Automatic Segmentation of Cardiac MRI

David T. Gering

GE Medical Systems, Applied Science Lab
P.O. Box 414, W875, Milwaukee, WI 53201
david.gering@med.ge.com

Abstract. A new method is proposed for automatic segmentation of cardiac MRI. The novelty is that we analyze motion to extract the information that other algorithms rely on users to input manually, such as an initial contour or seed points. The motion map, once computed, restricts subsequent processing to be constrained within a region of interest that envelops the heart. Focusing the computation in this manner resolves some of the ambiguity that causes the image segmentation problem to be an ill-posed one. The segmentation is then performed using a new contextual dependency network (CDN) that incorporates context via hierarchical processing. Within this framework, voxels are first classified as independent events with an EM approach to simultaneously correct for field inhomogeneity. The algorithm next considers neighborhood interactions using a Markov random field, followed by region-level properties, and finally, relationships between regions.

1 Introduction

Automated extraction of the left ventricular endocardium, epicardium, and myocardium for routine clinical image analysis is a difficult problem. This is mainly because the algorithm needs to be applicable for a wide variety of image situations where variability is caused due to patient health, patient movement, pathology, imaging noise and artifacts. Most automated algorithms available in the literature and in products fail to yield satisfactory results. These methods require extensive human intervention to extract clinically meaningful output. Therefore, there is a need for a robust, fast algorithm that requires minimal, if any, user interaction. Furthermore, an accurate endocardium and epicardium segmentation can facilitate determination of various parameters necessary to determine the cardiac health for diagnosis and treatment planning.

The field has a long and rich history. Of particular interest with cardiac imaging – in contrast to brain imaging, for example – is the availability of temporal information. Typical MRI scanning protocols acquire approximately 10 short-axis slices at approximately 20 phases of the cardiac cycle. Regardless, many segmentation algorithms in the literature discard the valuable asset of temporal information to focus

R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2878, pp. 524–532, 2003.

exclusively on spatial content. Obvious exceptions are those applications where tracking, rather than just segmentation, is the object. In such instances, temporal information is customarily employed to assist in propagating an initial contour to images from other phases of the cardiac cycle. Toward this end, Kalman filters [12] and optical flow [7] provide a means for estimating the motion between frames so that the result of the current frame can predict the initialization for the next frame [10]. Other applications desire to analyze motion based on the output of segmentation [3]. That is, segmentation is performed prior to motion analysis rather than visa versa, as proposed in this paper. However, when the application is computing parameters of cardiac health such as measurements of mass and ejection fraction, temporal information is typically ignored or merely exploited in the regularization process to assist in smoothing.

The novelty in the present work is that we analyze motion to extract the information that other algorithms customarily rely on users to input manually. We can therefore achieve an automatic method by not requiring the user to enter seed points or draw initial contours.

2 Method

2.1 Temporal Analysis

Motion is difficult to characterize in an efficient manner, so we exploit the simple fact that the motion of the left ventricular endocardium and the right ventricular wall creates extremely high variance measurements on MRI (FIESTA scanning protocol, GE Medical Systems, Milwaukee, WI). The stark contrast on FIESTA images between bright blood and dark muscle produce variance measurements that are an order of magnitude larger than those of the remaining – and generally motionless – portion of the image.

Temporal analysis is performed for each slice by processing all frames at a given slice location throughout the cardiac cycle. We compute the relative deviation of each voxel across the ensemble of cardiac phases (time dimension) as the variance normalized by the mean. A relative deviation image, once computed, can be segmented straightforwardly with basic threshold and morphological operations to extract a region of interest (ROI) that contains the heart exclusively. We segment our motion map by applying a threshold just above the noise floor followed by these morphological operations: island-removal within the foreground objects, erosion, island-removal within the background, and dilation.

Following this slice-by-slice processing, an ROI is extracted as the largest 3-D structure. Subsequent processing is constrained to occur within this ROI, thereby resolving some of the ambiguity that causes the image segmentation problem to be an ill-posed one.

2.2 Spatial Analysis Using a Contextual Dependency Network

Ambiguity necessitates the incorporation of contextual information into the segmentation process. Toward this end, we organized various levels of context into the hierarchical network of Table 1, and labeled it as a Contextual Dependency Network (CDN). The lowest level of context first classifies each voxel as if it were an independent event, and then higher-level layers – through their broader understanding of context –correct the classifications of the lower level. Just as a voxel-wise classification must be computed prior to a neighborhood-wise refinement, a voxel's region must be classified before features regarding the size and shape (or other intrinsic properties) of that region can be computed. This is a concept of predicated context, where high-level vision is performed based on aggregated information from low-level vision.

This general concept is no longer new to the field of computer vision, as the combination of Bayesian classification, Markov random fields, and deformable templates is evident in cardiac image segmentation [9]. Granlund [6] also developed hierarchical methods for computer vision. Our goal is to provide a framework for expressing and conceptualizing the predication. Our result is a mathematical formalism for designing efficient algorithms with fewer *ad hoc* heuristics than are often found in the cardiac segmentation literature [1] [11]. By considering each layer, we have suggested the computation listed in Table 1 as a very simple implementation of the CDN framework for this paper. Future work can substitute more sophisticated components.

Table 1. Contextual Dependency Network (CDN) for hierarchical processing

#	Layer	Example	Present Computation
5	User *(oracle)*	Corrective drawing	Manual corrections at conclusion
4	Inter-structure *(global)*	Relative position of a voxel's structure to other structures	Myocardium surrounds LV, but not RV
3	Intra-structure *(region)*	Relative position of a voxel within its own structure	RV and LV are first and second largest blood pools
2	Neighborhood *(local)*	Classification of a voxel's immediate neighbors	MRF with mean-field approximation
1	Voxel *(point)*	Classification based on voxel's intensity	Unsupervised Bayesian classification within an EM framework

Layers can communicate with each other bi-directionally, and this communication can be implemented multiple ways. We use the simplest here, which is to perform the first 2 layers of processing with a stationary prior, and use the 3rd and 4th layers to compute a spatially-varying prior. Then the communication of information from the upper layers back down to the lower layers can be achieved by re-executing the lower layers using the spatially-varying prior. More sophisticated inter-layer communication is possible using multi-level Markov random fields or Bayesian belief propaga-

tion networks. A more detailed discussion of CDN is available in [5] which develops the initial proposal in [4] for the application of segmenting brain tumors.

2.2.1 CDN Layer #1: EM-Based Classifier

We perform Bayesian classification using Gaussian tissue class models for 3 classes: blood, myocardium, and background. EM segmentation [2, 14] models the log-transformed image intensities as visible variables, *y*, tissue classifications as hidden variables, *w*, and the bias field as governed by model parameters, *b*. We would like to choose the parameters that maximize the log likelihood of the data, $\log p(y, w \mid b)$, but we do not know this likelihood because *w*'s invisibility renders $p(y, w \mid b)$ to be a random variable. Thus, although we cannot maximize it, we can *maximize* its *expectation*. This results in the following two iterative steps until convergence to a local minimum:

E-Step:
Compute the expectation $\sum_{w} p(w \mid y, b) \log p(w, y \mid b)$ using the current *b*.

M-Step:
Find new b^{t+1} to maximize the expectation, assuming $p(w \mid y, b^{t})$ is correct.

To make the algorithm unsupervised, we implemented "generalized EM" to estimate the model parameters (means of Gaussian distributions in our case) during the M-Step as follows [15]. Let *x* represent the bias-corrected intensities, and $\mathbf{\Phi}$ the Gaussian intensity model $\mathbf{\Phi} = \{(\mu_l, \sigma_l^2) \mid l \in L\}$, where *l* is a given label from the set of all possible segmentation labels, *L*. Then:

$$\mu_l = \frac{\sum_{i \in S} x_i P(w_i = l \mid x_i, \mathbf{\Phi})}{\sum_{i \in S} P(w_i = l \mid x_i, \mathbf{\Phi})} \tag{1}$$

2.2.2 CDN Layer #2: Mean-Field Markov Random Field

We then added CDN Layer 2 to effectively relax Layer 1's E-Step weights. The prior knowledge of spatial coherence over a configuration, *w*, of segmented voxels is modeled with a Gibbs distribution, *P(w),* which takes the following form:

$$P(w) = \frac{\exp\left(-U(w)\right)}{\sum_{w' \in W} \exp\left(-U(w')\right)} \tag{2}$$

This distribution's energy function, *U(w)* is an Ising model generalized to the case of discrete, multi-valued labels. The energy function is composed of clique potentials for singleton cliques V_1 and pair-wise cliques V_2:

$$U(w) = \sum_{i \in S_1} V_1(w_i) + \sum_{i \in S} \sum_{j \in N_i} V_2(w_i, w_j) \tag{3}$$

In other words, V_1 encodes our prior knowledge about an isolated voxel prior to viewing the image data. This prior knowledge is the tissue class prior probability, which is stationary during the first CDN iteration, and spatially-varying thereafter. V_2 is the potential over all cliques of size 2, and represents the tendency of two classified voxels to be neighbors. That tendency is encoded in the MxM Class Interaction Matrix, **J**, and it is computed from a segmented scan offered as training data.

To make the computation tractable, we used the mean field approximation to factorize the joint probability into a product of local conditional probabilities:

$$P(w) \approx \prod_{i \in S} P(w_i \mid \overline{w}_{N_i}) \tag{4}$$

Then, computation is straightforward using the local clique potentials. Given M possible label values, let $\mathbf{w}_i$ be an M-length binary vector of classification at the voxel indexed by *i*, and let $\overline{\mathbf{w}}_i$ be the vector of mean values (probabilities). Then:

$$\begin{aligned} V_1(\mathbf{w}_i) &= -P(\mathbf{w}_i) \\ V_2(\mathbf{w}_i, \overline{\mathbf{w}}_j) &= -\mathbf{w}_i^T \mathbf{J} \overline{\mathbf{w}}_j \end{aligned} \tag{5}$$

2.2.3 CDN Layer #3: Intra-structure Properties

The segmentation produced by the first two layers provides an initial estimate of the scene from which various properties can be computed on each region. In our initial simple implementation described in Table 1, we compute the size of blood pools with the *a priori* knowledge that the right ventricle (RV) and left ventricle (LV) correspond to the first and second largest 3-D pools within the ROI. (Recall that the ROI is computed from motion analysis prior to application of the CDN.) This enables us to create a spatially-varying prior for the next iteration of the first 2 CDN layers that can distinguish between RV and LV as well as exclude any fat that may be present within the confines of the ROI.

Recall that the classification of the first two layers involved 3 tissue classes: blood, myocardium, and background. Following the first application of CDN layer #3, the spatially-varying prior that it produces as output allows us to divide the blood class

into the two distinct classes of LV and RV. This prepares the way for CDN layer #4 to consider spatial relationships relative to these two anatomical structures.

2.2.4 CDN Layer #4: Inter-structure Properties

One approach to encoding contextual information is to discover features using unsupervised mathematics such as PCA. Alternatively, in a less general solution, we designers can explicitly designate the features of interest. Our goal in this section is to encode the computer with the knowledge that the myocardium envelops the LV, and that it separates the LV and RV. We express this with a form of topological atlas implemented by fitting a circular disc to each slice in order to maximize an appropriate energy function. The 3 parameters of the disc (radius and x,y coordinates of the focal point) are determined efficiently through making two iterations of searching the space of each parameter independently. Let "BG" denote the voxels classified by the first 3 CDN layers as belonging to the background, and "MYO" the myocardium. Define our energy function in terms of the number of voxels of each type under the disc:

$$\text{Energy} = (\#\,\text{MYO}) - 2*(\#\,\text{RV}) - 2*(\#\text{BG})$$

Despite the oversimplified constraint that our disc is not allowed to deform in this initial implementation, it is very useful in producing a spatially varying prior that excludes unwanted structures from consideration as left ventricular myocardium. From the binary masks that comprise the output of the segmentation, we complete the construction of the prior by blurring, combining, and normalizing the masks to sum to 1 everywhere.

3 Results

The results in Figure 1 were computed using 10 iterations of the EM algorithm with 3 iterations of the MRF mean-field approximation inside each E-Step. CDN layers #1-3 are computed on a 3-D basis, while layer #4 is computed on a slice-by-slice basis.

4 Discussion

The method presented herein is an efficient means of achieving fully automatic results, and our contribution is two-fold. First, we propose analyzing motion prior to performing the segmentation. This alleviates the dependence on user initialization. Second, we incorporate context into the segmentation process through a hierarchical framework that we refer to as a Contextual Dependency Network. This organizes the computation according to the concept of *predication*: image properties that are de-

pendent on their context cannot be computed until an initial guess of the context has been made.

Our work differs from a majority of cardiac segmentation algorithms in our choice of a region-based approach over a contour-based method. We experimented with applying level set methods to cardiac MRI, and we encountered difficulty in resolving ambiguity. For example, while region borders are clearly defined between the LV and RV, the portion of the epicardium opposite the RV is vulnerable to disappearing into the noise. It can blend seamlessly with the liver when image resolution is insufficient to resolve the diaphragm, or the patient is devoid of a layer of fat. In these cases, bias-field inhomogeneity correction is of enormous benefit. Due to the subtleties of the intensity differences, merely performing surface-coil intensity correction as a pre-processing step is insufficient – it must be incorporated within the iterations of the segmentation algorithm.

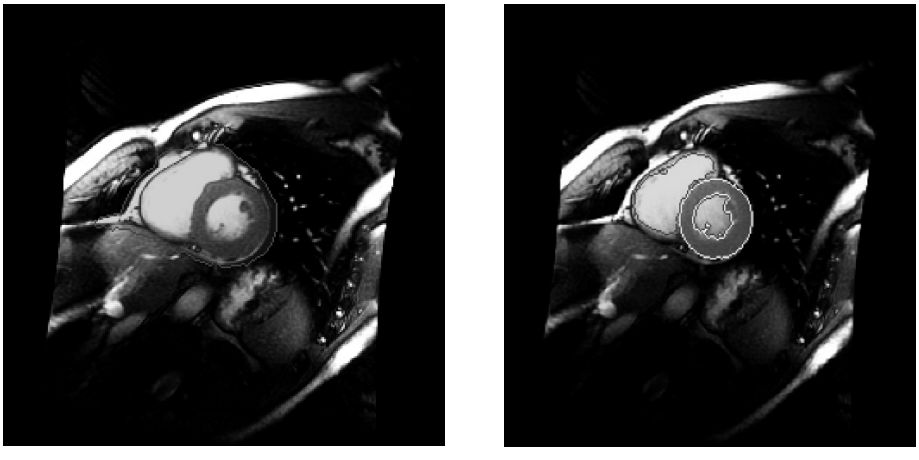

Fig. 1. LEFT: Input image with the ROI computed from temporal information overlaid in red. RIGHT: Final segmentation of myocardium (yellow) and ventricles (red). (Color version available at http://www.ai.mit.edu/people/gering/)

Equally severe is the adverse effects of image acquisition artifacts that are a by-product of the rapid scanning required for cardiac applications. These artifacts manifest as streaks that can be more robustly handled by a region-based approach than a contour-based method that is sensitive to all image gradients.

Furthermore, we differ from other region-based methods in that CDN layers #3-4 employ a *topological* atlas which has certain advantages in this application over registering a *geometric* atlas [8]. Specifically, cardiac MRI exams produce data sets of vast size (20 complete volumes distributed through time) and extensive variability. Patient-specific papillary muscles greatly influence the apparent myocardium, and pathology can produce a very irregularly shaped myocardium with significant deviations in thickness around its annulus.

We have presented a general framework with a simple implementation, but we emphasize the need for future work to substitute more sophisticated components into this framework. Toward this end, many of the other methods in the literature have valuable contributions, such as the more detailed shape models of [13], which can offer more value when incorporated into our framework. Our main concern is robustness, especially in cases involving pathology, and we are therefore conducting a clinical evaluation for a forthcoming publication. Additionally, our future work pursues more sophisticated implementations of the framework presented herein, especially the temporal processing.

References

1. C.W. Chen, J. Luo, K.J. Parker. "Image Segmentation via Adaptive K-Mean Clustering and Knowledge-Based Morphological Operations with Biomedical Applications". *IEEE Transactions on Medical Imaging* December 1998; 7:1673–1683.
2. A.P. Dempster, N.M. Laird, D.B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal Royal Statistical Society* 1977; 39:1–38.
3. J. Duncan, R. Owen, P. Anandan, L. Staib, T. McCauley, A. Salazar, F. Lee. "Shape-Based Tracking of Left Ventricular Wall Motion". *IEEE* 1991;41–44.
4. D.T. Gering, W.E.L. Grimson, R. Kikinis. "Recognizing Deviations from Normalcy for Brain Tumor Segmentation". In: T. Dohi, R. Kikinis, eds. *Medical Image Computing and Computer-Assisted Intervention*. Tokyo, Japan: Springer, 2002; 388–395.
5. D.T. Gering. "Recognizing Deviations from Normalcy for Brain Tumor Segmentation". Ph.D. Thesis, Massachusetts Institute of Technology, 2003 (http://www.ai.mit.edu/people/gering/).
6. G.H. Granlund. "Introduction to GOP Computer Vision". *Report LiTH-ISY-I-0849,* Dept. of Electrical Engineering, Linkoping University, May 5, 1987.
7. B. Horn, B. Schunck. "Determining Optical Flow". *Artificial Intelligence* 1981; 17:185–203.
8. M. Lorenzo-Valdes, G.I. Sanchez-Ortiz, R. Mohiaddin, D. Ruecker. "Atlas-based Segmentation of 4D Cardiac MR Sequences using Non-rigid Registration". In: T. Dohi, R. Kikinis, eds. *Medical Image Computing and Computer-Assisted Intervention*. Tokyo, Japan: Springer, 2002;.
9. M. Mignotte, J. Meunier, J. Tardif. "Endocardial Boundary Estimation and Tracking in Echocardiographic Images using Deformable Templates and Markov Random Fields". *Pattern Analysis and Applications* 2001; 4:256–271.
10. I. Mikic, S. Krucinski, J.D. Thomas. "Segmentation and Tracking in Echocardiographic Sequences: Active Contours Guided by Optical Flow Estimates". *IEEE Transactions on Medical Imaging* April 1998; 17:274–284.
11. A. Montillo, D. Metaxas, L. Axel. "Automated Segmentation of the Left and Right Ventricles in 4D Cardiac SPAMM Images". In: T. Dohi, R. Kikinis, eds. *Medical Image Computing and Computer-Assisted Intervention*. Tokyo, Japan: Springer, 2002; 620–633.
12. D. Terzopoulos, R. Szeliski. "Tracking with Kalman Snakes". In: A. Blake, A. Yuille, eds. *Active Vision*. Cambridge, MA: MIT Press, 1992; 3–20.

13. A. Tsai, A. Yezzi, W. Wells III, C. Tempany, D. Tucker, A. Fan, W.E.L. Grimson, A. Willsky. "Model-Based Curve Evolution Technique for Image Segmentation". In: *Computer Vision and Pattern Recognition (CVPR)*. Hawaii: 2001; .
14. W.M. Wells III, R. Kikinis, W.E.L. Grimson, F.A. Jolesz. "Adaptive Segmentation of MRI Data". *IEEE Trans Med Imaging* 1996;
15. J. Zhang. "The Mean Field Theory in EM Procedures for Markov Random Fields". *IEEE Transactions on Image Processing* 1992; 40:2570–2583.