A DIRECT APPROACH TOWARDS GLOBAL MINIMIZATION FOR MULTIPHASE LABELING AND SEGMENTATION PROBLEMS

YING GU†, LI-LIAN WANG† AND XUE-CHENG TAI‡

Abstract. This work intends to extend the results of [Bae, Yuan and Tai [1], *Global minimization for continuous multiphase partitioning problems using a dual approach*, IJCV, 2011] in several directions. First, we propose a direct primal-dual approach for global minimization of the continuous Potts model with applications to piecewise constant Mumford-Shah model for multiphase image segmentation. We provide sufficient and necessary conditions to guarantee a global minimum. The conditions for the global optimum are obtained from a direct binary setting without using convex relaxation. Moreover, some numerical schemes are proposed. The underlying algorithms involve almost the minimum number of parameters, are fast and easy to implement, and usually can produce global optimum.

1. Introduction

The multiclass labeling and multiphase segmentation problems share some similarity in nature, as typically both of them aim to find a partition of an image into m disjoint regions (phases or classes) according to some optimization rule. As a matter of fact, the former term is commonly used in computer vision, while the latter one often appears in the variational and PDE community. A typical model for piecewise constant multiphase image segmentation that minimizes the total interface of edge sets is the piecewise constant Mumford-Shah model (PCMSM) [21]:

$$\min_{\{c_i,\Omega_i\}_{i=1}^m} \Big\{ E_{\rm PCMS}\big(\{c_i\}_{i=1}^m,\Gamma\big) := \lambda \sum_{i=1}^m \int_{\Omega_i} |c_i - I|^2 d\boldsymbol{x} + |\partial\Omega_i| \Big\}, \quad \lambda > 0, \tag{1.1}$$

where $I : \Omega \to \mathbb{R}$ is the input image, $\{c_i\}$ are the optimal mean values and the subregions $\{\Omega_i\}$ forms a non-overlapping partition of Ω (with $\Gamma = \cup_{i=1}^m \partial\Omega_i$). With an appropriate numerical implementation, this model has many applications in segmenting images with nearly piecewise constant intensities or in finding a simplified "cartoon" approximation of a given image (see, e.g., [9, 17, 4] and the references therein). The development of fast and robust methods for multi-phase segmentation has attracted many recent attentions and been yet challenging. Indeed, even the intensity values $\{c_i\}$ are known a prior, minimizing the

2000 *Mathematics Subject Classification.* 65N22, 65N55, 74S20, 49J40.

Key words and phrases. Multiclass labelling, multiphase segmentation, Mumford-Shah model, Potts model, primal-dual formulation, global optimum, augmented Lagrangian method, Chambolle's algorithm.

This work is supported by Singapore MOE Grant T207B2202 and Singapore NRF2007IDM-IDM002-010.

†Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore.

‡ Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore, and Department of Mathematics, University of Bergen, Norway.

PCMSM is a hard task. In this situation, it turns out to be a special case of the (continuous) Potts model [27] for multiclass labeling without favoring ordering:

$$\begin{aligned}&\min_{\{\Omega_i\}_{i=1}^m}\Big\{E_{\rm PTS}\big(\Gamma\big):=\lambda\sum_{i=1}^m\int_{\Omega_i}f_i(\boldsymbol{x})\,d\boldsymbol{x}+\sum_{i=1}^m|\partial\Omega_i|\Big\},\\&\text{subject to}\ \ \cup_{i=1}^n\Omega_i=\Omega\ \ \text{with}\ \ \Omega_k\cap\Omega_l=\emptyset,\ \forall k\neq l,\end{aligned}\tag{1.2}$$

where the parameter $\lambda>0$, and $\{f_i\}_{i=1}^m$ can be viewed as intrinsic forces to enforce the criterion of classification and labeling. In general, the multi-class labeling problem is to assign each pixel $\boldsymbol{x}\in\Omega$ a unique label l from a set of m labels (or classes) $\{l_1,\cdots,l_m\}$. It is known that it is NP-hard when it is approached in a "discrete" manner, as the number of unknowns grows exponentially with the size of the problem. Accordingly, many recent attempts tackle this problem from a continuous point of view by minimizing the continuous Potts model (1.2) (see, e.g., [11, 26, 24, 5, 1] and the references therein). An important issue on solving (1.2) is to find a convex approximation of the non-convex problem:

$$\min_{\{u_i\}_{i=1}^m}\Big\{E\big(\{u_i\}_{i=1}^m\big):=\lambda\sum_{i=1}^m\int_{\Omega}f_i(\boldsymbol{x})u_i(\boldsymbol{x})d\boldsymbol{x}+\sum_{i=1}^m TV(u_i)\Big\},\tag{1.3}$$

subject to $u_i\in\{0,1\}$ and $\sum_{i=1}^m u_i=1$ on Ω, where $\{u_i\}$ are the labeling functions and the total variation (TV) is defined by

$$TV(u)=\int_\Omega|Du|=\sup_{\boldsymbol{p}\in S}\int_\Omega u\,\mathrm{div}\boldsymbol{p}\,d\boldsymbol{x},\tag{1.4}$$

with

$$S:=\Big\{\boldsymbol{p}=(p_1,p_2)\in C_c^1(\Omega;\mathbb{R}^2)\ :\ |\boldsymbol{p}|\le 1,\ \forall\boldsymbol{x}\in\Omega\Big\},\tag{1.5}$$

where $|\boldsymbol{p}|=\sqrt{p_1^2+p_2^2}$. The optimum u_i is expected to be the indictor function $\mathbf{1}_{\Omega_i}$ of Ω_i, i.e., a binary value.

Indeed, a major class of methods is based on the convex relaxation of the admissible set by allowing for the labeling functions to take "intermediate" values from the unit simplex:

$$\Delta_+:=\Big\{\vec{u}:=(u_1,\cdots,u_m)\in\mathbb{R}^m\ :\ u_i\in[0,1]\ \text{ and }\ \sum_{i=1}^m u_i=1,\ \forall\boldsymbol{x}\in\Omega\Big\}.\tag{1.6}$$

Here, we just mention a few contributions along this line. Zack et al. [36] minimized the energy in (1.3) over (1.6) by introducing additional variables to decouple the TV-term and the data term, while Lellmann et al. [11] replaced the TV-term by non-isotropic TV: $\int_\Omega(\sum_i|\nabla u_i|^2)^{1/2}d\boldsymbol{x}$, and applied a Douglas-Rachford splitting algorithm for solving the modified model (also refer to [12] for an up-to-date review). Pock et al. [26] proposed a primal-dual projected gradient algorithm for minimizing an energy functional different from (1.3) that produced a tighter bound on the energy and often (computationally) led to global minimizer. Li et al. [15, 14] applied the same labeling technique to the piecewise constant Mumford-Shah-type model with data fidelity involving fuzzy membership functions, and minimized the resulting model by using alternative minimization methods as in [4]. More recently, Bae, Yuan and Tai [1] proposed a dual-type method for the approximate model:

$$\min_{\vec{u}\in\Delta_+}E_\varepsilon(\vec{u}):=\sum_{i=1}^m\Big\{TV(u_i)+\int_\Omega\big(\lambda f_iu_i+\varepsilon u_i\log u_i\big)d\boldsymbol{x}\Big\},\tag{1.7}$$

where $\varepsilon > 0$ is a sufficiently small. The additional log-sum term can be essentially viewed as a penalization of the constraint $u_i = 0, 1$. Sufficient conditions for the existence of global optimum solution of the primal-dual model were derived. Moreover, the comprehensive comparison conducted in [1] demonstrated the dual-type approach outperformed most of the existing approaches mentioned above, and could achieve global minimizers at least computationally. In [19], the TV-term in (1.3) was replaced by another term which approximates the length term. Afterward, a special linearization technique was used to derive an algorithm that was shown to be fast without solving any equations (since all subproblems admitted explicit solutions, and certain narrow-band technique could be applied to accelerate the algorithm).

Another family of methods based on convex formulation via functional lifting and embedding in a higher-dimensional space has also attracted much recent attention, see, e.g., [22, 26, 5, 6, 25].

This paper is motivated by [1], and the main features and contributions can be summarized as follows.

- Different from most of the previous works, we stick with the binary setting (i.e., without relaxing the admissible set as in (1.6)) by seeking the minimizer from

$$\mathcal{A} := \Big\{\vec{u} = (u_1, u_2, \cdots, u_m) \,:\, u_i^2 = u_i,\ \sum_{i=1}^{m} u_i = 1\Big\}, \tag{1.8}$$

  and conduct a direct analysis for the primal-dual formulation of (1.3):

$$\min_{\vec{u}\in\mathcal{A}} \max_{\vec{\boldsymbol{p}}\in\boldsymbol{\mathcal{S}}} \Big\{E(\vec{u}, \vec{\boldsymbol{p}}) := \sum_{i=1}^{m} \int_{\Omega} u_i\big(\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big) d\boldsymbol{x}\Big\}, \tag{1.9}$$

  where $\vec{\boldsymbol{p}} = (\boldsymbol{p}_1, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_m)$ and $\boldsymbol{\mathcal{S}} = \mathcal{S}^m$. We provide the sufficient and necessary conditions for finding the global optimum of the primal-dual model (1.9). Although some results have been derived earlier in [1], the arguments and means are quite different.
- Based on the augmented Lagrangian method, we obtain an explicit expression of the primal variables in terms of the dual variables, where the primal variables take the binary values in most of the situations. As a result, the thresholding is not necessary in this case. Moreover, the iterative algorithm contains almost the minimum number of parameters and is easy to initialize. The essential step is very analogous to the Chambolle's algorithm [7] for image denoising, so the method is expected to be fast and robust.
- The approach can also be interpreted as a multiphase level-set method based on piecewise constant interpolation of the phases (or classes), so it is anticipated to be stabler than the classical level-set method [32] and the piecewise constant level-set method [17] (a global polynomial interpolation).

The rest of the paper is organized as follows. We formulate and analyze the algorithm in the forthcoming section. More precisely, we derive the sufficient and necessary conditions for finding global optimum of the primal-dual model (1.9) without using convex relaxation. We develop the algorithm which can oftentimes achieve the global optimum. We conduct

a comparison study and provide ample numerical results to show the strengths and performance of the proposed method for multiclass labeling and multiphase image segmentation in the last section.

2. Analysis of the primal-dual model and the algorithm

In this section, we conduct analysis of the global optimum of the primal-dual model (1.9), and introduce the fast algorithms. Due to similar nature of the labeling and segmentation problems, we shall not distinguish the words "class" and "phase", and likewise "multiclass" and "multiphase" throughout the paper.

2.1. Two-phase case.

To provide some insights of the algorithms for multiphase problems, we first discuss the two-phase case. In this situation, the admissible set (1.8) can be characterized by a single labeling function u_1 as $u_2 = 1 - u_1$. For convenience, we define $u = 2u_1 - 1$, so we have $u_1 = (1+u)/2$ and $u_2 = (1-u)/2$. Accordingly, the constraint on u_1 and u_2 becomes $u^2 = 1$, and the primal-dual formulation (1.9) is reduced to

$$\min_{u^2=1} \max_{\boldsymbol{p}\in\mathcal{S}} \big\{ L(u,\boldsymbol{p}) := \langle u, \mathrm{div}\boldsymbol{p} + \lambda g\rangle \big\}, \tag{2.1}$$

where $\langle u, v\rangle = \int_\Omega uv\, d\boldsymbol{x}$ and $g = (f_1 - f_2)/2$. The global optimal u is the binary value ± 1. The following characterization is essential for the development of algorithms.

Theorem 2.1. *The pair $(u^*, \boldsymbol{p}^*)$ is the optimum of the primal-dual problem (2.1), if and only if $(u^*, \boldsymbol{p}^*)$ satisfies*

$$|u^*| = 1,\ (\mathrm{div}\boldsymbol{p}^* + \lambda g) + |\mathrm{div}\boldsymbol{p}^* + \lambda g| u^* = 0 \ \textit{a.e. on}\ \Omega, \tag{2.2}$$

and

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}\in\mathcal{S}} \int_\Omega |\mathrm{div}\boldsymbol{p} + \lambda g| d\boldsymbol{x}. \tag{2.3}$$

Proof. Notice that

$$L(u,\boldsymbol{p}) \geq -\|u\|_{L^\infty(\Omega)} \|\mathrm{div}\boldsymbol{p} + \lambda g\|_{L^1(\Omega)},$$

where $\|\cdot\|_{L^p(\Omega)}$ is the usual L^p-norm. Therefore, for any u in the admissible set, we have

$$\max_{\boldsymbol{p}\in\mathcal{S}} L(u,\boldsymbol{p}) \geq \max_{\boldsymbol{p}\in\mathcal{S}} \big(-\|\mathrm{div}\boldsymbol{p} + \lambda g\|_{L^1(\Omega)} \big) = \min_{\boldsymbol{p}\in\mathcal{S}} \|\mathrm{div}\boldsymbol{p} + \lambda g\|_{L^1(\Omega)}.$$

Therefore, the optimum is attained if and only if the equality holds, that is, the second identity of (2.2) is true. □

Remark 2.1. It is worthwhile to point out that the solution of (2.1) is not unique. Indeed, the primal-dual problem (2.1) is equivalent to

$$\min_{u^2=1} \max_{\boldsymbol{p}\in\mathcal{S}\cap E_{\boldsymbol{p}}^c} \big\{ L(u,\boldsymbol{p}) := \langle u, \mathrm{div}\boldsymbol{p} + \lambda g\rangle \big\}, \tag{2.4}$$

where $E_{\boldsymbol{p}} := \{\boldsymbol{x} \in \Omega : \mathrm{div}\boldsymbol{p} + \lambda g = 0\}$. Therefore, u^* is undetermined on $E_{\boldsymbol{p}^*}$, but is well defined by

$$u^* = -\frac{\mathrm{div}\boldsymbol{p}^* + \lambda g}{|\mathrm{div}\boldsymbol{p}^* + \lambda g|}, \quad \forall \boldsymbol{x} \in E_{\boldsymbol{p}^*}^c, \tag{2.5}$$

which indicates that u^* takes the sign of $-(\mathrm{div}\boldsymbol{p}^* + \lambda g)$. □

Remark 2.2. The situation is reminiscent to the Chambolle's dual algorithm [7] for the ROF model [29]:

$$\min_u \Big\{ TV(u) + \frac{\mu}{2}\|u-f\|^2_{L^2(\Omega)} \Big\}, \quad \mu > 0, \tag{2.6}$$

where f is a given noisy image. In this case, the primal-dual problem takes the form

$$\min_u \max_{\boldsymbol{p}\in\mathcal{S}} \int_\Omega \Big(u \,\mathrm{div}\boldsymbol{p} + \frac{\mu}{2}|u-f|^2 \Big) d\boldsymbol{x}. \tag{2.7}$$

The dual algorithm in [7] is essentially based on minimizing the the dual problem:

$$\min_{\boldsymbol{p}\in\mathcal{S}} \int_\Omega |\mathrm{div}\boldsymbol{p} - \mu f|^2 d\boldsymbol{x}, \tag{2.8}$$

by solving the nonlinear equation:

$$-\nabla(\mathrm{div}\boldsymbol{p} - \mu f) + |\nabla(\mathrm{div}\boldsymbol{p} - \mu f)|\boldsymbol{p} = 0. \tag{2.9}$$

It is important to notice the difference between (2.3) and (2.8), that is, L^1-minimization versus L^2-minimization.

In [33], a different principle (cf. [4]) was adopted to derive a Chambolle-type algorithm for two-phase segmentation. However, the current approach appears more natural. □

Now, we are in a position to introduce the algorithm for the two-phase model (2.1). Following the Chambolle's dual algorithm [7], we use the augmented Lagrangian method to solve (2.3) and obtain the nonlinear equation analogous to (2.9):

$$-\nabla\Big(\frac{\mathrm{div}\boldsymbol{p} + \lambda g}{|\mathrm{div}\boldsymbol{p} + \lambda g|}\Big) + \Big|\nabla\Big(\frac{\mathrm{div}\boldsymbol{p} + \lambda g}{|\mathrm{div}\boldsymbol{p} + \lambda g|}\Big)\Big|\boldsymbol{p} = 0, \tag{2.10}$$

which can be solved by a gradient descent approach. More precisely, we consider the time-dependent counterpart of (2.10):

$$\frac{\partial \boldsymbol{p}}{\partial t} = -\nabla u - |\nabla u|\boldsymbol{p} \quad \text{with} \quad u := -\frac{\mathrm{div}\boldsymbol{p} + \lambda g}{|\mathrm{div}\boldsymbol{p} + \lambda g|}. \tag{2.11}$$

Let τ be the time step size and $\boldsymbol{p}^n$ be the approximation of $\boldsymbol{p}$ at $t = n\tau$. To avoid division by zero, we adopt the conventional regularization (cf. [23]) and define

$$u^n := -\frac{\mathrm{div}\boldsymbol{p}^n + \lambda g}{|\mathrm{div}\boldsymbol{p}^n + \lambda g|_\beta}, \tag{2.12}$$

where $|\mathrm{div}\boldsymbol{p}^n + \lambda g|_\beta = |\mathrm{div}\boldsymbol{p}^n + \lambda g| + \beta$ for a sufficiently small $\beta > 0$. Then, we resort to the semi-implicit discretization in time as in [7], to solve (2.11):

$$\frac{\boldsymbol{p}^{n+1} - \boldsymbol{p}^n}{\tau} = -\nabla u^n - |\nabla u^n|\boldsymbol{p}^{n+1} \;\; \Rightarrow \;\; \boldsymbol{p}^{n+1} = \frac{\boldsymbol{p}^n - \tau\nabla u^n}{1 + \tau|\nabla u^n|}. \tag{2.13}$$

It is seen that the L^1-minimization (2.3) induces additional nonlinearity, compared with the L^2-minimization (2.7).

In view of Remark 2.1, we adopt the MBO-type projection (see, e.g. [20, 30])

$$\mathcal{P}_{_B}(t) := \begin{cases} 1, & \text{if } \; t \geq 0, \\ -1, & \text{if } \; t < 0, \end{cases} \tag{2.14}$$

to obtain the binary value u.

We summarize the algorithm as follows, where the involved differential operators can be discretized as in [7].

Algorithm 1

1. Initialization: set $\boldsymbol{p}^0 = \mathbf{0}$ and choose $\beta, \tau > 0$;
2. For $n = 0, 1, \cdots$
   (i) Compute
   $$u^n := -\frac{\mathrm{div}\boldsymbol{p}^n + \lambda g}{|\mathrm{div}\boldsymbol{p}^n + \lambda g|_\beta};$$
   (ii) Update $\boldsymbol{p}$ by the Chambolle-type algorithm:
   $$\boldsymbol{p}^{n+1} = \frac{\boldsymbol{p}^n - \tau\nabla u^n}{1 + \tau|\nabla u^n|};$$
3. Endfor till some stopping rule meets;
4. Set
   $$u = \mathcal{P}_{_B}(u^n).$$

2.2. **Multi-phase case.** With the insights from two-phase model, we now consider the full primal-dual model (1.9). Observe that for any $\vec{u} \in \mathcal{A}$, we have

$$\begin{aligned}\sum_{i=1}^m u_i\big(\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big) &\geq \min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big\}\sum_{i=1}^m u_i \\ &= \min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big\}, \quad \forall \boldsymbol{x} \in \Omega,\end{aligned} \tag{2.15}$$

which implies

$$E(\vec{u}, \vec{\boldsymbol{p}}) \geq \int_\Omega \min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big\}\, d\boldsymbol{x}, \tag{2.16}$$

for all $(\vec{u}, \vec{\boldsymbol{p}}) \in \mathcal{A}\times\boldsymbol{S}$, where the energy functional $E(\cdot,\cdot)$ is given in (1.9). Hence, we deduce from (2.15) that

$$\max_{\vec{\boldsymbol{p}}\in\boldsymbol{S}} E(\vec{u}, \vec{\boldsymbol{p}}) \geq \max_{\vec{\boldsymbol{p}}\in\boldsymbol{S}}\Big\{E_{_D}(\vec{\boldsymbol{p}}) := \int_\Omega \min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i + \lambda f_i\big\}\, d\boldsymbol{x}\Big\},\ \forall \vec{u} \in \mathcal{A}. \tag{2.17}$$

Accordingly, *the global optimum* $(\vec{u}^*, \vec{\boldsymbol{p}}^*)$ *of the primal-dual model* (1.9) *is a pair in* $\mathcal{A}\times\boldsymbol{S}$ *such that the identity of* (2.16) *holds, and the global minimizer* $\vec{u}^*$ *of the original model* (1.3) *is in* $\mathcal{A}$ *such that the identity* (2.17) *holds.*

Similar to Theorem 1 in [1], we have the following result on the characterization of the global optimum $(\vec{u}^*, \vec{\boldsymbol{p}}^*)$.

Theorem 2.2. *Let*

$$\vec{\boldsymbol{p}}^* = \arg\max_{\vec{\boldsymbol{p}}\in\boldsymbol{S}} E_{_D}(\vec{\boldsymbol{p}}). \tag{2.18}$$

- *Suppose that* $\min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\big\}$ *has a unique minimum value for all* $\boldsymbol{x} \in \Omega$. *Then* $\vec{u}^*$ *must take the form*

$$u_k^* = \begin{cases} 1, & k = \arg\min_{1\leq i\leq m}\big\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\big\}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.19}$$

  for $1 \leq k \leq m$ *and* $\boldsymbol{x} \in \Omega$.

- *Suppose that* $\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}$ *has more than one minimum values at some* $\boldsymbol{x}\in\Omega$, *say, k values:* $\mathrm{div}\boldsymbol{p}_j^* + \lambda f_j, j\in\{j_1, j_2, \cdots, j_k\}$. *Then* $\vec{u}^*$ *must satisfy*

$$\sum_{i=1}^{k} u_{j_i}^* = 1 \quad \text{and} \quad u_j^* = 0,\ j\notin\{j_1,\cdots,j_k\}. \tag{2.20}$$

Proof. Since the first claim is a special case of the second one, so it suffices to prove the second statement. By (2.16), we have $E(\vec{u},\vec{\boldsymbol{p}}^*)\ge E_D(\vec{\boldsymbol{p}}^*)$ for any $\vec{u}\in\mathcal{A}$. Thus it is enough to show that $E(\vec{u}^*,\vec{\boldsymbol{p}}^*) = E_D(\vec{\boldsymbol{p}}^*)$. It is clear that if $\vec{u}^*$ defined by (2.20), then

$$\begin{aligned}\sum_{i=1}^{m} u_i^*(\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i) &= \sum_{i=1}^{k} u_{j_i}^*(\mathrm{div}\boldsymbol{p}_{j_i}^* + \lambda f_{j_i})\\ &= \sum_{i=1}^{k} u_{j_i}^* \min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\} = \min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\},\end{aligned}$$

for $\boldsymbol{x}\in\Omega$ and the corresponding $1\le k\le m$. Integrating over Ω leads to the identity $E(\vec{u}^*,\vec{\boldsymbol{p}}^*) = E_D(\vec{\boldsymbol{p}}^*)$.

Next, we prove that (2.20) is necessary. If there is a u_l^* with $l\in\{j_1,j_2,\cdots,j_k\}$, which does not correspond to a minimum value, i.e., $\mathrm{div}\boldsymbol{p}_l^* + \lambda f_l > \min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}$, then

$$\begin{aligned}\sum_{i=1}^{m} u_i^*(\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i) &= \sum_{i=1}^{k} u_{j_i}^*(\mathrm{div}\boldsymbol{p}_{j_i}^* + \lambda f_{j_i})\\ &= (1-u_l^*)\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\} + u_l^*\big(\mathrm{div}\boldsymbol{p}_l^* + \lambda f_l\big)\\ &= \min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\} + u_l^*\big(\mathrm{div}\boldsymbol{p}_l^* + \lambda f_l - \min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}\big).\end{aligned}$$

Therefore, the optimum value can be obtained if and only if $u_l^* = 0$. □

Remark 2.3. It is worthwhile to point out that the necessity of (2.19) and (2.20) is under the distributional sense. Indeed, we are free to change the value of $\vec{u}^*$ at a measure zero set of Ω without affecting the integral value. □

Remark 2.4. Similar analysis was conducted in [1] (see Theorem 1) based on convex relaxation. Here, we provided a direct concise proof, which did not require to switch the min-max using Karush-Kuhn-Tucher (KKT) conditions [10] and thresholding techniques. On the other hand, we showed that (2.20) is also necessary. □

We find from (2.18) that it is essential to solve the dual problem defined in (2.15). However, it is nonsmooth, so it appears very challenging to work directly on this dual model. The following smoothed-dual model was used in [1]:

$$\max_{\vec{\boldsymbol{p}}\in\boldsymbol{S}}\Big\{E_{D,\varepsilon}(\vec{\boldsymbol{p}}) := -\varepsilon\int_\Omega \log\sum_{i=1}^{m}\exp\Big(\frac{-\mathrm{div}\boldsymbol{p}_i - \lambda f_i}{\varepsilon}\Big)d\boldsymbol{x}\Big\}, \tag{2.21}$$

where $0<\varepsilon\ll 1$. It is important to note that (2.21) turns out to be the dual model of (1.7).

Hereafter, we shall take a different approach and derive the algorithm based on the augmented Lagrangian formulation. It is worthwhile to point out the augmented Lagrangian method has been recently widely used for solving minimization problems in image processing (see, e.g., [34, 31, 35] and the references therein). Here, we use the technique as a tool to derive the equations of the primal and dual variables so as to obtain an algorithm that

involves a minimum number of parameters and can be implemented as efficient as that in [7].

2.3. **Description of the algorithm.** We start with (1.9) and reformulate it into the following constrained problem:

$$\min_{\vec{\boldsymbol{q}},\vec{u}} \Big\{ \sum_{i=1}^{m} \int_{\Omega} |\boldsymbol{q}_i| d\boldsymbol{x} + \lambda \sum_{i=1}^{m} \langle f_i, u_i \rangle \Big\},$$
$$\text{subject to } \ \boldsymbol{q_i} = \nabla u_i,\ u_i^2 = u_i,\ \sum_{i=1}^{m} u_i = 1.$$

Using the notion of augmented Lagrangian method yields the unconstrained problem:

$$\begin{aligned}\min_{\vec{u},\vec{\boldsymbol{q}}} \max_{\vec{\boldsymbol{p}},\vec{\lambda}_1,\lambda_2} \Big\{ \mathcal{L}(\vec{u},\vec{\boldsymbol{q}},\vec{\boldsymbol{p}},\vec{\lambda}_1,\lambda_2) := \sum_{i=1}^{m} \Big(&\langle 1, |\boldsymbol{q_i}| \rangle + \lambda \langle f_i, u_i \rangle + \langle \boldsymbol{p_i}, \boldsymbol{q_i} - \nabla u_i \rangle \\ &+ \frac{r_1}{2} \langle \boldsymbol{q_i} - \nabla u_i, \boldsymbol{q_i} - \nabla u_i \rangle + \langle \lambda_{1i}, u_i^2 - u_i \rangle + \frac{r_2}{2} \langle u_i^2 - u_i, u_i^2 - u_i \rangle \Big) \\ &+ \Big\langle \lambda_2, \sum_{i=1}^{m} u_i - 1 \Big\rangle + \frac{r_3}{2} \Big\langle \sum_{i=1}^{m} u_i - 1, \sum_{i=1}^{m} u_i - 1 \Big\rangle \Big\},\end{aligned}$$

where $r_1, r_2, r_3 > 0$ are penalization constants and the vector-valued functions $\vec{\boldsymbol{p}} = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m)$, $\vec{\lambda}_1 = (\lambda_{11}, \cdots, \lambda_{1m})$ (with $\lambda_{1i} \geq 0$) and the scalar function λ_2 are Lagrange multipliers. The optimality conditions leads to the system:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial u_i} = \lambda f_i + \mathrm{div}\boldsymbol{p_i} + r_1 \mathrm{div}(\boldsymbol{q_i} - \nabla u_i) + \lambda_{1i}(2u_i - 1) \\ + r_2(u_i^2 - u_i)(2u_i - 1) + \lambda_2 + r_3 \Big(\sum_{i=1}^{m} u_i - 1 \Big) = 0, \qquad (2.22a)\end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{q_i}} = \frac{\boldsymbol{q_i}}{|\boldsymbol{q_i}|} + \boldsymbol{p_i} + r_1(\boldsymbol{q_i} - \nabla u_i) = 0, \qquad (2.22b)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{p_i}} = \boldsymbol{q_i} - \nabla u_i = 0, \qquad (2.22c)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_{1i}} = u_i^2 - u_i = 0, \qquad (2.22d)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = \sum_{i=1}^{m} u_i - 1 = 0. \qquad (2.22e)$$

Thus, it follows from (2.22b) and (2.22c) that

$$\boldsymbol{p_i} = -\frac{\nabla u_i}{|\nabla u_i|}, \quad 1 \leq i \leq m. \qquad (2.23)$$

In fact, $\vec{\boldsymbol{p}}$ turns out to be the dual variable. To develop an efficient algorithm, it is essential to express the primal variable $\vec{u}$ in terms of the dual variable $\vec{\boldsymbol{p}}$. We obtain from (2.22a), (2.22c), (2.22d) and (2.22e) that

$$(2u_i - 1)|h_i + \lambda_2| + (h_i + \lambda_2) = 0, \qquad (2.24)$$

where $h_i := \mathrm{div}\boldsymbol{p_i} + \lambda f_i$. It is seen that the use of the augment Lagrangian method allows us to derive the (simplified) primal-dual system (2.23)-(2.24) with a free variable λ_2. In contrast

with the usual Uzawa algorithm, this system involves a minimum number of parameters and variables and the iterative algorithm is easy to initialize, as to be shown shortly.

An essential step is to choose λ_2 and express the dual variables $\{u_i\}$ in terms of the primal variables $\{\boldsymbol{p}_i\}$. For this purpose, let h_k (resp. h_j) be (one of) the smallest (resp. the second smallest) value of $\{h_i\}_{i=1}^m$ (note: if all $\{h_i\}_{i=1}^m$ are equal, then $h_j = h_k$). By choosing

$$-\lambda_2 = h_k + \frac{h_j - h_k}{2}, \tag{2.25}$$

we find from (2.24) that $u_k = 1$ and $u_i = 0$ for all the i such that $h_i \le h_j$. It is worthwhile to point out that (2.24) can not identify u_i when $h_i + \lambda_2 = 0$, i.e., $h_j = h_k$. In the computation, we regularize (2.24) and find

$$u_i = \frac{1}{2} - \frac{h_i + \lambda_2}{2|h_i + \lambda_2|_\beta}, \tag{2.26}$$

where λ_2 is defined in (2.25), and $|h_i + \lambda_2|_\beta = |\mathrm{div}\boldsymbol{p}_i + \lambda f_i + \lambda_2| + \beta$ for sufficiently small $\beta > 0$ as in (2.12).

Next, for fixed $\vec{u}$, the equation (2.23) can be solved by applying the iterative algorithm in [7] to

$$\nabla u_i + |\nabla u_i|\boldsymbol{p_i} = \boldsymbol{0}, \quad 1 \le i \le m, \tag{2.27}$$

with $\vec{\boldsymbol{p}} \in \boldsymbol{\mathcal{S}}$, as described in (2.10)-(2.13).

Remark 2.5. As predicted by Theorem 2.2, if $\vec{\boldsymbol{p}}^*$ solves (2.18), then by defining $\vec{u}^*$ as

$$u_k^* = \begin{cases} 1, & k = \min\left\{\arg\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}\right\}, \\ 0, & \text{otherwise}, \end{cases}$$

for all $1 \le k \le m$ and $\boldsymbol{x} \in \Omega$, we have the following properties.

(i) $\vec{u}^*$ satisfies the optimality equation (2.24) with $\lambda_2 = -(\mathrm{div}\boldsymbol{p}_k^* + \lambda f_k)$.

(ii) Under the assumption of (2.19), that is, $\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}$ has a unique minimum value for all $\boldsymbol{x} \in \Omega$, we find from Theorem 2.2 that $\vec{u}^*$ is expected to be the unique global minimizer.

(iii) Only one component of $\vec{u}^*$ takes the value 1, while the other $m-1$ components are all zero. According to Theorem 2.2, the so-defined $\vec{u}^*$ is expected to be a global minimizer, since $\vec{u}^*$ satisfies (2.20). ■

Now, we are ready to present the full algorithm as follows.

Algorithm 2

1. Initialization: set $\vec{\boldsymbol{p}}^0 = \boldsymbol{0}$ and choose $\beta, \tau > 0$;
2. For $n = 0, 1, \cdots$
   (i) Compute λ_2 by (2.25);
   (ii) Compute $\vec{u}^n$ by (2.26);
   (iii) Compute $\vec{\boldsymbol{p}}^{n+1}$ by the Chambolle-type algorithm:

$$\boldsymbol{p}_i^{n+1} = \frac{\boldsymbol{p}_i^n - \tau\nabla u_i^n}{1 + \tau|\nabla u_i^n|}, \quad 1 \le i \le m;$$

3. Endfor till some stopping rule meets;

4. Set $\vec{\boldsymbol{p}}^{*} = \vec{\boldsymbol{p}}^{n+1}$ and define

$$u_k^* = \begin{cases} 1, & k = \min\big\{\arg\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}\big\}, \\ 0, & \text{otherwise}, \end{cases}$$

for all $1 \le k \le m$.

3. A comparison study and numerical results

In this section, we demonstrate the performance of the proposed algorithms by testing them on some typical images and comparing with some relevant methods.

3.1. **A comparison study.** To provide some more insights, we compare the model and proposed algorithms with the global smoothed-dual algorithm in [1], and examine the approach from the perspective of level set methodology.

3.1.1. *Global Smoothed-dual algorithm.* The dual approach in [1] is essentially based on the model (1.7), which approximates the original non-convex problem (1.3) when $0 < \varepsilon \ll 1$. The optimum of the dual variable is obtained by solving the model (2.21), which is a smoothed version of the non-smooth dual problem (2.18). This leads to the representation of the primal variables $\{u_i\}$ in terms of the dual variables $\{\boldsymbol{p}_i\}$:

$$u_i^{\varepsilon} = \frac{\exp\Big(-\frac{\mathrm{div}\boldsymbol{p}_i+\lambda f_i}{\varepsilon}\Big)}{\sum_{j=1}^m \exp\Big(-\frac{\mathrm{div}\boldsymbol{p}_j+\lambda f_j}{\varepsilon}\Big)} = \frac{1}{\sum_{j=1}^m \exp\Big(-\frac{(\mathrm{div}\boldsymbol{p}_j+\lambda f_j)-(\mathrm{div}\boldsymbol{p}_i+\lambda f_i)}{\varepsilon}\Big)}, \tag{3.1}$$

for $1 \le i \le m$ and $\boldsymbol{x} \in \Omega$. Apparently, the vector $\vec{u}^{\varepsilon}$ with components $\{u_i^{\varepsilon}\}$ is in the convex simplex Δ_+ in (1.6). Suppose that $\mathrm{div}\boldsymbol{p}_i + \lambda f_i$ is the unique minimum value of $\{\mathrm{div}\boldsymbol{p}_j + \lambda f_j\}_{j=1}^m$. Then we observe from (3.1) that $u_i^{\varepsilon} \to 1$ and $u_j^{\varepsilon} \to 0$ (for all $j \ne i$) as $\varepsilon \to 0^+$. Under this assumption, it is expected to obtain the global minimizer in the limiting process. Note that the expression (3.1) can be viewed as the counterpart of (2.24)-(2.26).

For comparison purpose, we recall the global smoothed-dual algorithm in [1].

Global Smoothed-Dual (GSD) Algorithm

1. Initialization: set $\vec{\boldsymbol{p}}^0 = \mathbf{0}$ and choose $\varepsilon, \delta > 0$;
2. For $n = 0, 1, \cdots$
   (i) Compute $\vec{u}^n$ by (3.1) with $\boldsymbol{p}_i^n$ in place of $\boldsymbol{p}_i$;
   (ii) Compute $\vec{\boldsymbol{p}}^{n+1}$ by the projection method:
$$\boldsymbol{p}_i^{n+1} = \mathrm{Proj}_{\boldsymbol{S}}(\boldsymbol{p}_i^n + \delta\nabla u_i^n), \quad 1 \le i \le m;$$
3. Endfor till some stopping rule meets;
4. Set $\vec{\boldsymbol{p}}^{*} = \vec{\boldsymbol{p}}^{n+1}$ and define
$$u_k^* = \begin{cases} 1, & k = \min\big\{\arg\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}\big\}, \\ 0, & \text{otherwise}, \end{cases}$$
for all $1 \le k \le m$.

3.1.2. *A perspective from the level set method.* It is interesting to interpret (1.3) from the perspective of level set method, where $\{u_i\}$ can be regarded as a piecewise constant interpolation of the phases (or subregions $\{\Omega_i\}$), that is, u_i takes binary values, and $u_i = 1$ for $\boldsymbol{x} \in \Omega_i$, and $u_i = 0$ elsewhere. This should be in contrast to the level set method [23] and its important variant [17] for image segmentation.

The multiphase level set framework in [32] labels the phases by the combinations of the signs of the level set functions (usually taken as the signed distance functions). Therefore, n level set functions $\{u_j\}_{j=1}^n$ can label $m = 2^n$ phases, and the corresponding indicator functions can be expressed as the Heaviside functions $H(u_j)$. Notice that the level set functions should satisfy the Ekiron equation: $|\nabla u_j| = 1$, and this constraint should be taken into account in the implementation. To avoid such a re-initialization, one may penalize this constraint as in [13, 18].

The piecewise constant level set method proposed in [17] uses one level set function to label multiple phases. It can be viewed as a global polynomial interpolation of the phases, as opposed to the piecewise constant interpolation in (1.3). More precisely, it labels the phases by

$$u = i \quad \text{if} \quad \boldsymbol{x} \in \Omega_i, \quad 1 \le i \le m, \tag{3.2}$$

and express the indicator function of Ω_i by the Lagrange basis polynomial:

$$\varphi_i = \prod_{1 \le j \le m; j \ne i} \frac{u - j}{i - j} = \frac{K_m(u)}{K_m'(i)(u - i)}, \tag{3.3}$$

i.e., $\varphi_i(i) = 1$ and $\varphi_i(j) = 0$ for $i \ne j$, where $K_m(u) = \prod_{j=1}^m (u - j)$. Under this setting, the model (1.3) is translated to

$$\min_{u \in \{1, \cdots, m\}} \Big\{ E_P\big(u\big) := \lambda \sum_{i=1}^m \int_\Omega f_i \varphi_i \, d\boldsymbol{x} + \sum_{i=1}^m TV(\varphi_i) \Big\}. \tag{3.4}$$

By imposing the constraint $K_m(u) = 0$, we obtain the primal-dual model (cf. [5]):

$$\min_{K_m = 0} \max_{\vec{\boldsymbol{p}} \in \mathbf{S}} \Big\{ F_P\big(u, \vec{\boldsymbol{p}}\big) := \sum_{i=1}^m \langle \varphi_i, \, \mathrm{div} \boldsymbol{p}_i + \lambda f_i \rangle \Big\}. \tag{3.5}$$

We see the resemblance between the primal-dual models (1.9) and (3.5). Notice that φ_i is a polynomial of u of degree m. As a result, the reduction using the augmented Lagrangian method similar to the procedure in Subsection 2.3 leads to the coupled system, where it is not possible to derive the expression of u in terms of the dual variables. Hence, the algorithm becomes extremely complicated.

3.2. **Numerical results.** In what follows, we present various numerical results to show the performance of the proposed algorithm, and compare it with the global smoothed-dual algorithm. The recent work [1] conducted a comprehensive comparison of the global smoothed-dual algorithm with several popular methods including the Alpha expansion and Alpha-Beta swap [2, 3], the method of Pock et al. [26], and the algorithm in Lellmann et al. [11]. The global smoothed-dual algorithm outperformed these algorithms in almost all the comparison tests in terms of quality of classification/segmentation, decay rate of numerical energy, efficiency of computation and ease of implementation. Hence, it somehow suffices for us to compare our method with the global smoothed-dual algorithm in [1] (see the GSD algorithm).

To set up a relatively fair criterion for comparison, we choose some typical images used in those papers for testing the algorithms, unify the choice of the parameters of the model and compare the numerical energy associated with the original model (1.3). Moreover, we use the relative dynamic error:

$$E_{l^1}(\boldsymbol{p}^{n+1}, \boldsymbol{p}^n) := \frac{\|\boldsymbol{p}^{n+1} - \boldsymbol{p}^n\|_{l^1}}{\|\boldsymbol{p}^{n+1}\|_{l^1}} \leq \eta, \tag{3.6}$$

where $\|\boldsymbol{p}\|_{l^1} = \sum |\boldsymbol{p}|$ and the summation is over all the pixels, for a prescribed tolerance $\eta > 0$ as the stopping rule, for all the comparisons below.

3.2.1. *Two-class/phase case.* As opposed to the global smoothed-dual and Algorithm 2 with $m = 2$, Algorithm 1 only requires to evolve one pair of primal and dual variables. Accordingly, the computational cost can be halved and the algorithm is expected to be more stable. Moreover, it can be viewed as a very analogy of the Chambolle's dual algorithm [7]. Indeed, we shall show that for the time step size $\tau \leq 1/8$ (the theoretical prediction by [7]), the method works well, while the global smooth-dual algorithm is relatively restrictive with the time step.

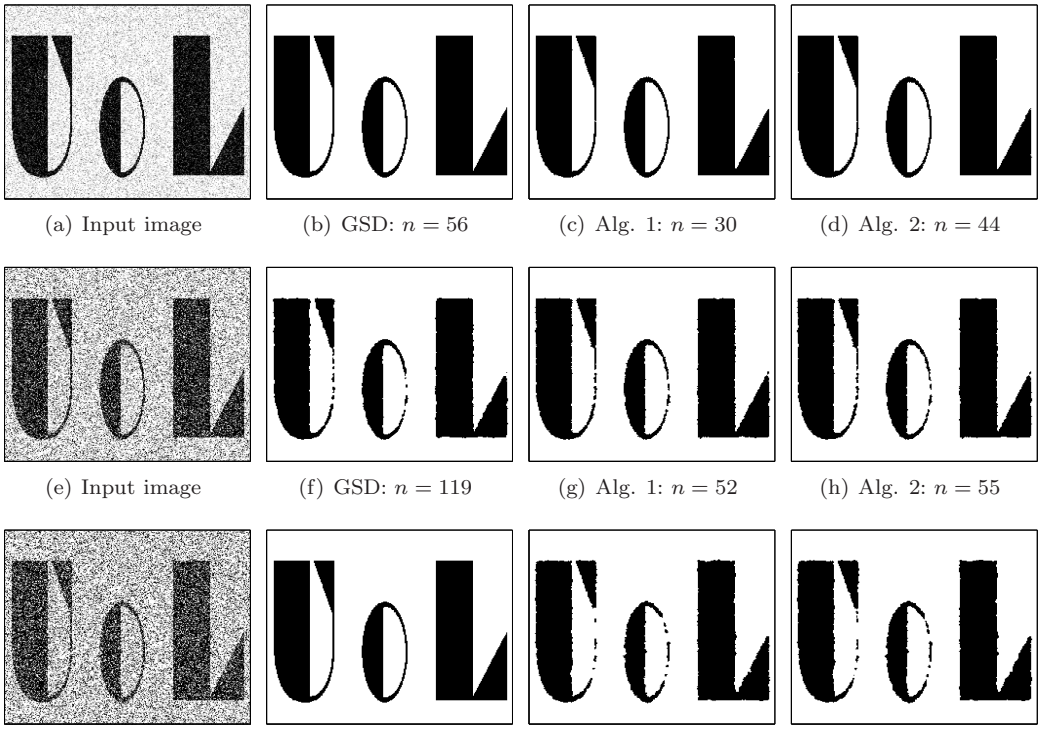
We first assume that the intensity values c_1 and c_2 are given and $f_i = |c_i - I|^2 (i = 1, 2)$, where I is the input image. The parameters, time step sizes and noise levels for three sets of tests are listed as follows. Recall that λ is the parameter in the continuous Potts model (1.3), and τ, β (resp. δ, ε) are involved in the Algorithms 1-2 (resp. the global smoothed-dual algorithm). We first test the input "UOL" image of size 256×256 with noise level from low to high. Notice that in all tests, the noise is of "Gaussian" type with zero mean and different variance d. We adopt the following setup:

- **Set 1.** Take $\lambda = 10^{-4}, \tau = \delta = 10^{-1}, \beta = 10^{-5}$, $\varepsilon = 10^{-2}$, and noise level: $d = 0.05$.
- **Set 2.** Take $\lambda = 7 \times 10^{-5}, \tau = 0.1, \delta = 0.05, \beta = 10^{-5}$, $\varepsilon = 10^{-2}$, and noise level: $d = 0.3$.
- **Set 3.** Take $\lambda = 5 \times 10^{-5}, \tau = 0.1, \delta = 0.01, \beta = 10^{-5}$, $\varepsilon = 0.05$, and noise level: $d = 0.5$.

We present in Figure 3.1 the input images with noise and segmented edge sets by three algorithms at the iteration terminated by (3.6) with $\eta = 10^{-2}$. We observe from Figure 3.1 (a)-(d) that when the noise level is low, three algorithms exhibit a similar performance, and converge quite fast. However, when we increase the noise level, see Figure 3.1 (e)-(h), Algorithm 1 yields much accurate segmentation (or classification) and converges relatively faster. However, the time step size δ in the global smoothed-dual algorithm should be chosen much smaller than τ in Algorithm 1 and Algorithm 2. We also point out the algorithm is not sensitive to the parameter β.

We plot the decay of numerical energy corresponding to the original energy functional (1.3), i.e., $E(\vec{u}^n)$ in Figure 3.2 for the comparison tests: **Sets 1-2.** Observe that Algorithms 1-2 enjoy a faster decay of energy than that of global smoothed-dual algorithm. It is also consistent with the results in Figure 3.1.

3.2.2. *Multiphase case.* We now turn to the comparison of Algorithm 2 and the global smoothed-dual algorithm for multiphase images. In this case, we take $f_i = |c_i - I|$ with $1 \leq i \leq 4$ with given $\{c_i\}$, and test the four-phase image (cf. Figure 3.3 (a)) of size 90×90



(a) Input image (b) GSD: $n = 56$ (c) Alg. 1: $n = 30$ (d) Alg. 2: $n = 44$

(e) Input image (f) GSD: $n = 119$ (g) Alg. 1: $n = 52$ (h) Alg. 2: $n = 55$

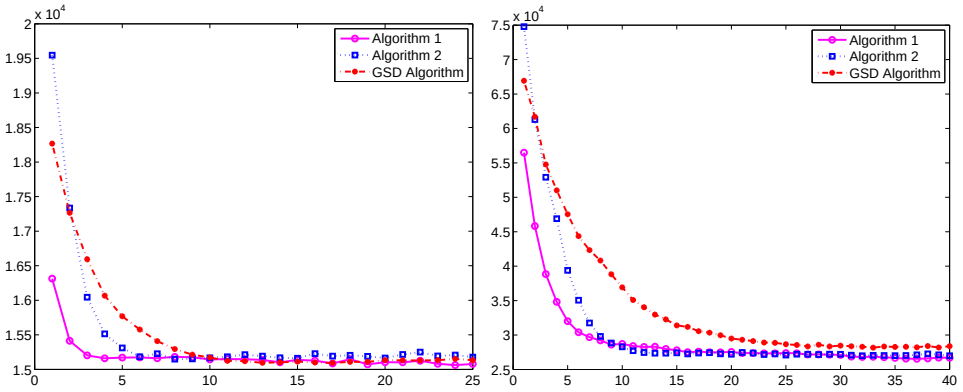
(i) Input image (j) GSD: $n = 124$ (k) Alg. 1: $n = 105$ (l) Alg. 2: $n = 116$

FIGURE 3.1. Comparison of Algorithms 1-2 and the global smooth-dual (GSD) method for two-phase case. Row 1 ((a)-(d)) for **Set 1**; Row 2 ((e)-(h)) for **Set 2**, and Row 3 ((g)-(l)) for **Set 3**. The iteration is stopped by (3.6) with $\eta = 10^{-2}$. The Algorithms 1-2 converge faster, allow to use larger time step size, and are not sensitive to the choice of the parameter β.

with three typical geometric objects and different noise levels. We consider two sets of tests with the following setup:

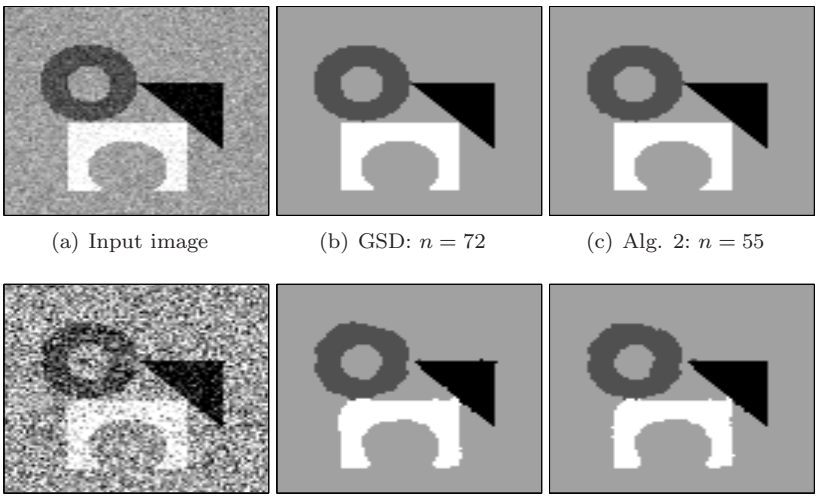
- **Set 4.** Take $\lambda = 0.2, \beta = 10^{-5}, \tau = 0.1, \delta = 0.05$, and $\varepsilon = 0.1$.
- **Set 5.** Take $\lambda = 0.05, \beta = 0.01, \tau = 0.1, \delta = 0.05$ and $\varepsilon = 0.1$.

In Figure 3.3, we plot the segmented edge sets by the two algorithms with stopping rule $\eta = 5 \times 10^{-3}$. Once again, we observe the advantages of Algorithm 2. The comparison of numerical energy decay is depicted in Figure 3.4, and we see a faster decay rate for the proposed algorithm. Indeed, we find that Algorithm 2 is robust for noise (in general, we choose the parameter β bigger if the noise level is high), and works for large time step size as the Chambolle's algorithm. In Table 3.1, we tabulate the number of iterations (to meet the stopping rule (3.6) with $\eta = 5 \times 10^{-3}$), the computational time and numerical energy at the last step of iteration, for two algorithms. In fact, the cost for per iteration of two algorithms is almost the same, but Algorithm 2 allows to use a large time step size, so it saves some computational time.



(a) Decay of energy (**Set 1**) (b) Decay of energy (**Set 2**)

FIGURE 3.2. Comparison of decay of numerical energy for **Set 1** and **Set 2**.



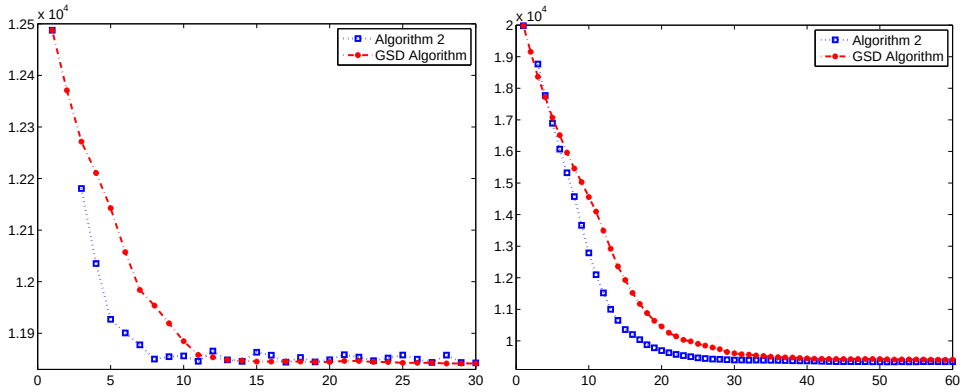
(a) Input image (b) GSD: $n = 72$ (c) Alg. 2: $n = 55$

(d) Input image (e) GSD: $n = 103$ (f) Alg. 2: $n = 75$

FIGURE 3.3. Comparison of Algorithms 2 and the global smooth-dual algorithm (GSD) for four-phase case. Row 1 ((a)-(d)) for **Set 4** and Row 2 ((e)-(h)) for **Set 5**. The iteration is terminated by (3.6) with $\eta = 5 \times 10^{-3}$. Both algorithms work for large noise level. As before, the Algorithm 2 converges faster, and produces slightly better results with larger time step.

TABLE 3.1. Comparison of Algorithm 2 and GSD.

	iterations		time (second)		energy	
	GSD	Alg. 2	GSD	Alg. 2	GSD	Alg. 2
Set 4	72	55	2.3	1.8	1.093×10^4	1.094×10^4
Set 5	103	75	3.6	2.6	7.177×10^3	7.163×10^3



(a) Decay of energy (**Set 4**) (b) Decay of energy (**Set 5**)

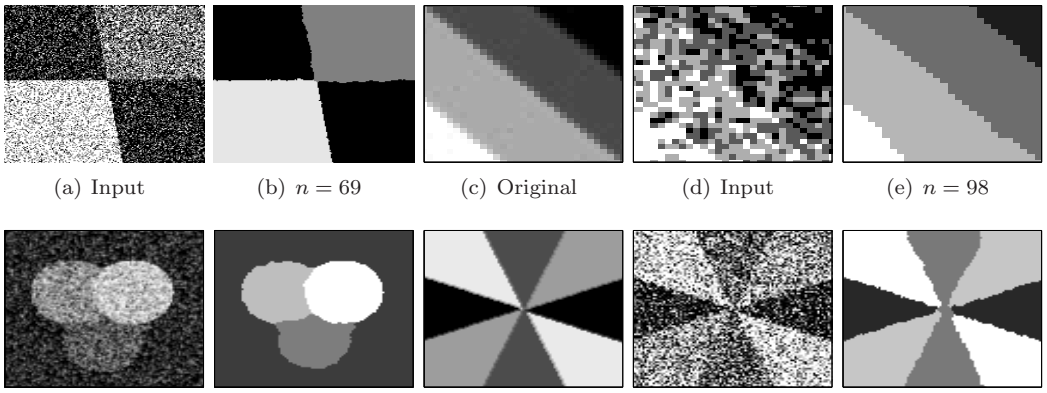
FIGURE 3.4. Comparison of decay of numerical energy for test **Set 4** and **Set 5**.

We next test some more images which are used as examples in most of the papers that we mentioned before. We refer to the caption of Figure 3.5 for the choices of the parameters in Algorithm 2 with stopping rule $\eta = 5 \times 10^{-3}$. In [1] (see Figures 5-8), the comparison with several methods is conducted for the same images to demonstrate that the global smoothed-dual algorithm outperforms the other methods. As with the previous tests, we find that the proposed Algorithm 2 has an even better performance in terms of quality of classification, ease of implementation, computational efficiency and robustness to noise.

3.2.3. *Triple-junction experiments.* Next, we test a very typical example relative to triple-junctions, which was considered by many authors, see, e.g., [36, 24, 11, 8, 1, 19]. The task is to inpaint the edges (or boundaries) of the subregion covered by the disk in Figure 3.6 (a) and (c) of size 290×290, which are expected to generate a triple-junction, that is, the completed three edges suppose to form three 120^0 angles at the junction. The data terms f_i inside the disk are taken to be zero, and $f_i = |c_i - I|$ outside the disk, and the parameters in Algorithm 2 are chosen as $\lambda = 10^{-4}, \beta = 0.5$ and $\tau = 0.1$. In this case, the non-smooth problem $\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i\}$ may not have a unique minimum value (see the second case of Theorem 2.2) in the covered subregion, so in theory the algorithm may fail to find the global optimum. However, by slightly increasing the diffusion effect tuned by the parameter β, we are able to obtain very satisfactory completion of the edges inside the disk, even with large portions of incomplete (or missing) edges (see Figure 3.6 (b) and (d)), namely, achieve the global optimum. This also indicates that the total variation of the labeling functions (or level-set functions) provides a good characterization of the edge set, and the algorithm produces very accurate approximation.

3.2.4. *Applications to image segmentation.* Finally, we apply the methods to the piecewise constant Mumford-Shah model (1.1) where $\{c_i\}$ are unknowns, and can be computed from the mean values:

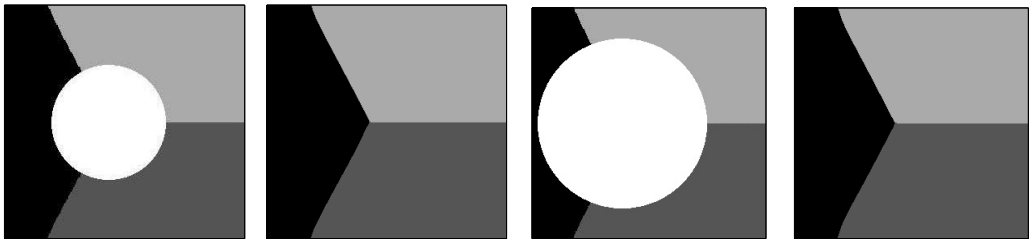
$$c_i = \frac{\int_\Omega I u_i d\boldsymbol{x}}{\int_\Omega u_i d\boldsymbol{x}}, \quad 1 \le i \le m. \tag{3.7}$$



(a) Input (b) $n = 69$ (c) Original (d) Input (e) $n = 98$

(f) Input (g) $n = 89$ (h) Original (i) Input (j) $n = 507$

FIGURE 3.5. More tests on Algorithm 2, where in all tests, $\tau = 0.1$, $f_i = |c_i - I|^2$ with fixed c_i, and other parameters: (a)-(b) (three phases): $\lambda = 10^{-4}, \beta = 0.05$, size: 168×168 ; (c)-(e) (four phases): $\lambda = 10^{-4}, \beta = 10^{-3}$, size: 33×33; (f)-(g) (four phases): $\lambda = 5 \times 10^{-4}, \beta = 10^{-2}$, size: 100×100 and (h)-(j) (four phases): $\lambda = 3 \times 10^{-4}, \beta = 0.05$, size: 100×100. The iteration is terminated by (3.6) with $\eta = 5 \times 10^{-3}$. Algorithm 2 can produce satisfactory results with large time step size. Also the value of β should be relatively larger for the image with larger noise. We also refer to [1] for similar tests on GSD.



(a) Input (b) $n = 300$ (c) Input (d) $n = 600$

FIGURE 3.6. Completion of edges by Algorithm 2. It shows that the proposed algorithm can find the global optimum.

For fixed c_i, the model (1.1) is special case of (1.2) with $f_i = |c_i - I|^2$. Therefore, with a slight modification of Algorithm 2, we are able to obtain the following algorithm for the multiphase piecewise constant image segmentation.

Algorithm 3

1. Initialization: set $\vec{\boldsymbol{p}}^0 = \mathbf{0}$ and choose $\beta, \tau > 0$, and $\vec{c}^0$;
2. For $n = 0, 1, \cdots$
   (i) Compute λ_2^n by (2.25) with $f_i^n = |c_i^n - I|^2$ in place of f_i;
   (ii) Compute $\vec{u}^n$ by (2.26) with $h_i = \mathrm{div}\boldsymbol{p}_i^n + \lambda f_i^n$ and $\lambda_2 = \lambda_2^n$;

(iii) Compute $\vec{\boldsymbol{p}}^{n+1}$ by

$$\boldsymbol{p}_i^{n+1} = \frac{\boldsymbol{p}_i^n - \tau \nabla u_i^n}{1 + \tau |\nabla u_i^n|}, \quad 1 \le i \le m;$$

(iv) Compute $\vec{c}^{n+1}$ by

$$c_i^{n+1} = \frac{\int_\Omega I u_i^n d\boldsymbol{x}}{\int_\Omega u_i^n d\boldsymbol{x}}, \quad 1 \le i \le m.$$

3. Endfor till some stopping rule meets;
4. Set $\vec{\boldsymbol{p}}^* = \vec{\boldsymbol{p}}^{n+1}, \vec{c}^* = \vec{c}^{n+1}$ and define

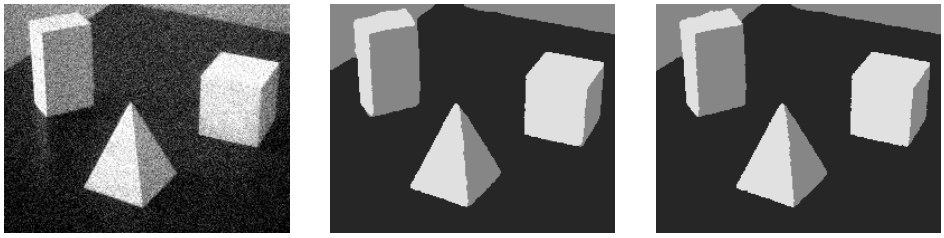
$$u_k^* = \begin{cases} 1, & k = \min\left\{\arg\min_{1\le i\le m}\{\mathrm{div}\boldsymbol{p}_i^* + \lambda f_i^*\}\right\}, \\ 0, & \text{otherwise}, \end{cases}$$

for all $1 \le k \le m$, where $f_i^* = |c_i^* - I|^2$.

Similarly, the global smoothed-dual algorithm can be modified for the segmentation problem (see Algorithm 2 in [1]), and for clarity, we term it as GSD-Seg in short.

We first compare Algorithm 3 with GSD-Seg for segmenting the "desk" image in Figure 3.7 (a) of size 205×255 containing three typical geometric objects with three phases. For this example, it is not an easy task to segment the faces of the objects, for instance, the two faces of the tetrahedron. In the test, both algorithms are terminated by (3.6) with $\eta = 5 \times 10^{-3}$ (refer to the caption of Figure 3.7 for the selection of parameters and time step size). We observe similar performances as before. Both of them produce quite satisfactory and accurate segmentation, while Algorithm 3 is relatively faster. It is faster than most of the approaches in e.g., [32, 17, 16, 28, 3, 11].

We also tabulate in Table 3.2, the number of iterations (to meet the stopping rule (3.6) with $\eta = 5\times10^{-3}$), the computational time and numerical energy at the last step of iteration, for two algorithms. Once again, the cost for per iteration of two algorithms is almost the same, but Algorithm 3 allows to use a large time step size.



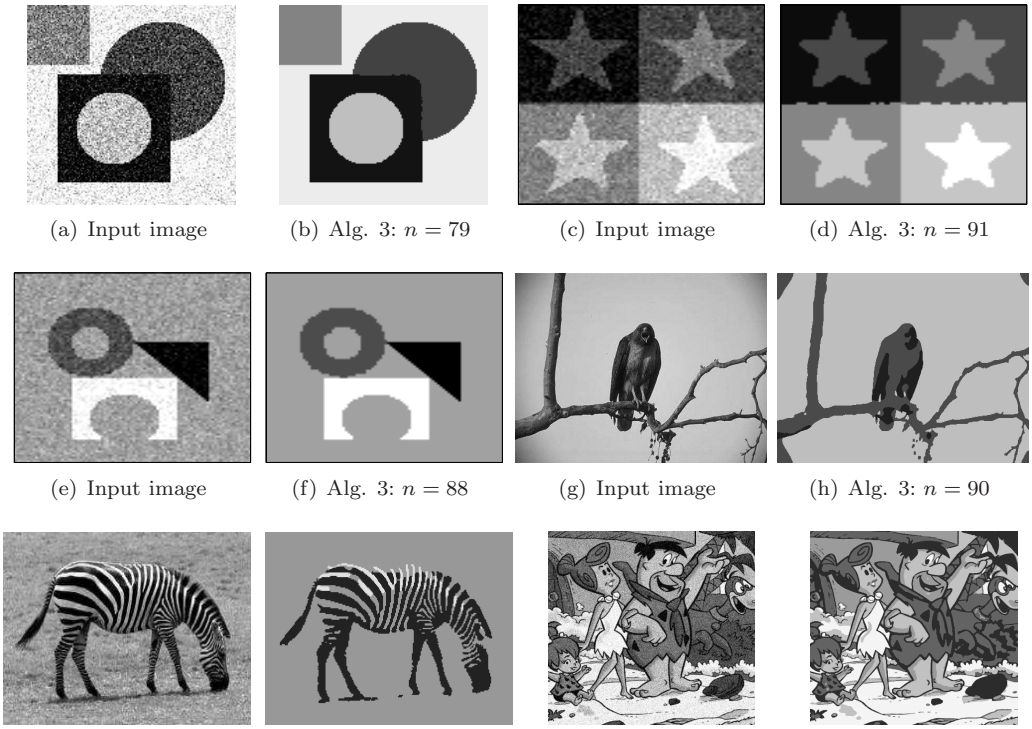
(a) Input image (b) GSD-Seg: $n = 115$ (c) Alg. 3: $n = 87$

FIGURE 3.7. Comparison of Algorithm 3 and GSD-Seg, where $\tau = 0.1$, $\delta = 0.02$, $f_i = |c_i - I|^2$ and other parameters: (a)-(c) (three phases of size 205×255): $\lambda = 5 \times 10^{-4}, \beta = 10^{-2}, \varepsilon = 0.1$. Given this noise input image, it is challenging to segment the faces of the objects. The iteration of both algorithms is terminated by (3.6) with $\eta = 5 \times 10^{-3}$. Algorithm 3 slightly outperforms GSD-Seg.

TABLE 3.2. Comparison of Algorithm 3 and GSD-Seg.

iterations		time (second)		energy	
GSD-Seg	Alg. 3	GSD-Seg	Alg. 3	GSD-Seg	Alg. 3
115	87	15	11	1.320×10^4	1.314×10^4

We next test some more images which are used as examples in most of the papers that we mentioned before. We refer to the caption of Figure 3.8 for the choice of the parameters in Algorithm 3. In all cases, the algorithm is terminated by (3.6) with $\eta = 5 \times 10^{-3}$. We see that in all tests, Algorithm 3 provides quite satisfactory segmentation and converges fast.



(a) Input image (b) Alg. 3: $n = 79$ (c) Input image (d) Alg. 3: $n = 91$

(e) Input image (f) Alg. 3: $n = 88$ (g) Input image (h) Alg. 3: $n = 90$

(i) Input image (j) Alg. 3: $n = 76$ (k) Input image (l) Alg. 3: $n = 89$

FIGURE 3.8. More tests on Algorithm 3, where $\tau = 0.1$, $f_i = |c_i - I|^2$ and other parameters: (a)-(b) (five phases): $\lambda = 10^{-4}, \beta = 10^{-3}$, size: 150×150; (c)-(d) (five phases): $\lambda = 10^{-3}, \beta = 10^{-2}$, size: 90×90; (e)-(f) (four phases): $\lambda = 10^{-3}, \beta = 10^{-5}$, size: 90×90; (g)-(h) (three phases): $\lambda = 10^{-3}, \beta = 10^{-2}$, size 321×481; (i)-(j) (three phases): $\lambda = 10^{-3}, \beta = 10^{-2}$, size: 167×250 and (k)-(l) (five phases): $\lambda = 10^{-3}, \beta = 10^{-2}$, size: 512×512. The iteration is terminated by (3.6) with $\eta = 5 \times 10^{-3}$. Algorithm 3 can produce satisfactory results with large time step size.

4. Concluding remarks

In this paper, we proposed a direct primal-dual approach towards a global minimization of the continuous Potts model for multiclass labeling problems with applications to multiphase image segmentation. Different from the existing works, the underlying analysis and algorithms were based on a binary setting which did not require the convex relaxation. Using the augmented Lagrangian technique, we were able to find the relations of the primal and dual variables that involve almost the minimum number of parameters. The proposed algorithms could be viewed as the counterparts of the Chambolle's algorithm in the context of labeling and segmentation. Indeed, the time step size could be taken as large as that in the Chambolle's algorithm, and the computational cost turned out to be almost of the same amount. Moreover, the second smoothing parameter involved in the algorithm was quite loose to choose. Various numerical results demonstrated the advantages of the methods over the existing approaches.

References

[1] E. Bae, J. Yuan, and X.C. Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[4] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.

[5] E.S. Brown, T.F. Chan, and X. Bresson. Convex formulation and exact global solutions for multi-phase piecewise constant Mumford-Shah image segmentation. *UCLA, Applied Mathematics, CAM-report-09-66*, July, 2009.

[6] E.S. Brown, T.F. Chan, and X. Bresson. A convex relaxation method for a class of vector-valued minimization problems with applications to mumford-shah image segmentation. *UCLA, Applied Mathematics, CAM-report-10-43*, July, 2010.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.

[8] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[9] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[10] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms: Fundamentals*. Springer, 1993.

[11] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrainted total variation. *Technical report, HCI*, 2008.

[12] J. Lellmann and C. Schnörr. Continuous multiclass labeling approaches and algorithms. *Tech. rep., Univ. of Heidelberg*, 2010.

[13] C. Li, C. Xu, C. Gui, and M.D. Fox. Level set evolution without re-initialization: A new variational formulation. *In IEEE Conference on Computer Vision and Pattern Recognition (CPVR)*, 2005.

[14] F. Li and M.K. Ng. Kernel density estimation based multiphase fuzzy region competition method for texture image segmentation. *Communications in Computational Physics*, 8(3):623–641, 2010.

[15] F. Li, M.K. Ng, T.Y. Zeng, and L.C. Shen. A multiphase image segmentation method based on fuzzy region competition. *SIAM Journal on Imaging Sciences*, 3(3):277–299, 2010.

[16] H.W. Li and X.C. Tai. Piecewise constant level set methods for interface problems. *International Series of Numerical Mathematics*, 154:307–316, 2007.

[17] J. Lie, M. Lysaker, and X.C. Tai. A variant of the level set method and applications to image segmentation. *Mathematics of Computation*, 75(255):1155–1174, 2006.

[18] C.X. Liu, F.F. Dong, S.F. Zhu, D.K. Kong, and K.F. Liu. New variational formulations for level set evolution without reinitialization with applications to image segmentation. *Journal of Mathematical Imaging and Vision*, available online since March 2011, DOI: 10.1007/s10851-011-0269-2.

[19] J. Liu, X.C. Tai, H.Y. Huang, and Z.D. Huan. A fast segmentation method based on constraint optimization and its applications: Intensity inhomogeneity and texture segmentation. *Pattern Recognition*, available online since March 2011, DOI: 10.1016/j.patcog.2011.02.022.

[20] B. Merriman, J.K. Bence, and S. Osher. Motion of multiple functions: a level set approach. *Journal of Computational Physics*, 112(2):334–363, 1994.

[21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[22] M. Nikolova, T.F. Chan, and S. Esedoglu. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[23] S. Osher and R.P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

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

[25] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 3(4):1122–1145, 2010.

[26] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. *In European Conference on Computer Vision (ECCV)*, 2008.

[27] R.B. Potts. Some generalized order-disorder transformations. *In Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

[28] Z.J. Rong, L.L. Wang, and X.C. Tai. Adaptive wavelet collocation methods for image segmentation using TV-Allen-Cahn type models. *Accepted by Advances in Computational Mathematics*, 2011.

[29] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[30] X.C. Tai, O. Christiansen, P. Lin, and I. Skjælaaen. Image segmentation using some piecewise constant level set methods with mbo type of projection. *International Journal of Computer Vision*, 73(1):61–76, 2007.

[31] X.C. Tai and C. Wu. Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. In *Scale Space and Variational Methods in Computer Vision-LNCS 5567*, pages 502–514. Springer, Berlin, 2009.

[32] L.A. Vese and T.F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[33] L.L. Wang and Y. Gu. Efficient dual algortihms for image segmentation using TV-Allen-Cahn type models. *Communications in Computational Physics*, 9(4):859–877, 2011.

[34] Y.L Wang, J.F. Yang, W.T. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[35] C. Wu and X.C. Tai. Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM Journal on Imaging Sciences*, 3:300–339, 2010.

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