Multi-Label Image Segmentation for Medical Applications Based on Graph-Theoretic Electrical Potentials*

Leo Grady** and Gareth Funka-Lea

Siemens Corporate Research — Department of Imaging and Visualization

Abstract. A novel method is proposed for performing multi-label, semi-automated image segmentation. Given a small number of pixels with user-defined labels, one can analytically (and quickly) determine the probability that a random walker starting at each unlabeled pixel will first reach one of the pre-labeled pixels. By assigning each pixel to the label for which the greatest probability is calculated, a high-quality image segmentation may be obtained. Theoretical properties of this algorithm are developed along with the corresponding connections to discrete potential theory and electrical circuits. This algorithm is formulated in discrete space (i.e., on a graph) using combinatorial analogues of standard operators and principles from continuous potential theory, allowing it to be applied in arbitrary dimension.

1 Introduction

One of the greatest challenges of automated medical image analysis is how to best take advantage of input from the technician or medical practitioner who will be performing the analysis. The current state-of-the-art is such that few automated image analysis techniques can be applied fully autonomously with reliable results. Consequently a person must evaluate the results of any automated process. In general, this person is currently performing the analysis manually. The goal of the automation is to shorten the time and improve the reliability of the analysis.

Frequently a medical image acquisition device is used for multiple medical indications and studies. In this case the goal of image analysis varies with the patient symptoms and history. For example, a cardiac image admits several "correct" segmentations depending on the nature of the region of interest: the entire heart or a specific ventricle or atria. The knowledge of the goal for the analysis is entirely with the user but by specifying the number of segments and by providing small samples of each segment the goal of the segmentation can be better defined for an automated approach.

In this paper we propose a novel algorithm to perform semi-automated image segmentation, given medical practitioner or computer pre-specified **labels**. Assume that the medical practitioner has provided K labeled voxels (hereafter referred to as **seed points** or **seeds**). For each unlabeled pixel, we could now ask: Given a random walker starting at this location, what is the probability that it first reaches each of the K seed points? It will be shown that this calculation may be performed exactly without the simulation of a random walk. By performing this calculation, we assign a K-tuple vector to each pixel that specifies the probability that a random walker starting from each unlabeled pixel will first reach each of the K seed points. A final segmentation may be derived from these K-tuples by selecting for each pixel the most probable seed destination for a random walker. In a uniform image (e.g., all black), a segmentation will be obtained that roughly corresponds to Voronoi cells for each set of seed points. We term this segmentation the **neutral** segmentation since it does not take into account any information from the image. By biasing the random walker to avoid crossing sharp intensity gradients, a quality segmentation is obtained that respects object boundaries (including weak boundaries). Analytical properties of the algorithm are listed at the end of the introduction.

* Published in Proceedings of the 8th ECCV04, Workshop on Computer Vision Approaches to Medical Image Analysis and Mathematical Methods in Biomedical Image Analysis, May 15th, 2004, Prague, Czech Republic, Springer-Verlag. More information at: http://cns.bu.edu/~lgrady/

** Leo Grady is with Siemens Corporate Research, Department of Imaging and Visualization, 755 College Road East, Princeton, NJ 08540, E-mail: Leo.Grady@scr.siemens.com

In our approach, we treat an image (or volume) as a purely discrete object — a graph with a fixed number of vertices and edges. Each edge is assigned a real-valued weight corresponding to the likelihood that a random walker will cross that edge (e.g., a weight of zero means that the walker may not move along that edge). Formulation of the algorithm on a graph allows the application of the algorithm to surface meshes or space-variant images [1, 2]. Regardless of the dimensions of the data, we will use the term *pixel* throughout this paper to refer to the basic picture element in the context of its intensity values. In contrast, the term *node* will be used in the context of a graph-theoretical discussion.

It has been established [3, 4] that the probability a random walker first reaches a seed point exactly equals the solution to the Dirichlet problem [5] with boundary conditions at the locations of the seed points and the seed point in question fixed to unity, while the others are set to zero. The solution to the discrete Dirichlet problem on an arbitrary graph is given exactly by the distribution of electric potentials on the nodes of an electrical circuit with resistors representing the inverse of the weights (i.e., the weights represent conductance) and the "boundary conditions" given by voltage sources fixing the electric potential at the "boundary nodes". For the remainder of the paper, we will adopt the terminology of circuit theory to describe the algorithm, with a potential, x_i^s, indicating the probability that a walker starting at node v_i first reaches a seed point with label s. A function that solves the Dirichlet problem for a given set of boundary conditions (i.e., the random walker probabilities) is known as **harmonic**. Figure 1 illustrates the harmonic functions (and subsequent segmentation) obtained for a 4×4 graph with unit weights in the presence of three seeds with different labels.

In light of the connection between random walks on graphs and discrete potential theory, one may calculate the probability that a random walker starting at pixel v_i first reaches a seed with label s, by solving the circuit theory problem that corresponds to a discrete analog of the Dirichlet problem [6]. Ground (i.e., fix the potential to zero) all seed points belonging to labels other than s and establish a unit voltage source with ground that fixes the s-labeled seeds to have a unit potential. The electric potentials established at each unlabeled node provide the probabilities that a walker originating from that node will first reach the seed with label s. These electric potentials may be calculated through the solution of a system of sparse linear equations, as described in section 2.2. The full K-tuple may be calculated by finding the potentials established through switching "on" (providing a unit voltage source to) each labeled collection of nodes and "off" (grounding) the remaining labeled nodes. Therefore, $K-1$ systems of linear equations must be solved. By linearity (i.e., the principle of superposition in circuit theory), the potentials so calculated must sum to unity. This allows us to avoid solving for one of the systems by subtracting the sum of the calculated potentials from unity to find the last entry in the full K-tuple.

There exists a small literature that addresses the problem of multi-label image segmentation without resorting to recursively applied binary segmentation (see [7] for a discussion on some concerns about recursive bisection) using both automated [8–12] and semi-automated [13–15] methods. Existing automated methods operate by defining K-way clustering heuristics over the image values or values derived from another processes (e.g., spectral coefficients). These methods attempt to cluster the image into K clusters with either a predefined K or a K chosen to satisfy a given criterion. As fully automated methods, there is no user interaction. In addition, spectral methods are not guaranteed to produce a unique solution and, indeed, may lead to a fully degenerate problem [2]. One should not be confused by the random walker interpretation of spectral methods [16] and the present method. In the former, it is illustrated that spectral methods may be interpreted as finding a subset of the nodes with cardinality equal to half the nodes in the graph, such that a random walker starting in the set is least likely to cross into the set complement. As mentioned above, this subset may not be unique, no user constraints are introduced and only a binary (i.e., 2-way) segmentation is obtained, without resorting to recursion or clustering on the spectral coefficients.

Semi-automatic methods with a similarity to our proposed algorithm also exist [17, 15]. The K-way max-flow/min-cut algorithm of [17] attempts to find the cuts with smallest value (as defined by image gradients) that separate each labeled region from all others using K-way graph cuts. Although this problem is NP-Hard, one may obtain a solution that is within a proven bound of the optimal. Our proposed algorithm has two main advantages over this semi-automated approach. First, given single pixel labels (or labels with a small boundary cost), graph cuts will produce small partitions that segment solely the labeled region from the remainder of the image. Second, our algorithm provides a confidence rating of each pixel's membership in the segmentation. An additional difference between our random walker approach and the graph cuts

algorithm is that the solution obtained by the random walker approach is guaranteed to be unique, by the uniqueness principle for harmonic functions [4], while the minimum boundary cost criterion of max-flow/min-cut may have multiple solutions (although the graph cuts algorithm of [17] always finds one of these solutions). An algorithm independently developed for machine learning by Zhu et. al [15] also finds clusters based upon harmonic functions, using boundary conditions set by a few "known" points (i.e., user-specified labels). However, by the nature of their problem domain, they employ different methods for finding similarity measures between points in feature space than are appropriate for computer vision and medical image analysis.

Additional properties of our approach that will be established later include:

1. Each segment is guaranteed to be connected to seed points with the same label, i.e., there are no isolated regions of a particular label that contain no seed points.
2. The K-tuple of probabilities at each pixel is equal to the weighted average of the K-tuples of neighboring pixels, with the weights given by walker biases.
3. The solution for the potentials is unique.
4. The expected value of the probabilities for an image of pure noise, given by identically distributed (not necessarily independent) random variables, is equal to those obtained in the neutral segmentation.
5. The expected value of the probabilities in the presence of random, uncorrelated walker biases is equal to the probabilities obtained by using walker biases equal to the mean of each bias.

This paper is organized as follows. Section 2 gives a simple weighting function to use, derives the set of linear equations that must be solved, establishes some theoretical results about the algorithm and provides implementation details. Section 3 illustrates the behavior of the algorithm in the presence of weak object boundaries and demonstrates results on real-world medical images. We conclude in Section 4 with a summary of the algorithm presented and a discussion of future work.

2 Exposition of the algorithm

This section describes four aspects of the algorithm: Generating the graph weights, establishing the system of equations to solve, proving theoretical results about the algorithm performance and the practical details of implementation.

We begin by defining a precise notion for a graph. A **graph** [18] consists of a pair $G = (V, E)$ with **vertices (nodes)** $v \in V$ and **edges** $e \in E \subseteq V \times V$. An edge, e, spanning two vertices, v_i and v_j, is denoted by e_{ij}. A **weighted graph** assigns a value to each edge called a **weight**. The weight of an edge, e_{ij}, is denoted by $w(e_{ij})$ or w_{ij}. The **degree** of a vertex is $d_i = \sum w(e_{ij})$ for all edges e_{ij} incident on v_i. In order to interpret w_{ij} as the bias affecting a random walker's choice, we require that $w_{ij} > 0$. The following will also assume that our graph is connected.

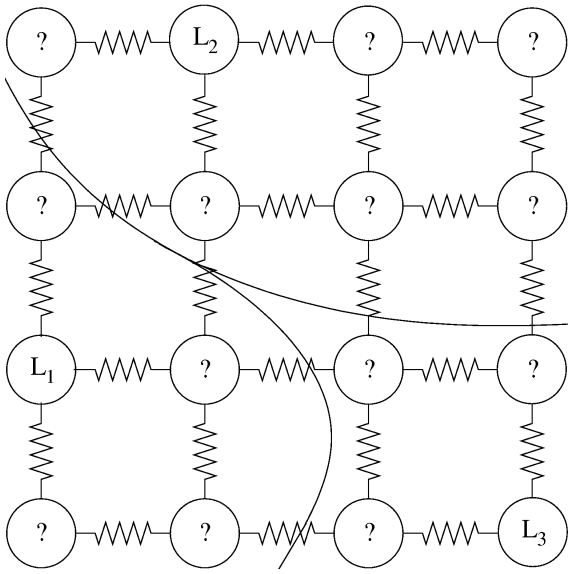
2.1 Edge weights

In order to represent the image structure (given at the pixels) by random walker biases (i.e., edge weights), one must define a function that maps a change in image intensities to weights. Since this is a common feature of graph based algorithms for image analysis, several weighting functions are commonly used in the literature [19, 17, 2]. Additionally, it was proposed in [15] to use a function that maximizes the entropy of the resulting weights. In this work we have preferred (for empirical reasons) the typical Gaussian weighting function given by

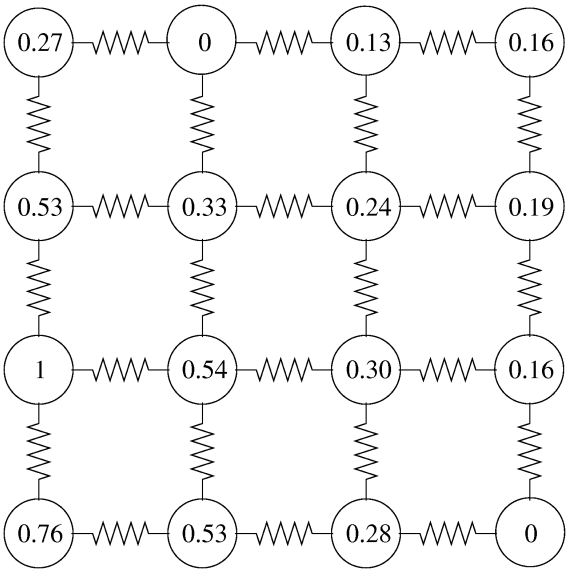
$$w_{ij} = \exp\left(-\beta(g_i - g_j)^2\right), \tag{1}$$

where g_i indicates the image intensity at pixel i. The value of β represents the only free parameter in this algorithm.

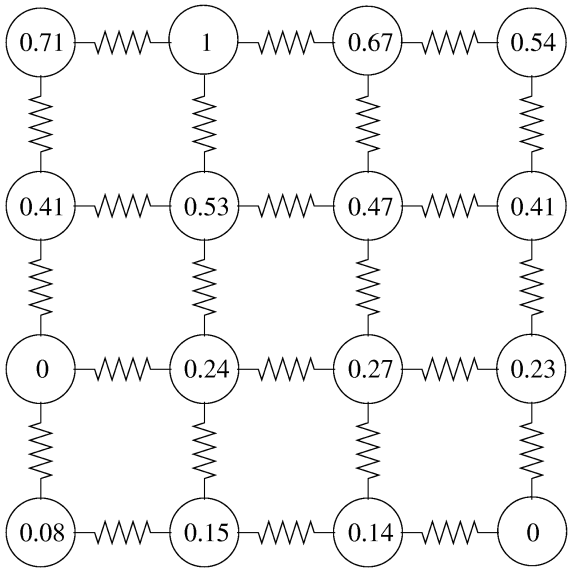
Later, we will discuss some principles for the design of an ideal weighting function for this algorithm.



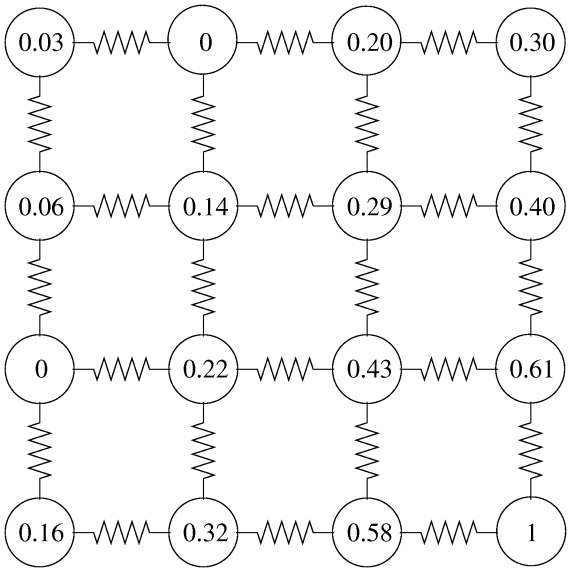
(a) Seed points with segmentation



(b) Probability that a random walker starting from each node first reaches seed L_1



(c) Probability that a random walker starting from each node first reaches seed L_2



(d) Probability that a random walker starting from each node first reaches seed L_3

Fig. 1. Illustration of the approach to segmentation. With three seed points representing three different labels (denoted L_1, L_2, L_3), alternately fix the potential of each label to unity (i.e., with a voltage source tied to ground) and set to zero (i.e., ground) the remaining nodes. The electric potentials calculated represent the probability that a random walker starting at each node first reaches the seed point currently set to unity. Figure 1(a) shows the initial seed points and the segmentation resulting from assigning each node the label that corresponds to its greatest probability. For illustration, all the weights (resistors) were set to unity. In the case of an image, these resistors would be a function of the intensity gradient. The reader can verify that the probabilities at each node sum to unity (up to rounding).

2.2 Discrete Dirichlet problem

The discrete Dirichlet problem has been discussed thoroughly in the literature [20, 4] and a convenient form for the solution in the context that we are concerned with is given in [21]. We will now review the method of solution.

Define the discrete Laplacian matrix [22] as

$$L_{v_i v_j} = \begin{cases} d_{v_i} & \text{if } i = j, \\ -w_{ij} & \text{if } v_i \text{ and } v_j \text{ are adjacent nodes,} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $L_{v_i v_j}$ is used to indicate that the matrix L is indexed by vertices v_i and v_j.

Partition the vertices into two sets, V_M (marked/seed nodes) and V_U (unmarked nodes) such that $V_M \cup V_U = V$ and $V_M \cap V_U = \emptyset$. Note that V_M contains all seed points, regardless of their label. Then, we may reorder the matrix L to reflect the subsets

$$L = \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}. \tag{3}$$

Denote the probability (potential) assumed at each node, v_i, for each label, s, by x_i^s. Define the set of labels for the seed points as a function $Q(v_j) = s$, $\forall v_j \in V_M$, where $s \in \mathbb{Z}, 0 < s \leq K$. Define the $|V_M| \times 1$ (where $|\cdot|$ denotes cardinality) marked vector for each label, s, at node $v_j \in V_M$ as

$$m_j^s = \begin{cases} 1 & \text{if } Q(v_j) = s, \\ 0 & \text{if } Q(v_j) \neq s. \end{cases} \tag{4}$$

As demonstrated in [21], the solution to the combinatorial Dirichlet problem may be found by solving

$$L_U x^s = -B m^s, \tag{5}$$

which is just a sparse, symmetric, positive-definite, system of linear equations with $|V_U|$ number of equations and the number of nonzero entries equal to $2|E|$. Since L_U is guaranteed to be nonsingular for a connected graph [6], the solution, x^s, is guaranteed to exist and be unique. Therefore, the potentials for all the labels may be found by solving the system

$$L_U X = -BM, \tag{6}$$

where X has columns taken by each x^s and M has columns given by each m^s. Therefore, there are $K - 1$ sparse linear systems to solve, where K is the total number of labels.

2.3 Theoretical properties of the algorithm

We now establish the properties of the algorithm outlined in the introduction.

The following propositions have a few practical consequences. First, if an "interpolation" is desired between the solution for a particular image and the neutral solution, this may be achieved through the addition of a constant to the weights. This situation might occur if the image data was known to be very poor and an almost semi-manual segmentation was desired by the medical practitioner. Second, the ideal weighting function would produce weights, such that the presence of independent random noise at the pixel level would produce uncorrelated multiplicative noise at the level of the weights. If such a weighting function were used, then the expected value of the potentials (and hence, the solution) in the presence of independent noise would equal the potentials found in the presence of no noise. The weighting function used in (1) does convert additive pixel noise to multiplicative noise on the weights. However, those weights are not, in general, uncorrelated. Finally, we know that the segmentation obtained with this algorithm in the case of pure noise (and presumably very close to pure noise) is the neutral segmentation, which seems appropriate.

The following two properties are discrete analogues of properties of continuous harmonic functions [5] and may be seen directly by viewing the solution to the combinatorial Dirichlet problem as a solution to the discrete Laplace equation (with Dirichlet boundary conditions), where the potential of each unlabeled node must satisfy

$$x_i^s = \frac{1}{d_i} \sum_{e_{ij} \in E} w(e_{ij}) x_j^s, \tag{7}$$

where the $x_j^s \in V$ (i.e., may include seed points).

1. A potential $0 \leq x_i^s \leq 1$, $\forall\, i, s$ (maximum/minimum principle)
2. The potential of each unlabeled node assumes the weighted average of its neighboring nodes (the mean-value theorem).

Proposition 1. *If the final segmentation is determined from the potentials using the above rule (i.e., node v_i is assigned to segment, s, only if $x_i^s > x_i^f \ \forall f \neq s$), then each node assigned to segment s is connected through a path of nodes also assigned to segment s to at least one of the seed points with label s.*

A restatement of this proposition is that the connected components generated by the final segmentation must contain at least one seed point bearing that label.

Proof. Note, this proof is similar to that given in [2].

The result follows if it can be shown that any connected subset, $P \subseteq V_U$, assigned to segment s must be connected to at least one node that is also labeled s.

A block matrix form of (7) may be written

$$L_P x_P^s = -R_P x_{\overline{P}}^s, \tag{8}$$

where $x^s = [x_P^s, x_{\overline{P}}^s]^T$,$L$ has been decomposed into the block form

$$L = \begin{bmatrix} L_P & R_P \\ R_P^T & L_{\overline{P}} \end{bmatrix}, \tag{9}$$

and $\overline{P}$ denotes the set complement of P in V. For example, in the case of $P = \{v_i\}$ in (7), $L_P = d_i$ and $-R_P x_{\overline{P}}^s = \sum_{e_{ij} \in E} w(e_{ij}) x_j^s$.

If $x_P^s > x_P^f \ \ \forall f \neq s$, then $x_P^s - x_P^f > 0$ and $-L_P^{-1} R_P (x_{\overline{P}}^s - x_{\overline{P}}^f) > 0$. The entries of R_P are nonpositive by definition of L. Since L is an M-matrix, any block diagonal submatrix of an M-matrix is also an M-matrix, and the inverse of an M-matrix has nonnegative entries (see [23] for the previous three facts), then $-L_P^{-1} R$ has nonnegative entries and therefore, some $x_i^s \in \overline{P}$ must be greater than $x_i^f \in \overline{P}$. Furthermore, since the entries of R_P are zero for nodes not connected to P, the nodes in $\overline{P}$ satisfying the inequality must be connected to a node in P. □

Proof of the remaining propositions rest on following lemma

Lemma 1. *For random variables, X, A and B, such that $X = \frac{A}{B}$, E[X]= 0 if E[A]= 0 and $B > 0$.*

Proof. By the Hölder inequality [24], $\mathrm{E}[A] = \mathrm{E}[XB] \leq \mathrm{E}[X]\mathrm{E}[B]$. By the same inequality, $\mathrm{E}[X] = \mathrm{E}[\frac{A}{B}] \leq \mathrm{E}[A]\mathrm{E}[\frac{1}{B}]$. Therefore, $\frac{\mathrm{E}[A]}{\mathrm{E}[B]} \leq \mathrm{E}[X] \leq \mathrm{E}[A]\mathrm{E}[\frac{1}{B}]$, and the result is proved. □

Since the time of Krichhoff [25], it has been known that there is a relationship between the potentials solved for in (5) and the weighted tree structure of the graph. The following relationship for the potential at node v_i in the presence of unit voltage sources (tied to ground) is given in [6, 26]

$$x_i^s = \frac{\sum_{TT \in TT_i} \prod_{e_{ij} \in TT} w(e_{ij})}{\sum_{TT \in TT_G} \prod_{e_{ij} \in TT} w(e_{ij})}, \tag{10}$$

where TT_i is the set of 2-trees present in the graph, such that node v_i is connected to a seed with label s, and TT_G is the set of all possible 2-trees in the graph. A **2-tree** is defined as a tree with one edge removed. Note that $TT_i \subseteq TT_G \ \forall v_i$ with equality holding if v_i is a seed point labeled as s. Therefore, to restate (10) in prose, if you sum over the product of the weights in every 2-tree that has v_i connected to a seed with label s and divide that sum by the sum of the product of the weights of the edges in every 2-tree that exists in the graph, that ratio is equal to the potential found by solution of (5). Although (10) is not very useful for computation of (5) (due to the massive number of 2-trees involved in any sizable graph), we can use it to prove some interesting results about the behavior of x_i^s under different choices of weights.

If the weights are uniform, the neutral case, by (10), yields potentials satisfying

$$x_i^s = \frac{|TT_i|}{|TT_G|}. \tag{11}$$

Now we are in position to prove a series of propositions about x^s under different conditions. We also note that if the weights are all multiplied by a constant, k, there will be no effect on x_i^s, since it is clear by (10) that the constant will divide out of the numerator and denominator.

Proposition 2. *If the set of weights, w_{ij}, are identically distributed (not necessarily independent) random variables, with $w_{ij} > 0$, then E[x_i^s] equals the potential obtained in the neutral segmentation.*

Proof. We proceed by simply verifying this proposition using Lemma 1. Denote the potential for label s at v_i for the neutral segmentation by n_i^s. Denote the complement of TT_i in TT_G as TT_C, such that $TT_i \cup TT_C = TT_G$ and $TT_i \cap TT_C = \emptyset$. For brevity of notation, denote $S_{TT_i} = \sum_{TT \in TT_i} \prod_{e_{ij} \in TT} w(e_{ij})$.

$$\mathrm{E}[x_i^s - n_i^s] = \mathrm{E}\left[\frac{S_{TT_i}}{S_{TT_i} + S_{TT_C}} - \frac{|TT_i|}{|TT_i| + |TT_C|}\right]. \tag{12}$$

Since each of the 2-trees will contain an equal number of edges, $(n-2)$, and all of the weights are identically distributed, S_{TT_i} will contain the sum of $|TT_i|$ identically distributed random variables. Let μ denote the mean of the distribution of these new variables.

After combining terms, the numerator of (12) is given by

$$\mathrm{E}[S_{TT_i}(|TT_i| + |TT_C|) - |TT_i|(S_{TT_i} + S_{TT_C})] = \\ \mu|TT_i|(|TT_i| + |TT_C|) - |TT_i|(\mu|TT_i| + \mu|TT_C|) = 0, \tag{13}$$

and the denominator of (12) must be strictly positive, since all the w_{ij} are guaranteed to be positive by construction.

Therefore, the conditions of Lemma 1 are satisfied for the left hand side of (12) to be equal to zero, and $\mathrm{E}[x_i^s] = n_i^s$. □

Since the same technique as above may be used to verify the following two propositions, the proofs are left to the reader.

Proposition 3. *If the set of weights, w_{ij}, are uncorrelated (not necessarily independent) random variables with corresponding means μ_{ij}, then E[x_i^s] equals the potential obtained by setting $w_{ij} = \mu_{ij}$.*

Proposition 4. *If $w_{ij} = k_{ij}y_{ij}$, where the k_{ij} are (not necessarily equal) constants and y_{ij} are identically distributed random variables, such that $y_{ij} > 0$, then E[x_i^s] equals the potential obtained by setting $w_{ij} = k_{ij}$.*

With a similar approach, we may prove

Proposition 5. *If $w_{ij} = k_{ij} + r$ where k_{ij} are (not necessarily equal) constants and r is a constant added to all weights, $\lim_{r\to\infty} x_i^s = n_i^s$, where n_i^s is the potential obtained in the neutral segmentation.*

Proof. We may write the potential for node v_i as

$$x_i^s = \frac{\sum_{TT \in TT_i} \prod_{e_{ij} \in TT} (k_{ij} + r)}{\sum_{TT \in TT_G} \prod_{e_{ij} \in TT} (k_{ij} + r)} = \frac{|TT_i| r^{N-2} + \mathcal{O}(r^{N-3})}{|TT_G| r^{N-2} + \mathcal{O}(r^{N-3})}, \tag{14}$$

where $\mathcal{O}(\cdot)$ indicates a term of order no greater than the argument. By l'Hôpital's Rule,

$$\lim_{r \to \infty} x_i^s = \frac{|TT_i|}{|TT_G|} = n_i^s. \tag{15}$$

□

2.4 Numerical practicalities

Many good sources exist on the solution to large, sparse, symmetric, linear systems of equations (e.g., [27, 28]). A direct method, such as LU decomposition with partial pivoting has the advantage that the computation necessary to solve (6) is only negligibly increased over the amount of work required to solve (5). Unfortunately, current medical data volumes frequently exceed $256 \times 256 \times 256 \approx 16\text{e}^6$ voxels, and hence require the solution of an equal number of equations. Furthermore, there is no reason to believe that the resolution will not continue to increase. The memory capabilities of most contemporary computers simply do not have enough memory to allow an LU decomposition with that number of equations.

The standard alternative to the class of direct solvers for large, sparse systems is the class of iterative solvers [29]. These solvers have the advantages of a small memory requirement and the ability to represent the matrix-vector multiplication as a function. In particular since, for a lattice, the matrix L_U has a circulant nonzero structure (although the coefficients are changing), one may avoid storing the matrix entirely. Instead, a vector of weights may be stored (or computed on the fly, if memory is at a premium) and the operation $L_U x_U^s$ may be performed very cheaply. Furthermore, sparse matrix operations (like those required for conjugate gradients) may be efficiently parallelized [30, 31], e.g., for use on a GPU. Because of the relationship of (5) to a finite differences approach to solving the Dirichlet problem on a hypercube domain, the techniques of numerical solution to PDEs may also be applied. Most notably, the algebraic multigrid method [32, 33] achieves near-optimal performance for the solution to equations like (5).

For simplicity, we have implemented the standard conjugate gradients algorithm with Jacobi preconditioning, representing the matrix-vector multiplication implicitly, as described above on an ™Intel ™Xeon 2.40GHz dual-processor with 1GB of RAM. Solution of (5) using conjugate gradients (tolerance = 1e^{-4}, sufficient for the algorithm) for a 256×256 image with two randomly placed seed points required 4.831 seconds.

2.5 Algorithm summary

To summarize, the steps of the algorithm are:

1. Obtain a set, V_M, of marked (labeled) pixels with K labels, either automatically or through the intervention of a medical practitioner.
2. Using (1), map the image intensities to edge weights in the lattice.
3. Solve (6) outright for the potentials or solve (5) for each label except the final one, f (for computational efficiency). Set $x_i^f = 1 - \sum_{s<f} x_i^s$.
4. Obtain a final segmentation by assigning to each node, v_i, the label corresponding to $\max_s (x_i^s)$.

We note that other options might be explored for assigning a label to each pixel based on the potentials (e.g., applying a clustering algorithm to the K-dimensional vectors at each node).

3 Algorithmic results

3.1 Weak boundaries

We will prefer the term object *boundary* to the traditional computer vision term *edge* (e.g., "edge detection") to avoid confusion with the edge set of the graph (e.g., $e_{ij} \in E$). Unlike region growing approaches, one aspect of the random walker motivation for this algorithm is that weak object boundaries will be found when they are part of a consistent boundary. Consider the situation of Figure 2, where a random walker starting on one side of a weak boundary wanders until first striking one of the two labeled nodes. On a four-connected lattice, the walker has three initial steps that keep it on one side of the boundary. Since other nodes on that side of the boundary are all very likely to reach seed 1 (filled circle), this walker is also very likely to first reach seed 1. For the same reasons, a walker on the other side of the weak boundary is also very likely to first reach seed 2 (open circle). Consequently, there will be a sharp potential drop (i.e., voltage) over the entire boundary, resulting in the correct segmentation. Figure 3 shows the segmentation obtained for a synthetic image with four areas of varying sizes and convexity with weak boundaries and few labeled nodes.

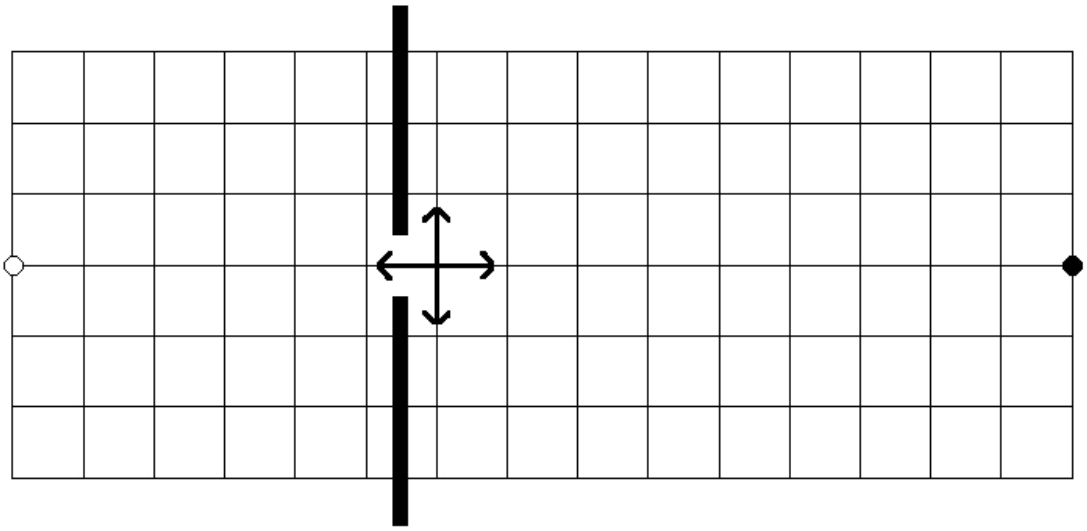


Fig. 2. Illustration of why the segmentation obeys weak image boundaries. Consider the 16×7 image consisting of just one hard boundary with a hole, represented by the thick black line, and two seed points placed at the white and black circles at the far ends of the image. A random walker starting at the pixel next to the weakness in the boundary (the center of the arrows) has 3 out of 4 chances on its initial step to enter into the region that is likely to be labeled as belonging to the black circle. Since the same holds true on the other side of the weak boundary, there will be a sharp drop in the probabilities and consequently, the segmentation will respect the boundary, even though it is weak.

3.2 Segmentation of real images

Figure 4 shows the segmentation results on two CT cardiac images and two MR brain images. The images and seeds were chosen to demonstrate the general applicability of the semi-automated segmentation approach on objects of varying uniformity, size, shape, contrast and topology. The first two figures, 4(a) and 4(b), demonstrate the segmentation results on two CT cardiac images. Figure 4(a) uses four groups of seeds, corresponding to the left ventricle, right ventricle, myocardial wall and the background, while Figure 4(b)

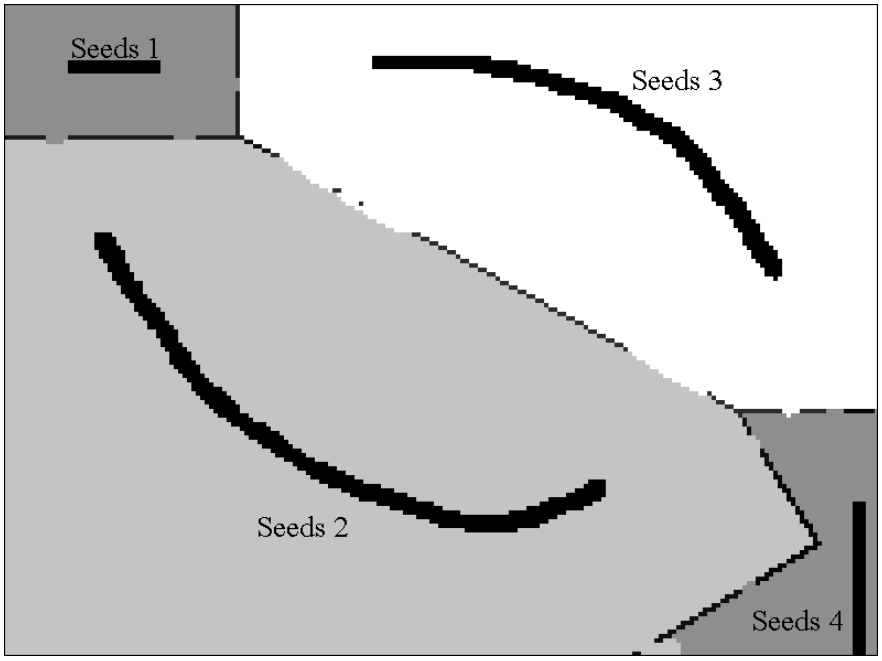


Fig. 3. Demonstration of the algorithm response to weak boundaries of different types, large/small regions and nonconvex regions on a synthetic image consisting of only black and white pixels. The thin black lines represent the "object" boundaries in the image, the thick black patches represent the labeled seeds of four different types and the shaded regions correspond to the resulting segmentation.

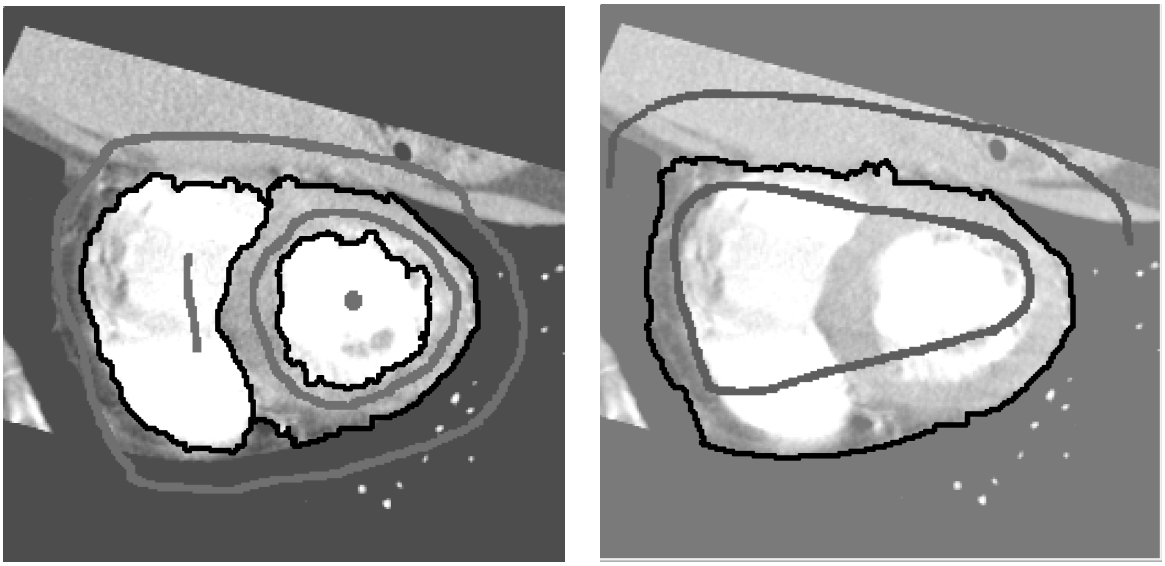
segments the entire heart from the background. The two neural MR images have three segments each. Figure 4(c) shows the corpus callosum and cerebellum segmented from the background and Figure 4(d) shows the thalamus and cerebellum segmented from the background. In each segmentation, the value of the one free parameter, β in (1), was kept constant, despite the different image characteristics of CT and MR images. Space constraints limit us from showing 3D segmentations, but we stress that the algorithm is defined on a lattice of arbitrary dimension.

4 Conclusion

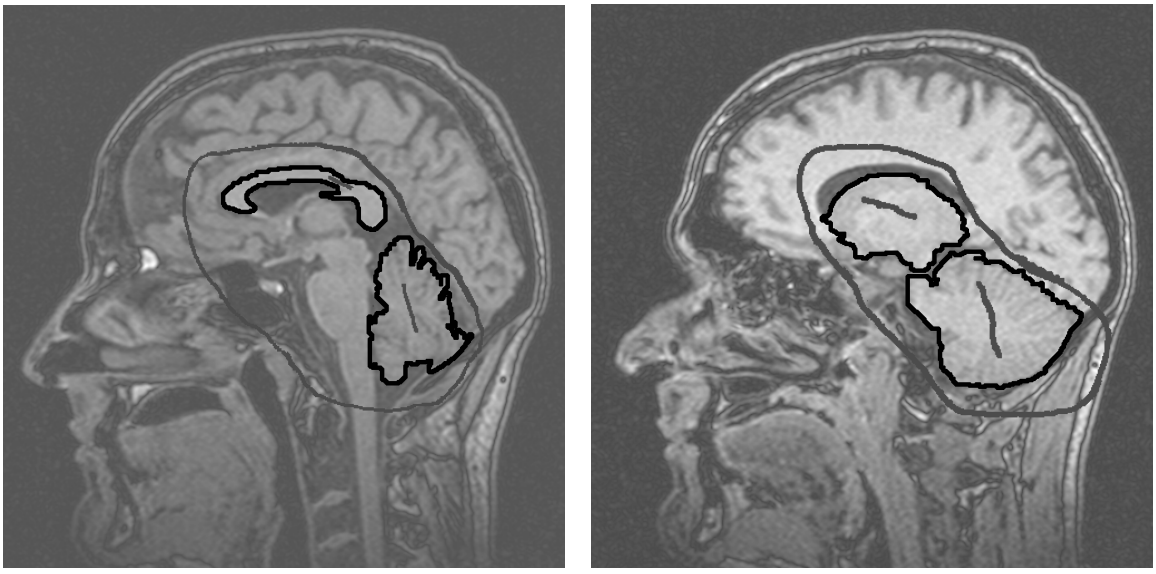
We have presented a novel algorithm for general image segmentation based on a small set of pre-labeled pixels. These pre-labeled pixels may be either generated automatically for a particular purpose, or they may be given directly by a medical practitioner. The algorithm functions by assigning each unlabeled pixel to the label of the seed point that a random walker starting from that pixel would be most likely to reach first, given that it is biased to avoid crossing object boundaries (i.e., intensity gradients). Because the algorithm is formulated on a general graph, and produces segmentations based on the separation of quantities defined at the nodes (i.e., potentials), the graph (lattice) may represent any dimension or topology.

We have demonstrated this approach on real images and shown that it provides a unique, quality, solution that is robust to weak object boundaries and that the solution respects the medical practitioner's pre-labeling choices. Furthermore, there is only a single free parameter, β in (1), and all of the segmentations shown in this paper were used with the same choice of that parameter. Of course, this approach could also be combined with pre-filters (e.g., median) or post-filters (e.g., clustering) to produce enhanced, problem-specific results.

The connections between random walks, combinatorial potential theory and electric circuits allowed us to prove that the segments are guaranteed to be connected (i.e., smooth), and that the segmentation approaches the neutral segmentation (i.e., Voronoi-like) as increased amounts of random noise are introduced. Furthermore, the direct correspondence with analog electric circuits opens the possibility for a hardware (e.g.,



(a) (b)



(c) (d)

Fig. 4. Examples of segmentations on CT cardiac data and MR brain images. The thick, dark gray lines (chosen to maximize contrast) represent the seed points and the thick black lines represent the segment boundaries. For clarity of exposition, seed points of each type are connected, although this is not necessary. $\beta = 900$ for all segmentations.

VLSI) implementation of the algorithm, where the physics of the circuit perform the same "computation" as the standard CPU, except at the extremely fast speed of the natural world. Finally, since our PDE (effectively the Laplace equation with Dirichlet boundary conditions) is formulated on a graph, there are no concerns about discretization errors or variations in implementation that sometimes cause problems for other PDE-based approaches.

Future work will concentrate on a faster implementation, clinical evaluation, systematic comparison with other techniques, the use of prior information in the segmentation and leveraging the theoretical results to produce a more effective weighting function.

Acknowledgments

The authors would like to thank Marie-Pierre Jolly and Yuri Boykov for advice and criticism during the preparation of this manuscript.

References

1. Wallace, R., Ong, P.W., Schwartz, E.: Space variant image processing. International Journal of Computer Vision **13** (1994) 71–90
2. Grady, L.: Space-Variant Computer Vision: A Graph-Theoretic Approach. PhD thesis, Boston University, Boston, MA (2004)
3. Kakutani, S.: Markov processes and the Dirichlet problem. Proc. Jap. Acad. **21** (1945) 227–233
4. Doyle, P., Snell, L.: Random walks and electric networks. Number 22 in Carus mathematical monographs. Mathematical Association of America, Washington, D.C. (1984)
5. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Volume 2. John Wiley and Sons (1989)
6. Biggs, N.: Algebraic Graph Theory. Number 67 in Cambridge Tracts in Mathematics. Cambridge University Press (1974)
7. Simon, H.D., Teng, S.H.: How good is recursive bisection? SIAM Journal of Scientific Computing **18** (1997) 1436–1445
8. Chan, P.K., Schlag, M.D.F., Zien, J.Y.: Spectral k-way ratio-cut partitioning and clustering. IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems **13** (1994) 1088–1096
9. Alpert, C.J., Yao, S.Z.: Spectral partitioning: the more eigenvectors, the better. In: Proceedings of the 32nd ACM/IEEE conference on Design automation conference, ACM/IEEE, ACM Press (1995) 195–200
10. Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **18** (1996) 884–900
11. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: Ninth IEEE International Conference on Computer Vision. Volume 1 of International Conference on Computer Vision., Nice, IEEE Computer Society, IEEE COmputer Society (2003) 313–319
12. Bar-Hillel, A., Weinshall, D.: Learning with equivalence constraints and the relation to multiclass learning. In: 16th Conference on Learning Theory, Washington DC (2003) 640–654
13. Boykov, Y., Jolly, M.P.: *Interactive graph cuts* for optimal boundary & region segmentation of objects in N-D images. In: International Conference on Computer Vision. Volume I. (2001) 105–112
14. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 1222–1239
15. Zhu, X., Lafferty, J., Ghahramani, Z.: Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the ICML 2003 workshop on The Continuum from Labeled to Unlabel Data in Machine Learning and Data Mining. (2003) 58–65
16. Meilă, M., Shi, J.: Learning segmentation by random walks. In: Advances in Neural Information Processing Systems. Volume 13. (2000) 873–879
17. Boykov, Y., Veksler, O., Zabih, R.: A new algorithm for energy minimization with discontinuities. In Pelillo-M., H.E.R., ed.: Energy Minimization Methods in Computer Vision and Pattern Recognition. Second International Workshop, EMMCVPR'99, York, UK, 26-29 July 1999. (1999) 205–220
18. Harary, F.: Graph Theory. Addison-Wesley (1994)
19. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 888–905
20. Biggs, N.: Algebraic potential theory on graphs. Bulletin of London Mathematics Society **29** (1997) 641–682

21. Grady, L., Schwartz, E.: Anisotropic interpolation on graphs: The combinatorial Dirichlet problem. Technical Report CAS/CNS-TR-03-014, Department of Cognitive and Neural Systems, Boston University, Boston, MA (2003)
22. Merris, R.: Laplacian matrices of graphs: A survey. Linear Algebra and its Applications **197,198** (1994) 143–176
23. Fiedler, M.: Special matrices and their applications in numerical mathematics. Martinus Nijhoff Publishers (1986)
24. Feller, W.: An Introduction to Probability Theory and its Applications. Second edn. Volume II. John Wiley and Sons, Inc. (1971)
25. Kirchhoff, G.: Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. Poggendorf's Annalen der Physik Chemie **72** (1847) 497–508
26. Chen, W.K.: Applied Graph Theory: Graphs and Electrical Networks. 2nd edn. Applied Mathematics and Mechanics. North-Holland Publishing Company (1976)
27. Golub, G., Van Loan, C.: Matrix Computations. 3rd edn. The Johns Hopkins University Press (1996)
28. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing. 2nd edn. Cambridge University Press (2002)
29. Hackbusch, W.: Iterative Solution of Large Sparse Systems of Equations. Springer-Verlag (1994)
30. Dongarra, J.J., Duff, I.S., Sorenson, D.C., van der Vorst, H.A.: Solving Linear Systems on Vector and Shared Memory Computers. Society for Industrial and Applied Mathematics, Philadelphia (1991)
31. Gremban, K.: Combinatorial preconditioners for sparse, symmetric diagonally dominant linear systems. PhD thesis, Carnegie Mellon University, Pittsburgh, PA (1996)
32. Shapira, Y.: Matrix-Based Multigrid: Theory and Applications. Volume 2 of Numerical Methods and Algorithms. Kluwer Academic Publishers (2003)
33. Dendy, J.E.: Black box multigrid. Journal of Computational Physics **48** (1982) 366–386