Automatic Delineation of Left and Right Ventricles in Cardiac MRI Sequences Using a Joint Ventricular Model

Xiaoguang Lu[1,⋆], Yang Wang[1], Bogdan Georgescu[1], Arne Littman[2], and Dorin Comaniciu[1]

[1] Siemens Corporate Research, Princeton, NJ, USA
xiaoguang.lu@siemens.com
[2] Magnetic Resonance, Siemens Healthcare, Erlangen, Germany

Abstract. Cardiac magnetic resonance imaging (MRI) has advanced to become a powerful tool in clinical practice. Extraction of morphological and functional features from cardiac MR imaging for diagnosis and disease monitoring remains a time-consuming task for clinicians. We present a fully automatic approach to extracting the structures and dynamics for both left and right ventricles. The cine short-axis stack of a cardiac MR scan is used to reconstruct a 3D volume sequence. A joint LV-RV model is introduced to delineate the boundaries of both left and right ventricles in each frame, and to combine both spatial and temporal context to track the chamber boundary motion over cardiac cycles. Both qualitative and quantitative results show promise of the proposed method.

1 Introduction

Accurate morphological and functional measurements of the heart anatomies are essential in clinical applications for diagnosis, prognostic, and therapeutic decisions. Magnetic resonance imaging (MRI) allows morphological characterization of heart structures with precision. An accurate identification of the borders of the structures to be analyzed is needed in order to extract physiologically meaningful quantitative information from the images. Potential applications of cardiac segmentation and tracking include the calculation of volume and mass, blood ejection fraction, analysis of contraction and wall motion as well as the 3D visualization of cardiac anatomy [1].

Advantages of cardiac MRI include a wide topological field of view with visualization of the heart and its internal morphology and surrounding mediastinal structures. It has a high soft-tissue contrast discrimination between the flowing blood and myocardium without the need for contrast medium or invasive techniques. Cardiac MR is able to perform multiple non-harmful and accurate scans required for disease monitoring. In such a scenario, fast, reproducible and accurate extraction of clinical features is essential for all decision support systems.

⋆ Corresponding author.

D.N. Metaxas and L. Axel (Eds.): FIMH 2011, LNCS 6666, pp. 250–258, 2011.

In addition to assessing left ventricle (LV) functions, recent research emphasizes the importance of right ventricle (RV) function in the prognosis of a variety of cardiopulmonary diseases. This indicates that there is a growing interest in the clinical relevance of both LV and RV; in particular for congenital diseases and that more routine quantification of RV function is warranted under most clinical circumstances. Because of ventricular interactions, RV filling influences LV performance, and similarly, the LV affects RV function through the interventricular septum. Dynamics of the RV can also infer a large amount of clinical information [2,3,4]. The normal RV anatomy is a complex crescent-shaped structure wrapped around the LV. The RV can in some cases (especially in diseased patients) be characterized by its non-uniform shape and high degree of trabaeculations. Due to the complexity of shape and dynamics, much of the research on the LV cannot be easily transferred to the RV.

In current clinical practice, assessment of RV structure and function remains mostly qualitative, which involves manually delineating the inner wall (endocardium) of the RV, and requires a great deal of user interaction with generally no a-priori information. Correlation between LV and RV shows promise to improve RV segmentation, as explored in [5], where five landmarks need to be manually identified for the subsequent automatic segmentation. We proposed an automatic approach to delineating LV and RV without any user interactions.

2 Methodology

Due to the large amount of available data, analysis such as segmentation of cardiac images is time consuming and error-prone for human operators, which needs to be automated in order to be clinically valuable. We present an automatic approach to extracting relevant morphological and dynamic parameters of both LV and RV from MRI data over a cardiac cycle, as demonstrated in Fig. 1. Conventional volume measurements by cardiac MRI are independent of the cavity shape, with the area from contiguous slices integrated over the chamber of interest. However, direct 3D extraction may take advantage of the chamber shape and provide heart movement measurement in three dimensions, leading to more accurate and realistic representations.

The proposed method segments chambers on the first frame, followed by dynamics extraction across the entire sequence, as shown in Fig. 2. Chamber segmentation includes two stages: at the first stage, the position, orientation, and scale of the heart chamber (LV/RV) in a 3D volume are determined to initialize the joint model for chamber segmentation; at the second stage, local deformation of the detected model is processed in order to fit the model to both LV and RV boundaries.

2.1 Joint Anatomy Model

In order to accurately model the complexity of the anatomy, a representation of the anatomic shape is created using a database of reconstructed 3D volumes that are manually segmented. For left ventricle, the model includes LV endocardium,

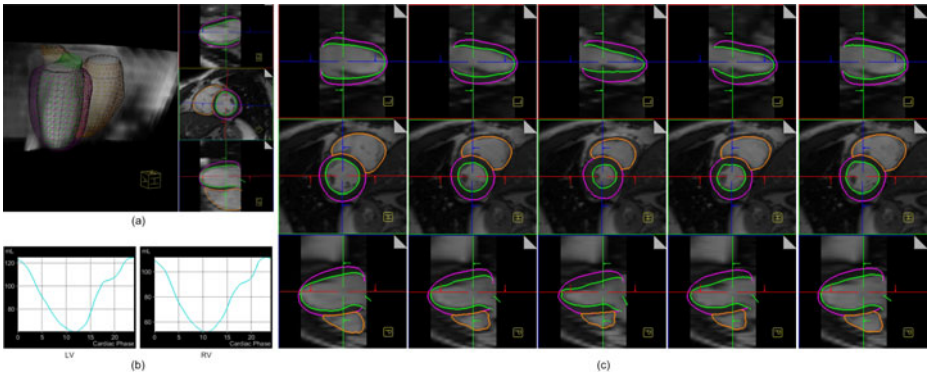

Fig. 1. Models of LV/RV fitted to a 3D reconstructed cardiac MRI volume sequence. (a) Estimated 3D model. (b) Volume measurement across time computed based on the fitted models. (C) 2D views of frame 1, 6, 11, 16, 21 of a single heartbeat cycle (25 frames in total).

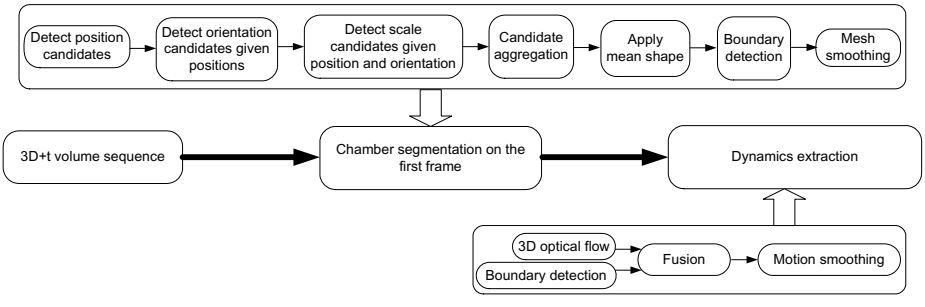

Fig. 2. Workflow of our automatic LV and RV detection and tracking system

LV epicardium, and LV outflow tract (LVOT). The right ventricle model consists of the RV blood pool cavity, the RV outflow tract (RVOT) as far as the pulmonary valve and the tricuspid valve opening. Both LV and RV models are triangular meshes as shown in Fig. 3. They are used to fit a given 3D cardiac volume to delineate corresponding anatomical structures. The joint model unifies the interventricular septum between LV the RV.

2.2 Learning-Based Model Fitting

A typical cardiac MR scan to examine the LV/RV morphology and functionality contains a short axis stack, which consists of image slices captured at the different positions along the short axis of heart chambers (e.g., LV). These image slices can be aligned using the physical coordinates (location and orientation) recorded during acquisition. A 3D volume is reconstructed from this stack of aligned image slices. If each image slice is captured in a time sequence and synchronized to

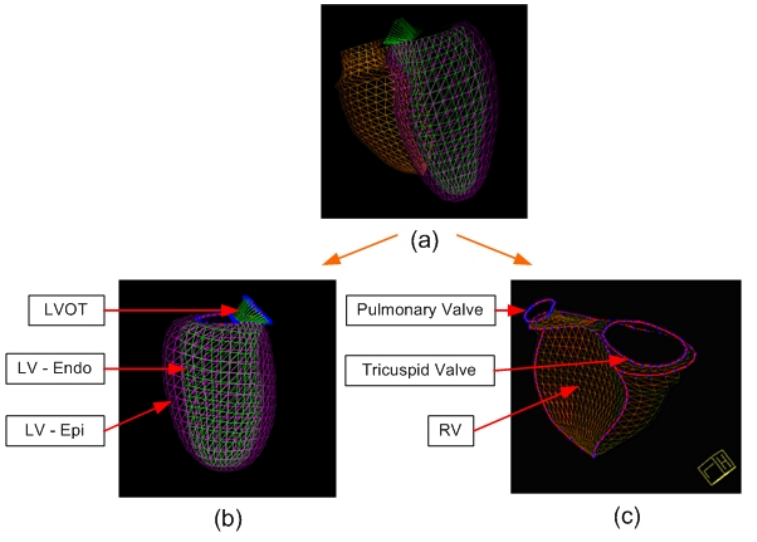

Fig. 3. Joint model (a) of LV (b) and RV (c). The RV inner boundary and LV epicardium are jointly modeled, sharing the same triangular mesh parts.

each other, a 3D volume sequence is obtained, which is used for 3D chamber segmentation and dynamics extraction in our proposed system.

To estimate our joint anatomical models in the reconstructed volumetric data, we train a series of detectors to estimate the model pose (including translation, orientation, and scale) and boundaries on a large database with both LV and RV annotated. We use a probabilistic boosting tree (PBT) [6] for each detector, which selects a set of discriminative features that are used to distinguish the positive samples from negatives from a large pool of features. Image orientation information recorded during acquisition can also be used to help initialize pose estimation. For the detector at the translation stage, we choose 3D Haar wavelet-like features [7], which are calculated efficiently using integral image based techniques. For the detectors at the orientation and scale search stages, steerable features [8] are applied because they do not require volume rotation and re-scaling which are computationally expensive, especially when the search hypothesis space is large.

In order to detect the model pose, we need to solve for the nine-parameter space, including three translations, three orientations, and three scales, i.e.,

$$\theta = \{(c_x, c_y, c_z), (\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\alpha}_z), (s_x, s_y, s_z)\} \tag{1}$$

where (c_x, c_y, c_z), $(\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\alpha}_z)$, (s_x, s_y, s_z) are the position, orientation and scale parameters. To estimate the above parameters efficiently, we apply a marginal space search strategy [8], which groups the original parameters space into subsets of increasing marginal spaces such that the posterior probability can be expressed as:

$$\begin{aligned} p(\theta_t|I_t) = p(c_x, c_y, c_z|I_t)p(\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\alpha}_z|c_x, c_y, c_z, I_t) \\ p(s_x, s_y, s_z|\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\alpha}_z, c_x, c_y, c_z, I_t) \end{aligned}$$

We train a series of detectors that estimate parameters at a number of sequential stages in the order of complexity, i.e., translation, orientation, and scale. Different stages utilize different features computed from 3D volumetric data. Multiple hypotheses are maintained between stages, which quickly removes false hypotheses at the earlier stages while propagates the right hypothesis to the final stage. Only one hypothesis is selected as the final detection result.

With the model pose estimated, we align the mean shape (an average model of all annotations) with data to get an initial estimate of the object shape. To capture the true anatomical morphology of the target object (e.g., LV and RV), we deform the mean shape by searching the boundary for each vertex of the model. The boundary hypotheses are taken along the normal directions at each vertex of the mean model. Detection is achieved using a boundary detector using PBT with steerable features. The detected boundaries are constrained by projecting the detected model onto a shape subspace obtained by the annotated dataset, which was constructed using principal component analysis. Although more sophisticated representations, such as local affine models [9,10], can also be applied to constrain shape deformations, we choose the global PCA shape model due to its efficiency during online detection.

The joint LV-RV model provides a mechanism to improve search and fitting accuracies based on the geometric constraint between LV and RV. For example, when LV is robustly detected, the RV parameter search range can be inferred to be in a much smaller space than the original RV search range without any a-prior information. The unified septum boundary constraint (LV and RV must share the same septum boundary) improves model fitting. Therefore, the joint model leads to more robust and accurate detection and delineation of the anatomies.

2.3 Dynamics Extraction

In this section, we present our tracking method to extract dynamic shape deformation automatically from an MRI sequence, which includes three main steps: initialization, deformation propagation, and motion smoothing. In the initialization step, the learning-based model fitting approach is applied to the initial frame to detect the shape for both LV and RV, as described in Sec. 2.2.

Starting from the detection result at the initial frame, the model deformations are propagated to neighboring frames using both the learned features and the local image templates. To ensure temporal consistency and smooth motion and to avoid drifting and outliers, two collaborative trackers, an intensity-based matching tracker and a boundary detection tracker, are used in our method. The intensity-based matching tracker directly computes the temporal displacement for each point from one frame to the next based on the image intensity, while the detection tracker obtains the deformations in each frame with maximal probability [11]. The above two trackers are integrated into a single Bayesian framework:

$$\arg\max_{\boldsymbol{X_t}} p(\boldsymbol{X_t}|\boldsymbol{Y_{t-1:t}}) = \arg\max_{\boldsymbol{X_t}} p(\boldsymbol{Y_t}|\boldsymbol{X_t}) p(\boldsymbol{X_t}|\boldsymbol{Y_{t-1}}), \tag{2}$$

where $\boldsymbol{Y_{t-1:t}} = (\boldsymbol{Y_{t-1}}, \boldsymbol{Y_t})$ are the image intensity and local feature responses from the two neighboring frames $I_{t-1:t} = (I_{t-1}, I_t)$. For clarity, we use $\boldsymbol{X_t}$ to

denote a concatenation of the mesh point positions, $\boldsymbol{X_t} = [X_1, \cdots, X_n]$, which need to be estimated at the current time instance t, and n is the total number of points in the mesh model.

The likelihood term $p(\boldsymbol{Y_t}|\boldsymbol{X_t})$ is computed from both boundary detection and local image template matching as follows,

$$p(\boldsymbol{Y_t}|\boldsymbol{X_t}) = (1-\lambda)p(F_t|\boldsymbol{X_t}) + \lambda p(T_t|\boldsymbol{X_t}), \tag{3}$$

where F_t is the steerable feature response [8], T_t is the local image template centered at $\boldsymbol{X_{t-1}}$ in the previous frame I_{t-1}, and λ is the weighting coefficient of the matching term. Given the resulting shape $\boldsymbol{X_{t-1}}$ from the previous frame $t-1$, the prediction term $p(\boldsymbol{X_t}|\boldsymbol{Y_{t-1}})$ can be simplified as $p(\boldsymbol{X_t}|\boldsymbol{X_{t-1}})$. In our system $p(\boldsymbol{X_t}|\boldsymbol{X_{t-1}})$ is modeled as a Gaussian distribution based on the shape distance $d(\boldsymbol{X_t}, \boldsymbol{X_{t-1}}) = \parallel \boldsymbol{X_t} - \boldsymbol{X_{t-1}} \parallel$. The objective function (2) can be optimized by searching in a local neighborhood centered at $\boldsymbol{X_{t-1}}$. To speed up the computation, we apply the optical flow technique to search for the new position $\boldsymbol{X_t}$ along the local gradient direction in the current frame I_t.

The above deformation propagation step is repeated until the full 4D model is estimated for the complete sequence. In this way the collaborative trackers complement each other, as the intensity-based matching tracker provides temporally consistent results and its major issue of drifting is addressed by the boundary detection.

Finally to obtain a smooth motion field, the tracking is performed in both forward and backward directions given the periodic nature of the cardiac motion, and a Gaussian kernel is applied to both the LV and RV shapes in the neighboring frames, i.e., $\boldsymbol{X_t^{smooth}} = \sum_{i=-k}^{k} G(i)\boldsymbol{X_{t+i}}$, where $G(i)$ is a normalized Gaussian kernel $N(0, \sigma)$. In our experiments we typically choose $\sigma = 0.6$ and $k = 1$.

3 Experiments

We collected 100 reconstructed volumes from 70 patients with left ventricles annotated, among which 93 reconstructed volumes from 63 patients were also annotated on right ventricles. Volumes were selected to cover a large range of dynamic heart motion, including both end diastole and end systole. The original short-axis stack images have an average in-plane resolution of $1.35mm$, and the distance between slices is around $10mm$.

A 4-fold cross-validation scheme was applied for evaluation. The entire LV dataset was randomly partitioned into four quarters. For each fold evaluation, three quarters were combined for training and the remaining one was used as unseen data for testing. This procedure was repeated four times so that each volume has been used once for testing. The same evaluation protocol was applied for RV. For each segmented mesh, the distance from each vertex to the groundtruth mesh (manual annotation) was computed as point-to-mesh distance. The average distance from all vertices of the segmented mesh was used as the measurement. Three major components, i.e., LV endocardium, LV epicardium, and RV main

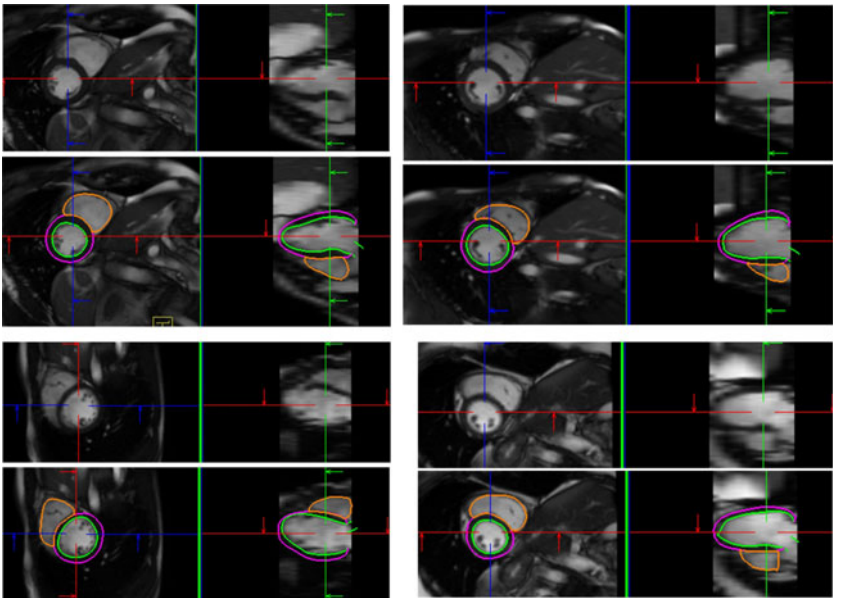

Fig. 4. Automatic delineation examples

Table 1. Point-to-mesh distance measurements obtained by a 4-fold cross validation based on the joint LV-RV model

measure (mm)	Mean	Std	Median
LV endocardium	2.95	4.85	1.84
LV epicardium	3.23	3.94	2.12
RV main	2.99	1.18	2.66

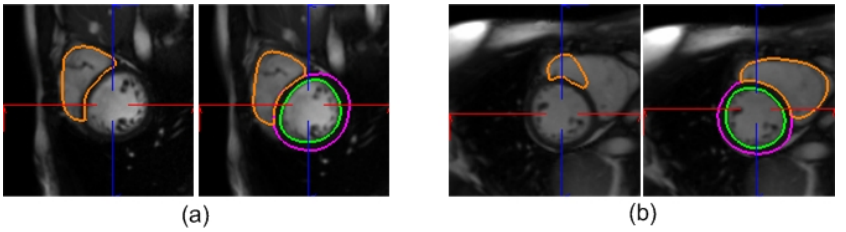

Fig. 5. Examples of superior performance of the joint LV-RV model over individual RV modeling. In both (a) and (b), the left ones are the delineation results obtained by the individual RV model and the right ones are from the proposed joint LV-RV model. The individual RV model overestimates the RV in (a) and significantly underestimates the RV in (b), while the joint LV-RV model provides correct delineation in both cases.

Table 2. Comparison of RV delineation results using joint LV-RV model against individual RV model. Point-to-mesh distance measurements are calculated.

measure (mm)	Mean	Std	Median
Individual RV model	4.12	6.28	2.65
Joint LV-RV model	2.99	1.18	2.66

cavity, were considered in our evaluation as listed in Table 1. Automatic delineation examples are provided in Fig. 4. Fig. 5 shows examples where joint LV-RV modeling provides superior performance to individual RV modeling. Table 2 summarizes the quantitative results obtained by joint modeling from the 4-fold cross validation in comparison with individual RV modeling.

On the average, it took about 3 seconds to segment both LV and RV from a single volume (e.g, $256\times256\times70$), and about 40 seconds to fully extract dynamics of the entire sequence (typically 20 frames) on a duo core 2.8GHz CPU.

4 Conclusions

We have presented a fully automatic method for segmenting LV and RV chambers from cardiac MRI images, and extracting the dynamics of both chamber movements. A joint ventricular model is used to delineate the boundaries of both LV and RV in each frame. Clinically relevant measurements, such as volumes and ejection fraction, can be calculated based on the fitted model.

References

1. Frangi, A., Niessen, W., Viergever, M.: Three-dimensional modeling for functional analysis of cardiac images: A review. IEEE Trans. on Medical Imaging 20(1) (2001)
2. Grothues, F., Moon, J., Bellenger, N., Smith, G., Klein, H., Pennell, D.: Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance. American Heart Journal 147(2), 218–223 (2004)
3. Corsi, C., Caiani, E., Catalano, O., Antonaci, S., Veronesi, F., Sarti, A., Lamberti, C.: Improved quantification of right ventricular volumes from cardiac magnetic resonance data. Computers in Cardiology 57(2), 37–40 (2005)
4. Haber, E., Metaxas, D., Axel, L., Wells, W., Colchester, A., Delp, S.: Motion analysis of the right ventricle from MRI images. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, p. 177. Springer, Heidelberg (1998)
5. Sun, H., Frangi, A.F., Wang, H., Sukno, F.M., Tobon-Gomez, C., Yushkevich, P.A.: Automatic cardiac MRI segmentation using a biventricular deformable medial model. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 468–475. Springer, Heidelberg (2010)
6. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, pp. 1589–1596 (2005)
7. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)

8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Transactions on Medical Imaging 27(11), 1668–1681 (2008)
9. Peters, J., Ecabert, O., Meyer, C., Kneser, R., Weese, J.: Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. Medical Image Analysis 14(1), 70–84 (2010)
10. Zhuang, X., Leung, K., Rhode, K., Razavi, R., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac mri. IEEE Transactions on Medical Imaging 29(9), 1612–1625 (2010)
11. Yang, L., Georgescu, B., Zheng, Y., Meer, P., Comaniciu, D.: 3D ultrasound tracking of the left ventricles using one-step forward prediction and data fusion of collaborative trackers. In: CVPR (2008)