Globally Optimal Segmentation of Multi-Region Objects

Andrew Delong
University of Western Ontario

Yuri Boykov
University of Western Ontario

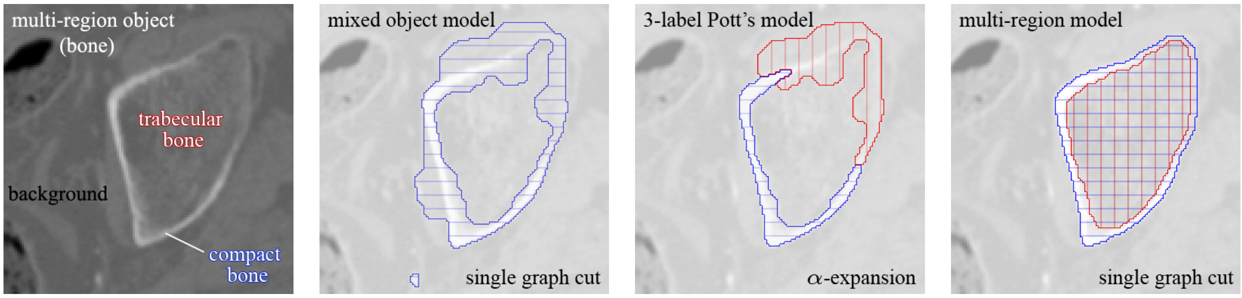


Figure 1. Our simplest motivating example. Standard binary [3, 22] and multi-label [5, 24] models fail because object/background colours are hard to separate. In the absence of user localization, above at center is the best result we can expect from such models. Now we can design multi-region models with geometric interactions to segment such objects more robustly in a single graph cut.

Abstract

Many objects contain spatially distinct regions, each with a unique colour/texture model. Mixture models ignore the spatial distribution of colours within an object, and thus cannot distinguish between coherent parts versus randomly distributed colours. We show how to encode geometric interactions between distinct region+boundary models, such as regions being interior/exterior to each other along with preferred distances between their boundaries. With a single graph cut, our method extracts only those multi-region objects that satisfy such a combined model. We show applications in medical segmentation and scene layout estimation. Unlike Li et al. [17] we do not need "domain unwrapping" nor do we have topological limits on shapes.

1. Introduction

State-of-the-art segmentation methods benefit from an appearance model of the object's interior and its boundary. Such methods include active contours, level sets, graph cuts, and random walker. With binary segmentation, the object's entire appearance must be incorporated into a single mixed model. Most real-world objects are better described by a combination of regions with distinct appearance models, and attempts to use multi-label segmentation reflect this, e.g. [9, 24]. Our new multi-region segmentation framework maintains a separate region+boundary model for each part of an object, and allows these parts to interact spatially.

Figure 1 shows the most basic type of object that we can deal with effectively, and suggests the main advantage we have over standard binary or Pott's-like models.

Our work is a few short steps from a number of existing techniques either from a conceptual or technical point of view. For example, what we call a multi-*region* model is ultimately a multi-label model, though we add simple yet important geometric constraints and then optimize with a single graph cut[1]. To help make our contribution clear, we begin by situating our work relative to other methods.

Pictorial structures. We briefly juxtapose our work with the well-known *pictorial structures* [6], not because our work is directly related, but because we address an analogous problem for objects of a completely different sort. Like their work, our models guarantee optimality only under certain conditions. The table below contrasts our works.

	pictorial struct. [6]	**this work**
shape of each part	fixed template	arbitrary region
spatial prior	relative part positions	boundary distances
optimization	dynamic programming	single graph cut
optimum guaranteed	if tree connectivity	if no "frustrated cycles"

Here "arbitrary region" means that each region does not itself have a specific preferred shape. Such part models can be good, or very bad, depending on the application. One can think of this work as introducing basic distance priors between shapes in a globally optimal way, though incorporating shape priors [29] themselves could be powerful.

[1] Our ideas may also apply in other optimization settings, e.g. [1, 20].

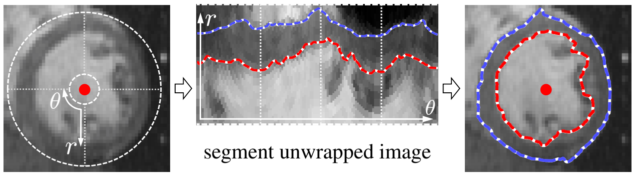


Figure 2. To segment an image, Li et al. [17] must work within a band that already follows the object's rough shape by estimating from a center-line/point. They then 'unwrap' the band into polar coordinates because their construction (Figure 3) requires it.

Multi-label segmentation. Our multi-region models are, generally speaking, a type of multi-label model. One superficial distinction is that an n-region model potentially has 2^n corresponding labels. The reason will be apparent from our graph construction, and we discuss a related idea called *log transformation* [21] toward the end of the paper.

Our first contribution, stated in terms of multi-label models, is to introduce priors on the distance between pairs of discontinuities (or "region boundaries" as we call them). This is achieved by certain long-range interactions between pixels, and stands in contrast to Pott's or random walk models, applied for example in [24] and [9] respectively.

Second, multi-label models often require approximate methods such as α-expansion [5]. We strive for an intuitive characterization of the conditions under which our models can be optimized by a single graph cut. A fully general characterization of when multi-label global optima are guaranteed [25] does not have a meaningful interpretation for specific problems. Elegant interpretations do exist for special cases however, such as Ishikawa's convex characterization [11]. Rather than testing multi-label models against abstract criteria [21, 25], we describe one way to design easy-to-optimize models in an intuitive piecewise manner.

Optimal nested surfaces. The multi-surface segmentation method of Li, Wu, Chen & Sonka [17] is actually what inspired our work. The main drawback of their method is that it is hard to use on anything except cylindrical objects; topological changes, bifurcations, or even strong curvature all require careful pre-segmentation. Figure 2 shows the underlying problem: their need to *unwrap* the image domain.

They start by assuming that a center-point (center-line) of an object in 2D (3D) is given. After casting outward rays and unwrapping them to obtain a polar representation of the image, they can segment multiple nested surfaces along the resulting columns. They model the segmentation as a *closure set* problem on a special graph, but our Figure 3 suggests an equivalent s-t min cut construction for the simplest case. They can encode a minimum and maximum distance constraint between consecutive surfaces. This all assumes that each surface intersects each ray at only one location. Their construction should also allow for soft spring-like forces, although they do not state this.

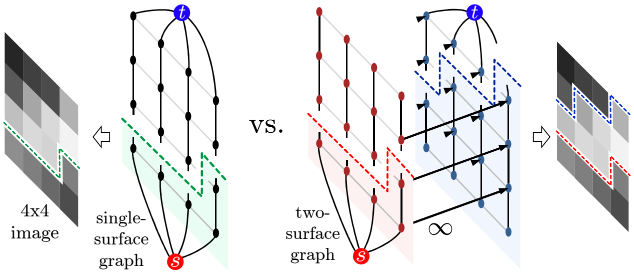


Figure 3. LEFT: s-t min cut construction corresponding to [17]; any cut must separate top row from bottom row. RIGHT: Basic idea from [17]. Each column separates top from bottom at two distinct locations, one forced to be strictly above the other.

Our graph construction sidesteps the unwrapping issue entirely. We do not need center-lines, have no topological constraints, and do not suffer from geometric distortion introduced by unwrapping. Briefly, our construction represents a multi-region object by a directed graph comprising an unordered set of layers, with one layer per region. Each layer has one vertex per image pixel[2]. Each layer by itself is just an independent binary graph cut problem familiar in binary segmentation [3]. We introduce inter-layer arcs in the graph that give effects analogous to [17] yet are easier to implement and useful in more general settings.

The paper is organized as follows. Section 2 introduces our multi-region segmentation framework, describing our energy, geometric interaction terms, and our regional terms. Section 3 demonstrates two applications: medical segmentation and scene layout estimation. Certain combinations of geometric interactions cannot be optimized by graph cuts, and Section 4 discusses ways to handle these cases. Section 5 concludes and suggests further applications.

2. Our Multi-Region Framework

We begin by describing three intuitive geometric interactions in their simplest form:

Containment. Region B must be inside region A, perhaps with repulsion force between boundaries.

Exclusion. Regions A and B cannot overlap at any pixel, perhaps with repulsion force between boundaries.

Attraction. Penalize the area $A - B$, exterior to B, by some cost $\alpha > 0$ per unit area. Thus A will prefer not to grow too far beyond the boundary of Y.

As suggested above, we can introduce a distance prior between region boundaries in the form of a hard or soft margin. The prior is enforced in the graph construction by an inter-layer neighbourhood at each pixel p. The local

[2] This assumption serves to make our notation more bearable. In general, the layers may represent an image at different resolution, matching the scale at which the corresponding part's features appear in the data.

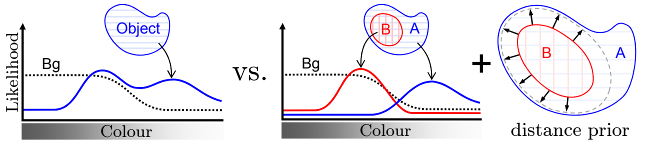


Figure 4. LEFT: Mixed colour model corresponding to Figure 1. RIGHT: Two-region model corresponding to the final result in Figure 1. Trabecular bone (B) is forced to be inside a band of compact bone (A) of some estimated thickness.

weight and shape for this neighbourhood can vary at each pixel. Figure 4 shows how these interactions combine to add discriminative power to object models in segmentation. In other words, these interactions combine to help fight the camouflage problem.

2.1. Multi-Region Energy

We define $\mathcal{P}$ to be the set of pixel indices and $\mathcal{L}$ to be the set of region indices. Our binary variables are $\mathbf{x} \in \mathbb{B}^{\mathcal{L}\times\mathcal{P}}$ which we index as $\mathbf{x}_p^i$ over pixels $p \in \mathcal{P}$ and over regions $i \in \mathcal{L}$. The set $\mathcal{L}$ is *not* ordered. For now we interpret $\mathbf{x}_p^i = 1$ to mean that pixel p is interior to region i. The notation $\mathbf{x}_p$ denotes a vector of *all* variables that correspond to pixel p, one for each of the $|\mathcal{L}|$ regions. If $\mathbf{x}_p = \mathbf{0}$ then pixel p is considered "background."

To express our multi-region energy, we start with two familiar components: data terms and regularization terms. Each pixel p has associated function D_p that defines a cost for every *combination* of regions. Each region i is regularized independently in a standard way by a collection of smoothness terms V^i defined as

$$V^i(\mathbf{x}^i) = \sum_{pq\in\mathcal{N}^i} V^i_{pq}(\mathbf{x}^i_p, \mathbf{x}^i_q) \qquad (1)$$

where each neighbourhood $\mathcal{N}^i$ typically defines nearest-neighbour grid connectivity.

Ideally each data cost $D_p(\mathbf{x}_p)$ could be arbitrary but, because D_p is a function of $|\mathcal{L}|$ binary variables, graph cuts requires that D_p be *submodular* [15]. Ramifications of this are discussed in Section 2.3. Each V^i plays the the same surface-regularization role as in standard binary segmentation. For the case $|\mathcal{L}|=1$ our D_p and V^i obviously describe a standard binary energy, solvable by graph cut [3].

When $\mathcal{L}$ indexes multiple regions, we can add a new category of energy terms to encode inter-region interactions. Our multi-region energy takes the overall form

$$E(\mathbf{x}) = \sum_{p\in\mathcal{P}} D_p(\mathbf{x}_p) + \sum_{i\in\mathcal{L}} V^i(\mathbf{x}^i) + \overbrace{\sum_{\substack{i,j\in\mathcal{L}\\ i\neq j}} W^{ij}(\mathbf{x}^i, \mathbf{x}^j)}^{\text{interaction terms}}. \quad (2)$$

where each W^{ij} encodes all geometric interactions between regions i and j.

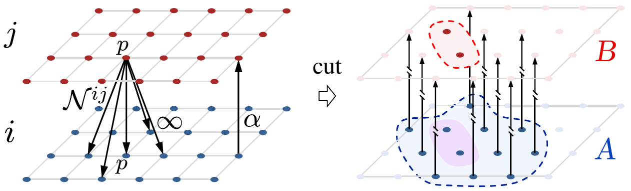


Figure 5. LEFT: Graph construction for region layers $i, j \in \mathcal{L}$ showing a subset of inter-region connectivity $\mathcal{N}^{ij}$. The ∞-cost arcs, shown emanating only from x_j^p, enforce a 1-pixel margin between region boundaries. RIGHT: The α-cost arcs attract the outer boundary by penalizing only the area $A - B$.

To understand how our interaction terms W^{ij} are indexed over both region pairs (i, j) and pixel pairs (p, q), it helps to consider Figure 5 along with the definition for one particular pair of regions

$$W^{ij}(\mathbf{x}^i, \mathbf{x}^j) = \sum_{pq\in\mathcal{N}^{ij}} W^{ij}_{pq}(\mathbf{x}^i_p, \mathbf{x}^j_q). \qquad (3)$$

The inter-region neighbourhood $\mathcal{N}^{ij}$ is the set of all pixels pairs (p, q) at which region i is assigned some geometric interaction with region j. We allow $(p, p) \in \mathcal{N}^{ij}$ because they refer to separate variables, unlike in $\mathcal{N}^i$. Note that W^{ii} and V^i would describe the same set of energy terms, but the conceptual distinction is just as important as the distinction between V_{pp} and D_p.

Section 2.2 details the energy terms and corresponding graph construction for our containment, exclusion, and attraction interactions. Section 2.3 then discusses limitations of our higher-order data terms.

2.2. Geometric Interactions

We now describe how our geometric interactions can be implemented with a single graph cut. The basic "i contains j" interaction is simplest, so we start there. All we do is introduce a term $W^{ij}_{pp}(0,1) = \infty$ at every pixel $p \in \mathcal{P}$. Those familiar with graph constructions may prefer to think of it as an ∞-cost arc from vertex x^j_p to x^i_p, thus prohibiting any cut that labels them 1 and 0 respectively. More generally we can add similar terms W^{ij}_{pq} for $p \neq q$. For example, to add a hard uniform margin to our containment constraint, we set $W^{ij}_{pq}(0,1) = \infty$ for all q within some radius of p.

The tables below list energy terms corresponding to our three main interactions.

i **contains** j		
$\mathbf{x}^i_p$	$\mathbf{x}^j_q$	W^{ij}_{pq}
0	0	0
0	1	∞
1	0	0
1	1	0

i **excludes** j		
$\mathbf{x}^i_p$	$\mathbf{x}^j_q$	W^{ij}_{pq}
0	0	0
0	1	0
1	0	0
1	1	∞

i **attracts** j		
$\mathbf{x}^i_p$	$\mathbf{x}^j_p$	W^{ij}_{pp}
0	0	0
0	1	0
1	0	α
1	1	0

(4)

Figure 5 shows the graph construction corresponding to the containment and attraction interactions. A soft containment cost $W^{ij}_{pq}(0,1) > 0$ for $p \neq q$ creates a spring-like repulsion force between the inner and outer boundaries. Note that our distinction between "containment" and "attraction" is largely artificial since they are the same type of constraint but with opposite orientation.

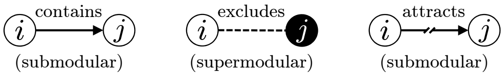
The exclusion interaction is more difficult because it cannot be optimized by graph cuts until we perform a simple transformation. The reason is because graph cuts can only optimize certain *submodular* functions [15]. A function $E(\mathbf{x})$ over binary $\mathbf{x}$ is submodular if it can be expressed as a sum of pairwise functions $E_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ that each satisfy

$$E_{ij}(0,0) + E_{ij}(1,1) \leq E_{ij}(0,1) + E_{ij}(1,0). \quad (5)$$

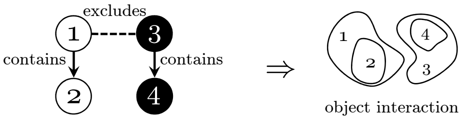
Our containment and attraction interactions are submodular, but for our exclusion terms W^{ij} in (4) clearly the reverse inequality holds, so exclusion is *super*modular. Because exclusion is everywhere supermodular, we can flip the meaning of layer j's variables so that $\mathbf{x}_p^j = 0$ designates the region's interior. Our exclusion terms $W^{ij}(\mathbf{x}^i, \bar{\mathbf{x}}^j)$ thus become submodular, so long as we can flip the variables.

The idea of flipping variable meanings among supermodular terms is not a new idea. It lies at the heart of *roof-duality* methods in quadratic pseudo-boolean optimization (QPBO) [2, 14, 23]. These methods are more sophisticated than graph cuts, consuming more time and memory, so we prefer not to rely on them unless necessary (Section 4).

Let us now explore the overall geometric interactions permitted by combining the three basic ones in (4). To aid the discussion, we introduce graphical depictions of each interaction between two objects i and j.

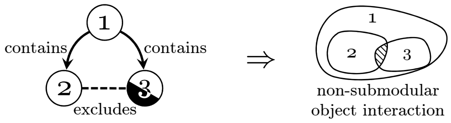


We can allow more sophisticated interactions, such as a hierarchy of nested regions or regions excluded from one another. The example below models two mutually exclusive regions, each with an interior part. A black circle indicates that the region's label is complemented in order for the overall problem to remain submodular.



(6)

There are many useful interactions that we cannot model with graph cuts. The example below describes two mutually exclusive regions, both contained within another region.



(7)

The above configuration cannot be trivially converted to a submodular energy. It introduces what is called a *frustrated cycle* among the overall pairwise energy terms. A cycle is called frustrated if it contains an odd number of non-submodular terms (see $\mathcal{P}3, \mathcal{P}4$ in [23]). This means that with graph cuts we can only model interactions that are bipartite with respect to exclusion, and submodular interactions cannot be added between layers that use opposite $0/1$ labels. If we step outside these constraints then global optima are no longer guaranteed, but approximations such as QPBO-I [23] or $\alpha\beta$-swap [5] may still be effective. (Section 4 explains why α-expansion often cannot be applied.)

2.3. Regional Data Terms

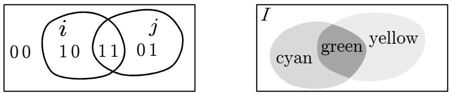
We begin by showing how the likelihoods in Figure 4 are used to drive the segmentation in Figure 1. We have $\mathcal{L} = \{A, B\}$ so each data term D_p defines up to 4 costs. Given image data I, each function D_p is described by the table below.

$\mathbf{x}_p^A$	$\mathbf{x}_p^B$	D_p
0	0	$-\log \Pr(Bg \mid I_p)$
0	1	K
1	0	$-\log \Pr(A \mid I_p)$
1	1	$-\log \Pr(B \mid I_p)$

(8)

The unspecified cost K brings us to an important point. The cost K is not driven by the image data itself, because the "A contains B" object model prohibits this configuration. For this particular model, each $D_p(\mathbf{x}_p)$ is added alongside pairwise term W_{pp}^{AB} having cost $W_{pp}^{AB}(0,1) = \infty$. The three likelihoods (8) can therefore be arbitrary for this object model, without concern for K or for submodularity. Submodularity of our overall energy (2) thus depends on a *combination* of data terms and interaction terms.

Suppose however that there were no geometric constraints between two layers i and j. The data terms $D_p(\mathbf{x}_p)$ must then be submodular (or supermodular if the label for j is flipped). To understand what this means intuitively, consider two regions i and j that represent subtractive colours.



(9)

Here, submodularity requires that each data term satisfy

$$D_p(0,0) + D_p(1,1) \leq D_p(0,1) + D_p(1,0). \quad (10)$$

One symmetric way to satisfy (10) is to say, for example, that D_p for a cyan pixel I_p does not simply encourage region i, but also *discourages* region j by an equal amount.

For models with strong geometric interactions, such as containment and exclusion, these constraints on D_p are usually satisfied for reasons suggested by (8).

Higher-order data terms. D_p may model three or more regions with dependent data costs, but graph cuts can only encode pairwise energy terms directly. Any function of three or more variables can be transformed into a combination of pairwise and unary terms in polynomial time [2]. Transforming a submodular 3rd-order term preserves submodularity among the resulting pairwise terms [15]. For a 4th-order term or higher there are submodularity-preserving

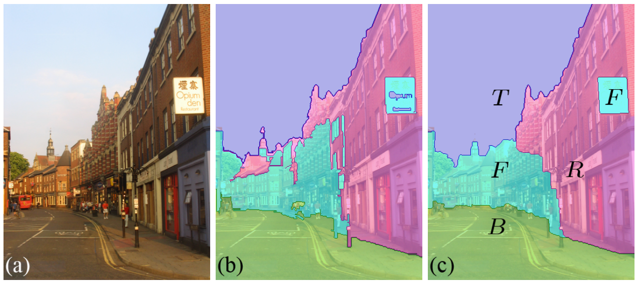


Figure 6. Scene layout estimation. Given a scene (a) we first generate data terms from local surface class confidences given by Hoiem et al. [10]. The maximum likelihood solution is shown in (b). With a single graph cut, our multi-region framework regularizes noise/gaps in the data (c) while keeping most important geometric classes (B, L, T, R, F) mutually exclusive throughout the image.

transformations only for certain cases [7, 30]. To solve the resulting pairwise problem with a single graph cut, one must truncate the non-submodular data terms to approximate the desired energy. (None of our medical examples needed truncation.) An alternative is to use QPBO [2] and its extensions [23] directly on the non-submodular energy.

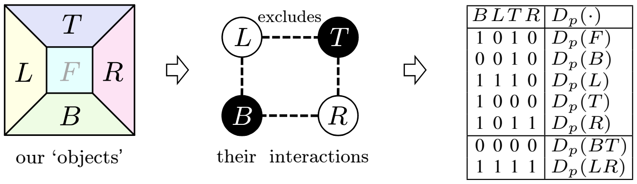
3. Applications

We choose two problems that we hope demonstrate the diverse applications of our framework. Section 3.1 shows how our multi-region energy (2) helps to model many objects in medical image segmentation. Section 3.2 proposes a novel way to regularize basic scene layout estimation using Hoiem-style[3] data terms [10].

3.1. Medical Segmentation

Medical image segmentation is a domain full of multi-part objects that are hard to detect with rigid part-models such as [6]. This is why so many state-of-the-art algorithms [1, 3, 9, 17] rely on region+boundary models over arbitrary shapes using mainly length/area priors. Of these techniques, only the recent work of Li, Wu, Chen & Sonka [17] attempts to globally optimize priors on the distance between multiple surfaces. As shown in Figures 2 and 3, they rely on accurate center-line estimation (a difficult problem in itself) and cannot handle complex topologies.

Figures 1, 8, 9 and 10 show experimental results of our multi-region framework using class-specific models (bone, knee, heart, kidney). The heart result was computed using QPBO-I, and the rest were computed in a single graph cut. Our early experiments are all 2D but they extend to N-D in a straight-forward manner. Using the Boykov-Kolmogorov max-flow algorithm [4] our running times are longer than binary graph cut in roughly linear proportion to the number of vertices and arcs added to the graph.



$B\,L\,T\,R$	$D_p(\cdot)$
1 0 1 0	$D_p(F)$
0 0 1 0	$D_p(B)$
1 1 1 0	$D_p(L)$
1 0 0 0	$D_p(T)$
1 0 1 1	$D_p(R)$
0 0 0 0	$D_p(BT)$
1 1 1 1	$D_p(LR)$

Figure 7. Our scene object interactions and corresponding higher-order data terms. Two unwanted labels are due to limitations imposed by frustrated cycles (Section 2.2). The configurations not listed have cost $\geq \infty$ due to the four exclusion constraints above.

3.2. Scene Layout Estimation

Given a photograph of a scene, we wish to break the image into rough geometric labels "bottom" (B), "top" (T), "left wall" (L), "right wall" (R) and "front-facing" (F). This application is described by Hoiem et al. [10], and we actually use data terms based on their local geometric class estimators. See Figure 6 for an example result. Instead of using α-expansion to find a local minimum of a Pott's energy, we design a set of interactions between class regions that can be optimized by a single graph cut.

In our setup, we let regions $\mathcal{L} = \{B, L, T, R\}$ and treat F as background. Ideally we want every pixel $p \in \mathcal{P}$ to be assigned a unique region, but adding this constraint introduces frustrated cycles (Section 2.2). We propose the subset of interactions and data terms portrayed in Figure 7.

To encourage the "box" layout seen in Figure 7 we borrow an idea from [18] and bias region B against cutting underneath itself using length terms V^B, and likewise for orientations L, T, R. Unlike [18] we do this with a soft penalty so that strong local data terms can override the prior, such as the front-facing sign in Figure 6c.

We still have two unwanted configurations BT and LR that have no corresponding likelihood. To discourage these labels we want to maximize corresponding D_p, but higher-order submodularity requires

$$\begin{aligned} D_p(BT) &\leq D_p(B) + D_p(T) - D_p(F), \text{ and} \\ D_p(LR) &\leq D_p(L) + D_p(R) - D_p(F). \end{aligned} \tag{11}$$

We truncate these terms to retain submodularity, potentially allowing either B to overlap T, or L to overlap R. Experimental results are shown in Figure 11.

Note that even if we did prohibit labels BT and LR, we would not be minimizing a Pott's energy. Instead, the equivalent multi-label formulation has labels $\mathcal{L} = \{l_\emptyset, l_1, \ldots, l_n\}$ where we designate $l_\emptyset$ the *null* label, corresponding to region F in our scene layout formulation. In this type of multi-label model, all pixel label pairs $f_p, f_q \neq l_\emptyset$ have

$$V_{pq}(f_p, f_q) = V_{pq}(f_p, l_\emptyset) + V_{pq}(l_\emptyset, f_q). \tag{12}$$

Because this model always penalizes (l_i, l_j) transitions more than $(l_i, l_\emptyset)$ transitions, over-smoothing creates gaps between l_i and l_j in regions with weak data, unlike [10, 18].

[3] We thank Derek Hoiem so, *so* much for making code [10] available.

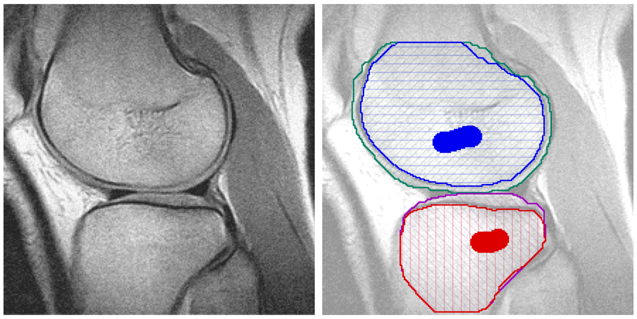


Figure 8. User-driven segmentation of knee joint, measuring thickness of cartilage. Above uses the 4-part submodular interaction portrayed in (6), and was computed by a single graph cut. Given the user seeds, bone and cartilage are segmented automatically using a combination of image gradients and anisotropic distance prior (margin) between surfaces. A two-part model, using these same seeds for either tibia or femur, gives poor results.

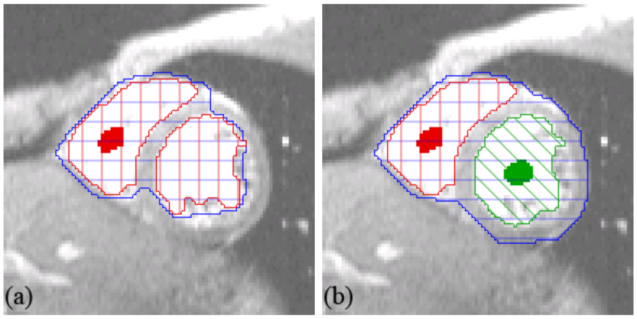


Figure 9. User-driven heart segmentation using the non-submodular interaction portrayed in (7), solved by QPBO-I. The user first adds seeds to mark the right ventricle (a), but the sampled colour model is attracted to both ventricles. The user then marks the left ventricle as a separate region (b). The outer wall is segmented automatically by compromising between image gradients and distance prior (margin). This cannot be done by Li et al. [17].

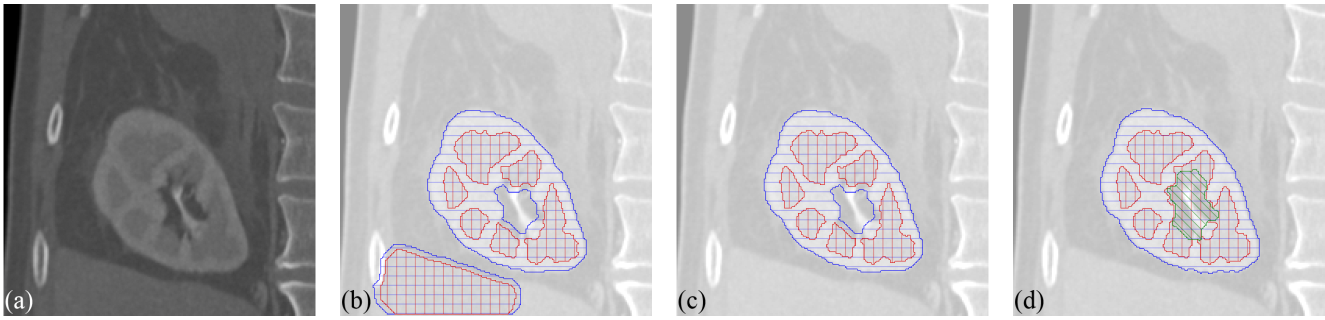


Figure 10. Kidney segmentation (a) is very difficult to automate due to low contrast and complex topology. Binary graph cuts simply cannot get reasonable results without heavy user interaction, and even multi-label methods need some form of localization [24, 9]. In (b–c) we model the kidney as medulla surrounded by a slightly brighter cortex of minimum thickness. On this challenging example our method is very sensitive to colour/geometric parameters, e.g. (b), but has discriminative power to extract only the correct object (c) without any localization. We also show an alternate 3-region object model (d) that eliminates the unwanted margin between medulla and collection cavity (dark/bright interior). This kind of topology would be impossible to segment using Li et al. [17] due to the unwrapping requirement.

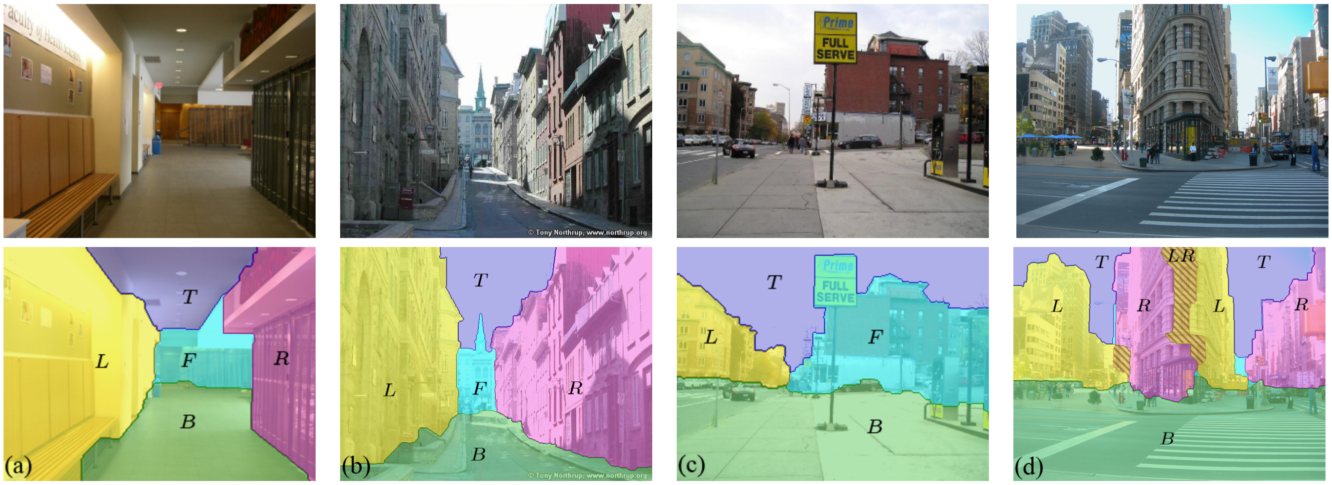
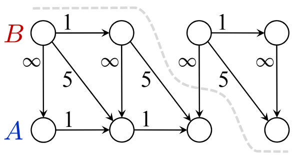


Figure 11. Scene layout results using our proposed interactions in Section 3.2, showing estimates for indoor (a) and outdoor (b–d) scenes. Smoothness parameters were tuned for each image. Diagonal shading on the Flatiron image (d) indicates that scene classes L and R overlap. This may happen when certain data terms conflict because our graph cut construction cannot simultaneously prohibit all classes from overlapping. Section 4 discusses ways to resolve this. See Figure 6b for an example of the data terms that drive this segmentation.



	costs $D_p(\cdot)$				labeling $\hat{f}$				$E(\hat{f})$
B	0	0	0	∞	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	3
A	4	0	0	∞	A	A	A	$\emptyset$	4
$\emptyset$	3	0	0	0	B	B	$\emptyset$	$\emptyset$	6

Figure 12. Example of how our interaction terms cause $\alpha\beta$-swap to get stuck in local minima. The graph and D_p encode a 4-pixel segmentation with "A contains B" constraint. Diagonal arcs encourage a 1-pixel margin between boundaries of B and A. Our s-t min cut construction finds global optimum $f^* = (B, B, A, \emptyset)$ with $E(f^*) = 0$, but the corresponding 3-label energy is hard for $\alpha\beta$-swap to optimize. The initial labeling $f_p = \emptyset$ is already a local minimum regardless of which labels are swapped (table at right).

4. Discussion

Given one of our multi-region models, one could apply $\alpha\beta$-swap to the corresponding multi-label energy. Unfortunately this provides no optimality guarantees, and Figure 12 suggests how our distance priors create local minima for $\alpha\beta$-swap. Often the α-expansion algorithm cannot even be applied because the equivalent multi-label energy is not a *metric* [5] and would create non-submodular terms at the expansion step. Specifically, let $V_{pq}(f_p, f_q)$ denote the pairwise cost corresponding to Figure 12. The costs here do not satisfy the triangle inequality because

$$V_{pq}(B, \emptyset) \nleq V_{pq}(B, A) + V_{pq}(A, \emptyset). \qquad (13)$$

The Pott's-like model suggested by (12) *is* a metric, however, and can be optimized effectively with α-expansion. On the few scene layout examples we tried, α-expansion either found or came close to the global optimum.

Multi-label constructions. Recall that our set of regions $\mathcal{L}$ is not ordered in any way. We are thus *not* building a 'layer cake' construction typical of discrete and continuous total-variation methods in multi-label optimization [11, 19, 20]. A special case of our multi-region energy (2) does coincide with a particular Ishikawa construction [11]. To construct a total-variation ($V_{pq} \propto |f_p - f_q|$) Ishikawa graph for n labels, order $n-1$ regions as $\mathcal{L} = \{1, \ldots, n-1\}$ and introduce hard "i contains $i+1$" constraints between subsequent layers.

Also recall that an n-region model represents up to 2^n corresponding labels, which is the ultimate objective of the *log transformation* [21]. They start with an energy over discrete variables $x_i \in \{1 \ldots m\}$ and try to represent each x_i using as close to $\log_2 m$ binary variables as possible. Their approach is much more general because they start from a multi-label energy and test it against a criterion for transformation to submodular binary encoding. The criterion itself is clear but it is not always obvious how to satisfy it when designing an energy for a particular application. In contrast, we *start* with binary variables and build up our multi-region models from intuitive pairwise interactions. We show that there are applications where such models are useful, without the need for an explicit transformation from multi-label.

Constructions along 'rays'. On page 2 we described a related construction by Li, Wu, Chen & Sonka [17] that optimizes along columns sampled from the image domain. Notice that because their columns are known *a priori* they can encode both a min and max distance prior, whereas our framework assumes rays are not known and can only encode a min distance[4]. Thus there is an advantage to their method when a good pre-segmentation is available.

On the subject of paper [17], we mention two connections between their work and existing works in vision. First, it is standard to convert their closure set problem into an equivalent s-t min cut, and we note that the corresponding min cut graph in their single-surface case happens to be a particular Ishikawa construction [11]. Their innovation can be thought of as building parallel Ishikawa constructions that influence one another. Second, there is a binary segmentation paper [27] that takes similar advantage of rays embedded in the image domain. Rather than unwrap the image domain and introduce geometric distortion of length/area, Veksler discretizes the rays and embeds them directly in the neighbourhood of a grid graph. One could implement multi-surface priors like Li et al. by extending Veksler's grid framework instead.

QPBO and approximations. There are many multi-region models that are useful yet contain frustrated cycles. Even the simple 3-region interaction portrayed in (7) and the scene layout application are two examples where the ideal set of interactions cannot be optimized with a single graph cut. We can still formulate the (potentially NP-hard) energy and apply global methods like QPBO-P [2] or a reasonably fast approximation like QPBO-I [23]. QPBO-I can give good results on examples like Figures 9 and 11d, in only 1–5 subsequent 'improve' attempts.

Given a model that contains frustrated cycles among region layers, it may also be possible to design move-making algorithms that operate on subsets of regions. This is in the spirit of "range-moves" [16, 26] where at each iteration we choose a large subset of interactions that can be trivially converted to submodular. Care must be taken to ensure that the energy of the labeling never increases, but application-specific moves can be developed in this way. For example, we have verified that we can implement the vertical/horizontal moves in [18] using a simple "L excludes R" construction with special D_p and V^i based on the current labeling.

[4] In our framework, it is actually possible to create a spring-like attraction force between boundaries of i and j via opposing "i attracts j" and "j attracts i" interactions of large radius. However, the strength of this attraction is unfortunately coupled with surface regularization strength, leading to unwanted oversmoothing for most applications.

5. Conclusions and Future Work

With our multi-region framework, not only can more difficult objects now be segmented, but designing tractable models is also quite easy. The main ideas were to keep a separate appearance model for each spatially distinct region, and to allow geometric priors between region boundaries. Along the way, we discussed many parallels between the works of Li et al. [17], Ishikawa [11] and Veksler [27], and we hope these comments were helpful. Our experiments suggest that more robust medical segmentation tools could be designed around these ideas.

There are many other applications that can potential be revisited with these ideas in mind. Particularly promising are a more sophisticated concept of shape priors [8, 29] and topological constraints [28], but also ratio minimization [13], EM-style algorithms like Grab-Cuts [22], and combining pictorial structures with segmentation [6, 12]. Complex objects can be modeled by a hierarchy of nested regions that interact, with each region potentially driven by different data.

Finally, we note that there has been much past success in transferring ideas from discrete optimization into continuous settings, e.g. [20]. We hypothesize that some of the ideas discussed in this paper may also apply in continuous settings.

Acknowledgements We thank Milan Sonka and Xiaodong Wu of the University of Iowa for interesting discussions regarding their related work. This work was supported by NSERC (Canada) Discovery Grant R3584A02 and in part by a NSERC PGS-D.

References

[1] B. Appleton and H. Talbot. Globally Minimal Surfaces by Continuous Maximal Flows. *IEEE TPAMI*, 28(1):106–118, 2006. 1, 5

[2] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Appl. Math.*, 123(1-3):155–225, 2002. 4, 5, 7

[3] Y. Boykov and M.-P. Jolly. Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images. *ICCV*, 1:105–112, 2001. 1, 2, 3, 5

[4] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE TPAMI*, 29(9):1124–1137, 2004. 5

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE TPAMI*, 23(11):1222–1239, 2001. 1, 2, 4, 7

[6] P. F. Felzenszwalb and D. Huttenlocher. Pictorial Structures for Object Recognition. *IJCV*, 61(1):55–79, 2005. 1, 5, 8

[7] D. Freedman and P. Drineas. Energy minimization via graph cuts: settling what is possible. In *CVPR*, 2005. 5

[8] D. Freedman and T. Zhang. Interactive Graph Cut Based Segmentation With Shape Priors. In *CVPR*, 2005. 8

[9] L. Grady. Multilabel Random Walker Image Segmentation Using Prior Models. In *CVPR*, June 2005. 1, 2, 5, 6

[10] D. Hoiem, A. A. Efros, and M. Hebert. Recovering Surface Layout from an Image. *IJCV*, 75(1), October 2007. 5

[11] H. Ishikawa. Exact Optimization for Markov Random Fields with Convex Priors. *IEEE TPAMI*, 25(10), 2003. 2, 7, 8

[12] P. Kohli, J. Rihan, M. Bray, and P. H. S. Torr. Simultaneous Segmentation and Pose Estimation of Humans Using Dynamic Graph Cuts. *IJCV*, 79(3):285–298, Sept 2008. 8

[13] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of Parametric Maxflow in Computer Vision. In *ICCV*, November 2007. 8

[14] V. Kolmogorov and C. Rother. Minimizing non-submodular functions with graph cuts—a review. *IEEE TPAMI*, 29(7), 2007. 4

[15] V. Kolmogorov and R. Zabih. What Energy Functions Can Be Optimized via Graph Cuts. *IEEE TPAMI*, 26(2):147–159, 2004. 3, 4

[16] P. Kumar and P. H. S. Torr. Improved Moves for Truncated Convex Models. In *NIPS*, volume 22, 2008. 7

[17] K. Li, X. Wu, D. Z. Chen, and M. Sonka. Optimal Surface Segmentation in Volumetric Images—A Graph-Theoretic Approach. *IEEE TPAMI*, 28(1), 2006. 1, 2, 5, 6, 7, 8

[18] X. Liu, O. Veksler, and J. Samarabandu. Graph Cut with Ordering Constraints on Labels and its Applications. In *CVPR*, June 2008. 5, 7

[19] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A Convex Relaxation Approach for Computing Minimal Partitions. In *CVPR*, June 2009. 7

[20] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A Convex Formulation of Continuous Multi-Label Problems. In *ECCV*, October 2008. 1, 7, 8

[21] S. Ramalingam, P. Kohli, K. Alahari, and P. Torr. Exact Inference in Multi-label CRFs with Higher Order Cliques. In *CVPR*, June 2008. 2, 7

[22] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. In *ACM SIGGRAPH*, 2004. 1, 8

[23] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing Binary MRFs via Extended Roof Duality. In *CVPR*, June 2007. 4, 5, 7

[24] H. Rusinek, Y. Boykov, M. Kaur, S. Wang, L. Bokacheva, J. Sajous, A. Huang, S. Heller, and V. Lee. Performance of an automated segmentation algorithm for 3D MR renography. *Magnetic Resonance in Medicine*, 2006. 1, 2, 6

[25] D. Schlesinger. Exact Solution of Permuted Submodular MinSum Problems. In *EMMCVPR*, 2007. 2

[26] O. Veksler. Graph Cut Based Optimization for MRFs with Truncated Convex Priors. In *CVPR*, June 2007. 7

[27] O. Veksler. Star Shape Prior for Graph-Cut Image Segmentation. In *ECCV*, 2008. 7, 8

[28] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *CVPR*, June 2008. 8

[29] N. Vu and B. Manjunath. Shape Prior Segmentation of Multiple Objects with Graph Cuts. In *CVPR*, 2008. 1, 8

[30] S. Živný, D. A. Cohen, and P. G. Jeavons. The Expressive Power of Binary Submodular Functions. *Discrete Applied Mathematics*, 2009. 5