A FAST CONTINUOUS MAX-FLOW APPROACH TO NON-CONVEX MULTILABELING PROBLEMS

EGIL BAE*, JING YUAN†, XUE-CHENG TAI‡, AND YURI BOYKOV§

Abstract. This paper deals with two optimization models for image labeling over a spatially continuous image domain, where the first one favors a linear ordering of the labels in the computation result and the second does not favor any particular ordering (Pott's model). We study convex reformulations and relaxations of these two non-convex labeling problems. Inspired by Ishikawa's multi-layered graph construction [34] for the same labeling problem over a discrete image domain, we propose novel continuous max-flow models and build up their dualities to the convex relaxed formulations of image labeling under a new variational perspective. Via such continuous max-flow formulations, we show that exact and global optimizers can be obtained to the original problem with linearly ordered labels. We also extend our studies to problems with continuous-valued labels and introduce a new theory to this problem. Finally, we show the proposed continuous max-flow models directly lead to new fast flow-maximization algorithmic schemes which significantly outperform previous approaches [53, 52] in terms of efficiency. Such continuous max-flow based algorithms can be validated by modern convex optimization theories and accelerated by modern parallel computational hardware.

Key words. image processing and segmentation, global optimization, convex relaxation, variational problems

AMS subject classifications. 49M20,49M29,65K10,68U10

1. Introduction. Many applications of image processing and computer vision can be mathematically modeled and solved by means of energy minimization. In this paper, we consider the following important low-level image labeling problem: each image pixel is assigned by one discrete label subject to a given optimality criterion. The energy function can be represented over a sequence of partitions or labeling functions, which directly leads to a non-convex optimization problem. Such image labeling problems potentially models many applications of image processing (see [49] for a good reference): image denoising [41, 57]; image segmentation [11, 2, 12]; image stereo reconstruction [40, 41]; multi-view reconstruction [45].

The optimization criterion for image labeling can be defined and formulated over either a spatially continuous image domain or a discrete image graph, which boils down to either a continuous non-convex optimization problem or a combinatorial optimization problem. In the spatially discrete setting, the optimization formulation can be constructed by the principle of Markov random field (MRF) over the discrete image graph. In this regard, many effective solvers have been proposed, e.g. graph-cuts [31, 9, 34], message passing [58, 39] and linear programming [42] etc., where graph-cut is one of the most efficient ways to exactly tackle the problem in case of a sub-modular energy function. In practice, most labeling problems involving more than two labels are NP-hard, therefore only approximate graph-based algorithms are available, such as [11]. However, for multi-labeling problems where the labels are linearly ordered and interaction penalties are convex, Ishikawa [34] showed that the exact solution can be obtained by computing the min-cut over a specially constructed graph with multiple layers.

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

†Jing Yuan, Computer Science Department, Middlesex College, University of Western Ontario, London Ontario, Canada N6A 5B7 (cn.yuanjing@gmail.com)

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

§Yuri Boykov, Computer Science Department, Middlesex College, University of Western Ontario, London Ontario, Canada N6A 5B7 (yuri@csd.uwo.ca)

Despite the efficiency of most graph-based approaches, their computation results are biased by the discrete grid, i.e. visible metrication errors [10, 38] are introduced in the results. Reducing such artifacts requires either considering more neighbor nodes [10, 38] or applying high-order potentials [37, 35]. However, this results in a high memory load of computation.

In this work, we focus on image labeling over a spatially continuous image domain and study the resulting continuous non-convex optimization problems. In contrast to the graph-based approaches, such continuous approaches properly avoid metrication errors and allows for computational result with high sub-grid accuracy. To this end, both the level set method [48, 16] and phase-field method [36, 6] were proposed. Another important approach is the piecewise-constant level set method proposed in [46, 47], which assigns a discrete label to each pixel of the image domain Ω by means of forcing some integral equality conditions. However, all these methods are based on minimizing a non-convex energy functional; therefore, only local optimizers can be obtained and the computation results highly depend on the initialization. In this paper, we show that the non-convex optimization problems of image labeling can be globally and efficiently solved by means of convex relaxation. In contrast to previous methods, the convex relaxation approach can yield globally or nearly globally optimal solutions to the original non-convex optimization problem. Furthermore, fast algorithms can be derived through modern convex optimization theories and their implementation require less memory and can be easily accelerated by modern parallel computational hardware.

1.1. A Short Review. By the milestone works of Strang [56] and Chan et al [17], it was realized that typical binary image labeling problems in the spatially continuous setting, in case of two labels, can be globally and exactly solved via a convex relaxation. Particularly, Chan et al. [17] considered the following optimization problem

$$\min_{u(x)\in\{0,1\}\forall x\in\Omega} \int_\Omega (1-u)\rho(\ell_2,x)\,dx + \int_\Omega u\rho(\ell_1,x)\,dx + \alpha\int_\Omega |\nabla u|\,dx \tag{1.1}$$

which is clearly non-convex due to its binary constraint $u(x)\in\{0,1\}$ for all $x\in\Omega$. Here $\rho(\ell_i,x)$ is the cost of assigning x to region Ω_i. The authors showed that such a binary constraint can be relaxed by $u(x)\in[0,1]$, then the computed result of the convex relaxation could be thresholded at any level in $(0,1]$ to yield a global minimum of the original binary constrained problem (1.1). Recently, such a convex relaxation approach has been further extended to the multiregion case as the continuous counterpart of Pott's model [54]. Pott's model describes optimal partition of the image domain Ω into n disjoint sub-regions $\{\Omega_i\}_{i=1}^n$ with minimal total perimeter as the solution of

$$\begin{aligned} \min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} \rho(\ell_i,x)\,dx \;+\; \alpha\sum_{i=1}^n |\partial\Omega_i| \\ \text{s.t.}\quad \cup_{i=1}^n \Omega_i \;=\; \Omega\,,\quad \Omega_k\cap\Omega_l \;=\; \emptyset\,,\quad \forall k\neq l\,. \end{aligned} \tag{1.2}$$

Convex relaxations for (1.2) was proposed and studied in [61, 44, 50, 13, 5]. Because the underlying optimization problem (1.2) is NP-hard, the relaxations are not generally exact, i.e. the reconstructed rounded integer-valued optimum can generally only be accepted as suboptimal. However, experimental results were promising in terms of the total energy and quality of the computed solutions. The tight relaxation for Potts model proposed in [50, 13] gives the best approximations to global minimums of the original Potts problem (1.2).

Another important image labeling problem is

$$\min_{u(x)\in\{\ell_1\ldots\ell_n\}} \int_\Omega \rho(u(x),x)\,dx + \int_\Omega C(x)|\nabla u(x)|\,dx\,, \tag{1.3}$$

where the n labels $\{\ell_1 \dots \ell_n\}$ are linearly ordered such that $\ell_1 < \dots < \ell_n$ and $\rho(u(x), x)$ is some bounded function, not necessarily convex in u. The last term of (1.3) regularizes the total (weighted) perimeter of the labeled partitions. The problem (1.3) can also express partitioning problems, by the convention $u = \ell_i$ in the region Ω_i. The piecewise constant level set method [47] has exactly the form of (1.3) with $\ell_i = i$ for $i = 1, ...n$. Note that the regularization term of (1.3) does not correspond to the length term in the more ideal Pott's model (1.2), because of its dependency on the size of the jumps of u. On the other hand, such a linear relationship on the size of the jump of u may be an advantage in other applications, like stereo reconstruction and image denoising where label values should potentially favor such a linear order.

To approach a continuous version of (1.3), where also the label values are constrained to a continuous set, [53, 52] generalized Ishikawa's work [34] to the spatially continuous setting, by representing the optimal labeling function as the discontinuity set of a binary function in a one-dimensional higher space, i.e. a spatially continuous min-cut. Such a lifting approach is related to earlier mathematical theories of calibrations and cartesian currents [8, 1]. Optimal labeling functions could be obtained by applying the result of Chan et al. in the higher dimensional space, i.e. first solve the relaxed binary problem and then threshold the result. Recently, the lifting approach was further applied to solve vector-valued problems [29] in the totally discrete setting.

1.2. Motivations and contributions. For discrete graphs, it is well known that the minimum cut problem is dual to the maximum flow problem by the *max-flow and min-cut theorem* [24]. Actually, the fastest graph cut algorithms are based on maximizing flow instead of computing the min-cut directly, e.g. the Ford-Fulkerson algorithm [23] and the push-relabel algorithm [28]. The minimal 'cut' is finally recovered along edges with 'saturated' flows, i.e. cuts appear at the flow-bottlenecked edges [18, 41]. In contrast, max-flow models and algorithms in the spatially continuous setting have been much less studied. Some work has appeared that deal with partitioning problems involving two regions: Strang [56] was the first to formulate max-flow and min-cut problems over a continuous domain; In [3], edge based max-flow and min-cut was formulated in which certain interior and exterior points must be specified in advance; Yuan et al [59, 60] proposed a direct continuous analogue of the typical discrete max-flow and min-cut models that are used for solving binary labeling problems in image processing and computer vision. In contrast, most previous works on labeling in the spatially continuous setting, e.g. [61, 29, 53, 12] etc, tried to conduct the energy minimization over the labeling functions directly.

To our knowledge, this is the first work to address continuous max-flow models for partitioning problems involving multiple regions. Motivated by Yuan et al. [59] and Ishikawa [34], we interpret (1.3) as a continuous min-cut problem over a mixed continuous/discrete domain and build up a novel continuous max-flow model in analogy with Ishikawa's discrete graph construction. The max-flow model can be used to produce global solutions of the non-convex problem (1.3) with discrete label values. In particular, it is shown that the max-flow model is dual to an exact convex relaxation of (1.3). Strict duality is also established between the max-flow model and the original problem, by extending the thresholding scheme of [17] from two to multiple regions. With aid of the proposed dualities, a new efficient continuous max-flow based algorithm is also derived.

The theory of the continuous max-flow approach for (1.3) is extended to two other problems of importance: a tight convex relaxation to Pott's model, and labeling with continuous-valued labels. We show significant advantages of the proposed continuous max-flow approach over previous work in terms of both theoretical elegance and efficiency in numerics.

Our main contributions can be summarized as follows

- We study a convex relaxation of the nonconvex labeling problem (1.3), the so-called *continuous min-cut model*. To this end, we build up a novel max-flow formulation over n linearly layered continuous image domains, which is in analogy with the discrete graph construction of Ishikawa. Duality between the proposed continuous max-flow model and its corresponding continuous min-cut model is shown upon a variational perspective.
- A thresholding scheme is derived for converting solutions of the convex relaxed problem into solutions of the non-convex problem (1.3) with discrete label values, extending the scheme proposed in [17] from two to multiple regions.
- New continuous max-flow based algorithms are proposed. Their efficiency and convergence can be verified by standard convex optimization theories. The labeling function is updated as an unconstrained lagrange multiplier each iteration, and does not need to be projected back onto any feasible set. Numerical experiments show a significantly faster convergence rate than the primal-dual algorithm in Pock et. al. [52, 53], especially at high precisions.
- A max-flow dual formulation of the convex relaxation of Pott's model [50] is proposed as a direct extension of the continuous max-flow model for (1.3). An algorithm is proposed which deals with all constraints on the labeling function implicitly and avoids expensive iterative computations of projections without closed form solution.

2. Preliminaries: Ishikawa's Work. Ishikawa [34] studied image labeling problems over an image graph which can be generally formulated as:

$$\min_{u\in U} \sum_{v\in\mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w)\in\mathcal{N}} g(u_v - u_w)\,, \tag{2.1}$$

where $\mathcal{P}$ denotes a discrete image grid in 2-D or N-D; $\mathcal{N} \subset \mathcal{P}\times\mathcal{P}$ is a neighborhood system on $\mathcal{P}$; $U = \{u:\ \mathcal{P}\mapsto L\}$ is the set of all feasible labeling functions. The potential prior $g(x)$ of (2.1) is assumed to be convex and ρ is any bounded function, but not necessarily convex. It was shown by [34] that the problems of the form (2.1) can be exactly optimized by finding the min-cut over a specially constructed multi-layered graph $G = (\mathcal{V}, \mathcal{E})$, where each layer corresponds to one label.

We adopt Ishikawa's notations [34] in this work and study the simplified graph which uses $n-1$ layers instead of n and avoids infinite capacities on the source edges [4] (see Fig. 2.1 for a 1-D example). The vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ are defined as follows:

$$\mathcal{V} \;=\; \mathcal{P}\times L \cup \{s,t\} \;=\; \{u_{v,i}\,|\,v\in\mathcal{P}\,;\ i=1,...,n-1\}\cup\{s,t\} \tag{2.2a}$$

$$\mathcal{E} \;=\; \mathcal{E}_D \cup \mathcal{E}_C \cup \mathcal{E}_P \tag{2.2b}$$

where the edge set $\mathcal{E}$ is composed of three types of edges

- Data edges $\mathcal{E}_D = \bigcup_{v\in\mathcal{P}} \mathcal{E}_D^v$, where

$$\mathcal{E}_D^v \;=\; (s, u_{v,1}) \cup \{(u_{v,i}, u_{v,i+1})\,|\,i=1,\ldots,n-2\} \cup (u_{v,n-1}, t)\,. \tag{2.3}$$

- Penalty edges $\mathcal{E}_P = \bigcup_{v\in\mathcal{P}} \mathcal{E}_C^v$, where

$$\mathcal{E}_C^v \;=\; \{(u_{v,i+1}, u_{v,i})\,|\,i=1,\ldots,n-2\}\,. \tag{2.4}$$

- Regularization edges $\mathcal{E}_R$:

$$\mathcal{E}_R \;=\; \{(u_{v,i}, u_{w,j})\,|\,(v,w)\in\mathcal{N}\,,\ i,j=1,...,n\}\,. \tag{2.5}$$

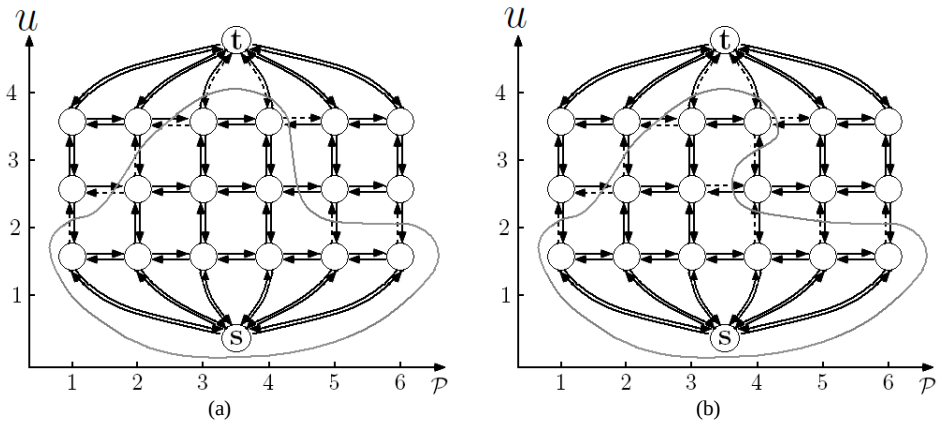


FIG. 2.1. *1D illustration: (a) Legal cut, (b) Illegal cut. Severed edges are depicted as dotted edges. The gray curve visualizes the cut. Vertices interior to the curve belongs to V_s while vertices exterior to the curve belongs to V_t. Severed edges are illustrated as dotted arrows.*

2.1. Anisotropic Total-Variation Regularization. When a pairwise prior $g(u_v - u_w) = C(u,w)\,|u_v - u_w|$ is given, (2.1) corresponds to an anisotropic total-variation regularized image labeling problem, i.e.

$$\min_{u\in U} \sum_{v\in\mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w)\in\mathcal{N}} C(v,w)\,|u_v - u_w| \tag{2.6}$$

which is the discrete counterpart of the total-variation based mutli-labeling problem (1.3).

Now we define flow configurations over the graph (2.2a) and (2.2b) such that its max-flow corresponds to the minimizer of (2.6):

- *Capacity of source flows:* the directed flow $p_1(v)$ along each edge from the source s to the node $u_{v,1}$ of the first layer, i.e. the edge $(s, u_{v,1})$, is constrained by

$$p_1(v) \le \rho(\ell_1, v)\,, \quad \forall v \in \mathcal{P}\,; \tag{2.7}$$

- *Capacity of flows between layers:* the directed flow $p_i(v)$ along each edge $(u_{v,i}, u_{v,i+1})$ from the node $u_{v,i}$ of the i-th layer to the node $u_{v,i+1}$ of the $i+1$-th layer is constrained by

$$p_i(v) \le \rho(\ell_i, v)\,, \quad \forall v \in \mathcal{P} \quad i = 1, ..., n-2 \tag{2.8}$$

- *Capacity of sink flows:* the directed flow $p_n(v)$ along each edge from the node $u_{v,n-1}$ of the last layer to the sink t is constrained by

$$p_n(v) \le \rho(\ell_n, v)\,, \quad \forall v \in \mathcal{P}\,; \tag{2.9}$$

- *Capacity of spatial flows at each layer:* the undirected flow $q_i(v,w)$ of each edge $(v,w) \in \mathcal{N}$ at the layer i, $i = 1, \ldots, n-1$, is constrained by

$$|q_i(v,w)| \le C(v,w)\,; \tag{2.10}$$

this actually amounts to the well-known anisotropic total-variation regularizer;

- *Conservation of flows:* flow conservation means that in-coming flows should be balanced by out-going flows at any node $v \in \mathcal{P}$ of each layer $i = 1, ..., n-1$, i.e.

$$\Big(\sum_{w:(w,v)\in\mathcal{N}} q_i(v,w) - \sum_{w:(v,w)\in\mathcal{N}} q_i(v,w) \Big) - p_i(v) + p_{i+1}(v) \;=\; 0\,. \qquad (2.11)$$

Since there is no lower bound on the flows (2.7)-(2.9), the flow capacities on the penalty edges (2.4) are infinite. This implies that each edge in the set $\mathcal{E}_D^v$ which links the source and sink can only be cut once, i.e. illegal cuts as shown in Fig. 2.1(b) have infinite cost and are not allowed.

Therefore, the max-flow problem over the graph is to find the largest amount of flow allowed to pass from the source s to sink t through the $n-1$ graph layers, i.e.

$$\max_{p,q} \sum_{v\in\mathcal{P}} p_1(v) \qquad (2.12)$$

subject to the flow constraints (2.7), (2.8), (2.9), (2.10) and (2.11).

It was proved that once the maximal flow is computed, a minimal cut can be extracted which corresponds to a minimizer of the problem (2.6).

3. Convex Relaxation and Continuous Max-Flow Models. In this section, we study the labeling problem (1.3) which is the continuous counterpart of (2.1) specialized to the classical total-variation regularizer:

$$\min_{u\in U} \int_\Omega \rho(u(x), x)\, dx + \int_\Omega C(x)|\nabla u(x)|\, dx\,, \qquad (3.1)$$

where $U = \{u :\ \Omega \mapsto \{\ell_1, ..., \ell_n\},\ \text{ s.t. } \int_\Omega |\nabla u|\, dx < \infty\}$ is the set of all feasible functions over the continuous image domain Ω; $\rho(u(x), x)$ is any uniformly bounded function, not necessarily convex in the element of u. The gradient magnitude is measured with the rotationally invariant 2-norm $|\nabla u|_2 = \sqrt{u_{x_1}^2 + ... + u_{x_m}^2}$, in contrast to the anisotropic graph representable 1-norm used in the discrete setting.

Inspired by Ishikawa's graph-cut work revisited in the last section, we propose a similar flow-maximization scheme in the spatially continuous setting, and build up the duality between such continuous ma ax-flow model and a convex relaxation of (3.1). Via the new max-flow model, we show the proposed convex relaxation model solves (3.1) exactly and globally.

3.1. Representations by Layer Functions. Let S_i, $i = 1, ..., n-1$, denote the $n-1$ upper level sets of the labeling function $u(x) \in U$ such that

$$S_i = \{x \in \Omega :\ u(x) > \ell_i\}\,. \qquad (3.2)$$

To ease exposition, we also define $S_0 = \Omega$ and $S_n = \emptyset$.

The characteristic functions $\lambda_i(x)$ of the upper level sets S_i $i = 1, ..., n-1$, also called *the layer functions* in this work, are defined by:

$$\lambda_i(x) \;=\; \begin{cases} 1 & \text{if } u(x) > \ell_i \\ 0 & \text{if } u(x) \le \ell_i \end{cases}\,, \quad i = 1, \ldots, n-1\,. \qquad (3.3)$$

Likewise, we define $\lambda_0(x) = 1$ and $\lambda_n(x) = 0$, $\forall x \in \Omega$, as the characteristic functions of the set S_0 and S_n respectively. We show how (3.1) can be expressed in terms of λ, as was done in the discrete setting in [19, 19, 15, 20].

As $\ell_1 < \ldots < \ell_n$, we have

$$\emptyset \,=\, S_n \,\subseteq\, \ldots \,\subseteq\, S_1 \,\subseteq\, S_0 \,=\, \Omega \tag{3.4}$$

and

$$0 \,=\, \lambda_n \,\leq\, \ldots \,\leq\, \lambda_1 \,\leq\, \lambda_0 \,=\, 1\,. \tag{3.5}$$

With help of the above notations, we can rewrite the optimization problem (3.1) in terms of the layer functions $\lambda_i(x)$, $i = 1, ..., n-1$, such that

$$\min_{\lambda_i(x)\in\{0,1\}} \sum_{i=1}^{n} \int_\Omega (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx \,+\, \sum_{i=1}^{n-1} \int_\Omega C_i(x) |\nabla \lambda_i(x)|\, dx \tag{3.6}$$

subject to the monotonically nonincreasing constraint (3.5), where $C_i(x) = (\ell_{i+1} - \ell_i)C(x)$, $i = 1, \ldots, n-1$.

To see this, the data term of (3.1) can be directly written as

$$\int_\Omega \rho(u(x), x)\, dx = \sum_{i=0}^{n} \int_{S_{i-1}\setminus S_i} \rho(\ell_i, x)\, dx = \sum_{i=0}^{n} \int_\Omega (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx\,. \tag{3.7}$$

Moreover, observe that any function $u(x) \in U$ can be written in terms of $\lambda_i(x)$, $i = 0, ..., n$, as

$$u(x) \,=\, \sum_{i=1}^{n} (\lambda_{i-1}(x) - \lambda_i(x))\ell_i \,=\, \ell_1 + \sum_{i=1}^{n-1} \lambda_i(x)(\ell_{i+1} - \ell_i)\,. \tag{3.8}$$

By the coarea formula [27], the regularization term of (3.1) amounts to

$$\int_\Omega C(x) |\nabla u(x)|\, dx \,=\, \sum_{i=1}^{n-1} \int_\Omega C_i(x) |\nabla \lambda_i(x)|\, dx\,. \tag{3.9}$$

Clearly, once the layer functions $\lambda_i(x)$ are computed, the labeling function $u(x)$ can be easily recovered by (3.8).

In this work, we focus on the case where $C(x) = \alpha$ is constant and $\ell_{i+1} - \ell_i = 1$, $i = 1, ..., n-1$ for simplicity. The results can be easily extended to other more general $C(x)$ and ℓ_i, $i = 1, ..., n$. Using the above results, (3.6) can equivalently be reformulated as

$$\min_{\lambda_i(x)\in\{0,1\}} \sum_{i=1}^{n} \int_\Omega (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx \,+\, \alpha \sum_{i=1}^{n-1} \int_\Omega |\nabla \lambda_i|\, dx \tag{3.10}$$

subject to the order constraint (3.5). (3.10) is nonconvex due to the binary setting of $\lambda_i(x) \in \{0, 1\}$, $i = 1, ..., n-1$.

3.2. Convex Relaxation Models. In the following parts, we show that the nonconvex optimization problem (3.10) can be globally and exactly solved via its convex relaxation:

$$\min_{\lambda_i(x)\in[0,1]} E^D(\lambda) \,=\, \sum_{i=1}^{n} \int_\Omega (\lambda_{i-1}(x) - \lambda_i(x))\rho(\ell_i, x)\, dx \,+\, \alpha \sum_{i=1}^{n-1} \int_\Omega |\nabla \lambda_i|\, dx \tag{3.11}$$

$$\text{s.t.} \quad 1 \,=\, \lambda_0(x) \,\geq\, \lambda_1(x) \,\geq\, \ldots \,\geq\, \lambda_{n-1}(x) \,\geq\, \lambda_n(x) \,=\, 0\,, \quad \forall x \in \Omega$$

where the binary constraints on the labeling functions $\lambda_i(x) \in \{0, 1\}$, $i = 1, \ldots, n-1$, are relaxed by the convex ones $\lambda_i(x) \in [0, 1]$, $i = 1, \ldots, n-1$. In this work, (3.11) is also called the *primal model* in comparison to its *dual formulation: the continuous max-flow model* proposed in the following section.

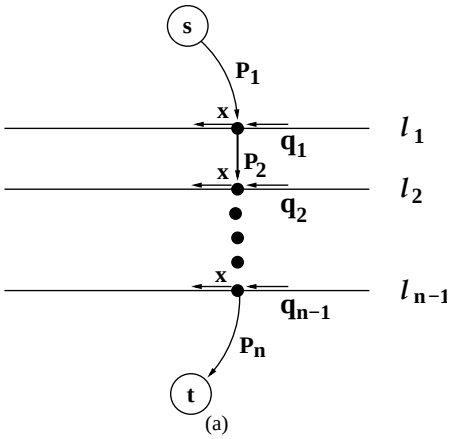


FIG. 3.1. *(a) Illustration of the max-flow problem defined over a mixed discrete/continuous domain.*

3.2.1. Dual Model: Continuous Max-Flow Formulation. Inspired by Ishikawa's graph configuration (2.2a) and (2.2b) reviewed in Sec. 2, we set up our spatially continuous settings in the same manner: $n-1$ copies of the image domain Ω are placed in sequential order between two terminals: the source s and the sink t (see Fig. 3.1); this mixed continuous/discrete setting can be defined as

$$\Omega \times \{1, ..., n-1\} \cup \{s\} \cup \{t\} = \{(x, i) \,|\, x \in \Omega,\ i = 1, ..., n-1\} \cup \{s\} \cup \{t\}\,. \quad (3.12)$$

Likewise, the continuous counterparts of edges, flows and capacities are given as follows (see Fig. 3.1 for an illustration):

- In view of (2.3), the data edges are defined as follows: for each $x \in \Omega$, the source s is linked to $(x, 1)$ of the first layer by the edge function $e_1(x)$; the points $(x, i-1)$ and (x, i) in two sequential image layers, $i = 2 \ldots n-1$, are linked by the edge function $e_i(x)$; at the last layer $(x, n-1)$ is linked to the sink t by the edge function $e_n(x)$.

- At each edge $e_i(x)$, $i = 1 \ldots n$, a flow function $p_i(x)$ is defined over all $x \in \Omega$.

- In analogue with the regularization edges (2.5), within each image layer $i = 1 \ldots n-1$, a spatial flow function is given by the vector field $q_i \in (C_0^\infty \Omega)^m$, where m is the dimension of Ω.

As the generalization of the flow constraints (2.7) - (2.11) given by the graph setting, we set the capacity and conservation constraints on the flow functions $p_i(x)$ and $q_i(x)$:

$$|q_i(x)| \,\leq\, C_i(x) \qquad \text{for } x \in \Omega\,,\ i = 1, \ldots, n-1 \qquad (3.13)$$

$$p_i(x) \,\leq\, \rho(\ell_i, x) \qquad \text{for } x \in \Omega\,,\ i = 1, \ldots, n \qquad (3.14)$$

$$\big(\operatorname{div} q_i - p_i + p_{i+1}\big)(x) \,=\, 0 \qquad \text{for } x \in \Omega\,,\ i = 1, \ldots, n-1 \qquad (3.15)$$

$$q_i \cdot n = 0 \qquad \text{on } \partial\Omega\,,\ i = 1, \ldots, n-1\,. \qquad (3.16)$$

[1]The notation a.e. stands for "for almost every", which means the constraint (3.15) should hold in the integrable and weak sense for every $x \in \Omega$, expect possibly a subset of zero measure.

Equivalently to Ishikawa's max-flow formulation (2.12), we propose a *continuous max-flow model* by maximizing the total amount of source flow

$$\sup_{p,q} E^P(p,q) = \int_\Omega p_1(x)\,dx \tag{3.17}$$

subject to the flow constraints (3.13), (3.14) and (3.15). In this work, we call (3.17) the *dual model*. We will prove it is equivalent or dual to the *primal model* (3.11) in the following sections.

3.2.2. Primal-Dual Model. Now we start from the proposed continuous max-flow model (3.11) and introduce the multiplier function $\lambda_i(x)$, $i = 1 \dots n-1$, to each linear equality constraint (3.15), i.e. the flow conservation condition. We, consequently, get the following primal-dual formulation:

$$\inf_{\lambda}\sup_{p,q} \quad E(p,q;\lambda) = \int_\Omega \big\{p_1 + \sum_{i=1}^{n-1} \lambda_i\big(\operatorname{div} q_i - p_i + p_{i+1}\big)\big\}\,dx \tag{3.18}$$

subject to (3.13) and (3.14). (3.18) is called the *primal-dual model*. It is equivalent to the continuous max-flow model (3.17) and can be rearranged and equally represented by

$$\begin{aligned} \inf_{\lambda}\sup_{p,q} \quad & E(p,q;\lambda) = \sum_{i=1}^{n}\int_\Omega (\lambda_{i-1}-\lambda_i)p_i\,dx + \sum_{i=1}^{n-1}\int_\Omega \lambda_i \operatorname{div} q_i\,dx \\ \text{s.t.} \quad & |q_i(x)| \le \alpha\,,\ i=1\dots n-1\,; \quad p_i(x) \le \rho(\ell_i,x)\,,\ i=1\dots n \quad \forall x\in\Omega\,. \end{aligned} \tag{3.19}$$

For the *primal-dual model* (3.18) introduced above, we have

PROPOSITION 3.1. *There exists at least one saddle-point for* (3.18)*, and the min and max operator of* (3.18) *are interchangeable, i.e.*

$$\sup_{p,q}\inf_{\lambda}\ E(p,q;\lambda) = \inf_{\lambda}\sup_{p,q}\ E(p,q;\lambda)\,. \tag{3.20}$$

To see this, we observe the following facts: for the primal-dual model (3.18), the conditions of the minimax theorem (see e.g., [21] Chapter 6, Proposition 2.4) are all satisfied: the constraints of flows are convex and the energy functional is linear over both the dual variables $\lambda_i(x)$, $i = 1 \dots n-1$ and the primal variables $p_i(x)$, $i = 1 \dots n$, $q_i(x)$, $i = 1 \dots n-1$. This also implies the existence of at least one saddle point [21].

Clearly, the optimization of (3.18) over the dual functions $\lambda_i(x)$, $i = 1 \dots n-1$, leads back to the primal max-flow model (3.17).

3.2.3. Duality btw. (3.17) **and** (3.11)**.** In this section, we build up the duality or equivalence between the *primal model* (3.11) and the continuous max-flow model (3.17):

PROPOSITION 3.2. *The continuous max-flow problem* (3.17) *and the continuous min-cut problem* (3.11) *are dual to each other.*

Proof. We first consider the optimization problem

$$f(v) = \sup_{w\le C} v\cdot w\,, \tag{3.21}$$

where v, w and C are scalars.

When $v < 0$, w can be negative infinity in order to maximize the value $v \cdot w$, i.e. $f(v) = +\infty$. It can also be easily seen that

$$\begin{cases} \text{if } v = 0\,, & \text{then } w \le C \text{ and } f(v) = 0, \\ \text{if } v > 0\,, & \text{then } w = C \text{ and } f(v) = v\cdot C \end{cases} .$$

Therefore, we have

$$f(v) = \begin{cases} v \cdot C & \text{if } v \geq 0 \\ \infty & \text{if } v < 0 \end{cases} \tag{3.22}$$

By the facts (3.21) and (3.22), the function $f(v)$ provides a prototype to maximize the primal-dual model (3.19) over the flow constraints (3.14), i.e. $p_i(x) \leq \rho(\ell_i, x)$, $i = 1 \dots n$.

Define

$$f_i(x) \;=\; \sup_{p_i(x) \leq \rho(\ell_i, x)} \left(\lambda_{i-1}(x) - \lambda_i(x)\right) p_i(x)\,, \quad i = 1 \dots n\,.$$

In view of of (3.22), we have

$$f_i(x) \;=\; \begin{cases} \left(\lambda_{i-1}(x) - \lambda_i(x)\right) \rho(\ell_i, x) & \text{if } \lambda_{i-1}(x) \geq \lambda_i(x) \\ \infty & \text{if } \lambda_{i-1}(x) < \lambda_i(x) \end{cases} \quad i = 1, ..., n \tag{3.23}$$

On the other hand, it is well known that [27]

$$\sup_{|q(x)| \leq \alpha} \int_\Omega \lambda \operatorname{div} q \, dx \;=\; \alpha \int_\Omega |\nabla \lambda| \, dx\,. \tag{3.24}$$

Given (3.23) and (3.24) for the primal-dual model (3.19), we then end up with the primal model (3.11) along with the constraints $\lambda_{i-1}(x) \geq \lambda_i(x)$, $i = 1 \dots n$, for all $x \in \Omega$. If these constraints on optimal λ are not met, the primal-dual energy is infinite and the solution doesn't exist. This contradicts the existence of at least one saddle point, see Prop. 3.1.

In view of the equivalence between the continuous max-flow model (3.17) and the primal-dual model (3.19), Prop. 3.2 is therefore proved. □

With the duality between (3.17) and (3.11) proposed by Prop. 3.2, it is easy to see that optimal layer functions $\lambda_i(x)$, $i = 1 \dots n$, to the convex relaxation model (3.11) just work as the optimal multipliers to the flow conservation condition (3.15) of the continuous max-flow model (3.17). This is the motivation for the new fast algorithm to compute the layer functions, proposed in Sec. 3.4, through the flow-maximization formulation (3.17).

3.3. Exact and Global Optimums. The functions λ_i, $i = 1 \dots n-1$, of the convex model (3.11) are relaxed to take values in the convex set $[0, 1]$, which is in contrast to the binary constraints of the original nonconvex formulation (3.10). The following proposition establishes a strong primal-dual relationship between the max-flow problem (3.17) and the original non-convex problem (3.10). By solving the max-flow problem (3.17) a set of optimizers to the original binary constrained problem (3.10) can be obtained by thresholding each layer function $\lambda_i^*(x)$. When the set of label values are continuous, an analogous thresholding scheme was shown to hold exactly in [52].

PROPOSITION 3.3. *Let $(p^*, q^*; \lambda^*)$ be any optimal saddle-point of* (3.18)*. Let $\{t_i\}_{i=1}^{n-1}$ be a sequence such that $0 < t_1 = t_2 = ... = t_{n-1} \leq 1$, define the level sets*

$$S_i^{t_i} = \{x \,:\, \lambda_i^*(x) \geq t_i\}\,, \quad i \,=\, 1 \dots n-1 \tag{3.25}$$

and let $\lambda_i^{t_i}(x)$ be the characteristic function of $S_i^{t_i}$, i.e.

$$\lambda_i^{t_i}(x) \,:=\, \begin{cases} 1\,, & \lambda_i^*(x) \geq t_i \\ 0\,, & \lambda_i^*(x) < t_i \end{cases} \quad .$$

then the set of binary functions $\lambda_i^{t_i}(x)$, $i = 1, \ldots, n-1$, is a global optimum of the original nonconvex multi-labeling problem (3.10)*. Moreover, the cut given by $\lambda_i^{t_i}(x)$, $i = 1, \ldots, n-1$, has an energy equal to the max flow energy in* (3.17)*, i.e.*

$$E^D(\lambda^t) = \int_\Omega p_1^*(x)\, dx = E^P(p^*).$$

Proof. Since p_i^*, $i = 1, ..., n$ and q_i^*, λ_i^*, $i = 1, ..., n-1$ is a global optimum of the primal-dual problem (3.18), then p_i^*, q_i^* optimize the dual problem (3.17) and $\lambda_i^*(x)$ optimizes (3.11).

For simplification reasons, define $t_0 = 0$ such that $S_0^{t_0} = \Omega$. Since l_i is increasing with i we must have

$$S_0^{t_0} \supseteq S_1^{t_1} \supseteq S_2^{t_2} \supseteq ... \supseteq S_{n-1}^{t_{n-1}}$$

Since the variables are optimal, the flow conservation condition (3.15) must hold, i.e

$$\operatorname{div} q_i^*(x) - p_i^*(x) + p_{i+1}^*(x) \;=\; 0\,, \quad \text{a.e. } x \in \Omega,\ i = 1, ..., n-1.$$

The proof is given by induction. For any $k \in \{1, ..., n-1\}$ define the function

$$E^k \;= \sum_{i=1}^{k} \int_{S_{i-1}^{\ell_{i-1}} \setminus S_i^{t_i}} \rho(\ell_i, x)\, dx + \int_{S_k^{\ell_k}} p_{k+1}^*(x)\, dx + \alpha \sum_{i=1}^{k} L_{S_i^{t_i}}$$

where $L_{S_i^{t_i}}$ is the length of the perimeter of the set $S_i^{t_i}$. We will prove $E^k = E^P(p^*)$ for any $k \in \{1, ..., n-1\}$ and start by considering $k = 1$. By the formula (3.23), it is easy to see that

$$p_1^*(x) = \rho(\ell_1, x), \quad \text{for any point } \ x \in \Omega \setminus S_1^{t_1} = S_0^{t_0} \setminus S_1^{t_1}$$

This, together with the fact that

$$p_1^*(x) = p_2^*(x) + \operatorname{div} q_1^*(x), \quad \text{a.e. } x \in S_1^{t_1}$$

implies that the total max-flow energy defined in (3.17) can be written

$$\begin{aligned} E^P(p^*) \;&= \int_{\Omega \setminus S_1^{t_1}} \rho(\ell_1, x)\, dx + \int_{S_1^{t_1}} \left(p_2^*(x) + \operatorname{div} q_1^*(x)\right) dx \\ &= \int_{\Omega \setminus S_1^{t_1}} \rho(\ell_1, x)\, dx + \int_{S_1^{t_1}} p_2^*(x)\, dx + \int_{S_1^{t_1}} \operatorname{div} q_1^*(x)\, dx \\ &= \int_{S_0^{t_0} \setminus S_1^{t_1}} \rho(\ell_1, x)\, dx + \int_{S_1^{t_1}} p_2^*(x)\, dx + \alpha L_{S_1^{t_1}} \;= E^1 \end{aligned}$$

The last term follows from Prop 4 of [5], or from the fact that $(q_i^* \cdot n)(x) = \alpha$ at all $x \in \partial S_i^{\ell_i}$ combined with the Gaussian theorem

$$\int_{S_i^{\ell_i}} \operatorname{div} q_i^*(x)\, dx \;=\; \int_{\partial S_i^{\ell_i}} q_i^* \cdot n\, ds \;=\; \alpha \left|\partial S^\ell\right|\,. \tag{3.26}$$

Assume now that $E^k = E^P(p^*)$ for some $k \in \{1, ..., n-2\}$, we will show this implies $E^{k+1} = E^P(p^*)$

$$E^P(p^*) = E^k \;= \sum_{i=1}^{k-1} \int_{S_{i-1}^{\ell_{i-1}} \setminus S_i^{t_i}} \rho(t_i, x)\, dx + \int_{S_{k-1}^{\ell_{k-1}}} p_k^*(x)\, dx + \alpha \sum_{i=1}^{k-1} L_{S_i^{t_i}}\,.$$

By the definition (3.25) it follows that $\lambda_{k-1}(x) - \lambda_k(x) > t_{k-1} - t_k = 0$ for all $x \in S_{k-1}^{t_{k-1}} \backslash S_k^{t_k}$. Therefore, by formula (3.23), for any point $x \in S_{k-1}^{t_{k-1}} \backslash S_k^{t_k}$ we must have $p_k^*(x) = \rho(\ell_k, x)$. Combining this with the fact that

$$p_k^*(x) = p_{k+1}^*(x) + \operatorname{div} q_k^*(x), \text{ a.e. } x \in \Omega$$

the above expression can be written

$$\begin{aligned} E^P(p^*) = E^k \; = & \sum_{i=1}^{k-1} \int_{S_{i-1}^{\ell_{i-1}} \backslash S_i^{t_i}} \rho(t_i, x)\, dx + \int_{S_{k-1}^{\ell_{k-1}} \backslash S_k^{\ell_k}} \rho(\ell_k, x)\, dx \\ & + \int_{S_k^{\ell_k}} p_{k+1}^*(x)\, dx + L_{S_k^{t_k}} + \alpha \sum_{i=1}^{k-1} L_{S_i^{t_i}} \; = E^{k+1}. \end{aligned} \tag{3.27}$$

Hence, we can conclude that also $E^{n-1} = E^P(p^*)$. By noting from (3.23) that for all $x \in S_{n-1}^{t_{n-1}}$ we must have $p_n^*(x) = \rho(\ell_n, x)$, the total max flow energy defined in (3.17) can be written

$$\begin{aligned} E^P(p^*) = E^{n-1} \; = & \int_{\Omega \backslash S_1^{t_1}} \rho(\ell_1, x)\, dx + \sum_{i=2}^{n-1} \int_{S_{i-1}^{\ell_{i-1}} \backslash S_i^{t_i}} \rho(t_i, x)\, dx \\ & + \int_{S_{n-1}^{t_{n-1}}} \rho(\ell_n, x)\, dx + \alpha \sum_{i=1}^{n-1} L_{S_i^{t_i}} \end{aligned} \tag{3.28}$$

By writing this expression in terms of the characteristic functions $\lambda_i^{t_i}$ of each region $S_i^{t_i}$, we get

$$E^P(p^*) = \sum_{i=1}^{n} \int_\Omega (\lambda_{i-1}^{\ell_{i-1}}(x) - \lambda_i^{t_i}(x))\, \rho(t_i, x)\, dx \; + \; \alpha \sum_{i=1}^{n-1} \int_\Omega |\nabla \lambda_i^{t_i}|\, dx = E^D(\lambda^\ell)$$

which is exactly the primal model energy (3.11) of the set of binary functions $\lambda_i^{t_i}$. Therefore, by duality between the max-flow problem (3.17) and the convex relaxation problem (3.11), $\lambda_i^{t_i}$ must be a global minimum of the min-cut problem (3.11) and therefore also a global minimum of the original problem (3.10). □

3.4. Multiplier-Based Max-Flow Algorithm. As shown in the previous section, the primal-dual energy formulation (3.18) is just the lagrangian functional of (3.17) and the multiplier functions $\lambda_i(x)$, $i = 1 \ldots n-1$, to the linear flow-conservation equalities (3.15) simply correspond to the layer/labeling functions. We now define the respective augmented lagrangian functional of (3.17) as

$$L_c(p, q, \lambda) \; := \; \int_\Omega p_1\, dx + \sum_{i=1}^{n-1} \int_\Omega \lambda_i (\operatorname{div} q_i + p_{i+1} - p_i)\, dx - \frac{c}{2} \sum_{i=1}^{n-1} \| \operatorname{div} q_i + p_{i+1} - p_i \|^2 , \tag{3.30}$$

where $c > 0$.

In this section, we assume the functions λ, p, q and operators $\int, \operatorname{div}, \nabla$ are discretized, but stick with the continuous notation for simplicity. An algorithm is constructed based on the classical augmented Lagrangian method (or alternating directions method of multipliers (ADMM)) [7], which alternatively maximizes the energy with respect to the flow variables $p(x)$ and $q(x)$ and finally updates the multiplier functions $\lambda_i(x)$, $i = 1 \ldots n-1$. Convergence

Algorithm 1 Multiplier-Based Maximal-Flow Algorithm

Choose some starting values for p^1, q^1 and λ^1, let $k, i = 1$ and start $k-$th iteration, which contains the following steps, until convergence:

- For each layer $i = 1 \dots n-1$, each step repeats as follows:
  - Optimize p_i by fixing other variables

$$\begin{aligned} p_i^{k+1} &:= \arg \max_{p_i(x) \le \rho(\ell_i, x)} L_c((p_{j<i}^{k+1}, p_i, p_{j>i}^k), (q_{j<i}^{k+1}, q_{j\ge i}^k), \lambda^k) \\ &:= \arg \max_{p_i(x) \le \rho(\ell_i, x)} -\frac{c}{2} \left\| p_i + \operatorname{div} q_{i-1}^{k+1} - p_{i-1}^{k+1} - \lambda_{i-1}^k / c \right\|^2 \\ &\quad - \frac{c}{2} \left\| p_i - (p_{i+1}^k + \operatorname{div} q_i^k) + \lambda_i^k / c \right\|^2 \end{aligned}$$

    which can be explicitly computed at each point $x \in \Omega$;
  - Optimize q_i, by introducing the new value of p_i^{k+1} and fixing other variables

$$\begin{aligned} q_i^{k+1} &:= \arg \max_{\|q\|_\infty \le \alpha} L_c((p_{i\le j}^{k+1}, p_{i>j}^k), (q_{j<i}^{k+1}, q_i, q_{j>i}^k), \lambda^k) \\ &:= \arg \max_{\|q\|_\infty \le \alpha} -\frac{c}{2} \left\| \operatorname{div} q_i + p_{i+1}^k - p_i^{k+1} - \lambda_i^k / c \right\|^2 , \end{aligned} \tag{3.29}$$

    which can either be solved iteratively by the projected-gradient algorithm [14], or approximately by one linearized step (3.31);
  - Optimize p_i again, by introducing the new values of q_i^{k+1} and fixing others

$$p_i^{k+1} := \arg \max_{p_i(x) \le \rho(\ell_i, x)} L_c((p_{j<i}^{k+1}, p_i, p_{j>i}^k), (q_{j\le i}^{k+1}, q_{j>i}^k), \lambda^k) ,$$

    which can be explicitly computed at each point $x \in \Omega$;
- At the first and last layer, it is a little different to update the flow functions p_1 and p_n which are given below:

$$\begin{aligned} p_1^{k+1} &:= \arg \max_{p_1(x) \le \rho(\ell_1, x)} L_c(p_1, p_2^k, ..., p_n^k, q^{k+1}, \lambda^k) \\ &:= \arg \max_{p_1(x) \le \rho(\ell_1, x)} \int_\Omega p_1 \, dx - \frac{c}{2} \left\| p_1 - (p_2^k + \operatorname{div} q_1^{k+1}) + \lambda_1^k / c \right\|^2 , \end{aligned}$$

  and

$$\begin{aligned} p_n^{k+1} &:= \arg \max_{p_n(x) \le \rho(\ell_n, x)} L_c(p_1^{k+1}, ..., p_{n-1}^{k+1}, p_n, q^{k+1}, \lambda^k) \\ &:= \arg \max_{p_n(x) \le \rho(\ell_n, x)} -\frac{c}{2} \left\| p_n + \operatorname{div} q_{n-1}^{k+1} - p_{n-1}^{k+1} - \lambda_{n-1}^k / c \right\|^2 . \end{aligned}$$

  Both can be computed explicitly;
- Update multipliers λ_i, $i = 1, \dots, n-1$, by

$$\lambda_i^{k+1} = \lambda_i^k - c \, (\operatorname{div} q_i^{k+1} - p_i^{k+1} + p_{i+1}^{k+1}) ;$$

- Repeat the above steps until convergence.

of such an algorithm can be validated by standard convex optimization theories. For the two-label case, a similar flow-maximization scheme for the continuous min-cut problem was

proposed in [59, 60] and demonstrated a significantly faster convergence than state of the art [12].

Instead of solving the sub-problem (3.29) iteratively by the projected-gradient algorithm [14], an inexact solution can be obtained by the linearization:

$$q_i^{k+1} = \Pi_\alpha\Big(q_i^k + c\nabla(\operatorname{div} q_i^k + p_{i+1}^k - p_i^{k+1} - \lambda_i^k/c).\Big) \tag{3.31}$$

where Π_α is the projection onto the convex set $C_\alpha = \{q \,|\|q\|_\infty \leq \alpha\}$. There are extended convergence results for such a linearization [22, 30] for closely related problems.

Both variants of the algorithm is demonstrated to converge significantly faster than the primal-dual algorithm [52] which will be discussed in Section 5.3.

4. Extension to Continuous Labels. Now we extend the material in Section 3 to the case where the feasible label values are constrained to the interval $[\ell_{\min}, \ell_{\max}]$, i.e. the total number of labels goes to infinity. We address such a continuous labeling problem by a direct extension of the continuous max-flow model (3.17). In this section, we first propose the novel max-flow model, then derive its equivalent min-cut formulation. Finally, we compare with the work proposed by Pock et. al. [53].

4.1. Max-Flow Model. As the number of labels goes to the limit of infinity, the max-flow problem (3.17) with the flow constraints (3.13)-(3.15) turns into

$$\sup_{p,q} \quad \int_\Omega p(\ell_{\min}, x)\, dx \tag{4.1}$$

$$\text{s.t.} \quad p(\ell, x) \,\leq\, \rho(\ell, x)\,, \quad |q(\ell, x)| \,\leq\, \alpha, \qquad \forall x \in \Omega, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}] \tag{4.2}$$

$$\operatorname{div}_x q(\ell, x) + \partial_\ell\, p(\ell, x) \,=\, 0\,, \qquad \text{a.e. } x \in \Omega, \quad \ell \in [\ell_{\min}, \ell_{\max}]. \tag{4.3}$$

$$q(., \ell) \cdot n = 0, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}] \tag{4.4}$$

where $\ell \in [\ell_{\min}, \ell_{\max}]$ is the set of all feasible continuous-valued labels. The flow functions $p(x)$ and $q(x)$ are defined in the one dimensional higher space $[\ell_{\min}, \ell_{\max}] \times \Omega$.

4.2. Min-Cut Model. Let $\lambda(\ell, x)$ be the multiplier function to the flow conservation constraint (5.3). The equivalent primal-dual model to (5.1) can be written as

$$\sup_{p,q} \inf_{\lambda} \int_\Omega p(\ell_{\min}, x)\, dx + \int_{\ell_{\min}}^{\ell_{\max}} \int_\Omega \big\{ \operatorname{div}_x q(\ell, x) + \partial_\ell\, p(\ell, x) \big\}\, \lambda(\ell, x)\, dx\, d\ell \tag{4.5}$$

subject to (5.2).

Likewise, we can also prove duality through the following proposition

PROPOSITION 4.1. *The max-flow model* (5.1) *with continuous label-values is dual / equivalent to the following min-cut model over* $[\ell_{\min}, \ell_{\max}] \times \Omega$*:*

$$\begin{aligned} &\min_{\lambda(\ell,x)\in[0,1]} \int_{\ell_{\min}}^{\ell_{\max}} \int_\Omega \big\{\alpha\, |\nabla_x \lambda| \,-\, \rho(\ell, x) \partial_\ell\, \lambda(\ell, x)\big\}\, dx d\ell \\ &+ \int_\Omega (1 - \lambda(\ell_{\min}, x))\rho(\ell_{\min}, x) + \lambda(\ell_{\max}, x)\rho(\ell_{\max}, x)\, dx \end{aligned} \tag{4.6}$$

subject to

$$\partial_\ell\, \lambda(\ell, x) \,\leq\, 0\,, \quad \lambda(\ell_{\min}, x) \leq 1\,, \quad \lambda(\ell_{\max}, x) \geq 0\,, \quad \forall x \in \Omega, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}]\,. \tag{4.7}$$

Proof. By using integration by parts in ℓ, the primal-dual formulation (5.5) can be rearranged as

$$\sup_{p,q}\inf_{\lambda} \quad + \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} \left\{ \lambda(\ell,x)\ \mathrm{div}_x\, q(\ell,x) - p(\ell,x)\partial_\ell \lambda(\ell,x) \right\} dx\, d\ell$$
$$+ \int_{\Omega} (1-\lambda(\ell_{\min},x))p(\ell_{\min},x) + \lambda(\ell_{\max},x)p(\ell_{\max},x)\, dx. \tag{4.8}$$

subject to (5.2). Observe that optimal λ must satisfy $\lambda(\ell_{\min},x) \leq 1$ and $\lambda(\ell_{\max},x) \geq 0$ for all $x \in \Omega$, otherwise the energy becomes arbitrarily large as $p(\ell_{\min},x)$ and $p(\ell_{\max},x)$ are chosen arbitrarily close to $-\infty$. In the same vein $\partial_\ell \lambda(\ell,x) \leq 0$ for all $\ell \in [\ell_{\min},\ell_{\max}]$ and all $x \in \Omega$, otherwise the energy tends to infinity as $p(\ell,x) \to -\infty$. Hence by maximizing the primal-dual energy (5.8) w.r.t the flow functions p and q, we obtain (5.6) with the constraints (5.7).

The leftmost constraint in (5.7) forces the function $\lambda(\ell,x)$ to be monotonically nonincreasing in ℓ. It corresponds to the constraint (3.5) for discrete label values.
□

In analogue with (3.8), the labeling function $u(x)$ can finally be reconstructed from the binary function $\lambda(\ell,x)$ by

$$u(x) \;=\; \ell_{\min} + \int_{\ell_{\min}}^{\ell_{\max}} \lambda(\ell,x)\, d\ell\,.$$

4.3. Comparisons to Pock et al [53]. In [53], Pock et al gave a different continuous formulation of Ishikawa's construction, as the minimization problem over a binary function in $[\ell_{\min},\ell_{\max}] \times \Omega$

$$\min_{\lambda(\ell,x)\in\{0,1\}} \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} \left\{ \alpha\, |\nabla_x \lambda| + \rho(\ell,x)\, |\partial_\ell \lambda(\ell,x)| \right\} dx d\ell\,. \tag{4.9}$$

subject to

$$\lambda(\ell_{\min},x) = 1\,, \quad \lambda(\ell_{\max},x) = 0\,. \tag{4.10}$$

In order to solve this non-convex binary problem, the convex relaxation of [17] was adopted by minimizing over $\lambda(x,\ell) \in [0,1]$. By applying the thresholding result of [17], binary optimums could be obtained by thresholding the computed result.

The main differences between our formulation (5.6), (5.7) and Pock et al's formulation (5.9), (5.10) can be summarized as follows:

First, the constraint $\partial_\ell \lambda(\ell,x) \leq 0$ is not forced explicitly in [53]. However, it turns out the presence of the absolute value of the term $\rho(\ell,x)\,|\partial_\ell \lambda(\ell,x)|$ forces this constraint to hold. Observe that if $\rho(\ell,x) < 0$ is negative, the formulation of (5.9) becomes non-convex, which cannot be solved globally. This is in contrast to our formulation (5.6), which is convex also in this case.

In the more recent work of Pock et. al. [52], a more strict derivation resulted in a little different formulation. In this formulation, the integrand of the energy functional is infinite if $\partial_\ell \lambda(\ell,x) \leq 0$, hence this constraint is forced to hold. Their derivations rely heavily on results from the theory of calibrations [1] and cartesian currents [25, 26]. Label values ranged over the whole real line $\mathbb{R}$ was assumed, which required to impose limits at infinity: $\lim_{\ell\mapsto+\infty} \lambda(\ell,x) = 0$ and $\lim_{\ell\mapsto-\infty} \lambda(\ell,x) = 1$. On the other hand, our studies reveal

an alternative simple theory to this problem based the two dual optimization problems: the max-flow problem (5.1) and the min-cut problem (5.6).

We eventually stick to a finite label value set in practice. After discretization, the label space also becomes discrete in [52]. However, it is has not been proved that all properties, such at the thresholding scheme and monotonicity constraint hold exactly after discretization. In contrast, these properties were proved to hold exactly for our model with discrete label values developed in Section 3.

Last but not the least, a primal-dual algorithm was proposed in [52], which consists of taking ascent steps over the dual variables $p(x)$ and $q(x)$ and descent step over the primal variable $\lambda(x)$, followed by projections of all the variables onto the nearest points of the feasible sets iteratively until convergence.

The algorithms proposed in this work are instead based on the new max-flow formulations (5.1) which are dual to their respective convex relaxation problems. Experiments demonstrate a significant faster convergence rate than the primal-dual algorithm proposed in [52]. All constraints on the labeling function λ are handled implicitly, and consequently λ does not need to be projected onto the feasible set every iteration.

5. Tight Continuous Max-Flow Approach to Pott's Model. In this section, we focus on a tight convex relaxation for Potts model (1.2) as proposed in [50, 13].

5.1. Tight Convex Relaxed Pott's Model. Consider the problem (3.10) formulated with dual variables as follows

$$\min_{\lambda} \sup_{q} \; = \sum_{i=1}^{n} \int_{\Omega} (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx \; + \; \alpha \sum_{i=1}^{n-1} \int_{\Omega} \lambda_i \operatorname{div} q_i \, dx \tag{5.1}$$

subject to

$$\lambda_i(x) \in \{0, 1\}, \quad \forall x \in \Omega \quad i = 1 \ldots n-1\,; \tag{5.2}$$

$$0 \; = \; \lambda_n(x) \; \leq \; \ldots \; \leq \; \lambda_1(x) \; \leq \; \lambda_0(x) \; = \; 1 \quad \forall x \in \Omega; \tag{5.3}$$

$$|q_i(x)| \leq \alpha, \;\; \forall x \in \Omega, \quad i = 1 \ldots n-1\,. \tag{5.4}$$

As discussed in the previous sections, (4.1) can be used to partition the image domain into n sub-regions by $\Omega_i = \{x \in \Omega \quad \text{s.t. } \lambda_{i-1}(x) - \lambda_i(x) = 1\}$, $i = 1 \ldots, n-1$. However, the regularization term in (4.1) does not correspond to the length term as in the Pott's model due to the linear dependence on the size of the jumps.

Recently, a tight convex relaxation for Pott's model was presented in [50, 13], by optimizing (4.1) and replacing the constraint (4.4) with the following convex constraint set C^P:

$$q(x) \in C^P \; = \; \left\{ \, \forall q \in \mathbb{R}^{n \times m} \; | \quad |\sum_{i=i_1}^{i_2} q_i| \leq \alpha \; ; \quad \forall \, (i_1, i_2) \; , \quad 1 \leq i_1 \leq i_2 \leq n-1 \, \right\}, \tag{5.5}$$

$\forall x \in \Omega$ and applying the relaxation $\lambda_i(x) \in [0, 1]$, $i = 1 \ldots n$. It boils down to optimizing

$$\min_{\lambda} \sup_{q} \; = \sum_{i=1}^{n} \int_{\Omega} (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx \; + \; \alpha \sum_{i=1}^{n-1} \int_{\Omega} \lambda_i \operatorname{div} q_i \, dx \tag{5.6}$$

subject to

$$\lambda_i \in [0,1]\,, \qquad i = 1 \dots n-1\,; \quad 0 \,=\, \lambda_n \,\leq\, \dots \,\leq\, \lambda_1 \,\leq\, \lambda_0 \,=\, 1\,; \tag{5.7}$$

and (4.5) .

In comparison to other convex relaxations, such a relaxation is more 'tight' due to the larger constraint set C^P on the dual variables $q_i(x)$, $i = 1 \dots n-1$. Therefore the problem (4.6) is called the *tight convex relaxed Pott's model* in this paper. It was also observed in [50, 13] that this convex model produces optimal layer functions $\{\lambda_i(x)\}_{i=1}^n$ which are binary nearly everywhere, therefore very close to the real global optimum of the Pott's model.

5.2. Tight Continuous Max-Flow Model and Duality. In this section, we investigate the tight convex relaxed Pott's model (4.6) In contrast to the previous works [50, 13], we follow the same continuous flow-maximization idea presented in Sec. 3.2.1 and propose its new continuous max-flow formulation, so-called *tight continuous max-flow model* as follows:

$$\sup_{p,q} \int_\Omega p_1(x)\, dx \tag{5.8}$$

subject to

$$q(x) \in C^P \qquad , \quad \forall x \in \Omega; \tag{5.9}$$

$$p_i(x) \,\leq\, \rho(\ell_i, x)\,, \qquad i = 1 \dots n\,, \quad \forall x \in \Omega; \tag{5.10}$$

$$\big(\operatorname{div} q_i - p_i + p_{i+1}\big)(x) \,=\, 0\,, \quad i = 1 \dots n-1, \quad \text{a.e. } x \in \Omega\,, \tag{5.11}$$

$$q_i \cdot n = 0 \qquad i = 1 \dots n-1, \quad \text{on } \partial\Omega\,. \tag{5.12}$$

Correspondingly, by the introduction of the multiplier functions $\lambda_i(x)$, $i = 1 \dots n-1$, to the flow conservation equalities (4.11), we obtain the equivalent *primal-dual model* corresponding to the *tight continuous max-flow model* as follows:

$$\inf_{\lambda} \sup_{p,q} \int_\Omega \big\{ p_1 + \sum_{i=1}^{n-1} \lambda_i \big(\operatorname{div} q_i - p_i + p_{i+1} \big) \big\}\, dx \tag{5.13}$$

subject to

$$p_i(x) \,\leq\, \rho(\ell_i, x)\,, \quad i = 1 \dots n\,; \quad q(x) \in C^P\,. \tag{5.14}$$

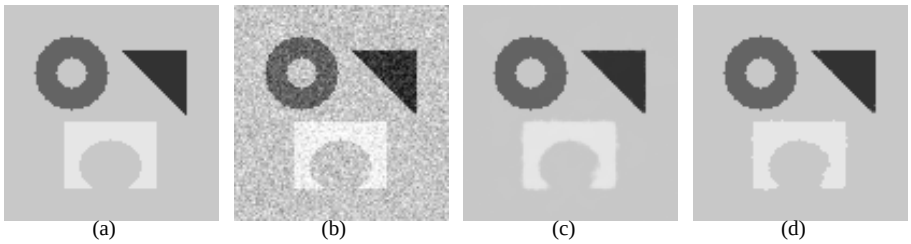
Following the same steps presented in Sec. 3.2.3, we have

PROPOSITION 5.1. *The tight continuous max-flow problem* (4.8) *and the tight convex relaxed Pott's problem* (4.6) *are dual to each other.*

The proof is identical to the proof of Prop. 3.2 and is omitted here.

In view of (4.13), we see that the labeling functions $\lambda_i(x)$, $i = 1 \dots n-1$, work as multipliers to the linear equalities of flow conservation in the tight continuous max-flow model (4.8). Likewise, this paves the way to derive a new max-flow based algorithm to Pott's problem.

5.3. Tight Continuous Max-Flow Based Algorithm for Pott's Model. Observe that the two continuous max-flow problems (3.17) and (4.8) are nearly the same except the constraints on the flow functions $q_i(x)$, $i = 1 \dots n-1$. Therefore, for the tight convex relaxed Pott's model (4.6), it is natural to consider a similar scheme in numerics as Alg. 3.4 as a new algorithm for (4.6), which is called the *tight continuous max-flow based algorithm* in this work. The only difference is that the constraint for $q_i(x)$, $i = 1 \dots n-1$, at any pixel



(a) (b) (c) (d)

FIG. 6.1. *(a) Ground truth, (b) input, (c) Rescaled labeling function before threshold, (d) Rescaled labeling function after thresholding each λ_i at* 0.5.

$x \in \Omega$ is no more the simple $\alpha-$ball: $|q_i(x)| \leq \alpha$, but the more complicated convex set C^P. Therefore, for the new *tight continuous max-flow based algorithm*, we take the very same steps as Alg. 3.4, except the step (3.29) which is replaced by maximizing the same energy over the different constraint set C^P as follows:

- Optimize q_i, $i = 1 \dots n-1$, by fixing the other variables

$$\begin{aligned} q_i^{k+1} &:= \arg\max_{q \in C^P} L_c((p_{i\leq j}^{k+1}, p_{i>j}^k), (q_{j<i}^{k+1}, q_i, q_{j>i}^k), \lambda^k) \\ &:= \arg\max_{q \in C^P} -\frac{c}{2} \left\| \operatorname{div} q_i + p_{i+1}^k - p_i^{k+1} - \lambda_i^k / c \right\|^2 . \end{aligned} \tag{5.15}$$

Clearly, (4.15) can be optimized by the projected-descent step. In this regard, the projection of a vector $(q_1, ..., q_{n-1}) \in \mathbb{R}^{m \times n-1}$ to the convex set C^P becomes our focus, which has no closed form solution in general. However, we can show that for $(q_1, ..., q_{n-1})$, when all components but one are fixed, the projection can be computed analytically. This exactly solves the projection step of (4.15). We state our result as the following proposition:

PROPOSITION 5.2. *Given the set*

$$C_i^P(\bar{q}) := \{q_i \in \mathbb{R}^m \mid (\bar{q}_1, ..., \bar{q}_{i-1}, q_i, \bar{q}_{i+1}, ..., \bar{q}_{n-1}) \in C^P\}, \quad i = 1 \dots n-1 . \tag{5.16}$$

Each $C_i^P(\bar{q})$, $i = 1 \dots n-1$, consist of the intersection of a set of spheres in $\mathbb{R}^m$ with the same radius α. Moreover, for any vector $q_i \in \mathbb{R}^m$, its projection to the set $C_i^P(\bar{q})$ can be computed analytically.

The proof of Prop. 4.2 is omitted here to ease reading, but given in Appendix A.

A primal-dual algorithm was proposed in [52] for optimizing (4.1), which consists of taking ascent steps over the dual variables and descent step over the primal variable $\lambda(x)$, followed by projections of all the variables onto the nearest points of the feasible sets iteratively until convergence. Our continuous max-flow algorithm has the following advantages:

- By Prop. 4.2, we can compute the projection onto the convex set C^P analytically. In contrast, an iterative algorithm (Dyjkstra's algorithm) was proposed for approximately computing the projection in [13], which slows down the algorithm's convergence.
- It avoids the steps to project the layer functions $\lambda_i(x)$, $i = 1 \dots n-1$, onto the convex set $\lambda_i(x) \in [0, 1]$ and the linear-order-constraint set (4.3) at every iteration. According to the theory of the *tight continuous max-flow model*, we can force such constraints on the layer functions implicitly by the simpler constraints on the flow functions $p_i(x)$ (4.10).

6. Numerical Experiments. In this work, we focus on applications of the model (1.3) and (1.2) to image segmentation and stereo reconstruction. Comparisons are made to the discrete approach [34] and the approach proposed by Pock et. al. [53].

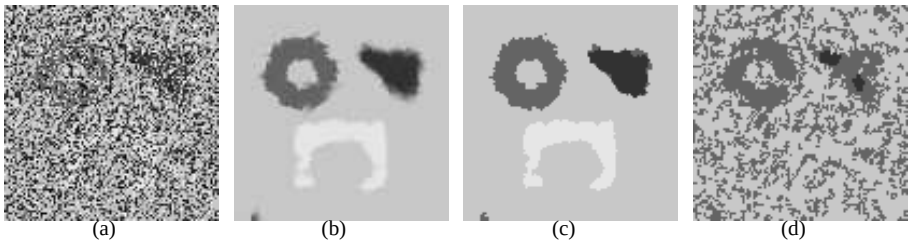


FIG. 6.2. *(a) Input image damaged by impulse noise; (b) reconstructed labeling function with non-convex data term* (6.3) *before threshold, (c) labeling function after thresholding each* λ_i *at* 0.5*, (d) reconstructed labeling function with convex data term* (6.1) *and* $\beta = 1$.

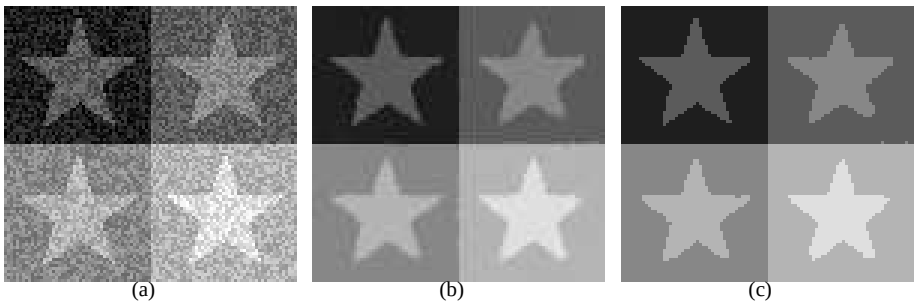


FIG. 6.3. *(a) Input, (b) Labeling function before threshold (c) Labeling function after thresholding each* λ_i *at* 0.5.

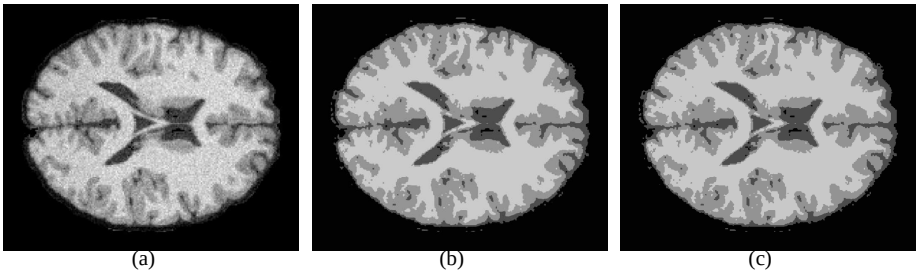
6.1. Image Segmentation. The discrete-valued labeling function $u(x)$ can be used to partition the image into n regions by the convention $u = i$ in region i. Hence $\rho(u(x), x)$ is the cost of assigning the point x to region u. One possibility for such a data term is

$$\rho(i, x) = |I(x) - c_i|^\beta, \quad i = 1, ..., n \tag{6.1}$$

where I is the input image and c_i is the average intensity value of region i. They are assumed to be fixed in this work, although a simple updating scheme can also be constructed for finding a local minimum with respect to **c** as in [4]. Such a data term is convex for $\beta \geq 1$ and non-convex for $\beta < 1$. The term $\alpha \int_\Omega |\nabla u| \, dx$ is used to regularize u. It does not penalize the jump from each region to the next equally, like the more ideal Pott's model. However, for relatively simple images and when the number of regions is not too large, it works quite well. In addition, image segmentation is good for illustrative purposes of the method, since the results are easily visualized. Figure 6.1, 6.4 and 6.3 show results. For ease of visualization, we have rescaled the labeling function u such that u takes the value c_i in region i (instead of the value i), i.e.

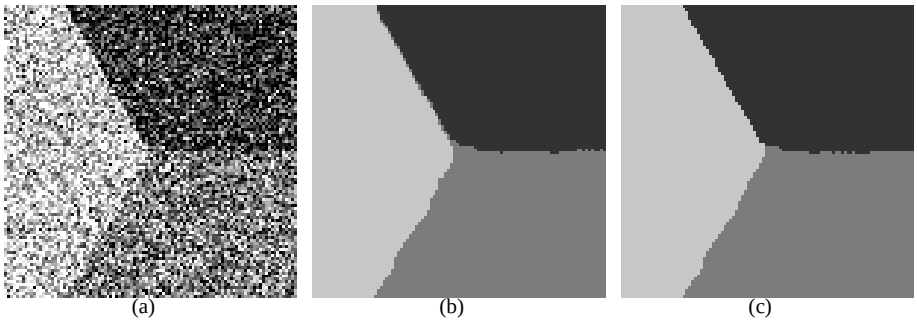
$$u = c_1 + \sum_{i=1}^{n-1} (c_{i+1} - c_i) \lambda_i^* \, . \tag{6.2}$$

Subfigure (b) shows the resulting u before thresholding each $\lambda_i^*(x)$. As expected such a solution may not be binary. Subfigure (c) shows the discrete valued solution after thresholding each $\lambda_i^*(x)$ according to Prop. 3.3. We also demonstrate image segmentation with a non-convex data term. The ground truth image from Figure 6.1 (a) has been damaged by impulse noise in Figure 6.2 (a). More specifically, 70% of the pixels have been randomly selected



(a) (b) (c)

FIG. 6.4. *(a) Input, (b) Labeling function before threshold (c) Labeling function after thresholding each* λ_i *at* 0.5.



(a) (b) (c)

FIG. 6.5. *(a) Input, (b) segmentation with total variation regularized model* (1.3) *(after threshold), (c) segmentation with convex relaxed Pott's model (after threshold). The total variation regularized model results in misclassifications along the boundary between region 1 (darkest) and region 3 (brightest) and does not reconstruct the triple junction properly.*

and given a random number between 0 and 255 (max gray value). For this type of noise, the convex data terms does not perform well, as shown in Figure 6.2 (d) where we have selected (6.1) with $\beta = 1$. Instead the following non-convex data term can be used

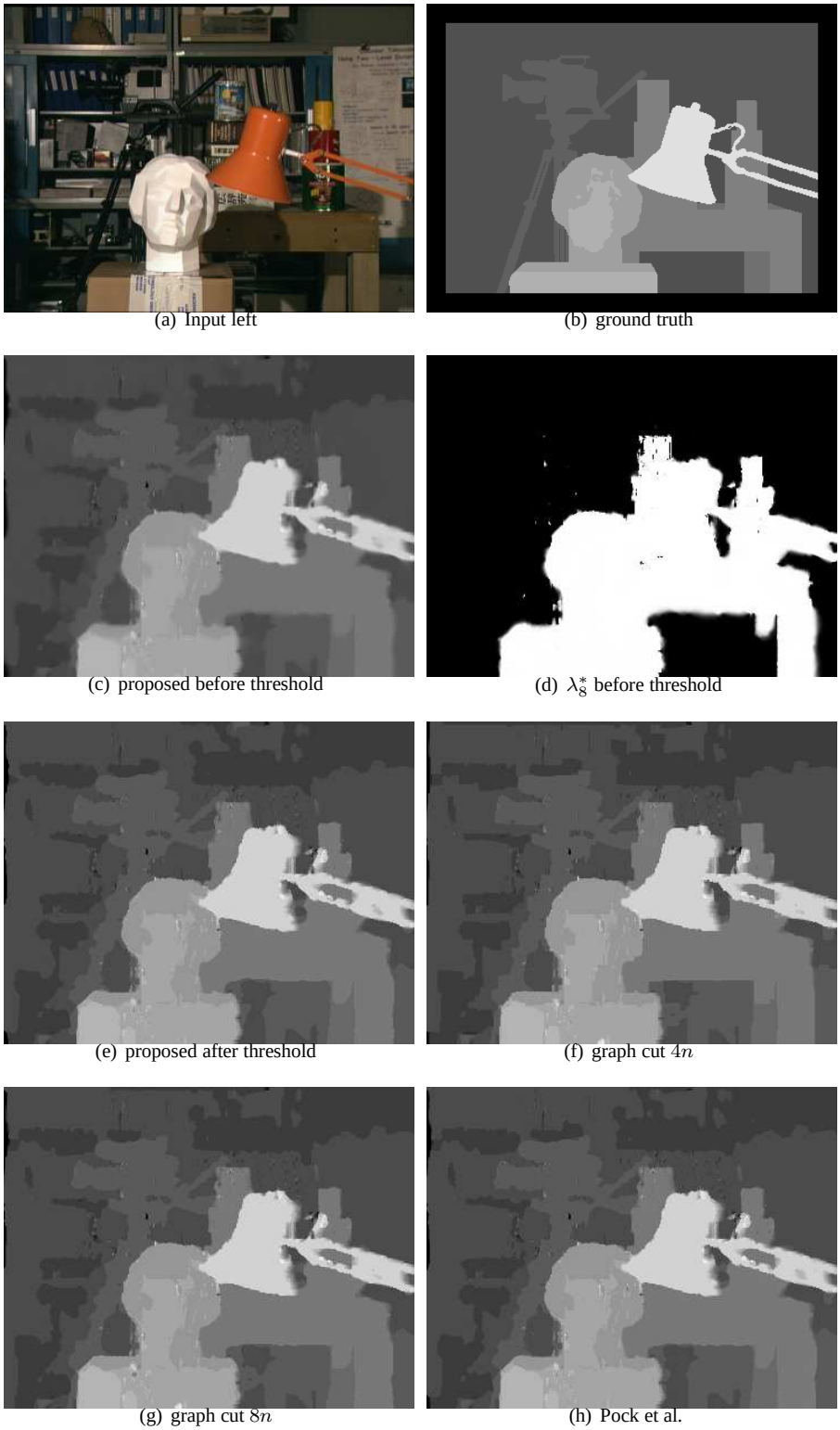
$$\rho(i,x) := \begin{cases} 0\,, & \text{if } i = \text{argmin}_k \, |I(x) - c_k| \\ 1\,, & \text{else} \end{cases} \quad . \tag{6.3}$$

This non-convex problem can be solved globally by our method, the result is shown in Figure 6.2 (b) before threshold and 6.2 (c) after thresholds.

We next apply our algorithm for the convex relaxed Pott's model of [50] from section 4.3. The image in Figure (6.5) (a) has been segmented with the total variation regularized model in (b) and convex relaxed Pott's model in (c). As we see, total variation results in misclassifications along the boundary between region 1 (white) and region 3 (dark) and cannot reconstruct the triple junction properly.

6.2. Stereo reconstruction. We now consider stereo reconstruction with data from the Tsukuba stereo set [55]. Given two color images I_L and I_R of a scene taken from horizontally slightly different viewpoints, we would like to reconstruct the depth map u. The quality of the matching between I_L and I_R for a depth value u is measured by using the following ρ in the data term of (3.1)

$$\rho(u,x) = \sum_{j=1}^{3} |I_L^j(x) - I_R^j(x + (u,0)^T)|. \tag{6.4}$$



(a) Input left (b) ground truth

(c) proposed before threshold (d) λ_8^* before threshold

(e) proposed after threshold (f) graph cut $4n$

(g) graph cut $8n$ (h) Pock et al.

FIG. 6.6. *(a) Left input image, (b) ground truth, (c) non-integer solution* $u = \sum_{i=1}^{n-1} \lambda_i^*$*, (d)* λ_8^* *before threshold, (e) Integer valued solution after threshold, (f) Graph cut 4 neighbors, (g) Graph cut 8 neighbors, (h) Pock et. al.*

	Energy precision $\varepsilon < 10^{-3}$			Energy precision $\varepsilon < 10^{-4}$		
	Primal-dual [52]	Proposed 1	Proposed 2	Primal-dual [52]	Proposed 1	Proposed 2
Brain	280	50 ($\times$ 5)	110	430	65 ($\times$ 5)	280
Figure 6.1	295	35 ($\times$ 5)	115	640	65 ($\times$ 5)	290
Stereo	4055	550 ($\times$ 5)	1070	14305	920 ($\times$ 5)	3905

TABLE 6.1

Iteration counts for each experiment. Number of iterations to reach an energy precision of 10^{-3} and 10^{-4} are shown. Proposed 1 stands for algorithm 2 where the subproblem is solved by 5 iterations of Chambolle's algorithm each outer iteration (indicated by the number in the parenthesis). Proposed 2 stands for Algorithm 2 with the subproblems solved inexactly in one step through the linearization (3.31)*.*

	Energy precision $\varepsilon < 10^{-4}$		Energy precision $\varepsilon < 10^{-5}$		Energy precision $\varepsilon < 10^{-6}$	
	Primal-dual [52]	Proposed 1	Primal-dual [52]	Proposed 1	Primal-dual [52]	Proposed 1
Stereo	14305	920 ($\times$ 5)	> 30000	1310 ($\times$ 5)	> 30000	1635 ($\times$ 5)

TABLE 6.2

Iteration counts for stereo experiment. Number of iterations to reach an energy precision of 10^{-4}, 10^{-5} and 10^{-6} are shown.

Here $I^j(x)$ denotes the $j-th$ component of the color vector $I(x)$. The above data term (6.4) is obviously highly non-convex. The term $\alpha \int_\Omega |\nabla u|\, dx$ is used to regularize u. The strength increases linearly with the size of the jump of u. This is reasonable in stereo reconstruction, since u describes the "depth", which is a physical entity arranged linearly in a third dimension perpendicular to the image planes. Figure 6.6 shows results on a standard example. We have used $\alpha = 0.03$ and scaled images between 0 and 1. As suggested in [55] we have set $n = 17$ and used the discrete label set $\{0, ..., 16\}$. This integer optimization problem over a continuous domain can be formulated exactly with our approach. Solving (3.18) will result in optimal functions λ_i^* that are not necessarily binary. In fact they are not expected to be binary in case the solution to the original problem is not unique. The result of [52] with threshold level 0.5 is depicted in Figure 6.6(h).

We also compare with graph cut using a neighborhood system of 4 and 8. Graph cut produces a single non-unique solution which is shown in Fig 6.6(f) and (g) with 4 and 8 neighbors respectively. As we see, such solutions suffer from metrication artifacts because of the discrete grid bias.

	Primal-dual [52]			Proposed 2		
	iterations	flops pr. iteration	flops	iterations	flops pr. iteration	flops
Triple	280	$1.2 * 10^6$	$3.4 * 10^8$	130	$2.3 * 10^5$	$3.0 * 10^7$
Fig 6.1	230	$2.6 * 10^7$	$6.0 * 10^9$	110	$3.9 * 10^6$	$4.3 * 10^8$

TABLE 6.3

Number of iterations and flops pr iteration to reach energy precision of $\epsilon < 10^{-3}$ for convex relaxed Potts model on the example in Figure 6.5. The proposed algorithm converges in less number of iterations. In addition, the computational cost each iteration is much lower since the expensive iterative projection algorithm onto C^P is avoided. Overall, the proposed algorithm converges around 10 times faster than [52].

6.3. Evaluation of convergence. Iteration counts for all experiments are presented in Table 6.1. The two variants of Algorithm 2 are evaluated against the primal-dual method of Pock et. al. [52]. The relative energy precision at iteration i is given by

$$\varepsilon = \frac{E^i - E^*}{E^*}, \tag{6.5}$$

where E^i is the energy at iteration i and E^* is the final energy. A good estimate of E^* is obtained by using a huge amount of iterations of each method and each experiment. The table shows how many iterations are required to reach an energy precision of 10^{-3} and 10^{-4}. Our algorithms are implemented with a mimetic finite difference spatial discretization [33, 32]. In order to make the comparison as accurate as possible, the primal-dual algorithm [52] is also implemented with such a mimetic finite difference discretization, although a slightly different forward scheme for the gradient and backward scheme for the divergence was used in [52].

The first variant of Algorithm 2 solves the subproblem (3.29) iteratively by Chambolle's algorithm [14]. Since the previous solution is available as a good initialization, not many iterations of this algorithm is required. In our experiments, 5 inner iterations was used for each step. Increasing the number of inner iterations beyond 5 did not seem to have any impact on the convergence rate in our experience.

The primal-dual method of [52] avoids the inner problem, but as we see requires significantly more iterations to reach the same energy precisions. Our algorithm also requires less total number of iterations (inner times outer iterations). The difference becomes progressively clearer with higher energy precision. For the stereo example, which is by far most difficult computationally, our approach reached an energy precision of $\epsilon < 10^{-5}$ after 1310 iterations, $\epsilon < 10^{-6}$ after 1635 iterations and $\epsilon < 10^{-7}$ after 2340 iteration. The primal-dual algorithm [52] failed to ever reach an energy precision of 10^{-5} or lower within our predetermined number of maximum iterations (30000). We believe this difference is due to the fact that our approach avoids the iterative projections of the labeling function and hence progresses in the exact steepest descent direction every iteration.

The second variant of the Algorithm 2 instead computes an inexact solution to (3.29) through the linearization (3.31) and hence avoids the inner iterations. However, the penalty parameter c must be set lower to maintain convergence, hence more outer iterations are required. Overall it converges a little faster than the first variant and outperforms the primal-dual algorithm [52] for all the experiments.

The new algorithm for Pott's model also converges in significantly less iterations than [52] as seen in Table 6.2, where the linearization (3.31) has been used to solve subproblem (3.29). The proposed algorithm also avoids the expensive projection step onto C^P. In contrast, [52] needs to project the dual variables q onto C^P by Dyjkstra's iterative algorithm each iteration, which can only be solved approximately and is the bottleneck of the overall algorithm. In consequence, our algorithm requires significantly less number of floating point operations pr. iteration.

Compared to the highly optimized c++ implementation of discrete max-flow [9], the C implementation of our algorithm converges around 4 times slower. However, our algorithm consists mainly of floating point matrix and vector arithmetic and is therefore highly suited for massive parallel implementation on GPU. Traditional max-flow algorithms have a much more serial nature, which makes them more dependent on an efficient serial CPU. A GPU implementation of the algorithm of Pock et. al. has already been compared to discrete graph cut in [53], showing a speed up factor of about 30. In the near future, hardware improvements are also expected to be largely of the parallel aspect. Hence, we see our work as more suited for the current and future generation of hardware.

7. Conclusions and Future topics. In this paper we proposed and investigated a novel max-flow formulation of multilabeling problems over a continuous image domain. It is a direct mapping of Ishikawa's graph-based configuration to the spatially continuous setting. The multilabeling problem was interpreted as a min-cut problem, which we proved was dual to the proposed continuous max-flow model. In addition, we derived new and reliable multiplier-based max-flow algorithms whose convergence could be verified by standard optimization

theories. Experiments showed that the algorithms outperformed earlier approach in terms of convergence rate. Due to the continuous convex formulation, the algorithm can be more easily speeded up by multi-grid or parallel implementation than graph-based methods, and its memory requirement is not so high.

In comparison to [53] and its improvement [52], our continuous max-flow approach presented a new theoretical framework based on the max-flow dual formulation of discrete-valued constrained problems of the form (1.3); a thresholding scheme was derived which was shown to hold exactly for discrete labels; experiments showed that the max-flow based algorithms converged significantly faster than the primal-dual method proposed in [52]. The algorithm could also be extended to the convex relaxation of Pott's model [50], thereby avoiding expensive iterative projections without closed form solution. In a future work we will also extend this algorithm to the convex relaxation of the piecewise smooth Mumford-Shah model [51], speed up and fine tune the projection algorithm of Section A. Comparisons with a simultaneous work [43] which presented another algorithm for minimizing the energy in the convex formulation of [52] will also be subject of future research.

Appendix A. Projection onto $C_i^P(\bar{q})$.

Observe that $C_i^P(\bar{q})$ is an intersection of spheres in $\mathbb{R}^m$. The centers of the spheres are denoted $\bar{q}_k^j$ for $(k,j) \in I = \{(k,j) \text{ s.t. } 1 \leq k < i < j \leq n\}$ and are defined as $\bar{q}_k^j = \sum_{\ell=k,\ell\neq i}^{j} \bar{q}_\ell$. Let $\mathbb{S}(\mathbf{c},\alpha)$ denote the sphere of center $\mathbf{c} \in \mathbb{R}^m$ and radius α and define $S_k^j(\alpha) = \mathbb{S}(\bar{q}_k^j, \alpha)$. Then $C_i^P(\bar{q})$ is

$$C_i^P(\bar{q}) = \cap_{k=1}^{i-1} \cap_{j=i+1}^{n} S_k^j(\alpha) \tag{A.1}$$

To obtain an analytical expression for the projection onto $C_i^P(\bar{q})$, observe first that

PROPOSITION A.1. *Let $q_k^j = \Pi_{S_k^j(\alpha)} q_i$ be the projection of q_i onto the sphere $S_k^j(\alpha)$. Assume that for some $(k,j) \in I$, $q_k^j \in C_i^P(\bar{q})$, then $q_i^* = \arg\min_{q_k^j \in C_i^P(\bar{q}),(k,j)\in I} |q_i - q_k^j|$ is a projection of q_i onto $C_i^P(\bar{q})$.*

Proof. Let $(K,J) = \arg\min_{(k,j)\in I \text{ s.t. } q_k^j \in C_i^P(\bar{q})} |q_k^j - q_i|$. Assume there exists a q^* with $q^* \in C_i^P(\bar{q})$ and $|q^* - q_i| < |q_K^J - q_i|$. Then $q^* \in S_K^J(\alpha)$ and $|q^* - q_i| < \Pi_{S_K^J(\alpha)} q_i = |q_K^J - q_i|$, a contradiction. □

If $q_k^j \notin C_i^P(\bar{q})$ for all $(k,j) \in I$, the projection onto $C_i^P(\bar{q})$ must necessarily lie on the intersection of the *boundaries* of $S_k^j(\alpha)$ as the next proposition shows. We focus on two dimensional images in $\mathbb{R}^2$ for simplicity, i.e. $m = 2$. In that case, intersections of *boundaries* of $S_k^j(\alpha)$ are just isolated points in $\mathbb{R}^2$. The boundaries of $S_k^j(\alpha)$ are denoted $\partial S_k^j(\alpha)$, i.e.

$$\partial S_k^j(\alpha) = \{x \in \mathbb{R}^m \text{ s.t. } |x - \bar{q}_k^j| = \alpha\}. \tag{A.2}$$

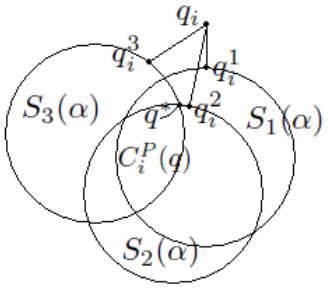
PROPOSITION A.2. *Assume $q_k^j \notin C_i^P(\bar{q})$ for all $(k,j) \in I$. Denote the set of intersections*

$$Q = \left\{x \in \mathbb{R}^2 \text{ s.t. } x \in \partial S_{k'}^{j'}(\alpha) \cap \partial S_k^j(\alpha), \text{ for some } (k',j') \neq (k,j) \in I\right\}. \tag{A.3}$$

Then $\Pi_{C_i^P(\bar{q})} q_i \in Q$.

Proof. Let $q^* = \Pi_{C_i^P(\bar{q})} q_i$. Observe that the projection q^* must lie on the boundary of the set $C_i^P(\bar{q})$, therefore $q^* \in \partial S_k^j(\alpha)$ for some $(k,j) \in I$, say $q^* \in \partial S_K^J(\alpha)$. Since $q^* \in S_K^J(\alpha)$ it follows that $|q^* - q_i| > |q_K^J - q_i|$.

Assume that $q^* \notin Q$. Consider the part of the circle $s \subset \partial S_K^J(\alpha)$, which is the open curve with end points in q^* and q_K^J of minimum length (since there are two possibilities).



(a)

FIG. A.1. *(a) Projection of q_i onto $C_i^P(q)$. The projections q_i^1, q_i^2 and q_i^3 onto $S_1(\alpha), S_2(\alpha)$ and $S_3(\alpha)$ are not contained $C_i^P(q)$, therefore the projection q^* onto $C_i^P(q)$ must lie on the intersection of the boundaries of $S_1(\alpha), S_2(\alpha)$ and $S_3(\alpha)$.*

Since $q^* \in C_i^P(\bar{q})$ and $q_K^J \notin C_i^P(\bar{q})$ it follows that there exists a point $\tilde{q} \in s$ such that $\tilde{q} \in Q$ and $\tilde{q} \in C_i^P(\bar{q})$. Then $|\tilde{q} - q_i| < |q^* - q_i|$, a contradiction to $q^* = \Pi_{C_i^P(\bar{q})} q_i$.
□

When $m = 3$ (3D images), then Q is itself a set of circles in $\mathbb{R}^3$ (and isolated points). The projection onto Q can be computed analytically, but we omit the details.

In is not necessary to check the projection onto every $S_k^j(\alpha)$ for $(k, j) \in I$. As the next result shows, it suffices to check the sphere with largest euclidian distance to q_i.

PROPOSITION A.3. *Let $(K, J) = \arg\max_{(k,j)\in I} |q_i - \bar{q}_k^j|$. If $q_K^J = \Pi_{S_K^J(\alpha)} q_i \in C_i^P(\bar{q})$, then $q_K^J = \Pi_{C_i^P(\bar{q})} q_i$. If $q_K^J \notin C_i^P(\bar{q})$, then $\Pi_{C_i^P(\bar{q})} q_i \in Q$*

Proof. Observe that $|q_i - q^*| = |q_i - \Pi_{C_i^P(\bar{q})} q_i| \geq \max_{(k,j)\in I} |q_i - \Pi_{S_k^j(\alpha)} q_i|$. The inequality follows since $q^* \in S_k^j(\alpha)$ for all $(k, j) \in I$. Let $(K, J) = \arg\max_{(k,j)\in I} q_i - \bar{q}_k^j$. Then $|q_i - \Pi_{S_K^J(\alpha)} q_i| = \max_{(k,j)\in I} |q_i - \Pi_{S_k^j(\alpha)} q_i|$. If $\Pi_{S_K^J(\alpha)} q_i \in C_i^P(\bar{q})$, then by the above inequality $q^* = \Pi_{S_K^J(\alpha)} q_i$. This shows the first part of the proposition. If $\Pi_{S_K^J(\alpha)} q_i \notin C_i^P(\bar{q})$, then by the above inequality $\Pi_{S_k^j(\alpha)} q_i \notin C_i^P(\bar{q})$ for all $(k, j) \in I$, hence $q^* = \Pi_{C_i^P(\bar{q})} q_i \in Q$.
□

Further simplifications can be made. We stick to 2D dimensional problems from now on, i.e. $m = 2$. It is not necessary to check every point in $q^* \in Q$, to find the one in $C_i^P(\bar{q})$ with smallest distance to q_i. The centers of the disks $S_k^j(\alpha)$ are all assumed to be contained in $C_i^P(\bar{q})$ by the construction, i.e.

$$\sum_{\ell=k, \ell\neq i}^{j} \bar{q}_\ell \in C_i^P(\bar{q}), \quad \forall (k, j) \in I, \tag{A.4}$$

which makes the calculation especially simple

PROPOSITION A.4. *Assume $q_k^j \notin C_i^P(\bar{q})$ for all $(i, j) \in I$ and assume* (A.4) *holds. Let $(K, J) = \arg\max_{(k,j)\in I} |q_k^j - q_i|$ and $(K', J') = \arg\max_{(k,j)\in I\setminus(K,J)} |q_k^j - q_i|$ (second largest). If (K, J) is unique then $q^* \in \partial S_K^J(\alpha) \cap \partial S_{k'}^{j'}(\alpha)$ for some $(k', j') \in (K', J')$, if (K, J) is not unique $q^* \in \partial S_k^j(\alpha) \cap \partial S_{k'}^{j'}(\alpha)$ for some $(k, j), (k', j') \in (K, J)$.*

This observation reduces the number intersecting points that needs to be checked. If both the largest and second largest distance is unique then $q^* \in S_K^J(\alpha) \cap \partial S_{K'}^{J'}(\alpha)$ which consists of two points.

Algorithm 2 Exact projection algorithm onto $C_i^P(\bar{q})$

- Compute $(K, J) = \arg\max_{(k,j)\in I} |q_i - q_k^j|$
- if $q_K^J = \Pi_{S_K^J(\alpha)} q_i \in C_i^P(\bar{q})$, then
  $q^* = \Pi_{S_K^J(\alpha)} q_i$
- else
  $(K, J) = \arg\max_{(k,j)\in I} |q_k^j - q_i|$, $(K', J') = \arg\max_{(k,j)\in I\setminus(K,J)} |q_k^j - q_i|$
  $q^* = \arg\min_{q\in\partial C_K^J(\alpha)\cap\partial C_{K'}^{J'}(\alpha),\ (k,j)\neq(K,J)\in I} |q - q_i|$.

A simple algorithm can then be constructed for computing $q^* = \Pi_{C_i^P(\bar{q})} q_i$, see Alg. A. There may be several ways to accelerate the algorithm. In practice, it is expected the boundary of the set $C_i^P(\bar{q})$ is composed of only a few elements of $\partial S_k^j(\alpha)$, so called active elements. Furthermore, the set of active elements $\partial S_k^j(\alpha)$ are known when advancing from one layer to the next, and does not need to be recalculated. The algorithm would only need to work with this set of relevant $C_k^j(\alpha)$.

REFERENCES

[1] G. Alberti, G. Bouchitté, and G. Dal Maso. The calibration method for the mumford-shah functional and free-discontinuity problems. *Calc. Var. Partial Differential Equations*, 16(3):299–333, 2003.

[2] Ben Appleton and Hugues Talbot. Globally optimal surfaces by continuous maximal flows. In *DICTA*, pages 987–996, 2003.

[3] Ben Appleton and Hugues Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):106–118, 2006.

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

[5] 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), 2011.

[6] Andrea L Bertozzi, Selim Esedoglu, and Alan Gillette. Inpainting of binary images using the cahn-hilliard equation. *IEEE Trans Image Process*, 16(1):285–91, 2007.

[7] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, September 1999.

[8] G. Bouchitt'e. Recent convexity arguments in the calculus of variations. In *Lecture notes from the 3rd Int. Summer School on the Calculus of Variations*. Pisa, 1998.

[9] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:359–374, 2001.

[10] Yuri Boykov and Vladimir Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV*, pages 26–33, 2003.

[11] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222 – 1239, 2001.

[12] 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.

[13] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany, November 2008.

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

[15] Antonin Chambolle. Total variation minimization and a class of binary mrf models. In *EMMCVPR*, pages 136–152, 2005.

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

[17] Tony F. Chan, Selim Esedo ̄glu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648 (electronic), 2006.

[18] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, second edition, 2001.

[19] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part i: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, 2006.

[20] Jérôme Darbon. Global optimization for first order markov random fields with submodular priors. *Discrete Appl. Math.*, 157(16):3412–3423, 2009.

[21] Ivar Ekeland and Roger Téman. *Convex analysis and variational problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[22] John Ernest Esser. Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting. (Ph.D. thesis, UCLA CAM-report 10-31), April 2010.

[23] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404.

[24] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[25] Mariano Giaquinta, Giuseppe Modica, and Jiri Soucek. *Cartesian Currents in the Calculus of Variations I, volume 37 of Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge A Series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 1998.

[26] Mariano Giaquinta, Giuseppe Modica, and Jiri Soucek. *Cartesian Currents in the Calculus of Variations II, volume 38 of Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge A Series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 1998.

[27] Enrico Giusti. *Minimal surfaces and functions of bounded variation*. Australian National University, Canberra, 1977.

[28] Andrew V. Goldberg and Robert E. Tarjan. A new approach to the maximum-flow problem. *J. ACM*, 35(4):921–940, October 1988.

[29] Tom Goldstein, Xavier Bresson, and Stanley Osher. Global minimization of markov random fields with applications to optical flow. *UCLA cam-report 09-77*, 2009.

[30] Tom Goldstein, Xavier Bresson, and Stanley Osher. Geometric applications of the split bregman method: Segmentation and surface reconstruction. *J. Sci. Comput.*, 45(1-3):272–293, 2010.

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

[32] J. M. Hyman and M. J. Shashkov. Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids. *Appl. Numer. Math.*, 25(4):413–442, 1997.

[33] J. M. Hyman and M. J. Shashkov. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.*, 33(4):81–104, 1997.

[34] Hiroshi Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1333–1336, 2003.

[35] Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *CVPR*, pages 2993–3000, 2009.

[36] Y.M. Jung, S.H. Kang, and J. Shen. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM Journal on Applied Mathematics*, 67(5):1213–1232, 2007.

[37] Pushmeet Kohli, M. Pawan Kumar, and Philip H.S. Torr. p^3 and beyond: Move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1645–1656, 2009.

[38] Vladimir Kolmogorov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *ICCV*, pages 564–571, 2005.

[39] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006.

[40] Vladimir Kolmogorov and Ramin Zabih. Multi-camera scene reconstruction via graph cuts. In *in European Conference on Computer Vision*, pages 82–96, 2002.

[41] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:65–81, 2004.

[42] Nikos Komodakis and Georgios Tziritas. Approximate labeling via graph-cuts based on linear programming. In *In Pattern Analysis and Machine Intelligence*, page 2007, 2007.

[43] J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal-dual iterations for variational problems in computer vision. *European Conference on Computer Vision (ECCV)*, 2010.

[44] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, HCI, IWR, Uni. Heidelberg, IWR, Uni. Heidelberg, November 2008.

[45] Victor Lempitsky and Yuri Boykov. Global optimization for shape fitting. In *CVPR*, 2007.

[46] H. Li and X.C. Tai. Piecewise constant level set methods for multiphase motion. *International Journal of Numerical Analysis and Modeling*, 4:274–293, 2007.

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

[48] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[49] Nikos Paragios, Yunmei Chen, and Olivier Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

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

[51] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

[52] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. Global solutions of variational models with convex regularization. Technical report, Institute for Computer Graphics and Vision, Graz University of Technology, 2010.

[53] Thomas Pock, Thomas Schoenemann, Gottfried Graber, Horst Bischof, and Daniel Cremers. A convex formulation of continuous multi-label problems. In *ECCV '08*, pages 792–805, 2008.

[54] Renfrey B. Potts. Some generalized order-disorder transformations. *In Proceedings of the Cambridge Philosophical Society, Vol. 48*, pages 106–109, 1952.

[55] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2001.

[56] Gil. Strang. Maximal flow through a domain. *Mathematical Programming*, 26:123–143, 1983.

[57] Olga Veksler. Graph cut based optimization for mrfs with truncated convex priors. In *CVPR*. IEEE Computer Society, 2007.

[58] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, 51:3697–3717, 2002.

[59] J. Yuan, E. Bae, and X.C. Tai. A study on continuous max-flow and min-cut approaches. In *CVPR*, USA, San Francisco, 2010.

[60] J. Yuan, E. Bae, X.C. Tai, and Y. Boykov. A study on continuous max-flow and min-cut approaches. Technical report CAM10-61, UCLA, CAM, September 2010.

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