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

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

Abstract. This work addresses a class of multilabeling problems over a spatially continuous image domain, where the data fidelity term can be any bounded function, not necessarily convex. Two total variation based regularization terms are considered, the first favoring a linear relationship between the labels and the second independent of the label values (Pott's model). In the spatially discrete setting, Ishikawa [33] showed that the first of these labeling problems can be solved exactly by standard max-flow and min-cut algorithms over specially designed graphs. We will propose a continuous analogue of Ishikawa's graph construction [33] by formulating continuous max-flow and min-cut models over a specially designed domain. These max-flow and min-cut models are equivalent under a primal-dual perspective. They can be seen as exact convex relaxations of the original problem and can be used to compute global solutions. Fast continuous max-flow based algorithms are proposed based on the max-flow models whose efficiency and reliability can be validated by both standard optimization theories and experiments. In comparison to previous work [53, 52] on continuous generalization of Ishikawa's construction, our approach differs in the max-flow dual treatment which leads to the following main advantages: A new theoretical framework which embeds the label order constraints implicitly and naturally results in optimal labeling functions taking values in any predefined finite label set; A more general thresholding theorem which, under some conditions, allows to produce a larger set of non-unique solutions to the original problem; Numerical experiments show the new max-flow algorithms converge faster than the fast primal-dual algorithm of [53, 52]. The speedup factor is especially significant at high precisions. In the end, our dual formulation and algorithms are extended to a recently proposed convex relaxation of Pott's model [50], thereby avoiding expensive iterative computations of projections without closed form solution.

Key words. image processing and segmentation, continuous max-flow / min-cut, optimization

AMS subject classifications. . . .

1. Introduction. Many problems in image processing and computer vision can be modeled as energy minimization problems. In image restoration, such minimization problems may be defined over a set of functions which indicate the gray value of the restored image at each pixel. In image segmentation, the minimization problem can be defined over a set of partitions of the image domain. More generally, such problems can be formulated in terms of a labeling function. Examples include image denoising [41, 57] where gray-scale values are directly regarded as labels, image segmentation [13, 2, 14] for which each label represents a region, two-view stereo reconstruction [40, 41] where discrete-valued depths are used as labels, multi-view reconstruction [45] where inside and outside are simply indicated by two labels (see [49] for a good reference to more applications).

Such optimization problems can be addressed by regarding the spatial image domain as either discrete or continuous, leading to respectively variational problems or combinatorial optimization problems. Typically, such problems are modeled as the minimization of an energy which is compromised of a data fitting term and a regularization term. Often the most desirable models are the most difficult to handle from an optimization perspective as they may be non-convex in the continuous setting, or NP-hard in the discrete setting.

In the spatially discrete setting, many such optimization problem can be stated as a Markov random field (MRF) over a discrete image graph. Many techniques have been pro-

---

[*]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)

posed for solving such optimization problems, e.g. graph-cuts [30, 11, 33], message passing [59, 38] and linear programming [42] etc. Graph-cut is an efficient technique which can tackle many such combinatorial optimization problems in case they can be represented as the minimum cut problem over a graph. Greig et al [30] were the first to observe that the min-cut strategy can be applied to exactly solve binary labeling problems in computer vision. However, most labeling problems involving more than two labels are NP-hard, therefore only approximate graph cut algorithms are available, like [13]. However, for a particular set of multilabeling problems with linearly ordered labels and convex interaction penalties, Ishikawa [33] showed that exact solutions can be obtained by graph-cut, even if the data term is not convex. Another exception, proposed in [6], is a discrete variant of the Chan-Vese model [58] with 4 regions. Despite the efficiency of graph based approaches, their computation results are often rough and biased by the discrete grid, i.e. metrication errors [12, 37] are introduced. Reducing such visual artifacts requires either considering more neighboring nodes [12, 37] or applying high-order potentials [36, 34]. However, this either results in a heavy memory load and high computation cost, or amounts to more complex algorithmic schemes, e.g. QPBO [9, 39].

If the image domain Ω is regarded as continuous, the objective is an energy functional and the minimization problem is a variational problem. Numerical solutions can be obtained by instead discretizing and solving the corresponding Euler-Lagrange PDEs. Such a continuous approach avoids the metrication errors, allows to produce results that are rotationally invariant and allows for subgrid accuracy. If discrete label values are assumed, the image is partitioned into several regions. Traditionally, the level set methods [48, 17] and phase field methods [35, 7] were used to approach such partitioning problems. The piecewise constant level set method [46, 47] was later proposed, which assigns an integer label at each point of the image domain. However, both the level set and phase field methods are based on minimization of nonconvex energy functionals. Hence only local optimums can be obtained and the computation results highly depend on the initialization.

1.1. Convex Relaxation Approaches for Pott's model. On the other hand, in the pioneer works of Strang [56] and Chan et. al. [18], it was realized that typical binary image labeling problems in the spatially continuous setting could be formulated and exactly solved by means of convex relaxations. Their approach allows to compute global minimizers of the Pott's model restricted to two phases. The continuous variant of the Potts model [54] describes a partition of the continuous domain Ω into n disjoint subdomains $\{\Omega_i\}_{i=1}^n$ as the minimum of a weighted sum of data fidelity and the length of the partition boundaries

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} \rho(i,x)\,dx + \alpha \sum_{i=1}^n |\partial\Omega_i| \tag{1.1}$$
$$\text{s.t. } \cup_{i=1}^n \Omega_i = \Omega\,, \quad \Omega_k \cap \Omega_l = \emptyset\,, \ \forall k \neq l\,,$$

For instance, if $\rho(i,x) = |I(x) - c_i|^2$, then (1.1) corresponds to the piecewise constant Mumford-Shah model of image segmentation. Restricted to $n = 2$, (1.1) can be written in terms of the characteristic function u of Ω_1 and $1 - u$ of Ω_2 as

$$\min_{u(x)\in\{0,1\}} \int_\Omega (1-u)\rho(2,x)\,dx + \int_\Omega u\rho(1,x)\,dx + \alpha \int_\Omega |\nabla u|\,dx\,. \tag{1.2}$$

It was shown in [18] that the non-convex binary constraint $u(x) \in \{0,1\}$ could be relaxed and replaced by the convex constraint $u(x) \in [0,1]$. Global and exact binary optimums could be obtained by thresholding the result of the convex relaxed problem at almost any level in

2

the interval $[0, 1]$. In [60], such a convex relaxation scheme was redefined under a novel continuous max-flow/min-cut perspective and studied by an elegant variational theory.

Similar relaxation schemes have also been applied to a continuous version of Pott's model with $n > 2$ [63, 44, 50, 15]. Recently, [5, 61] investigated the equivalent dual model of the convex relaxation formulation studied in [63, 44] and derived a simple and fast algorithm in the entropy maximization style. A max flow interpretation of this dual model was given in our recent paper [62], along with a new algorithm. However, as the underlying optimization problem is NP-hard, these relaxations of the Pott's model are not exact, i.e. their reconstructed integer-valued solutions can generally only be accepted as suboptimal, even though the experimental results are promising in terms of the total energy and quality. The relaxation [50, 15] is tightest, meaning it will most often result in global minimums of the original problem (1.1). On the other hand, it is more complicated than [63, 44], especially from a computational perspective, because the number of side constraints grow quadratically in n. Further, projections must be computed onto the feasible set every iteration of the algorithm. Since no closed form solution exist, the projections must be computed by an iterative algorithm, which slows down convergence. As part of this work, a significantly faster algorithm for solving the relaxed problem [50, 15] is proposed.

In contrast to level set and phase field methods, the convex relaxation approach can yield global solutions; fast algorithms can be designed by standard convex optimization theories. In addition, experiments showed that lower energies could be achieved. Their results are therefore expected to be closer to the global integer optimums of the original problem. In contrast to the graph-based methods, the convex relaxation approach yields subgrid accuracy, because of the spatially continuous setting, and perfectly avoid metrication errors due to its crucial rotation invariance. In addition, the reduced numerical schemes can be easily accelerated by multigrid or parallel implementation and require less memory.

1.2. Labeling with Linearly Ordered Labels. In this work, we will start by focusing on image labeling problems of the form

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

where $\rho(u(x), x)$ is any bounded function, not necessarily convex in u. The last term of (1.3) regularizes and is called the (weighted) total variation of u. The applications of (1.3) are numerous. For instance u may represent the gray value in image denoising. The problem (1.3) can also model partitioning problems, like image segmentation, by the convention $u = i$ in region Ω_i and where $\rho(u(x), x)$ is the data cost of assigning x to region Ω_u. However, the regularization term (1.3) does not correspond to the length term in the more ideal Pott's model (1.1), because of its dependence 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. The piecewise constant level set method [47] also has the form of (1.3) if $\ell_1, \ldots, \ell_n = 1, \ldots, n$. In order to minimize the non-convex energy (1.3), is was proposed to represent the integer constraints by polynomials and minimize the resulting non-convex lagrangian functional to a local minima.

In the spatially discrete setting, Ishikawa [33] showed that such labeling problems could be solved globally and exactly by computing the minimum cut over a specially designed graph (see Sec. 2.2 for more details). [53, 52] generalized this result to the totally continuous setting, where both the image domain and label values are continuous, by representing the optimal labeling function as the discontinuity set of a binary function in a one-dimensional higher space. Such a lifting approach is related to earlier mathematical theories of calibrations and cartesian currents [10, 1]. Optimal labeling functions could be obtained by applying the

3

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 generalized to solve vector valued problems [29] in the totally discrete setting. The resulting relaxation is not exact, therefore global solutions cannot in general be obtained. Since the underlying minimization problem is NP-hard, global minimization is not feasible in general for such problems.

1.3. Motivations and Contributions. We will interpret (1.3) as a continuous min-cut problem and derive a continuous max-flow model based on Ishikawa's graph construction, which is shown to be dual to an exact convex relaxation of (1.3). In Section 4, the models are extended to problems with continuous label values and comparisons to [53, 52] show some interesting differences between our relaxed problem and the relaxation of [53, 52]. However, the main conceptual difference is our derivations of the max-flow models, which are dual formulations of the convex relaxed problem. These max-flow formulations implicitly deal with the constraints on the labeling function. By duality theory a more general thresholding scheme can be derived for producing a larger set of solutions to the original non-convex problem from solutions of the relaxed problem than [53, 52]. Furthermore, efficient algorithms can be build up based on our max-flow formulations, which are presented in Section 6.

The max-flow dual formulation and algorithms are further extended to the tightest convex relaxation of Pott's model [50] and will be presented in Section 5. All constraints on the labeling function are handled implicitly, and the algorithm avoids expensive iterative computations of projections of the dual variables without closed form solution as in [50].

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 [19, 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 binary labeling problem: 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; Most related to ours is the work of Yuan et al [60, 61], which 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. [63, 29, 53, 14] etc, tried to conduct the energy minimization over the labeling functions directly.

Motivated by Yuan et al [60] and Ishikawa [33], 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 graph construction. 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). Under some conditions, this scheme allows to produce a larger set of non-unique solutions to the original

4

problem than [53, 52].

- 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 fast 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. Related Works. To motivate works in the following sections, we first revisit the duality correspondence between max-flow and min-cut in the spatially continuous context proposed by Yuan et al [60]. We then give a short review of the graph construction of Ishikawa [33].

2.1. Max-Flow and Minimum s-t Cut. Given a graph $G = (\mathcal{V}, \mathcal{E})$ composed of the vertex set $\mathcal{V}$ and the edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The vertex set $\mathcal{V}$ includes the nodes of a 2-D or 3-D image grid together with two terminal vertices: the source s and the sink t. The edge set $\mathcal{E}$ contains two types of edges: the spatial edges $e_n = (p, q)$ where $p, q \in \mathcal{V} \backslash \{s, t\}$, and the terminal edges or data edges: $e_s = (s, p)$ and $e_t = (p, t)$, where $p \in \mathcal{V} \backslash \{s, t\}$. We assign the cost $C(e)$ to each edge $e \in \mathcal{E}$, which is assumed to be nonnegative, i.e. $C(e) \geq 0$.

An s-t cut assigns two disjoint partitions of $\mathcal{V} \backslash \{s, t\}$ to the source s and the sink t respectively, which divides the nodes of $\mathcal{V} \backslash \{s, t\}$ into two disjoint groups $\mathcal{V}_s$ and $\mathcal{V}_t$:

$$\mathcal{V} = \mathcal{V}_s \bigcup \mathcal{V}_t, \quad \mathcal{V}_s \cap \mathcal{V}_t = \emptyset.$$

To each cut, an energy is defined as the sum of the costs $C(e)$ of edges $e \in \mathcal{E}_{st} \subset \mathcal{E}$, with one end point in $\mathcal{V}_s$ and the other in $\mathcal{V}_t$. The minimal s-t cut problem is to find such a partition V_s of $\mathcal{V}$ with the minimal cut-energy:

$$\min_{\mathcal{E}_{st} \subset \mathcal{E}} \sum_{e \in \mathcal{E}_{st}} C(e). \tag{2.1}$$

On the other hand, each edge $e \in \mathcal{E}$ can be viewed as a pipe and its edge cost $C(e)$ can be regarded as the capacity of this pipe. In this flow 'network', the max-flow problem is to find the largest amount of flow allowed to pass from the source s:

$$\max_{p_s} \sum_{v \in \mathcal{V} \backslash \{s, t\}} p_s(v) \tag{2.2}$$

subject to the following flow constraints:

- *Capacity of Spatial Flows p:* for each spatial edge $e_n := (p, q) \in \mathcal{E}$ where $p, q \in \mathcal{V} \backslash \{s, t\}$, the spatial flow $p(e_n)$ is undirected and constrained by:

$$-C_p(e_n) \leq p(e_n) \leq C_q(e_n)$$

where $C_p(e_n), C_q(e_n) \geq 0$. We study the special case $C_p(e_n) = C_q(e_n) = C(e_n)$ for simplicities, i.e.

$$|p(e_n)| \leq C(e_n); \tag{2.3}$$

5

- *Capacity of Source Flows p_s:* for each edge $(s, v) \in \mathcal{E}$ linking the terminal s to the node $v \in \mathcal{V} \backslash \{s, t\}$, the source flow $p_s(v)$ is directed from s to v. Its capacity $C_s(v)$ indicates that

$$0 \leq p_s(v) \leq C_s(v) \, ; \tag{2.4}$$

- *Capacity of Sink Flows p_t:* for each edge $(v, t) \in \mathcal{E}$ linking the node $v \in \mathcal{V} \backslash \{s, t\}$ to the terminal t, the sink flow $p_t(v)$ is directed from v to t. Its capacity $C_t(v)$ indicates that

$$0 \leq p_t(v) \leq C_t(v) \, ; \tag{2.5}$$

- *Conservation of Flows:* at each node $v \in \mathcal{V} \backslash \{s, t\}$, in-coming flows should be balanced by out-going flows, i.e. all the flows passing v, which includes spatial flows $p(e_n)$ along spatial edges e_n around v, the source flow $p_s(v)$ and the sink flow $p_t(v)$, should be constrained by

$$\left(\sum_{e_n \in \mathcal{E}} p(e_n) \right) - p_s(v) + p_t(v) = 0 \, . \tag{2.6}$$

It is well known that the max-flow formulation (2.2) is equivalent to the min-cut problem (2.1), where the flows are saturated uniformly on the cut edges, i.e. the total flow is bottle-necked by the 'saturated pipes'. By the graph-cut terminologies, when a flow $p(e)$ on the edge $e \in \mathcal{E}$ reaches its corresponding capacity $C(e)$, given in (2.3), (2.4) or (2.5), we call it 'saturated'; otherwise, 'unsaturated'.

2.1.1. Max-Flow and Min-Cut in Continuous Setting. In the recent studies of Yuan et al [60], a direct analogue of the equivalence between the max-flow model (2.2) and the min-cut (2.1) model was discovered in the spatially continuous context. Given the continuous image domain Ω together with two terminals: the source s and the sink t which are linked to each pixel of Ω. Three types of flows are defined at each pixel $x \in \Omega$: the source flow $p_s(x)$ directed from the source s to x, the sink flow $p_t(x)$ directed from x to the sink t and the spatial flow $p(x)$ within the image plane Ω. In analogue with (2.3), (2.4) and (2.5), the three flow fields are constrained by the capacities C_s, C_t and C:

$$p_s(x) \leq C_s(x) \, , \quad p_t(x) \leq C_t(x) \, , \quad |p(x)| \leq C(x) \, ; \quad \forall x \in \Omega \, . \tag{2.7}$$

With the above notations, the flow conservation condition (2.6) can be rewritten in the spatially continuous setting as

$$p_t(x) - p_s(x) + \text{div}\, p(x) = 0 \, , \quad \text{a.e. } x \in \Omega \tag{2.8}$$

where the divergence $\text{div}\, p(x)$ evaluates the excess of the spatial flow around x.

The continuous max-flow problem can therefore be formulated by maximizing the total flow from the source s:

$$\sup_{p_s, p_t, p} \int_\Omega p_s \, dx \tag{2.9}$$

subject to the flow constraints (2.7) and (2.8).

Yuan et al [60] proved that such a continuous max-flow formulation (2.9) is equivalent to the continuous minimum *s-t* cut problem proposed in [18, 14]:

$$\min_{u(x) \in [0,1]} \int_\Omega (1-u)C_s \, dx + \int_\Omega u C_t \, dx + \int_\Omega C(x) \, |\nabla u| \, dx \, . \tag{2.10}$$
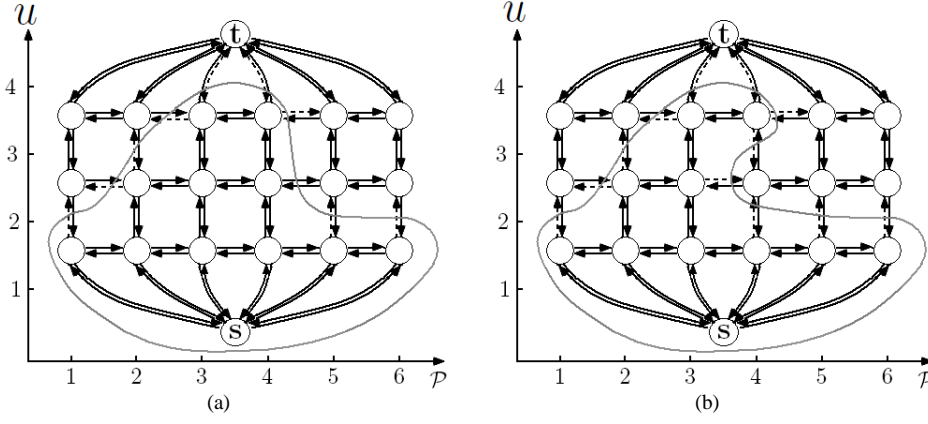
6

FIG. 2.1. *(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.*

More specifically, the max-flow problem (2.9) and the min-cut problem (2.10) are dual to each other and the labeling function $u(x)$ works as a Lagrange multiplier to the flow conservation condition (2.8). In addition, connections between the flow functions $p_s(x), p_t(x), p(x)$ and the labeling function $u(x)$ can be reinterpreted under an elegant variational perspective. An efficient and reliable max-flow based algorithm could be derived and built up based on the continuous max-flow model (2.9) [60].

2.2. Revisit Ishikawa's Work. Ishikawa [33] studied image labeling problems which can be generally formulated as:

$$\min_{u \in U} \sum_{v \in \mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w) \in \mathcal{N}} g(u_v - u_w), \qquad (2.11)$$

where $\mathcal{P}$ denotes a discrete image grid in 2-D or N-D and $\mathcal{H} \subset \mathcal{P} \times \mathcal{P}$ is a neighborhood system on $\mathcal{P}$; $U = \{u : \mathcal{P} \mapsto L\}$ is the set of feasible labeling functions. The potential prior $g(x)$ of (2.11) is assumed to be convex and ρ is any bounded function, but not necessarily convex in u. It was shown that problems of the form (2.11) could be exactly optimized by finding the min-cut on a specially constructed multi-layer graph $G = (\mathcal{V}, \mathcal{E})$, where each layer corresponds to one label.

We adopt Ishikawa's notations [33] in this chapter and make use of 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 simple 1-D example. The vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ are defined as follows:

$$\mathcal{V} = \mathcal{P} \times L \cup \{s, t\} = \{u_{v,i} \mid v \in \mathcal{P}; \ i = 1, ..., n - 1\} \cup \{s, t\} \qquad (2.12a)$$

$$\mathcal{E} = \mathcal{E}_D \cup \mathcal{E}_C \cup \mathcal{E}_P \qquad (2.12b)$$

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

$$\mathcal{E}_D^v = (s, u_{v,1}) \cup \{(u_{v,i}, u_{v,i+1}) \mid i = 1, \ldots, n - 2\} \cup (u_{v,n-1}, t). \qquad (2.13)$$

7

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

$$\mathcal{E}_C^v = \{(u_{v,i+1}, u_{v,i}) \,|\, i = 1, \dots, n-2\}. \tag{2.14}$$

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

$$\mathcal{E}_R = \{(u_{v,i}, u_{w,j}) \,|\, (v,w) \in \mathcal{H}, \ i,j = 1, ..., n\}, \tag{2.15}$$

2.2.1. Anisotropic Total-Variation Regularized Labeling. When a pairwise prior $g(u_v - u_w) = C(u,w) \,|u_v - u_w|$ is given, (2.11) corresponds to an anisotropic total-variation regularized image labeling problem, i.e.

$$\min_{u \in U} \sum_{v \in \mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w) \in \mathcal{N}} C(v,w) \,|u_v - u_w| \tag{2.16}$$

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

We now define capacities on the edges such that the minimum cut on the graph corresponds to a minimizer of (2.16):

- *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

$$p_1(v) \le \rho(\ell_1, v), \quad \forall v \in \mathcal{P}; \tag{2.17}$$

- *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

$$p_i(v) \le \rho(\ell_i, v), \quad \forall v \in \mathcal{P} \quad i = 1, ..., n-2 \tag{2.18}$$

- *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

$$p_n(v) \le \rho(\ell_n, v), \quad \forall v \in \mathcal{P}; \tag{2.19}$$

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

$$|q_i(v,w)| \le C(v,w); \tag{2.20}$$

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.

$$\Big(\sum_{(v,w) \in \mathcal{H}} q_i(v,w) \Big) - p_i(v) + p_{i+1}(v) = 0. \tag{2.21}$$

Since there is no lower bound on the flows (2.17)-(2.19), the capacity on the penalty edges (2.14) is 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.

The max-flow problem is to find the largest amount of flows allowed to pass from the source s to sink t, and can therefore be formulated as

$$\max_{p,q} \sum_{v \in \mathcal{P}} p_1(v) \tag{2.22}$$

subject to the flow constraints (2.17), (2.18), (2.19), (2.20) and (2.21).

Once the maximal flow is computed, a minimal cut can be extracted which corresponds to a minimizer of the problem (2.16).

8

3. Continuous Max-Flow and Min-Cut Models. We now consider the labeling problem (1.3), i.e. the continuous counterpart of (2.11) specialized to the classical total-variation regularizer:

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

where $U = \{u : \ \Omega \mapsto \{\ell_1, ..., \ell_n\}, \quad \text{s.t.} \quad \int_\Omega |\nabla u|\, dx \ < \ \infty\}$ is the set of all feasible functions defined over the continuous image domain Ω. $\rho(u(x), x)$ is any uniformly bounded function, not necessarily convex in the element of u. We show (3.1) can be exactly solved by generalizing the discrete max-flow and min-cut models discussed in the last section to the continuous setting.

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

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

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 function* in this work, are defined by:

$$\lambda_i(x) = \left\{ \begin{array}{ll} 1 & \text{if } u(x) > \ell_i \\ 0 & \text{if } u(x) \le \ell_i \end{array} \right. , \quad i = 1, \ldots, n - 1\,. \tag{3.3}$$

Likewise, we also define $\lambda_0(x) = 1$ and $\lambda_n(x) = 0$, $\forall x \in \Omega$, as the characteristic functins of the set S_0 and S_n respectively.

Since $\ell_1 < \ldots < \ell_n$ we have

$$0 = \lambda_n \ \le \ \ldots \ \le \ \lambda_1 \ \le \ \lambda_0 \ = \ 1 \tag{3.4}$$

and

$$\emptyset \ = \ S_n \ \subseteq \ \ldots \ \subseteq \ S_1 \ \subseteq \ S_0 \ = \ \Omega. \tag{3.5}$$

Now we rewrite the optimization problem (3.1) in terms of the layer functions $\lambda_i(x)$, $i = 1, ..., n - 1$. Such a formulation has also been presented in [20] in the discrete setting for the more general (2.11). In [50] such a formulation was given in the continuous setting in connection with a convex relaxation of Pott's model.

Observe that the subdomain of Ω where $u(x) = \ell_i$ is given by $S_i \backslash S_{i-1}$ and its characteristic function is $\lambda_{i-1}(x) - \lambda_i(x)$. Any function $u(x) \in U$ can therefore be written in terms of $\lambda_i(x)$, $i = 0, ..., n$ as

$$u(x) \ = \ \sum_{i=1}^{n} (\lambda_{i-1}(x) - \lambda_i(x))\ell_i \ = \ \ell_1 + \sum_{i=1}^{n-1} \lambda_i(x)(\ell_{i+1} - \ell_i)\,. \tag{3.6}$$

The data term can written as

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

By the coarea formula [27], for the function $u(x) \in U$ we also have

$$\int_\Omega C(x)|\nabla u(x)|\, dx \ = \ \sum_{i=1}^{n-1} \int_\Omega C_i(x)|\nabla \lambda_i(x)|\, dx \tag{3.8}$$
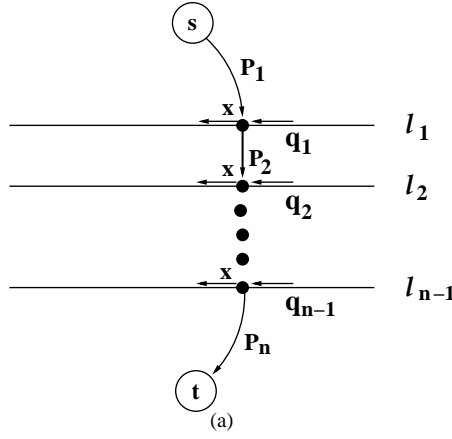
9

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

where $C_i(x) = (\ell_{i+1} - \ell_i)C(x)$, $i = 1, \ldots, n - 1$.

In view of (3.7) and (3.8), the problem (3.1) amounts to search for the optimal layer functions $\lambda_i \in \{0, 1\}$, $i = 1, \ldots, n - 1$ which minimizes

$$\min_{\lambda_i(x) \in \{0,1\}} \sum_{i=1}^{n} \int_{\Omega} (\lambda_{i-1}(x) - \lambda_i(x))\, \rho(\ell_i, x)\, dx \; + \; \sum_{i=1}^{n-1} \int_{\Omega} C_i(x)|\nabla \lambda_i(x)|\, dx \qquad (3.9)$$

subject to the monotonically nonincreasing constraint (3.4). Once the layer functions $\lambda_i(x)$ are computed, the labeling function $u(x)$ can be recovered by (3.6).

We focus on the case where $C(x) = \alpha$ is constant and $\ell_{i+1} - \ell_i = 1$, $i = 1, ..., n - 1$. Of course, the results can be easily extended to other more general $C(x)$ and ℓ_i, $i = 1, ..., n$. In this regard, (3.1) or (3.9) can be equally reformulated as

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

subject to the order constraint (3.4). (3.10) is obviously nonconvex due to the binary setting of $\lambda_i(x) \in \{0, 1\}$, $i = 1, ..., n - 1$.

3.2. Convex Relaxed Models. In the following parts of this section, we show that the nonconvex optimization problem (3.10) can be globally and exactly solved via its convex relaxed version:

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

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

where the binary constraints on the labeling functions $\lambda_i(x) \in \{0, 1\}$, $i = 1, \ldots, n - 1$, are relaxed by the convex constraints $\lambda_i(x) \in [0, 1]$, $i = 1, \ldots, n - 1$. In this work, we call (3.11) the *primal model*.

3.2.1. Continuous Max-Flow. Motivated by the discrete graph configuration (2.12a) and (2.12b) in Sec. 2.2, $n - 1$ copies of the image domain Ω are placed in sequential order

between two terminals: the source s and the sink t. This mixed continuous/discrete domain can be defined as

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

The continuous counterparts of edges, flows and capacities can now be defined (see Fig. 3.1 for an illustration)

- In view of (2.13), 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 $e_i(x)$ a flow function $p_i(x)$ is defined for $i = 1 \ldots n$ and for all $x \in \Omega$. Mathematically, $p_i : \Omega \mapsto \mathbb{R}$ for $i = 1, ..., n$.
- In analogue with the regularization edges (2.15), within each image layer $i = 1 \ldots n-1$, a spatial flow function is given by the vector field $q_i : \Omega \mapsto \mathbb{R}^m$, where m is the dimension of Ω. More specifically, we will optimize over $q_i \in (C^\infty(\Omega))^m$, $i = 1, ..., n$.

As a generalization of the flow constraints (2.17) - (2.21) in the discrete setting, we now give constraints on the flow functions $p_i(x)$ and $q_i(x)$:

$$|q_i(x)| \leq \alpha, \qquad i = 1 \ldots n-1, \quad \forall x \in \Omega; \qquad (3.13)$$

$$p_i(x) \leq \rho(\ell_i, x), \qquad i = 1 \ldots n-1, \quad \forall x \in \Omega; \qquad (3.14)$$

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

Then, in analogue with Ishikawa's max-flow model (2.22), we propose a *continuous max-flow model* by maximizing the total amount of source flow :

$$\sup_{p,q} E^P(p, q) = \int_\Omega p_1(x)\, dx \qquad (3.16)$$

subject to the flow constraints (3.13), (3.14) and (3.15). In this work, we call (3.16) the *dual model*.

3.2.2. Primal-Dual Model. Introduce the multiplier functions $\lambda_i(x)$, $i = 1 \ldots n-1$, to each linear equality constraint of the flow conservation condition (3.15), we then get the equivalent *primal-dual model* of (3.16):

$$\inf_\lambda \sup_{p,q} \ E(p, q; \lambda) = \int_\Omega \left\{ p_1 + \sum_{i=1}^{n-1} \lambda_i \big(\operatorname{div} q_i - p_i + p_{i+1}\big) \right\} dx \qquad (3.17)$$

subject to (3.13) and (3.14).

(3.17) can be rearranged and equally represented by

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

$$\text{s.t.} \quad |q_i(x)| \leq \alpha,\ i = 1 \ldots n-1;\quad p_i(x) \leq \rho(\ell_i, x),\ i = 1 \ldots n.$$

---

[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.

Note that for the primal-dual model (3.17), 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 $\lambda_i(x)$, $i = 1 \ldots n - 1$ and the primal variables $p_i(x)$, $i = 1 \ldots n$, $q_i(x)$, $i = 1 \ldots n - 1$. This also implies the existence of at least one saddle point, see [21]. It also follows that the min and max operators of the primal-dual formulation (3.17) or (3.18) can be interchanged, i.e.

$$\sup_{p,q} \inf_{\lambda} \; E(p, q; \lambda) \; = \; \inf_{\lambda} \sup_{p,q} \; E(p, q; \lambda) \,. \tag{3.19}$$

Clearly, the optimization of (3.17) over the dual functions $\lambda_i(x)$, $i = 1 \ldots n - 1$, leads back to the primal max-flow model (3.16).

3.2.3. Continuous Min-Cut. Now we show that optimization of (3.17) over all flow functions $p(x)$ and $q(x)$, i.e. the righthand side of (3.19), leads to (3.11) as its equivalent *primal model*. Then by the fact that the primal-dual model (3.18) is equivalent to the continuous max-flow problem (3.16), we have

PROPOSITION 3.1. *The continuous max-flow problem* (3.16) *and the continuous min-cut problem* (3.11) *are dual problems.*

The proof is based on the following observations:

We consider the optimization problem

$$f(v) \; = \; \sup_{w \leq C} \; v \cdot w \,, \tag{3.20}$$

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

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

Therefore, we have in general

$$f(v) = \begin{cases} v \cdot C & \text{if } v \geq 0 \\ \infty & \text{if } v < 0 \end{cases} \tag{3.21}$$

By the facts (3.20) and (3.21), the function $f(v)$ provides a prototype to maximize the primal-dual model (3.18) over the flow functions $p_i(x)$, $i = 1 \ldots n$, together with their constraints $p_i(x) \leq \rho(\ell_i, x)$, i.e. (3.14).

For each $x \in \Omega$, define

$$f_i(x) \; = \; \sup_{p_i(x) \leq \rho(\ell_i, x)} \; (\lambda_{i-1}(x) - \lambda_i(x)) \, p_i(x) \,, \quad i = 0 \ldots n \,.$$

In view of of (3.21), we have

$$f_i(x) = \begin{cases} (\lambda_{i-1}(x) - \lambda_i(x)) \, \rho(\ell_i, x) & \text{if } \lambda_{i-1}(x) \geq \lambda_i(x) \\ \infty & \text{if } \lambda_{i-1}(x) < \lambda_i(x) \end{cases} \quad i = 0, ..., n \tag{3.22}$$

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

$$\sup_{|q(x)| \leq \alpha} \int_{\Omega} \lambda \operatorname{div} q \, dx \; = \; \alpha \int_{\Omega} |\nabla \lambda| \, dx \,. \tag{3.23}$$

Insert (3.22) and (3.23) in the primal-dual model (3.18), and we then end up with the primal model (3.11), together the constraints $\lambda_{i-1}(x) \geq \lambda_i(x)$ for all $x \in \Omega$ and $i = 1, ..., n$. If these constraints on optimal λ were not met, the primal-dual energy would be infinite, contradicting the existence of at least one saddle point. Prop. 3.1 is therefore proved.

3.3. Exact and Global Optimums. The functions $\lambda_i(x)$, $i = 1, \ldots, n-1$, of the continuous min-cut model (3.11) are relaxed to take values in the convex set $[0, 1]$, which is in contrast to the binary constraints of the nonconvex formulation (3.10). Both the max-flow problem (3.16) and relaxed problem (3.11) are convex optimization problems which can be solved globally. Furthermore, the following proposition shows that the *continuous min-cut model* (3.11) can be applied to produce global and exact optimums of its original nonconvex counterpart (3.10) through simple thresholdings.

PROPOSITION 3.2. *Let* $(p^*, q^*; \lambda^*)$ *be any optimal primal-dual pair of* (3.17). *Let* $\{t_i\}_{i=1}^{n-1}$ *be a sequence such that* $0 < t_1 \leq t_2 \leq \ldots \leq t_{n-1} \leq 1$, *let* $\lambda_i^{t_i}(x)$ *denote the function of* t_i *upper level set of* λ_i^*, *i.e.*

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

and define the level sets

$$S_i^{t_i} = \{x : \lambda_i^* \geq t_i\}. \tag{3.24}$$

If $t_1 = t_2 = \ldots = t_n$, *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* (3.16), *i.e.*

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

In general, if for every k *such that* $t_{k-1} < t_k$

$$\lambda_{k-1}^{t_{k-1}}(x) \neq \lambda_k^{t_k}(x), \quad a.e.\ x \in S_{k-1}^{t_{k-1}} \backslash S_k^{t_k}, \tag{3.25}$$

then $\lambda_i^{t_i}(x)$, $i = 1, \ldots, n-1$, *is a global binary optimum of* (3.10).

Proof. Since p_i^*, $i = 1, \ldots, n$ and q_i^*, λ_i^*, $i = 1, \ldots, n-1$ is a global optimum of the primal-dual problem (3.17), then p_i^*, q_i^* optimize the dual problem (3.16) 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

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

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

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

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

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

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 $k \in \{1, \ldots, n-1\}$ and start by considering $k = 1$. By the formula (3.22), it is easy to see that

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

13

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 (3.16) 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 $E^{k+1} = E^P(p^*)$.

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

Assume first $t_{k-1} = t_k$. By the definition (3.24) it follows that $\lambda_{k-1}(x) - \lambda_k(x) > t_{k-1} - t_k = 0$ for all $x \in S_{k-1}^{t_{k-1}} \setminus S_k^{t_k}$. Therefore, by formula (3.22), for any point $x \in S_{k-1}^{t_{k-1}} \setminus S_k^{t_k}$ we must have $p_k^*(x) = \rho(\ell_k, x)$. Combining this with the fact that

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

the above expression can be written

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

If $t_{k-1} > t_k$, the same argument shows the above equality provided (3.25) holds. Hence, we can conclude that also $E^{n-1} = E^P(p^*)$. By noting from (3.22) 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 (3.16) can be written

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

14

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 (3.16) and the min-cut 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). $\square$

This proposition establishes a strong primal-dual relationship between the max-flow problem (3.16) and the original non-convex problem (3.10). By solving the max-flow problem (3.16), a large set of non-unique optimums to the original problem can be obtained by thresholding each λ_i^*. In Section 6, such a max-flow algorithm is designed which can be used to compute solutions of (3.10) very efficiently.

4. Extension to Continuous Labelings. The above discussions on labeling with n linearly ordered labels can be further extended to the case where the feasible label values are ranged in a continuous interval $[\ell_{\min}, \ell_{\max}]$, i.e. the total number n of labels goes to infinity. We address such a continuous labeling problem by a natural extension of the continuous max-flow model (3.16). To this end, we first propose a novel max-flow model, then derive its equivalent min-cut formulation. Finally, we compare the proposed models to Pock et. al. [53] in details.

4.1. Max-Flow Model. In the continuous limit, as the number of labels goes to infinity, the max-flow problem (3.16) with the flow constraints (3.13)-(3.15) turns into

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

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

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

where $\ell \in [\ell_{\min}, \ell_{\max}]$ is the set of feasible continuous-valued labels. The flow functions p and q 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 (4.3). The primal-dual model can then be written

$$\sup_{p,q} \inf_\lambda \int_\Omega p(\ell_{\min}, x)\, dx + \int_{\ell_{\min}}^{\ell_{\max}} \int_\Omega \big\{ \operatorname{div}_x q(\ell, x) - \partial_\ell\, p(\ell, x) \big\}\, \lambda(\ell, x)\, dx\, d\ell \tag{4.4}$$

subject to (4.2).

By using integration by parts in ℓ, the above formulation can be rearranged as

$$\sup_{p,q} \inf_\lambda \int_{\ell_{\min}}^{\ell_{\max}} \int_\Omega \big\{ \alpha\, |\nabla_x \lambda| + p(\ell, x) \partial_\ell \lambda(\ell, x) \big\}\, dx\, d\ell$$

$$+ \int_\Omega (1 - \lambda(\ell_{\min}, x)) p(\ell_{\min}, x) + \lambda(\ell_{\max}, x) p(\ell_{\max}, x)\, dx. \tag{4.5}$$

subject to (4.2).

15

Consequently, by maximizing w.r.t the flow functions p and q, we obtain the continuous min-cut model

$$\min_{\lambda(\ell,x)\in[0,1]} \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} \left\{ \alpha \left| \nabla_x \lambda \right| + \rho(\ell,x) \partial_\ell \lambda(\ell,x) \right\} dx d\ell \qquad (4.6)$$

subject to

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

The leftmost constraint in (4.7) forces the function $\lambda(\ell,x)$ to be monotonically nonincreasing in ℓ. It corresponds to the constraint (3.4) for discrete label values.

Observe that by imposing infinite capacities on the source and sink edges, i.e. $p(\ell_{\min},x)$ and $p(\ell_{\max},x)$ unbounded above, the constraints would instead become

$$\partial_\ell \lambda(\ell,x) \leq 0, \quad \lambda(\ell_{\min},x) = 1, \quad \lambda(\ell_{\max},x) = 0, \quad \forall x \in \Omega, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}].$$
$$(4.8)$$

Both (4.7) and (4.8) are equivalent in the continuous setting.

In analogue with (3.6), 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 contrast, Pock et al [53] gave a different continuous formulation of Ishikawa's construction, as the minimization problem over a binary function in $[\ell_{\min}, \ell_{\max}] \times \Omega$

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

subject to

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

In order to solve this non-convex binary problem, the relaxation approach from [18] was adopted in [53]. Minimization was instead carried out over the convex set $\lambda(x,\ell) \in [0,1]$, then a globally optimal binary function could be obtained by thresholding the result at any level in $[0,1]$.

Some differences can be observed between our min-cut formulation (4.6), (4.7) and the formulation (4.9), (4.10). 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 in the term $\rho(\ell,x) \left| \partial_\ell \lambda(\ell,x) \right|$ forces this constraint to hold. Observe that if $\rho(\ell,x) < 0$ is negative, (4.9) is nonconvex and cannot be solved globally, which is in contrast to our formulation (4.6). In a more recent technical report of Pock et. al. [52], a more strict mathematical 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 $\lim_{\ell \mapsto -\infty} \lambda(\ell,x) = 1$. Our simple derivations show that the problem can also be formulated as the convex max-flow problem (4.1) or min-cut problem (4.6), (4.7).

16

By applying the dual formulation of total variation (3.23), the relaxed version of (4.9) can also be written as a primal-dual saddle point problem

$$\min_{\lambda} \sup_{q,p} \int_{\Omega} \lambda \operatorname{div} q_x(x, \ell) + \partial_{\ell} p \lambda \, dx \tag{4.11}$$

subject to

$$\lambda(\ell, x) \in [0, 1], \quad \lambda(\ell_{\min}, x) = 1, \quad \lambda(\ell_{\max}, x) = 0, \forall x \in \Omega, \ \ell \in [\ell_{\min}, \ell_{\max}], \tag{4.12}$$

$$|q(x, \ell)| \leq \alpha, \ |p(x, \ell)| \leq \rho(x, \ell), \quad \forall (x, \ell) \in \Omega \times [\ell_{\min}, \ell_{\max}] \ \}. \tag{4.13}$$

The most important difference to our primal-dual formulation (4.4) is that λ is unconstrained in our formulation (4.4). This allows to build up a very efficient algorithm in the augmented lagrangian framework.

We eventually stick to a finite label value set in practice. It correctly leads to an approximation of a continuously valued $u(x)$, since we always have that $\lambda(\ell, x)$ is monotonically non-increasing in ℓ by (4.7), i.e. for $x \in \Omega$

$$\lambda(\ell_{\min}, x) \geq \ldots \geq \lambda(\ell_{\max}, x).$$

In practice, the model of [53, 52] also needs to be discretized to find a numerical solution. After discretization, the label space of course also becomes discrete in (4.9). A fast primal-dual algorithm was proposed in [52]. It consists of optimizing (4.11) by taking ascent steps in the dual variables and descent steps in the primal variables followed by projections of all the variables onto the nearest points of the feasible sets iteratively until convergence. Let λ^h, q^h and p^h denote discrete counterparts of λ, q and p, the algorithm consists of choosing two time steps σ, τ and solving for $k = 1, \ldots$

$$(\lambda^h)^{k+1} = \Pi_{\mathcal{L}}((\lambda^h)^k + \sigma \nabla^h((\bar{q}^h, \bar{p}^h)^k) \tag{4.14}$$

$$(q^h, p^h)^{k+1} = \Pi_{\mathcal{C}_{\alpha}}((q^h, p^h)^k - \tau(\operatorname{div}^h(\lambda^h)^{k+1})) \tag{4.15}$$

$$(\bar{q}^h, \bar{p}^h)^{k+1} = 2(q^h, p^h)^{k+1} - (q^h, p^h)^k \tag{4.16}$$

where $\mathcal{L} = \{\lambda^h$ which satisfies (4.12)$\}$ and $C_{\alpha} = \{(q, p)$ which satisfies (4.13)$\}$ and $\Pi_{\mathcal{L}}, \Pi_{\mathcal{C}_{\alpha}}$ are projections onto $\mathcal{L}$ and $\mathcal{C}_{\alpha}$ respectively. The algorithm we present in the Section 6 is instead based on our new max-flow formulation (4.1) and is shown to outperform the fast primal-dual algorithm proposed in [52] in experiments.

5. Convex relaxation of Pott's model. A convex relaxation of Pott's model was presented in [50, 15] based on the formulation (3.10). By writing (3.10) with dual variables we obtain

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

subject to

$$\lambda_i \in \{0, 1\}, \quad i = 1, \ldots, n, \tag{5.2}$$

$$0 = \lambda_n \leq \ldots \leq \lambda_1 \leq \lambda_0 = 1, \tag{5.3}$$

17

and

$$|q_i| \leq \alpha, \quad i = 1, ..., n \tag{5.4}$$

As stated in Section 1.2, (5.1) can be used to partition the image domain into n regions $\{\Omega_i\}_{i=1}^n$ by $\Omega_i = \{x \in \Omega \text{ s.t. } \lambda_{i-1}(x) - \lambda_i(x) = 1\}$. The regularization term in (5.1) does not correspond to the length term of the Pott's model due to the linear dependence on the size of the jumps. In [50, 15] it was observed that the Pott's model can be written as (5.1) by constraining the dual variables to a smaller convex set

$$q \in C^P = \{q_i \in (C^\infty(\Omega))^m, \ i = 1, ..., n,$$

$$\text{s.t. } \{\forall x, \in \Omega \mid \sum_{i=i_1}^{i_2} q_i(x)| \leq \alpha \ \forall (i_1, i_2) \text{ s.t. } 1 \leq i_1 \leq i_2 \leq n \} \}. \tag{5.5}$$

The problem (5.1) subject to (5.3), (5.2) and (5.5) is non-convex due to the binary constraints (5.2). A convex relaxation was formulated by instead optimizing over $\lambda_i(x) \in [0, 1]$, $i = 1, ..., n$. If the solution $\{\lambda_i\}_{i=1}^n$, is binary everywhere, it is also optimal to the Pott's problem (5.3). It was observed in [50, 15] that this was most often the case. The same fast primal-dual algorithm was proposed for solving the problem, except for two differences:
- The primal variables also have to be projected onto the set (5.3) every iteration, in addition to the set $\lambda_i(x) \in [0, 1]$.
- The dual variables have to be projected onto the set (5.5) after the ascent step every iteration. No exact closed form solution exist for such a complex projection, therefore an iterative algorithm must be applied, which slows down convergence (Dyjkstra's algorithm was suggested).

We can formulate a "max-flow" dual problem of this relaxation by instead constraining the flow field to the set (5.5). Consider the problem of maximizing (3.16) subject to (3.14),(3.15) and (5.5). By following the same steps as in Section 3, we obtain the primal-dual model

$$\inf_\lambda \sup_{p,q} \int_\Omega \{p_1 + \sum_{i=1}^{n-1} \lambda_i (\text{div } q_i - p_i + p_{i+1})\} \, dx \tag{5.6}$$

subject to

$$p_i(x) \leq \rho(\ell_i, x), \ i = 1 \ldots n, \quad q \in C^P . \tag{5.7}$$

Following the same arguments as in Section 3.2.3, by maximizing (5.6) for p we obtain (5.1) subject to (5.5), (5.3) and $\lambda_i \in [0, 1]$, $i = 1, ..., n$. Observe that all constraints on λ are handled implicitly in (5.6), (5.7), including the order constraint (5.3). The algorithm in the next section also avoids the iterative inexact projections of the flow fields onto C^P.

6. Max-Flow Algorithms.

6.1. Multiplier-Based Max-Flow Algorithm. As stated in the previous section, the energy formulation (3.17) is just the lagrangian functional of (3.16) and λ_i, $i = 1, \ldots, n - 1$, are the multiplier functions. To this end, we define its respective augmented lagrangian functional as

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

$$\tag{6.3}$$

18

Algorithm 1 Multiplier-Based Maximal-Flow Algorithm I

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

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

$$q_i^{k+1} := \arg \max_{\|q\|_\infty \leq \alpha} L_c(p^k, q, \lambda^k).$$

$$:= \arg \max_{\|q\|_\infty \leq \alpha} -\frac{c}{2} \left\| \operatorname{div} q_i(x) + p_{i+1}^k - p_i^k - \lambda_i^k/c \right\|^2, , \qquad (6.1)$$

  The above formulation can either be solved iteratively by Chambolle's projection algorithm [16], or approximately in one step by (6.4).

- Optimize p_1 by fixing other variables

$$p_1^{k+1} := \arg \max_{p_1(x) \leq \rho(\ell_1, x)} L_c(p_1, p_2^k, \ldots, p_n^k, q^{k+1}, \lambda^k)$$

$$:= \arg \max_{p_1(x) \leq \rho(\ell_1, x)} \int_\Omega p_1 \, dx - \frac{c}{2} \left\| p_1 - (p_2^k + \operatorname{div} q_1^{k+1}) + \lambda_1^k/c \right\|^2$$

  Optimization of p_1 can be easily computed at each $x \in \Omega$ pointwise;

- Optimize p_i, $i = 2, \ldots, n-1$ by fixing other variables

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

$$:= \arg \max_{p_i(x) \leq \rho(\ell_i, x)} -\frac{c}{2} \left\| p_i + \operatorname{div} q_{i-1}^{k+1} - p_{i-1}^{k+1} - \lambda_{i-1}^k/c \right\|^2$$

$$-\frac{c}{2} \left\| p_i - (p_{i+1}^k + \operatorname{div} q_i^{k+1}) + \lambda_i^k/c \right\|^2$$

  which can also be easily computed at each $x \in \Omega$ pointwise

- Optimize p_n by fixing other variables

$$p_n^{k+1} := \arg \max_{p_n(x) \leq \rho(\ell_n, x)} L_c(p_1^{k+1}, \ldots, p_{n-1}^{k+1}, p_n, q^{k+1}, \lambda^k)$$

$$:= \arg \max_{p_n(x) \leq \rho(\ell_1, x)} -\frac{c}{2} \left\| p_n + \operatorname{div} q_{n-1}^{k+1} - p_{n-1}^{k+1} - \lambda_{n-1}^k/c \right\|^2,$$

  p_n^{k+1} can be simply updated pointwise;

- Update multipliers λ_i, $i = 1, \ldots, n-1$, by

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

- Let $k = k + 1$ go to the $k + 1$-th iteration until converge.

---

where $c > 0$.

An algorithm can be formulated, see Algorithm 1, for the continuous maximal flow problem (3.16) based on the alternating direction method of multipliers [8], where each flow function is optimized independently and λ_i, $i = 1, \ldots, n-1$, is updated as a multiplier at each iteration. A similar algorithm was presented for two label problems in [60, 61] and showed a significantly faster convergence rate than previous state of the art work [14].

The flow functions p_i, $i = 1, \ldots, n$ are optimized one component at a time, starting from the source advancing one layer at a time to the sink. Convergence of such algorithms is

Algorithm 2 Multiplier-Based Maximal-Flow Algorithm II

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

- For each layer $i = 1, ..., n$:
  - Optimize p_i by fixing other variables

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

  - Optimize q_i, by fixing other variables

$$q_i^{k+1} := \arg \max_{\|q\|_\infty \leq \alpha} L_c((p_{i \leq j}^{k+1}, p_{i>j}^k), q_i, \lambda^k) \tag{6.2}$$

  - Optimize p_i, by fixing other variables

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

- Update multipliers λ_i, $i = 1, \ldots, n - 1$, by

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

- Let $k = k + 1$ go to the $k + 1$-th iteration until converge, set $i = 1$.

---

validated by classical optimization theories. Another variant of this algorithm, see Algorithm 2, also optimizes the flow functions q_i along with p_i one layer at a time and is more stable with respect to the penalty parameter c. This algorithm is used in our experiments.

Some interesting things can be observed about these algorithms. The constraints (3.3) on the labeling function λ are automatically satisfied, therefore λ does not need to be projected onto the feasible set. The penalty parameter c controls how fast the algorithm will convergence as converge will be faster as c increases. In our experiments c can be set quite high without altering the convergence or final result, e.g. $c = 3$. As a consequence, Algorithm 2 converges in a relatively few outer iterations, significantly less than the fast primal-dual algorithm of [52]. On the flip side our algorithm contains a more difficult subproblem (6.1) which can be solved iteratively by Chambolle's algorithm [16]. Since the previous solution is available as a good initialization, not many iterations of this algorithm is required, see the experiment section for detail.

Instead of solving the subproblem (6.1) iteratively, an inexact solution can be generated by the linearization

$$q_i^{k+1} = \Pi_\alpha \left(q_i^k + c\nabla(\operatorname{div} q_i^k - F_i^k). \right) \tag{6.4}$$

where Π_α is the projection onto the convex set $C_\alpha = \{q \,|\|q\|_\infty \leq \alpha\}$. There are extended convergence theories for such a linearization, which was studied and proved in [22] for closely related problems. However, this approximation places some more restrictions on the penalty parameter c, which must be set lower to maintain convergence. For instance $c = 0.2$ seems to be a good choice.

6.2. Algorithm for convex relaxed Pott's model. The convex relaxed Pott's model in (5) can be optimized by a small modification of Algorithm 1 or 2. The flow fields q_i, $i = 1, ..., n$ are instead optimized over the set (5.5). As stated in Section 5, no closed form solution
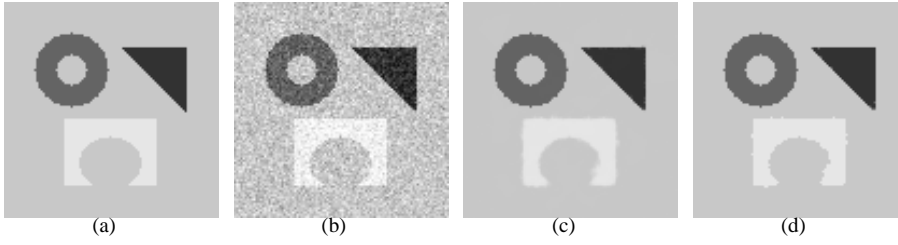
FIG. 7.1. *(a) Ground truth, (b) input, (c) Rescaled labeling function before threshold, (d) Rescaled labeling function after thresholding each λ_i at 0.5.*

exists for projecting a vector $(q_1, ..., q_n) \in \mathbb{R}^{m \times n}$ onto C^P. However, if all components but one are fixed, the projection can be computed analytically. Given a vector $\bar{q} \in C^P$, define the sets

$$C_i^P(\bar{q}) = \{q_i \in (C^\infty(\Omega))^m \text{ s.t. } (\bar{q}_1, ..., \bar{q}_{i-1}, q_i, \bar{q}_{i+1}, ..., \bar{q}_n) \in C^P\}, \quad i = 1, ..., n \quad (6.5)$$

This set $C_i^P(\bar{q})$ is just an intersection of spheres in $\mathbb{R}^m$ of radius α and different centers. For any $q_i \in \mathbb{R}^m$ the projection onto $C_i^P(\bar{q})$ can be computed analytically, the details are given in appendix A. Applied to the convex relaxed problem (5.6), Algorithm 1 and 2 does not change except the feasible set in the subproblem (6.1) of Alg 1 and (6.2) of Alg 2 are replaced by $C_i^P(q_{j<i}^{k+1}, q_{j\geq i}^k)$, i.e. subproblem (6.1) and (6.2) become

$$q_i^{k+1} := \arg \max_{q_i \in C_i^P((q_{j<i}^{k+1}, q_{j\geq i}^k))} L_c(p^k, (q_{j<i}^{k+1}, q_i, q_{j>i}^k), \lambda^k). \quad (6.6)$$

To optimize (6.6) one may either apply the linearization (6.4) to q_i followed by a projection onto $C_i^P((q_{j<i}^{k+1}, q_{j\geq i}^k))$, or an iterative ascent/project algorithm in q_i with projections onto $C_i^P((q_{j<i}^{k+1}, q_{j\geq i}^k))$.

The projection onto $C_i^P((q_{j<i}^{k+1}, q_{j\geq i}^k))$ can be computed exactly, but of course costs more floating point evaluations than projections onto the simpler set C_α. The algorithm in the appendix has a worst case complexity of $O(n^2)$ to find the projection exactly (in the worst case, the euclidian distance between n^2 points must be computed for 2D images). The algorithm is applied for each of the n labels, hence the total complexity of computing projections for each iteration is $O(n^3)$.

The primal-dual algorithm needs to project the vector $(q_1, ..., q_n) \in \mathbb{R}^{m \times n}$ onto C^P during each iteration. Since no closed form solution exists, it was suggested in [15] to apply Dyjkstra's iterative algorithm to compute the projection approximately. Each iteration of this algorithm applied to the problem has a complexity of $O(n^3)$. As far as we can see, one iteration of Dyjkstra's algorithm requires around the same number of floating point evaluations as our exact projection algorithm summed over each label.

TABLE 7.1
Percentage of misclassified pixels for stereo example

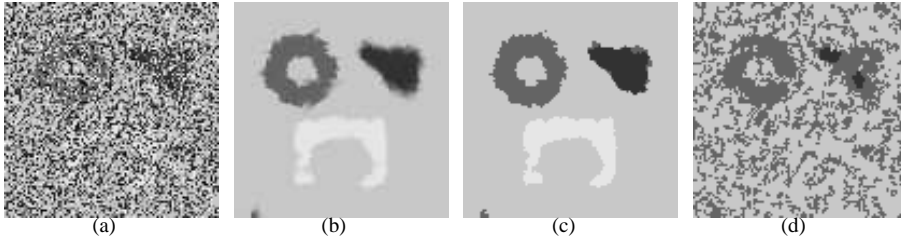| Graph Cut 4-neighbors	Graph Cut 8-neighbors	Pock et. al.	Proposed T_1	Proposed T_2
30.74	30.5	30.20	29.95	27.59

FIG. 7.2. *(a) Input image damaged by impulse noise; (b) reconstructed labeling function with non-convex data term (7.3) before threshold, (c) labeling function after thresholding each λ_i at 0.5, (d) reconstructed labeling function with convex data term (7.1) and $\beta = 1$.*
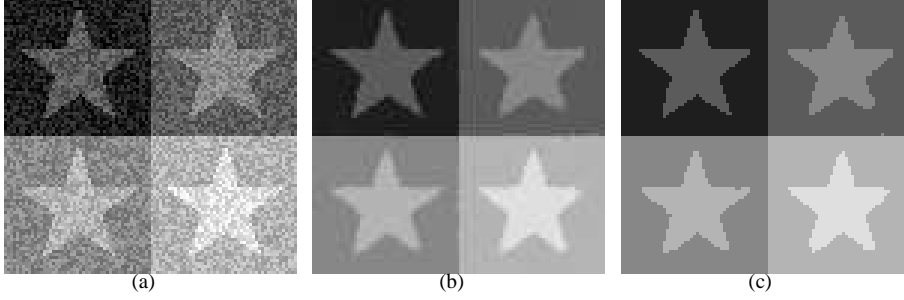

FIG. 7.3. *(a) Input, (b) Labeling function before threshold (c) Labeling function after thresholding each λ_i at 0.5.*

7. Numerical Experiments.

In this work, we focus on applications of the model (1.3) and (1.1) to image segmentation and stereo reconstruction. Comparisons are made to the discrete approach [33] and the approach proposed by Pock et. al. [53].

7.1. 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

$$\rho(i, x) = |I(x) - c_i|^\beta, \quad i = 1, ..., n \tag{7.1}$$

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 7.1, 7.4 and 7.3 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.

$$u = c_1 + \sum_{i=1}^{n-1} (c_{i+1} - c_i)\lambda_i^*. \tag{7.2}$$

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
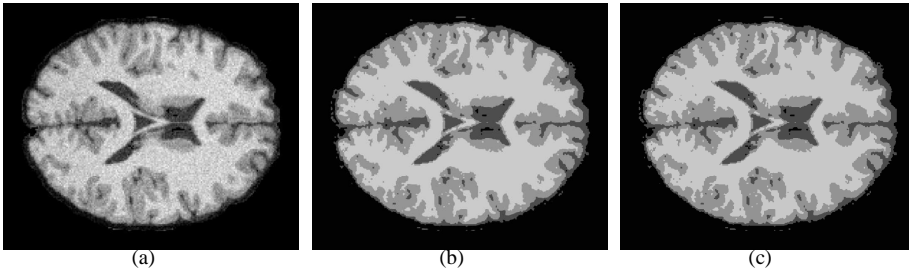
FIG. 7.4. *(a) Input, (b) Labeling function before threshold (c) Labeling function after thresholding each λ_i at 0.5.*
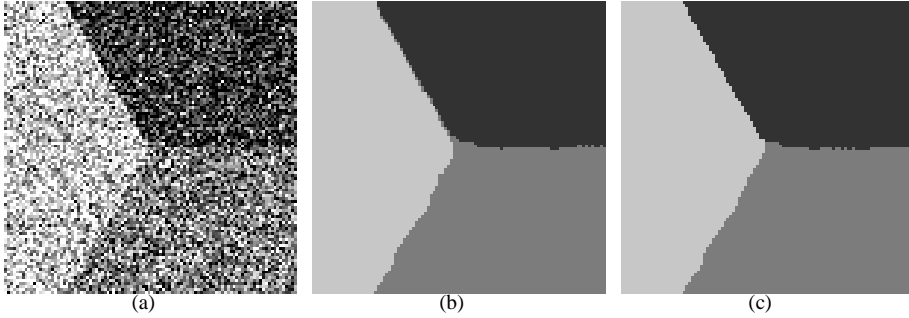

FIG. 7.5. *(a) Input, (b) segmentation with total variation regularized model* (1.3) *(after threshold), (c) segmentation with convex relaxed Pott's model (after threshold). The total variation regularized model results in misclassifications along the boundary between region 1 (darkest) and region 3 (brightest) and does not reconstruct the triple junction properly.*

each $\lambda_i^*(x)$ according to Prop. 3.2. We also demonstrate image segmentation with a non-convex data term. The ground truth image from Figure 7.1 (a) has been damaged by impulse noise in Figure 7.2 (a). More specifically, 70% of the pixels have been randomly selected 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 7.2 (d) where we have selected (7.1) with $\beta = 1$. Instead the following non-convex data term can be used

$$\rho(i, x) := \begin{cases} 0, & \text{if } i = \text{argmin}_k |I(x) - c_k| \\ 1, & \text{else} \end{cases}.$$ (7.3)

This non-convex problem can be solved globally by our method, the result is shown in Figure 7.2 (b) before threshold and 7.2 (c) after thresholds.

We next apply our algorithm for the convex relaxed Pott's model of [50] from section 6.2. The image in Figure (7.5) (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.

7.2. 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
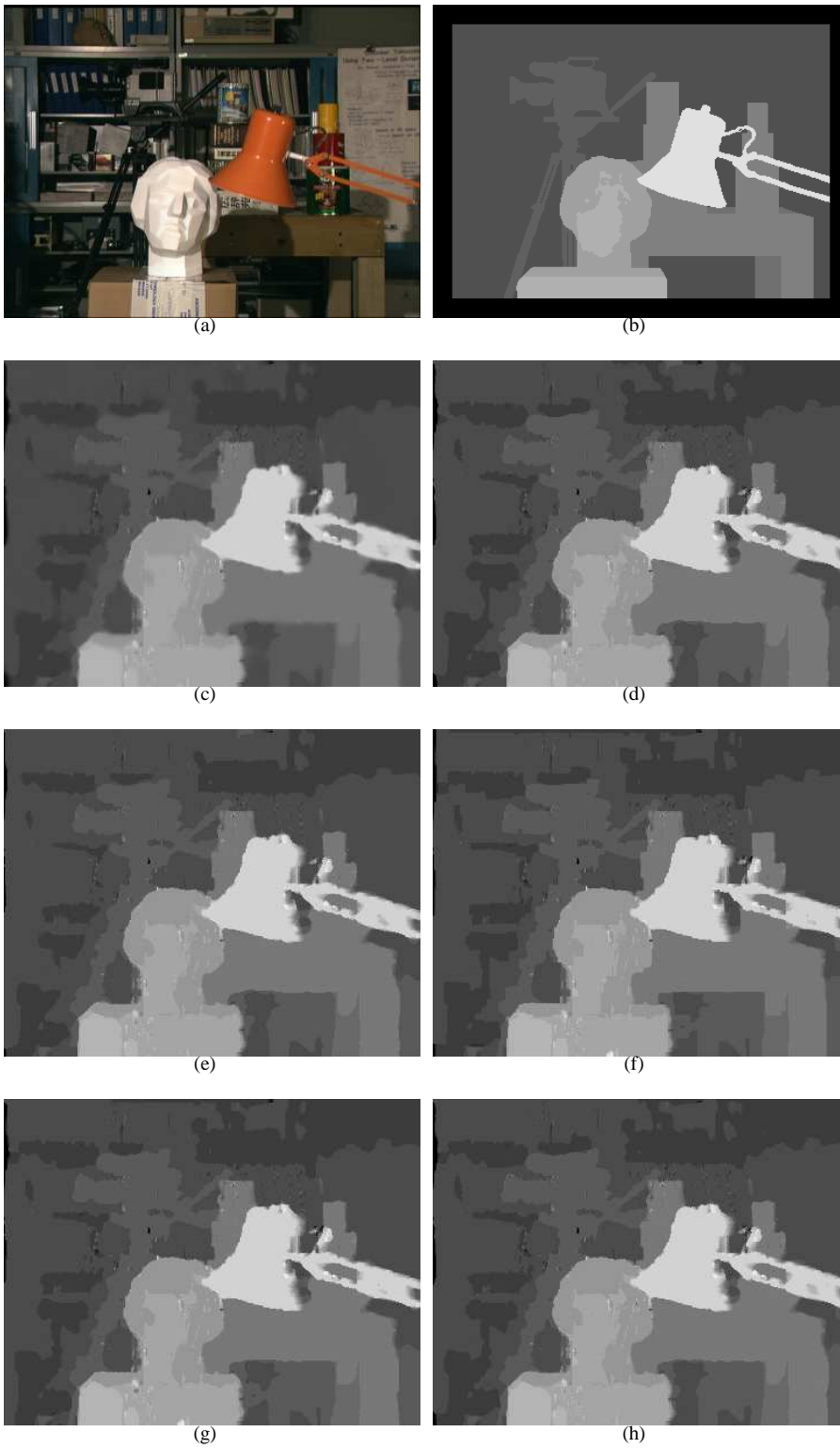
FIG. 7.6. *(a) Left input image, (b) ground truth, (c) non-integer solution $u = \sum_{i=1}^{n-1} \lambda_i^*$, (d) Integer valued solution after threshold T_1, (e) Integer valued solution after threshold T_2, (f) Graph cut 4 neighbors, (g) Graph cut 8 neighbors, (h) Pock et. al.*

(a)   (b)

FIG. 7.7. *The solution λ^* may not be binary. (a) depicts λ_8^* and figure (b) depicts λ_9^* from the stereo example in Figure 7.6*

	Energy precision $\varepsilon < 10^{-3}$			Energy precision $\varepsilon < 10^{-4}$		
	Fast primal-dual [52]	Proposed 1	Proposed 2	Fast primal-dual [52]	Proposed 1	Proposed 2
Brain	280	50 ($\times$ 5)	110	430	65 ($\times$ 5)	280
Figure 7.1	295	35 ($\times$ 5)	115	640	65 ($\times$ 5)	290
Stereo	4055	550 ($\times$ 5)	1070	14305	920 ($\times$ 5)	3905

TABLE 7.2

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 (6.4).

the data term of (3.1)

$$\rho(u, x) = \sum_{j=1}^{3} |I_L^j(x) - I_R^j(x + (u, 0)^T)|. \qquad (7.4)$$

Here $I^j(x)$ denotes the $j - th$ component of the color vector $I(x)$. The above data term (7.4) 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 7.6 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 (3.17) 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. As stated in Prop 3.2, a sequence of optimal integer valued solutions can be obtained by thresholding each λ_i^* at different levels t_i with $0 < t_1 \leq ... \leq t_n \leq 1$, provided (3.25) holds. In Figure 7.6(c), we depict such a non-binary solution by plotting $u = \sum_{i=1}^{n-1} \lambda_i^*$. Observe that λ_i^* is not binary at locations where a unique solution is not expected, such as the upper right corner and underneath the table. At such locations the data terms are weak. By varying the threshold level, different integer valued solutions can be obtained. For instance, we have chosen $t_i = 0.5$ for all i in Figure 7.6(d), which is called threshold T_1. In Figure 7.6(e) we have chosen $t_i = 0.15$ for $i = 1, ..., 5$ and $t_i = 0.5$ for $i = 6, ..., 16$, called threshold T_2 (one can check (3.25) is

25

valid for $k = 6$). This last threshold option allows to remove some of the misclassifications in the background. In contrast, the approach of Pock et. al. obtains the labeling function by one single threshold of the optimal higher dimensional indicator function. Their result with threshold level 0.5 (which is optimal in terms of misclassified pixels) is depicted in Figure 7.6(h). As we see in the figures and in Table 7.1, our approach produces a solution of less misclassified pixels compared to the ground truth, Figure 7.6(b). Let us reiterate that both approaches are global and converge to the same minimal energy, but our approach allows for a larger set of non-unique solutions.

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 7.6(f) and (g) with 4 and 8 neighbors respectively. As we see, such solutions suffer from metrication artifacts because of the discrete grid bias.

7.3. Evaluation of convergence. Iteration counts for all experiments are presented in Table 7.2. The two variants of Algorithm 2 are evaluated against the fast primal-dual method of Pock et. al. [52]. The relative energy precision at iteration i is given by

$$\varepsilon = \frac{E^i - E^*}{E^*}, \tag{7.5}$$

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 [32, 31]. In order to make the comparison as accurate as possible, the fast 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 (6.1) iteratively by Chambolle's algorithm [16]. 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 each time. Increasing the number of inner iterations beyond 5 did not seem to have any impact on the convergence rate in our experience.

The fast 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 fast primal-dual algorithm [52] failed to ever reach an energy precision of 10^{-5} or lower within our predetermined number of maximum iterations (20000). 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 (6.1) by the linearization (6.4) and hence avoids the inner iterations. However, the penalty parameter c must be set lower to have convergence, hence more outer iterations are required. Overall it converges a little faster than the first variant and outperforms the fast primal-dual algorithm [52] for all the experiments.

Regarding the algorithm for Pott's model in Section 6.2, to reach an energy precision of 10^{-4} on the image in Figure 7.7, 130 outer iterations were required (with 5 inner iterations to solve subproblem (6.6)). In comparison, 120 outer iterations (with 5 inner iterations) were

26

required for the total variational regularized model (1.3), indicating the number of iterations in the convex relaxed Pott's model does not significantly increase. This is in contrast to the primal-dual algorithm of [50], where for each iteration the dual variable q must be projected onto C^P by an iterative algorithm (e.g. Dijkstra's algorithm), which is the bottle neck of the overall algorithm.

Comparison to discrete graph cut [11] is more complex. Our algorithms are implemented in matlab, in contrast to the optimized c++ discrete max-flow implementation of [11]. 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.

8. 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 can verified by standard optimization theories. Experiments showed that the algorithms outperform the existing approach both in terms of convergence rate and reliability. 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 new thresholding scheme could be derived for producing a larger set of exact and global optimums; experiments showed that the max-flow based algorithms converged significantly faster and more accurately than the fast 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 and compare extensively with [50, 51].

During the completion of this work, we discovered another inexact one step scheme for solving the subproblem 6.1, which allows for larger step sizes, resulting in even faster convergence. We plan to investigate this further and release another paper on fast implementations and more comparisons in the future. We also became aware of a simultaneous work [43] which gives another algorithm for minimizing the energy in the convex formulation of [52]. Comparison with this work 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 $(k,j) \in I = \{(k,j) \text{ s.t. } 1 \le k < i < j \le n\}$ 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
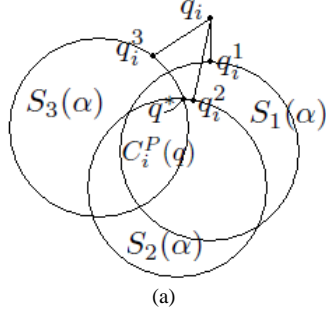
(a)

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 $S_3(\alpha)$.*

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. $\square$

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*

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

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). 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 $q^* = \Pi_{C_i^P(\bar{q})} q_i$. $\square$

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

28

Algorithm 3 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
  $(K, J) = \arg\max_{(k,j)\in I}|q_k^j - q_i|, (K', J') = \arg\max_{(k,j)\in I\setminus(K,J)}|q_k^j - q_i|$
  $q^* = \arg\min_{q\in\partial C_K^J(\alpha)\cap\partial C_{K'}^{J'}(\alpha),\ (k,j)\neq(K,J)\in I}|q - q_i|$.

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

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.

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 compromised 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. Technical report CAM09-75, UCLA, CAM, September 2009.

[6] Egil Bae and Xue-Cheng Tai. Efficient global minimization for the multiphase Chan-Vese model of image segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 28–41, 2009.

[7] 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.

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

[9] Endre Boros and Peter L. Hammer. Pseudo-boolean optimization. *Discrete Appl. Math.*, 123(1-3):155–225, 2002.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

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

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

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

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

[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] 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.

[31] 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.

[32] 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.

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

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

[35] 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.

[36] 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.

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

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

[39] Vladimir Kolmogorov and Carsten Rother. Minimizing nonsubmodular functions with graph cuts-a review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1274–1279, 2007.

[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] L. A. Vese and T. F. Chan. A new multiphase level set framework for image segmentation via the mumford and shah model. *International Journal of Computer Vision*, 50:271–293, 2002.

[59] 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.

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

[61] 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.

[62] Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov. A continuous max-flow approach to potts model. In *ECCV*, 2010.

[63] 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.