

A Fast Continuous Max-Flow Approach to Non-convex Multi-labeling Problems

Egil Bae¹(✉), Jing Yuan², Xue-Cheng Tai³, and Yuri Boykov²

¹ Department of Mathematics, University of California, Los Angeles, USA
ebae@math.ucla.edu

² Computer Science Department, Middlesex College,
University of Western Ontario, London, ON, Canada
cn.yuanjing@gmail.com, yuri@csd.uwo.ca

³ Department of Mathematics, University of Bergen, Bergen, Norway
tai@math.uib.no

Abstract. This paper studies continuous image labeling problems with an arbitrary data term and a total variation regularizer, where the labels are constrained to a finite set of real numbers. Inspired by Ishikawa's multi-layered graph construction for the same labeling problem over a discrete image domain, we propose a novel continuous max-flow model and build up its duality to a convex relaxed formulation of image labeling under a new variational perspective. Via such continuous max-flow formulations, we show that exact and global optimizers can be obtained to the original non-convex labeling problem. We also extend the studies to problems with continuous-valued labels and introduce a new theory to this problem. Finally, we show the proposed continuous max-flow models directly lead to new fast flow-maximization algorithmic schemes which outperform previous approaches in terms of efficiency. Such continuous max-flow based algorithms can be validated by convex optimization theories and accelerated by modern parallel computational hardware.

1 Introduction

Many practical problems in image processing and computer vision can be modeled as multilabel problems, where the task is to optimally assign the unknown variable l , chosen from some finite set $\{l_1, \dots, l_n\}$, at each point of the image domain Ω . It has become an important paradigm to formulate such labeling problems as the optimization of an energy function/functional which mathematically encodes all the information needed for the specified imaging and vision task. Such optimization problems can be formulated by either regarding the image domain as discrete or continuous.

In the spatially discrete setting, graph cut has become one of the most important and efficient techniques to tackle such problems, by computing max-flow or min-cut on appropriately constructed graphs. Applications of max-flow/min-cut in computer vision range from image segmentation or labeling [1, 2], stereo [3, 4], 3D reconstruction [5] etc. Unfortunately, most minimization problems

involving more than two labels are NP-hard, therefore only approximate algorithms are available [1, 4]. However, for a particular set of multilabeling problems with convex interaction penalty, Ishikawa [6] showed that exact solutions can be computed by max-flow and min-cut. Such energy functions are important in e.g. stereo reconstruction. Despite the efficiencies of graph-based methods, their computation results are often comparatively rough and biased by the discrete graph setting, i.e. metrication errors occur. Reducing such visual artifacts requires either considering more neighbour nodes, which increases memory burden largely, or applying more complex schemes such as high-order potentials [7].

Recently, the variational approach has become more and more popular for obtaining optimal labelings in the spatially continuous setting, where the problem is formulated as the minimization of a continuous energy functional. In contrast to the graph-based model, there are many advantages of the variational approach: the variational model perfectly avoids metrication errors due to its crucial rotation invariance; moreover, its reduced continuous numerical scheme is reliable, tractable, and can be easily implemented and accelerated in many different ways, e.g. parallel, multigrid or GPU hardwares; last but not least, the continuous models require far less memory in computation.

The application of variational methods to optimal labelings is often to relax the combinatorial constraints to a proper convex set. It leads to a constrained convex minimization problem such that global and exact optimums, in some special cases, are available. For example, Chan et al. [8] showed that global and exact binary optimizers can be obtained by thresholding the computation result of the convex relaxed model; therefore, a sequence of so-called continuous min-cuts can be obtained. [9, 10] generalized Ishikawa's work [6] to the spatially 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, i.e. a spatially continuous min-cut. Such a lifting approach is related to earlier mathematical theories of calibrations and Cartesian currents [11, 12]. Optimal labeling functions could be obtained by applying the result of Chan et al. in the higher dimensional space, i.e. first solve the relaxed binary problem and then threshold the result.

Motivations and Contributions

For discrete graphs, it is well known that the minimum cut problem is dual to the maximum flow problem by the *max-flow and min-cut theorem* [13]. Actually, the fastest graph cut algorithms are based on maximizing flow instead of computing the min-cut directly, e.g. the Ford-Fulkerson algorithm [14] and the push-relabel algorithm [15]. The minimal 'cut' is finally recovered along edges with 'saturated' flows, i.e. cuts appear at the flow-bottlenecked edges [4, 16]. In contrast, max-flow models and algorithms in the spatially continuous setting have been much less studied. Some work has appeared that deal with partitioning problems involving two regions: Strang [17] was the first to formulate max-flow and min-cut problems over a continuous domain; In [18], edge based max-flow and min-cut was formulated in which certain interior and exterior points must be

specified in advance; Yuan et al. [19,20] 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. Max-flow and min-cut interpretations of recent convex relaxations for Potts model have been made in [21]. However in these cases there is generally a duality gap and the original problems can only be solved approximately.

To our knowledge, this is the first work to address continuous max-flow and min-cut duality for problems where the labeling function can take several discrete values. Motivated by Ishikawa [6] and Yuan et al. [19], we interpret the problem as a continuous min-cut problem over a mixed continuous/discrete domain and build up a novel continuous max-flow model in analogy with Ishikawa’s discrete graph construction. The max-flow model can be used to produce global solutions of the original non-convex problem with discrete label values. In particular, it is shown that the max-flow model is dual to an exact convex relaxation of the original problem. Strict duality is also established between the max-flow model and the original problem, by extending the thresholding scheme of [8] from two to several regions.

A new continuous max-flow based algorithm is proposed. Its efficiency and convergence can be verified by standard convex optimization theories. The labeling function is updated as an unconstrained lagrange multiplier each iteration, and does not need to be projected back onto any feasible set. Numerical experiments show a significantly faster convergence rate than the primal-dual algorithm in Pock et al. [9,10] and later [22], especially at high precisions.

A significantly extended version of this paper is available at [23], which contains extensions to other regularizers and more experiments. This conference paper contains some novelties which are not in [23], such as the discussion on saturated/unsaturated edges in Sect. 3.5.

2 Preliminaries: Ishikawa’s Work

Ishikawa [6] studied image labeling problems over an image graph which can be generally formulated as:

$$\min_{u \in U} \sum_{v \in \mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w) \in \mathcal{N}} g(u_v - u_w), \quad (1)$$

where \mathcal{P} denotes a discrete image grid in 2-D or N-D; $\mathcal{N} \subset \mathcal{P} \times \mathcal{P}$ is a neighborhood system on \mathcal{P} ; $U = \{u : \mathcal{P} \mapsto L\}$ is the set of all feasible labeling functions, where $L = \{\ell_1, \dots, \ell_n\}$. The potential prior $g(x)$ in (1) is assumed to be convex and ρ is any bounded function, but not necessarily convex. It was shown by [6] that problems of the form (1) can be exactly optimized by finding the minimal cut over a specially constructed multi-layered graph $G = (\mathcal{V}, \mathcal{E})$, where each layer corresponds to one label.

We adopt Ishikawa’s notations [6] in this work and study the simplified graph which uses $n - 1$ layers instead of n and avoids infinite capacities on the source edges [24] (see Fig. 1 for a 1-D example). The vertex set \mathcal{V} and the edge set \mathcal{E} are defined as follows:

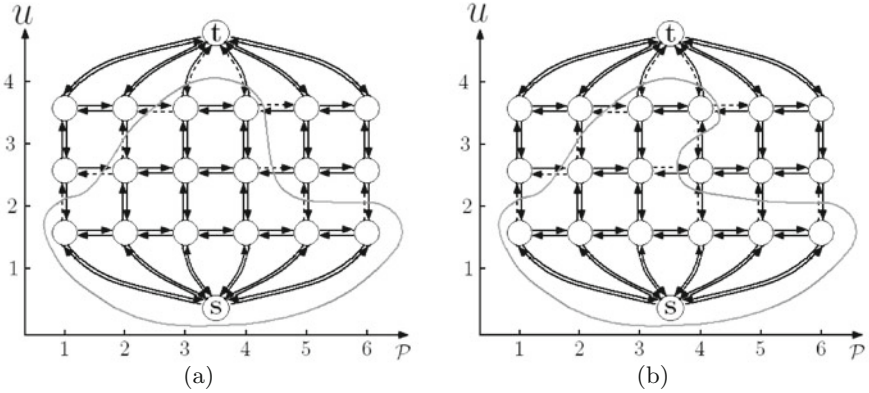


Fig. 1. 1D illustration: (a) Legal cut, (b) Illegal cut. Severed edges are depicted as dotted arrows. The gray curve visualizes the cut. Vertices interior to the curve belongs to V_s while vertices exterior to the curve belongs to V_t .

$$\mathcal{V} = \mathcal{P} \times L \cup \{s, t\} = \{u_{v,i} \mid v \in \mathcal{P}; i = 1, \dots, n-1\} \cup \{s, t\} \quad (2a)$$

$$\mathcal{E} = \mathcal{E}_D \cup \mathcal{E}_C \cup \mathcal{E}_P \quad (2b)$$

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

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

$$\mathcal{E}_D^v = (s, u_{v,1}) \cup \{(u_{v,i}, u_{v,i+1}) \mid i = 1, \dots, n-2\} \cup (u_{v,n-1}, t). \quad (3)$$

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

$$\mathcal{E}_C^v = (u_{v,1}, s) \cup \{(u_{v,i+1}, u_{v,i}) \cup (t, u_{v,n-1}) \mid i = 1, \dots, n-2\}. \quad (4)$$

- Regularization edges \mathcal{E}_R :

$$\mathcal{E}_R = \{(u_{v,i}, u_{w,j}) \mid (v, w) \in \mathcal{N}, i, j = 1, \dots, n\}. \quad (5)$$

2.1 Anisotropic Total-Variation Regularization

When a pairwise prior $g(u_v - u_w) = C(v, w) |u_v - u_w|$ is given, (1) corresponds to an anisotropic total-variation regularized image labeling problem, i.e.

$$\min_{u \in U} \sum_{v \in \mathcal{P}} \rho(u_v, v) + \alpha \sum_{(v,w) \in \mathcal{N}} C(v, w) |u_v - u_w| \quad (6)$$

which is a discrete counterpart of the total-variation regularizer.

Now we define flow configurations over the graph (2a) and (2b) such that its max-flow corresponds to the minimizer of (6):

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

$$p_1(v) \leq \rho(\ell_1, v), \quad \forall v \in \mathcal{P}; \quad (7)$$

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

$$p_i(v) \leq \rho(\ell_i, v), \quad \forall v \in \mathcal{P} \quad i = 1, \dots, n - 2 \quad (8)$$

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

$$p_n(v) \leq \rho(\ell_n, v), \quad \forall v \in \mathcal{P}; \quad (9)$$

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

$$|q_i(v, w)| \leq C(v, w); \quad (10)$$

this corresponds to the well-known anisotropic total-variation regularizer in case of a 4 nearest neighborhood system \mathcal{N} ;

- *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 $i = 1, \dots, n - 1$, i.e.

$$\left(\sum_{w:(w,v) \in \mathcal{N}} q_i(v, w) - \sum_{w:(v,w) \in \mathcal{N}} q_i(v, w) \right) - p_i(v) + p_{i+1}(v) = 0. \quad (11)$$

Since there is no lower bound on the flows (7)–(9), the flow on the penalty edges (4) can become arbitrarily large. 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. 1(b) have infinite cost and are not allowed.

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

$$\max_{p, q} \sum_{v \in \mathcal{P}} p_1(v) \quad (12)$$

subject to the flow constraints (7), (8), (9), (10) and (11).

Due to duality between the max-flow and min-cut problem [13], one can solve the max-flow problem and then extract a solution to the min-cut problem (6).

3 Multilabeling by Continuous Max-Flow and Min-Cut

Define the feasible set of functions as $U = \{u : \Omega \mapsto \{\ell_1, \dots, \ell_n\} \text{ s.t. } \int_{\Omega} |\nabla u| \leq \infty\}$, where $\ell_1 < \dots < \ell_n$ are real numbers. The continuous counterpart of (1) can be formulated as

$$\min_{u \in U} \int_{\Omega} \rho(u(x), x) dx + \int_{\Omega} C(x) |\nabla u(x)| dx, \quad (13)$$

where $\rho : \mathbb{R} \times \Omega \mapsto \mathbb{R}$ is any bounded function, not necessarily convex. The set U is a non-convex set of discrete valued labeling functions. This is in contrast to Pock et al. who considered a convex feasible set of continuous valued labeling functions. We show this problem can be regarded as a continuous min-cut problem by following the ideas of Ishikawa.

We start by rewriting (13) in terms of the upper level sets of $u \in U$

$$\lambda_i(x) = \begin{cases} 1, & \text{if } u(x) > \ell_i \\ 0, & \text{if } u(x) \leq \ell_i \end{cases}, \forall x \in \Omega \quad i = 1, \dots, n - 1.$$

Let $\lambda_0(x) = 1$ and $\lambda_n(x) = 0$, a.e. $x \in \Omega$. Clearly, we have

$$1 = \lambda_0(x) \geq \lambda_1(x) \geq \lambda_2(x) \geq \dots \geq \lambda_{n-1}(x) \geq \lambda_n(x) = 0 \text{ a.e. } x \in \Omega. \quad (14)$$

By the coarea formula, we have for any function $u \in U$ that

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

where $C_i(x) = (\ell_{i+1} - \ell_i)C(x)$, $i = 1, \dots, n - 1$. In this work, we will mostly focus on the case where $C(x) = \alpha$ is constant for simplicity.

Therefore, (13) can be equivalently rewritten as

$$\min_{\{\lambda_i\}_{i=1}^{n-1} \in \mathcal{B}} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) \rho(\ell_i, x) dx + \alpha \sum_{i=1}^{n-1} (\ell_{i+1} - \ell_i) \int_{\Omega} |\nabla \lambda_i| dx \quad (15)$$

subject to the constraint (14), where the binary constraint \mathcal{B} is defined as

$$\mathcal{B} = \{\phi : \Omega \mapsto \{0, 1\}, \text{ s.t. } \int_{\Omega} |\nabla \phi| < \infty\} \quad (16)$$

The problem (15) is obviously a nonconvex optimization problem due to the binary constraints (16).

After solving (15), the labeling function u can be recovered from λ_i by $u = \sum_{i=1}^n (\lambda_{i-1} - \lambda_i) \ell_i$.

3.1 Primal Model: Continuous Max-Flow

In this section, we build up a max-flow model in continuous settings, which simulates Ishikawa’s graph configuration. It will be shown that solutions of (15) and (13) can be obtained by exploring the dual of this maximization problem.

To this end, we place $n - 1$ image domains Ω_i , $i = 1, \dots, n - 1$ with $\Omega_i = \Omega$, layered in a sequential order between two terminals: the source s and the sink t . The source s is linked to each image pixel x of the first layer Ω_1 by an edge $e_1(x)$; the same image pixel x between two sequential image layers Ω_{i-1} and Ω_i , $i = 2, \dots, n - 1$, is linked by the edge $e_i(x)$; and the pixel x at the last layer Ω_{n-1} is also linked to the sink t by the edge $e_n(x)$. Define flow functions $p_i : \Omega \mapsto \mathbb{R}$

corresponding to each edge function e_i , $i = 1, \dots, n$. Within each image layer Ω_i , $i = 1, \dots, n-1$, the spatial flow functions are given by $q_i \in C^\infty(\Omega)^N$, where N is the dimension of the image domain.

As a generalization of the discrete constraints (7)–(11), we now give constraints on flow functions $p_i \in L^1(\Omega)$, $i = 1, \dots, n$, and $q_i \in C^\infty(\Omega)^N$, $i = 1, \dots, n-1$

$$|q_i(x)| \leq C_i(x) \quad \text{for } x \in \Omega, \quad i = 1, \dots, n-1 \quad (17)$$

$$p_i(x) \leq \rho(\ell_i, x) \quad \text{for } x \in \Omega, \quad i = 1, \dots, n \quad (18)$$

$$(\operatorname{div} q_i - p_i + p_{i+1})(x) = 0 \quad \text{for } x \in \Omega, \quad i = 1, \dots, n-1 \quad (19)$$

$$q_i \cdot n = 0 \quad \text{on } \partial\Omega, \quad i = 1, \dots, n-1. \quad (20)$$

Therefore, the continuous max-flow model, in analogue with Ishikawa’s discrete one (12), can be formulated by

$$\sup_{p,q} E^P(p) = \int_{\Omega} p_1(x) dx \quad (21)$$

subject to the constraints (17)–(20). In this work, we call (21) the *primal model*. Observe the maximization problem (21) is bounded above by $\int_{\Omega} \rho(\ell_1(x), x) dx$ due to the constraint (18).

3.2 Primal-Dual Model

By introducing multiplier functions $\lambda_i(x)$, $i = 1, \dots, n-1$, to the linear equality constraints of flow conservation (19), we have the equivalent *primal-dual model* of (21):

$$\inf_{\lambda} \sup_{p,q} E(p, q; \lambda) = \int_{\Omega} \left\{ p_1 + \sum_{i=1}^{n-1} \lambda_i (\operatorname{div} q_i - p_i + p_{i+1}) \right\} dx \quad (22)$$

subject to (17), (18) and (20).

After rearrangement, the above primal-dual formulation (22) can be equivalently written as

$$\inf_{\lambda} \sup_{p,q} E(p, q; \lambda) = \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) p_i dx + \sum_{i=1}^{n-1} \int_{\Omega} \lambda_i \operatorname{div} q_i dx \quad (23)$$

subject to (17), (18) and (20).

3.3 Dual Model: Continuous Min-Cut

Now we show that optimizing the primal-dual model (23) over all the flow functions p and q leads to the equivalent *dual model*, i.e. the continuous min-cut model:

$$\inf_{\lambda} E^D(\lambda) = \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) \rho(\ell_i, x) dx + \sum_{i=1}^{n-1} \int_{\Omega} C_i(x) |\nabla \lambda_i| dx \quad (24)$$

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

Optimization of Flow Functions. In this regard, we consider the optimization problem

$$f(v) = \sup_{w \leq C} v \cdot w, \tag{25}$$

where v , w and C are scalars. When $v < 0$, w can be arbitrarily large in order to maximize the value $v \cdot w$, i.e. $f(v) = +\infty$. Therefore, we must have $v \geq 0$ so as to make the function $f(v)$ meaningful and

$$\begin{cases} \text{if } v = 0, \text{ then } w < C \text{ and } f(v) \text{ reaches its maximum } 0 \\ \text{if } v > 0, \text{ then } w = C \text{ and } f(v) \text{ reaches its maximum } v \cdot C \end{cases} .$$

Therefore, we have

$$f(v) = \begin{cases} v \cdot C, & v \geq 0, \\ \infty & v < 0 \end{cases} . \tag{26}$$

The function $f(v)$ given in (25) provides us with a prototype to maximize the flow functions $p_i(x)$, $i = 1, \dots, n$, in the primal-dual model (23).

For each $x \in \Omega$, consider

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

In view of (26), we have

$$f_i(x) = \begin{cases} (\lambda_{i-1}(x) - \lambda_i(x)) \rho(\ell_i, x), & \lambda_{i-1}(x) \geq \lambda_i(x) \\ \infty & \lambda_{i-1}(x) < \lambda_i(x) \end{cases}, \quad i = 1, \dots, n. \tag{27}$$

On the other hand, it is well known that for any $\lambda_i \in BV(\Omega)$

$$\sup_{q_i} \int_{\Omega} \lambda_i(x) \operatorname{div} q_i(x) dx = \int_{\Omega} C_i(x) |\nabla \lambda_i(x)| dx, \tag{28}$$

when q_i is optimized over the set (17) and (20). In view of (27) and (28), maximizing (23) over all the flow functions p and q leads directly to the equivalent dual model (24). The constraints (14) must be satisfied for an optimal ϕ , otherwise the energy would be infinite, contradicting boundedness of the max-flow problem from above.

Note that a solution to the problem (24) exists since (24) is convex, lower semi-continuous and bounded from below and the constraints (24) are convex. Regarding existence of a solution to the max-flow problem (21), due to boundedness from above a maximizing sequence $\{p^i, q^i\}_{i=1}^{\infty}$ exists to the problem (21). However, it may not admit a maximizing subsequence w.r.t. q^i which converges to a $q^* \in C_\nu$ because the supremum may be attained for a discontinuous q^* which lies in the closure of the set of smooth vector fields $C^\infty(\Omega)^N$ and not in the set itself. In this paper we still speak of (p^*, q^*) as a primal-dual solution even though q^* may be discontinuous to ease readability. A more formal presentation can be given if arguments involving (p^*, q^*) are replaced with $\lim_{i \rightarrow \infty} \{p^i, q^i\}_{i=1}^{\infty}$ for the maximizing sequence $\{p^i, q^i\}_{i=1}^{\infty}$.

3.4 Exact and Global Optimums

The functions λ_i , $i = 1 \dots n - 1$, of the convex model (24) are relaxed to take values in the convex set $[0, 1]$, which is in contrast to the binary constraints of the original nonconvex formulation (15). The following proposition establishes a primal-dual relationship between the max-flow problem (21) and the original non-convex problem (15). By solving the max-flow problem (21) a set of optimizers to the original binary constrained problem (15) can be obtained by thresholding each layer function λ_i^* .

Proposition 1. *Assume ϕ^* is a minimizer of (24) and let $\{t_i\}_{i=1}^{n-1}$ be a sequence such that $0 < t_1 = t_2 = \dots = t_{n-1} \leq 1$. Define the level sets*

$$S_i^{t_i} = \{x : \lambda_i^*(x) \geq t_i\}, \quad i = 1 \dots n - 1 \tag{29}$$

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

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

then the set of binary functions $\lambda_i^{t_i}(x)$, $i = 1, \dots, n - 1$, is a global optimum of the original nonconvex multi-labeling problem (15). Furthermore, if $(p^*, q^*; \lambda^*)$ is any optimal primal-dual solution of (22), the cut given by $\lambda_i^{t_i}(x)$, $i = 1, \dots, n - 1$, has an energy equal to the max flow energy in (21), i.e.

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

Proof. Since p_i^* , $i = 1, \dots, n$ and q_i^*, λ_i^* , $i = 1, \dots, n - 1$ is a global optimum of the primal-dual problem (22), then p_i^*, q_i^* optimize the dual problem (21) and $\lambda_i^*(x)$ optimizes (24).

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

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

Since the variables are optimal, the flow conservation condition (19) must hold, i.e

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

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

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

where $L_{S_i^{t_i}}$ is the length $|\partial S_i^{t_i} \setminus (\partial S_i^{t_i} \cap \partial \Omega)|$. We will prove $E^k = E^P(p^*)$ for any $k \in \{1, \dots, n - 1\}$ and start by considering $k = 1$. By the formula (27), it follows that

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

This, together with the fact that

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

implies that the total max-flow energy defined in (21) can be written

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

The last term follows because

$$\int_{S_i^{\ell_i}} \operatorname{div} q_i^*(x) \, dx = \int_{\Omega} \lambda_i^* \operatorname{div} q_i^* \, dx = \alpha \int_{\Omega} |\nabla \lambda_i^{\ell_i}| \, dx = \alpha \left| \partial S_i^{\ell_i} \setminus (\partial S_i^{t_i} \cap \partial \Omega) \right|. \tag{30}$$

where the second equality is due to Prop. 4 of [25]. Note that the boundary length $L_{S_1^{t_1}}$ is necessarily finite, otherwise the energy would be infinite, contradicting boundedness from above.

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

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

By the definition (29) 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 (27), 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

$$\begin{aligned} E^P(p^*) = E^k &= \sum_{i=1}^{k-1} \int_{S_{i-1}^{\ell_{i-1}} \setminus S_i^{t_i}} \rho(t_i, x) \, dx + \int_{S_{k-1}^{\ell_{k-1}} \setminus S_k^{\ell_k}} \rho(\ell_k, x) \, dx \\ &\quad + \int_{S_k^{\ell_k}} p_{k+1}^*(x) \, dx + L_{S_k^{t_k}} + \alpha \sum_{i=1}^{k-1} L_{S_i^{t_i}} = E^{k+1}. \end{aligned} \tag{31}$$

Hence, we can conclude that also $E^{n-1} = E^P(p^*)$. By noting from (27) 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 (21) can be written

$$\begin{aligned}
 E^P(p^*) &= E^{n-1} = \int_{\Omega \setminus S_1^{t_1}} \rho(\ell_1, x) dx + \sum_{i=2}^{n-1} \int_{S_{i-1}^{\ell_{i-1}} \setminus S_i^{t_i}} \rho(t_i, x) dx \\
 &\quad + \int_{S_{n-1}^{t_{n-1}}} \rho(\ell_n, x) dx + \alpha \sum_{i=1}^{n-1} L_{S_i^{t_i}}
 \end{aligned} \tag{32}$$

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 (24) of the set of binary functions $\lambda_i^{t_i}$. Therefore, by duality between the max-flow problem (21) and the convex relaxed problem (24), $\lambda_i^{t_i}$ must be a global minimum of the min-cut problem (24) and therefore also a global minimum of the original problem (15).

3.5 ‘Saturated’/‘Unsaturated’ Edges

In the discrete setting, it is well known that the minimum cut severs edges that are saturated in the max-flow problem. This section attempts to give a variational explanation to the phenomena for the continuous max-flow and min-cut problems studied in this work. Let $\lambda_1^*, \dots, \lambda_{n-1}^*$ be optimal to the dual problem (24). Assume that for some $x \in \Omega$ and $i \in 1, \dots, n$ $\lambda_i(x) > t > \lambda_{i+1}(x)$, where $t \in (0, 1)$. Thresholding at t will generate the binary solution $\lambda_0(x) = \dots = \lambda_i(x) = 1$ and $\lambda_{i+1}(x), \dots, \lambda_n(x) = 0$. Therefore the cut generated by the binary function ‘severs’ the edge $e_{i+1}(x)$ between layer i and $i+1$. Since $\lambda_i(x) > \lambda_{i+1}(x)$ it follows by (27) that the optimal flow function must satisfy $p_i^*(x) = \rho(\ell_i, x)$, i.e. the edge $e_i(x)$ is saturated. Assume on the contrary that for some $x \in \Omega$ and $i \in 1, \dots, n$ $p_i^*(x) < \rho(\ell_i, x)$. In this case $\lambda_i^*(x) = \lambda_{i+1}^*(x)$, otherwise $p_i^*(x)$ would not be optimal since increasing $p_i^*(x)$ would also increase the energy. Consequently, for any threshold level $t \in (0, 1]$, $\lambda_i^t(x) = \lambda_{i+1}^t(x)$, i.e. the edge $e_i(x)$ is not severed by the cut.

Similar interpretations of the spatial flow can be made by using the identity

$$\int_{\Omega} \lambda \operatorname{div} q dx = \int_{\Omega} q \cdot \nabla \lambda dx \tag{33}$$

If for some $x \in \Omega$ and a neighborhood $\mathcal{N}_\epsilon(x) = \{y \in \Omega : \|y - x\| < \epsilon\}$, $|q_i^*(y)| < \alpha$ for all $y \in \mathcal{N}_\epsilon(x)$, we say the spatial flow is unsaturated in $\mathcal{N}_\epsilon(x)$. Then λ_i is constant in $\mathcal{N}_\epsilon(x)$. Consequently, for any threshold $t \in (0, 1]$, $\lambda_i^t(y)$ is either identically 0 or 1 in $\mathcal{N}_\epsilon(x)$ and the cut will not sever the spatial domain $\mathcal{N}_\epsilon(x)$ at the i -th layer. Assume $\nabla \lambda_i \neq 0$ in some domain $S \subset \Omega$, then by (33) $|q_i^*(x)| = |\alpha \nabla \lambda_i^*| / |\nabla \lambda_i^*| = \alpha$ a.e. $x \in S$. Consequently, for any threshold $t \in (0, 1]$, $|q_i^*| = \alpha$ whenever $\nabla \lambda_i^t \neq 0$ in the distributional sense.

3.6 Extension to Continuous Labelings

Assume now that the feasible label values are constrained to the continuous interval $[\ell_{\min}, \ell_{\max}]$. As the number of labels goes to the limit of infinity, the max-flow problem (21) with the flow constraints (17)–(19) turns into

$$\sup_{p,q} \int_{\Omega} p(\ell_{\min}, x) dx \quad (34)$$

$$\text{s.t. } p(\ell, x) \leq \rho(\ell, x), \quad |q(\ell, x)| \leq \alpha, \quad \forall x \in \Omega, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}] \quad (35)$$

$$\operatorname{div}_x q(\ell, x) + \partial_{\ell} p(\ell, x) = 0, \quad \text{a.e. } x \in \Omega, \quad \ell \in [\ell_{\min}, \ell_{\max}]. \quad (36)$$

where $\ell \in [\ell_{\min}, \ell_{\max}]$ is the set of all feasible continuous-valued labels. The flow functions $p(x)$ and $q(x)$ are defined in the one dimensional higher space $[\ell_{\min}, \ell_{\max}] \times \Omega$. By carrying out similar steps as in the last section, the following dual problem can be derived

Proposition 2. *The max-flow model (34) with continuous label-values is dual / equivalent to the following min-cut model over $[\ell_{\min}, \ell_{\max}] \times \Omega$:*

$$\begin{aligned} & \min_{\lambda(\ell,x) \in [0,1]} \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} \{ \alpha |\nabla_x \lambda| - \rho(\ell, x) \partial_{\ell} \lambda(\ell, x) \} dx d\ell \\ & + \int_{\Omega} (1 - \lambda(\ell_{\min}, x)) \rho(\ell_{\min}, x) + \lambda(\ell_{\max}, x) \rho(\ell_{\max}, x) dx \end{aligned} \quad (37)$$

subject to

$$\partial_{\ell} \lambda(\ell, x) \leq 0, \quad \lambda(\ell_{\min}, x) \leq 1, \quad \lambda(\ell_{\max}, x) \geq 0, \quad \forall x \in \Omega, \quad \forall \ell \in [\ell_{\min}, \ell_{\max}]. \quad (38)$$

The proof can be found in [23].

The labeling function $u(x)$ can finally be reconstructed from the binary function $\lambda(\ell, x)$ by $u(x) = \ell_{\min} + \int_{\ell_{\min}}^{\ell_{\max}} \lambda(\ell, x) d\ell$.

In [9], Pock et al. gave a similar formulation of continuous labeling problems, as the search for a binary function defined over $[\ell_{\min}, \ell_{\max}] \times \Omega$, which minimizes

$$\min_{\lambda(\ell,x) \in \{0,1\}} \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} \{ \alpha |\nabla_x \lambda| + \rho(\ell, x) |\partial_{\ell} \lambda(\ell, x)| \} dx d\ell. \quad (39)$$

subject to

$$\lambda(\ell_{\min}, x) = 1, \quad \lambda(\ell_{\max}, x) = 0, \quad x \in \Omega \quad (40)$$

In order to solve this non-convex binary problem, the convex relaxation of [8] was adopted by minimizing over $\lambda(x, \ell) \in [0, 1]$. By applying the thresholding result of [8], binary optimizers could be obtained by thresholding the computed result.

Some differences can be observed between our formulation (37), (38) and the formulation (39), (40): The constraint $\partial_{\ell} \lambda(\ell, x) \leq 0$ is not forced explicitly in [9]. However, it turns out the presence of the absolute value of the term

$\rho(\ell, x) |\partial_\ell \lambda(\ell, x)|$ forces this constraint to hold. Observe that if $\rho(\ell, x) < 0$ is negative, the formulation of (39) is non-convex, and can therefore not be solved globally. This is in contrast to our formulation (37), which is convex also in this case. The functional (39) could be made convex by adding a sufficiently large number to the data term at every $x \in \Omega$. In the more recent work of Pock et al. [10], a more strict derivation resulted in a little different formulation. In this formulation, the integrand of the energy functional is infinite if $\partial_\ell \lambda(\ell, x) \leq 0$, hence this constraint is forced to hold. Their derivations rely heavily on results from the theory of calibrations [12] and cartesian currents [26, 27]. Label values ranged over the whole real line \mathbb{R} was assumed, which required to impose limits at infinity: $\lim_{\ell \rightarrow +\infty} \lambda(\ell, x) = 0$ and $\lim_{\ell \rightarrow -\infty} \lambda(\ell, x) = 1$.

We eventually stick to a finite label value set in practice. After discretization, the label space also becomes discrete in [10]. However, it has not been proven if all properties, such as the thresholding scheme and monotonicity constraint hold exactly after discretization. In contrast, these properties were proved to hold exactly for our model with discrete label values developed in Sect. 3.

Last but not the least, a primal-dual algorithm was proposed in [10], which consists of iteratively taking ascent steps over the dual variables p and q and descent step over the primal variable λ , followed by projections of all the variables onto the nearest points of the feasible sets iteratively until convergence.

4 Algorithms

4.1 Multiplier-Based Max-Flow Algorithm

In this section, it is assumed that the image domain Ω is discrete and the differential operators are discretized, such that the optimization problems become finite dimensional. We stick to the continuous notation, using \int , ∇ and \div to ease readability. As stated in the previous section, the energy formulation of (22) is just the Lagrangian function of (21) and λ_i , $i = 1, \dots, n-1$, are the multiplier functions. To this end, we define its respective augmented Lagrangian function as

$$L_c(p, q, \lambda) := \int_{\Omega} p_1 + \sum_{i=1}^{n-1} \lambda_i (\operatorname{div} p_i + p_{i+1} - p_i) - \frac{c}{2} |\operatorname{div} p_i + p_{i+1} - p_i|^2 dx, \quad (44)$$

where $c > 0$.

We propose an algorithm for the continuous maximal flow problem (21) based on the augmented Lagrangian method [29], see Algorithm 3.6. Algorithm 3.6 is an example of an alternating direction method of multipliers, where (44) is maximized alternatively with respect to each variable p_i, q , followed by an updating of the Lagrange multipliers λ_i , $i = 1, \dots, n-1$ at each iteration. For the two-label case, a similar flow-maximization scheme for the continuous min-cut problem was proposed in [19, 20].

Algorithm 1 Multiplier-Based Maximal-Flow Algorithm

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

- For each layer $i = 1 \dots n$ solve
 - Optimize p_i by fixing other variables

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

which can be explicitly computed at each point $x \in \Omega$; At the first and last layer, $i = 1$ and $i = n - 1$ the update formulas are a little different and are given in (42) and (43) below.

- Optimize q_i , by introducing the new value of p_i^{k+1} and fixing other variables

$$\begin{aligned} q_i^{k+1} &:= \arg \max_{\|q\|_\infty \leq \alpha} L_c((\tilde{p}_{i \leq j}^{k+1}, p_i^k, p_{j > i}^k), (q_{j < i}^{k+1}, q_i, q_{j > i}^k), \lambda^k) \\ &:= \arg \max_{\|q\|_\infty \leq \alpha} -\frac{c}{2} \left\| \operatorname{div} q_i + p_{i+1}^k - \tilde{p}_i^{k+1} - \lambda_i^k / c \right\|^2, \end{aligned} \quad (41)$$

which can either be solved iteratively by the projected-gradient algorithm [28], or approximately by one linearized step (45);

- Optimize p_i again, by introducing the new values of q_i^{k+1} and fixing others

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

which can be explicitly computed at each point $x \in \Omega$;

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

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

- Set $k \leftarrow k + 1$ and repeat until convergence.

At the first and last layer $i = 1$ and $i = n$ the update formulas for p_1 and p_n are:

$$\begin{aligned} p_1^{k+1} &:= \arg \max_{p_1(x) \leq \rho(\ell_1, x)} L_c(p_1, p_2^k, \dots, p_n^k, q^{k+1}, \lambda^k) \\ &:= \arg \max_{p_1(x) \leq \rho(\ell_1, x)} \int_\Omega p_1 dx - \frac{c}{2} \left\| p_1 - (p_2^k + \operatorname{div} q_1^{k+1}) + \lambda_1^k / c \right\|^2, \end{aligned} \quad (42)$$

and

$$\begin{aligned} p_n^{k+1} &:= \arg \max_{p_n(x) \leq \rho(\ell_n, x)} L_c(p_1^{k+1}, \dots, p_{n-1}^{k+1}, p_n, q^{k+1}, \lambda^k) \\ &:= \arg \max_{p_n(x) \leq \rho(\ell_n, x)} -\frac{c}{2} \left\| p_n + \operatorname{div} q_{n-1}^{k+1} - p_{n-1}^{k+1} - \lambda_{n-1}^k / c \right\|^2. \end{aligned} \quad (43)$$

Both can be computed explicitly;

Instead of solving the sub-problem (41) iteratively by the projected-gradient algorithm [28], an inexact solution can be obtained by the linearization:

$$q_i^{k+1} = \Pi_\alpha \left(q_i^k + c \nabla (\operatorname{div} q_i^k + p_{i+1}^k - p_i^{k+1} - \lambda_i^k / c) \right) \quad (45)$$

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

5 Numerical Experiments

In this work, we focus on applications to image segmentation and stereo reconstruction. Comparisons are made to the discrete approach [6] and the primal-dual algorithm of [9].

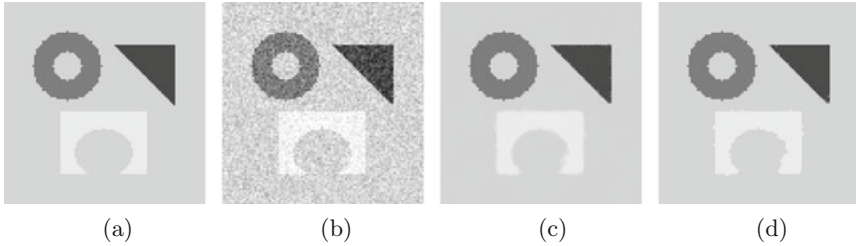


Fig. 2. (a) Ground truth, (b) input, (c) Rescaled labeling function before threshold, (d) Rescaled labeling function after thresholding each λ_i at 0.5.

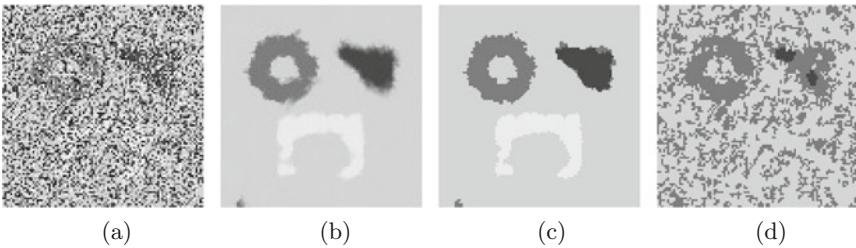


Fig. 3. (a) Input image damaged by impulse noise; (b) reconstructed labeling function with non-convex data term (47) before threshold, (c) labeling function after thresholding each λ_i at 0.5, (d) reconstructed labeling function with convex data term (46) and $\beta = 1$.

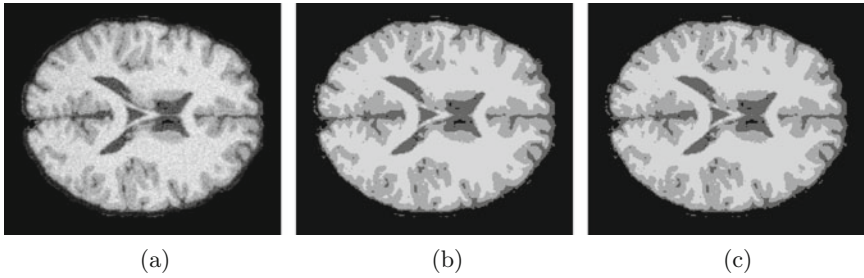


Fig. 4. (a) Input, (b) Labeling function before threshold (c) Labeling function after thresholding each λ_i at 0.5.

In case of image segmentation we assume $\ell_i = i$, $i = 1, \dots, n$ and n is the number of regions. $\rho(i, x)$ is the data cost of assigning pixel x to region i . One possibility is

$$\rho(i, x) = |I(x) - c_i|^\beta, \quad i = 1, \dots, n \quad (46)$$

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. Such a data term is convex for $\beta \geq 1$ and non-convex for $\beta < 1$. Results with $\beta = 2$ are shown in Figs. 2, 4. We also demonstrate image segmentation with a non-convex data term in Fig. 3. The ground truth image from Fig. 2(a) has been damaged by impulse noise in Fig. 3(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 Fig. 3(d) where we have selected (46) with $\beta = 1$. Instead the following non-convex data term can be used

$$\rho(i, x) := \begin{cases} 0, & \text{if } i = \operatorname{argmin}_k |I(x) - c_k| \\ 1, & \text{else} \end{cases} \quad (47)$$

In the stereo application we are given two color images I_L and I_R of a scene taken from horizontally slightly different viewpoints and would like to reconstruct the depth map u . The quality of the matching between I_L and I_R for a depth value u is measured by using the following ρ in the data term of (13)

$$\rho(u, x) = \sum_{j=1}^3 |I_L^j(x) - I_R^j(x + (u, 0)^T)|. \quad (48)$$

Here $I^j(x)$ denotes the j th component of the color vector $I(x)$. The above data term (48) is obviously highly non-convex. The results on a standard example are shown in Fig. 5, where comparison are also given [10] and graph cut with a neighborhood system of 4 and 8. Graph cut produces a single non-unique solution which is shown in Fig. 5(f) and (g) with 4 and 8 neighbors respectively. As we see, such solutions suffer from metrication artifacts because of the discrete grid bias.

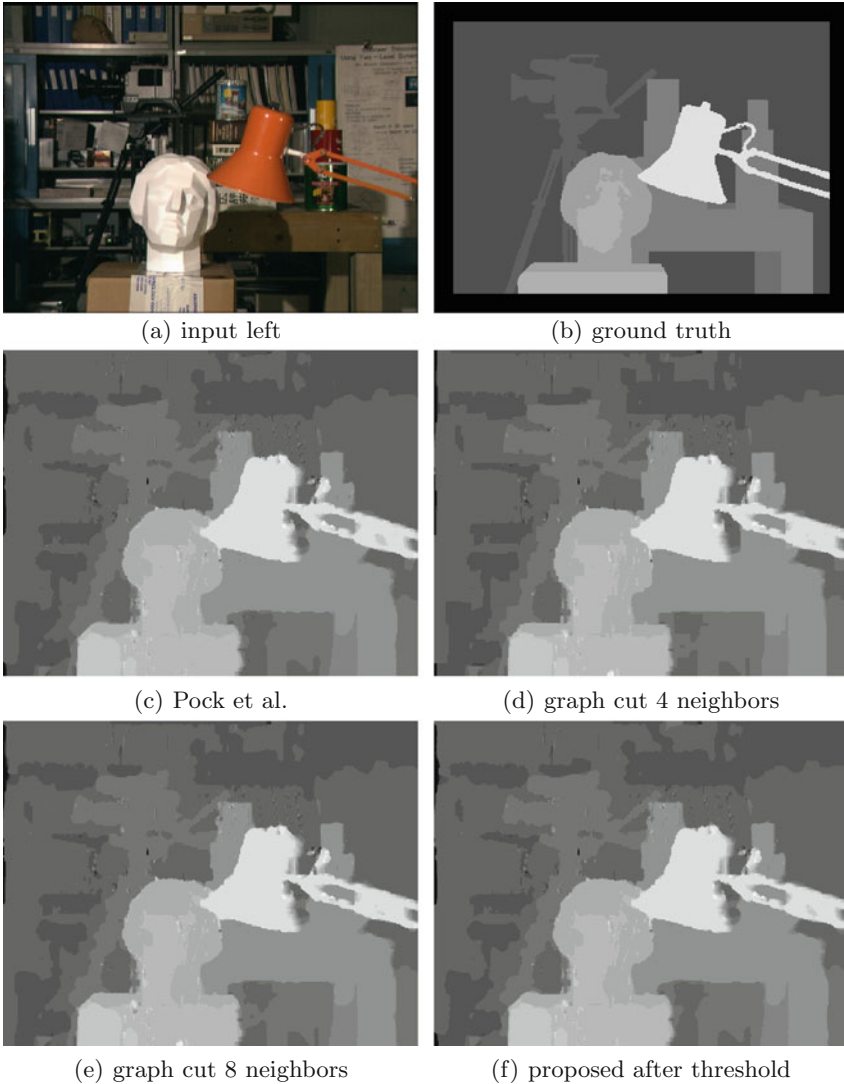


Fig. 5. Stereo depth estimation.

Iteration counts for all experiments are presented in Table 1 and CPU times are shown in Table 2. The two variants of Algorithm 1 are evaluated against the primal-dual method of Pock et al. [10]. The relative energy precision at iteration i is given by

$$\varepsilon = \frac{E^i - E^*}{E^*}, \quad (49)$$

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

Table 1. Iteration counts for each experiment. Number of iterations to reach an energy precision of 10^{-3} and 10^{-4} are shown. PD = Primal-dual. Proposed 1 stands for Algorithm 1 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 1 with the subproblems solved inexactly in one step through the linearization (45).

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

Table 2. CPU time in seconds for each experiment for reaching an energy precision of 10^{-3} and 10^{-4} . PD = Primal-dual. Proposed 1 stands for Algorithm 1 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 1 with the subproblems solved inexactly in one step through the linearization (45).

	Energy precision $\varepsilon < 10^{-3}$			Energy precision $\varepsilon < 10^{-4}$		
	PD [10]	Proposed 1	Proposed 2	PD [10]	Proposed 1	Proposed 2
Brain	86	68	38	132	89	96
Figure 2	1.34	0.64	0.47	2.61	1.18	1.32
Stereo	2027	1214	598	7153	2029	2182

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 [31,32]. In order to make the comparison as accurate as possible, the primal-dual algorithm [10] 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 [10].

The first variant of Algorithm 3.6 solves the subproblem (41) iteratively by Chambolle’s algorithm [28]. Since the previous solution is available as a good initialization, not many iterations of this algorithm is required. In our experiments, 5 inner iterations was used for each step. Increasing the number of inner iterations beyond 5 did not seem to have any impact on the convergence rate in our experience.

The primal-dual method of [10] 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 $\varepsilon < 10^{-5}$ after 1310 iterations, $\varepsilon < 10^{-6}$ after 1635 iterations and $\varepsilon < 10^{-7}$ after 2340 iteration. The primal-dual

Table 3. Iteration counts for stereo experiment. Number of iterations to reach an energy precision of 10^{-4} , 10^{-5} and 10^{-6} are shown. PD = Primal-dual.

	Energy prec. $\varepsilon < 10^{-4}$		Energy prec. $\varepsilon < 10^{-5}$		Energy prec. $\varepsilon < 10^{-6}$	
	PD [10]	Proposed 1	PD [10]	Proposed 1	PD [10]	Proposed 1
Stereo	14305	920 ($\times 5$)	> 30000	1310 ($\times 5$)	> 30000	1635 ($\times 5$)

algorithm [10] failed to ever reach an energy precision of 10^{-5} or lower within our predetermined number of maximum iterations (30000). We believe this difference is due to the fact that our approach avoids the iterative projections of the labeling function and hence progresses in the exact steepest descent direction every iteration.

The second variant of the Algorithm 1 instead computes an inexact solution to (41) through the linearization (45) and hence avoids the inner iterations. However, the penalty parameter c must be set lower to maintain convergence, hence more outer iterations are required. Overall it converges a little faster than the first variant and outperforms the primal-dual algorithm [10] for all the experiments.

Comparison to discrete graph cut [33] is more complicated. Our algorithms are implemented in matlab, in contrast to the optimized c++ discrete max-flow implementation of [33]. 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. 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.

6 Conclusions

In this paper we proposed and investigated a novel max-flow formulation of multilabelings in the continuous setting. It is a direct mapping of Ishikawa’s graph-based configuration to the continuous setting. We proved the maximization problem is dual to an equivalent min-cut formulation by variational analysis. In addition, we proposed a new and reliable multiplier-based max-flow algorithm with convergence that can be verified by optimization theories, which was demonstrated to significantly outperform earlier approaches. Due to its continuous formulation, the algorithm can easily be speeded up by a multigrid or parallel implementation, in contrast to graph-based methods. The memory requirement is also not as strict.

Acknowledgements. This research has been supported by the Norwegian Research Council eVita project 214889 and eVita project 166075 and Natural Sciences and Engineering Research Council of Canada (NSERC) Accelerator Grant R3584A04.

References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1222–1239 (2001)
2. Appleton, B., Talbot, H.: Globally optimal surfaces by continuous maximal flows. In: *DICTA*, pp. 987–996 (2003)
3. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002, Part III*. LNCS, vol. 2352, pp. 82–96. Springer, Heidelberg (2002)
4. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 65–81 (2004)
5. Lempitsky, V., Boykov, Y.: Global optimization for shape fitting. In: *CVPR (2007)*
6. Ishikawa, H.: Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1333–1336 (2003)
7. Kohli, P., Kumar, M.P., Torr, P.H.: p^3 and beyond: move making algorithms for solving higher order functions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 1645–1656 (2009)
8. Chan, T.F., Esedoğlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **66**, 1632–1648 (2006). (electronic)
9. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part III*. LNCS, vol. 5304, pp. 792–805. Springer, Heidelberg (2008)
10. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. Technical report, Institute for Computer Graphics and Vision, Graz University of Technology (2010)
11. Bouchitté, G.: Recent convexity arguments in the calculus of variations. In: *Lecture Notes from the 3rd International Summer School on the Calculus of Variations*, Pisa (1998)
12. Alberti, G., Bouchitté, G., Maso, G.D.: The calibration method for the mumford-shah functional and free-discontinuity problems. *Calc. Var. Partial Differ. Eqn.* **16**, 299–333 (2003)
13. Ford, L.R., Fulkerson, D.R.: *Flows in Networks*. Princeton University Press, Princeton (1962)
14. Ford, L.R., Fulkerson, D.R.: Maximal flow through a network. *Can. J. Math.* **8**, 399–404 (1956)
15. Goldberg, A.V., Tarjan, R.E.: A new approach to the maximum-flow problem. *J. ACM* **35**, 921–940 (1988)
16. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: *Introduction to Algorithms*, 2nd edn. MIT Press, Cambridge (2001)
17. Strang, G.: Maximal flow through a domain. *Math. Program.* **26**, 123–143 (1983)
18. Appleton, B., Talbot, H.: Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 106–118 (2006)
19. Yuan, J., Bae, E., Tai, X.: A study on continuous max-flow and min-cut approaches. In: *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, pp. 2217–2224 (2010)
20. Yuan, J., Bae, E., Tai, X., Boykov, Y.: A study on continuous max-flow and min-cut approaches. Technical report CAM10-61, UCLA, CAM (2010)

21. Yuan, J., Bae, E., Tai, X.-C., Boykov, Y.: A continuous max-flow approach to potts model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 379–392. Springer, Heidelberg (2010)
22. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
23. Bae, E., Yuan, J., Tai, X., Boykov, Y.: A fast continuous max-flow approach to non-convex multilabeling problems. Technical report, UCLA, CAM-report 10-62 (2010)
24. Bae, E., Tai, X.-C.: Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) SSVM 2009. LNCS, vol. 5567, pp. 1–13. Springer, Heidelberg (2009)
25. Bae, E., Yuan, J., Tai, X.C.: Global minimization for continuous multiphase partitioning problems using a dual approach. *Int. J. Comput. Vision* **92**, 112–129 (2011)
26. Giaquinta, M., Modica, G., Soucek, J.: Cartesian Currents in the Calculus of Variations I. *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge A Series of Modern Surveys in Mathematics*, vol. 37. Springer, Heidelberg (1998)
27. Giaquinta, M., Modica, G., Soucek, J.: Cartesian Currents in the Calculus of Variations II. *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge A Series of Modern Surveys in Mathematics*, vol. 38. Springer, Heidelberg (1998)
28. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
29. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
30. Esser, J.E.: Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting (2010)
31. Hyman, J.M., Shashkov, M.J.: Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.* **33**, 81–104 (1997)
32. Hyman, J.M., Shashkov, M.J.: Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids. *Appl. Numer. Math.* **25**, 413–442 (1997)
33. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 359–374 (2001)