Increasing Accuracy of Optimal Surfaces Using Min-Marginal Energies

Jens Petersen, Andres M. Arias-Lorza, Raghavendra Selvan, Daniel Bos, Aad van der Lugt, Jesper H. Pedersen, Mads Nielsen, and Marleen de Bruijne

***Abstract*— Optimal surface methods are a class of graph cut methods posing surface estimation as an n-ary ordered labeling problem. They are used in medical imaging to find interacting and layered surfaces optimally and in low order polynomial time. Representing continuous surfaces with discrete sets of labels, however, leads to discretization errors and, if graph representations are made dense, excessive memory usage. Limiting memory usage and computation time of graph cut methods are important and graphs that locally adapt to the problem has been proposed as a solution. Min-marginal energies computed using dynamic graph cuts offer a way to estimate solution uncertainty and these uncertainties have been used to decide where graphs should be adapted. Adaptive graphs, however, introduce extra parameters, complexity, and heuristics. We propose a way to use min-marginal energies to estimate continuous solution labels that does not introduce extra parameters and show empirically on synthetic and medical imaging datasets that it leads to improved accuracy. The increase in accuracy was consistent and in many cases comparable with accuracy otherwise obtained with graphs up to eight times denser, but with proportionally less memory usage and improvements in computation time.**

***Index Terms*— Segmentation, graph algorithms, airways, carotid arteries.**

I. Introduction

GRAPH cut methods are widely used in medical image analysis to solve problems such as segmentation, registration, and reconstruction. Problems are modeled in terms of finding labels for variables minimizing a measure of energy. The variables and their dependencies form a Markov random field (MRF) and the result is, for certain classes of problems, the maximum a posteriori (MAP) estimate of the underlying posterior distribution [1], [2]. Solutions can be found in low-order polynomial time by formulating the energy minimization as a maximum flow/minimum cut problem [3]. For this to be possible, the energy functions have to be graph representable [4], which is the case for many binary labeling problems. Optimal solutions, in the sense of being MAP, can also be efficiently found for some n-ary ordered labeling problems where dependencies are pairwise and convex [1]. In other cases, heuristics involving repeated use of graph cuts with binary labels often approximate optimal solutions well [4].

Optimal surface methods [5], [6] are a class of graph cut methods, that pose the problem of finding surfaces as an n-ary ordered labeling problem. They have been used extensively in medical imaging because of their ability to find multiple interacting and layered surfaces. Some examples include segmentation of: airways [7], carotid arteries [8], intraretinal layers [9], knee joint cartilage [10], hip cartilage [11], bladder and prostate [12], and heart atrium wall [13]. One can also use them to model interaction of surfaces in different images, for instance across modalities [14], or across time [15], [16]. In optimal surface methods, labels encode possible positions for points on the sought surfaces. Surfaces are, however, difficult to accurately represent with such discrete sets of positions. If the sought surfaces can fall into a wide range of positions, then many labels have to be used or the result is likely to be inaccurate. A greater number points and labels, however, lead to increased computation time and memory requirements.

Memory requirements that escalate with accuracy demands is a problem not limited to optimal surface methods, but with graph cuts in general. Methods that seek to limit memory usage are actively pursued [17]–[20]. However, if the problem is large enough, one may be forced to split it up into smaller sub-problems and solve them either sequentially or in parallel [19]–[22]. This is difficult [19], [22] and heuristical methods are often required. A different approach altogether is to make graphs smaller by locally adapting their resolution using multiresolution banded schemes [23]–[25].

Dynamic graph cuts [26] are relevant for locally adaptive graphs as they allow small updates of the minimum cut graph to be computed fast. However, they also allow the minimum energy of solutions where given variables take on different

Manuscript received October 6, 2018; revised December 19, 2018; accepted December 21, 2018. Date of publication January 1, 2019; date of current version June 28, 2019. This work was supported in part by the Danish Council for Independent Research and in part by the Netherlands Organization for Scientific Research. *(Corresponding author: Jens Petersen.)*

J. Petersen, R. Selvan, and M. Nielsen are with the Department of Computer Science, University of Copenhagen, 2100 Copenhagen, Denmark (e-mail: phup@diku.dk).

A. M. Arias-Lorza is with the Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, 3015 Rotterdam, The Netherlands.

D. Bos and A. van der Lugt are with the Department of Radiology, Erasmus MC, 3015 Rotterdam, The Netherlands.

J. H. Pedersen is with the Department of Thoracic Surgery, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

M. de Bruijne is with the Department of Computer Science, University of Copenhagen, 2100 Copenhagen, Denmark, and also with the Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, 3015 Rotterdam, The Netherlands.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2018.2890386

labels to be computed. These so-called min-marginal energies can be used to estimate solution uncertainty, by associating high uncertainty with variable assignments that can be changed from their optimal configurations with little added energy cost. Uncertainty measured in this way [27] is a probabilistic value indicating how sure we are that a particular variable has a particular label. Uncertainty is useful in choosing where to refine or adapt graphs [28], leading to more accurate solutions at similar graph sizes. Similar approaches were used to increase the sparsity of graphs and labels in concurrent segmentation and atlas registration [29] and in optimal flow estimation [30].

The works of [28]–[30] show that min-marginal energies can be used to estimate labels more accurately. However, the min-marginal information is used indirectly, to decide how to locally change and adapt the resolution of the graphs. Whenever a graph is changed, the underlying problem is inherently modeled differently as dependencies between neighboring variables at different resolution levels are different. There are also extra free parameters used to decide the details of the heuristics involved. Applying these methods may therefore require extra parameter tuning and training data. Instead, the information provided by min-marginal energies may be more generally useful, if it can be exploited to improve accuracies for a given resolution graph directly.

In this work, we use min-marginal energies to compute a continuous solution, which locally approximates the minimum energy of the corresponding continuously valued MRF. We focus on optimal surface methods and investigate empirically if these continuous labels increase accuracy by decreasing discretization errors. We also investigate if possible accuracy improvements are justified in terms of computation time and memory usage compared to results which could otherwise be obtained by increasing graph resolution. The work addresses two important medical image segmentation problems, where state of the art solutions have been obtained with optimal surfaces, CT images of airways and MR images of carotid arteries, as well as synthetically generated data.

II. Background

The purpose of this section is to provide just enough background information on graph cuts with ordered labels, optimal surface methods, min-marginal energies, and certainties to understand the present work. We refer to the original works [1], [6], [28] and the myriad of application papers mentioned in the introduction for further possibilities and graph constructions. Optimal surface methods were introduced by [6], but a similar graph construction and one that can solve the same set of problems has been suggested by [1]. Reference [6] specifically target surface finding, whereas [1] describes the method in the context of finding optimal label configurations for random variables in Markov random fields. We follow the latter to a large degree and describe the methods as solving labeling problems, to remain as general as possible.

A. An Introduction to Graph Cuts With Ordered Labels

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with vertices $\mathcal{V}$ and edges $\mathcal{E}$, we seek the configuration of a set of random variables

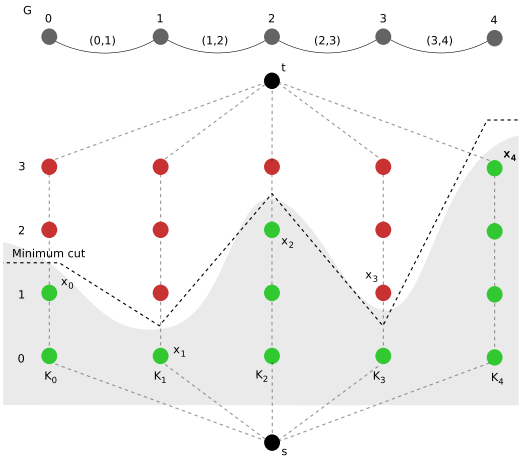

Fig. 1. Illustrating contour approximated by optimal surfaces. The example graph $\mathcal{G}$ (top) has vertices $\mathcal{V} = \{0, 1, 2, 3, 4\}$ and edges $\mathcal{E} = \{(0, 1), (1, 2), (2, 3), (3, 4)\}$. Each variable can take one of four labels associated with a column position (K_v, $v \in \mathcal{V}$) in the graph $\mathcal{G}_{s,t}$ (bottom) (note: edges are left out for readability). Also shown are the source (s), source set (green), sink (t), sink set (red), minimum cut, and implied labelling given by $x_0 = 1, x_1 = 0, x_2 = 2, x_3 = 0$ and $x_4 = 3$ with the energy $f_0(1) + f_1(0) + f_2(2) + f_3(0) + f_4(3) + g_{0,1}(1) + g_{1,2}(2) + g_{2,3}(2) + g_{3,4}(3)$.

$\mathcal{X} = \{x_v | v \in \mathcal{V}\}$, which minimizes the energy:

$$E(\mathcal{X}) = \sum_{v \in \mathcal{V}} f_v(x_v) + \sum_{(u,v) \in \mathcal{E}} g_{u,v}(|x_u - x_v|), \qquad (1)$$

where $g_{u,v}(x)$ is a convex function, $x_v \in K_v = \{L_v, L_v + 1, \cdots, U_v\}$, and L_v and U_v are integers and $L_v < U_v$.

The functions f_v give the cost of assigning a particular label to the particular variable. This is typically derived from data. In image segmentation, for instance, it may be based on information from image gradients [7], [8]. The edges $\mathcal{E}$ of the graph, $\mathcal{G}$ are used to model dependencies in the labels of vertices, for instance, to regularize solutions. For each pair of vertices u and v, $g_{u,v}$ assigns a pairwise cost to the solution given by x_u and x_v, a function which only depends on the absolute difference in x_u and x_v. Fig. 1 shows a simple optimal surface problem.

The minimisation of (Eq. 1) can be done by solving it as a maximum flow/minimum cut problem. To do this, a minimum cut graph $\mathcal{G}_{s,t} = (\mathcal{V}_{s,t}, \mathcal{E}_{s,t})$ is constructed, which is a directed graph with two special vertices, the source s and the sink t. Each edge in $(u, v) \in \mathcal{E}_{s,t}$, is associated with a capacity, $c(u, v) \geq 0$. An $s-t$-cut, henceforth just called cut, in such a graph, is a partition of the graph vertices into two disjoint sets, such that the two terminal vertices, the source s and the sink t vertices, are in different sets. The cost of a cut is equal to the sum of the capacities of the edges going from the source set, the set where s is in, to the sink set, the set t is in. $\mathcal{G}_{s,t}$ is constructed such that the cost of a minimum cut is only a constant from the energy (Eq. 1) of the configuration implied by the cut. In this way, the minimum cut will also give the minimal energy configuration.

To understand how a configuration is implied by the cut, we will first need to know a little more about $\mathcal{G}_{s,t}$. For each label in the columns $K_0, K_1, \cdots, K_{|\mathcal{V}|}$ there is a corresponding vertex in $\mathcal{G}_{s,t}$ and the capacity of edges between these

vertices and the terminal vertices are carefully chosen such that the f_v and $g_{u,v}$ terms are added to the cost of the cut [1], [6]. The graph is constructed such that the minimum cut intersects each corresponding column of vertices in $\mathcal{G}_{s,t}$ exactly once and this intersection point uniquely identifies the optimal label for each x_v. Note that, all we get from the minimum cut solution is whether each vertex of $\mathcal{G}_{s,t}$ belongs to the source or sink set. For any given column, the intersection point is between the vertex of the source set whose corresponding label is the largest and the vertex of the sink set whose corresponding label is the smallest. The intersection point thus corresponds to a unique entry k in the column of possible labels K_v for each $v \in \mathcal{V}$. Fig. 1 illustrates these concepts.

Computing the minimum cut can be done using many different algorithms. In this work, we use the algorithm described in [3]. Its worst-case running time of $O(\mathcal{E}_{s,t}\mathcal{V}_{s,t}^2|C|)$, where C is the cost of the minimum cut, is not the lowest known, however, it is relevant as it is particularly efficient for many problems encountered in computer vision [3].

B. Optimal Surface Methods in Medical Imaging

In finding surfaces, the approach that is typically followed is to let the vertices $\mathcal{V}$ of the graph $\mathcal{G}$ parameterize the sought surface(s). Each column K_v of possible labels for x_v represents indices into a given vector of possible surface positions at the position parameterized by $v \in \mathcal{V}$.

In some applications such as macular segmentation in optical coherence tomography images [9], [31], the sought surfaces can be assumed to be roughly aligned with image axes. The graph can then be constructed such that each vertex of $\mathcal{V}$ corresponds to an x-y grid coordinate of the image and each associated column consists of indices to the particular z-axis voxel column at this x-y coordinate. The optimal label for a given v then gives the particular z coordinate where the optimal surface is located for the associated x-y coordinate. One can also use the technique to refine initial rough segmentations by forming the graph $\mathcal{G}$ from the initial segmentation surface. Possible surface positions can be extracted along columns running approximately orthogonal to the initial surface at each $v \in \mathcal{V}$. In this way, a set of alternative positions can be explored [7], [8], [10]–[13].

Besides, there are many interesting possibilities, which we will exploit in the experiments, but otherwise not explain in detail here, as they do not affect the way min-marginal energies or certainties are computed, such as the possibility of looking for multiple coupled surfaces at the same time [32], adding hard constraints [6] and regional terms [31]. One purpose of hard constraints, when looking for multiple surfaces, is to enforce that surfaces are layered and that they do not overlap. To understand what is meant by regional terms, one should realize that the term $f_v(x_v)$ in (Eq. 1) can be thought of as the cost of placing the surface at the index x_v. Regional terms, in contrast, define a cost of particular positions either being inside, outside or between specific pairs of surfaces. In medical imaging, surfaces often mark changes in tissue type, and regional terms thus make it possible to model appearances of different tissue types more directly. More details on these aspects can be found in the supplementary material.

C. Computing Min-Marginals and Certainties

In [27], a measure of certainty in a given label assignment is computed from its associated min-marginal energy. In the following we will reiterate the definitions of [27] and show how it can be done in optimal surface graphs.

The min-marginal energy of a particular label assignment $x_v = j$, is defined as

$$\begin{aligned}\psi(x_v = j) &= \min_{\mathcal{X}, x_v = j} E(\mathcal{X}) \\ &= \min_{x_1,\ldots,x_{v-1},x_{v+1},\ldots,x_n} E(x_1, \ldots, x_{v-1}, j, x_{v+1}, \ldots, x_n),\end{aligned} \quad (2)$$

that is, it is the minimal energy configuration where x_v is fixed to j. Confidence or certainty in the label assignment $x_v = j$ is then estimated, using the softmax function and the min-marginal energy of all possible labels for x_v as

$$\phi(x_v = j) = \frac{\exp\big(-\psi(x_v = j)\big)}{\sum_{l \in K_v} \exp\big(-\psi(x_v = l)\big)}. \quad (3)$$

Computing certainty of variable assignments (Eq. 3) involves computing a minimum cut for each possible assignment. In each of these cuts a variable is fixed to a particular value by forcing the corresponding vertex in $\mathcal{G}_{s,t}$ to be part of respectively the source or the sink set. This is done by adding infinite cost edges either from the source to the corresponding vertex or from the corresponding vertex to the sink. Recomputing the minimum cut for each assignment is a large computational task, however, dynamic graph cuts [26] allow small graph changes to be computed fast. This is because the time to recompute the minimum cut after an update of the graph is proportional to the change in minimum cut cost [26], which is typically small for a single label assignment.

III. Method

To obtain more accurate and continuous labels, the first thing one may think of is to interpret certainty $\phi(x_v = j)$ as a probability of variable x_v having label j. In this way the possible labels still belong to the discrete set K_v, but it is possible to weigh each of the labels' contribution to the overall result. In many cases, however, this would be a bad idea because uncertainties as computed by Eq. 3 are not exact marginal probabilities and can be exponentially far from these [33]. Our strategy is instead to use certainties to derive a continuous label for x_v, in the interval $[\min(K_v), \max(K_v)]$ such that certainty of the actual value of x_v being either lesser or greater than the derived continuous label is both 0.5.

Eq. 3 gives the certainty of a particular n-ary label assignment, however, we will look at with what certainty we know that a given variable is greater than or equal to a given label, that is,

$$\phi(x_v \geq j) = \frac{\exp\big(-\psi(x_v \geq j)/T\big)}{\exp\big(-\psi(x_v \geq j)/T\big) + \exp\big(-\psi(x_v < j)/T\big)}, \quad (4)$$

where the relevant min-marginal energies $\psi(x_v < j)$ and $\psi(x_v \geq j)$ are given by:

$$\psi(x_v < j) = \min_{\mathcal{X}, x_v < j} E(\mathcal{X}) \quad (5)$$

and

$$\psi(x_v \geq j) = \min_{\mathcal{X}, x_v \geq j} E(\mathcal{X}), \tag{6}$$

for $v \in \mathcal{V}$. Note that compared to the certainty defined by [27], we also add the parameter $T > 0$ to make it possible to adjust how differences in min-marginal energies relate to differences in certainties. It is analogous to the temperature parameter used in simulated annealing methods, in that, large T tend to flatten the distribution of certainties without affecting the ranking of the configurations. Alternatively, small T tends to increase the peaks of the distribution. Implicitly assuming $T = 1$, as [27], in situations where the energies are of a different scale, can lead to either too much or too little certainty in the optimal solution, that is, assuming k is the optimal discrete label for x_v, we may get results like $\phi(x_v \geq j) \approx 1$ for $j \leq k$ and $\phi(x_v \geq j) \approx 0$ for $j > k$, or $\phi(x_v \geq j) \approx 0.5$ for all j.

It follows from the definition that the certainty $\phi(x_v \geq j)$ for a given column (Eq. 4) is monotonously decreasing with the index j. Let us in the following assume that there is a unique configuration minimizing the energy in (Eq. 1), and get back to the alternative later. The optimal solution k for a given variable x_v then satisfies $\phi(x_v \geq j) > 0.5$ for $j \leq k$ and $\phi(x_v \geq j) < 0.5$ for $j > k$. It follows that $\phi(x_v \geq j) = 0.5$ does not exist for any $j \in K_v$, but we will interpolate it, by fitting a continuous function to the label and certainty pairs. Assuming that each label in the continuous setting represents the midpoint of equal sized intervals, the interpolated continuous label where this function equals 0.5 should be somewhere between $k - 0.5$ and $k + 0.5$. Using linear interpolation the value is expressed as:

$$\phi_{1/2}(x_v) = k - 0.5 + \frac{0.5 - \phi(x_v \geq k)}{\phi(x_v \geq k+1) - \phi(x_v \geq k)}. \tag{7}$$

To compute this value, we only need $\psi(x_v < k)$, $\psi(x_v \geq k)$, and $\psi(x_v \geq k+1)$, because $\psi(x_v < k+1) = \psi(x_v \geq k)$. That is, besides the overall minimum energy, $\psi(x_v \geq k)$, we need to compute the min-marginal energy of configurations where the cut intersects column K_v below k, and above k. In cases where the labels are discrete samples of a continuous quantity, it seems reasonable to expect that this new continuous label is a better solution than the optimal discrete label.

Getting back to the situation, where there are multiple discrete label choices of x_v minimizing (Eq. 1). Suppose $x_v = k$ is one such label choice and it is the one implied by the minimum cut, we then know that at least one of the configurations where x_v is larger or smaller than k has the same min-marginal energy as the optimal configuration. In other words, we know that $\psi(x_v \geq k-1) = \psi(x_v \geq k)$ and/or $\psi(x_v \geq k+1) = \psi(x_v \geq k)$. This implies that $\phi(x_v \geq k) = 0.5$ and/or $\phi(x_v \geq k+1) = 0.5$. From this we conclude, that if we have multiple optimal configurations for a given variable, then the 0.5 certainty value exists at one or more of the discrete label choices for this variable. One of the labels will be the optimum selected by the minimum cut method, and we may simply stick to this solution.

As the aforementioned certainties are not exact marginal probabilities [33], a source of error could be introduced by replacing the discrete solution with one decided by certainties. A crucial question is thus, how big can this error be? Assuming that k is the optimal label for x_v, the interpolated label $\phi_{1/2}(x_v)$ where certainty is 0.5, satisfies $k - 0.5 \leq \phi_{1/2}(x_v) \leq k + 0.5$. So compared to the discrete solution, the error is bounded by half the size of the chosen graph resolution.

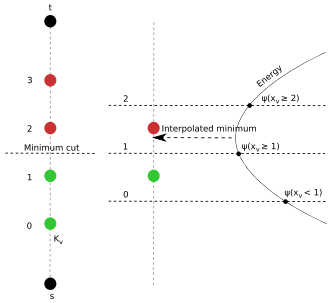

Fig. 2. Using min-marginal energies around the optimal discrete label to estimate a continuous label at the minimum of a parabola fit to the energy of the discrete optimal solution ($\psi(x_V \geq 1)$) and the min-marginal energy of solutions above ($\psi(x_V \geq 2)$) and below ($\psi(x_V < 1)$) the optimum.

A. Effect of the Temperature Parameter

We have yet to explain how the temperature parameter T (Eq. 4) affects the proposed continuous label solution. Suppose $x_v = k$ is a unique optimal configuration for a particular vertex $v \in \mathcal{V}$. The following properties are easy to derive:

$$\lim_{T \to 0} \phi(x_v \geq j) = \begin{cases} 1 & \text{if } j \leq k \\ 0 & \text{if } j > k \end{cases}$$

and

$$\lim_{T \to \infty} \phi(x_v \geq j) = \begin{cases} 0.5^+ & \text{if } j \leq k \\ 0.5^- & \text{if } j > k. \end{cases}$$

Here the $+$ and $-$ are a slight misuse of notation to indicate that the 0.5 value is approached from the positive and negative sides, respectively. Computing where $\phi(x_v \geq j)$ equals 0.5 in the limit at $T = 0$ using (Eq. 7), thus simply gives k. It is a bit more complicated to determine what happens at the limit at $T = \infty$. Inserting (Eq. 4) into (Eq. 7) and letting T approach infinity we are left with the following[1]

$$\lim_{T \to \infty} \phi_{1/2}(x_v) = k - 0.5 + \frac{\psi(x_v < k) - \psi(x_v \geq k)}{\psi(x_v \geq k+1) + \psi(x_v < k) - 2\psi(x_v \geq k)}. \tag{8}$$

This value is equivalent to the position of the minimum of a parabola fit to the coordinates $\big(j-1, \psi(x_v < k)\big)$, $\big(j, \psi(x_v \geq k)\big)$, and $\big(j+1, \psi(x_v \geq k+1)\big)$. That is, the proposed continuous label solution coincides with the minimum of a second order interpolation of the min-marginal energies, when T approaches ∞. Fig. 2 illustrates this.

In letting Eq. 7 define the solution, T can thus be thought of as a weighting of the contributions of the discrete optimal configuration ($T \to 0$) and a continuous label configuration ($T \to \infty$), which minimizes a second order function fit to the min-marginal energy.

[1] Derived using Maple 2017.

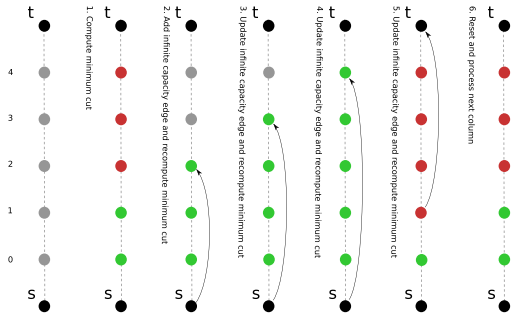

Fig. 3. Vertex visiting order in an example column when computing min-marginals in steps from left to right. The variable is constrained with infinite capacity edges from and to the source (*s*) and sink (*t*). In steps 2, 3, and 4 the variable is forced to be greater than or equal to 2, 3, and 4, respectively. In step 5, the variable is forced to be less than 1. In each step, green or red color indicate whether a vertex belongs to the source or sink set, whereas gray color indicates that a vertex can belong to either set.

B. A Few Notes on Implementation

When a vertex is fixed as being part of the source set, then all vertices with smaller indices in the column are also part of the source set. Similarly, when a vertex is fixed as being part of the sink set, then all vertices with larger indices in the column are also part of the sink set. The number of vertices changing set and therefore the total change in cost is thus typically smaller if certainties are computed by visiting vertices in the direction of the columns. This is illustrated in Fig. 3. Note that to interpolate the position where certainty equals 0.5 (Eq. 7), of the steps shown in the figure only steps 1, 2, 5, and 6 are needed. This is because, as explained previously, besides the overall minimum energy, only the min-marginal energies of the two configurations where the cut intersects the column above and below k are needed. An algorithmic summary and more details can be found in the supplementary material.

IV. Experiments and Results

We will investigate the accuracy and computation time of labels corresponding to an interpolated 0.5 certainty compared to the discrete optimal labels.

We experiment with solving three different problems: one synthetic and two real medical imaging problems. The latter two are airway wall segmentation in chest CT images, and carotid artery wall segmentation in MR images. Optimal surface methods have been used for both these problems previously with state of the art results [7], [8] and we adopt the approach of these papers, including the extraction of initial surfaces and image preprocessing, and only give a short overview here. Both methods solve two-surface problems by constructing the graph from a rough initial segmentation. A closed triangular mesh is constructed on the surface of the initial segmentation. We use the meshing algorithm of [34], which gives more smooth results than the mesh used in the original works [7], [8]. The graph consists of two subgraphs, one for each sought surface. The columns of possible surface positions, or labels, are sampled along flow lines in a Gaussian smoothed version of the initial segmentation, inward and outward from each of the initial surface mesh vertices. The data terms $f_v(x_v)$ are derived from weightings of the first and second order derivative of the image intensity in the direction of the columns/flow lines. Pairwise costs, specific to each surface, of $g_{u,v}(|x_u - x_v|) = p_m|x_u - x_v|, m \in \{\text{interior, exterior}\}$ are added. Here larger values of p_m make the surfaces vary less among neighboring mesh vertices, so it is referred to as the interior and exterior surface smoothness. Constraints are added to force the exterior surface to be found outside the interior.[2]

We evaluate the two real applications by comparing to manual annotations provided as images. To do this we convert the output meshes to voxel representations, by computing the fraction of each voxel contained in the corresponding mesh. Results are evaluated taking partial volume into account by using the generalized version of Dice's coefficient $2(A \cdot B)/(A^2 + B^2)$, where A and B are vectorized versions of the images resulting from the automatic segmentation and the manual annotation. We also quantify absolute relative error as $|A - B|_1/|B|_1$.

A. Data

The synthetic data was specifically constructed to investigate accuracy of surfaces found with the help of information from min-marginal energies. It consists of single surface problems, where surfaces are randomly shaped and positioned in 3D images, such that each x-y coordinate of the images is associated with precisely one surface height. That is, the surfaces intersect each voxel column once but not necessarily at integer positions. We want to examine the effect of image resolution and select it in a recursive manner $R_n = 2\ R_{n-1} - 1$, where R_n is the number of voxels along each of the axes at level n and $R_1 = 2$. Ground truth surface position is made spatially dependent at multiple scales, by letting it be the result of Gaussian noise with zero mean and $1/R_n$ standard deviation at each level except level n. The result is n different resolution levels, which are upscaled and summed to produce a $R_n \times R_n$ grid of surface heights. As a final step, the surface heights are linearly transformed to lie in the interval $[0, R_n - 1]$. From the ground truth surfaces (Fig. 5) images of resolution $R_n \times R_n \times R_n$ are produced to function as input data to the methods. Each voxel in these images is set to one if it is below the ground truth surface and zero if it is above. Partial volume is recorded at boundary voxels by setting their value to the fraction that is inside the ground truth surface. Gaussian noise with zero mean and varying standard deviations are added, to model noise in data sources.

The data for the airway wall segmentation problem is identical to that used in [7]. It consists of low dose chest CT (120 kV and 40 mAs) from the Danish lung cancer screening trial [35]. Scans were reconstructed with approximately 0.78 mm $\times$ 0.78 mm $\times$ 1 mm resolution. 649 two-dimensional cross-sectional images perpendicular to and centered on airways from 15 randomly selected scans/subjects were annotated and split into two folds, such that 329 and 319 annotations from 8 and 7 subjects were in each. The annotations were done by marking each pixel of 0.5 mm $\times$ 0.5 mm as being

[2]The code used to mesh the initial surfaces and segment the surfaces can be found at https://bitbucket.org/opfront/mesh and https://bitbucket.org/opfront/opfront

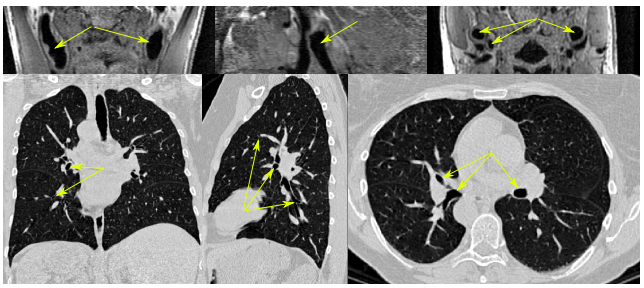

Fig. 4. Example coronal, sagittal, and transverse planes of the carotid artery neck MRI (top) and airway chest CT (bottom) segmented in this work. The arrows point to example carotid arteries and airways, respectively.

airway interior, wall, background, or unknown. Unknown was used to mark noncentral airways. Examples of data and segmentations can be seen in Fig. 9.

The data for the carotid artery wall segmentation problem is identical to that used in [8]. It consists of MRI of the left and right carotid arteries in the neck from the Rotterdam study [36]. The Black-Blood Proton Density Weighted sequences were used and reconstructed with 0.507 mm $\times$ 0.507 mm $\times$ 0.9 mm resolution and cropped to contain 12.5 mm above and below the carotid artery bifurcation point. 57 arteries from 31 subjects were annotated and split subject-wise in three folds. The annotations were done by tracing the interior and exterior carotid artery contours in two-dimensional images reconstructed perpendicular to the artery centrelines at regular intervals. These sets of two-dimensional contours were converted to three-dimensional masks, by fitting an implicit function following the approach of [37].

Fig. 4 shows example images of the two medical imaging datasets.

B. Synthetic Data Experiment

The synthetic data experiment investigates the effect of varying temperature (T in Eq. 4) on surface accuracy at different graph resolutions. For this we randomly generate 20 ground truth surfaces and associated noise corrupted (Gaussian, with a standard deviation of 0.2) input images with a resolution corresponding to $n = 7$, that is, grids of 65×65 surface positions in the range of $[0, 65 - 1]$ and input images of resolution $65 \times 65 \times 65$ with values in $[0, 1]$. The surface position in the input image was then recovered by an optimal surface method using graphs, $\mathcal{G}$ of decreasing resolution, with 65^2, 33^2, 17^2, and 9^2 vertices and columns the size of 65, 33, 17, and 9, respectively. The corresponding minimum cut graphs, $\mathcal{G}_{s,t}$, then have vertices evenly distributed at every 1^3, 2^3, 4^3, and 8^3 input image voxels. The voxel values of the input images at each of the graph vertex positions were directly used as a regional term. To regularize, we add pairwise cost for each pair of neighboring graph vertices. For the particular experiment, $g_{u,v}(|x_u - x_v|) = 0.1|x_u - x_v|$ worked well. Fig. 5 shows example surfaces at the tested resolutions.

The accuracy of the found surfaces was quantified by measuring the average minimum distance to and from the ground truth surface as shown in Fig. 6. Increasing the temperature parameter from the limit at zero leads to more accurate surfaces. This converging behavior is visible in every single example, except a few of the low-resolution graphs, where random effects play a larger role. In other words, the most accurate labels are found not at the discrete optimum ($T \to 0$), nor at a finite value of the temperature parameter, $T \in]0, \infty[$, but rather at the interpolated minimal min-marginal energy $T \to \infty$. This suggests, that if we are not interested in the certainties themselves, we do not need to consider T. Solution accuracy can be improved without introducing extra parameters by computing the min-marginal energies in a band around the discrete solution and using these values to interpolate the minimal min-marginal energy. We will refer to this as the minimal min-marginal surface(s) or label(s).

Fig. 7 shows the recovery of a simple Bell-shaped surface. Data was generated similarly to the random synthetic dataset, but with surface heights determined by a two-dimensional Gaussian function. The result illustrates how the approach reduces discretization errors.

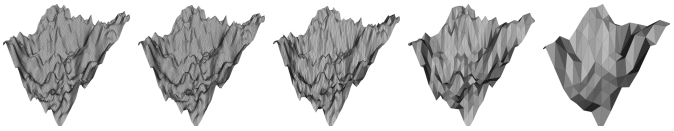

Fig. 5. Synthetic data experiment: left to right, example ground truth surface and found surfaces at graph resolutions of 1, 2, 4, and 8 voxel widths.

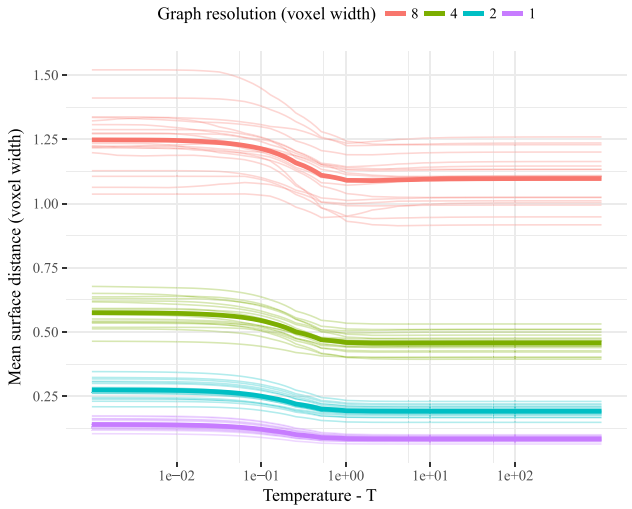

Fig. 6. Synthetic data experiment: effect of graph resolution and the temperature on accuracy. Each thin and lighter color curve is a specific example surface, the thicker curves are the mean curves. Increasing the influence of min-marginal energies on the solution leads to more accurate surfaces. Note: horizontal axis is logarithmic.

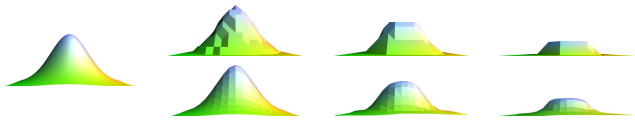

Fig. 7. Recovering a two-dimensional Bell shape (middle-left) using optimal surfaces (top) and minimal min-marginal surfaces (bottom). The pairwise term is increased from left to right with increasingly regularized surfaces as a result. Note the smoother results of the minimal min-marginal surfaces.

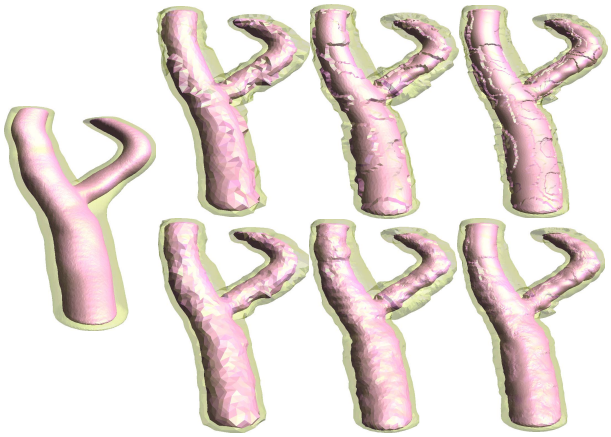

Fig. 8. Example carotid artery surfaces visualized with manual annotation (middle-left), discrete optimal surfaces (top), minimal min-marginal surfaces (bottom) and graph resolution left to right of 1 mm, 0.5 mm, and 0.25 mm. The minimal min-marginal surfaces have fewer discretization errors.

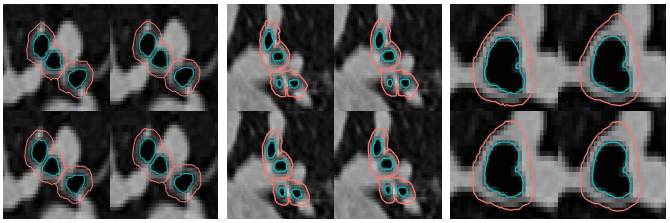

Fig. 9. Example airway wall surfaces overlaid CT with discrete optimal surfaces (column 1, 3, and 5) and minimal min-marginal surfaces (column 2, 4, and 6) at graph resolution 0.5 mm (top) and 0.25 mm (bottom). Differences are subtle but the optimal surface solutions are more irregular particularly at the lower graph resolution than the corresponding minimal min-marginal surface solutions.

C. Medical Data Experiments - Parameter Selection

We experimented with graphs at multiple resolutions for each medical dataset. The density of vertices within the initial mesh and the distance between column positions (number of labels) define the graph resolution for a particular problem. We varied these two settings in unison, by generating meshes with edges whose lengths roughly correspond to the examined column distances. For the airway and carotid artery dataset we examined settings of respectively $0.25, 0.375, \cdots, 0.75$ and $0.25, 0.50, \cdots, 1.50$ mm. Note that the original works [7] and [8] used settings of respectively 0.5 and 0.35 mm. As in the original works, we tune the 5 parameters: interior and exterior surface smoothness, interior, and exterior cost function weights and the variance of the Gaussian kernel used to construct the flow lines. Unlike the previous works, however, we use the Covariance Matrix Adaptation Evolution Strategy [38] for tuning parameters, by maximizing average Dice's coefficient over the area inside the two surfaces on each training fold. The obtained parameters for each cross-validation fold can be found in the supplementary material.

D. Medical Data Experiments - Qualitative and Quantitative Results

Segmentations obtained at the minimal min-marginal energy label, as seen in Fig. 8 and 9, have fewer discretization errors and fit more smoothly to image contours than their discrete optimal surface counterparts. Visually the effect is most apparent with lower resolution graphs, but improvements can also be seen at the highest resolutions.

Fig. 10 shows the average similarity with manual annotations measured with Dice's coefficient for different graph resolutions. Similarity increases as a function of graph resolution and, except for the carotid artery exterior, there is a converging behavior such that gains from increasing the resolution further decrease as the resolution is increased. Minimal min-marginal surfaces consistently outperform discrete optimal surfaces, with better results ($p < 0.01$) at each combination of resolution level and surface. In many cases, gains from using minimal min-marginal surfaces compared to discrete optimal surfaces are comparable to what would be obtained by more than a doubling of the graph resolution, which would otherwise require a factor of 2^3 increase in memory. The results of the two methods become increasingly similar with increasing resolution, which is expected as the resolution of the employed graphs bounds the difference. However, even when the resolution is high and discretization errors are small, computing minimal min-marginal surfaces remains a worthwhile alternative compared to increasing graph resolution further.

Fig. 11 shows the relative reduction in error from using the min-marginal information at each resolution level, defined as $(RE_{\mathrm{D}} - RE_{\mathrm{M}})/RE_{\mathrm{D}}$, where RE_{D} and RE_{M} denotes the average absolute relative error for discrete optimal surfaces and minimal min-marginal surfaces, respectively. Improvements are consistent and there are few examples of scans where results become worse. The largest relative reductions in error happen with the interior surface at graph resolutions around 0.5 mm for airways and 1.25 mm for carotid arteries, where median errors are reduced by 7.2 % and 15.9 %, respectively.

Fig. 12 shows how well the extracted surfaces reproduce the manual annotations as a function of the total time used computing minimum cuts. To remove random oscillations, each computed minimum cut time was based on the average of 5 single threaded runs of each data point and parameter setting. An AMD Opteron(tm) Processor 6380 running at 2500 MHz was used. Computation time was increased by a factor of 1.4 ± 0.3 (range was $0.8 - 2.4$) for airways and 5.6 ± 3.4 (range was $1.7 - 38.1$) for carotid arteries when min-marginal energies are computed in the band around the optimal solution. The extra time taken to compute the min-marginal energies is mostly justified by corresponding improvements in surface accuracy. Slight differences can be seen between the airways and the carotid arteries. In the case of airways, if we look at the discrete optimal surface results at the highest resolution graphs tested, 0.375 mm and 0.25 mm, the results can be simultaneously improved and obtained faster by lowering resolution and exploiting min-marginal information in the proposed manner ($p < 0.01$). That is, the Dice's coefficients obtained with minimal min-marginal surfaces at resolutions of 0.5 and 0.375 mm are 0.001 ± 0.006 and 0.004 ± 0.002 larger than results obtained at resolutions of 0.25 mm with optimal surfaces and they were obtained in

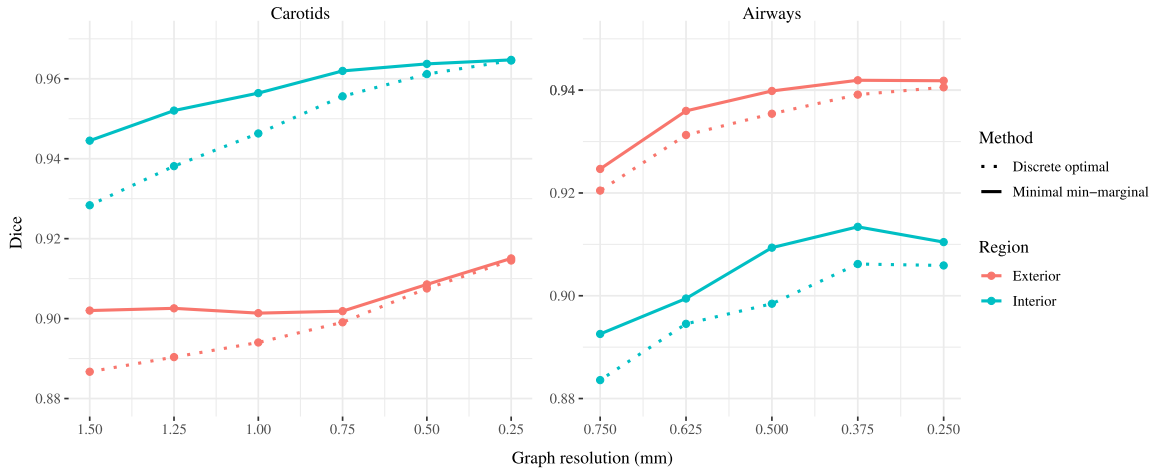

Fig. 10. Similarity of the found surfaces with the manual annotations at the examined graph resolution levels. Results of the interior (blue) and exterior (red) carotid artery (left) and airway (right) surfaces using minimal min-marginal surfaces (full line) and discrete optimal surfaces (dotted line) are shown.

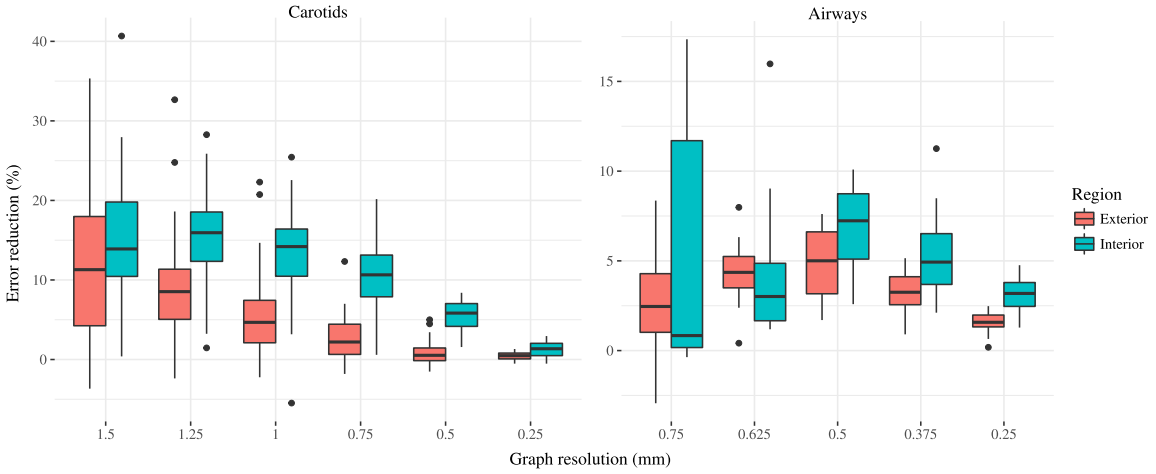

Fig. 11. Relative reduction in error as a result of using minimal min-marginal surfaces compared to optimal surfaces at the examined graph resolution levels. Results of the interior (blue) and exterior (red) carotid artery (left) airway (right) surfaces are shown.

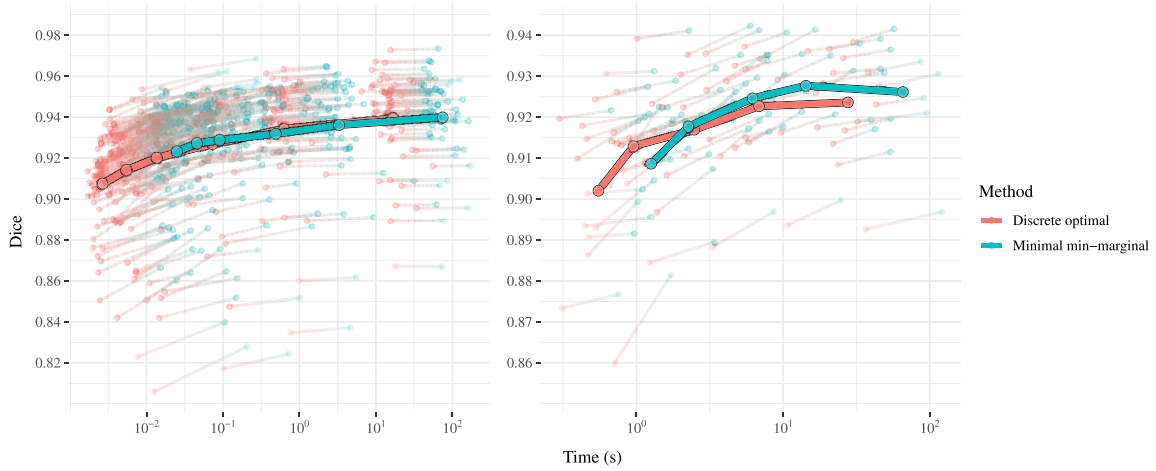

Fig. 12. The average similarity with manual annotations (opaque curves) plotted as a function of running time for the examined graph resolution levels. Results of the carotid artery (left) and airway (right) surfaces using minimal min-marginal surfaces (blue) and discrete optimal surfaces (red) are shown. Each artery/airway data point is plotted (partially transparent) and lines indicate corresponding points across methods.

respectively 26 ± 4 % and 57 ± 19 % of the time it took to compute the optimal surface result. In the case of the carotid arteries, the two methods can produce roughly the same quality results given the same amount of time.

V. Discussion and Conclusion

In this work, we proposed a way to get continuous labels in MRF models with discrete graph cuts. Given a surface finding method and a method to estimate its certainty, points precisely

on the found surface should ideally have the property of being equally certain of being on either side of the surface. This is generally not the case with optimal surface methods and certainty defined as in Eq. 4 as shown in the method section. This lead to the definition (Eq. 7) and investigation of the proposed continuous surface solution which approximates this property. A parameter T was introduced to accurately estimate the certainties, but empirical evidence suggests that the most accurate solutions are found at $T \to \infty$. Interestingly at this limit, Eq. 7 simplifies to Eq. 8, which corresponds to the minimum of a parabola fit to the local min-marginal energies. This minimal min-marginal surface/label solution was consistently seen to improve results compared to existing discrete optimal surfaces methods.

It should be mentioned that continuous label solutions can also be obtained with continuous optimization for similarly formulated problems. For some special cases of $g_{u,v}$ (Eq. 1) this can be done optimally, see for instance [39] and [40], for others, approximate solutions can be derived using methods such as convex relaxation. Solutions to continuous counterparts to a given discrete problem can be mathematically tricky to derive, their implementations are typically slower than discrete graph cuts on comparable hardware but generally easier to parallelize. Computation time is highly application dependent, but factors of 2 to 9 times slower than graph cuts was reported on an example problem in [40]. In comparison, our approach is approximate and can be seen as a simple way to refine the results of existing discrete models. It is not much slower than discrete graph cuts (1.4 ± 0.3 times slower on the airways and 5.6 ± 3.4 times on the carotid arteries) at the same resolution and it works for all convex $g_{u,v}$. The extra computation time used by our approach, compared to discrete graph cuts, is almost exclusively used on computing min-marginal energies, which themselves are useful for other purposes like estimating solution uncertainty.

Fig. 10 and 11 shows the largest improvements from using min-marginal surfaces occur at different graph resolutions for airways (around 0.5 mm) and carotid arteries (around 1.25 mm). These differences are likely an effect of how close it is possible to get to the ground truth solutions, given inaccuracies in these, noise in the data and the particular model used. It is worth noting that, despite that the airway and carotid artery scans were reconstructed at roughly the same resolution, the Black-Blood Proton Density Weighted MRI images used for the carotid arteries do appear less sharp than the low dose CT images used for the airways. Lower effective resolution of the carotid artery images therefore likely explains part of why the biggest gains occur at a lower resolution for the carotid arteries. The single case where Dice's coefficients drop significantly as the resolution is increased ($p = 0.009$) should be noted. This happens for the interior minimal min-marginal surface of the airways (Fig. 10). It is likely a consequence of the specific round of parameter tuning finding a poor local optimum. If adjusted for the number of comparisons done, the p-value does not remain significant at the 0.05 level.

Our work shows that in many cases the minimal min-marginal energy surfaces in lower resolution graphs are better solutions to the examined problems, regarding accuracy, computation time, and particularly memory usage, than corresponding discrete optimal surfaces in higher resolution graphs. However, a disadvantage of using lower resolution graphs is that the underlying data is less densely sampled, and essential intermediary data points can, therefore, be missed. It is possible to use a more strategic placement of the graph vertices, by examining the data term $f_v(x_v)$ locally [41]. The resulting irregularly sampled graph can then be used to find optimal surfaces considering the underlying data more effectively. The technique could be used together with min-marginal information to make the effect of graph resolution on found surface accuracy even smaller.

An advantage of using minimal min-marginal surfaces/labels, compared to other ways of decreasing discretization errors, such as increasing resolution or employing multiresolution and banded schemes [23]–[25], [28], is that it introduces no extra parameters and the underlying graph is preserved. This means that any problem, which was previously addressed with optimal surfaces, can benefit from min-marginal information and solved with increasing accuracy as a result. There is no need to retrain parameters nor to compute min-marginal energies during possibly time-consuming training procedures. Multiresolution schemes may themselves benefit from the extra information in min-marginal energies. [28] already showed that uncertainty could be used in heuristics of where to refine graphs, but we expect that further gains may be possible if the position of the graph vertices in these refinements are chosen under consideration of the interpolated minimal min-marginal energy position.

To compute min-marginals in a constant size band around the optimal solution we need to compute $O(|\mathcal{V}|)$ minimum cuts. In the worst case, each of these minimum cuts takes as long as estimating the optimal configuration in the first place. This could be a source for concern, however, in practice, our results mirror what prior works have also reported, that is, computing min-marginal energies for all variables seems to be possible within a constant factor of the time it took to compute the optimal configuration. We expect that poor performance would only arise in very highly regularized energies, where the pairwise terms $g_{u,v}(|x_u - x_v|)$ are very large. In such cases every single label changing would lead to large parts of the solution being updated. Such energy models could be needed if one was interested in using a very high-resolution graph on very noisy data, however, this is certainly not a common situation, as model complexity is typically chosen taking data quality into account.

Graph resolution and, therewith, optimal surface solution quality is, in practice, often limited by memory requirements. An advantage of using min-marginals in the suggested manner is that improvements come with an inconsequential extra cost in memory and an acceptable increase in running time. Errors were reduced by as much as 15.9% and improvements in segmentation quality, as seen in Fig. 10 are in many cases comparable to what would otherwise be gained by a factor 2^3 increase in graph size and memory usage. Computation of min-marginals should, therefore, be considered as a way to

increase accuracy and/or lower memory usage, in addition to providing useful extra information that can be used to estimate segmentation uncertainty.

REFERENCES

[1] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1333–1336, Oct. 2003.

[2] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum a posteriori estimation for binary images," *J. Roy. Stat. Soc. B (Methodological)*, vol. 51, no. 2, pp. 271–279, 1989.

[3] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[4] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[5] K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images-a graph-theoretic approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 119–134, Jan. 2006.

[6] X. Wu and D. Z. Chen, "Optimal net surface problems with applications," in *Automata, Languages and Programming* (Lecture Notes in Computer Science). vol. 2380. Berlin, Germany: Springer, 2002, p. 775.

[7] J. Petersen *et al.*, "Optimal surface segmentation using flow lines to quantify airway abnormalities in chronic obstructive pulmonary disease," *Med. Image Anal.*, vol. 18, pp. 531–541, Apr. 2014.

[8] A. M. Arias-Lorza *et al.*, "Carotid artery wall segmentation in multispectral MRI by coupled optimal surface graph cuts," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 901–911, Mar. 2015.

[9] M. K. Garvin, M. D. Abramoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, "Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search," *IEEE Trans. Med. Imag.*, vol. 27, no. 10, pp. 1495–1505, Oct. 2008.

[10] Y. Yin, X. Zhang, R. Williams, X. Wu, D. D. Anderson, and M. Sonka, "LOGISMOS—Layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint," *IEEE Trans. Med. Imag.*, vol. 29, no. 12, pp. 2023–2037, Dec. 2010.

[11] Y. Xia, S. S. Chandra, C. Engstrom, M. W. Strudwick, S. Crozier, and J. Fripp, "Automatic hip cartilage segmentation from 3D MR images using arc-weighted graph searching," *Phys. Med. Biol.*, vol. 59, no. 23, pp. 7245–7266, 2014.

[12] Q. Song, X. Wu, Y. Liu, M. Smith, J. Buatti, and M. Sonka, "Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2009, pp. 827–835.

[13] G. Veni, Z. Fu, S. P. Awate, and R. T. Whitaker, "Bayesian segmentation of atrium wall using globally-optimal graph cuts on 3D meshes," in *Information Processing in Medical Imaging*, (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2013, pp. 656–667.

[14] Q. Song *et al.*, "Optimal co-segmentation of tumor in PET-CT images with context information," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1685–1697, Sep. 2013.

[15] J. Petersen, M. Modat, M. J. Cardoso, A. Dirksen, S. Ourselin, and M. de Bruijne, "Quantitative airway analysis in longitudinal studies using groupwise registration and 4D optimal surfaces," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2013, pp. 287–294.

[16] F. Zhao *et al.*, "Congenital aortic disease: 4D magnetic resonance segmentation and quantitative analysis," *Med. Image Anal.*, vol. 13, no. 3, pp. 483–493, 2009.

[17] T. Ajanthan, R. Hartley, and M. Salzmann, "Memory efficient max flow for multi-label submodular MRFs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5867–5876.

[18] O. JamriÅ¡ka, D. Sýkora, and A. Hornung, "Cache-efficient graph cuts on structured grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 3673–3680.

[19] A. Shekhovtsov and V. Hlaváč, "A distributed mincut/maxflow algorithm combining path augmentation and push-relabel," *Int. J. Comput. Vis.*, vol. 104, no. 3, pp. 315–342, 2013.

[20] M. Yu, S. Shen, and Z. Hu, "Dynamic parallel and distributed graph cuts," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5511–5525, Dec. 2016.

[21] A. Delong and Y. Boykov, "A scalable graph-cut algorithm for N-D grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[22] P. Strandmark and F. Kahl, "Parallel and distributed graph cuts by dual decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2085–2092.

[23] V. Lempitsky and Y. Boykov, "Global optimization for shape fitting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[24] H. Lombaert, Y. Sun, L. Grady, and C. Xu, "A multilevel banded graph cuts method for fast image segmentation," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2005, pp. 259–265.

[25] A. K. Sinop and L. Grady, "Accurate banded graph cut segmentation of thin structures using Laplacian pyramids," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2006, pp. 896–903.

[26] P. Kohli and P. H. S. Torr, "Efficiently solving dynamic Markov random fields using graph cuts," in *Proc. ICCV*, vol. 2, Oct. 2005, pp. 922–929.

[27] P. Kohli and Philip H. S. Torr, "Measuring uncertainty in graph cut solutions," *Comput. Vis. Image Understand.*, vol. 112, no. 1, pp. 30–38, 2008.

[28] P. Kohli, V. Lempitsky, and C. Rother, "Uncertainty driven multi-scale optimization," in *Pattern Recognition* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2010, pp. 242–251.

[29] S. Parisot, W. Wells, S. Chemouny, H. Duffau, and N. Paragios, "Uncertainty-driven efficiently-sampled sparse graphical models for concurrent tumor segmentation and atlas registration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 641–648.

[30] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab, "Optical flow estimation with uncertainties through dynamic MRFs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[31] M. Haeker, X. Wu, M. Abràmoff, R. Kardon, and M. Sonka, "Incorporation of regional information in optimal 3-D graph search with application for intraretinal layer segmentation of optical coherence tomography images," in *Information Processing in Medical Imaging*, (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2007, pp. 607–618.

[32] K. Li, S. Millington, X. Wu, D. Z. Chen, and M. Sonka, "Simultaneous segmentation of multiple closed surfaces using optimal graph searching," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer, 2005, pp. 406–417.

[33] D. Tarlow and R. P. Adams, "Revisiting uncertainty in graph cut solutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2440–2447.

[34] J.-D. Boissonnat and S. Oudot, "Provably good sampling and meshing of surfaces," *Graph. Models*, vol. 67, no. 5, pp. 405–451, 2005.

[35] J. H. Pedersen *et al.*, "The danish randomized lung cancer CT screening trial—Overall design and results of the prevalence round," *J. Thoracic Oncol.*, vol. 4, no. 5, pp. 608–614, May 2009.

[36] Q. J. van den Bouwhuijsen, M. W. Vernooij, A. Hofman, G. P. Krestin, A. van der Lugt, and J. C. Witteman, "Determinants of magnetic resonance imaging detected carotid plaque components: The rotterdam study," *Eur. Heart J.*, vol. 33, no. 2, pp. 221–229, 2011.

[37] F. Heckel, O. Konrad, H. K. Hahn, and H.-O. Peitgen, "Interactive 3D medical image segmentation with energy-minimizing implicit functions," *Comput. Graph.*, vol. 35, no. 2, pp. 275–287, 2011.

[38] N. Hansen and A. Ostermeier, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," in *Proc. IEEE Int. Conf. Evol. Comput.*, May 1996, pp. 312–317.

[39] E. Bae, J. Yuan, X.-C. Tai, and Y. Boykov, "A fast continuous max-flow approach to non-convex multi-labeling problems," in *Efficient Algorithms for Global Optimization Methods in Computer Vision* (Lecture Notes in Computer Science). vol. 8293. Berlin, Germany: Springer, 2014, pp. 134–154.

[40] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, "A convex formulation of continuous multi-label problems," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 792–805.

[41] M. D. Abràmoff, X. Wu, K. Lee, and L. Tang, "Subvoxel accurate graph search using non-Euclidean graph space," *PLoS ONE*, vol. 9, no. 10, pp. e107763, Oct. 2014.