3D CARDIAC SEGMENTATION WITH POSE-INVARIANT HIGHER-ORDER MRFS

Bo Xiang[1,2], *Chaohui Wang*[1,2], *Jean-Francois Deux*[3], *Alain Rahmouni*[3], *Nikos Paragios*[1,2]

[1] Center of Visual Computing, Ecole Centrale de Paris, France
[2] Equipe GALEN, INRIA Saclay - Île de France, France
[3] Radiology Department, Henri Mondor Hospital, Créteil, France

ABSTRACT

This paper proposes a novel pose-invariant segmentation approach for left ventricle in 3D CT images. The proposed formulation is modular with respect to the image support (i.e. landmarks, edges and regional statistics). The prior is represented as a third-order Markov Random Field (MRF) where triplets of points result to a low-rank statistical prior while inheriting invariance to global transformations. The ventricle surface is determined through triangulation where image discontinuities can be easily evaluated and the Divergence theorem provides an exact calculation of regional statistics acting on the image or a derived feature space. Promising results using boosting along with the learned prior demonstrate the potential of our method.

Index Terms— Left ventricle segmentation, pose-invariant shape prior, higher-order MRF

1. INTRODUCTION

Left ventricle (LV) segmentation is a prerequisite for quantitative analysis on diagnosing cardiovascular diseases. It is a well studied problem in biomedical imaging and most of the existing methods combine anatomical prior knowledge with observed visual support towards automatic delineation of LV.

Early approaches have considered slice-by-slice segmentation [1] using active contours.The use of active shape models [2] as well as active appearance models was a step forward where geometric and photometric variations were modeled using linear sub-spaces either at the frame level or during the entire cardiac cycle. Level set methods [3] were also considered for LV segmentation towards providing accurate integration of boundary and visual support. Non-parametric statistics was an alternative to linear subspaces towards capturing more complex variations. Graph-based methods [4] have also emerged due to their computational efficiency combined with certain user interaction. Spatio-temporal models involving poly-parametric class of transformations were recently investigated. Last but not least, the biomechanical prior models [5] modeled the complete cardiac behavior which was calibrated through the image observations.

All the above methods are modeled in a common subspace where statistics are learned in order to describe the variability among individuals. Such a process introduces strong statistical bias and at the same time makes the applicability of the model problematic when referring to diseased subjects. Furthermore, the construction of prior models is problematic due to the lack of sufficient number of training examples given the dimensionality of the model. Last but not least, these methods account partially for visual support with some of them acting on the boundaries, some of them on the regional statistics and some on anatomical landmarks. The aim of this paper is to propose a unified formulation that can integrate all existing sources of information, while being pose-invariant and able to cope with linear and non-linear statistics.

To this end, we adopt a Point Distribution Model associated with higher-order MRF. Triangulation on the surface provides third-order terms where an exact estimation of the boundary support and regional statistics is feasible through the Divergence theorem [6]. Furthermore, third-order cliques are also considered to encode prior knowledge [7]. Incremental deformation of the model coupled with landmark constraints is considered within a dual-decomposition approach to recover the optimal solution into a new image. Our approach is most closely related to the work [6, 7]. The problem solved in [7] is simply graph matching (corresponding to a certain extend to the landmark part of the proposed formulation), while the 2D applied method in [6] neither inherits global pose invariance nor encompasses the use of landmarks.

The reminder of the paper is organized as follows. Section 2 presents the pose-invariant model and its associated statistical shape prior. In Section 3 we discuss the integration of visual and prior terms in a sound theoretical framework. Experimental results are presented in Section 4 and discussion in Section 5 concludes the paper.

2. POSE-INVARIANT POINT DISTRIBUTION MODEL

An essential task in knowledge-based image segmentation is to determine a statistical shape model as a prior knowledge which is then combined with visual support in an

inference process. We represent the shape as a Point Distribution Model $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ (Fig.1), consisting of a set $\mathcal{V} = \{1, \cdots, n\}$ of n control points (marked as blue dots in Fig.1(a)) lying on the boundary. Let $\mathbf{x}_i (i \in \mathcal{V})$ denote the 3D position of the control point i. Meanwhile a subset $\mathcal{V}_l \subset \mathcal{V}$ of points are considered as anatomical landmarks (marked as green crosses in Fig.1(a)). Furthermore, we associate this model with a set of triangles $\mathcal{T} = \{\mathbf{t}_{ijk} = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) | i, j, k \in \mathcal{V}, i \neq j, j \neq k, k \neq i\}$ connecting any three points i, j, k. A subset $\mathcal{S} \subset \mathcal{T}$ of triangles are defined as the surface triangles (Fig.1(b)) which compose a closed mesh (2-manifold).

Based on the shape representation, the next task is to build a statistic model from a training set in order to construct the prior. The existing global approaches, such as ASM, often suffer from their lack of flexibility, while they also require a significant number of training samples to capture statistics due to their high-dimensional representation. To overcome these limitations, we propose a similarity-invariant model based on local interactions.

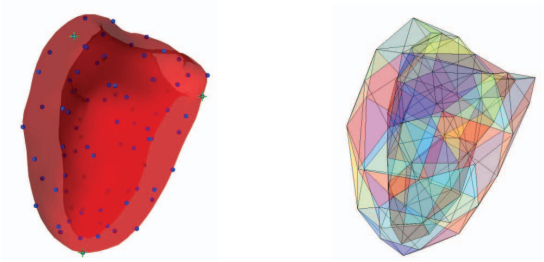
Let us consider a triplet of points i, j and k, the geometrical constraint between their positions $\mathbf{t}_{ijk} = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ can be charactered in a pose-invariant manner using two inner angles $\{\alpha = \angle(\mathbf{x}_i\mathbf{x}_j\mathbf{x}_k), \beta = \angle(\mathbf{x}_j\mathbf{x}_k\mathbf{x}_i)\}$ of the triangle $\mathbf{t}_{ijk}$, while the third angle is a linear combination of the two. The inner angle can be computed as follows:

$$\alpha = \arccos \frac{\overrightarrow{\mathbf{x}_i\mathbf{x}_j} \cdot \overrightarrow{\mathbf{x}_j\mathbf{x}_k}}{\|\mathbf{x}_i\mathbf{x}_j\| \, \|\mathbf{x}_j\mathbf{x}_k\|} \tag{1}$$

An important property of the angle measurement is that it leads to a shape model invariant with respect to the global pose (i.e. translation, rotation and scale).

Given a training set $\mathbf{X} = \{\mathcal{X}_1, \cdots, \mathcal{X}_m\}$ of m instances of the shape, each example consists of n control points which are located in correspondences. For $\forall i \in \mathcal{V}$, we have the set $\mathbf{X}_i = \{\mathbf{x}_i^1, \cdots, \mathbf{x}_i^m\}$ representing the instances of the i^{th} control point in the shape. Then using a statistical model, we learn from the training set the probability density distributions $p_{ijk}(\alpha, \beta)$ of the angles which are determined by a triplet of control points i, j and k. Let $\mathcal{C} = \{(i, j, k) | i, j, k \in \mathcal{V}, i \neq j, j \neq k, k \neq i\}$ denote the triplet set corresponding to the triangle set $\mathcal{T}$. Thus $\forall c \in \mathcal{C}$, a distribution $p_c(\alpha, \beta)$ is learned to describe the point position dependencies in the triplet c. With the accumulation of all these local distribution densities, the prior manifold is constructed through all triplet interactions. In this manner, our shape model encodes global consistency as well as local variations. Without loss of generality, both simple and complex distribution models can be used to learn the statistics. We use a Gaussian distribution model in practice.

So far, we build our pose-invariant point distribution model $\mathbf{M} = (\mathcal{X}, \mathcal{C}, \mathcal{P})$. It consists of a set of control points $\mathcal{X}$, a triplet set $\mathcal{C}$, and the statistical shape prior $\mathcal{P} = \{p_c(\alpha, \beta) | c \in \mathcal{C}\}$. Based on this representation, knowledge-based segmentation can be easily encoded in a MRF model towards efficient optimization.



(a) Distribution of the control points (b) Triangulated mesh

Fig. 1. 3D myocardium Point Distribution Model.

3. KNOWLEDGE-BASED SEGMENTATION

Cardiac computed tomography (CT) image segmentation is challenging due to its complex background and low contrast. We aim to segment the myocardium structure which is the muscle between endocardium and epicardium. However, the boundaries of myocardium are not clearly visible because: (1) endocardium detection can be disturbed by the papillary muscles in the blood pool; (2) epicardium is not distinct between the myocardium and the right ventricle. To deal with these problems, we combine the anatomical landmarks, regional and boundary support to achieve a good segmentation.

We formulate the segmentation problem as an energy minimization problem. Given an initial position of the shape model, we aim to search for the optimal control point displacements to best compromise between the visual support and shape-fitness in an observed image. To this end, we use a higher-order Markov Random Field (MRF) formulation, in which a node models a control point and a third-order clique models the interaction by a triplet.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{C})$, $\mathcal{V}$ denotes the node set and $\mathcal{C}$ denotes the three-order clique set. We also denote a node set $\mathcal{V}_l \subset \mathcal{V}$ corresponding to the anatomical landmarks and a clique set $\mathcal{C}_s \subset \mathcal{C}$ corresponding to the surface triangles. Let X_i $(i \in \mathcal{V})$ denote the latent variable (i.e. the 3D position) of node i, and $\mathcal{U}_i$ denote the candidate space for the configuration x_i. To this end, the segmentation problem is formulated as the inference of the configuration $\Theta = (x_i)_{i \in \mathcal{V}}$ of all the nodes over the candidate space $\mathcal{U} = \prod_{i \in \mathcal{V}} \mathcal{U}_i$:

$$\Theta^{\text{opt}} = \arg\min_{\Theta \in \mathcal{U}} E(\Theta) \tag{2}$$

The MRF energy $E(\Theta)$ is defined as a combination of landmark energy ϕ, region and boundary energy $\psi^{(1)}$ and shape prior energy $\psi^{(2)}$, while the first is a singleton term and the other two are higher-order (third-order) terms:

$$E(\Theta) = \sum_{i \in \mathcal{V}_l} \phi_i(x_i) + \sum_{c \in \mathcal{C}_s} \psi_c^{(1)}(x_c) + \sum_{c \in \mathcal{C}} \psi_c^{(2)}(x_c) \tag{3}$$

The definitions of each energy are described in following.

3.1. Landmark Correspondences

In the proposed graphical-model framework, we consider a set of salient points/anatomical landmarks (e.g. an apical control point) among the vertices of the surface model. They can be localized by a detector based on machine learning techniques such as Randomized Forests [8]. A singleton term is then introduced to enforce the consistency between the detected position and the estimated one for each landmark i:

$$\phi_i(x_i) = C_0 \cdot |x_i - \hat{x}_i|^2 \tag{4}$$

where $\hat{x}_i$ denotes the detected position for landmark i.

3.2. Regional Statistics & Boundary Support

Region-driven and boundary-based terms are based on the hypothesis that the populations in the image can be separated by their statistical properties. To represent the statistics, voxel intensity as a low-level representation has the limitation on the capability to recognize different objects, for example when the intensities of objects are overlapped. In this context, we use a high-dimensional feature vector for each voxel instead of an intensity value. The feature vector consists of patches of intensities and Gabor features [9] in view of their abilities to capture texture features in different scales and orientations. In our case, a $5 \times 5 \times 5$ patch and Gabor features with 3 scales and 6 orientations are used.

Based on the feature images, we learn an Adaboost classifier [10] for the object and background and we apply the classifier responses to obtain a likelihood image of the testing image (Fig.2). Each voxel is represented by a signed score in [-1, 1], whose sign symbol indicates the pixel label (i.e. negative for object and positive for background) and absolute value indicates the possibility to be object or background (distinguished in blue and red in Fig.2(b)). The likelihood image is used for both region and boundary related potentials.

With regard to the regional potential, the exact calculation of regional likelihood on the volume can be factorized by using Divergence theorem, which states that volume V integral can be transferred into surface S integral.

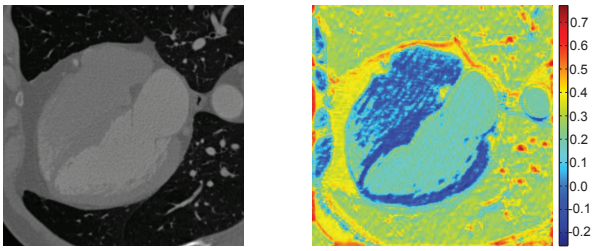
$$\iiint_V f \, dv = \oiint_S (\mathrm{F} \cdot \mathrm{n}) \, ds = \sum_{\mathrm{c} \in \mathcal{C}_\mathrm{s}} \gamma_\mathrm{c}^{\text{region}}(\mathrm{x_c}) \tag{5}$$

where $f = \nabla \cdot \mathrm{F}$ is the likelihood image, n is the outward pointing unit normal of the boundary surface S. For surface triplet $c \in \mathcal{C}_s$, the regional potential is defined as follows:

$$\gamma_c^{\text{region}}(x_c) = C_1 \cdot \iint_{s_c} (\mathrm{F} \cdot \mathrm{n}) \, ds \tag{6}$$

where s_c denotes the surface area determined by triplet c. Furthermore, we define the boundary related energy as a surface integral on surface triplet $c \in \mathcal{C}_s$:

$$\gamma_c^{\text{boundary}}(x_c) = C_2 \cdot \iint_{s_c} M(x) \, ds \tag{7}$$



(a) Original image (b) Likelihood colormap

Fig. 2. Image Likelihood.

where $M(x)$ is a boundary map which is the output of an boundary detector. For simplicity, we measure it by the dot product of the unit facet normal and the likelihood gradient at the position x, i.e. $M(x) = -\mathrm{n} \cdot \nabla f$. This weak boundary detector is sufficient since we combine other image cues and shape prior in our framework. To this end, the data related energy defined on surface triplet $c \in \mathcal{C}_s$ is:

$$\psi_c^{(1)}(x_c) = \gamma_c^{\text{region}}(x_c) + \gamma_c^{\text{boundary}}(x_c) \tag{8}$$

3.3. Prior Knowledge Constraints

The prior term encodes the spatial constraints of the control points configuration with respect to the shape manifold. The energy of local interaction on triplet $c \in \mathcal{C}$ is defined as:

$$\psi_c^{(2)}(x_c) = -\log p_c(\alpha, \beta) \tag{9}$$

where $p_c(\alpha, \beta)$ denotes the possibility density of angles in the triplet c, which is obtained in the statistical shape model.

4. EXPERIMENTAL VALIDATION

We validate the proposed method on a dataset of 20 3D CT cardiac volumes. The volumes from different subjects have an approximate mean size of $512 \times 512 \times 250$ voxels, where the voxel size is about $0.35 \times 0.35 \times 0.5mm^3$. Manual segmentation on the dataset is available and is considered as ground truth using for both learning and validation procedures.

We perform a leave-one-out cross-validation on the dataset. A dual-decomposition optimization framework [11] is adopted to perform the Maximum-a-Posteriori (MAP) inference for the proposed higher-order MRF. The weight parameters C_0, C_1, C_2 in MRF energy potentials are once adjusted manually to balance the corresponding terms for a good performance and then are kept for all the testing volumes. For quantitative evaluation, we compute the Dice coefficient between the segmented volume and the ground truth, which is one of the most common similarity measures between two volumes. We compare the proposed method with a Random Walks algorithm [4] and a baseline ASM [2] using the same likelihood image. In Fig.3 we present the distributions of the Dice coefficients obtained by the three methods

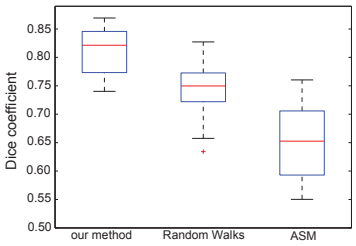


Fig. 3. Boxplot of the Dice coefficient statistics.

evaluating on the whole dataset. Noted that a higher Dice coefficient implies a more accurate segmentation performance, our method outperforms the related work. Visual examples of experimental results are presented in Fig.4, showing that the proposed method deals well with the shape variations and poor image quality. Both quantitative and qualitative analyses demonstrate the potential of our method.

5. CONCLUSION

In this paper, we propose a novel segmentation method based on graph theory with its explicit representations and have tested it using cardiac images. To the best of our knowledge, this is the first formulation that is pose invariant and can integrate common visual supports of any nature. It is achieved through the use of higher-order MRF where third-order cliques encode pose-invariant prior knowledge as well as boundary and regional statistics.

The main limitation of the method is the inherited computational complexity that is proportional to the number of higher-order cliques. We are investigating a more compact manner to encode prior knowledge that requires the smallest possible number of third-order cliques without altering the ability to express the shape manifold. Combined segmentation with landmark selection is another promising direction since geometric methods are known to be robust to the initial conditions. Last but not least, the extension of the method to account for the temporal nature of the cardiac motion through higher-order spatio-temporal priors is under investigation.

6. REFERENCES

[1] M.P. Jolly, "Automatic segmentation of the left ventricle in cardiac MR and CT images," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 151–163, 2006.

[2] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, et al., "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[3] X. Huang and D.N. Metaxas, "Metamorphs: Deformable shape and appearance models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 1444–1459, 2007.

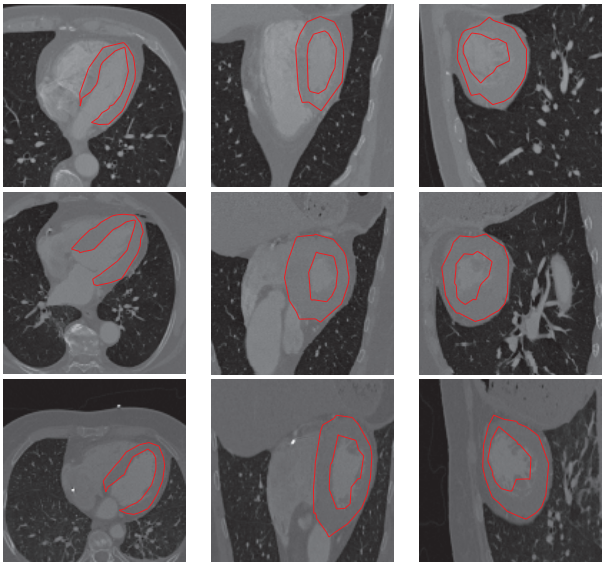


Fig. 4. Examples of segmentation results on 3D CT volumes. Each row represents a volume with three orthogonal views.

[4] Leo Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[5] M. Sermesant, H. Delingette, and N. Ayache, "An electromechanical model of the heart for image analysis and simulation," *IEEE Trans. Medical Imaging*, vol. 25, no. 5, pp. 612–625, 2006.

[6] B. Xiang, C. Wang, J.F. Deux, A. Rahmouni, and N. Paragios, "Tagged cardiac mr image segmentation using boundary & regional-support and graph-based deformable priors," *ISBI*, pp. 1706–1711, 2011.

[7] C. Wang, O. Teboul, F. Michel, and N. Paragios, "3D knowledge-based segmentation using pose-invariant higher-order graphs," *MICCAI*, pp. 189–196, 2010.

[8] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] A. Bernardino and J. Santos-Victor, "Fast iir isotropic 2-d complex gabor filters with boundary initialization," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3338–3348, 2006.

[10] M. Collins, R.E. Schapire, and Y. Singer, "Logistic regression, AdaBoost and Bregman distances," *Machine Learning*, vol. 48, no. 1, pp. 253–285, 2002.

[11] N. Komodakis, N. Paragios, and G. Tziritas, "MRF optimization via dual decomposition: Message-passing revisited," *ICCV*, 2007.