

Segmentation of Dynamic N-D Data Sets via Graph Cuts using Markov Models

Yuri Boykov¹, Vivian S. Lee², Henry Rusinek², and Ravi Bansal¹

¹ Siemens Corporate Research, Imaging & Visualization, Princeton NJ 08540, USA
yuri@csd.uwo.ca, ravi@scr.siemens.com

² NYU School of Medicine, Radiology, 550 First Avenue, New York NY 10016, USA
lee@mri.med.nyu.edu, hr18@homemail.nyu.edu

Abstract. This paper describes a new segmentation technique for multi-dimensional dynamic data. One example of such data is a perfusion sequence where a number of 3D MRI volumes shows the dynamics of a contrast agent inside the kidney or heart at end-diastole. We assume that the volumes are registered. If not, we register consecutive volumes via mutual information maximization. The sequence of n registered volumes is regarded as a single volume where each voxel holds an n -dimensional vector of intensities, or *intensity curve*. Our approach is to segment this volume directly based on voxels intensity curves using a generalization of the graph cut techniques in [7, 2]. These techniques use a spatial Markov model to describe correlations between voxels. Our contribution is in introducing a temporal Markov model to describe the desired dynamic properties of segments. Graph cuts obtain a globally optimal segmentation with the best balance between boundary and regional properties among all segmentations satisfying user placed hard constraints. Flexibility, coherent theoretical formulation, and the possibility of a globally optimal solution are attractive features of our method that gracefully handles even low quality data. We demonstrate results for 3D kidney and 2D heart perfusion sequences.

1 Introduction

Dynamic data show how the intensity of different tissues changes with time. Typical examples of dynamic data are perfusion sequences of heart, brain, kidney, etc. The observed voxel intensity changes as a contrast agent propagates through an organ. Such dynamic contrast-enhanced data are commonly used in medicine to analyze blood circulation and to examine proper functioning of organs. Functional MR of brains is another example of dynamic data where intensity of certain groups of neurons change while a patient performs given tasks. Dynamic data reveal geometric structure of an organ in a fashion distinct from most other types of data in computer vision; intensity dynamics vary in different parts of the organ while no visible motion takes place.

In most applications segmentation is a necessary step to analyze dynamic data. Segmentation is a process in which image elements representing the same

tissue class are grouped together and labeled. Such tissue classification is used for monitoring of volume changes (atrophy, tumor growth), 3D rendering and visualization, measurement of tracer concentration etc.

The list of most commonly used segmentation techniques includes thresholding, region growing, split-and-merge, segmentation by clustering, snakes, level set methods, and others. Multi-dimensional dynamic data presents significant challenges for segmentation techniques. Each volume of a dynamic sequence has to be quickly obtained over a short period of time. As a result, the data sets may have very low resolution with strong "partial voluming" effects. Certain perfusion applications [11, 13] use low doses of contrast-enhancing agent which may further reduce the contrast in each volume. Many segmentation techniques can easily fail by "leaking" through a large number of weak object boundaries.

In this paper we present a segmentation technique that addresses the difficulties of dynamic data. Our main idea is as follows. Note that a dynamic sequence of n volumes with grey-level data describes the same geometric structures. Normally, the sequence is either registered or can be registered. If necessary, we obtain such registration via mutual information maximization between pairs of consecutive volumes. The registered dynamic sequence is treated as a single volume where each voxel has an n -dimensional intensity vector, i.e. an *intensity curve*. Then we directly segment this single volume of intensity curves. We approach this potentially treacherous multi-dimensional segmentation problem via graph cut methods [7, 2] that offer a solid theoretical framework based on posterior energy minimization. These methods model correlation between voxels using spatial Markov models.

The graph cut method in [7] allows us to find a globally optimal binary segmentation in case of the Potts energy that combines certain regional and boundary properties of segments. Exact minimization of an appropriate energy solves many of the problems due to low resolution and weak contrast. The technique works on images or volumes. The method in [2] adds a possibility of intuitive user inputs to impose hard constraints for energy minimization. These hard constraints give an efficient way to correct any imperfections in the results. In practice, segmentation imperfections are almost guaranteed in case of low resolution/contrast dynamic data and any completely automatic technique is very likely to fail in such cases.

The graph based energy minimization techniques in [7, 2] use data gradients to describe boundary properties of segments. The desirable regional properties of segments are summarized by intensity distributions or histograms $\Pr(I|L)$ for all regions (labels L) of interest. There are certain difficulties in applying these methods to a volume of intensity curves that we get with dynamic data. In particular, each voxel has n -dimensional intensity vector rather than a simple gray-scale value. The histograms $\Pr(I|L)$ become very difficult to handle numerically when the dimension of I is larger than 4 or 5. Given that a typical perfusion sequence can easily have n greater than 15 or more, we introduced a temporal Markov model for intensity curves. In fact, this model effectively simplifies the

n -dimensional histograms $\Pr(I|L)$ into a product of n 2-dimensional transitional histograms.

Graph based energy minimization methods are quite flexible, easy to implement, work for N -dimensional volumes, and generate globally optimal segmentation that can have arbitrary topological properties. These methods are used in a large number of applications previously reported in computer vision literature, e.g. [12, 9, 1, 16, 14, 15]. In the context of dynamic medical data, [10] demonstrated that the graph based method in [7] can be used to segment functional MRI of brains. They implicitly assume that each intensity curve is a sequence of independent observations. The regional properties of segments are described by mutual information between intensity curves and the time-line of "tasks" given to a patient. In fact, their assumption of independence strips intensity curves of most of its useful "dynamic" content. Our Markov assumption is much more general and allows a wider scope of problems with dynamic data.

The rest of the paper is structured as follows. In Section 2 we provide background information on energy minimization and graph cut techniques that we use. Details of our segmentation method are described in Section 3. In particular, our temporal Markov model for intensity curves is explained in Section 3.2. In Section 4 we show experimental results for 2D heart and 3D kidney perfusion sequences.

2 Background on Potts Energy Minimization

Greig et. al. [7] was first to discover that powerful graph cut algorithms from combinatorial optimization [6, 4] can be used to minimize certain important energy functions in vision. The energies addressed by Greig et. al. and by later graph based methods (e.g. [9, 1, 2]) can be represented as a posterior energy in a standard MAP-MRF¹ framework that assumes a spatial Markov property for a volume labeling. The typical a posteriori energy function is

$$E(L) = \sum_{p \in \mathcal{P}} D_p(L_p) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(L_p, L_q), \quad (1)$$

where $L = \{L_p | p \in \mathcal{P}\}$ is a labeling (segmentation) of volume \mathcal{P} , $D_p(\cdot)$ is a data term, $V_{p,q}$ is an interaction potential, and \mathcal{N} is a set of all pairs of neighboring pixels/voxels. The first and the second terms in (1) represent regional and boundary properties of segments, correspondingly.

Greig et.al. developed a technique based on graph cuts that gives a globally optimal binary segmentation L in case of the Potts model of interaction in (1):

$$V_{p,q}(L_p, L_q) = K_{(p,q)} \cdot T(L_p \neq L_q) \quad (2)$$

where $K_{(p,q)}$ is a discontinuity penalty and $T(\cdot)$ is 1 if condition inside the parenthesis is true and 0 otherwise. This method was generalized in [2] to include

¹ Maximum A Posteriori estimation of a Markov Random Field.

additional hard constraints (seeds) that may be placed by a user. The same seeds can also be used as sample points to summarize regional properties of desirable segments. In [2] such properties are represented by intensity histograms $\Pr(I|L)$ for each possible label value L . The underlying MAP-MRF formulation in [7, 2] suggests the data penalty function based on the likelihood $\Pr(I_p|L_p)$ of intensity I_p at pixel p :

$$D_p(L_p) = -\ln \Pr(I_p|L_p). \quad (3)$$

A fast implementation of graph based energy minimization methods in [7, 2] is possible via min-cut/max flow algorithm discussed in [3].

3 Our Technique for Dynamic Data

Here we provide details of our segmentation technique. In Section 3.1 we describe our registration method in case when original dynamic sequence can not be viewed as "perfectly" registered. After registration we treat dynamic sequence of n volumes as a single volume of intensity curves $I_p = \{I_p^t | 1 \leq t \leq n\}$. We use graph cut techniques in [7, 2] to find a globally optimal segmentation $L = \{L_p | p \in \mathcal{P}\}$ where each voxel label L_p is either "object" or "background". Our energy is given by equations (1), (2), and (3). The data term, D_p in (3), is set by assuming a Markov property for the distributions $\Pr(I_p|L_p)$ of intensity curves. This Markov model is explained in Section 3.2. Our choice of parameters for the Potts interaction penalty, $V_{p,q}$ in (2), is explained in Section 3.3.

3.1 Mutual Information Based Registration

Let $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$ denote the random field from which one of the volumes, called the *float* volume, is sampled. Similarly, let $\mathcal{R} = \{R_1, R_2, \dots, R_m\}$ be the random field from which the other volume, called the *reference* volume, is sampled. Assuming that the random variables $\{F_1, F_2, \dots, F_n\}$ are independently and identically distributed (i.i.d.), let F be the random variable which represents the voxel intensities in the *float* volume. Similarly, assuming that $\{R_1, R_2, \dots, R_m\}$ are i.i.d, let R be the random variable representing the voxel intensities in the *reference* volume. Then the mutual information $I(F, R)$ between the two random variables F and R with marginal probability density functions $P_F(f)$, $P_R(r)$ and joint probability density function $P_{F,R}(f, r)$ is defined to be $I(F, R) = \sum_{f,r} P_{F,R}(f, r) \log_2 \left(\frac{P_{F,R}(f,r)}{P_F(f) \times P_R(r)} \right)$.

Let α denote the set of rigid transformation parameters, three translations and three rotations, by which the *float* volume is being transformed. Then the probability density function of the transformed *float* volume, as a function of α is denoted as $P_{F,\alpha}(f)$. Similarly, the joint density function will be denoted by $P_{F,R,\alpha}(f, r)$. The probability density function, $P_{R,\alpha}(r)$, is a function of α because we are evaluating the mutual information on the overlapping region of the two volumes. The various joint and marginal probability density functions

are estimated by using the Parzen window method [5]. Using these notations, the mutual information, $I_\alpha(F, R)$, is evaluated as a function of α as

$$I_\alpha(F, R) = \sum_{f,r} P_{F,R,\alpha}(f, r) \log_2 \left(\frac{P_{F,R,\alpha}(f, r)}{P_{F,\alpha}(f) \times P_{R,\alpha}(r)} \right). \quad (4)$$

The optimal set of transformation parameters α , which maximize $I_\alpha(F, R)$, are then found using the stochastic gradient descent approach [8, 17].

3.2 Temporal Markov Model for Intensity Curves

If we have dynamic sequence of n volumes then intensity I_p at voxel p becomes an n -dimensional vector $I_p = (I_p^1, \dots, I_p^n)$, that we call an intensity curve. Remember that the general n -dimensional distribution $\Pr(I_p|L_p)$ in (3) cannot be handled in practice for n greater than 4 or 5 and additional assumptions should be made about intensity curves I_p .

Assuming independence of $\{I_p^t\}$ for $1 \leq t \leq n$ reduces $\Pr(I_p|L_p)$ to a product of n one dimensional distributions $\Pr(I_p^t|L_p)$ and may look as an attractive solution. In fact, this can significantly oversimplify dynamic information contained in intensity curves I_p . For example, information on "continuity" between I_p^t for consecutive values of t is completely lost if I_p^t are treated as independent. If an object of interest contains several types of tissue with different intensity dynamics then independence based probability model can assign high probability to a random curve that "jumps" between curves representing these different tissues. In such a case probability distribution is not very meaningful.

In our formulation we make a more general assumption that intensity curves I_p have a Markov property. This property means that for any time instance t the conditional distribution of intensity I_p^t given the whole history of intensities I_p^1, \dots, I_p^{t-1} at the previous time instances, in fact, depends only on the most recent observation I_p^{t-1} , i.e. $\Pr(I_p^t|I_p^1, \dots, I_p^{t-1}) = \Pr(I_p^t|I_p^{t-1})$. This assumption is quite reasonable in the context of dynamic medical data like perfusion or functional MR. It is easy to show that the Markov property implies that the joint distribution $\Pr(I_p|L_p)$ can be written as a product of two-dimensional distributions

$$\Pr(I_p|L_p) = \Pr(I_p^1|L_p) \cdot \prod_{t=2}^n \Pr(I_p^t|I_p^{t-1}, L_p). \quad (5)$$

From computational point of view, two dimensional histograms $\Pr(I_p^t|I_p^{t-1}, L_p)$ are quite manageable. We use equation (5) to describe desirable "dynamical" properties of segments summarized in the data term (3) of energy (1).

3.3 Setting up the Potts Interactions

It remains to explain our choice of parameters for the Potts interaction term (2) of energy (1). We can choose the discontinuity penalty $K_{(p,q)}$ to vary over

different pairs of neighboring voxels. The goal is to encourage segmentation discontinuities where image gradient forces are large and to discourage discontinuities inside regions with low texture. One typical example of such discontinuity penalty is used in a number of graph based techniques in vision

$$K_{(p,q)} = \lambda \cdot \exp\left(-\frac{\|I_p - I_q\|^2}{\sigma^2}\right).$$

The fact that I_p is an n -dimensional vector makes one interpret $\|\cdot\|$ as a vector norm, e.g. Euclidean norm.

In general, we can have $K_{(p,q)} \neq K_{(q,p)}$. In the context of gray-scale volume segmentation, [2] showed that this can help to differentiate between the boundary from "bright" to "dark" and the boundary from "dark" to "bright". If the properties of the object-to-background boundary are known this may help to draw the boundary in the right place. For multi-dimensional intensity curves I_p we can use asymmetrical function

$$K_{(p,q)} = \beta \cdot (C(I_p) - C(I_q)) + \lambda \cdot \exp\left(-\frac{\|I_p - I_q\|^2}{\sigma^2}\right).$$

The scalar function $C(\cdot)$ can represent, for example, a "center of mass" of the corresponding curve, or its average intensity, or any other property of intensity curve that can characterize the desirable properties of the object-to-background segmentation boundary.

4 Experimental Results

We have two choices when estimating the histograms/distributions (5) for object and background intensity dynamics. We can use some historic histograms that were "learned" from intensity curves in the previous segmentation/classification experiments. In this case the initial segmentation can be computed automatically and additional user placed seeds (hard constraints) would be needed only to correct imperfections, if any. Such a version can be created for a specific application where the data does not vary too much.

Here we present results for a more flexible alternative approach that can be used to segment a variety of data sets. A user places seeds even for initial segmentation. These seeds indicate a few voxels that he is certain to be object (red seeds) or background (blue seeds). Similar to [2] we make a dual use of these seeds. First, they work as hard constraints for energy minimization. Second, they give an "expert" sample of intensity curves for object and background. These samples are used to set the histograms in equation (5). Then, an optimal segmentation can be computed. Any imperfections can be still efficiently removed by placing additional correcting seeds (hard constraints). Normally, the results are robust and do not depend on exact positioning of seeds. Below we show some of the results for 2D heart and 3D kidney perfusion data sequences.

In Figure 1(a-d) we show a 2D segmentation example for a single slice of perfusion data of heart at end-diastole. The sequence of 35 images was well

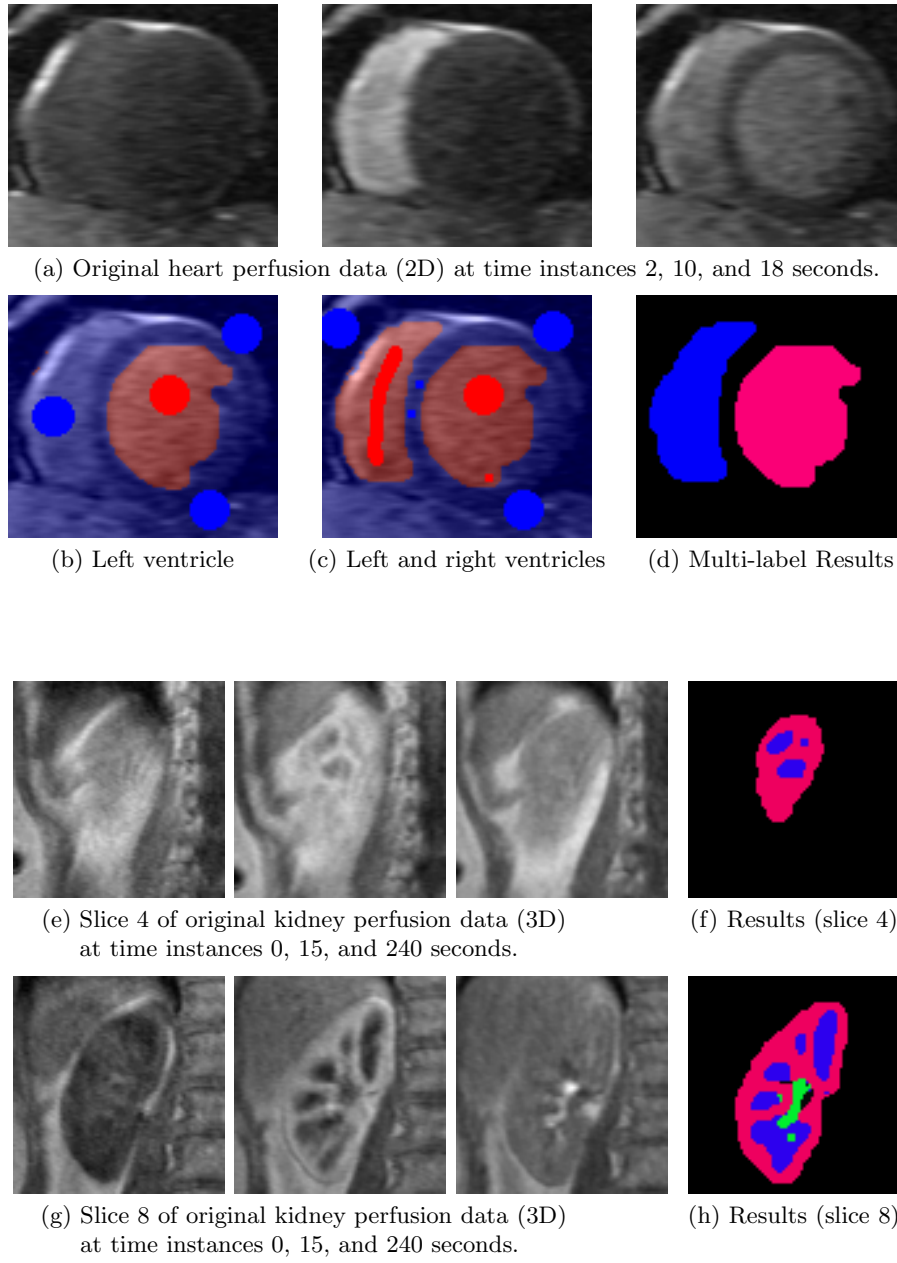


Fig. 1. 2D (heart) and 3D (kidney) examples of dynamic data segmentation.

registered and MI based registration (Section 3.1) was not necessary. In separate experiments we segment the left ventricle alone (b) and both the left and the right ventricles together (c). The segmentation results in (b) and (c) are shown by highlighting the object and background segments with, correspondingly, red and blue colors. Note that multi-label segmentation is possible if the method is used iteratively. An additional binary segmentation can separate the left and the right ventricles starting from the results in (c). The resulting multi-label (red, blue, black) segmentation is shown in (d). All results are obtain in under 10 seconds including time that it takes a user to place seeds (Pentium III, 333MHz).

In Figure 1(e-h) we show an example of kidney perfusion data (a sequence of 18 volumes of size $100 \times 100 \times 17$). The original sequence requires registration. The resolution in each volume is $1 \times 1 \times 3.6$ mm and there are severe defects due to partial voluming. The goal is to segment the cortex, the medulla, and the collecting system of the kidney. Given nontrivial topological properties of the tissues and extremely poor data quality (to date, the best possible) no automatic segmentation seems possible and the use of an interactive technique looks very appropriate.

Registration using MI maximization takes about 3-4 minutes. Then we obtain a multi-label segmentation in three iterations. First, the whole kidney is separated from the background. Then we separate the medulla versus the cortex and the collecting system. The last step is to cut the collecting system from the cortex. Each iteration takes around one minute (50% user input and 50% computation). The seeds were placed in two or three (out of 17) slices in the volume. In (f) and (h) of Figure 1 we show 3D segmentation results. We use "red" for cortex, "blue" for medulla, "green" for collecting system, and "black" for background.

References

1. Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.
2. Yuri Boykov and Marie-Pierre Jolly. *Interactive graph cuts* for optimal boundary & region segmentation of objects in N-D images. In *International Conference on Computer Vision*, July 2001.
3. Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *3rd. Intl. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*. Springer-Verlag, September 2001.
4. William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial Optimization*. John Wiley & Sons, 1998.
5. R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.
6. L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.
7. D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, 1989.

8. S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan College Publishing, 1994.
9. H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 125–131, 1998.
10. Junmo Kim, John W. Fisher III, Andy Tsai, Cindy Wible, Alan S. Willsky, and William M. Wells III. Incorporating spatial priors into an information theoretic approach for fMRI data analysis. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 62–71, 2000.
11. V.S. Lee, H. Rusinek, G. Johnson, N. Rofsky, G.A. Krinsky, and J.C. Weinreb. Ultra-low dose Gadolinium-DTPA MR renography for the diagnosis of renovascular disease. *Radiology*, 2001, in press.
12. Sebastien Roy and Ingemar Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *IEEE Proc. of Int. Conference on Computer Vision*, pages 492–499, 1998.
13. H. Rusinek, V.S. Lee, and G. Johnson. Optimal dose of Gd-DTPA for dynamic MR studies. *Magnetic Resonance in Medicine*, 2001, in press.
14. Dan Snow, Paul Viola, and Ramin Zabih. Exact voxel occupancy with graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 345–352, 2000.
15. B. Thirion, B. Bascle, V. Ramesh, and N. Navab. Fusion of color, shading and boundary information for factory pipe segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 349–356, 2000.
16. Olga Veksler. Image segmentation by nested cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 339–344, 2000.
17. P. Viola and W. M. Wells. Alignment by maximization of mutual information. In *International Conference on Computer Vision*, pages 16–23, 1995.