

Multilabel MRFs with Label Adjacency Constraint: Globally Optimal Solutions and Applications

Nhat Vu · Vignesh Jagadeesh · B.S.Manjunath*

Received: date / Accepted: date

Abstract This work addresses the problem of optimally solving Markov Random Fields (MRFs) in which labels obey a certain topology constraint. Utilizing prior information, such as domain knowledge about the appearance, shape, or spatial configuration of objects in a scene can greatly improve the accuracy of segmentation algorithms in the presence of noise, clutter, and occlusion. Nowhere is this more evident than in the segmentation of biomedical images, where typically the spatial relationships among the image regions inherently reflect those of the anatomical structures being imaged. In this work, we propose a new methodology to segment a special class of images, which exhibit nested layer topologies often encountered in biomedical applications. The segmentation problem is modeled using multi-label Markov Random Fields with an additional label adjacency constraint (LAC). The multi-label MRF energy with LAC is transformed via boolean variables encoding into an equivalent function of binary variables. We show this boolean function is sub-modular, graph representable, and can be minimized exactly and efficiently with graph cut techniques. Our experimental results on both synthetic and real images demonstrate the utility of the proposed LAC segmentation algorithm.

Keywords Markov random fields · graph cuts · nested layer segmentation · label constraint

1 Introduction

The segmentation of a typical image, such as one depicting an office desk or outdoor scene, tends to produce junction points where three or more regions with different labels

*Department of ECE, University of California, Santa Barbara
E-mail: NV (astronhat@yahoo.com), VJ (vignesh@ece.ucsb.edu), BSM (manj@ece.ucsb.edu)
Ph:8058932526
Nhat Vu is now affiliated with Quantcast Inc.

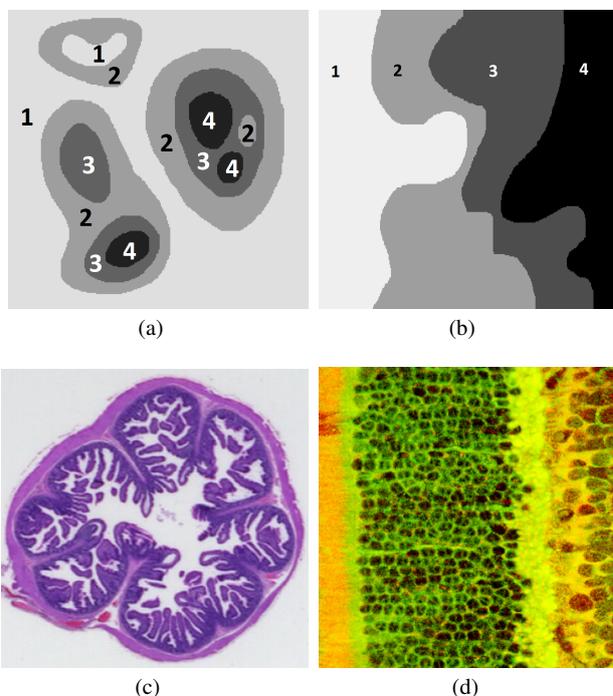


Fig. 1 Images with nested layer topologies. (a,b) Each intensity value in (1-highest, 4-lowest) indicates a distinct label (4 labels total). (c) Image of jejunum cross-section. (d) Confocal image of a mammalian retina.

touch. This is not surprising since there are no restrictions on the spacial layout of the different objects or regions in the scene. However in many biomedical applications, the spatial relationships among the image regions inherently reflect those of the anatomical structures being imaged. For a special subclass of these images, the regions exhibit a nesting relationship where it is possible to partition the image regions such that no junctions exist. That is, given an ordered set of labels, it is possible to assign labels to the image

regions such that adjacent regions have consecutive labels from the set. Such a segmentation prevents junctions where three or more regions with nonconsecutive labels touch.

Images with nested regions are uncommon in natural or uncontrived scenes (Martin et al (2001)), but occur frequently in biomedical datasets. The images in Fig. 1 illustrate some common nested regional relationships often observed in the biomedical image domain. The regions in Fig. 1(a) exhibits an “inside-outside” nesting relationship where one label can be said to encapsulate another label. An the example of this type of layer nesting is shown in Fig. 1(c). Fig. 1(b) shows an example of a “before-after” nesting relationship that is frequently observed in images of tissue cross-sections, such as the skin epidermis or the mammalian retina, e.g. Fig. 1(d). Images with nested layers are not solely restricted to the biomedical domain, but can also be found in other areas, such as the geosciences.

Since segmentation is an ill-posed problem in general, we propose to use the nested layer relationship to improve the segmentation accuracy. Previous works have shown that incorporating domain priors, such as an object’s shape, can dramatically enhance an algorithm’s ability to correctly segment the image (Leventon et al (2000); Rousson and Paragios (2002); Tsai et al (2003)). As our results demonstrate, utilizing the nested layer relationship will greatly improve the segmentation. An additional benefit of using this prior information is that we are able to constrain the segmentation to return only those results that are structurally or anatomically feasible.

1.1 Nested Layer Segmentation Formulation

Given an image with pixel set $\mathcal{P} = \{1, 2, \dots, N\}$ and an ordered label set \mathcal{L} , we seek a labeling \mathbf{y} that minimizes the MRF (Geman and Geman (1984)) energy

$$E(\mathbf{y}) = \sum_{p \in \mathcal{P}} V_p(y_p) + \sum_{p \in \mathcal{P}, q \in \mathcal{N}_p} V_{pq}(y_p, y_q) \quad (1)$$

subject to the Label Adjacency Constraint (LAC) to prevent junctions

$$|y_p - y_q| \leq 1, \quad \forall y_p, y_q \in \mathcal{L}. \quad (2)$$

Here \mathcal{N}_p denotes the set of pixels that are neighbors of p , and without loss of generality, we specify the label set to be $\mathcal{L} = \{1, 2, \dots, K\}$. The term $V_p(y_p)$ is the unary potential associated with labeling p with y_p , and $V_{pq}(y_p, y_q)$ is the pairwise potential associated with labeling neighboring pixel pair p and q with labels y_p and y_q , respectively. Our notation do not explicitly indicate the dependency of the MRF energy on image data, but it is assumed that this dependency exists. Notice condition (2) restricts neighboring pixel pairs

from having labels that differs by more than one. This condition effectively forces the regions in the final segmentation to have a nested relationship. The problem of solving the energy in equation (1) has been extensively researched using graph cuts (Boykov et al (2001); Boykov and Kolmogorov (2004)), message passing (Kolmogorov (2006); Wainwright et al (2005)) and annealing (Geman and Geman (1984)). Our work specifically focusses on optimally solving the energy in equation (1) constrained by LAC in equation (2) using graph cut techniques.

The primary contributions of this work are:

- *Globally Optimal Solution to the Nested Layer Segmentation (NLS) Problem*
- *Memory and Computationally Efficient Graph Construction for NLS*
- *Identification and Experimental Validation of several NLS scenarios*

1.2 Related Works

The work by (Chung and Vese (2005)) is most relevant in terms the problem of segmenting images with nested layers. In (Chung and Vese (2005)), the authors proposed a multilayer level set approach motivated by island dynamics in epitaxial growth. In traditional level set methods, the zero level set is used to embed an evolving front or curve that divides the image into positive and negative regions (two labels). The multilayer level set method uses multiple level lines of a single level set function to represent the boundaries separating regions with adjacent labels. This is analogous to using the contour lines of a topographic map to indicate the elevation levels. Effectively, the different regions are implicitly nested, and this relationship is maintained throughout the level set evolution. An advantage of using the multi-level formulation is that only one level set function is needed to represent multiple (> 2) labels, unlike multiphase approaches, which require more than one level set function and is less efficient (Zhao et al (1996); Vese and Chan (2002)).

Yet, the multilayer level set approach has several limitations. As pointed out by the authors, the algorithm does not guarantee a globally optimal solution and is sensitive to the initialization. Since this method relies on performing gradient descent on an energy functional, these limitations are to be expected. Secondly, the user manually specifies the levels used for embedding the boundaries. While this task may not pose a problem for images with few labels, it may potentially become difficult to choose the right set of values when the number of labels becomes large.

In the graph cut formulation, Ishikawa presented a method to find the exact minimizer of equation (1) given that the label set is ordered and the pairwise clique potential is a

convex function of the label difference (Ishikawa (2003)). While it appears that this framework would be applicable to the nested layer problem, it remains a challenge to find a practical convex function that also satisfies the LAC condition. For example, we can use the convex function

$$V_{pq}(y_p, y_q) = c \cdot f(y_p - y_q) = c \cdot |y_p - y_q|^\gamma, \quad (3)$$

for $\gamma \rightarrow \infty$ and c a quantity that depends on the image data. Yet in the implementation, there are limitations that affect the efficiency and accuracy of the segmentation. The most noticeable issue is the scalability of the graph required for optimization. When $\gamma \geq 2$ in equation (3), the number of edges in the graph is on the order of $\mathcal{O}(NK^2)$, assuming the total number of neighborhood edges is $\mathcal{O}(N)$. Even for modest image sizes (256^2), the graph size can grow quickly as the number of labels increases. As we will show, the graph construction for our method reduces the number of edges to be on the order of $\mathcal{O}(NK)$. A second limitation deals with the numerical overflow of the energy calculation during graph cut optimization. In the graph structure of (Ishikawa (2003)), the edge weights between neighboring nodes are proportional to the second difference of $f(y_p - y_q)$, i.e. $f(y_p - y_q + 1) - 2f(y_p - y_q) + f(y_p - y_q - 1)$, and as a result, the majority of these edge weights have very large values. Most feasible minimum cost cuts will inevitably sever a large number of these edges, and consequently the energy computation suffers from numerical overflow and returns suboptimal results.

Another notable graph cut approach with label constraint was proposed by (Liu et al (2008)), where a pair of order-preserving moves are used for geometric class scene labeling. Their algorithm uses a series of horizontal and vertical moves to iteratively segment the image into five regions: center, top, bottom, left, and right. Though each horizontal or vertical move is optimal, the final result is not guaranteed to be globally optimal. Note also that this approach is not directly applicable to the task of nested layer segmentation.

In essence, our work formally proposes the problem of nested layer segmentation and identifies it as a subclass of problems that can be solved using (Ishikawa (2003)). Subsequently, an efficient graph construction for circumventing memory and overflow issues arising from the construction in (Ishikawa (2003)) is proposed.

Recently, (Felzenszwalb and Veksler (2010)) proposed a tiered scene labeling construction that exactly solves the labeling problem for images having a tiered structure. By their definition of tiered structures, each column has a top, bottom and middle layer. In this work, we impose no such restrictions on the number of labels a column can occupy. It is also known that (Felzenszwalb and Veksler (2010)) slows down considerably as the number of tiers increases. We used the source code provided by (Felzenszwalb and Veksler (2010))

and confirmed the high running times of tiered scene labeling in comparison to our formulation. (Zheng et al (2012)) extend the tiered scene labeling problem to multiple tiers by proposing a fast approximate inference scheme. In contrast, the formulation proposed in this work can solve multiple tiered scene labeling problems that obey LAC exactly! It is useful to note that all visual results demonstrated in (Zheng et al (2012)) obey the LAC.

Further, (Garvin et al (2008)) model intra retinal layer segmentation as a graph search problem. While their work focussed primarily on retinal image segmentation, our work attempts to solve a generic layer constrained segmentation problem using graph cuts.

Finally, while we have identified many more NLS scenarios where the technique proposed in this work has been applied, they are not included in this draft due to space constraints and to primarily focus on the problem formulation. A discussion of other applications considered, along with the source code is downloadable at (<http://vision.ece.ucsb.edu/segmentation/nls/>).

1.3 Paper Outline

In the remainder of this paper, we will show that the globally optimal solution for equation (1) subject to equation (2) can be found efficiently via graph cut techniques. In section 2, we review some graph cut preliminaries and provide an important theorem in section 3, which states the condition for optimizing equation (1) exactly. Then using the boolean encoding scheme described in section 3, we transform the original multi-label MRF energy into an equivalent function of boolean variables in section 4. We show that the resulting boolean energy is submodular, graph representable, and can be minimized exactly and efficiently with graph cut. In section 5, we show experimental results on both synthetic and real images to demonstrate the utility of our algorithm, and briefly describe potential future extensions to this work in section 6. We would like to mention that a preliminary version of this work appears in (Vu (2008)).

2 Optimization with Graph Cuts

Solving for the exact minimizer of the multi-label MRF energy in equation (1) is NP-hard in general (Kolmogorov and Zabih (2004)). However there exist certain forms of the energy function that allow for an exact solution. In this section, we discuss several cases where exact solutions can be found and the subsequent section provides a useful theorem, which states the conditions for exact minimization via graph cut techniques. We start by briefly describing some necessary background on graph cut.

2.1 Graph Cut Preliminaries

A weighted directed graph $G(\mathcal{V}, \mathcal{E})$ is composed of a set of nodes \mathcal{V} and a set of directed edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ connecting the nodes. The node set includes two special terminal nodes, the source s and the sink t , and the remaining nodes are considered neighborhood nodes. A directed edge $(p, q) \in \mathcal{E}$ connects node $p \in \mathcal{V}$ to node $q \in \mathcal{V}$ and has nonnegative capacity or weight w_{pq} . Note that because the graph is directed, $(p, q) \neq (q, p)$.

A subset of edges $\mathcal{E}_c \subset \mathcal{E}$ is called an st-cut if the terminal nodes are completely separated in the induced graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E} - \mathcal{E}_c)$. That is there are no forward paths from source s to sink t when all edges in the cut are removed. Hence, the cut partitions the nodes into disjoint subsets \mathcal{S} and \mathcal{T} where $s \in \mathcal{S}$ and $t \in \mathcal{T}$. In our convention, an edge (p, q) is in the cut if $p \in \mathcal{S}$ and $q \in \mathcal{T}$. For simplicity, we will refer to the st-cut simply as a cut. The cost of the cut is the sum of all the edge weights in \mathcal{E}_c . For a given graph, the minimum cost cut (mincut) can be found by solving an equivalent maximum flow (maxflow) problem (Ford and Fulkerson (1962)).

The goal of using graph cuts for energy minimization is to construct a graph such that there is a one-to-one mapping between cuts in the graph and labelings of the image pixels. Moreover the minimum cut cost should equal to (up to a constant) the minimum energy labeling. As mentioned, not all energy functions are graph representable, but below we discuss several forms that allow for optimization via graph cut.

2.2 Energies with Exact Solutions

There are several unique forms of equation (1) where the exact solutions can be found. The most common case is when $\mathcal{L} = \{0, 1\}$ and the second order potential $V_{pq}(y_p, y_q)$ is submodular (Greig et al (1989); Kolmogorov and Zabih (2004); Freedman and Drineas (2005)). For $|\mathcal{L}| > 2$ and \mathcal{L} an ordered set, Ishikawa gave conditions for the pairwise cost $V_{pq}(y_p, y_q)$ to be exactly minimized (Ishikawa (2003)). Along the same line, using pseudo-boolean optimization (Boros and Hammer (2002)), Schlesinger and Flach also showed that multi-label MRFs with convex energy functions of order two can be minimized exactly in polynomial time (Schlesinger and Flach (2006)). Subsequently, Ramalingam et al. outlined a more extensive set of conditions for exactly minimizing multi-label MRF energies with higher order potentials (Ramalingam et al (2008)). In addition, they described a principled framework for transforming the class of submodular multi-label k th order functions into an equivalent class of submodular second order boolean functions, which can be solved exactly using graph cuts. We will discuss their framework in more detail in section 3.

3 Boolean Encoding of Multi-label Variables

In this section, we then review the boolean transformation techniques presented in (Ramalingam et al (2008)), which we use to transform the LAC formulation in the next section. The key idea in transforming a multi-label function into one of binary variables is to use two or more boolean variables to encode the states of a single multi-label variable. The transformation is accomplished by defining a set of encoding functions, which replace all occurrences of the multi-label variable with that of the encoding boolean variables. For a given multi-label k th order function $E(\mathbf{y})$ with $\mathbf{y} \in \mathcal{Y}$, the transformation $T : \mathcal{Y} \rightarrow \mathcal{Z}$ will result in a boolean k th order function $E_{\text{bin}}(\mathbf{z})$, where \mathbf{z} belongs to the space of boolean labelings \mathcal{Z} . In addition, the transformation must be bijective and the minimum value of $E(\mathbf{y})$ over \mathbf{y} must equal to the minimum value of $E_{\text{bin}}(\mathbf{z})$ over \mathbf{z} . Refer to (Ramalingam et al (2008)) for more details on these conditions.

3.1 Encoding Unary Multi-label Variables

The unary potential in equation (1) can be rewritten as

$$V_p(y_p) = \sum_{i \in \mathcal{L}} \theta_{p;i} \delta(y_p, i), \quad (4)$$

where $\theta_{p;i}$ is the potential for assigning label $y_p = i$ to pixel p , and

$$\delta(y_p, i) = \begin{cases} 1 & \text{if } y_p = i \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

In order to make equation (4) a function of boolean variables, the multi-label terms $\delta(y_p, i)$ should be replaced with boolean functions $f_{y_p;i}(\mathbf{z}_p)$. Here $\mathbf{z}_p = \{z_p^1, z_p^2, \dots, z_p^M\}$, with $z_p^i \in \{0, 1\}$, and M is the number of boolean variables used to encode one multi-label variable.

Schlesinger and Flach (Schlesinger and Flach (2006)) proposed encoding a K -label variable y_p using $K-1$ boolean variables $\{z_p^1, z_p^2, \dots, z_p^{K-1}\}$ such that

$$\{y_p = i\} \leftrightarrow \{z_p^1 z_p^2 \dots z_p^{K-1} = \{\mathbf{0}_{(i-1)} \mathbf{1}_{(K-i)}\}\} \quad (6)$$

where we use the notation

$$\mathbf{0}_{(i-1)} = \underbrace{00 \dots 0}_{i-1} \text{ and } \mathbf{1}_{(K-i)} = \underbrace{11 \dots 1}_{K-i}. \quad (7)$$

It is fairly straightforward to deduce that the function $f_{y_p;i}(\mathbf{z}_p)$ satisfies

$$f_{y_p;i}(\mathbf{z}_p) = \begin{cases} z_p^1, & i = 1 \\ z_p^i - z_p^{i-1}, & 2 \leq i \leq K-1 \\ 1 - z_p^{K-1}, & i = K. \end{cases} \quad (8)$$

A more detailed derivation can be found in (Ramalingam et al (2008)). To carry out the boolean transformation, $f_{y_p;i}(\mathbf{z}_p)$ can be substituted for every instance of $\delta(y_p, i)$ in equation (4), thereby transforming the multi-label function $V_p(y_p)$ into one of boolean variables.

3.2 Encoding Pairwise Multi-label Variables

Similar to the unary potential case, the pairwise potential in equation (1) can be expressed as

$$V_{pq}(y_p, y_q) = \sum_{i,j \in \mathcal{L}} \theta_{pq;i,j} \delta(y_p, i) \delta(y_q, j), \quad (9)$$

where $\theta_{pq;i,j}$ is the potential associated with the pairwise label assignments of $y_p = i$ and $y_q = j$ to neighboring sites p and q , respectively. By substituting equation (8) into equation (9), the pairwise multi-label potential becomes the boolean potential

$$V_{\text{bin}}(\mathbf{z}_p, \mathbf{z}_q) = \sum_{i,j \in \mathcal{L}_{\{-1\}}} \alpha_{ij} z_p^i z_q^j + L_1, \quad (10)$$

with $\mathcal{L}_{\{-1\}} = \{1, 2, \dots, K-1\}$ and the coefficients

$$\alpha_{ij} = \theta_{pq;i,j} - \theta_{pq;(i+1)j} - \theta_{pq;i(j+1)} + \theta_{pq;(i+1)(j+1)}. \quad (11)$$

The term L_1 is the sum of the first order terms and constants. The following theorem states the condition for an exact solution of a second order boolean function via graph cuts. For a proof, refer to (Freedman and Drineas (2005)).

Theorem 1 Let $z_i \in \{0, 1\}$ and let

$E(z_1, \dots, z_n) = \sum_{i,j} a_{ij} z_i z_j + L$, where L represents terms that are linear in z_i plus any constants. Then E can be minimized via graph cut techniques if and only if $a_{ij} \leq 0$ for all i, j .

As a consequent, as long as our boolean transformed energy satisfies Theorem 1, we can use graph cut constructions to exactly solve for the minimum energy labeling. In order to minimize equation (10) exactly the coefficients must satisfy $\alpha_{ij} \leq 0$ (submodular condition).

3.3 Graph Construction for Boolean Encoding

As an example (Ramalingam et al (2008)), for a 4-label variable $y_p \in \mathcal{L} = \{1, 2, 3, 4\}$, the encoding is given by

$$\begin{aligned} \{y_p = 1\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{111\}\} \\ \{y_p = 2\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{011\}\} \\ \{y_p = 3\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{001\}\} \\ \{y_p = 4\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{000\}\}. \end{aligned} \quad (12)$$

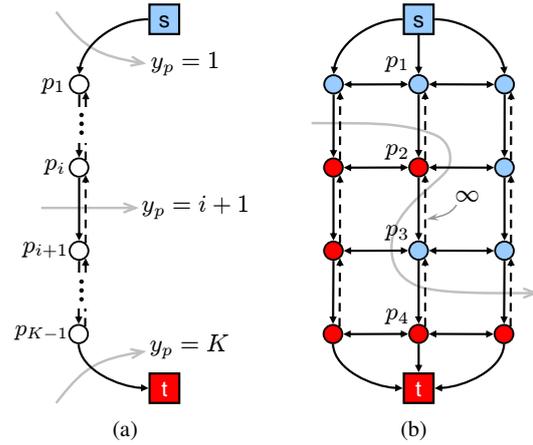


Fig. 2 (a) Graph construction for unary variable encoding. (b) Example of an infeasible cut (gray arrow). Blue nodes belong to the source set \mathcal{S} and red nodes belong to the sink set \mathcal{T} . Edge (p_3, p_2) is in the cut and has infinite weight, making the cut cost prohibitively high.

The graph construction corresponding to this encoding is shown in figure 2(a). Each multi-label variable y_p is encoded by $K-1$ nodes $\{p_1, p_2, \dots, p_{K-1}\}$. Using the convention that, after the cut, $p_i \in \mathcal{S}$ implies $z_p^i = 0$ and $p_i \in \mathcal{T}$ implies $z_p^i = 1$, the cuts corresponding to $y_p = i$ for $i \in \mathcal{L}$ result in the binary labelings in equation (6). Furthermore, to ensure that each cut has a corresponding cost equal to the unary energy in equation (4), the edge weights are assigned as:

$$w_p^{s,1} = \theta_{p,1} \quad (13a)$$

$$w_p^{i,i+1} = \theta_{p;i+1} \quad (13b)$$

$$w_p^{K-1,t} = \theta_{p,K}. \quad (13c)$$

Here $w_p^{s,1}$, $w_p^{i,i+1}$, and $w_p^{K-1,t}$ are the weights of directed edges (s, p_1) , (p_i, p_{i+1}) , and (p_{K-1}, t) , respectively.

Notice that for the above example, the three boolean variables can encode a maximum of $2^3 = 8$ labelings. However, the labelings $z_p^1 z_p^2 z_p^3 = \{010, 100, 101, 110\}$ are unused, and cuts resulting in these labelings must be made infeasible. This is accomplished by adding infinite capacity edges (p_{i+1}, p_i) for $i = 1, 2, \dots, K-2$, which makes the cuts corresponding to the unused labelings have prohibitively high costs. These edges are shown as dashed arrows in figure 2(a). Figure 2(b) shows an example of an infeasible cut, where according to our graph cut convention, the edge (p_3, p_2) is in the cut, and the boolean encoding for y_p is $z_p^1 z_p^2 z_p^3 z_p^4 = \{0101\}$. However, since this edge has infinite weight, such a cut is prevented.

The graph structure for encoding a pair of variables $\{y_p, y_q\}$ for neighboring pixels p and q is dependent on the pairwise potential $\theta_{pq;i,j}$. An example that satisfies the submodularity condition is the potential given in (Ishikawa (2003)), i.e.

$$\theta_{pq;i,j} = c \cdot |y_p - y_q|^\gamma, \text{ for } k > 0. \quad (14)$$

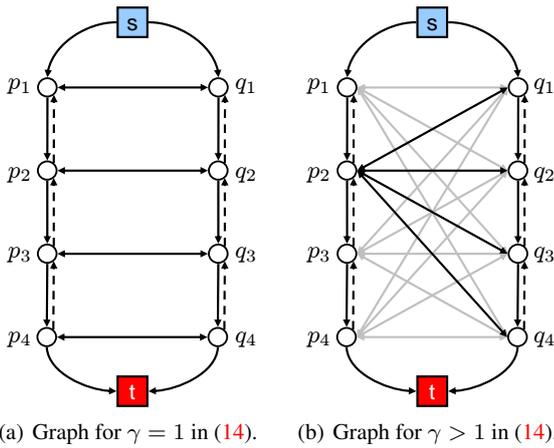


Fig. 3 Graph constructions for the pairwise variable encoding according to (Ishikawa (2003); Schlesinger and Flach (2006)). For simplicity, a bidirectional edge connecting p_i and q_j is used to represent the two directed edges (p_i, q_j) and (q_j, p_i) . The number of edges grows according to $\mathcal{O}(N \cdot K^2)$ for $\gamma > 1$.

Figure 3 shows the graph construction for $\gamma = 1$ (left) and $\gamma > 1$ (right). In each graph, the left and right columns of nodes encode the variables y_p and y_q , respectively. The nodes p_i and q_j are connected via two directed edges (p_i, q_j) and (q_j, p_i) , but for simplicity these edges are represented by a single bidirectional edge. Refer to (Ishikawa (2003)) for more detail on the weight assignments.

The combined graph construction for the unary and pairwise potentials in the above examples allows for an exact solution via graph cuts. With the appropriate weighting assignments, a feasible cut on the graph has a cost equal to the sum of the unary and pairwise potentials of cut edges, and the mincut corresponds to the labeling with lowest energy. We would like to note that other boolean encoding schemes are also possible (Ramalingam et al (2008)), but our results rely on the one presented here.

Proposition 1 *Given a constraint on labels of the form $\mathcal{C}_d : |y_p - y_q| \leq d, \forall p, q \in V, \{y_p, y_q\} \in \mathcal{L}, 1 \leq d \leq |\mathcal{L}| - 1$, the set of edges each with infinite weight $\mathcal{E}_d = \bigcup_{i=1}^{|\mathcal{L}|-1-d} \{(q_{i+d}, p_i), (p_{i+d}, q_i)\}$, guarantees resulting labels adhere to \mathcal{C}_d .*

Proof Sketch: Consider two adjacent sites in the random field p, q , which have edges (p_{i+d}, q_i) and (q_{i+d}, p_i) having infinity costs. We now consider four cases.

Case 1: Cut entering through (p_{i-1}, p_i) and exiting through (q_{j-1}, q_j) , $i + d + 1 \leq j \leq |\mathcal{L}| - 1$

These cuts are infeasible and are prevented by (q_{i+d}, p_i) , see second column Figure 4.

Case 2: Cut entering through (p_{i-1}, p_i) and exiting through (q_{j-1}, q_j) , $i - 1 \leq j \leq i + d$

These are feasible cuts since they do not violate \mathcal{C}_d , see first column in Figure 4.

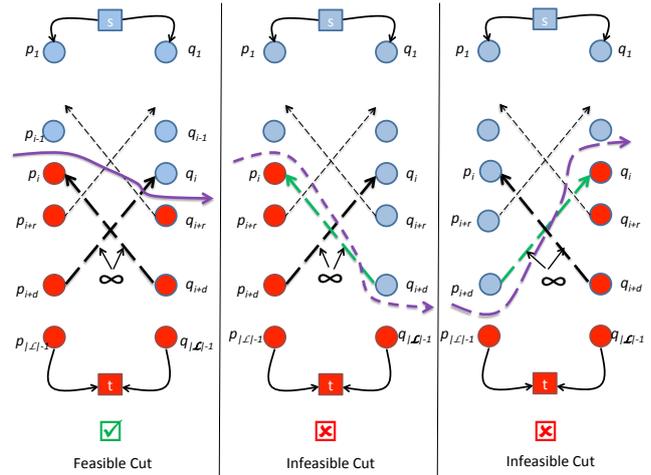


Fig. 4 Illustration of various feasible and infeasible cuts for the generic case for a label constraint of the form $|y_p - y_q| \leq d$. Considering two sites p and q and a specific label assignment i , there are two edges having dark shades (p_{i+d}, q_i) and (q_{i+d}, p_i) that prevent infeasible labelings that violate $|y_p - y_q| \leq d$. The first column shows a feasible cut since both edges with dark shades are not in the cut. However, in the second and third columns one of the dark shaded edge switches color to green if it is in the cut. Since the cost of cuts in the second and third columns are exorbitantly high, they are infeasible solutions of the energy minimization.

Case 3: Cut entering through (p_{i+d}, p_{i+d+1}) and exiting through (q_{j-1}, q_j) , $2 \leq j \leq i - 1$

These cuts are infeasible and are prevented by (p_{i+d}, q_i) , see third column in Figure 4.

Case 4: Cut entering through (p_{i+d}, p_{i+d+1}) and exiting through (q_{j-1}, q_j) , $i - 1 \leq j \leq i + d$

These are feasible cuts since they do not violate \mathcal{C}_d .

Since any cut can be classified into one of the four cases and \mathcal{C}_d is not violated in any of the above cases, thus verifying the proposition.

The above proposition provides the intuition behind generating solutions that adhere to a label adjacency constraint. In this context, the proposed approach can be viewed as a restriction of the original Ishikawa construction. The above proposition though intuitive, does not provide a concrete way to map energies (specifically, the interaction terms while feasible cuts occur) to cuts in the graph. We now turn our attention to a formulation using pseudo-boolean optimization that derives the above proposition and yields a graph construction that exactly maps solutions in the energy to cuts in the graph when $|y_p - y_q| \leq d$, when $d = 1$.

4 Multi-label MRF with Label Adjacency Constraint

We now turn our attention to transforming the LAC in equation (2) for neighboring pairs of multi-label variables y_p and y_q into an equivalent constraint for the corresponding pair of boolean variables z_p and z_q . We will show that the MRF

energy $E(\mathbf{y})$, subject to condition (2), can be minimized exactly when

$$\theta_{pq;ii} - \theta_{pq;(i+1)i} - \theta_{pq;i(i+1)} + \theta_{pq;(i+1)(i+1)} \leq 0, \quad \forall i \in \mathcal{L}. \quad (15)$$

Additionally, we will describe the graph construction, which enforces the label constraint condition and for which the mincut cost yields the minimum energy labeling.

4.1 Boolean Encoding of LAC

Recall that the set of labels is the ordered set $\mathcal{L} = \{1, 2, \dots, K\}$. The constraint $|y_p - y_q| \leq 1$ forces two neighboring sites p and q to have either the same label or consecutive labels from \mathcal{L} . Without loss of generality, assume $y_p \leq y_q$. As an example for $K = 4$, the set of labelings $\{y_p, y_q\} = \{(1, 3), (1, 4), (2, 4)\}$ violates this constraint. The boolean encodings using the scheme in equation (6) for these three cases are:

$$\begin{aligned} \{y_p = 1, y_q = 3\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{111\}, z_q^1 z_q^2 z_q^3 = \{001\}\} \\ \{y_p = 1, y_q = 4\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{111\}, z_q^1 z_q^2 z_q^3 = \{000\}\} \\ \{y_p = 2, y_q = 4\} &\leftrightarrow \{z_p^1 z_p^2 z_p^3 = \{011\}, z_q^1 z_q^2 z_q^3 = \{000\}\}. \end{aligned} \quad (16)$$

Observe that these infeasible labelings all have in common at least one instance where a boolean variable pair $\{z_p^i, z_q^j\} = \{1, 0\}$ for $j \geq i + 1$. In general for any boolean variable pair, this assignment implies that if $y_p = i$ then $y_q > i + 1$. However it is clear that such an assignment violates the LAC.

We can state the LAC for the boolean variables more precisely. Given the boolean encodings $\mathbf{z}_p = \{z_p^1, z_p^2, \dots, z_p^{K-1}\}$ and $\mathbf{z}_q = \{z_q^1, z_q^2, \dots, z_q^{K-1}\}$ for the multi-label variable pair y_p and y_q , the constraint in equation (2) is equivalent to

$$z_p^i \bar{z}_p^j + \bar{z}_p^j z_q^i = 0 \text{ for } i \in \mathcal{L}_{\{-2\}}, j > i, \quad (17)$$

where $\bar{z} = 1 - z$ and $\mathcal{L}_{\{-2\}} = \{1, 2, \dots, K - 2\}$. The penalty for these infeasible pairwise boolean encodings can be expressed as

$$P(\mathbf{z}_p, \mathbf{z}_q) = \sum_{i \in \mathcal{L}_{\{-2\}}, j > i} \lambda (z_p^i \bar{z}_p^j + \bar{z}_p^j z_q^i), \quad (18)$$

where $\lambda \rightarrow \infty$. Combining the boolean transformations for the multi-label MRF energy and the LAC, we arrive at boolean energy

$$E_{\text{bin}}(\mathbf{z}) = \sum_{p \in \mathcal{P}} V_{\text{bin}}(\mathbf{z}_p) + \sum_{p \in \mathcal{P}, q \in \mathcal{N}_p} \left(V_{\text{bin}}(\mathbf{z}_p, \mathbf{z}_q) + P(\mathbf{z}_p, \mathbf{z}_q) \right). \quad (19)$$

4.2 Conditions for Exact Solutions

In order to find the exact global minimum of equation (19), we have to show that the second summation, expressed in the form of equation (10), satisfies Theorem 1. That is, all the second order terms involving $z_p^i z_q^j$ must have coefficients $\alpha_{ij} \leq 0$. Observe that the boolean potential in equation (10) can be reexpressed as

$$\begin{aligned} V_{\text{bin}}(\mathbf{z}_p, \mathbf{z}_q) &= \sum_{i \in \mathcal{L}_{\{-2\}}, j > i} (\alpha_{ij} z_p^i z_q^j + \alpha_{ji} \bar{z}_p^j z_q^i) \\ &\quad + \sum_{i \in \mathcal{L}_{\{-1\}}} \alpha_{ii} z_p^i z_q^i + L_1 \\ &= \sum_{i \in \mathcal{L}_{\{-2\}}, j > i} (\alpha'_{ij} z_p^i \bar{z}_p^j + \alpha'_{ji} \bar{z}_p^j z_q^i) \\ &\quad + \sum_{i \in \mathcal{L}_{\{-1\}}} \alpha_{ii} z_p^i z_q^i + L'_1, \end{aligned} \quad (20)$$

where $\alpha'_{ij} = -\alpha_{ij}$ and

$$L'_1 = L_1 + \sum_{i \in \mathcal{L}_{\{-2\}}, j > i} (\alpha'_{ij} z_p^i + \alpha'_{ji} z_q^j). \quad (21)$$

With the reformulation above, exact minimization of equation (20) requires $\alpha_{ii} \leq 0$, $\alpha'_{ij} \geq 0$, and $\alpha'_{ji} \geq 0$. Exact minimization of L'_1 is guaranteed since it is composed of first order terms and constants (Freedman and Drineas (2005)).

The boolean pairwise potential $V_{\text{bin}}(\mathbf{z}_p, \mathbf{z}_q)$ and the boolean penalty $P(\mathbf{z}_p, \mathbf{z}_q)$ can be combined to give

$$\begin{aligned} V_{\text{bin}}(\mathbf{z}_p, \mathbf{z}_q) + P(\mathbf{z}_p, \mathbf{z}_q) &= \sum_{i \in \mathcal{L}_{\{-2\}}, j > i} \left((\alpha'_{ij} + \lambda) z_p^i \bar{z}_p^j + (\alpha'_{ji} + \lambda) \bar{z}_p^j z_q^i \right) \\ &\quad + \sum_{i \in \mathcal{L}_{\{-1\}}} \alpha_{ii} z_p^i z_q^i + L'_1. \end{aligned} \quad (22)$$

The next section provides a graph construction to ensure the coefficients of $z_p^i \bar{z}_p^j$ and $\bar{z}_p^j z_q^i$ satisfy $(\alpha'_{ij} + \lambda) \geq 0$ and $(\alpha'_{ji} + \lambda) \geq 0$, guaranteeing that the first summation in equation (22) can be minimized exactly. Consequently, the only requirement for the exact minimization of equation (22) is for $\alpha_{ii} \leq 0$, $i \in \mathcal{L}_{\{-1\}}$. In summary, to minimize equation (19), we must define pairwise potentials to ensure

$$\theta_{pq;ii} - \theta_{pq;(i+1)i} - \theta_{pq;i(i+1)} + \theta_{pq;(i+1)(i+1)} \leq 0. \quad (23)$$

4.3 Label Adjacency Constraint Graph

Minimizing equation (22) with st-mincut techniques requires that all occurrences of the label pairs $z_p^i z_q^j = \{10\}$ and $z_p^j z_q^i = \{01\}$, where $i \in \mathcal{L}_{\{-2\}}$ and $j > i$ be made infeasible, since these labelings violate the LAC. Recall that

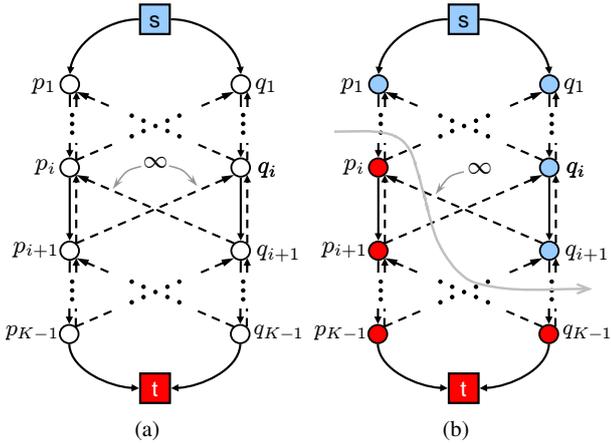


Fig. 5 (a) The constraint edges (p_{i+1}, q_i) and (q_{i+1}, p_i) have infinity weight and enforce the label adjacency condition. (b) Example of an infeasible cut (gray arrow). The cut assigns $z_q^{i+1} = 0$ and $z_p^i = 1$, which would violate the constraint in equation (2). The edge (q_{i+1}, p_i) is in the cut, making the cut cost prohibitively high.

according our graph cut convention, $z_q^j = 0$ if node $q_j \in \mathcal{S}$ and $z_p^i = 1$ if node $p_i \in \mathcal{T}$. Then to prevent the labelings $z_p^i z_q^j = \{10\}$, a set of directed edges (q_j, p_i) for $j > i$ with infinite weights should be added to the graph. Likewise, a set of directed edges (p_j, q_i) for $j > i$ with infinite weights should be added to prevent labelings $z_p^j z_q^i = \{01\}$. However observe that only the edge (q_{i+1}, p_i) is needed to prevent all labelings $z_p^i z_q^j = \{10\}$ for $j > i$. This is due to the encoding scheme in equation (6), where $z_q^j = 0$ implies that z_q^i must also equal 0 for $i < j$. Therefore preventing the labeling $z_p^i z_q^{i+1} = \{10\}$ will also prevent $z_p^i z_q^j = \{10\}$ for $j > i + 1$. By similar reasoning, only the edge (p_{i+1}, q_i) with infinite weight is needed to prevent all labelings $z_p^j z_q^i = \{01\}$ for $j > i$.

The graph in figure 5(a) shows the infinity weighted edges used to enforce the LAC. Note that these are directed edges. Figure 5(b) shows an example of an infeasible cut, where $p_i \in \mathcal{T}$ and $q_{i+1} \in \mathcal{S}$ resulting in $z_p^i z_q^{i+1} = \{10\}$. Since edge (q_{i+1}, p_i) with infinite weight is in the cut, this cut has a very high cost. Note that even if the cut assign $q_j \in \mathcal{T}$, $j > i + 1$ or $p_j \in \mathcal{S}$, $j < i$, edge (q_{i+1}, p_i) will still be in the cut.

Up to this point, we have not addressed the portion of the graph construction that is necessary to account for the costs $\theta_{pq;ij}$. The constrained edges (q_{i+1}, p_i) and (p_{i+1}, q_i) ensure that the first summation in equation (22) will be zero for all feasible cuts on the graph. The only remaining task is to add the necessary edges to minimize the second summation, *i.e.* the term involving α_{ii} . Subsequently, we assume that the pairwise potential $\theta_{pq;ij}$ is submodular with respect to all adjacent label pairs $\{y_p = i, y_q = i + 1\}$ so that $\alpha_{ii} \leq 0$. In figure 6(a), we show one possible edge weight assignment scheme, but there are other equivalent constructions, *e.g.* see

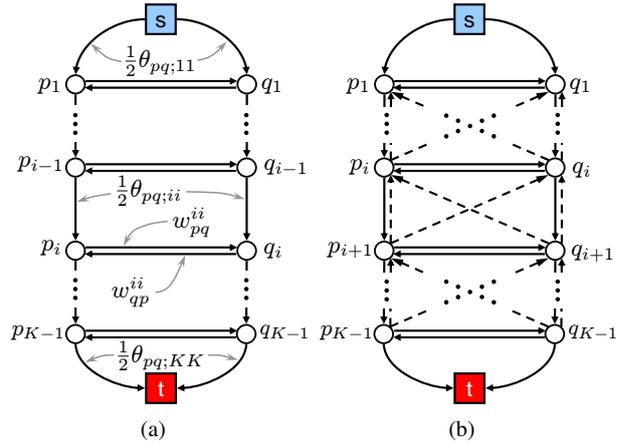


Fig. 6 Graph construction for the minimization of a multi-label variable pair with adjacent label constraint. (a) Edge weight assignments for the pairwise potential. (b) Final graph from additively combining the graphs in (a) and figure 5(a).

(Kolmogorov and Zabih (2004); Kohli and Torr (2007)), and their reparameterizations (Kolmogorov and Rother (2007)). The weights for edges (p_i, q_i) and (q_i, p_i) in the figure are

$$w_{pq}^{ii} = \theta_{pq;(i+1)i} - \frac{1}{2} \left(\theta_{pq;ii} + \theta_{pq;(i+1)(i+1)} \right) \quad (24a)$$

$$w_{qp}^{ii} = \theta_{pq;i(i+1)} - \frac{1}{2} \left(\theta_{pq;ii} + \theta_{pq;(i+1)(i+1)} \right). \quad (24b)$$

Although the graph shown in figure 6(a) may not be the most compact construction, it provides a straightforward and intuitive representation for encoding the energy in equation (1).

Utilizing the additive property of graphs (Kolmogorov and Zabih (2004)), the overall graph structure shown in figure 6(b) is produced by combining the graphs in figures 5(a) and 6(a), where the weights of directed edges linking the same nodes are added. The final graph has $(K - 1) \cdot N + 2$ nodes (including the terminals s and t), which is the same as the graph in figure 3(b). However the number of edges is $(K - 1) \cdot (M_n + M_t) + (K - 2) \cdot M_n$, where M_n and M_t are the total number of neighborhood and terminal edges, respectively, for the two label problem. Thus the number of edges is on the order of $\mathcal{O}(NK)$, which is a significant reduction from the number of edges in the graph construction proposed by (Ishikawa (2003)), shown in figure 3(b). Note that M_n depends on the neighborhood connectivity. The above construction (to the best of our knowledge) is the most efficient known technique for optimally solving the Nested Layer Segmentation problem.

Table 1 summarizes the weight assignments for the edges in the final graph. Note that it is possible for edges (p_i, q_i) and (q_i, p_i) to have negative weights. However, the reparameterization techniques in (Kohli and Torr (2007); Kolmogorov and Rother (2007)) can be used to transform the graph so that these edges will have nonnegative weights. In

Table 1 Edge weight assignments for label adjacency constraint graph.

Edge	Weight
(s, p_1)	$\theta_{p;1} + \frac{1}{2}\theta_{pq;11}$
(p_i, p_{i+1})	$\theta_{p;i+1} + \frac{1}{2}\theta_{pq;(i+1)(i+1)}$
(p_{K-1}, t)	$\theta_{p;K} + \frac{1}{2}\theta_{pq;KK}$
(p_{i+1}, p_i)	∞
(p_i, q_i)	$\theta_{pq;(i+1)i} - \frac{1}{2}(\theta_{pq;ii} + \theta_{pq;(i+1)(i+1)})$
(q_i, p_i)	$\theta_{pq;i(i+1)} - \frac{1}{2}(\theta_{pq;ii} + \theta_{pq;(i+1)(i+1)})$
(p_{i+1}, q_i)	∞
(q_{i+1}, p_i)	∞

this work we use a pairwise potential where $\theta_{pq,ii} = 0, \forall i \in \mathcal{L}$, and hence these edge weights are always nonnegative. The final graph construction in figure 6(b) allows the exact minimization of the energy in equation (1), subject to the LAC (2), in polynomial time using st-mincut techniques.

5 Experiments and Results

In this section, we first describe the segmentation workflow for our proposed method, starting with the user input and ending with completion of the maxflow/mincut. We provide details for computing the unary and pairwise potentials as well as the set of image features used in the experiments. We then show the results of applying our LAC algorithm on a set of 2D and 3D images, and briefly compare some results with the α -expansion and $\alpha\beta$ -swap algorithms (Boykov et al (2001)) and results using Ishikawa’s construction (Ishikawa (2003)).

5.1 Segmentation Workflow

Figure 8 shows the segmentation of an immunofluorescence image of a retinal cross section using 7 labels. The highly inhomogeneous textures in several of the regions require large exemplar regions for training. This image was provided by Dr. Geoff Lewis from the Neuroscience Research Institute, U.C., Santa Barbara.

For a given image, the segmentation begins with interactive user input. The user selectively marks one or more exemplar regions from each layer and indicates the layer nesting order, *e.g.* by the order in which each region was marked. As will be explained shortly, the pixel features in these exemplar regions are used for density estimation and subsequently to compute the unary potentials. Next, the algorithm constructs the LAC graph and computes the corresponding edge weight assignments. Finally, performing the

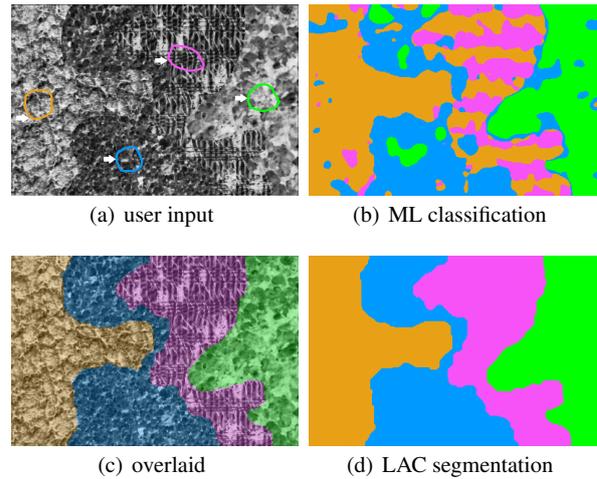


Fig. 7 LAC segmentation of a texture image. (a) Using 4 labels, the user selects an exemplar region for each label in order of layer nesting, either top-to-bottom or bottom-to-top for this example. From density estimation using the exemplar region features, the maximum likelihood (ML) classification is shown in (b). The LAC segmentation result (d) is overlaid or superimposed on top of the original image (c). Parameters: $\text{conn} = 8, \lambda_x = 1$, texture features.

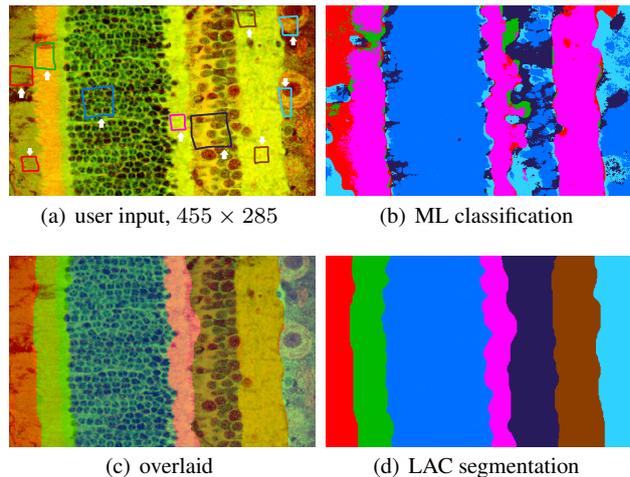


Fig. 8 Segmentation of immunofluorescence image of retinal cross section with 7 labels. Parameters: $\text{conn} = 16, \lambda_x = 0.1$, texture features.

maxflow/mincut on the LAC graph results in the globally minimal labeling.

Figure 7 illustrates the segmentation workflow for an image with four texture regions, some of which are visually very similar. Figure 7(a) shows the exemplar regions selected by the user, with the ordering specified from either top-to-bottom or bottom-to-top. For this example, the selected regions include all pixels enclosed inside the markings. Figure 7(b) shows the maximum likelihood (ML) classification for the labels after density estimation (see section 5.1 below). Though normally one would not choose ML classification for segmentation, this result illustrates that the

combination of inadequate image features and/or poor density estimation do not necessarily provide an accurate indication of the layer labels. Nonetheless the LAC segmentation algorithm is powerful enough to correct for this shortcoming. Figure 7(c) shows the result of our algorithm overlaid on top of the original image, and figure 7(d) shows the LAC result alone.

Figure 8 shows the segmentation of an immunofluorescence image of a retinal cross section using 7 labels. The highly inhomogeneous textures in several of the regions require large exemplar regions for training. This image was provided by Dr. Geoff Lewis from the Neuroscience Research Institute, U.C., Santa Barbara.

In our work, we use the MATLAB mex interface to run the maxflow algorithm of Boykov and Kolmogorov (Boykov and Kolmogorov (2004)) written in C++ and is available online (<http://vision.csd.uwo.ca/code/>). After user input, density estimation is performed using the Fast Gauss Transform (Yang et al (2003); Elgammal et al (2003)).

Unary and Pairwise Potentials: Let $x_p \in \mathbb{R}^d$ be a d -dimensional feature vector at pixel location p . We use the following cost for the unary potential in equation (4) $\theta_{p,i} = -\log \Pr(x_p | y_p = i)$, $\forall i \in \mathcal{L}$. This is simply the negative likelihood of x_p given that the label $y_p = i$. This cost favors the class label that best explains the observation x_p .

Given a set of pixel features from user selection, we use kernel density estimation to calculate the likelihood. Let $\{x_{1|i}, x_{2|i}, \dots, x_{n_i|i}\}$ be the set of training image features for the n_i pixels labeled $i \in \mathcal{L}$ by the user. The kernel density estimate for a pixel p is

$$\Pr(x_p | y_p = i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|x_p - x_{j|i}\|^2}{2\sigma^2}\right). \quad (25)$$

In all experiments, the bandwidth parameter σ is set to $\sqrt{2d}/10$ for $x_p \in [0, 1]^d$, similar to that in (Malcolm et al (2007)).

We define the pairwise potential $\theta_{pq,ij}$ as

$$\theta_{pq,ij} = \begin{cases} 0 & \text{if } y_p = y_q \\ g(x_p, x_q) & \text{if } |y_p - y_q| = 1 \\ \infty & \text{otherwise.} \end{cases} \quad (26)$$

The data dependent function $g(x_p, x_q)$ is

$$g(x_p, x_q) = \frac{1}{|p - q|} \left(\lambda_{\min} + \lambda_x \cdot \exp\left(-\frac{\|x_p - x_q\|^2}{2\sigma_x^2}\right) \right) \quad (27)$$

and acts to penalize pairwise label changes by an amount dependent on the difference between the features x_p and x_q (plus a constant λ_{\min}) (Boykov and Jolly (2001); Rother et al (2004); Kohli et al (2008)). Here, $|p - q|$ is the Euclidian distance between pixel p and q , which is not a constant when

the neighborhood connectivity is 8 or greater. The parameter λ_{\min} ensures that there is some minimum penalty for a label difference. The parameter $\lambda_x \in [0.1, 1]$ controls the degree of smoothness of the layer boundaries. Smaller values favor more jagged boundaries.

In practice, we only compute the cost $\theta_{pq,ij}$ when $|y_p - y_q| = 1$ and set the weight for the constraint edges in the graph to some large value, e.g. 10^6 . For all experiments, the parameter σ_x , which controls the contrast sensitivity, is set to be the square root of the average square norm

$$\sigma_x = \sqrt{\frac{1}{|\mathcal{E}_{RF}|} \sum_{(p,q) \in \mathcal{E}_{RF}} \|x_p - x_q\|^2}, \quad (28)$$

and λ_{\min} is set as

$$\lambda_{\min} = \frac{1}{|\mathcal{E}_{RF}|} \sum_{(p,q) \in \mathcal{E}_{RF}} \exp\left(-\frac{\|x_p - x_q\|^2}{2\sigma_x^2}\right). \quad (29)$$

Here \mathcal{E}_{RF} is the set of edges in the random field model.

Image Features: We use a combination grayscale, color values, and texture descriptors for the image features in our experiments. However, we want to emphasize that choice of image features is often application dependent and our method is not restricted to only the features presented herein. The RGB color images are converted into the three dimensional Luv colorspace, which we denote $[I \ I_u \ I_v]$. For texture discrimination, we use the diffusion based texture features proposed in (Rousson et al (2003)). These features are computed from the joint nonlinear diffusion of the structure tensor components, resulting in a three dimensional vector $[I_{xx} \ I_{xy} \ I_{yy}]$. Then to characterize the scale of the texture, we use the TV flow based local scale measure I_s , which is one dimensional (Brox and Weickert (2004)). There are other popular texture descriptors that can be used, such as the Gabor features (Manjunath and Ma (1996)). However, these features are often high dimensional and can negatively affect the accuracy of density estimation, especially when user input is sparse. The combined texture descriptor used in our experiments is only four dimensional, but it has been shown to perform comparably with a 12-dimensional Gabor feature (Brox and Weickert (2006)).

5.2 LAC Segmentation Results: Quantitative Evaluation

We compare the proposed approach with the Ishikawa construction and the α expansion algorithm. First we pause to discuss the issues with both competing methods, followed by a thorough quantitative evaluation.

Ishikawa Construction: As stated in section 1.2, the graph construction of Ishikawa (Ishikawa (2003)) can also be used to solve the nested layer segmentation problem. However, there are practical challenges that make Ishikawa's method

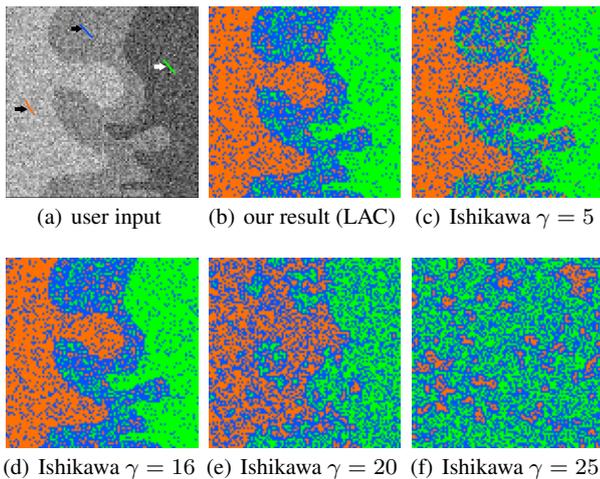


Fig. 9 Comparison with Ishikawa method. For low values of γ , the results violate the LAC. As the value of k increases, the Ishikawa method encounters numerical errors and produces increasingly less accurate results.

less suitable for tackling these problems. Besides requiring a larger graph compared to our method, the Ishikawa algorithm is prone to numerical overflow when enforcing the LAC, as shown in figure 9. Using the pairwise potential (26), we ran the Ishikawa algorithm using equation (3) for $\gamma = \{1, 2, \dots, 35\}$. Figure 9(b) shows the result of our method using the same parameters, and the results of the Ishikawa method for four values of γ are shown in the last two rows of figure 9.

Notice that for sufficiently small γ , the Ishikawa segmentation violates the LAC resulting in a large energy equation (1) subject to the LAC. The result in figure 9(c) for $\gamma = 5$ shows numerous instances of this violation. As γ increases to around 16, where the energy is lowest, these violations are reduced. As seen in figure 9(d) there are no label violations and the result is visually similar to ours. However the segmentation energy remains slightly higher, indicating a near but non-optimal solution. For sufficiently large values of γ , the algorithm encounters numerical overflow problems when computing the maximum flow, and the results are no longer accurate. Figures 9(e) and 9(f) show the result for $\gamma = 20$ and $\gamma = 25$, respectively. At these values of γ , the segmentation energies are far from the optimal.

α -Expansion: We also compare our LAC algorithm to the α -expansion and $\alpha\beta$ -swap algorithms (Boykov et al (2001)), which are state of the art algorithms used to obtain approximate solutions to the multi-label MRF problem. These algorithms iteratively makes labeling moves at each iteration to decrease the MRF energy. As noted by Liu et al. (Liu et al (2008)), both the expansion and swap algorithms are more likely to get stuck in local minima when ordering constraints are used, and we observed this behavior frequently in our experiments. We test these two algorithms

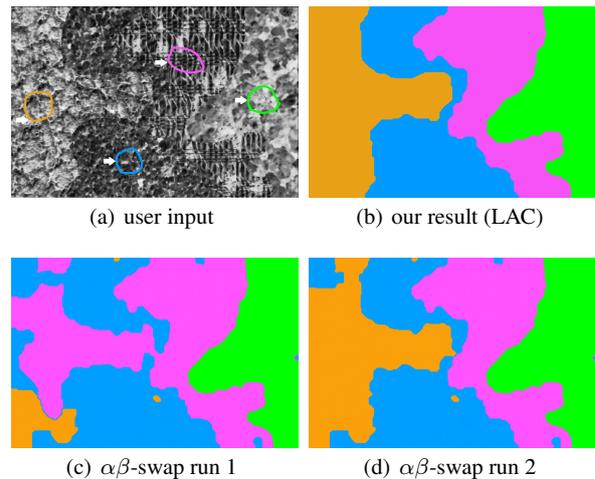


Fig. 10 Results from two separate runs of the $\alpha\beta$ -swap algorithm with LAC. Although there are no constraint violations, the MRF energies for the labelings in 10(c) and 10(d) are much higher than the shown using LAC segmentation in 10(b).

using the same parameter settings on the image in figure 7 and set the adjacency constraint penalty to 10^6 . Since both algorithms compute the solution iteratively starting from a random initializations, the final solutions are often different for different runs. The result of two separate runs for the $\alpha\beta$ -swap algorithm is shown in figure 10. The expansion algorithm also produced similar results. Although these segmentations do not violate the LAC, the MRF energies for these labelings are much higher than the LAC segmentation shown in figure 10(b). We observed similar results for the other images in our experiments.

5.3 Quantitative Evaluation:

We conducted experiments to quantitatively compare our algorithm against Ishikawa’s method and the α -expansion algorithm. The first set of experiments compare the segmentation performance on a set of synthetic nested layer images, similar to that in figure 7. The second set of experiments evaluate performance on a set of confocal retinal images. For both cases, pixel wise ground truth segmentation is compared to results produced by the algorithm using the hamming distance and rand index, a metric recently proposed for comparing segmentations (Unnikrishnan et al (2007)). We report hamming distance as the percentage of pixels in the segmentation result that are not in agreement with the ground truth. Hence, a larger Hamming distance percentage reported would correspond to greater disagreement of segmentation with the ground truth. Finally, results on segmentation of images acquired for retinal detachment studies over four time intervals are reported.

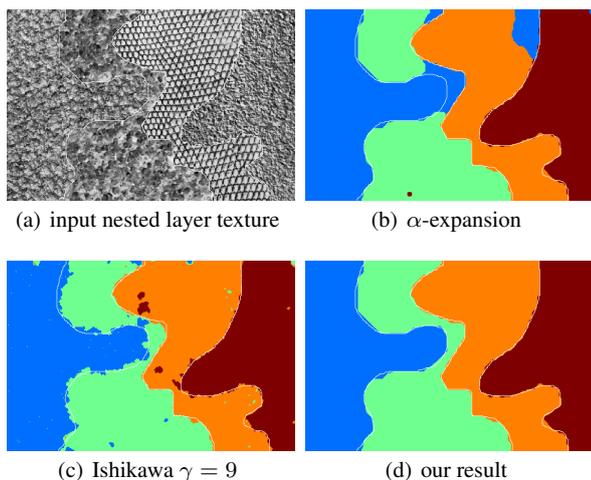


Fig. 11 (Best Viewed in Color) Nested Texture segmentation experiment described in section 5.3.1. Results of the proposed approach do not violate label ordering and produce results that are visually and quantitatively superior to other competing techniques.

5.3.1 Experiments on Nested Texture Dataset

We created a database of 200 nested texture images, similar in structure to that in figure 7. For each image, we randomly sample four textures from the Brodatz texture album (Brodatz (1966)). An example is shown in figure 11. The unary potentials are computed using TV flow texture features, and the interaction potentials are constructed as described previously for the LAC and Ishikawa methods. Here, we also show the hamming distance and rand scores for the maximum likelihood (ML) segmentation, as well as the Ishikawa's construction (IC) and α -expansion algorithm (AE). In all experiments, we attempted two variations. In the first case, the strokes offered to the algorithm were hardcoded (typical of interactive settings, strong unaries), while strokes were not hardcoded in the second case. We discovered that LAC and IC performed much better than ML and AE. Further, the performance of LAC and IC almost overlapped, in spite of IC requiring a lot more edges. Further, the unary potentials for the texture dataset are pretty strong and the total number of layers is only four. The subsequent discussion will explicitly illustrate advantages over IC when unaries are weak and the number of labels scale. The quantitative results for this case are illustrated in figures 13, 15, 14, 16 and tables 2, 3.

It is important to note that the hamming distance/rand index is not designed to capture violations in nested layer constraints. As a result, even minor differences in the hamming distance/rand index (in layered textures in table 2) leads to major visual differences in the segmentation results (see figure 11). Since we are unaware of a metric that penalizes violations in label constraints, we resort to the widely used validation metrics.

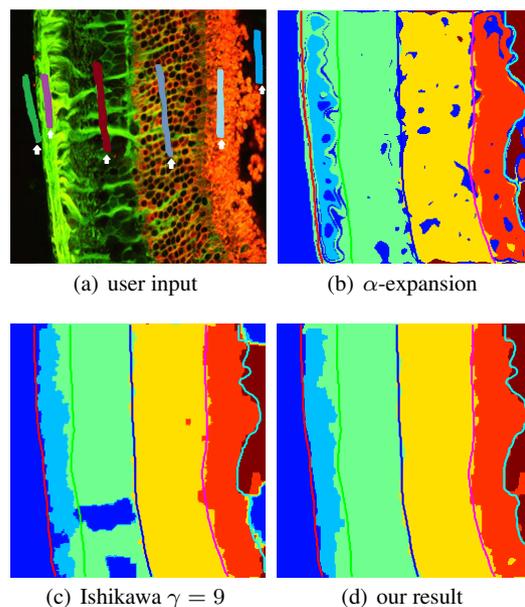


Fig. 12 (Best Viewed in Color) Retinal layer segmentation experiment described in section 5.3.2. The label ordering is preserved for our result.

5.3.2 Experiments on Nested Retina Dataset

The second experiment utilizes 11 confocal images of the retina taken 3 days after retinal detachment. An example is shown in figure 12. These images contain a total of 4 layers of interest and 2 background layers (6 layers total). Note that the textures in the layers are challenging to adequately characterize, and we resort to using color features. In addition, a distance transform constructed from user input strokes is utilized to modulate the unary potential, which discourages pixels far away from a stroke to be assigned that stroke's label in addition to hardcoding labels on pixels contained in a stroke. We refer to the above procedure as stroke hardcoding. With stroke hardcoding, see Table 2, LAC has a Hamming distance of 13 from ground truth while the nearest competitor is AE with a distance of 16. On a more challenging scenario with no stroke hardcoding, see Table 3, LAC still has a Hamming distance of 13 from ground truth while AE has a distance of 29, a difference of 16 between the two results! It can be observed that the Ishikawa construction degrades in performance with increasing labels since it has distances of 27 and 33 in comparison to LAC's 13 and 13.

5.3.3 Analysis of Runtime and Performance across Datasets

The results in figure 17 present the running time of various techniques in seconds. A striking observation is the long time taken by the Ishikawa construction since the number of edges grows quadratically with the number of la-

bels. Further, the proposed approach comes very close to the running time of α -expansions even though the memory requirements for α -expansions is lower. Further, since the proposed approach requires a single maxflow computation, parallel and distributed solvers could potentially offer significant speedups.

Figure 18 illustrates performance of the various techniques with datasets on the x-axis. We find that for retinal detachment studies, the proposed approach consistently outperforms α -expansions. However, an interesting trend on performance with respect to the datasets emerge. For all techniques considered, 3-day detached retinas are the easiest to segment while normal retinas are the toughest to segment. The ML and Ishikawa constructions are worst affected by changing datasets. This finding could be utilized to collect more careful user annotations for tougher (Normal and 28-day detached retina) datasets.

In the case of nested texture and nested retina datasets, we find that the proposed approach still performs the best. However, the Ishikawa construction edges out α expansions in the nested texture dataset while α -expansions still ranks second for the nested retina dataset. This behavior can be attributed to the stronger unary terms present in the texture datasets and lower number of layers, in contrast to the weaker unary terms and more layers for retinal datasets.

In summary, the quantitative experiments have demonstrated the advantages offered by the proposed approach on diverse texture and retinal layer segmentation datasets. In spite of the non-availability of a validation metric that captures label constraint violations, the proposed approach ranks better than state of the art on the rand scores. We have experimented with various other applications in 2D and 3D segmentation where nested layer constraints hold. The results of auxiliary experiments can be viewed at <http://vision.ece.ucsb.edu/segmentation/nls/>.

5.3.4 Experiments on Retinal Detachment Dataset

The final experiment we performed utilizes data from retinal detachment studies. It is of great biological importance to study the effects of retinal detachment (Fisher et al (2005)) over time. For this purpose, confocal microscopic images are captured at time regular time intervals. In the dataset we considered, the intervals were normal retina, 3 days, 7 days and 28 days after detachment. We utilized ten images randomly sampled from each time interval with pixel wise ground truth for validating the proposed approach.

Table 23 illustrates the comparative performance of the proposed approach with competing techniques on the Retinal Detachment dataset. We observe that the proposed approach (colored red) performs better than all competing methods. It is interesting to note that α -expansions comes close to the proposed approach with stroke hardcoding. For exam-

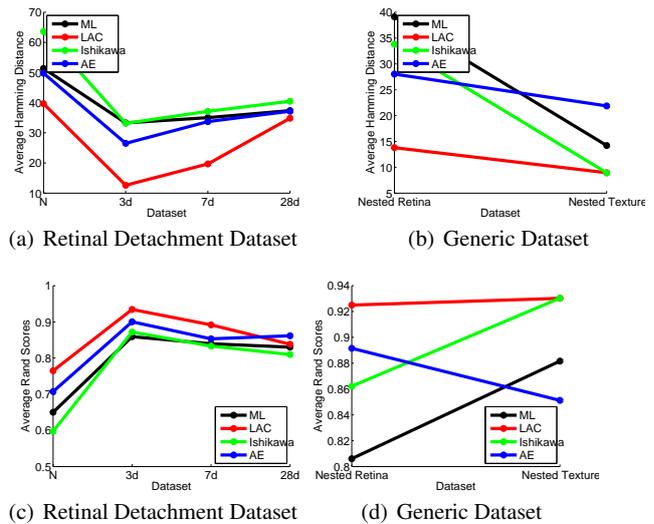


Fig. 18 (Best Viewed in Color) Performance on the Retinal Detachment and Generic Datasets with no Stroke Hardcoding on the Hamming distance (top row) and Rand scores (bottom row)

ple LAC beats AE by 3 percent in the 3day detached retinas in Table 2. We emphasize that though the difference is about 3 percent in favour of our method, α -expansions yields biologically implausible results by violating label constraints. Further, ML and the Ishikawa constructions yield much higher distances in comparison to the proposed approach. On the more challenging scenario of no stroke hardcoding, LAC performs significantly better than competing methods and the margin in 3day detached retinas grows to 11!, see Table 3. A graphical illustration of the comparative performance of the various approaches on the rand scores and hamming distance is shown in figures 13,15,14,16 and tables 2,3.

5.3.5 Failure Case

A probable failure case of LAC is when there are weak and misleading unary potentials, or noise that is correlated with the constraints. In such rare scenarios LAC might wipe out layers, especially near the image corners. We very rarely observed such cases in our experiments, and even in such cases the average performance of LAC was comparable to its competitors. For example, without stroke hardcoding on the 28 day detached case the performance of LAC drops though it beats all other methods on the average Hamming distance. By strengthening the unaries with hardcoding LAC is able to increase its performance margin, see Table 23.

6 Future Work and Conclusion

There are several future directions that we wish to explore. Despite using a straightforwardly simple graph construction,

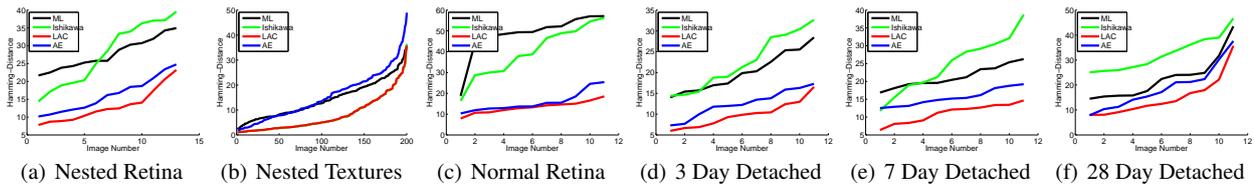


Fig. 13 (Best Viewed in Color, **Strokes hardcoded**) Hamming distances (Lower is Better) corresponding to different datasets considered.

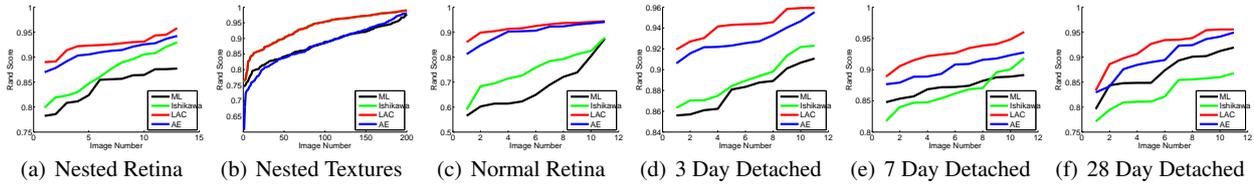


Fig. 14 (Best Viewed in Color, **Strokes hardcoded**) Rand Scores (Higher is Better) corresponding to different datasets considered.

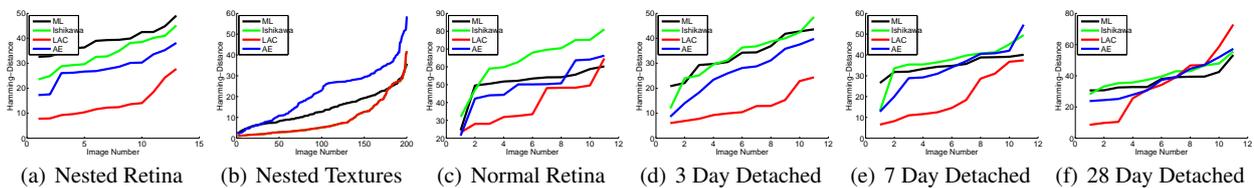


Fig. 15 (Best Viewed in Color, **Strokes not hardcoded**) Hamming distances (Lower is Better) corresponding to different datasets considered.

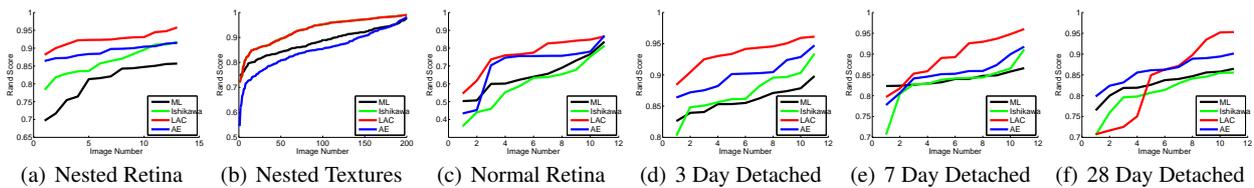


Fig. 16 (Best Viewed in Color, **Strokes not hardcoded**) Rand Scores (Higher is Better) corresponding to different datasets considered.

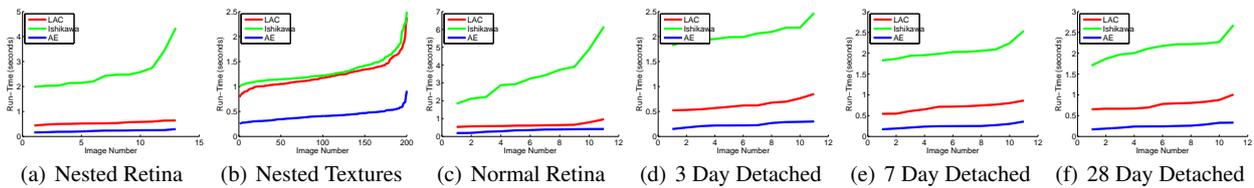


Fig. 17 (Best Viewed in Color) Running times corresponding to the different datasets considered

the LAC segmentation is still limited to relatively small data sizes, especially in 3D. This limitation can be quite significant, since biomedical datasets are typically very large. Fortunately, there are research efforts in developing maxflow algorithms for large vision graphs (DeLong and Boykov (2008)) and methods to run graph cuts on large graphs using GPUs (Schoenemann and Cremers (2007)). Second, incorporating higher order Potts potentials (Kohli et al (2007)) into the proposed MRF energy should further improve the algorithm's ability to capture larger spatial dependencies among pixel groups, *e.g.* where the texture is inhomogeneous or have large scale. These types of textures are difficult to characterize using existing texture descriptors. A third direction is to use the LAC segmentation algorithm to segment im-

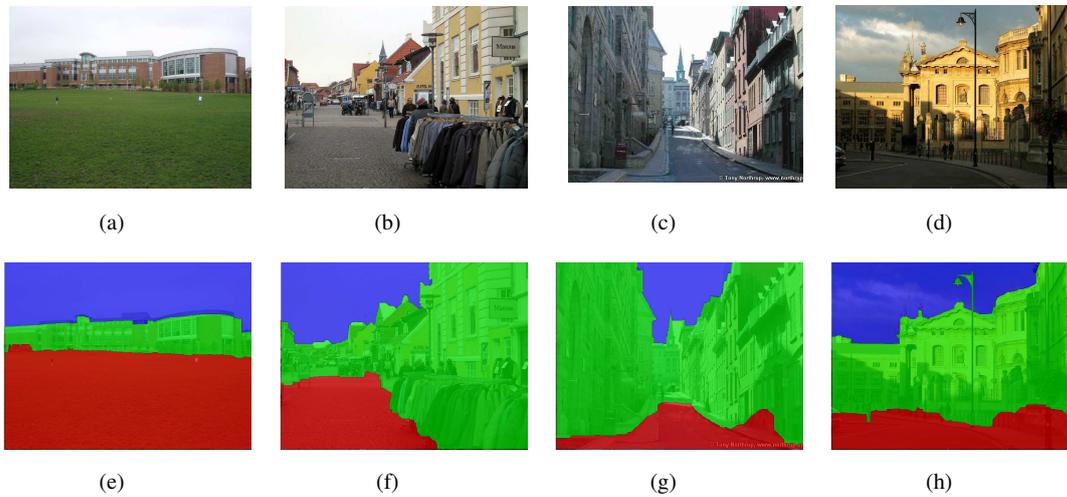
ages that exhibit a hierarchical nested layer relationship. We would also like to note that our construction can be applied to the geometric scene labeling (Hoiem et al (2007)) problem where a tiered label structure exists as observed by (Zheng et al (2012)). Preliminary results obtained by applying LAC are shown in Figure 19. We observed that the unary potentials provided by (Hoiem et al (2007)) are of good quality in comparison to the unary potentials we employed for retinal layer segmentation. A detailed analysis of the benefits LAC could offer for geometric scene labeling, investigation of new datasets where unary potentials are not very reliable, and comprehensive benchmarking are part work we are actively pursuing. We have requested (Zheng et al (2012)) for datasets used in their paper, and were told that we could have

Table 2 Average Hamming Distance(Lower is Good)/Rand Score(Higher is Good) - Strokes are Hardcoded

Method	Edges Required	Layered Texture	Nested Retina	Normal	Detached(3d)	Detached(7d)	Detached(28d)
ML	N/A	14/.88	27/.84	48/.68	20/.88	21/.87	22/.87
LAC	$\mathcal{O}(NK)$	7/.93	13/.92	13/.91	9/.94	11/.92	15/.92
IC	$\mathcal{O}(NK^2)$	7/.93	27/.86	39/.75	22/.89	24/.86	32/.82
AE	$\mathcal{O}(N)$	15/.87	16/.91	16/.89	12/.92	15/.90	18/.90

Table 3 Average Hamming Distances(Lower is Good)/Rand Scores(Higher is Good) - No Stroke Hardcoding

Method	Edges Required	Layered Texture	Nested Retina	Normal	Detached(3d)	Detached(7d)	Detached(28d)
ML	N/A	14/.88	39/.80	51/.65	33/.85	35/.83	37/.83
LAC	$\mathcal{O}(NK)$	8/.93	13/.92	39/.76	12/.93	19/.89	34/.83
IC	$\mathcal{O}(NK^2)$	8/.93	33/.86	63/.59	33/.87	37/.83	40/.80
AE	$\mathcal{O}(N)$	22/.85	29/.89	53/.70	23/.90	35/.85	38/.85


Fig. 19 (Best Viewed in Color) Application of LAC to the geometric scene labeling problem. (a)-(d) illustrate the input images and (e)-(h) correspond to labels resulting from applying graph cuts using the LAC constraint.

access within the next month. As a result, we are unable to currently provide quantitative metrics that compare our formulation with theirs, though our solution solves the problem exactly. We will report the quantitative metrics as soon as we receive the datasets from (Zheng et al (2012)).

In this work, we described a novel method for nested layer segmentation and showed that the additional label adjacency constraint in the MRF framework allows for efficient global solutions. More specifically we showed that the globally optimal solution for the MRF energy in equation (1) subject to the LAC in equation (2) can be found efficiently via graph cut techniques. Using boolean encoding, we transformed the original multilabel MRF energy into an equivalent function of boolean variables, which is submodular, graph representable, and can be minimized exactly and efficiently with graph cut. Our experimental results on both synthetic and real images demonstrate the utility of our proposed segmentation framework.

Acknowledgements This work was supported by grants from the National Science Foundation, NSF-III-0808772 and NSF-OIA-0941717. The authors would like to thank Dr. Nedim C. Buyukmihci, Dr. Darl Ray Swartz, Dr. Geoff Lewis, and Dr. Steven Fisher for providing data utilized in experiments

References

- Boros E, Hammer PL (2002) Pseudo-boolean optimization. *Discrete Applied Mathematics* 123(1):155–225, DOI 10.1016/S0166-218X(01)00341-9 4
- Boykov Y, Jolly MP (2001) Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: *Proc. IEEE Int'l Conf. Computer Vision*, Vancouver, BC, Canada, vol 1, pp 105–112, DOI 10.1109/ICCV.2001.937505 10
- Boykov Y, Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE TPAMI* 26(9):1124–1137, DOI 10.1109/TPAMI.2004.60 2, 10

- Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. *IEEE TPAMI* 23(11):1222–1239, DOI 10.1109/34.969114 [2](#), [9](#), [11](#)
- Brodatz P (1966) *Textures: A Photographic Album for Artists & Designers*. Dover [12](#)
- Brox T, Weickert J (2004) A tv flow based local scale measure for texture discrimination. In: *Proc. Euro. Conf. Computer Vision* [10](#)
- Brox T, Weickert J (2006) A TV flow based local scale estimate and its application to texture discrimination. *J Vis Commun Image R* 17(5):1053–1073, DOI 10.1016/j.jvcir.2005.06.001 [10](#)
- Chung G, Vese LA (2005) Energy minimization based segmentation and denoising using a multilayer level set approach. In: *Proc. Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, FL, United States, vol 3757, pp 439–455 [2](#)
- DeLong A, Boykov Y (2008) A scalable graph-cut algorithm for n-d grids. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp 1–8 [14](#)
- Elgammal A, Duraiswami R, Davis LS (2003) Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking. *IEEE TPAMI* 25(11):1499–1504, DOI 10.1109/TPAMI.2003.1240123 [10](#)
- Felzenszwalb P, Veksler O (2010) Tiered scene labeling with dynamic programming. In: *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on, IEEE, pp 3097–3104 [3](#)
- Fisher S, Lewis G, Linberg K, Verardo M (2005) Cellular remodeling in mammalian retina: results from studies of experimental retinal detachment. *Progress in retinal and eye research* 24(3):395–431 [13](#)
- Ford LR, Fulkerson DR (1962) *Flows in Networks*. Princeton University Press [4](#)
- Freedman D, Drineas P (2005) Energy minimization via graph cuts: settling what is possible. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol 2, pp 939–946, DOI 10.1109/CVPR.2005.143 [4](#), [5](#), [7](#)
- Garvin M, Abramoff M, Kardon R, Russell S, Wu X, Sonka M (2008) Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *Medical Imaging*, *IEEE Transactions on* 27(10):1495–1505 [3](#)
- Geman S, Geman D (1984) Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE TPAMI* 6(6):721–741 [2](#)
- Greig DM, Porteous BT, Seheult AH (1989) Exact maximum a posteriori estimation for binary images. *J R Statist Soc B* 51:271–279 [4](#)
- Hoiem D, Efros A, Hebert M (2007) Recovering surface layout from an image. *International Journal of Computer Vision* 75(1):151–172 [14](#)
- Ishikawa H (2003) Exact optimization for markov random fields with convex priors. *IEEE TPAMI* 25(10):1333–1336, DOI 10.1109/TPAMI.2003.1233908 [3](#), [4](#), [5](#), [6](#), [8](#), [9](#), [10](#)
- Kohli P, Torr PHS (2007) Dynamic graph cuts for efficient inference in markov random fields. *IEEE TPAMI* 29(12):2079–2088, DOI 10.1109/TPAMI.2007.1128 [8](#)
- Kohli P, Kumar MP, Torr PHS (2007) P3 & beyond: Solving energies with higher order cliques. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition.*, pp 1–8, DOI 10.1109/CVPR.2007.383204 [14](#)
- Kohli P, Ladicky L, Torr PHS (2008) Robust higher order potentials for enforcing label consistency. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition.*, pp 1–8 [10](#)
- Kolmogorov V (2006) Convergent tree-reweighted message passing for energy minimization. *IEEE TPAMI* 28(10):1568–1583, DOI 10.1109/TPAMI.2006.200 [2](#)
- Kolmogorov V, Rother C (2007) Minimizing nonsubmodular functions with graph cuts—a review. *IEEE TPAMI* 29(7):1274–1279, DOI 10.1109/TPAMI.2007.1031 [8](#)
- Kolmogorov V, Zabih R (2004) What energy functions can be minimized via graph cuts? *IEEE TPAMI* 26(2):147–159, DOI 10.1109/TPAMI.2004.1262177 [3](#), [4](#), [8](#)
- Leventon ME, Grimson WEL, Faugeras O (2000) Statistical shape influence in geodesic active contours. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, vol 1, pp 316–323, DOI 10.1109/CVPR.2000.855835 [2](#)
- Liu X, Veksler O, Samarabandu J (2008) Graph cut with ordering constraints on labels and its applications. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp 1–8 [3](#), [11](#)
- Malcolm J, Rathi Y, Tannenbaum A (2007) A graph cut approach to image segmentation in tensor space. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp 1–8, DOI 10.1109/CVPR.2007.383404 [10](#)
- Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. *IEEE TPAMI* 18(8):837–842, DOI 10.1109/34.531803 [10](#)
- Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proc. IEEE Int'l Conf. Computer Vision*, Vancouver, BC, Canada, vol 2, pp 416–423, DOI 10.1109/ICCV.2001.937655 [2](#)
- Ramalingam S, Kohli P, Alahari K, Torr PHS (2008) Exact inference in multi-label crfs with higher order cliques. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp 1–8 [4](#), [5](#), [6](#)
- Rother C, Kolmogorov V, Blake A (2004) Grabcut: Interactive foreground extraction using iterated graph cuts. In: *Proc. SIGGRAPH*, vol 23, pp 309–314 [10](#)

- Rousson M, Paragios N (2002) Shape priors for level set representations. In: Proc. Euro. Conf. Computer Vision **2**
- Rousson M, Brox T, Deriche R (2003) Active unsupervised texture segmentation on a diffusion based feature space. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol 2, pp 699–704, DOI 10.1109/CVPR.2003.1211535 **10**
- Schlesinger D, Flach B (2006) Transforming an arbitrary minsum problem into a binary one. Tech. rep., Technical Report TUD-FI06-01, Dresden University of Technology **4, 6**
- Schoenemann T, Cremers D (2007) Globally optimal image segmentation with an elastic shape prior. In: Proc. IEEE Int'l. Conf. Computer Vision **14**
- Tsai A, Yezzi J A, Wells W, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A (2003) A shape-based approach to the segmentation of medical imagery using level sets 22(2):137–154, DOI 10.1109/TMI.2002.808355 **2**
- Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. IEEE TPAMI 29(6):929–943, DOI 10.1109/TPAMI.2006.12 **11**
- Vese LA, Chan TF (2002) A multiphase level set framework for image segmentation using the mumford and shah model. Int'l J Computer Vision 50(3):271–293, DOI 10.1023/A:1020874308076 **2**
- Vu N (2008) Image segmentation with semantic priors: A graph cut approach. PhD thesis, University of California, Santa Barbara **3**
- Wainwright MJ, Jaakkola TS, Willsky AS (2005) MAP estimation via agreement on trees: message-passing and linear programming 51(11):3697–3717, DOI 10.1109/TIT.2005.856938 **2**
- Yang C, Duraiswami R, Gumerov NA, Davis L (2003) Improved fast gauss transform and efficient kernel density estimation. In: Proc. IEEE Int'l Conf. Computer Vision, pp 664–671, DOI 10.1109/ICCV.2003.1238383 **10**
- Zhao HK, Chan T, Merriman B, Osher S (1996) A variational level set approach to multiphase motion. J of Comp Phys 127:179–195 **2**
- Zheng Y, Gu S, Tomasi C (2012) Fast tiered labeling with topological priors. European Conference on Computer Vision **3, 14, 15**