SPATIO-TEMPORAL SEGMENTATION OF THE HEART IN 4D MRI IMAGES USING GRAPH CUTS WITH MOTION CUES

H. Lombaert, F. Cheriet

École Polytechnique de Montréal, Montreal, Canada

ABSTRACT

With the increasing availability of 4D cardiac imaging technologies, the need for efficient spatio-temporal segmentation algorithms for the heart is growing. We propose a new method for heart segmentation in 4D data sets. We efficiently use the established graph cut method for the segmentation of the heart by simultaneously exploiting motion and region cues. We construct a 4D graph designed to find a moving object with a uniform intensity from a static background. This method has useful applications ranging from qualitative tasks such as direct visualization of the heart by removing its surrounding structures, to quantitative tasks such as measurements and analysis of the total heart volume. The method has been tested on cardiac MRI sequences with successful results.

Index Terms— Image segmentation, Motion pictures, Magnetic resonance imaging

1. INTRODUCTION

Cardiac motion provides additional physiological information vital to the diagnosis and treatment of cardiac dysfunctions. Traditionally 2D imaging such as fluoroscopy and echocardiography were used to sense the cardiac motion. Lately, improvements in MRI and CT imaging technologies allow the visualization of a moving 3D heart. One important step in analyzing the cardiac motion is the segmentation or isolation of the heart throughout the whole image sequence.

Many different approaches have been proposed for 3D segmentation of the heart. Popular methods (see [1] for a review) include variational approaches ([2]), and model based approaches ([3], [4], [5]), a growing interest also exists in graph based approaches ([6], [7]). Efficient graph cut algorithms allow the use of larger volumes ([8]) in faster time ([9]). Most of the works in cardiac segmentation actually focus on the blood pool, that is the inner boundary of the different cardiac structures such as the left or right ventricle. Very few works exist ([10]) in the segmentation of the total heart as a single object including all chambers and myocardial muscles.

Ingenious works ([11], [12], [13]) tackle the segmentation problem in the spatio-temporal domain. Temporal information improves the coherence of the segmentation process throughout the whole image sequence. 4D approaches usually try to recover the motion over time and later, moving structures are segmented based on the results of the motion recovery. In [14] and [15], although computed iteratively, both steps are formulated in the same functional to group image layers of similar velocities. We propose to tackle the problem by simultaneously segmenting the heart and exploiting motion information in a cooperative scheme. The method was designed to be fast and robust for a majority of cases, while allowing automatic processing as well as easy fine tuning of the results.

2. METHOD

Our method is based on the graph cut algorithm ([6], [7]). We incorporate motion cues in the algorithm in order to obtain a smooth segmentation over time. The user can additionally edit zones to be included or excluded in the final segmentation. First, we explain the graph cut algorithm. Secondly, we show how to use motion cues extracted from the temporal domain into the algorithm. Lastly, we give details on how to exploit user inputs.

2.1. Graph Cuts

In the graph cut algorithm, a graph $\mathcal{G}$ (Fig. 1) is constructed from an image where each node in the graph corresponds with a pixel $p \in I$ in the image. All nodes are connected to two extra special nodes, a source node specifying the *object* terminal, and a sink node specifying the *background* terminal.

A cut $\mathcal{C}$ separating both terminals isolates the nodes in two sets: those connected to *object*, and those connected to *background*. The sum of the weights of all severed edges $(p, q) \in \mathcal{C}$ gives its cost $|\mathcal{C}|$. Finding the minimum cost $|\hat{\mathcal{C}}|$ yields the optimal segmentation $\hat{f}$ of an object from its background. It has been proven ([16]) that such problem is equivalent to finding the maximum flow of a graph, and polynomial time complexity algorithms exist to find a global optimal solution.

The energy to be minimized $E(f)$ (or $|\mathcal{C}|$) contains two terms: the data term $E_{data}(f)$ which tells how well the solution f fits with the observed data, and the smoothness term $E_{smooth}(f)$ weighted by a parameter λ which imposes a smoothness constraint on the solution f,

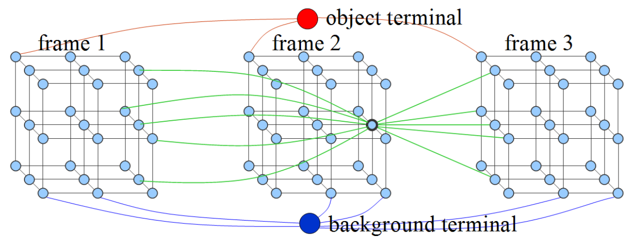


Fig. 1. 4D graph corresponding to 3 frames of a 3^3 volume. Spatial links are all shown, and only temporal links of the highlighted node are shown.

$$E(f) = E_{data}(f) + \lambda E_{smooth}(f), \text{where}$$

$$\begin{aligned} E_{data}(f) &= \sum_{p \in I} D_p(f_p) \\ E_{smooth}(f) &= \sum_{\substack{p,q \in \mathcal{N} \\ f_p \neq f_q}} V_{p,q}(f_p, f_q) \end{aligned}$$

The data term will be discussed in the next sections and will contain motion cues as well as user input information. The smoothness term ensures that our segmentation f remains smooth. An edge between two neighboring pixel nodes $p, q \in \mathcal{N}$ having intensities $I(p)$ and $I(q)$ will have a weight

$$V_{p,q} = \exp\left(-\frac{(I(p) - I(q))^2}{2\sigma^2}\right),$$

where the parameter σ controls the weight of the edge. The smoothness term will thus favor segmented regions with piecewise homogeneous intensities.

2.2. Motion Cues

In our method, we are actually looking for a moving object surrounded by a static background. The terminal nodes should therefore be named *moving object* and *static background*. Each pixel in the graph will be classified as either being part of a moving object or part of a stationary background.

As in the optical flow method, a moving pixel is assumed to have a varying intensity over time, and a stationary pixel will have a constant intensity. The likelihood $D_{p_i}(obj)$ of pixel at time i for being part of an object will thus be good if its pixel intensity variation between two frames i and $i+1$ differs more than a value θ_{motion}. Similarly, the likelihood $D_{p_i}(bkg)$ for being part of the background will be low if the intensity variation is below θ_{static}.

These will be the weights of the edges connecting all nodes to both terminals, *moving object* and *static background*:

$$D_{p_i}(obj) = \mu_{obj}(p_i) \cdot \exp\left((I(p_i) - I(p_{i+1}))^2 - \theta^2_{motion}\right),$$
$$D_{p_i}(bkg) = \mu_{bkg}(p_i) \cdot \exp\left(\theta^2_{static} - (I(p_i) - I(p_{i+1}))^2\right)$$

Note that our graph is designed to separate moving pixels from stationary pixels and favor regions with uniform intensities. This is why $V_{p,q}$ only smooths intensity, motion cues are given as *a priori* information. Both parameters $\mu_{obj}(p_i)$ and $\mu_{bkg}(p_i)$ are factors affected by user inputs as explained in the next section.

2.3. User Initialization

Motion appears in many regions. The heart is our target structure, but the chest also moves due to respiration, the lungs being a very soft tissue presents motion, and the diaphragm causes the respiration motion. If no additional information is provided, the heart cannot be located, and large moving regions having similar intensities will be grouped during segmentation.

Typical graph cuts based methods for segmentation require the user to mark a few pixels as object and a few others as background. Graph edges from these pixel nodes to either terminal nodes will have infinite weights while all other pixel nodes will have no links to the terminals.

We use a similar idea and we gradually affect the terminal edge weight of a pixel with the geodesic distance of that pixel to the set of marked pixels. The closer the pixel is to the set of marked pixels and the more similar the pixel intensity is to that set, the shorter its distance is. The geodesic distance between two points is the accumulation of the pixel intensities along the shortest path between these two points. In our implementation we used a region growing with a priority queue.

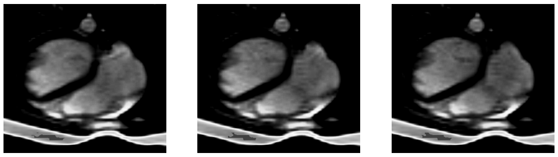
Our previous factors $\mu_{obj}(p)$ and $\mu_{bkg}(p)$ tell us how influent our motion cues should be at pixel p. For instance, if a pixel is far away (in a geodesic sense) from the set of marked object pixels, our terminal edge weight $D_p(obj)$ should be low, similarly, if a pixel is close to the set, it is more likely to be part of the object. We use:

$$\mu_f(p) = \exp\left(-\frac{d_f(p)^2}{2\delta^2}\right),$$

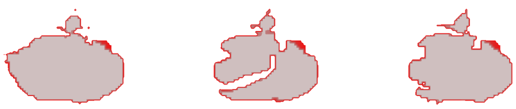
where $d_{obj}(p)$ or $d_{bkg}(p)$ is the geodesic distance from point p to the set of marked object or background pixels, and δ defines a radius of influence around each set of marked pixels. If no pixels are initially marked as object nor background, factors $\mu_f(p)$ could be set to 1.

3. RESULTS

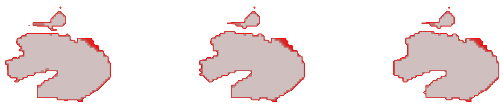
The method has been tested on 5 MRI data sets. The first two data sets have 11 and 14 frames of $88 \times 128 \times 16$ volumes presenting two human hearts. The next two data sets have



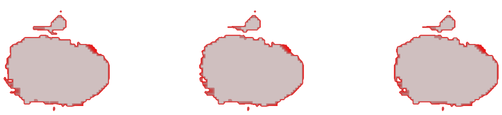
(a) Frames a time t_0, t_1, and t_2



(b) 3D Segmentations



(c) 4D Segmentations



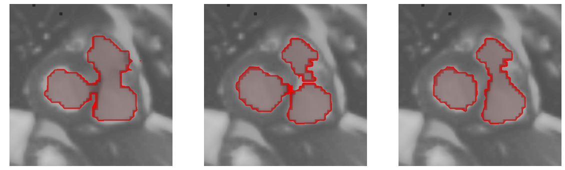
(d) 4D Segmentations using motion

Fig. 2. (a) Slices of 3 consecutive frames, and segmentation results using (b) independent 3D graph cuts, the boundary jumps from the outer to the inner myocardium wall, (c) 4D graph cuts with no motion information, the myocardium is not segmented, and (d) with motion information, the moving myocardium is segmented.

both 25 frames of $256 \times 256 \times 18$ presenting the same heart from a long axis view and from a short axis view. The last data set has 16 frames of $120 \times 128 \times 16$ volumes presenting only the apex of the heart.

We manually marked both ventricles with a single stroke of paint, and added a second stroke of paint marking surrounding areas such as the lungs or the organs below the heart. We compared our results with 3 different algorithms, all using the same parameters, user inputs, and graph connectivity.

The first method involves independent segmentations of each frame using separate 3D graph cuts. We want to show that segmentation using sole spatial information does not guarantee temporal smoothness as shown in figure 2(b). In these 3 consecutive frames of our first test volume (Fig. 2(a)), the myocardium muscle is hardly distinguishable due to the imaging conditions. That causes the segmentation boundary to jump from the myocardium outer and inner walls between two consecutive frames. The second method involves a spatio-temporal segmentation of all frames using a single 4D graph cuts. As traditionally used, terminal edges to the marked object and background pixels are set with infinite weight. The results in figure 2(c) show a more coherent segmentation over time. The segmentation however fails to



(a) 3D (b) 4D (c) 4D+motion

Fig. 4. Slice of data set 4 with segmentation using (a) independant 3D graph cuts, (b) 4D graph cuts with no motion information, and (c) with motion information.

include the moving surrounding myocardium. Our method appearing on the last row of figure 2(d) detects the moving myocardium while being coherent over time.

In all our cases, 3D graph cuts yield non coherent segmentations over time. 4D approaches overcome this problem, and our method using motion cues correctly adds additional moving cardiac structures to the segmentation. The figure 3 shows all our 5 cases with the used markers.

In case 4 (Fig. 4), our method failed to include the myocardium. When looking in 3D, the myocardium pixel intensities are indeed similar to the surrounding organs pixel intensities below the heart. Although it could be recovered with an additional marking, we can observe that by looking at the details on figure 4, our method still includes additional moving surrounding tissues. As no ground truth is available, and as it is hard to perform a precise human expert segmentation coherent over time, it is difficult for us to state that our method (Fig. 4(c)) clearly outperforms a 4D segmentation with sole spatial information (Fig. 4(b)). In this case, a visual appreciation tends to show that using motion cues leads to a better segmentation.

Running separate 3D graph cuts is faster (from 3.38 sec to 15.98 sec on a Core2 2.4GHz CPU) than running 4D graph cuts (from 9.09 sec to 106.94 sec). In our implementation, both methods using 4D graph cuts with and without motion cues contain the same amount of nodes and links, so both methods have similar running times. Our method also has a small extra cost (from 1.73 sec to 6.73 sec) for computing geodesic distances from the set of marked pixels.

4. DISCUSSIONS

In our method, we achieved utilizing motion cues directly in the graph used for segmentation. Recovering a smooth motion information in a separate step has been shown to be difficult using the graph cuts method, layers of similar velocities are rather recovered ([14]). But in this paper we try to show that motion cues can still be successfully exploited in a joint problem, where here segmentation and motion have to be processed at the same time. Our graph has indeed been designed to find a moving object with a uniform intensity in a single step.

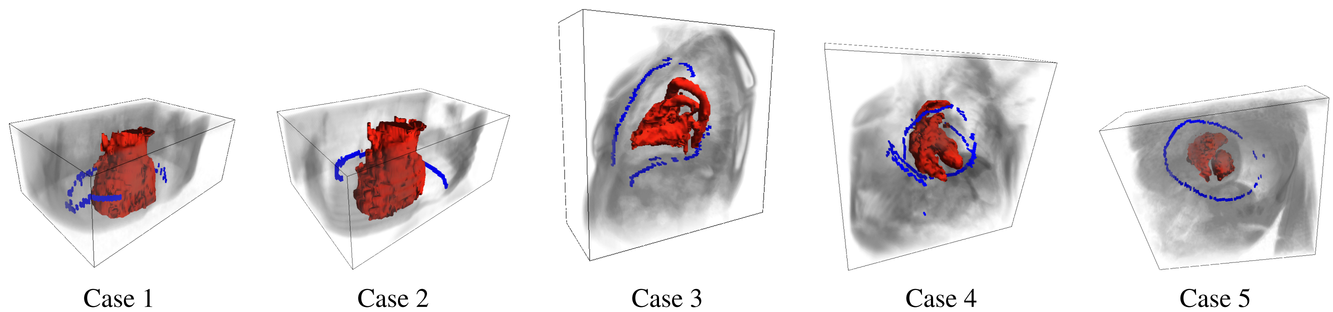


Fig. 3. Heart segmentations in red of our 5 data sets, overlaid by the user marked voxels.

Future work include experiments on CT data where image noise and blur bring challenging conditions, automating the algorithm in a similar manner to [10] (i.e., finding the heart with a growing ellipsoid). When advances in modern imaging technology will permit, real time applications would be possible with little modifications (i.e. adding new volume frames as new nodes to the graph, removing old frames as time goes on). Shape and motion in the previous frames indeed matter, and the segmentation would benefit from our 4D method using motion cues. It will be smooth and coherent over time.

4.1. Aknowledgments

We wish to thank Siemens Corporate Research, Princeton, NJ, for support and for helping us with the data sets and the National Science and Environment Research Council (NSERC) for funding.

References

[1] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomed. Eng.*, vol. 2, no. 1, pp. 315–337, 2000.

[2] N. Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *IEEE Trans. Med. Imaging*, vol. 22, no. 6, pp. 773–776, 2003.

[3] A. F. Frangi, W. J. Niessen, and M. A. Viergever, "Three-dimensional modeling for functional analysis of cardiac images: a review.," *IEEE Trans. Med. Imaging*, vol. 20, no. 1, pp. 2–25, Jan. 2001.

[4] J. Montagnat, H. Delingette, and N. Ayache, "A review of deformable surfaces: topology, geometry and deformation," *Image and Vision Computing*, vol. 19, no. 14, pp. 1023–1040, Dec. 2001.

[5] J. Montagnat and H. Delingette, "4d deformable models with temporal constraints: application to 4d cardiac image segmentation," *Med. Image Analysis*, vol. 9, no. 1, pp. 87–100, Feb. 2005.

[6] Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in *IEEE Int. Conf. Computer Vision*, 2001, vol. 1, pp. 105–112.

[7] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Computer Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[8] H. Lombaert, Y. Sun, L. Grady, and C. Xu, "A multilevel banded graph cuts method for fast image segmentation," in *IEEE Int. Conf. Computer Vision*, 2005, pp. 259–265.

[9] A. Delong and Y. Boykov, "A scalable graph-cut algorithm for N-D grids," in *IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[10] G. Funka-Lea, Y. Boykov, C. Florin, M. P. Jolly, R. Moreau-Gobard, R. Ramaraj, and D. Rinck, "Automatic heart isolation for ct coronary visualization using graph-cuts," in *IEEE Int. Symp. Biomed. Imaging*, 2006, pp. 614–617.

[11] M. Waschbüsch, S. Würmlin, and M. Gross, "Interactive 3d video editing," *Vision Computing*, vol. 22, no. 9, pp. 631–641, Sep. 2006.

[12] X. Bai, J. Wang, D. Simons, and G. Sapiro, "Video snapcut: robust video object cutout using localized classifiers," in *ACM SIGGRAPH*, 2009, pp. 1–11.

[13] T. Vaudrey, D. Gruber, A. Wedel, and J. Klappstein, "Space-time multi-resolution banded graph-cut for fast segmentation," in *DAGM Symp. Pattern Recognition*, 2008, pp. 203–213.

[14] T. Schoenemann and D. Cremers, "Near real-time motion segmentation using graph cuts," in *Annual Symp. of the German Assoc. for Pattern Recognition*, 2006, vol. 4174, pp. 455–464.

[15] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *IEEE Int. Conf. Computer Vision*, 1998, pp. 1154–1160.

[16] D. Greig, B. Porteous, and A. Seheult, "Exact maximum a posteriori estimation for binary images," *Royal J. on Stat. Society*, vol. 51, no. 2, pp. 271–279, 1989.