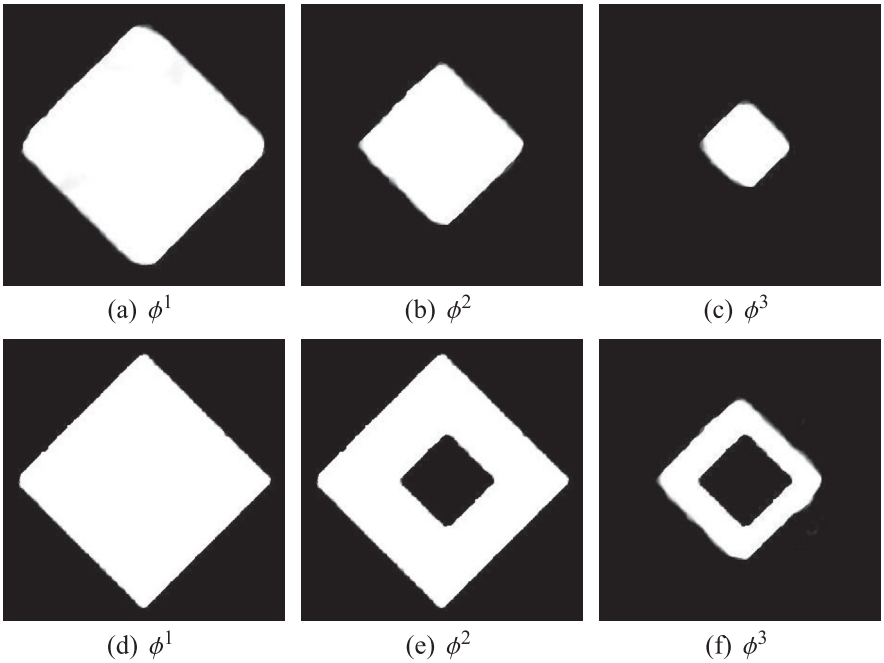
Proof. Please see Appendix D. □

3.4. Comparison with some related methods

We first point out the difference between the proposed graph and Ishikawa's [28]. Just as shown in Fig. 1, in Ishikawa's graph, only the vertices in first layer are connected with the source s, and the vertices connected with sink t are all located at the last layer. But in the proposed graph, every vertices are both connected source and sink. This means that paths from source to sink in our graph are alway much shorter than Ishikawa's, especially when K is large. The efficiency of many max-flow algorithms depend highly on the length of these paths. Generally speaking, the shorter the path is, the higher efficiency for the algorithm. Thus, to find the max-flow with algorithm [25] on the proposed graph would cost much less computational time than Ishikawa's. We will show this in the numerical experiments. Because of the different structures, the weights d^k assigned to the edges are totally different in the two graphs, see the second figure in Fig. 1 for details. Of course, for the two-phases segmentation, i.e. $K = 2$, these two graphs would be the same.

The proposed graph is similar to Darbon's in [29]. The slight difference of the two graph is that the proposed graph is a directed graph but [29] is an undirected one. Here, we give a continuous max-flow algorithm of [29]. With the proposed continuous dual model, the isotropic TV and the length of the boundary can be exactly penalized in the segmentation. This will greatly improve the quality of the segmentation result.

There is another graph cut based method proposed by Bae, *etc.* [13,14] for the 4-phases Chan–Vese model. However, in this method, it is not easy to extend it to any K-phase segmentation other than 4 clusters because the weights of the edges in the graph is difficult to be determined in the general cases. Moreover, there is a mild convex condition for the data term. In [13], the authors have proposed an approximation method to handle the cases when the convex condition is failed. In our method, the model is convex without any conditions and thus it can be optimized by the graph cut algorithm for any positive data terms such as L^1, L^2 and other norms.



(a) ϕ^1 (b) ϕ^2 (c) ϕ^3

(d) ϕ^1 (e) ϕ^2 (f) ϕ^3

Fig. 4. The output super-level set functions ϕ^1, ϕ^2, ϕ^3 in Algorithm 3 for Figs. 2 and 3.

Table 1
The segmentation accuracy (SA) for different algorithms in Figs. 2 and 3. Bold values mean the highest segmentation accuracy (CA) values.

	SA		
	Graph cut	Algorithm 1	Algorithm 3
Fig. 2	0.9782	**0.9850**	**0.9850**
Fig. 3	0.9774	0.9845	**0.9900**

Recently, Yuan, *etc.* [39] have proposed a continuous max flow model for the Potts model. In their particular graph, the vertices are copied as K layers for the K-phase segmentation, see Fig. 1 for the structure of the graph. To obtain the continuous Potts model, the weights of edges connected to source are constrained as the same but without any maximum capability for these edges in their graph. To be more precisely, the weights of the t-links connected to source in the graph cannot be determined in advance of the computation, thus in the discrete case, the max-flow on such a graph cannot be found by the existing max-flow algorithms. Compared to their graph, the maximum flow of the proposed graph can be efficiently found by the augmented path based max-flow algorithm. In the continuous case, our method solves one less unknown variable function than theirs because of the super-level set representation.

In [17], Pock, *etc.* proposed a functional lifting method (FLM) for the multiphase label problem using the super-level set representation. In FLM, the data term is L^1 norm and there is no linear inequality constraint for the super-level set. Compared to FLM, the data term in our method is linear and thus it is easier to minimize. Moreover, just as mentioned earlier, the regularization term in the FLM is different from the Potts model. To represent the length of the clusters boundaries more precisely, in [21], the authors used a tight approximation for the regularization term. For the K-phase segmentation, it requires to find the intersection of projections of the dual variables on $\frac{K(K+1)}{2}$ convex sets, which does not have a closed-form solution and thus the calculation of such a projection is often time-consuming. Our graph cut method can be regraded as a direct graph cut implementation for the FLM continuous convex method in some sense. The continuous max-flow algorithm can also be regarded as the dual algorithm of FLM. For our improved continuous methods, we use another dual representation for TV of a vector-valued function whose component functions satisfy a partition condition. Such a method

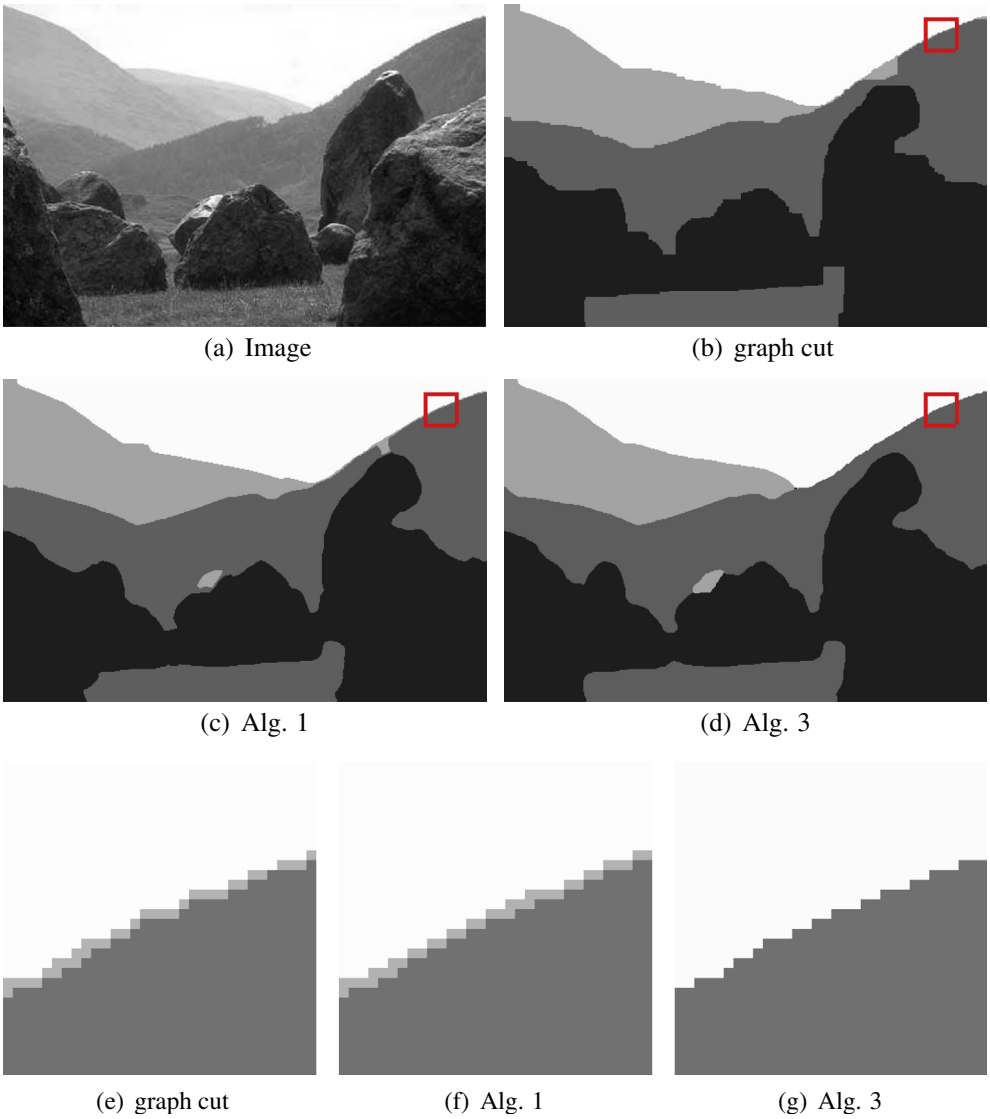


Fig. 5. Segmentation results with different algorithms. The number of the clusters $K = 4$. The last row is the corresponding enlargement of the red rectangle area in the segmentation results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

can enable us to compute the projections of the dual variables on one convex set, which is easier than [21].

4. Experimental results

In this paper, the data term is set as $d^k(x) = |I(x) - c^k|^\lambda,\ \lambda = 1, 2$, where $I(x) \in [0, 1]$ is image. The intensity means c^k can be obtained by K-means algorithm or chosen as some uniformly distributed numbers in interval $[0, 1]$, and c^k also can be updated by the alternating algorithm. In the following experiments, the convergence criterion of the algorithms is chosen as $||\boldsymbol{\phi}^{(v+1)} - \boldsymbol{\phi}^{(v)}||^2 < \epsilon||\boldsymbol{\phi}^{(v)}||^2$, where ϵ is alway set as 1.0×10^{-6}.

For the discrete model (16), we employ the algorithm in [25] to implement the max-flow algorithm.

From the layer cake formula (11), we can represent the segmentation results by label function

$$l(x) = \sum_{k=1}^{K-1} \phi^k(x).$$

In the following experiments, we will use it to show the results.

All the tested natural images are chosen from the Berkeley segmentation dataset. We test lots of them and only show some results in this paper. Others have the similar results as displayed here.

4.1. Choice of parameters

In the following experiments, μ is manually chosen. The time step parameter τ_ϕ for $\boldsymbol{\phi}$ is set as 1.0 in both Algorithms 1 and 3. The penalty parameter for augmented Lagrangian method in Algorithm 1 is chosen as $r = 20\mu$, while the time step parameter for dual variable $\boldsymbol{g}$ in Algorithm 3 is set as $\tau_{\boldsymbol{g}} = 0.1$.

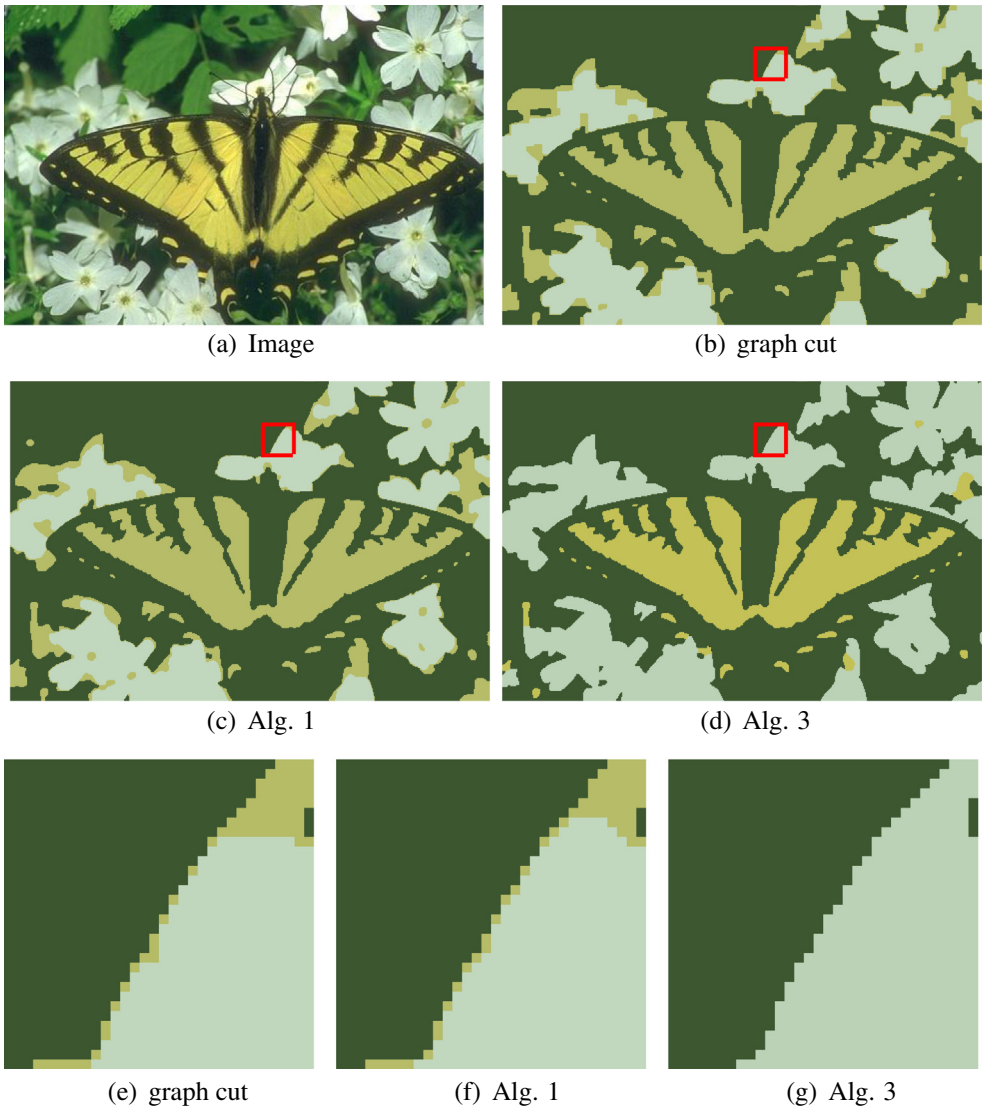
In the next, we shall make some comparisons among the proposed algorithms and several related methods.

4.2. Comparisons

4.2.1. Comparison of the discrete model and continuous model

The first example is to illustrate the difference between discrete models and continuous models. The main difference for the two methods is the regularization term. As mentioned earlier, the TV of the label function may have different weights for different boundaries, and it is not exactly equal the length of boundaries. This would lead to some undesirable segmentation results.

In Fig. 2(a), a synthesized piecewise constant image is shown. In this image, from the boundary to center, the intensity values of the areas in the image are set as 0, $\frac{60}{255}$, $\frac{120}{255}$, $\frac{180}{255}$, respectively. The noisy image corrupted by heavy Gaussian noise with variance $\sigma^2 = 0.2$ is



(a) Image (b) graph cut (c) Alg. 1 (d) Alg. 3 (e) graph cut (f) Alg. 1 (g) Alg. 3

Fig. 6. Segmentation results with different algorithms. The number of the clusters $K = 3$. The last row is the corresponding enlargement of the red rectangle area in the segmentation results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

demonstrated in Fig. 2(b). We implement different algorithms to partition the noisy image into 4 clusters by fixed means $\boldsymbol{c} = (0.2, 0.4, 0.6, 0.8)$. The segmentation results with the discrete graph cut method, dual continuous methods (Algorithms 1 and 3) are displayed in Fig. 2(c)–(e). The regularization parameters are all set as $\mu = 0.3$ in these algorithms. It can be found that there are more zigzags in the result provide by graph cut than continuous method's. The reason of this phenomena is that the discrete graph cut method suffers from metrication errors caused by anisotropic TV. The isotropic TV in continuous method has a better performance. In this case, since each region is nested and the label function l takes the values 0, 1, 2, 3 from outside to inside, thus $\int_\Omega |\nabla l| dx$ is exactly equal the length of the boundaries and there is no undesirable phenomena expect for the zigzags near the boundaries. But if we changed the original image intensity values as 0, $\frac{120}{255}$, $\frac{180}{255}$, $\frac{60}{255}$ from outside to inside and other settings are all kept to be the same as in the previous case, then we get the results displayed in Fig. 3. Now, the label function $l(x)$ take the values 0, 2, 3, 1 from outside to inside. Besides the zigzags, it can be seen from this figure that the results provided by graph cut and Algorithm 1 have a narrow bands (label 1) near the boundary between labels 0 and 2. The similar phenomena also occurs in the boundary between the areas labeled with 1 and 3. But it does not take place near the boundary between label 2 and 3. Please see the local areas of the red rectangle in the segmentation results. For details, we show the enlargement of them in Fig. 3(f)–(i). This phenomena is caused by the different weights of the boundaries in the regularization terms for graph cut method and Algorithm 1. As can be seen from the label function, the length of boundaries between labels 0 and 2, 1 and 3 is computed 2 times in the regularization term $\int_\Omega |\nabla l| dx = \sum_{k=1}^{3} \int_\Omega |\nabla \phi^k| dx$, but only 1 time for boundary between label 2 and 3. These different weights produce spitting boundaries and they would not be disappeared by increasing the value of the regularization parameter. The finally calculated super-level set function ϕ (in Algorithm 3, others are similar) for both of the cases are illustrated in Fig. 4. It can be seen that they are all nearly binary in practice.

To quantitatively evaluate the segmentation accuracy (SA) of different results, we use an index

$$\mathrm{SA} = \frac{N_c}{N_t}$$

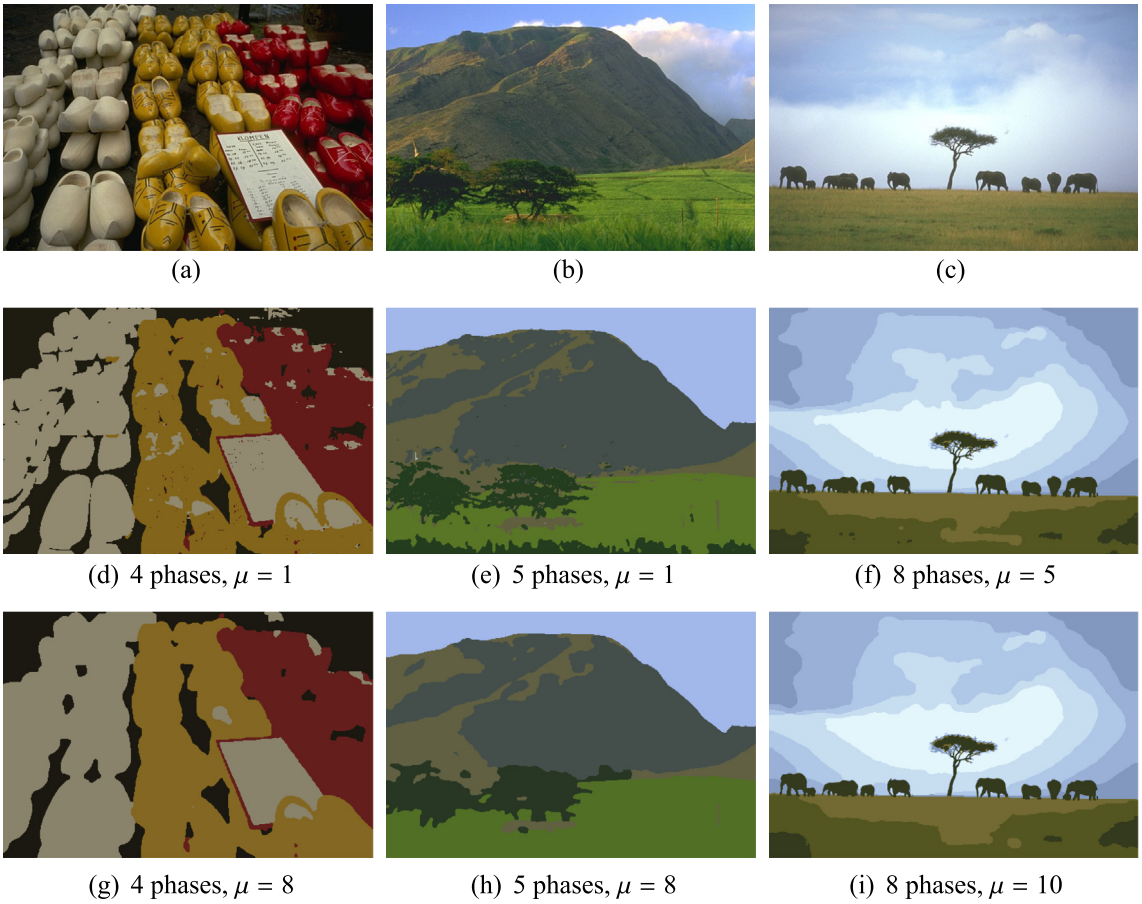


Fig. 7. Segmentation results of the proposed Algorithm 3 for some natural images.

Table 2
The average CPU time (seconds) for solving maximum flows of the Ishikawa's graph [28] and the proposed graph using the max-flow algorithm [25]. (The sizes of Figs. 5(a) and 6(a) are both 321×481).

	Fig. 5(a)					Fig. 6(a)		
	$K = 2$	$K = 3$	$K = 4$	$K = 8$	$K = 10$	$K = 2$	$K = 3$	$K = 4$
Ishikawa's [28]	0.049	0.350	1.843	3.727	4.907	0.049	0.208	0.479
Proposed	0.047	0.106	0.183	0.716	1.384	0.044	0.098	0.196

to indicate the quality of the results. Here N_c is the number of the correctly segmented pixels by the segmentation algorithms and N_t is the total number of the pixels in the images. The SA values of the segmentation results in this experiment are all reported in Table 1. It also can be found that the continuous methods have better performance (higher SA values) than the discrete approach.

Other results for segmenting natural images with different algorithms are displayed in Figs. 5–7. For color images, we choose the segmentation data term as the vector-valued L^2 norm in RGB color channels. Since the ground truth segmentations of the natural images are not available, and thus the SA values cannot be calculated in this case. However, it can be seen from these examples that the continuous max-flow algorithm 3 looks like to produce the best segmentation results.

4.2.2. Comparison of the proposed discrete algorithm and Ishikawa's method [28]

Since every vertices of the proposed graph are connected to both source s and sink t, compared to Ishikawa's graph [28], our approach would be faster to search trees from source s to sink t in the augmenting paths based max-flow algorithms such as [25]. Hence, to solve the max-flow on the proposed graph is much faster than Ishikawa's, especially when the number of clusters K is large. In this experiment, we employ the max-flow algorithm [25] to solve the maximum flows. Our experimental platform is a laptop with 2.3 GHz Intel Core i5 CPU and Matlab R2011b. We partition the images into different clusters with updating the intensity mean c^k. When the segmentation is finished, we record the CPU time of the max-flow algorithm in each iteration and then take the average CPU time to make a comparison. Table 2 shows the results of the CPU time of solving max-flow on the proposed graph and on the Ishikawa's graph. It can be seen that both of the two methods require almost the same CPU time when $K = 2$. This is reasonable since both of the graphs would be reduced to the same one when $K = 2$. However, when $K > 2$, searching the maximum flows on the proposed graph is much faster than that by [28]. Generally speaking, based on our experiments, our method is more than 2 times faster than [28] when the number of clusters $K > 2$.

The proposed discrete method would have the same segmentation results as Ishikawa's method [28] since they are both to minimize the same energy (different with a constant term). Compared to Ishikawa's graph, the only advantage of the proposed graph is the higher computational efficiency. However, the proposed continuous max-flow Algorithms 1 and 3 can produce better segmentation results than [28]. Please see Figs. 3, 5 and 6 for examples.

4.3. Comparison of the multi-phase Chan–Vese model and proposed method

In the multi-phase Chan–Vese model [12], any K-phase segmentation can be represented by $\log_2 K$ level set functions. For the 4-phase Chan–Vese model, a relaxation convexity method has been proposed in [13]. This model can be optimized by the graph cut method provided that the data term in the energy satisfies a mild convex condition (please see [13] for details). When this condition fails, the authors have also proposed an approximation convex energy to replace the original Chan–Vese model. Here we compare the proposed method with the graph cut based 4-phase Chan–Vese model. Fig. 8 shows the results with different methods to partition some brain MR images to 4 clusters. In Fig. 8(b), the image contains Gaussian noise and thus we apply the L^2 data term to segment it. Since the level of noise is not very high and the data term is L^2 norm, thus the convex mild condition usually holds and the results produced by [13] is good. The results produced by the Chan–Vese model and the proposed algorithms are very similar except for a little smoothness difference caused by the choice of the regularization parameters. But in Fig. 8(g), the image was corrupted by heavy impulse noise and we need to use L^1 data term to classify the pixels. In this case, the convex mild condition fails and thus the segmentation result (Fig. 8(h)) with graph cut based Chan–Vese is not as good as the proposed algorithms (Fig. 8(i), (j)).

4.4. Application to stereo benchmark

The stereo benchmark problem (see e.g. [40]) is to calculate the depth mapping l between a pair of color images $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ taken

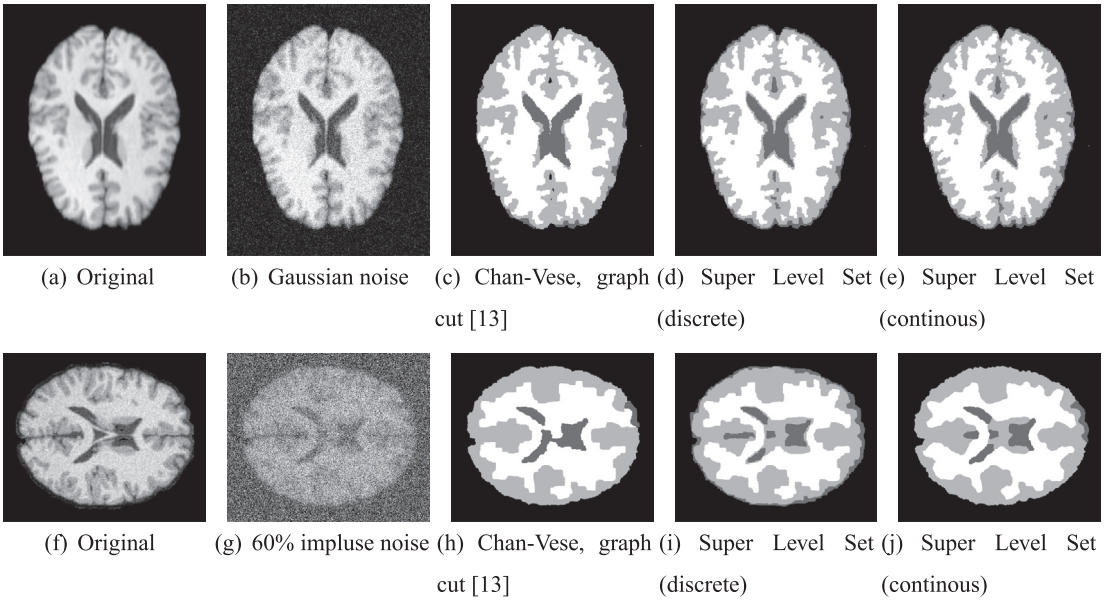


Fig. 8. Comparison of the graph cut based 4-phase Chan–Vese model [13] and the proposed algorithms. First column, original images; second column, the noisy images with different types noise; third column, the segmentation results with 4-phase Chan–Vese model using graph cut algorithm [13]; fourth column, the clusters with the super level set representation using the proposed graph cut method; fifth column, the segmentation results with the super level set representation by the proposed continuous dual model.

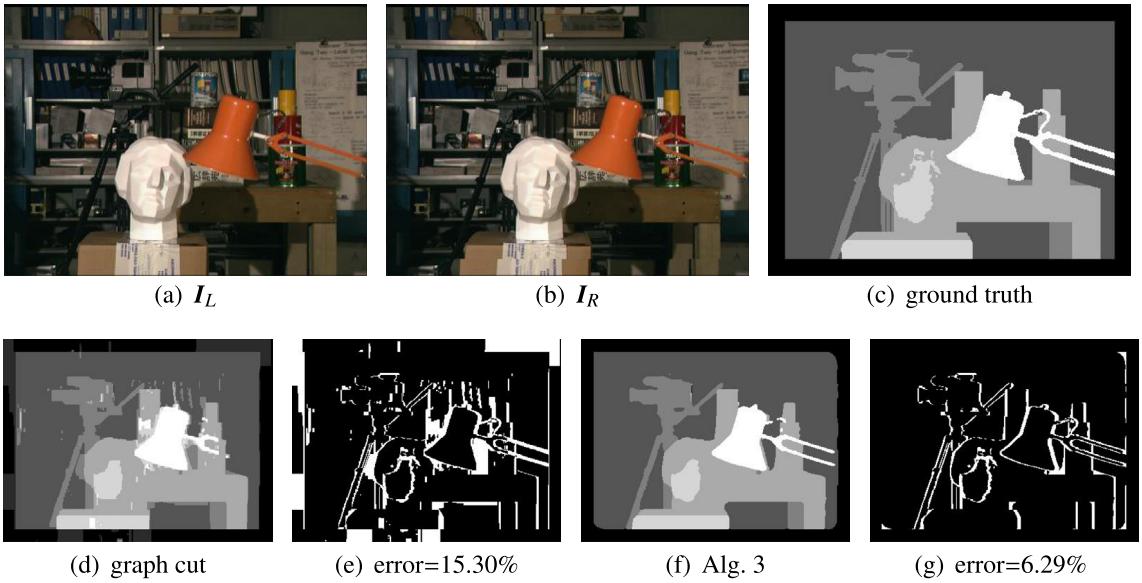


Fig. 9. Results of depth mapping estimation. The regularization parameter $\mu = 0.09$ and the number of the clusters $K = 17$ for both of the two methods.

from horizontally different viewpoints. Such a depth mapping l can be obtained by minimizing the non-convex data term

$$D(l) = \sum_{\boldsymbol{x}\in\Omega} ||\boldsymbol{I}_L(\boldsymbol{x}) - \boldsymbol{I}_R(\boldsymbol{x} + (l,0)^T)||_1 = \sum_{\boldsymbol{x}\in\Omega}\sum_{i=1}^{3} |I_L^i(\boldsymbol{x}) - I_R^i(\boldsymbol{x} + (l,0)^T)|,$$

where $I_L^i,\ I_R^i,\ i = 1,\ 2,\ 3$ stands for the ith component function of $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ in RGB color channels. Usually, a regularization term such as μTV (l) should be added to force the depth mapping to be a piecewise constant function. If l only takes integers from 1 to K, then

$$D(l) = \sum_{\boldsymbol{x}\in\Omega}\sum_{i=1}^{3}\sum_{k=1}^{K} |I_L^i(\boldsymbol{x}) - I_R^i(\boldsymbol{x} + (k,0)^T)|\delta_{k,l}.$$

By (13), $D(l)$ can be represented by a series of super-level set functions $\boldsymbol{\phi}$ and thus $D(\boldsymbol{\phi})$ is convex. Here, we use the proposed Algorithm 3 and the proposed discrete graph cut method to solve this problem.

Fig. 9 shows the results of depth mapping estimation using our method. We choose the stereo from middlebury stereo datasets. In this experiment, $\boldsymbol{I}_R$ is produced by the given input image $\boldsymbol{I}_L$ according to the ground truth depth mapping shown in Fig. 9(c). Then we use the image pair $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ to reconstruct the depth mapping. We set the phases $K = 17$ according to [40], and the regularization parameter μ is set as 0.09 for both of the discrete and continuous methods. The final results with the graph cut method and continuous Algorithm 3 are displayed in Fig. 9(d) and (f), respectively. It can be found that the result produced by Algorithm 3 is better than the discrete method's according to the errors of these two methods which are displayed in Fig. 9(e) and (g). This is because the regularization term in the continuous method is more suitable for the smoothness constraints in this problem.

5. Conclusion and discussion

In this paper, we have proposed a graph cut based continuous max-flow for the multi-phase segmentation, which is associated to the super-level set representation in continuous method. Due to the energy of the model is convex, in the discrete case, its global minimization can be exactly solved by searching a maximum flow on a special constructed graph. We experimentally show that finding the maximum flow on the proposed graph is faster than the earlier methods such as Ishikawa's [28]. Meanwhile, we mathematically show that the min cut of the proposed graph is corresponding to the super-level set representation of the PCLSM model. To overcome the drawback of PCLSM model that each boundary of the clusters may have a non-uniform weight, we propose a continuous max-flow model by using $K-1$ super-level set functions and K dual variables to partition any image into K phases. Compared to some existing methods, experimental results have illustrated that our method can improve the segmentation accuracy or computational efficiency.

Although the general Potts model cannot be exactly solved by graph cut, but a tight approximation is possible. As discussed in the paper, in the continuous model, since the regularization term in the Potts model must have some connections between $k-1$ and k layers vertices, one could improve the discrete graph cut method by considering such a relationship.

We emphasize that we can obtain the global minimization under the condition that c^k is known (corresponding to d^k is known). However, in many applications, the real intensity means c^k may be unknown. In this case, one can use the alternating algorithm to update c^k and get some good results. Theoretically, the segmentation model is no longer convex by considering c^k together, especially when the number of clusters K is unknown. Therefore, how to get a totally convex relaxation model including c^k and K is one of our future work.

Acknowledgments

Liu is partially supported by National Natural Science Foundation of China (No. 11201032) and the Fundamental Research Funds for the Central Universities, Huang is partially supported by National Natural Science Foundation of China (No. 11071023), Leung is partially supported by the HKUST Grant RPC11SC06.

Appendix A. Proof of Theorem 1

Proof. For any cut $(\mathbb{V}_s, \mathbb{V}_t)$, we let

$$\phi_p^k = \begin{cases} 0, & \text{if } v_p^k \in \mathbb{V}_s, \\ 1, & \text{if } v_p^k \in \mathbb{V}_t, \end{cases} \tag{A.1}$$

where $k = 1, 2, \ldots, K-1$. This means that there is a one-to-one relationship between the cut of $\mathbb{G}$ and $\boldsymbol{\phi} = (\phi^1, \ldots, \phi^{K-1})$ without any conditions. Denote the set of the edges in the cut as $\mathbb{E}^C = \{(v_1, v_2) \in \mathbb{E} : v_1 \in \mathbb{V}_s, v_2 \in \mathbb{V}_t\}$ and $|\mathbb{E}^C| = C(\mathbb{V}_s, \mathbb{V}_t)$. If $\boldsymbol{\phi} \notin \widetilde{\mathbb{B}}$, i.e. the condition $\phi_p^1 \geqslant \cdots \geqslant \phi_p^{K-1}$ fails, then there must be a k $(k = 1, 2, \ldots, K-2)$ such that $\phi_p^k < \phi_p^{k+1}$. Since both ϕ_p^k and ϕ_p^{k+1} are binary, we have $\phi_p^k = 0, \phi_p^{k+1} = 1$. Then in the cut $(\mathbb{V}_s, \mathbb{V}_t)$ defined by (A.1), we get $v_p^k \in \mathbb{V}_s$, $v_p^{k+1} = \mathbb{V}_t$. By the definition of the cut, we conclude $(v_p^k, v_p^{k+1}) \in \mathbb{E}^C$ and $|\mathbb{E}^C| = +\infty$. Hence such a cut is not a feasible cut. Conversely, if $\boldsymbol{\phi} \in \widetilde{\mathbb{B}}$, a similar discussion can show $(v_p^k, v_p^{k+1}) \notin \mathbb{E}^C$ and thus $\mathbb{E}^C < +\infty$, which indicates this cut is a feasible cut. The first part of the theorem has been proven.

For Eq. (20), we only need to prove that the cost of the t-links in a feasible cut is equal to the data term in $\mathcal{E}^{Lab-sup-d}$ plus a constant term since the connections between the n-links and regularization term part can be handled by a standard discussion just as [26,28,27,16]. In the constructed graph, for each vertex v_p^k, there is one and only one of the t-links (s, v_p^k) and (v_p^k, t) belongs to the edges set $\mathbb{E}^C$ of a feasible cut. For a fixed p, all the possible K cases of the feasible cut are listed as following:

1. $(v_p^1, t), \ldots, (v_p^{K-1}, t) \in \mathbb{E}^C(p)$.
   In this case, the cost of the t-links of the cut $|\mathbb{E}_t^C(p)| = \sum_{k=1}^{K-1} C(v_p^k, t) = \sum_{k=1}^{K-1} d_p^k$ and the associated $\phi_p^k = 0$. Therefore we can write
   $$|\mathbb{E}_t^C(p)| = \sum_{k=1}^{K-1} d_p^k = \sum_{k=1}^{K-1} \left(d_p^{k+1} - d_p^k\right)\phi_p^k + \sum_{k=1}^{K-1} d_p^k.$$
2. $(s, v_p^1), \ldots, (s, v_p^{K-1}) \in \mathbb{E}^C(p)$.
   In this case, $|\mathbb{E}_t^C(p)| = \sum_{k=1}^{K-1} C(s, v_p^k) = \sum_{k=1}^{K-1} d_p^{k+1}$ and the associated $\phi_p^k = 1$. Therefore we can write
   $$|\mathbb{E}_t^C(p)| = \sum_{k=1}^{K-1} d_p^{k+1} = \sum_{k=1}^{K-1} \left(d_p^{k+1} - d_p^k\right)\phi_p^k + \sum_{k=1}^{K-1} d_p^k.$$
3. For $k = 1, \ldots, K-2, (s, v_p^1), \ldots, (s, v_p^k), (v_p^{k+1}, t), \ldots, (v_p^{K-1}, t) \in \mathbb{E}^C(p)$.
   In this case, $|\mathbb{E}_t^C(p)| = \sum_{k'=1}^{k} C(s, v_p^{k'}) + \sum_{k'=k+1}^{K-1} C(v_p^{k'}, t) = \sum_{k'=1}^{k} d_p^{k'+1} + \sum_{k'=k+1}^{K-1} d_p^{k'}$ and the associated $\phi_p^{k'} = 1$ when $1 \leqslant k' \leqslant k$ and $\phi_p^{k'} = 0$ when $K-1 \geqslant k' > k$. It can be verified that we also can write
   $$|\mathbb{E}_t^C(p)| = \sum_{k=1}^{K-1} \left(d_p^{k+1} - d_p^k\right)\phi_p^k + \sum_{k=1}^{K-1} d_p^k.$$

Combining the above K cases, we have

$$\left|\bigcup_{p \in \mathbb{P}} \mathbb{E}_t^C(p)\right| = \sum_{k=1}^{K-1} \sum_{p \in \mathbb{P}} \left(d_p^{k+1} - d_p^k\right)\phi_p^k + \sum_{k=1}^{K-1} \sum_{p \in \mathbb{P}} d_p^k.$$

Please note the first term is just the data term in $\mathcal{E}^{Lab-sup-d}$. A standard discussion about the n-links and the regularization term can guarantee the conclusion as needed. □

Appendix B. Proof of Proposition 1

Proof. The proof is similar as in [30]. By introducing $K+1$ Lagrangian multiplier functions $\boldsymbol{\phi}(x) = (\phi^k(x), \ldots, \phi^K(x))$, the max-flow problem (21) can be rewritten as

$$\left\{ \min_{\boldsymbol{\phi}} \max_{f_t^k, f_s^k, f^{k,k+1}, \boldsymbol{g}^k} \left\{ \begin{array}{ll} \sum_{k=0}^{K} \int_\Omega (1 - \phi^k(x)) f_t^k(x) \mathrm{d}x & + \sum_{k=0}^{K} \int_\Omega \phi^k(x) f_s^k(x) \mathrm{d}x \\ \sum_{k=0}^{K-1} \int_\Omega (\phi^{k+1}(x) - \phi^k(x)) f^{k,k+1}(x) \mathrm{d}x & - \sum_{k=0}^{K} \int_\Omega \phi^k(x) \nabla \cdot \boldsymbol{g}^k(x) \mathrm{d}x \end{array} \right\} \right.$$
$$\left. f_s^k(x) \leqslant d^{k+1}(x), f_t^k(x) \leqslant d^k(x), f^{k,k+1}(x) \geqslant \alpha, ||\boldsymbol{g}^k(x)||_2 \leqslant \mu, k = 0, 1 \ldots, K, x \in \Omega. \right. \tag{B.1}$$

The min and max operators can be interchanged because (B.1) satisfies all the conditions of the min–max theorem [41]. It follows that there is at least one saddle point for (B.1). The first three terms in (B.1) can be reformulated as

$$\max_{f_t^k(x) \leqslant d^k(x)} (1 - \phi^k(x)) f_t^k(x) = \begin{cases} (1 - \phi^k(x)) d^k(x), & \phi^k(x) \leqslant 1, \\ +\infty, & \phi^k(x) > 1, \end{cases} \tag{B.2}$$

$$\max_{f_s^k(x) \leqslant d^{k+1}(x)} \phi^k(x) f_s^k(x) = \begin{cases} \phi^k(x) d^{k+1}(x), & \phi^k(x) \geqslant 0, \\ +\infty, & \phi^k(x) < 0, \end{cases} \tag{B.3}$$

$$\max_{f^{k,k+1}(x) \geqslant \alpha} (\phi^{k+1}(x) - \phi^k(x)) f^{k,k+1}(x) = \begin{cases} (\phi^{k+1}(x) - \phi^k(x))\alpha, & \phi^k(x) \geqslant \phi^{k+1}(x), \\ +\infty, & \phi^k(x) \leqslant \phi^{k+1}(x). \end{cases} \tag{B.4}$$

From B.2, B.3 and B.4, it follows that $1 \geqslant \phi^0 \geqslant \phi^1 \geqslant \cdots \geqslant \phi^K \geqslant 0$ because the energy of (B.1) would be infinite if this condition is failed and this contradicts with the existence of a saddle point. To combine with the dual norm of TV,

$$\max_{\boldsymbol{g} \in \mathbb{C}} - \sum_{k=0}^{K} \int_\Omega \phi^k(x) \nabla \cdot \boldsymbol{g}^k(x) \mathrm{d}x = \mu \sum_{k=0}^{K} \int_\Omega |\nabla \phi^k(x)| \mathrm{d}x,$$

the problem (B.1) becomes

$$\min \left\{ \begin{array}{ll} \sum_{k=0}^{K} \int_\Omega (1 - \phi^k(x)) d^k(x) \mathrm{d}x + & \sum_{k=0}^{K} \int_\Omega \phi^k(x) d^{k+1}(x) \mathrm{d}x \\ + \alpha \int_\Omega (\phi^K(x) - \phi^0(x)) \mathrm{d}x + & \mu \sum_{k=0}^{K} \int_\Omega |\nabla \phi^k(x)| \mathrm{d}x \end{array} \right\}.$$

Since α can be very large, the minimizer of above problem must satisfy $\phi^K = 0$ and $\phi^0 = 1$. Rearranging the expression, we get the problem

$$\min_{\boldsymbol{\phi} \in \widetilde{\mathbb{B}}^1} \sum_{k=1}^{K} \int_\Omega (\phi^{k-1} - \phi^k(x)) d^k(x) \mathrm{d}x + \mu \sum_{k=1}^{K-1} \int_\Omega |\nabla \phi^k(x)| \mathrm{d}x + \sum_{k=1}^{K} \int_\Omega d^k(x) \mathrm{d}x - \alpha |\Omega|.$$

Therefore, the conclusion holds. □

Appendix C. Proof of Theorem 2

Proof. From Proposition 1, the problem (22) is equivalent to the dual problem of convex relaxation model $\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}^1}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$. Thus $\boldsymbol{\phi}^*$ is a global minimizer of this convex relaxation model by the duality and convexity. In the next, we only need to prove that the binary function $\boldsymbol{\psi}^*_\gamma$ is the global minimizer of the binary problem $\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$. First, it is easy to verify $\boldsymbol{\psi}^*_\gamma\in\widetilde{\mathbb{B}}$ because of $\boldsymbol{\phi}^*\in\widetilde{\mathbb{B}}^1$.

By the layer cake formula [7] and coarea formula [31], we have

$$\begin{aligned}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi}) &= \sum_{k=1}^{K}\int_\Omega(\phi^{k-1}(x)-\phi^k(x))d^k(x)\mathrm{d}x+\mu\sum_{k=1}^{K-1}\int_\Omega|\nabla\phi^k(x)|\mathrm{d}x\\ &= \int_0^1\left\{\sum_{k=1}^{K}\int_\Omega(\psi_\gamma^{k-1}(x)-\psi_\gamma^k(x))d^k(x)\mathrm{d}x\right\}\mathrm{d}\gamma\\ &\quad+\mu\int_0^1\left\{\sum_{k=1}^{K-1}\int_\Omega|\nabla\psi_\gamma^k(x)|\mathrm{d}x\right\}\mathrm{d}\gamma\\ &= \int_0^1\mathcal{E}^{Lab-sup}(\boldsymbol{\psi}_\gamma)\mathrm{d}\gamma.\end{aligned}$$

Assume to the contrary that $\boldsymbol{\psi}^*_\gamma$ is not the global minimizer of $\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$, then there must exist a binary function $\hat{\boldsymbol{\psi}}\in\widetilde{\mathbb{B}}\subset\widetilde{\mathbb{B}}^1$ such that $\mathcal{E}^{Lab-sup}(\hat{\boldsymbol{\psi}})<\mathcal{E}^{Lab-sup}(\boldsymbol{\psi}^*_\gamma)$. This directly implies that

$$\mathcal{E}^{Lab-sup}(\hat{\boldsymbol{\psi}})=\int_0^1\mathcal{E}^{Lab-sup}(\hat{\boldsymbol{\psi}})\mathrm{d}\gamma<\int_0^1\mathcal{E}^{Lab-sup}(\boldsymbol{\psi}^*_\gamma)\mathrm{d}\gamma=\mathcal{E}^{Lab-sup}(\boldsymbol{\phi}^*),$$

which means that $\boldsymbol{\phi}^*$ is not a minimizer of the problem $\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}^1}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$. This contradicts with our previous condition. Therefore, we conclude that $\boldsymbol{\phi}^*_\gamma$ is a global minimizer of the binary problem $\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}}\mathcal{E}^{Lab-sup}(\boldsymbol{\phi})$. □

Appendix D. Proof of Theorem 3

Proof. To simplify the notations, we still use $\phi^0=1$ and $\phi^K=0$. Obviously, the primal problem of the dual model (23) is

$$\min_{\boldsymbol{\phi}\in\widetilde{\mathbb{B}}^1}\mathcal{E}^p(\boldsymbol{\phi})=\sum_{k=1}^{K}\int_\Omega(\phi^{k-1}-\phi^k)d^k\mathrm{d}x+\frac{\mu}{2}\sum_{k=1}^{K}\int_\Omega|\nabla\phi^{k-1}-\nabla\phi^k|\mathrm{d}x,\tag{D.1}$$

thus $\boldsymbol{\phi}^*$ is a global minimizer of (D.1) by the duality.

Let $m(x)=\min_{k\in\{1,\dots,K\}}\{d^k(x)+\nabla(g^*)^k(x)\},\forall\boldsymbol{\phi}\in\widetilde{\mathbb{B}}^1$,

$$\begin{aligned}\mathcal{E}^{Conv-sup}(\boldsymbol{\phi},\mathbf{g}^*)&=\sum_{k=1}^{K}\int_\Omega\Big((\phi)^{k-1}(x)-(\phi)^k(x)\Big)\Big(d^k(x)+\nabla\cdot(\mathbf{g}^*)^k(x)\Big)\mathrm{d}x\\ &\geqslant\sum_{k=1}^{K}\int_\Omega\Big((\phi)^{k-1}(x)-(\phi)^k(x)\Big)m(x)\mathrm{d}x=\int_\Omega m(x)\mathrm{d}x.\end{aligned}$$

Since $(\boldsymbol{\phi}^*,\mathbf{g}^*)$ is a saddle of problem (23) and the minimizer of $\min\{d^1(x)+\nabla\cdot(\mathbf{g}^*)^1(x),\dots,d^K(x)+\nabla\cdot(\mathbf{g}^*)^K(x)\}$ is unique, we have that

$$(\phi^*)^{k-1}(x)-(\phi^*)^k(x)=\begin{cases}1, & \text{when } k=\arg\min_{i\in\{1,\dots,K\}}\{d^i(x)+\nabla(g^*)^i(x)\},\\ 0, & \text{else},\end{cases}$$

is the unique minimizer of $\mathcal{E}^{Conv-sup}(\boldsymbol{\phi},\mathbf{g}^*)$ with respect to $\boldsymbol{\phi}$.

Let $\psi^k=\phi^{k-1}-\phi^k$, $(\psi^*)^k=(\phi^*)^{k-1}-(\phi^*)^k$. Since

$$\begin{pmatrix}\psi^1\\ \psi^2\\ \psi^3\\ \vdots\\ \psi^K\end{pmatrix}=\begin{pmatrix}1&-1&0&\cdots&0&0\\ 0&1&-1&\cdots&0&0\\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots\\ 0&0&0&\cdots&1&-1\\ 0&0&0&\cdots&0&1\end{pmatrix}\begin{pmatrix}1\\ \phi^1\\ \phi^2\\ \vdots\\ \phi^{K-1}\end{pmatrix},$$

this linear transformation is reversible and thus there is a one-to-one relationship between ψ^k and ϕ^k, which means the problem (D.1) is equivalent to the convex relaxation Potts model

$$\min_{\boldsymbol{\psi}\in\widetilde{\mathbb{B}}}\mathcal{E}^{Potts}(\boldsymbol{\psi})=\sum_{k=1}^{K}\int_\Omega\psi^k d^k\mathrm{d}x+\frac{\mu}{2}\sum_{k=1}^{K}\int_\Omega|\nabla\psi^k|\mathrm{d}x.\tag{D.2}$$

Hence, $\boldsymbol{\psi}^*=((\psi^*)^1,\dots,(\psi^*)^K)$ is the global minimizer of problem (D.2). By the binary of ψ^*, we conclude that it is also the global minimizer of the binary Potts model. □

References

[1] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Commun. Pure Appl. Math. 42 (1989) 577–685.
[2] X. Cai, R. Chan, T. Zeng, A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding, SIAM J. Imag. Sci. 6 (1) (2013) 368–390.
[3] Y. Li, J. Kim, Multiphase image segmentation using a phase-field model, Comput. Math. Appl. 62 (2) (2011) 737–745.
[4] Y.M. Jung, S.H. Kang, J. Shen, Multiphase image segmentation via Modica–Mortola phase transition, SIAM J. Appl. Math. 67 (5) (2007) 1213–1232.
[5] T.F. Chan, L.A. Vese, Active contours without edges, IEEE Trans. Image Process. 10 (2) (2001) 266–277.
[6] J. Lie, M. Lysaker, X. Tai, A binary level set model and some applications to Mumford–Shah image segmentation, IEEE Trans. Image Process. 15 (5) (2006) 1171–1181.
[7] T.F. Chan, S. Esedoglu, M. Nikolova, Algorithms for finding global minimizers of image segmentation and denoising models, SIAM J. Appl. Math. 66 (5) (2006) 1632–1648.
[8] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imag. Vision 20 (1) (2004) 89–97.
[9] T. Goldstein, S. Osher, The split Bregman method for L1 regularized problems, SIAM J. Imag. Sci. 2 (2009) 323–343.
[10] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, S. Osher, Fast global minimization of the active contour/snake model, J. Math. Imag. Vision 28 (2) (2007) 151–167.
[11] T. Goldstein, X. Bresson, S. Osher, Geometric applications of the split Bregman method: segmentation and surface reconstruction, J. Sci. Comput. 45 (2010) 272–293.
[12] L. Vese, T. Chan, A multiphase level set framework for image segmentation using the Mumford and Shah model, Int. J. Comput. Vision 50 (3) (2002) 271–293.
[13] E. Bae, X. Tai, Efficient global minimization for the multiphase Chan–Vese model of image segmentation, in: Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, Springer, 2009, pp. 28–41.
[14] E. Bae, X. Tai, Efficient Global Minimization Methods for Image Segmentation Models with Four Regions, Tech. Rep, 11-82, UCLA CAM Report, 2011.
[15] J. Lie, M. Lysaker, X. Tai, A variant of the level set method and applications to image segmentation, Math. Comput. 75 (255) (2006) 1155–1174.
[16] E. Bae, X. Tai, Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation, in: Scale Space and Variational Mehotds in Computer Vision, 2009, pp. 1–13.
[17] T. Pock, T. Schoenemann, G. Graber, H. Bischof, D. Cremers, A convex formulation of continuous multi-label problems, in: European Conference on Computer Vision 2008 (ECCV2008), 2008, pp. 792–805.
[18] L. Bertelli, S. Chandrasekaran, F. Gibou, B. Manjunath, On the length and area regularization for multiphase level set segmentation, Int. J. Comput. Vision 90 (3) (2010) 267–282.
[19] R. Potts, Some generalized order-disorder transformations, Math. Proc. Cambridge Philos. Soc. 48 (1952) 106–109.
[20] E. Bae, J. Yuan, X. Tai, Global minimization for continuous multiphase partitioning problems using a dual approach, Int. J. Comput. Vision 92 (1) (2011) 112–129.
[21] T. Pock, A. Chambolle, D. Cremers, H. Bischof, A convex relaxation approach for computing minimal partitions, in: IEEE CVPR2009, 2009.
[22] J. Yuan, E. Bae, X. Tai, Y. Boykov, A continuous max-flow approach to Potts model, in: ECCV'10 Proceedings of the 11th European conference on Computer vision: Part VI, 2010, pp. 379–392.

[23] E. Bae, J. Yuan, X. Tai, Y. Boykov, A Fast Continuous Max-Flow Approach to Non-Convex Multilabeling Problems, Tech. Rep. 10-62, UCLA CAM Report, 2010.
[24] D. Greig, B. Porteous, A. Seheult, Exact maximum a posteriori estimation for binary images, J. Roy. Stat. Soc. Ser. B (1989) 271–279.
[25] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2001) 359–374.
[26] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1–18.
[27] V. Kolmogorov, R. Zabih, What energy functions can be minimized via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2) (2004) 147–159.
[28] H. Ishikawa, Exact optimization for markov random fields with convex priors, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1333–1336.
[29] J. Darbon, M. Sigelle, Image restoration with discrete constrained total variation Part II: levelable functions, convex priors and non-convex cases, J. Math. Imag. Vision 26 (3) (2006) 277–291.
[30] J. Yuan, E. Bae, X. Tai, A study on continuous max-flow and min-cut approaches, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA, 2010, pp. 2217–2224.
[31] W. Fleming, R. Rishel, An integral formula for total gradient variation, Arch. Math. 11 (1960) 218–222.
[32] J. Liu, H. Zhang, Image segmentation using a local GMM in a variational framework, J. Math. Imag. Vision. http://dx.doi.org/10.1007/s10851-012-0376-5.
[33] C. Zach, D. Gallup, J.M. Frahm, M. Niethammer, Fast global labeling for real-time stereo using multiple plane sweeps, in: Vision, modeling and visualization workshop (VMV), 2008.
[34] J. Lellmann, J. Kappes, J. Yuan, F. Becker, C. Schnorr, Convex multi-class image labeling by simplex-constrained total variation, in: Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision (SSVM09), 2009, pp. 150–162.
[35] J. Liu, Y. Ku, S. Leung, Expectation–maximization algorithm with total variation regularization for vector-valued image segmentation, J. Visual Commun. Image Represent. 23 (8) (2012) 1234–1244.
[36] X. Tai, C. Wu, Augmented lagrangian method, dual methods and split Bregman iteration for ROF model, in: Proceeding SSVM '09 Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision, 2009, pp. 502–513.
[37] C. Wu, X. Tai, Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models, SIAM J. Imag. Sci. 3 (3) (2010) 300–339.
[38] A. Chambolle, D. Cremers, T. Pock, A Convex Approach to Minimal Partitions, Tech. Rep., Technical Report, Ecole Polytechnique, France, 2011.
[39] J. Yuan, E. Bae, X. Tai, Y. Boykov, A Fast Continuous Max-Flow Approach to Potts Model, Tech. Rep. 11-85, UCLA CAM Report, 2011.
[40] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vision 47 (1–3) (2002) 7–42.
[41] I. Ekeland, R. Temam, Convex Analysis and Variational Problems, SIAM Classics in Applied Mathematics, SIAM, Philadelphia, USA, 1999.