Segmentation with non-linear regional constraints via line-search cuts

Lena Gorelick[1], Frank R. Schmidt[2], Yuri Boykov[1],
Andrew Delong[1], and Aaron Ward[1]

[1] University of Western Ontario, Canada
[2] Université Paris Est, France

Abstract. This paper is concerned with energy-based image segmentation problems. We introduce a general class of *regional functionals* defined as an arbitrary non-linear combination of regional unary terms. Such (high-order) functionals are very useful in vision and medical applications and some special cases appear in prior art. For example, our general class of functionals includes but is not restricted to soft constraints on segment volume, its appearance histogram, or shape.

Our overall segmentation energy combines *regional functionals* with standard length-based regularizers and/or other submodular terms. In general, *regional functionals* make the corresponding energy minimization NP-hard. We propose a new greedy algorithm based on iterative *line search*. A parametric max-flow technique efficiently explores all solutions along the direction (line) of the steepest descent of the energy. We compute the best "step size", *i.e.* the globally optimal solution along the line. This algorithm can make large moves escaping weak local minima, as demonstrated on many real images.

1 Introduction

Many existing image segmentation methods optimize segments S with respect to quality functionals of the form

$$E(S) = U(S) + Q(S), \tag{1}$$

where term $U(S)$ describes some regional properties of the segments and $Q(S)$ is a regularization prior for the segmentation boundary or shape. Such functionals have many weak local minima on real images and global optimization methods are preferred. Many powerful optimization methods are limited to special forms of $U(S)$ and $Q(S)$. For example, in continuous formulations the segmentation

---

S is a subset of some continuous image domain $\Omega \subset \mathbb{R}^n$. Current optimization methods computing globally optimal S require that term $U(S)$ is linear with respect to characteristic function 1_S and regularization term $Q(S)$ is a convex TV-based functional [1–3]. In discrete formulations where segmentation is defined as a vector of binary variables $S = \{s_p | p \in \Omega\}$ for a finite set of pixels $\Omega \subset \mathbb{Z}^n$, efficient combinatorial methods for global optimization require that $U(S)$ be a *unary term* and $Q(S)$ be a quadratic *submodular term* [4, 5].

This paper proposes a general class of *non-linear* regional functionals $R(S)$ that can impose significantly more powerful constraints over segment regions than standard unary or linear functionals $U(S)$. Our regional functionals $R(S)$ could still be combined with some boundary regularization. Thus, our overall segmentation energy includes two terms

$$E(S) = R(S) + Q(S), \qquad (2)$$

where $Q(S)$ is a standard quadratic submodular energy (in the discrete case) or a convex TV-based functional (in the continuous case). For example, $Q(S)$ could represent segmentation boundary length.

Below we review standard unary/linear terms $U(S)$ and motivate an extension to a general class of non-linear regional functionals $R(S)$. The introduction concludes with a summary of contributions in this paper.

1.1   Prior art on regional constraints in segmentation

The notion *unary term* comes from MAP-MRF labeling literature and corresponds to a sum of individual pixel potentials inside a posterior energy. For example, common MRF-based segmentation energies [6, 7] include a unary term representing an appearance model as the sum of log-likelihoods for intensities I_p

$$\sum_{p \in \Omega} - \ln \Pr(I_p | L_p) \qquad (3)$$

where variables L_p are pixel labels (object or background), and probability distributions $\Pr(\cdot|obj)$, $\Pr(\cdot|bkg)$ are known models for intensities inside two segments. Using binary variables s_p (values 1 and 0 represent segment interior and exterior, respectively) this unary log-likelihood term can be equivalently represented as

$$U(S) = \sum_{p \in \Omega} f_p \cdot s_p \qquad (4)$$

for $f_p = - \ln \left(\frac{\Pr(I_p|obj)}{\Pr(I_p|bkg)} \right)$. In general, any unary energy term can be written as expression (4) for some potentials f_p. Since (4) is linear w.r.t. binary variables s_p, unary energy terms may also be called *linear*. In a continuous segmentation framework the analogue for the unary term (4) is a linear functional w.r.t characteristic function 1_S of segment S

$$U(S) = \int_S f \, dx = \int_\Omega f \cdot 1_S \, dx \qquad (5)$$

where $f : \Omega \to \mathbb{R}$ is a given scalar potential function.

Linear functionals like (4) or (5) are widely used in computer vision [6–8] to express various regional constraints on segments (e.g. appearance model, volumetric ballooning, etc). The main advantage of linear regional functionals is that many efficient global optimization algorithms [5, 1–3] can easily incorporate them. However, linear functionals can enforce only a fairly limited set of properties on segments. For example, standard linear *ballooning* term

$$U(S) = -V(S), \qquad \text{where} \qquad V(S) = \int_{\Omega} 1_S \, dx$$

can encourage larger size. But practically useful (e.g. in medical applications) bias towards a specific target value V_0 for segment size, *i.e.*

$$R(S) = (V(S) - V_0)^2$$

cannot be expressed linearly. It is also known (see Boykov-Jolly in Figs. 3-7) that linear log-likelihood term (3) is far from ideal for enforcing appearance models. It is particularly problematic in gray-scale medical images (see Figs. 4,5,9) due to significant overlap between the object and background intensity distributions. Limitations of unary/linear terms motivated some researchers to use non-linear regional appearance models. For example, [9–11] minimize some distances (*i.e.* KL or Bhattacharyya) between the observed and target intensity distributions. There are other examples of prior work using specific non-linear regional functionals. For example, [12] use mutual information as an appearance model and [13, 14] enforce non-linear constraints on segment's area.

In general, optimization of non-linear regional terms is NP-hard and cannot be addressed by standard global optimization methods. Some earlier papers developed specialized techniques for particular forms of non-linear regional functional. For example, the algorithm in [11] was developed for minimizing Bhattacharyya distance between distributions and a dual-decomposition approach in [14] applies to convex piece-wise linear regional functionals[3]. An active-contour technique in [9] addresses arbitrary distance measures between distributions, but any gradient flow is likely to lead to a weak local minimum; see *local line search* in Figs. 4,6. Combinatorial techniques in [12, 10] apply to general non-linear regional functionals, as defined in this paper, and they can make *large moves* by globally minimizing some approximating energy. However, we show that greedy algorithms in [12, 10] may stop at solutions that are not even a local minimum of the actual non-linear functional; see Fig. 8 and a discussion in Sec. 3.4.

1.2   Contributions

In Section 2 we introduce a general class of *non-linear regional functionals* $R(S)$ defined as an arbitrary non-linear combination of any regional unary terms. All examples of regional functionals mentioned in the previous section are special

---

[3] E.g., [14] can use distances between bin counts, but not normalized distributions.

cases of our general class. Our functionals can enforce constraints on area or volume and penalize any distance metric (KL, Bhattacharyya, L_1 etc) between a segment's intensity distribution and some target distribution. Other practically useful examples of our regional functionals are discussed in future work Section 5.

In Section 3 we propose a general *line-search cuts* algorithm for minimizing energies (2) combining any non-linear regional functional $R(S)$ with an arbitrary submodular term $Q(S)$. At each iteration the method explores a "line" of solutions along the gradient direction and selects a global minimum on this line. We use an efficient parametric max-flow technique [15] to exhaustively traverse this line[4]. The algorithm can guarantee only a local minimum at convergence, however "large moves" allow to escape many weak solutions. Our method follows the *exact (global) line-search* principle [17]. While many possible variants of our technique, e.g. standard *backtracking line-search*, may reduce the runtime, in general, they are more likely to get stuck in local minima.

In contrast to many specialized techniques, our optimization method applies to arbitrary non-linear regional functionals. Section 4 shows our results for different applications on synthetic, medical, and natural images. We compare our *global line-search* approach to basic gradient descent (see *local line-search* in Figs. 4,6) and discuss limitations of earlier "large move" methods [12, 10]. These two methods are also applicable to general non-linear regional functionals, but unlike our method may stop at solutions that are not even a local minimum of the actual non-linear functional (see Fig. 8 and Sec. 3.4). Future work and extensions for non-linear regional functionals are reviewed in Sec. 5.

2   Regional Functionals

This section discusses the term *regional functionals*. Let $I : \Omega \to \mathbb{R}^m$ be an image with domain $\Omega \subset \mathbb{R}^n$ and m-dimensional colors. As discussed in Section 1.1, the most common type of regional terms used in segmentation is a linear functional $U(S)$ which can be represented via arbitrary scalar function $f : \Omega \to \mathbb{R}$

$$U(S) = \int_S f(x)\,\mathrm{d}x = \int_\Omega f(x) \cdot 1_S(x)\,\mathrm{d}x =: \langle f, S \rangle . \tag{6}$$

Usually, f corresponds to some *appearance model* based on intensities observed in image I, see equations (3)-(5). The integral in (6) can be seen as a dot product between scalar function f and the characteristic function of set S. Thus, in the rest of this paper we use notation $\langle f, S \rangle$ to refer to such linear functionals.

Now we introduce a general class of non-linear regional functionals. In particular, we focus on arbitrary non-linear combinations of linear functionals. We assume several scalar functions $f_1, \ldots, f_k : \Omega \to \mathbb{R}$, each defining a linear functional of type (6), and one non-linear function $F : \mathbb{R}^k \to \mathbb{R}$ that combines these linear functionals in the overall *non-linear regional functional*

$$R(S) = R^F_{\{f_1, \ldots, f_k\}}(S) = F(\langle f_1, S \rangle, \ldots, \langle f_k, S \rangle). \tag{7}$$

---

[4] Our greedy line-search method can also be implemented in continuous framework for convex TV-based $Q(S)$ using ideas in [16].

2.1   Examples of Regional Functionals

The following examples describe regional functionals that are very useful in computer vision applications:

Volume Constraint The volume constraint penalizes a segmentation when its size deviates from a specific target volume V_1:

$$f_1(x) = 1 \qquad\qquad F(v_1) = (v_1 - V_1)^2, \qquad\qquad (8)$$

corresponding to $R(S) = R^F_{\{f_1\}}(S) = F(\langle f_1, S\rangle) = (\langle f_1, S\rangle - V_1)^2$ in (7).

Bin Count Constraint The bin count constraint penalizes a deviation between the number of segment pixels in a certain bin (appearance model) and a target count. To that end, we pre-compute k indicator functions f_i for each bin, and in combination with target bin counts $V_1, \ldots, V_k$, we consider the following regional functional:

$$F(v_1, \ldots, v_k) = \sum_{i=1}^{k} (v_i - V_i)^2 \qquad\qquad (9)$$

Histogram Constraint In addition to penalizing the bin counts, we can enforce standard distance measures for normalized histograms. To do this, we need k indicator functions f_i for every bin and an additional function $f_{k+1}(x) = 1$ encoding the size of the whole segment. Given a target histogram $q_1, \ldots, q_k$ the regional functional is defined:

$$F(v_1, \ldots, v_{k+1}) = -\log\left(\sum_{i=1}^{k} \sqrt{\frac{v_i q_i}{v_{k+1}}}\right) \qquad\qquad (10)$$

for the *Bhattacharyya* distance or

$$F(v_1, \ldots, v_{k+1}) = \sum_{i=1}^{k} \frac{v_i}{v_{k+1}} \log\left(\frac{v_i}{v_{k+1} q_i}\right) \qquad\qquad (11)$$

for the *Kullback-Leibler* divergence.

We will show in Section 3.3 how regional functionals can be approximated linearly. It turns out that the linear approximation can be translated into a unary potential functional. In particular, for all the given examples we will also provide the exact Taylor approximation in terms of these unary potentials.

3   Line-Search Cuts

Below we describe a general optimization method for segmentation energies (2) with arbitrary non-linear regional functionals $R(S)$ defined in (7). Section 3.1 reviews standard *line-search* techniques [17]. Section 3.2 shows how a *global line-search* for energy (2) can be approximated by minimizing a Lagrangian of (2) using efficient parametric max-flow techniques [15, 16]. At each iteration, our Lagrangian uses a linear approximation of $R(S)$ detailed in Section 3.3. Many general properties of our greedy descent algorithm are summarized in Section 3.4.

(a) *Line-search* in the space of segments S (b) Distance between segments
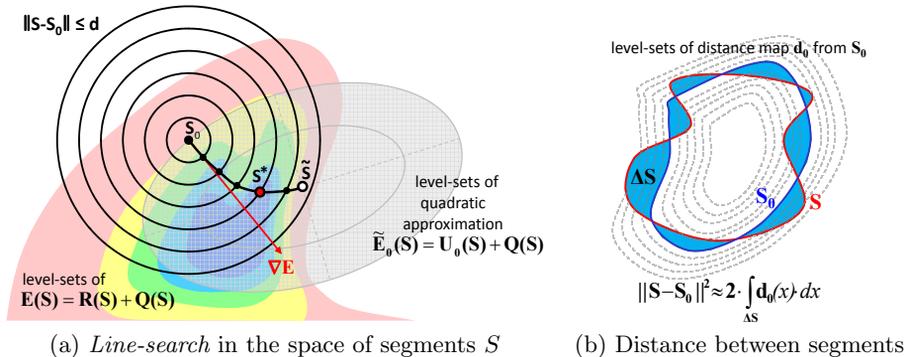
Fig. 1. Overview of *line-search cuts*. At each iteration, the algorithm finds global minimum S^* for energy $E(S)$ along a curve approximating the line of gradient descent ∇E from current solution S_0. This follows the *exact (global) line-search* principle [17].

3.1   Standard *line-search* techniques

Standard iterative *line-search* methods for minimizing a function explore the direction of gradient descent at a current solution and use different strategies for selecting a step size (see [17]). For example, consider energy function $E(S)$ and current solution S_0 illustrated in Fig. 1(a). Commonly used *backtracking line-search* makes a fixed size step in the direction of the gradient descent ∇E. If the energy goes up, then the algorithm backtracks and a smaller step size is tested. In case where the whole line can be explored efficiently, *exact line-search* [17] can be used to jump to the global minimum of energy $E(S)$ on the line. In this paper we refer to this specific technique as *global line-search*.

3.2   *Global line-search* in the space of segments

We will explore the *global line-search* idea to develop a "large move" optimization algorithm avoiding weak local minima. It is possible to traverse the gradient descent line explicitly either by scaling a gradient-flow vector field for contour S_0 or by scaling an *embedding function* within the *level-sets* framework. However, it is unclear at what intervals to probe the points on this line: large intervals can miss a good solution, while probing at small intervals could be too slow. Instead, we propose to explore some implicitly defined "line" closely approximating the gradient descent direction (see black curve in Fig. 1), which can be fully explored by parametric max-flow techniques automatically computing all intervals.

We define *line* as one-dimensional set $\{S(d)|d\geq0\}$ parametrized by scalar d

$$S(d) = \underset{||S-S_0||<d}{\operatorname{argmin}} \tilde{E}_0(S), \qquad \text{for} \qquad \tilde{E}_0(S) := U_0(S) + Q(S) \qquad (12)$$

where each $S(d)$ is a minimum of a quadratic approximation of energy $E(S)$ within distance d from S_0 and $U_0(S)$ is a first-order Taylor approximation of non-linear regional functional $R(S)$ near S_0, as detailed in Section 3.3. In case of discrete labelings S set $\{S(d)|d\geq0\}$ is finite. We call it a line by analogy with

the case of continuous S. Following [19], higher grid resolution makes finite set $\{S(d)|d \geq 0\}$ a denser approximation of a continuous line in the space of segments.

Extending ideas in [18, 16], line $\{S(d)|d \geq 0\}$ can be re-parameterized as $\{S(t)|t \geq t_0\}$ via unconstrained Lagrangian $L_t(S)$ and time parameter t

$$S(t) = \operatorname{argmin} L_t(S), \quad \text{for} \quad L_t(S) := \tilde{E}_0(S) + \frac{1}{2(t - t_0)} \cdot ||S - S_0||^2. \quad (13)$$

The L_2 distance $||S - S_0||^2$ in (13) can be approximated by a linear term

$$||S - S_0||^2 \approx 2 \cdot \int_{\Delta S} d_0(x) \cdot dx = 2 \cdot \int_{\Omega} d_0'(x)(1_S - 1_{S_0}) \cdot dx \quad (14)$$

using distance map d_0 for segment S_0 or its signed version d_0', see [18] and Fig. 1(b). Since Lagrangian (13) has only linear and quadratic submodular terms, all solutions $S(t)$ can be efficiently explored by parametric max-flow methods [15, 16]. Our *global line-search* selects the best solution S^* with respect to original energy E in (2); see Fig. 1(a).

This approach can make "large moves". However, the only guarantee at convergence is that we have a local minimum of the original energy $E(S)$. Indeed, since all S on our curve are minima of Lagrangian (13) then for any $t > t_0$

$$0 = \frac{d\tilde{E}_0}{dS} + \frac{S - S_0}{t - t_0} \quad \underset{t \to t_0}{\Longrightarrow} \quad \frac{\partial S}{\partial t} = -\frac{dU_0}{dS} - \frac{dQ}{dS} = -\frac{dR}{dS} - \frac{dQ}{dS} = -\frac{dE}{dS}.$$

The second part is true since U_0 is a first-order approximation of R at S_0; see Sec. 3.3. Thus, our "line" follows gradient flow of $E(S)$ near t_0. At convergence $S = S_0$ for all $t \geq t_0$. Thus, $\frac{\partial S}{\partial t} = 0$ and S_0 is a local minimum where $\frac{dE}{dS} = 0$.

Our approach combines the benefits of "large moves" and local gradient descent. Quadratic approximation energy $\tilde{E}_0(S)$ monotonically decreases along our "line" from S_0 to its global minimum at $\tilde{S}$; see Fig. 1(a). But, true energy $E(S)$ is guaranteed to decrease along line $S(t)$ only near t_0 where U_0 approximates R sufficiently well. Values of $E(S)$ can go up or down for larger $t \gg t_0$.

3.3 Approximating Regional Functionals for Line-Search Cuts

To apply *line-search cut* to energy $E = R + Q$ in (2) with non-linear regional term R we need approximation $\tilde{E}_0 = U_0 + Q$, see (12). Functional R in (7) can be approximated with linear term U_0 of type (6) using Taylor expansion. One way to derive such first-order approximation is to relax $R(S)$ by replacing S with real-valued function $u : \Omega \to [0; 1]$. The Gateâux derivative of $R(u)$ is

$$\frac{\partial R}{\partial u} = \frac{\partial R^F_{\{f_1, \dots, f_k\}}}{\partial u} = \sum_{i=1}^{k} \frac{\partial F}{\partial v_i} \cdot f_i \quad (15)$$

and the Taylor approximation $\widetilde{R}$ of R developed at u_0 becomes

$$\widetilde{R}^F_{\{f_i\}_{i=1}^k}(u) = R^F_{\{f_i\}_{i=1}^k}(u_0) + \sum_{i=1}^{k} \underbrace{\frac{\partial F}{\partial v_i} (\langle f_1, u_0 \rangle, \dots, \langle f_k, u_0 \rangle)}_{(*)} \cdot \langle f_i, u - u_0 \rangle. \quad (16)$$

Constraint $R(S)$	Scalar Function $g(x)$ for Linear Approximation $U_0(S)$
Volume (8)	$2\left(\langle f_1, S_0 \rangle - V_1\right)$
Bin Count (9)	$2\sum_{i=1}^{k}\left(\langle f_i, S_0 \rangle - V_i\right) \cdot f_i(x)$
Histogram distance *Bhattacharyya* (10)	$\dfrac{\sum_{i=1}^{k}\sqrt{\frac{\langle f_i, S_0 \rangle q_i}{\langle 1, S_0 \rangle^3}} - \sqrt{\frac{q_i}{\langle 1, S_0 \rangle \langle f_i, S_0 \rangle}}\, f_i(x)}{2\sum_{i=1}^{k}\sqrt{\frac{\langle f_i, S_0 \rangle q_i}{\langle 1, S_0 \rangle}}}$
Histogram distance *Kullback-Leibler* (11)	$\sum_{i=1}^{k}\left[\log\left(\frac{\langle f_i, S_0 \rangle}{\langle 1, S_0 \rangle q_i}\right) + 1\right] \cdot \left[\frac{f_i(x)}{\langle 1, S_0 \rangle} - \frac{\langle f_i, S_0 \rangle}{\langle 1, S_0 \rangle^2}\right]$

Table 1. Scalar function $g(x)$ defining first-order approximation $U_0(S)$ in (19) for several examples of non-linear regional functional $R(S)$ presented in Section 2.1.

Calling the expression $(*)$ as $\frac{\partial F}{\partial v_i}(u_0)$, (16) can be written as

$$\widetilde{R}^F_{\{f_i\}_{i=1}^k}(u) = \underbrace{R^F_{\{f_i\}_{i=1}^k}(u_0) - \sum_{i=1}^{k}\frac{\partial F}{\partial v_i}(u_0)\,\langle f_i, u_0 \rangle}_{\text{constant wrt. } u} + \underbrace{\left\langle \sum_{i=1}^{k}\frac{\partial F}{\partial v_i}(u_0)f_i, u \right\rangle}_{\text{linear term}}. \quad (17)$$

The same expression can also be derived without referring to this relaxation approach. In that case, one has to consider $R(S)$ as a functional from the $\mathbb{F}_2$-vector space of subsets into $\mathbb{R}$. For simplicity, we derived this expression only for the relaxed version of the energy. The analog linear approximation for the set functional $R^F_{\{f_i\}_{i=1}^k}(S)$ developed at S_0 then becomes

$$\widetilde{R}^F_{\{f_i\}_{i=1}^k}(S) = \underbrace{R^F_{\{f_i\}_{i=1}^k}(S_0) - \sum_{i=1}^{k}\frac{\partial F}{\partial v_i}(S_0)\,\langle f_i, S_0 \rangle}_{\text{constant wrt. } S} + \underbrace{\left\langle \sum_{i=1}^{k}\frac{\partial F}{\partial v_i}(S_0)f_i, S \right\rangle}_{\text{linear term}}. \quad (18)$$

During energy minimization, we can ignore the constant term of (18). Denoting $g(x) := \sum_{i=1}^{k}\frac{\partial F}{\partial v_i}(S_0)f_i(x)$, the linear approximation of R can be written as

$$\widetilde{R}^F_{\{f_i\}_{i=1}^k}(S) = \int_{\Omega} g \cdot 1_S\, \mathrm{d}x + \text{const} = U_0(S) \quad (19)$$

Thus, the first-order Taylor approximation of our general regional functional $R(S)$ is a linear functional defined by some scalar function $g(x)$. This linear approximation U_0 in (19) can be integrated into Lagrangian (13) and minimized by parametric max-flow methods, as detailed in Section 3.2. Table 1 specifies $g(x)$ corresponding to different examples of regional constraints mentioned in Section 2.1.

(a) "Manhattan" L_1 metric (4-neighborhood)

(b) "Octagonal" metric (8-neighborhood)

(b) Near-Euclidean metric (16-neighborhood)
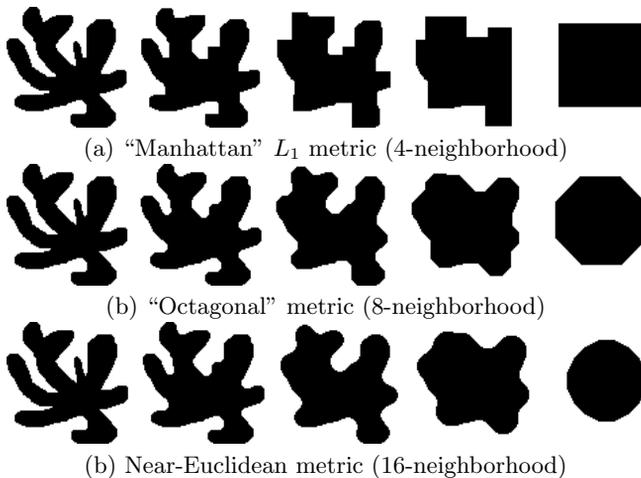
Fig. 2. Synthetic examples: *line-search cuts* can generate a "flow" of a 2D contour under energy E combining quadratic volume constraint (8) and boundary length for Manhattan (a), octagonal (b), and near-Euclidean (c) homogeneous metrics. Target volume is set to the initial volume of the shape. The last result is at convergence.

3.4 Properties of line-search cuts

Our *line-search cuts* method is a *second-order descent* method related to Newton's technique. However, there are two important differences. First, our quadratic functional $\tilde{E} = U_0 + Q$ is not a proper second-order Taylor approximation for $E = R + Q$ since U_0 is only a first-order term for R. Note that a second-order term for R could be non-submodular for general non-linear regional functionals and parametric max-flow [15] may not apply[5]. Second, Newton's descent finds global minimum $\tilde{S}$ for approximation $\tilde{E}$ while we find the global minimum of exact energy E on a line between S_0 and $\tilde{S}$; see Fig. 1(a).

The *line search* framework is dual to the classical *trust region* approach to optimization. The *trust region* problem [17] is defined as minimization of the second-order Taylor approximation of the energy over some ball around S_0 where the approximation is "trusted"; see Fig. 1(a). The ball's radius d is a fixed *step size* parameter. The trust region approach is a conservative variant of Newton's descent which is often too greedy, particularly when current solution S_0 is too far from a local minimum of the true energy. The constrained optimization defining a point on our line (12) is a standard formulation of the trust region problem except that $\tilde{E}$ is not exactly the 2nd-order Taylor approximation. Note that smaller radii d in (12) yield steps approximating the gradient descent well. We refer to an algorithm making steps of some fixed small size d as the *local line-search*. In contrast, our *global line-search* proposed in Section 3.2 finds the optimal step size giving the largest decrease of true energy E along the line. Such large moves can escape many weak local minima, as shown in Figs. 4-6.

---

[5] Some approaches with weaker guarantees, e.g. using QPBO, might still be possible.
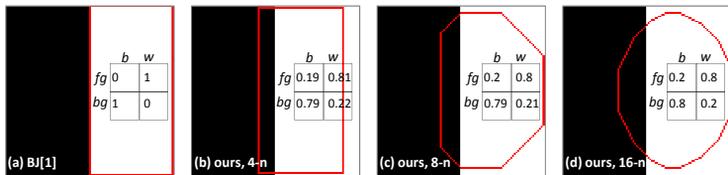
Fig. 3. Synthetic example: *line-search cuts* can match object and background intensity distributions to target distributions $(b, w)_{fg} = (0.2, 0.8)$ and $(b, w)_{bg} = (0.8, 0.2)$ by minimizing a combination of KL distance (11) and segmentation boundary length for Manhattan (b), octagonal (c), and near-Euclidean (d) homogeneous metrics. The result in (a) is the global minimum for Boykov-Jolly segmentation model with the standard linear log-likelihood term (3) using the same target distributions. When maximizing unary likelihoods (3) each pixel independently selects the most likely label regardless of the overall segment appearance. We use (a) to initialize *line-search cuts* in (b-d).

Two earlier methods applicable to general regional functionals use Newton's descent [12] and attempt to approximate trust region [10]. When minimizing approximation $\tilde{E}$ the method in [10] does not enforce the standard trust region constraint $||S - S_0|| < d$ which is critical for making our line follow the gradient descent direction; see Fig. 1(a). Instead, they combine $\tilde{E}$ with a *heuristic* linear term to damp the Newton's steps used in [12]. In contrast to us, their steps could be in the direction opposite to the gradient descent. Consequently, [10] often converges to solutions far from local minima[6] of E; see Fig. 8. In our tests, the proposed line-search approach outperformed [10] with respect to energy values and proximity to the ground truth at convergence; see Figs. 2-4,7.

For example, simple synthetic tests in Fig. 2 correspond to contour energy E combining quadratic volume constraint (8) and geometric boundary length under various grid approximations [19]. All local minima of E are "circles" of a certain radius. Our algorithm converges to such "circles" while [10] stops far away from a minimum for E; see Fig. 8(a). Similarly, Figs. 3(b-d) show our convergence results for segmentation energy E combining boundary length and KL constraint (11). The algorithm is initialized by the global optimum (Fig. 3(a)) of the Boykov-Jolly segmentation energy E (3), combining length and log-likelihoods for the same target distributions. Clearly, the non-linear regional term corresponding to KL distance works much better to enforce the desired intensity distributions inside the segments. The convergence result for [10] is shown in Fig. 8(b).

4   Applications

Below, we apply our method to segmentation of natural and medical images. Similarly to [15, 11], in all experiments the target intensity/color distributions for the object and the background were obtained from the ground truth segments.

---

[6] We refer to *local minima* in the continuous L_2 sense. [10] converges to local minima wrt. certain discrete moves which could be weaker than gradient descent moves.
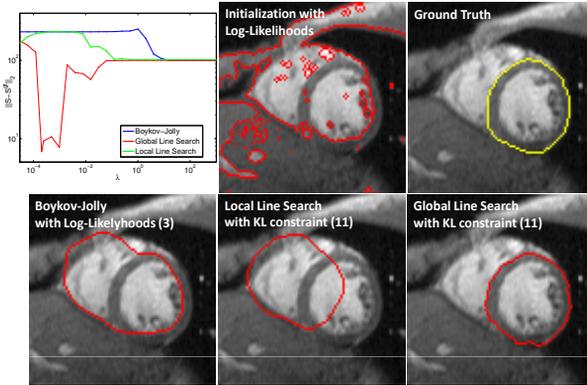
Fig. 4. Segmentation results for an MR image of a heart. Maximizing unary log-likelihoods (3) suffers from the same bias as in Fig. 3(a). Adding length-based regularization in Boykov-Jolly segmentation does not alleviate the problems of (3). The local-line search using more powerful KL constraint (11) gets stuck in a local minimum, whereas the global line-search makes large moves overcoming many weak local minima.
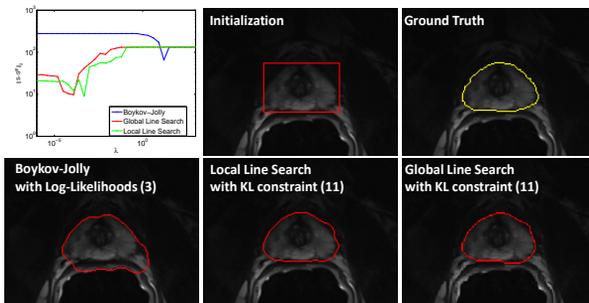
Fig. 5. Prostate midgland segmentation on T2W MRI, where fast and accurate segmentation is important to the planning of prostate cancer therapy. Both local and global line-search may obtain similarly good solutions for KL constraint (11), but the latter works for a wider range of λ.

In many applications, particularly in the field of medical imaging, such target distributions could be pre-learned. While the previous section's synthetic image in Fig. 3 used only two-bins histograms, in this section the histograms have 100 bins for gray-scale medical images in Figs. 4,5,9 and 10^3 bins (10 per channel) for color images in Figs. 6,7. We use the floating point precision in the standard code for graph-cuts [5] and for the parametric max-flow algorithm [15].

Examples in Figs. 4-7 compare the performance of Kullback-Leibler histogram constraint (11) to the linear log-likelihood functional (4) à la Boykov-Jolly [6]. Each of the functionals is combined with the same contrast-sensitive 16-neighborhood boundary regularization term $Q(S) = \lambda \sum_{(p,q)} w_{pq} \cdot \delta(s_p \neq s_q)$ and minimized for different values of smoothness parameter λ. We use global
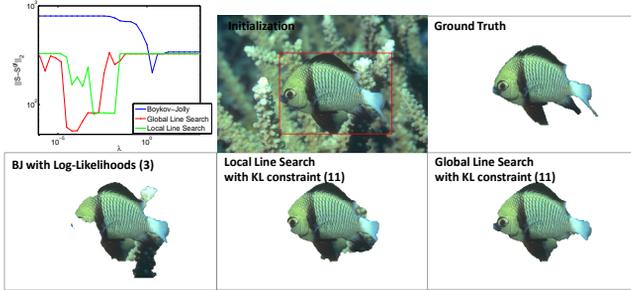
Fig. 6. Results for the "fish" from [7]. Note that Boykov-Jolly is very sensitive to λ.
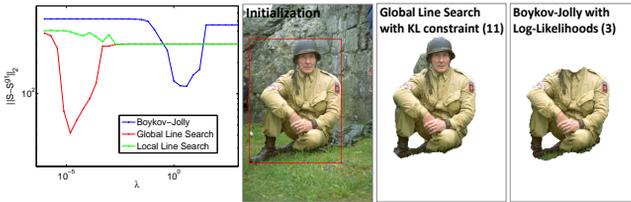
Fig. 7. Results for the "soldier" from [7]. The local-line search fails for all λ.

and local line-search methods (see Sec. 3) to optimize energy (2) with regional functional (11) and standard graph cuts [5] to optimize (1) with log-likelihoods (4). To initialize the line-search methods, we used the bounding boxes in Figs. 5-7 and the "max-likelihood" result in Fig. 4. Quantitative results in the top-left corner of the figures plot the L_2 distance (14) between the resulting contour and the ground truth as a function of the smoothness parameter λ. These plots are in log-scale. We show the results corresponding to the best λ in each plot.

Figs. 4,5 show 2D MRI examples of a heart and a prostate. In both cases the Boykov-Jolly solution is far away from the ground truth since the foreground and background intensity models significantly overlap. In Fig. 4 the local line-search fails to find a good solution for any smoothness parameter λ, while the global line-search finds reliable solutions for a wide range of λ in both Figs. 4 and 5.

Figs. 6,7 show examples of natural images from [7]. The results demonstrate the benefits of minimizing regional functional (11) over linear log-likelihoods (4). Note that the global line-search outperforms the local-line search over a wide range of smoothness parameter λ.

Figure 9 shows our segmentation of a carotid artery in a 2D ultrasound image. We compare two histogram constraints: Kullback-Leibler (11) and Bhattacharyya (10). We use the global line-search and evaluate the results against the ground truth for each smoothness parameter λ. We show separate experiments for segmenting the outer wall (left) and the lumen (right). The bottom row shows the results obtained with the best λ. The segmentation results using KL and Bhattacharyya are similar, while the latter is more "robust" with respect to the smoothness parameter.
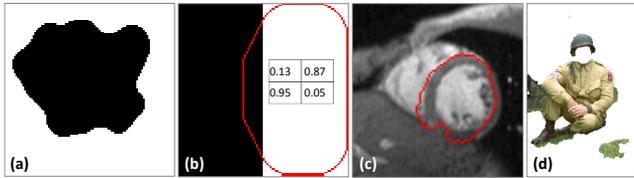
Fig. 8. Results of [10] at convergence on examples in Figures 2-4,7. Unlike our method, the algorithms in [10, 12] may stop at solutions that are not local minima of $E(S)$.
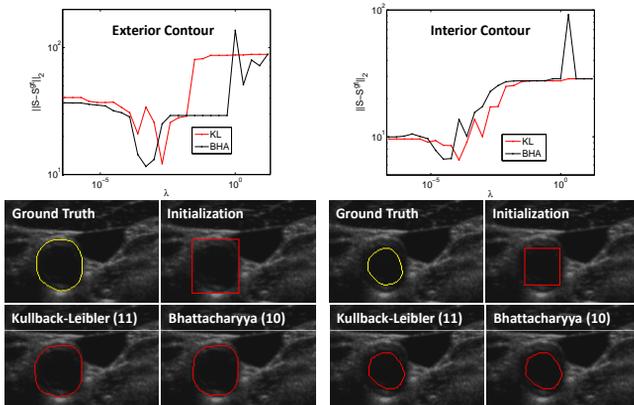
Fig. 9. Comparing the performance of Kullback-Leibler histogram constraint (11) (red) and Bhattacharyya histogram constraint (10) (black) in the task of 2D ultrasound carotid segmentation, with the outer wall shown on the left and lumen - on the right.

5   Conclusions and Future Work

We introduced a general class of non-linear *regional functionals* and a powerful *line-search cuts* optimization framework that can make large moves escaping many weak local minima. We showed examples enforcing quadratic constraints on segment volume, as well as KL and Bhattacharyya constraints on segment color distributions. There are other interesting examples of non-linear regional functionals that can be formulated as (7). In the future we plan to study shape prior for image segmentation. In particular, we observe that *geometric shape moments* [20] fall within our general class of non-linear regional functionals (7).

As mentioned in Section 3.2, our general line-search approach to optimizing non-linear regional functionals can be implemented using convex TV-based optimization methods [1–3, 16]. We plan to test such continuous variants of the technique in Section 3.2 as they should alleviate possible metrication artifacts of the discrete max-flow algorithms, see Figs. 2,3.

We also wish to explore other line-search strategies (e.g. backtracking). They may perform better than the *local line-search* while avoiding the computational burden of an exhaustive *global line-search*. Finally, it is possible to generalize our line-search technique from binary to multi-label segmentation problems.

References

1. Chan, T., Esedoḡlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Appl. Math **66** (2006) 1632–1648
2. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition. Number 3757, Springer (2005) 136–152
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. of Math. Imaging and Vision **40** (2011) 120–145
4. Kolmogorov, V., Zabih, R.: What Energy Functions Can Be Optimized via Graph Cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) **26** (2004) 147–159
5. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) **29** (2004) 1124–1137
6. Boykov, Y., Jolly, M.P.: Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images. In: Int. Conf. on Computer Vision (ICCV). (2001)
7. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. In: ACM SIGGRAPH. (2004)
8. Vese, L.A., Chan, T.F.: A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. Int. Journal of Computer Vision (IJCV) **50** (2002) 271–293
9. Freedman, D., Zhang, T.: Active contours for tracking distributions. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **13** (2004)
10. Rother, C., Kolmogorov, V., Minka, T., Blake, A.: Cosegmentation of Image Pairs by Histogram Matching - Incorporating a Global Constraint into MRFs. In: Computer Vision and Pattern Recognition (CVPR). (2006)
11. Ayed, I., Chen, H., Punithakumar, K., Ross, I., Shuo, L.: Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the Bhattacharyya measure. In: Computer Vision and Pattern Recognition (CVPR). (2010)
12. Kim, J., Kolmogorov, V., Zabih, R.: Visual correspondence using energy minimization and mutual information. In: Int. Conf. on Comp. Vision (ICCV). (2003)
13. Werner, T.: High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). In: Computer Vision and Pattern Recognition (CVPR). (2008)
14. Woodford, O.J., Rother, C., Kolmogorov, V.: A global perspective on MAP inference for low-level vision. In: Int. Conf. on Computer Vision (ICCV). (2009)
15. Kolmogorov, V., Boykov, Y., Rother, C.: Applications of Parametric Maxflow in Computer Vision. In: Int. Conf. on Computer Vision (ICCV). (2007)
16. Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. Int. Journal of Comp. Vision (IJCV) **84** (2009)
17. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge Univ. Press (2004)
18. Boykov, Y., Kolmogorov, V., Cremers, D., Delong, A.: An Integral Solution to Surface Evolution PDEs via Geo-Cuts. ECCV, LNCS 3953 **3** (2006) 409–422
19. Boykov, Y., Kolmogorov, V.: Computing Geodesics and Minimal Surfaces via Graph Cuts. In: Int. Conf. on Computer Vision (ICCV). (2003)
20. Gorelick, L., Galun, M., Sharon, E., Basri, R., Brandt, A.: Shape Representation and Classification Using the Poisson Equation. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) **28** (2006)