# Label Set Perturbation for MRF based Neuroimaging Segmentation

Dylan Hower[¶] Vikas Singh[†*] Sterling C. Johnson[†]
[¶]Epic Systems Inc. Madison, WI
[†]Dept. of Biostatistics & Medical Inform. Univ. of Wisconsin–Madison
[‡]Dept. of Medicine Univ. of Wisconsin–Madison

## Abstract

*Graph-cuts based algorithms are effective for a variety of segmentation tasks in computer vision. Ongoing research is focused toward making the algorithms even more general, as well as to better understand their behavior with respect to issues such as the choice of the weighting function and sensitivity to placement of seeds. In this paper, we investigate in the context of neuroimaging segmentation, the sensitivity/stability of the solution with respect to the input "labels" or "seeds". In particular, as a form of parameter learning, we are interested in the effect of allowing the given set of labels (and consequently, the response/statistics of the weighting function) to vary for obtaining lower energy segmentation solutions. This perturbation leads to a "refined" label set (or parameters) better suited to the input image, yielding segmentations that are less sensitive to the set of labels or seeds provided. Our proposed algorithm (using Parametric Pseudoflow) yields improvements over graph-cuts based segmentation with a fixed set of labels. We present experiments on about* $450$ *3-D brain image volumes demonstrating the efficacy of the algorithm.*

## 1. Introduction

Graph-cuts based methods have emerged as powerful tools for image segmentation, and are steadily gaining popularity in the medical imaging community. Here, segmentation is cast as a discrete labeling problem where we seek to reassign each pixel in an image to a 'label' from a given label set (derived from interactively placed seeds, for example). A good labeling can then be determined by estimating the maximum a posteriori (MAP) labeling solution via energy minimization. In some cases, this is equivalent to finding the maximum flow (or minimum cut) in an appropriate graph, for which efficient methods exist. Therefore, an attractive feature of graph-cuts is their running time – offering the possibility of fast and accurate segmentation of 3-D brain image volumes in neuroimaging applications.

In the graph-cuts framework, the segmentation problem is typically expressed as the minimization of Markov Random Field (MRF) energy in the following form.

$$E(f) = \min \sum_{p \in \mathcal{P}} \underbrace{D_p(f_p)}_{\text{data cost}} + \sum_{(p,q) \in \mathcal{N}} \underbrace{V_{p,q}(f_p, f_q)}_{\text{smoothness cost}}, \qquad (1)$$

where $f$ is the desired labeling for the pixels in $\mathcal{P}$. The labeling maps each pixel $p \in \mathcal{P}$ to a label $f_p$ in a set $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \cdots, \mathcal{L}_k\}$. The chosen neighborhood system is denoted as $\mathcal{N}$. The term, $D_p(\cdot)$, in (1) is the data or deviation cost, and consists of the set of penalties associated with the variation between the pixel and the assigned label $f_p \in \mathcal{L}$. The term, $V_{p,q}(\cdot,\cdot)$, in (1) represents the smoothness or separation cost, and corresponds to the pairwise interactions of pixels. When expressed in this form, the configuration, $f$, that leads to the global optimum of (1) is expected to correspond to a good segmentation, and we are concerned with efficient methods to obtain the global minima or good approximations. The optimization of (1) assumes that $D_p(\cdot)$, $V_{p,q}(\cdot,\cdot)$, and the label set, $\mathcal{L}$ are given – either as coefficients or in functional form (e.g., Gaussian). How, and the extent to which these *parameters* affect the quality of the solution [5] raises a number of important research questions studied within *parameter learning*. For segmentation, several recent papers are providing insights into issues such as the stability of graph-cut solutions (given $D_p(\cdot)$ and $V_{p,q}(\cdot,\cdot)$) [13] and means for deciding the relative influence of these two terms [19]. Our interest is to explore the uncertainty in the parameterized coefficients of the energy in (1) – specifically, how variations in the set of labels can be leveraged for improving the overall quality of the segmentation, and to study its applicability for segmenting different tissue types in 3-D MR brain images. In doing so, we allow fine-tuning of the labels by permitting $\mathcal{L}$ (specifically, the label intensity means) to "move around". The problem becomes: how to *efficiently* compute the specific perturbation, $\mathcal{L}'$, that gives the best overall labeling (in terms of lower energy), i.e., finding the labeling *and* the label set. We will discuss these aspects in detail shortly.

*Correspondence: V. Singh (www.biostat.wisc.edu/∼vsingh).
Research performed when DH spent a semester at UW-Madison funded by Morgridge Institute for Research and UW-CIBM. This project was supported by Wisconsin Comprehensive Memory Program, NIH CTSA award 1UL1RR025011, UW Graduate School, and NIH R01-AG021155.

### 1.1. Related Work

The growing popularity of graph-cuts based algorithms has led to methods for the incorporation of user interaction [2] and powerful hierarchical solutions for 3-D image volumes [11] – ideas relevant for the application domain in this paper. In the context of the underlying problem, it is useful to note recent work on the *stability* of the final solutions of graph-cuts based algorithms. For example, to provide an uncertainty measure associated with the labeling, Kohli and Torr presented an algorithm [14] for estimating marginal energies for each variable. Earlier, Yanover and Weiss [26] also used marginals to compute the $m$ most probable solutions. Another relevant topic relates to the body of work on learning the parameters for MRF models, which is useful for expressing priors in low-level vision [20, 23, 24, 16, 17]. In such formalizations, one typically seeks to train the MRF model (by optimizing its parameters) so that the optimal solution of the function matches with the training data. In general, this turns out to be a difficult problem, and various strategies have been proposed to search the space of solutions, see [23, 17]. With respect to regularization, Peng and Veksler [19] emphasized that an arbitrary choice of the relative influence of the smoothness and data terms leads to inferior solutions, and [27] explored the effect of spatial coherence in the feature space of clusters. On the other hand, the effect of the set (or placement) of initial seeds [21] (in an interactive mode) on the final solution has also been investigated. The placement of seeds may specify, for example, a representative pixel from the foreground (and the background), and can be used to construct the label set, $\mathcal{L}$, for the given problem instance. Because the final segmentation may be sensitive to the initial seed placement, allowing certain variations in the location or intensity values of the seeds can be expected to partly mitigate this sensitivity. *Label set perturbation* refers to this expectation. Indeed, since the labels are the parameters of the energy we wish to optimize, label set perturbation may be viewed as parameter learning for a certain class of models. Finally, in the context of parameter search, we note that [17] seeks to incorporate a non-local prior (e.g., shape template) into segmentation. The solution makes a branch-and-bound search of a *discrete set of parameters* (e.g., different template poses) more efficient, similar to the basic premise of our work (we thank an anonymous reviewer for this observation). The algorithm [17] permits a broader class of energies compared to this paper, but the flexibility comes with a substantially higher computational cost.

### 1.2. Problem Statement and Overview

Consider the label set, $\mathcal{L}$ for a segmentation or denoising problem where the forms of the functions, $V_{p,q}(\cdot,\cdot)$ and $D_p(\cdot)$ are given, e.g., $D_p(f_p) = (I(p) - f_p)^2$ where $I(p)$ gives the intensity of $p$. The set of labels may be specified by (i) seed points, (ii) priors generated from hand-segmented training data, or (iii) a set of cluster centers returned from a $k$-means clustering on the histogram of image intensity values; (ii) and (iii) are used when processing many image volumes in "batch" mode. However, errors in the estimation of the intensity means (i.e., labels) may propagate as errors in the segmentation. This is particularly true in brain image segmentation, where an initial label set (determined by $k$-means clustering) gives a good first estimate, but often, the segmentation can be progressively improved with manual fine-tuning of the labels in $\mathcal{L}$. As an illustration, consider this effect on an image from the Berkeley Segmentation Database [18] in Fig. 1. Here, we start with three labels as input for (1), each representative of a perceptually distinct region in the image. A manual fine-tuning of the labels (parameters), incrementally improves the segmentation (see Fig. 1, $I_1, \cdots, I_5$ relative to ground truth). Furthermore, this also corresponds to a decrease in the optimized value of the MRF energy (see Fig. 1, top row).

The example suggests that permitting the initial $\mathcal{L}$ to vary allows *decreasing* the optimized objective function value. Then, an interesting question is: if $|\mathcal{L}| = k$, then what is an $\mathcal{L}' \in \mathcal{S}_\delta(\mathcal{L})$ which gives strictly smaller objective function values. Here, $\mathcal{S}_\delta(u)$ for $u \in \Re^k$ denotes a $k$-D neighborhood of width $\delta$ centered on $u$ and specifies the extent of perturbation (in the parameter) permitted. Since $E_{\mathcal{L}}(f)$ is an indicator of the quality of the segmentation, the "relaxation" of $\mathcal{L}$ allows finding $E_{\mathcal{L}' \in \mathcal{S}_\delta(\mathcal{L})}(g^*)$, one that is at least as good as $E_{\mathcal{L}}(f^*)$. Here, $f^*$ and $g^*$ indicate optimal solutions (labeling configurations). If the energy is already the best possible, then $\mathcal{L}' \approx \mathcal{L}$ (i.e., the label set/parameter stays approximately the same). Otherwise, we can obtain an $\mathcal{L}'$ that improves the solution. We modify (1) to formally convey the problem in general form

$$\min \sum_{p \in \mathcal{P}} \underbrace{D_p(f_p)}_{} + \sum_{(p,q) \in \mathcal{N}} \underbrace{V_{p,q}(f_p, f_q)}_{}, \text{ where } f_u \in \mathcal{L}' \in \mathcal{S}_\delta(\mathcal{L})$$

This paper makes the following contributions: (I.) We provide an algorithm to search for an improved label set, $\mathcal{L}'$, $|\mathcal{L}'| = k$, in a specified neighborhood of $\mathcal{L}$ to reduce the minimum value of the objective function (solution w.r.t. $\mathcal{L}$). Here, the solution space includes the set of distinct labeling configurations (for fixed $\mathcal{L}$), and also the space of label sets, $\mathcal{L}' \in \mathcal{S}_\delta(\mathcal{L})$ (similar to the discrete set $\Omega$ in [17]). Rather than branch-and-bound search [17], our algorithm exploits some key properties of Parametric Pseudoflow [7]. Allowing perturbation adds only a small overhead to the run-time of one max-flow but offers robustness and improvements in accuracy. (II.) We report extensive evaluations on hundreds of 3-D brain image volumes including the ADNI data [10]. Our results demonstrate that in contrast to graph-cuts based segmentation (with fixed labels/parameters), permitting label set variations yields more accurate solutions than three popular neuroimaging segmentation software.

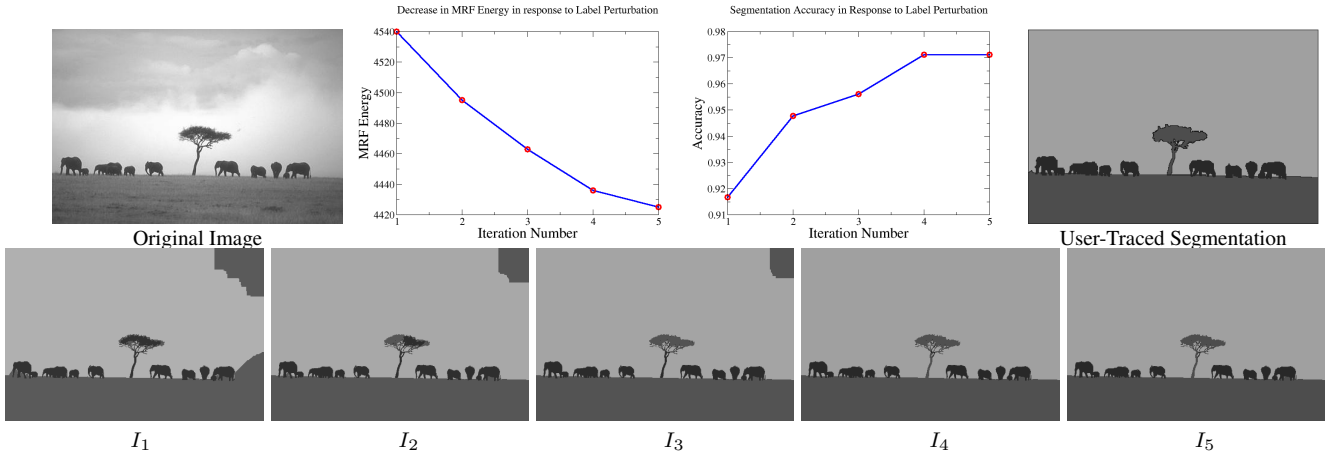



Figure 1. Images $I_1, \cdots, I_5$ (bottom row) indicate incremental segmentations as a result of manual fine-tuning of labels. The segmentation seems to improve qualitatively, and the plots in the first row indicate a corresponding decrease in the objective function value.

## 2. Main Ideas

To illustrate the ideas, we return to the Fig. 1 example and analyze changes in segmentations as we vary the labels. First, observe how the global optimum of (1) and the corresponding segmentation, $f$, respond as we slowly move (turn a knob) from $\mathcal{L}$ to another label set, $\mathcal{L}^{(1)}, \mathcal{L}^{(2)}, \cdots, \mathcal{L}^{(t)}$: while the absolute value of the optimal solution changes with such label perturbation, the induced configurations may or may not vary for a subset of label sets $\mathcal{L}' \in \mathcal{D}_\delta(\mathcal{L})$. This partitions the space of label sets around $\mathcal{L}$ (parameter space) into a set of *equivalence classes*.

**Definition 2.1** (Equivalence classes)**.** Label sets, $\mathcal{L}$ and $\mathcal{L}'$ are *equivalent* for a problem instance $I$ (for pre-specified forms of $D_p(\cdot)$ and $V_{p,q}(\cdot,\cdot)$) if their corresponding optimal configurations, $f^*$ and $g^*$ are equivalent: $\forall p \in \mathcal{P}, f_p^* = g_p^*$.

An infinitesimally small translation of a label set positioned on the boundary of one equivalence class may move it into a different equivalence class. Notice that the corresponding optimal segmentations for the original and perturbed label sets are *different*. The new (perturbed) label set, i.e., parameter settings, corresponds to a *breakpoint* (see [15, 8] for some related interpretations). In Fig. 1, a breakpoint corresponds to a small change in one of the labels that leads to *at least one pixel* switching affiliation and moving from one class to another. Therefore, rather than *all* label sets around $\mathcal{L}$ as possible solutions to our problem, we may focus only on the (possibly fewer) set of equivalence classes. In other words, a label set from the same equivalence class as the optimal label set, $\mathcal{L}^*$ will be sufficient and give *exactly* the same segmentation. Next, we invoke an important result [8, 15] to show that the set/number of breakpoints is not too large.

**Property 2.2.** *The total number of breakpoints $b$ for a given problem instance $I$ can be no more than $|\mathcal{P}| = n$ where $\mathcal{P}$ is the set of pixels in the image.*

In practice, this upper bound is conservative and the number of breakpoints are quite small. Our objective now is to use these results to efficiently find the breakpoints, and identify a good $\mathcal{L}^*$ (the parameter learning/search procedure) and the corresponding segmentation for an image.

## 3. Graph Construction

The optimization of (1) is generally performed by constructing a graph to encode properties of the given image. The desired segmentation can be obtained easily given the solution of the maximum flow problem on this graph [3]. We start by describing the graph construction for binary segmentation, and then discuss the changes that make it suitable for the problem at hand.

The graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices (or nodes) and $\mathcal{E}$ is the set of edges (or arcs). The set $\mathcal{V}$ includes one node for each pixel in $\mathcal{P}$ as well as two distinguished nodes: a source $s$ and a sink $t$, see Fig. 2. The pairwise interactions between the pixels (i.e., smoothness cost) is incorporated as directed edges: Fig. 2 shows the edges with a capacity, $V_{p,q}(f_p, f_q)$. Edges connecting the source to a node or a node to the sink represent the data cost in (1). The maximum flow problem on this graph can be solved efficiently which gives a set of 'saturated' edges (where no additional flow can be pushed). The maximum flow is equivalent to the minimum cut of this graph given by removing the set of saturated edges. The label assignment (segmentation) is determined by assigning each node to zero (if detached from the source) or one (otherwise).

### 3.1. Graph for Multiple Labels

The graph we will use for problem instances with more than two labels is a simple "parametric" extension of the one shown in Fig. 2. The set of pairwise edges between neighboring nodes (for smoothness cost) and the set of vertices does not change. However, each edge connecting the

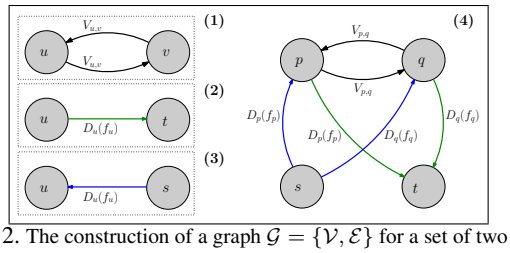



Figure 2. The construction of a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ for a set of two pixels $\mathcal{P} = \{p, q\}$ in an image, with a source, $s$, and sink, $t$.

node to the source or the sink is now replaced with its parametric variant, see Fig. 3 (top row) and also [6]. To illustrate this parameterization, the first step is to decide on a certain *ordering* on the set of labels. In the context of our application where the imaging modalities (like MR, CT) are all gray-scale, the ordering on the labels is explicit. Even in the more general case of natural images where the labels may have a semantic meaning (e.g., sky, sand) such an ordering can be easily derived via a minimal distortion embedding on the number line, for which simple yet powerful solutions are available [25]. Now, let us consider the case where $|\mathcal{L}| = k$. This gives $k-1$ *intervals* denoted as $\lambda = \{\lambda_1, \cdots, \lambda_{k-1}\}$ where each interval maps to a label pair $(\mathcal{L}_\tau, \mathcal{L}_{\tau+1})$ (subsequent labels in order). For the parameterization, each edge connecting a node to the source/sink is replaced by $k-1$ edges, one for each interval in $\lambda$ (clearly, adding parametric edges generalizes the binary-label case where $k = 2$ and $|\lambda| = 1$). The weights (capacities) of the $k-1$ edges reflect the $k-1$ subgradients (or derivatives) of the function $D_p(\cdot)$ in those intervals, similar to [6]. For instance, if $\lambda_\tau \mapsto (\mathcal{L}_\tau, \mathcal{L}_{\tau+1})$ then for a node $p \in \mathcal{P}$, the edge capacities for interval $\lambda_\tau$ will be calculated from $D'_p(\lambda_\tau) = D_p(\mathcal{L}_\tau + \epsilon) - D_p(\mathcal{L}_\tau)$. More specifically, the edge (for parameter $\tau$) from the source to $p$ is assigned a weight, $-\min\{0, D'_p(\lambda_\tau)\}$, where as the weight of the edge from $p$ to the sink is $\max\{0, D'_p(\lambda_\tau)\}$. The intuition and consequences for assigning edge weights in this manner will become clear shortly. Toward this goal, we review some useful characteristics of this construction.

**Property 3.1.** *Every node $p \in \mathcal{P}$ will have a positive-capacity edge with exactly one of the nodes in $\{s, t\}$.*

**Class of MRF energies.** For clarity of presentation, the construction described deals with $D_x(\cdot)$ convex and $V_{x,y}(\cdot,\cdot)$ linear; the modification for $D_x(\cdot)$ arbitrary and $V_{x,y}(\cdot,\cdot)$ convex (see [9] for one of the first illustrations of such ideas) can be done with only a logarithmic increase in running time [7], and is well suited for a wide class of MRF energies. In practice, the procedure may still be applied to other MRF energies (albeit in a heuristic manner).

**Property 3.2** (Monotonicity)**.** *(a) The capacities from the source to each pixel are a monotonically non-increasing function of $\lambda_\tau$; (b) The capacities from each pixel to the sink are a monotonically non-decreasing function of $\lambda_\tau$.*

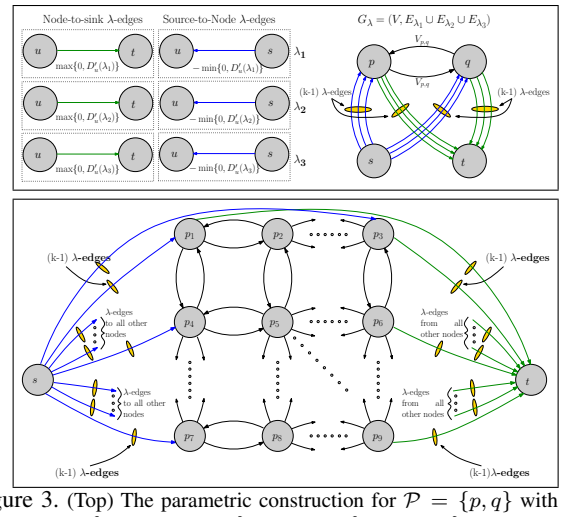



Figure 3. (Top) The parametric construction for $\mathcal{P} = \{p, q\}$ with four labels, $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4\}$ and $\lambda = \{\lambda_1, \lambda_2, \lambda_3\}$. (Bottom) The parametric graph for the image $G_\lambda = G_{\lambda=1} \cup G_{\lambda=2} \cup \cdots \cup G_{\lambda=k-1}$.

Let $\mathcal{G}_{\lambda_\tau} = (\mathcal{V}, \mathcal{E}_{\lambda_\tau})$ denote the graph for a specific parameter $\lambda_\tau \in \lambda$. Notice that this graph includes *only* the edges pertaining to $\lambda_\tau$. Let $f^*(\lambda_\tau)$ denote the solution to the maximum flow (and minimum cut) problem on this graph which induces a partition of the nodes into a source set $S_{\lambda_\tau}$ (nodes connected to the source) and the sink set $\bar{S}_{\lambda_\tau}$ (nodes connected to the sink). The following result from [8] lays the necessary groundwork for making use of parametric max-flow for our segmentation problem.

**Theorem 3.3** (Threshold Theorem)**.** *In a given parametric graph, $G_{\lambda_\tau}$, with $S^*_{\lambda_\tau}$ as the source set of $f^*(\lambda_\tau)$, there is an optimal solution $f^*$ to the image segmentation problem that satisfies $f^*_p \geq \lambda_\tau$ if $p \in S^*_{\lambda_\tau}$ and $f^*_p < \lambda_\tau$ if $p \in \bar{S}^*_{\lambda_\tau}$ where $p \in \mathcal{P}$ and $f^*_p$ gives $p$'s label in the optimal solution.*

## 3.2. Segmentation and Identification of Breakpoints

The properties of Theorem 3.3, and especially the inequalities, provide a framework to analyze a maximum flow solution of $G_\lambda$. First, consider the task of obtaining a multi-label segmentation of the image. If the maximum flow solutions for each graph $G_{\lambda_\tau}, \lambda_\tau \in \lambda$ were available, we could use Thm. 3.3 as a "yes/no" oracle to obtain a labeling for each pixel $p$ in $\mathcal{P}$. We ask: "Is the label for $p$ in an optimal solution less than $\mathcal{L}_\tau$?" Asking this at most $k-1$ times for each $p \in \mathcal{P}$ yields a segmentation where the first "yes" (when evaluating the labels in increasing order) specifies the label to be assigned to $p$. Second, consider the set of intervals, $\lambda$. For each $p \in \mathcal{P}$, we can identify a corresponding $\lambda_\tau$ such that $f^*_p \geq \mathcal{L}_\tau$ *and* $f^*_p < \mathcal{L}_{\tau+1}$. In addition, if the intervals in $\lambda$ are sufficiently small (resolution), for a $\lambda_\gamma \in \lambda$ there may only be one pixel (node) that satisfies the above property in $\lambda_\gamma$, i.e., for $\lambda_\tau < \lambda_\gamma$, $p$ is in the source set, and for $\lambda_\tau \geq \lambda_\gamma$, $p$ is in the sink set. That is, $\lambda_\gamma$ is a breakpoint where $p$ moves from the source set to the sink set, precisely the behavior we observed in Fig. 1 and §2. Observe that

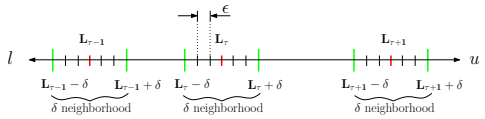



Figure 4. The range of labels with a $\delta$ neighborhood shown around $\mathcal{L}_{\tau-1}$, $\mathcal{L}_{\tau}$, and $\mathcal{L}_{\tau+1}$. Each integral grid point (at $\epsilon$-scale) in the refinement interval at a $\epsilon$ grid distance may be a breakpoint.

the desired label set, $\mathcal{L}' \in \mathcal{S}_\delta(\mathcal{L})$ should only be comprised of such breakpoints. Finally, Thm. 3.3 gives a nestedness property which states that the source set for $\lambda_{\tau+1} \in \lambda$ is nested in the source set of the previous parameter, $\lambda_\tau$ for all $\tau$ values. This establishes that a given pixel, $p \in \mathcal{P}$, will have at most one breakpoint, i.e., once it moves to the sink set, it does not move back to the source set. Second, by solving a parametric maximum flow problem on $G_\lambda$, we can determine which $\lambda_\tau \in \lambda$ are breakpoints without examining the breakpoints for each pixel individually. We denote the set of all breakpoints by $\beta = \{\lambda_\tau \mid \lambda_\tau \in \lambda, S_{\lambda_\tau} \subset S_{\lambda_{\tau-1}}\}$.

We can use any parametric max-flow (e.g., parametric Pseudoflow [7]) to find all integral node-shifting breakpoints in a parametric graph, $G_\lambda$, by dynamically computing cuts for each subsequent parameter in $\lambda$. Because of the ordering on $\mathcal{L}$, the amount of flow needed to saturate an edge that connects two nodes in $\mathcal{P}$ w.r.t. parameter $\lambda_\tau$, will be *less than* the amount of flow needed to saturate the same edge for parameter $\lambda_{\tau+1}$. Therefore, after finding the flow for a specific $\lambda_\tau$, pseudoflow can resume by using the computed flows (for the $\lambda_\tau$-th parameter) and continue to push more flow until the edges are saturated for parameter $\lambda_{\tau+1}$. This allows computing the min-cut for all parameters efficiently, and is known to work well in practice [11].

The parametric pseudoflow algorithm can be extended to find the set of *all* breakpoints, $\beta$ up to a desired level of precision, $\epsilon$ in a given range $[l, u]$, see Fig. 4. This is possible because the set of minimum cuts on a parametric graph can be found either in forward or reverse order. The backward direction merely involves reversing (i) the order of the labels, and (ii) the direction of the edges. Consider a parametric graph for the set of parameters $\{l, l+\epsilon, l+2\epsilon, \cdots, u-\epsilon, u\}$, and the corresponding reverse graph. We can compute the first pseudoflow iteration on the graph and also its reverse graph (i.e., for parameters $l$ and $u$). Now, if $S_l = S_u$ or $(u - l) \leq \epsilon$, we know that there are no breakpoints in the interval, $[l, u]$ and we can stop. Otherwise, we iterate using the next pair of parameters by decrementing (and incrementing) the parameter for the reverse (and forward) graph. We continue until we meet the stopping condition above.

## 4. Refining Labels to Non-integral Breakpoints

Once the breakpoints are computed using the procedure outlined above, without loss of generality we can restrict each $\mathcal{S}_\delta(\mathcal{L}_\tau) \in \mathcal{L}$ during label set refinement to only include breakpoints. For the refinement process, we proceed from $\mathcal{L}$ (initial label set) to $\mathcal{L}'$ (the refined label set) in $k$ steps. This refinement is analogous to moving to a node with a better lower bound in [17]. At each step, one of the labels $\mathcal{L}_\tau$ is refined to a breakpoint in its refinement interval, $[\mathcal{L}_\tau - \delta, \mathcal{L}_\tau + \delta]$, see Fig. 4. For each breakpoint in these intervals, the value of the corresponding max-flow/min-cut is known (from parametric pseudoflow). We also know the breakpoint with the smallest min-cut value in each interval ($\delta$-neighborhood). We refine each label to the *specific* breakpoint in its interval which has the smallest min-cut value. To summarize, the modification proceeds as follows:

$$\begin{aligned} \mathcal{L}^{(2)} &= \{\mathcal{L}'_1, \mathcal{L}_2, \cdots, \mathcal{L}_k\} \\ \mathcal{L}^{(3)} &= \{\mathcal{L}'_1, \mathcal{L}'_2, \cdots, \mathcal{L}_k\} \\ &\vdots \\ \mathcal{L}^{(k+1)} &= \{\mathcal{L}'_1, \mathcal{L}'_2, \cdots, \mathcal{L}'_k\} = \mathcal{L}' \end{aligned}$$

We can now present the following useful property:

**Theorem 4.1** (Monotonicity)**.** *In each iteration of label set refinement, the objective function value cannot increase.*

*Proof.* For a given breakpoint $\alpha$ in the refinement interval of label $\mathcal{L}_\tau$, consider the optimal source set, $S^*_\alpha$ from the pseudoflow algorithm. The source set, $S^*_\alpha$ is associated with the value of the cut corresponding to separating the pixels in the source set, $S^*_\alpha$, from the pixels not in the source set, $\bar{S}^*_\alpha$. Denote this cost by $C(\bar{S}^*_\alpha, S^*_\alpha)$, and note that such a cost is known for all breakpoints in the $\delta$-neighborhood of $\mathcal{L}_\tau$. Now, compare the value of $C(\cdot, \cdot)$ for $\alpha$ with that of the breakpoint $\alpha'$ that yields the smallest cut value in $[\mathcal{L}_\tau - \delta, \mathcal{L} + \delta]$. Since $\alpha$ and $\alpha'$ are both breakpoints, exactly one of the following must be true: $S^*_\alpha \subset S^*_{\alpha'}$ or $S^*_{\alpha'} \subset S^*_\alpha$. The value of their corresponding cuts must also be different. Let us assume $C(\bar{S}^*_\alpha, S^*_\alpha) + \Delta = C(\bar{S}^*_{\alpha'}, S^*_{\alpha'})$. But $C(\bar{S}^*_\alpha, S^*_\alpha) > C(\bar{S}^*_{\alpha'}, S^*_{\alpha'})$, which implies $\Delta \leq 0$.

**(Case 1).** When $(S^*_{\alpha'} \subset S^*_\alpha)$, let $S^*_\alpha \setminus S^*_{\alpha'}$ denote the set of pixels that change assignment. Let the optimal labeling when $\mathcal{L}_\tau = \alpha$ be denoted as $f^*$. When $\mathcal{L}_\tau = \alpha'$ as a result of the refinement, the new labeling, $f'$ is given as follows.

$$f'_p = \begin{cases} f^*_p + (\alpha' - \alpha) & p \in S^*_{\alpha'} \\ f^*_p & p \notin S^*_{\alpha'} \end{cases}$$

Since $\alpha' - \alpha > 0$, the objective function value when $\mathcal{L}_\tau = \alpha'$ must satisfy $E(f') \leq E(f^*) + (\alpha' - \alpha)\Delta \leq E(f^*)$.

**(Case 2).** When $S^*_\alpha \subset S^*_{\alpha'}$, let $S^*_{\alpha'} \setminus S^*_\alpha$ denote the set of pixels that change assignment. When $\mathcal{L}_\tau = \alpha'$ as a result of the refinement, the new labeling, $f'$ is given as follows.

$$f'_p = \begin{cases} f^*_p - (\alpha - \alpha') & p \in S^*_{\alpha'} \\ f^*_p & p \notin S^*_{\alpha'} \end{cases}$$

Because $\alpha - \alpha' > 0$, the objective function value when $\mathcal{L}_\tau = \alpha'$ must satisfy $E(f') \leq E(f^*) + (\alpha - \alpha')\Delta \leq E(f^*)$. In either case, with each modification, the energy function cannot increase which gives the desired result. □

## 5. Running Time Analysis

Assume that the parametric graph $G_\lambda$ has $n$ nodes and $m$ edges, and let $\delta$ and $\epsilon$ denote the size of the neighborhood and the resolution (see Fig. 4), where the number of edges, $|\mathcal{E}| = m$, increases as a function of $\frac{2\delta}{\epsilon}$. Detecting breakpoints at $\epsilon$ granularity takes $\mathcal{O}(mn \log(\frac{m^2}{n^2}))$ using the Parametric Pseudoflow algorithm [7]. The min-cut values at each breakpoint can be saved for future use. This allows label refinement to be completed in $\mathcal{O}(k \log(\frac{2\delta}{\epsilon}))$ time. This is negligible even (in worst case) when $\frac{2\delta}{\epsilon} = n$ since $k \log(n) \ll mn$. Once this is available, determination of the final labeling takes $\mathcal{O}(n)$ time. Therefore, the total running time of our algorithm is $\mathcal{O}(mn \log(\frac{m^2}{n^2}))$.

## 6. Experimental Results

Our experiments were focused toward obtaining accurate segmentations of T1-weighted magnetic resonance (MR) brain images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). This is important for any subsequent analysis of volumetric variations between groups of healthy controls and patients with a certain clinical condition.

While graph-cuts based segmentation is significantly faster compared to widely used neuroimaging software, their overall accuracy (using a fixed set of labels) is frequently inferior to popular tools such as SPM5 [1], FSL [22], and FreeSurfer [4]. Therefore, typically segmentation in neuroimaging continues to be based on parametric or atlas based approaches. Our experiments suggest that the proposed algorithm, in contrast, is either competitive with or improves upon these neuroimaging software while *preserving the running time benefits* of graph-cuts based approaches.

**Setup.** (i) We evaluate the quality of our segmentation on a large database of 3-D brain images (with given ground truth segmentation) to ones obtained from SPM5, FSL, FreeSurfer, and graph-cuts based segmentation (with fixed labels). (ii) We compare our segmentation to those obtained from graph-cuts on a recent aging-related dataset of $400$ 3-D brain images. Since ground truth segmentation is unavailable for this data, our comparisons focus on the optimized energy values and variations in the two segmentations.

### 6.1. BrainWeb Data

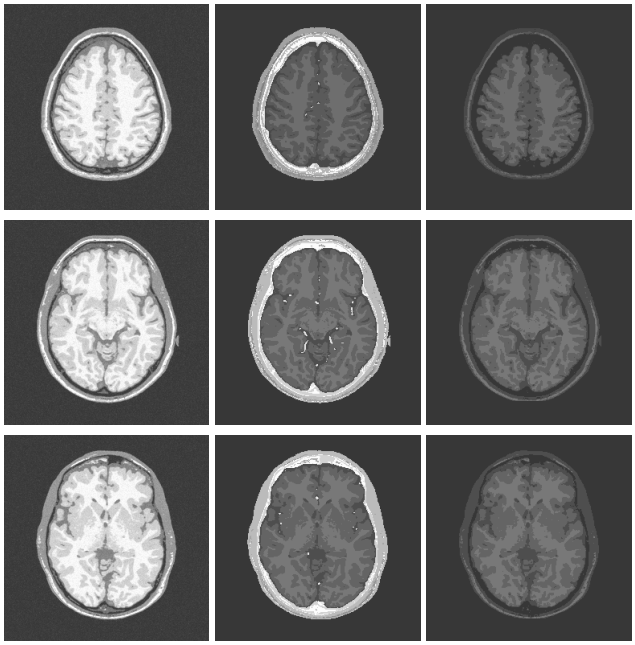
BrainWeb is a database of realistic brain MR image data generated by an MRI simulator (`http://www.bic.mni.mcgill.ca/brainweb/`). Since ground-truth segmentation of different tissue types is available, this dataset is used as a benchmark for evaluating the performance of different techniques [12]. The data consists of thirty T1-weighted image volumes in all, where ten image volumes have lower SNR (relative to the other $20$ images) as a function of varying degrees of introduced noise. The image sizes were $256^2 \times 190$ (images 1-20) and $181^2 \times 160$ (images 21-30). The initial label set, $\mathcal{L}$, for these images was determined by applying $k$-means on the image volume with $k = 4$. The refinement interval size, $\delta$ was set to $\frac{1}{3} \min \|\mathcal{L}_i - \mathcal{L}_j\|, \mathcal{L}_i, \mathcal{L}_j \in \mathcal{L}$. The granularity, $\epsilon$ was $\delta/10$.

Figure 5 (a) shows some representative segmentations obtained from our algorithm: the first, second, and third columns correspond to a 2-D axial slice of the input brain image volume, the corresponding ground truth segmentation, and our results, respectively. Fig. 5(b) summarizes the segmentation accuracy (for each tissue type) of our solution, FSL, SPM5, and FreeSurfer on this data. We see that for the two primary tissue types of interest, i.e., GM and WM, SPM5 gives the best results overall: $91\%$ (and $93\%$) for GM (and WM) for images 1-20, and $89\%$ (and $91\%$) for GM (and WM) for images 21-30 (with varying degrees of noise). FreeSurfer is only slightly less accurate than SPM5. FSL works well for WM segmentation, but gives modest results for GM [12]. The accuracy of graph-cuts (with fixed labels) was, on average, 5-7% worse than our algorithm, possibly because of inaccuracies in the initial label set specification. We noticed that the accuracy improved by manually adjusting either the labels or the weights of the data/smoothness costs. However, using parameter searching (label set perturbation) to decrease the objective function, our algorithm gives $91\%$ (and $96\%$) for GM (and WM) for images 1-20, and $89\%$ (and $93\%$) for GM (and WM) for images 21-30. This is useful because FSL, SPM5, and FreeSurfer take 8-20 minutes to segment each image volume where as our algorithm determines a high quality segmentation in about four minutes – up to a *four fold* improvement in running time with similar or better segmentations.

### 6.2. ADNI Data

Our second set of evaluations were performed on four hundred image volumes from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [10]. This study is still ongoing and ground truth segmentations are not available. Our evaluations primarily focused on the improvements in the MRF energy that can be obtained as a result of label set perturbation/parameter learning. We also evaluated the variation in the final segmentations (fixed labels and variable labels) in terms of the number of pixels with different label assignments. Other settings were the same as in §6.1.

Some representative segmentations from our algorithm are shown in Fig. 6. Columns one, two and three show a 2-D axial slice from the input image volume, results of graph-cuts based segmentation with fixed labels, and our results respectively. The fourth column includes histograms of (i) improvements in the MRF energy and (ii) variations in the final segmentation with/without label perturbation. Overall,




(a)

**Confusion Matrix of Multiple-Anatomical-Model images**

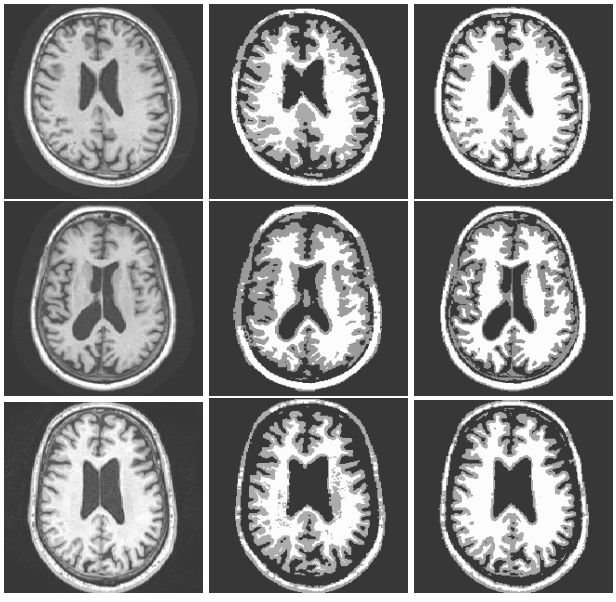
| | | gray matter | white matter | other |
|---|---|---|---|---|
| 1 | SPM5 | 0.913 | 0.031 | 0.056 |
| 2 | FreeSurfer | 0.832 | 0.138 | 0.029 |
| 3 | FSL | 0.904 | 0.020 | 0.076 |
| 4 | ours | 0.910 | 0.034 | 0.05 |
| 5 | SPM5 | 0.065 | 0.933 | 0.014 |
| 6 | FreeSurfer | 0.089 | 0.908 | 0.029 |
| 7 | FSL | 0.048 | 0.949 | 0.003 |
| 8 | ours | 0.044 | 0.956 | 0.000 |
| 9 | SPM5 | 0.001 | 0.000 | 0.999 |
| 10 | FreeSurfer | 0.005 | 0.000 | 0.999 |
| 11 | FSL | 0.000 | 0.000 | 1.000 |
| 12 | ours | 0.000 | 0.000 | 1.000 |

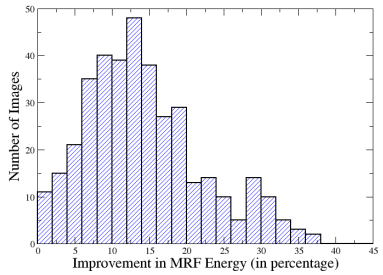
**Confusion Matrix for Variable Quality Images**

| | | gray matter | white matter | other |
|---|---|---|---|---|
| 1 | SPM5 | 0.891 | 0.068 | 0.041 |
| 2 | FreeSurfer | 0.803 | 0.129 | 0.068 |
| 3 | FSL | 0.828 | 0.069 | 0.104 |
| 4 | ours | 0.893 | 0.096 | 0.015 |
| 5 | SPM5 | 0.091 | 0.908 | 0.005 |
| 6 | FreeSurfer | 0.092 | 0.906 | 0.002 |
| 7 | FSL | 0.130 | 0.868 | 0.002 |
| 8 | ours | 0.074 | 0.926 | 0.000 |
| 9 | SPM5 | 0.008 | 0.000 | 0.991 |
| 10 | FreeSurfer | 0.005 | 0.000 | 0.995 |
| 11 | FSL | 0.009 | 0.000 | 0.991 |
| 12 | ours | 0.007 | 0.000 | 0.993 |

(b)

Figure 5. Example segmentations of BrainWeb images are shown in (a); the confusion matrices with the accuracy of our algorithm, graph-cuts with fixed labels, SPM5, FreeSurfer and FSL is shown in the (b). The table in (b)[top] summarizes the results for images 1-20, and the table in (b)[bottom] corresponds to images 21-30 (i.e., for varying degrees of additional noise). Here, rows 1-4, 5-8, and 9-12 correspond to GM, WM, and other matter labeling respectively.



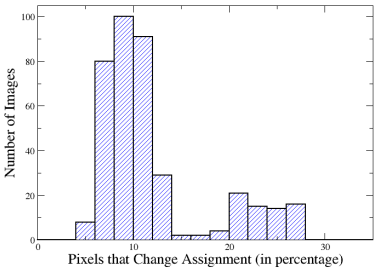




Figure 6. Example segmentations from the 400 image volumes from the ADNI dataset are shown on the left, along with histograms depicting differences between our method, and the graph-cuts solution with fixed labels.

the results look consistent and seem to give qualitative improvements over a standard graph-cuts approach. The first histogram (top row) in Fig. 6 shows that the *improvement* in the final MRF energy by permitting label perturbation is 5-25%. In general, this improvement also corresponds to a fraction of pixels (usually around the edges) changing their assignment to another label. The second histogram (bottom row) in Fig. 6 shows that label set perturbation leads to a reassignment of $\sim$10% of pixels. As in §6.1, each run on an ADNI image volume $(256^2 \times 166)$ takes about four minutes, which gives a run-time improvement of up to *four fold* over the neuroimaging segmentation software. In summary,

our results suggest that even with a relatively simple initialization of the initial label set, the algorithm obtains high quality segmentation of brain tissue types. This provides evidence that graph-cuts *with label perturbation* (parameter adjustment) is a viable (and attractive) alternative for neuroimaging segmentation. Our implementation will soon be available from `www.biostat.wisc.edu/~vsingh`.

## 7. Conclusions

In this paper, we show that the flexibility of parameter search/learning in the form of label set perturbation allows us to mitigate the sensitivity of the final solution to the initial label set (generated via $k$-means, seed indications, or priors). By exploiting useful properties of Parametric Pseudoflow, we present a general and efficient algorithm to solve this problem. We report extensive experiments on neuromaging segmentation. In the fixed label formulation of the problem, we see that the initial label set (or seeds) has a noticeable effect on the quality of the final segmentations. In contrast, the flexibility from label set perturbation mitigates these difficulties, and gives improved segmentations (by 5-7%). In a dataset where ground-truth segmentations are available, our algorithm outperforms three state-of-the-art segmentation algorithms in terms of accuracy, and gives *two to four fold* improvements in running time. On another dataset of 400 image volumes, we obtain robust and high-quality segmentations, also with run-time improvements. While our experiments focused on neuroimaging segmentation applications, the algorithm will likely be useful in other segmentation problems as well.

## References


[1] J. Ashburner and K. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.

[2] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. *Proc. of Intl. Conf. on Computer Vis.*, 1:105, 2001.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

[4] A. Dale, B. Fischl, and M. Sereno. Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[5] L. Grady and M. P. Jolly. Weights and topology: A study of the effects of graph construction on 3D image segmentation. In *Proc. of Med. Im. Comp. & Comp. Assisted Inter.*, 2008.

[6] D. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *J. of the ACM*, 48(4):686–701, 2001.

[7] D. Hochbaum. The pseudoflow algorithm: A new algorithm for the maximum-flow problem. *Operations Research*, 56(4):992–1009, 2008.

[8] D. Hochbaum and M. Queyranne. Minimizing a convex cost closure set. *SIAM J. Discr. Math.*, 16(2):192–207, 2003.

[9] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. *Lecture Notes in Computer Science*, 1406:232–248, 1998.

[10] C. Jack, M. Bernstein, N. Fox, P. Thompson, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. of Magnetic Resonance Imaging*, 2008.

[11] O. Juan and Y. Boykov. Active graph cuts. In *Proc. of Conf. on Computer Vis. and Pattern Recog.*, 2006.

[12] F. Klauschen, A. Goldman, V. Barra, A. Meyer-Lindberg, et al. Evaluation of automated brain MR image segmentation & volumetry methods. *Human Brain Mapping*, 2008.

[13] P. Kohli and P. Torr. Measuring Uncertainty in Graph Cut Solutions-Efficiently Computing Min-marginal Energies Using Dynamic Graph Cuts. *Lecture Notes in Computer Science*, 3952:30–42, 2006.

[14] P. Kohli and P. Torr. Measuring uncertainty in graph cut solutions. *Comput. Vis. Image Under.*, 112(1):30–38, 2008.

[15] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. In *Proc. of Intl. Conf. on Computer Vis.*, pages 1–8, 2007.

[16] S. Kumar, J. August, and M. Hebert. Exploiting Inference for approximate parameter learning in discriminative fields. In *Proc. of Conf. on Energy Minim. Met. in Computer Vis. and Pattern Recog.*, 2005.

[17] V. Lempitsky, A. Blake, and C. Rother. Image Segmentation by Branch-and-Mincut. *Lecture Notes in Computer Science*, 5305:15–29, 2008.

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and application to evaluating segmentation algorithms & measuring ecological statistics. In *Proc. of Intl. Conf. on Computer Vis.*, 2001.

[19] B. Peng and O. Veksler. Parameter selection for graph cut based image segmentation. In *Proc. of British Machine Vis. Conf.*, 2008.

[20] D. Scharstein and C. Pal. Learning Conditional Random Fields for stereo. In *Proc. of Conf. on Computer Vis. and Pattern Recog.*, 2007.

[21] A. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *Proc. of Intl. Conf. on Computer Vis.*, 2007.

[22] S. Smith, M. Jenkinson, M. Woolrich, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23:208–219, 2004.

[23] M. F. Tappen. Utilizing Variational Optimization to learn markov random fields. In *Proc. of Conf. on Computer Vis. and Pattern Recog.*, 2007.

[24] M. F. Tappen, C. Liu, E. Adelson, and W. Freeman. Learning Gaussian Conditional Fields for low-level vision. In *Proc. of Conf. on Computer Vis. and Pattern Recog.*, 2007.

[25] S. S. Vempala. *The Random Projection Method*. American Mathematical Society, 2004.

[26] C. Yanover and Y. Weiss. Finding the $M$ most probable configurations in arbitrary graphical models. In *Advances in Neural Information Processing Systems*, 2003.

[27] R. Zabih and V. Kolmogorov. Spatially coherent clustering using graph cuts. In *Proc. of Conf. on Computer Vis. and Pattern Recog.*, 2004.