A Dynamical Shape Prior for LV Segmentation from RT3D Echocardiography^{*}

Yun Zhu¹, Xenophon Papademetris^{1,2}, Albert J. Sinusas²,
and James S. Duncan^{1,2}

¹ Department of Biomedical Engineering, Yale University, USA
yun.zhu@yale.edu

² Department of Diagnostic Radiology, Yale University, USA

Abstract. Real-time three-dimensional (RT3D) echocardiography is the newest generation of three-dimensional (3-D) echocardiography. Segmentation of RT3D echocardiographic images is essential for determining many important diagnostic parameters. In cardiac imaging, since the heart is a moving organ, prior knowledge regarding its shape and motion patterns becomes an important component for the segmentation task. However, most previous cardiac models are either static models (SM), which neglect the temporal coherence of a cardiac sequence or generic dynamical models (GDM), which neglect the inter-subject variability of cardiac motion. In this paper, we present a subject-specific dynamical model (SSDM) which simultaneously handles inter-subject variability and cardiac dynamics (intra-subject variability). It can progressively predict the shape and motion patterns of a new sequence at the current frame based on the shapes observed in the past frames. The incorporation of this SSDM into the segmentation process is formulated in a recursive Bayesian framework. This results in a segmentation of each frame based on the intensity information of the current frame, as well as on the prediction from the previous frames. Quantitative results on 15 RT3D echocardiographic sequences show that automatic segmentation with SSDM is superior to that of either SM or GDM, and is comparable to manual segmentation.

1 Introduction

Real-time three-dimensional (RT3D) echocardiography is a new imaging modality that can capture the complex three-dimensional (3-D) shape and motion of the heart *in vivo*. To fully take advantage of the information offered by RT3D echocardiography, a robust and accurate automatic segmentation tool for tracking the dynamic shape of the heart is indispensable for quantitative analysis of cardiac function. Unfortunately, automatic segmentation of RT3D echocardiography is challenging and depends on image quality.

The use of shape and temporal priors has proven effective for segmenting images with missing and misleading image information [1]. The most widely used

^{*} This work is supported by the grant 5R01HL082640-03.

shape model is probably the Active Shape Model (ASM), which uses Principal Components Analysis (PCA) to describe the average shape and the most characteristic shape variations of a set of training shapes. However, ASM is a static model (SM) because it supplies a prior just for shape, but not for the motion of that shape. Temporal models can take a number of forms. The simplest form is one that insists on temporal coherence and smoothness. More complicated forms approximate cardiac motion using a parametric model. However, these temporal models do not include any prior knowledge of the shape.

To combine shape and temporal priors, researchers have proposed spatial-temporal statistical models. For example, Mitchell *et al.* extended ASM to the 2-D Active Appearance Motion Model (AAMM) [2], which includes both motion and appearance information. It is difficult, however, to extend 2-D AAMM to the segmentation of a full 3-D cardiac sequence because of the high dimensionality involved. Perperidis *et al.* constructed a 4-D atlas using two separate models, which accounted for inter-subject variability and cardiac temporal dynamics (intra-subject variability), respectively [3]. While these two models were successfully applied to the classification of cardiac images from normal volunteers and patients with hypertrophic cardiomyopathy, they are not related, making them suboptimal for left ventricular (LV) segmentation.

Dynamical shape model is a recently proposed spatial-temporal statistical model. It performs sequential segmentation using the cardiac dynamics learned from a set of training samples. It is highly flexible, and can be applied to segment a full 3-D sequence. For example, Jacob *et al.* proposed a second-order autoregressive model to approximate cardiac dynamics [4]. Sun *et al.* proposed learning cardiac dynamics using a second-order nonlinear model [5]. While these models are superior to SM, they are time homogeneous¹ and therefore inadequate for describing complex shape deformations, such as cardiac dynamics. In addition, since they supply a uniform model to all sequences, they ignore the subject variations in motion patterns. This makes them generic dynamical models (GDM).

In this paper, we present a subject-specific dynamical model (SSDM) to simultaneously account for the subject-specific variations in cardiac shape as well as inhomogeneous motion patterns. To build this SSDM, we need to differentiate two factors that cause cardiac shape variability. One is the inter-subject variability, and the other is temporal dynamics caused by cardiac deformation during a cardiac cycle, as shown in Figure 1. These two factors are interactive and cannot be separated into two independent statistical models. Because conventional PCA and Independent Component Analysis (ICA) can only focus on one factor at a time, we extend them to higher orders by utilizing Multilinear PCA (MPCA) [6] and Multilinear ICA (MICA) [7]. This allows us to decompose the training set and to describe the interaction of inter-subject variability and temporal dynamics. In addition, we design a dynamic prediction algorithm that can progressively

¹ A dynamical model is time-homogeneous if the conditional probability of state t given its previous states only depends on the time difference between those states, i.e. $\mathcal{P}(\mathbf{s}_t | \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots, \mathbf{s}_{t-m}) = \mathcal{P}(\mathbf{s}_{t+n} | \mathbf{s}_{t+n-1}, \mathbf{s}_{t+n-2}, \dots, \mathbf{s}_{t+n-m})$, for all n and m .

identify the subject vector associated with a new cardiac sequence and use this subject vector to predict the subject-specific dynamics from the segmentations observed in the previous frames. We formulate the integration of this SSDM into a recursive Bayesian framework. This framework models the evolution of the endocardial (ENDO) and epicardial (EPI) surfaces driven by both intensity information from the current frame as well as the dynamical shape prior inferred from the past segmentations, based on the knowledge learned from the training set.

2 Method

2.1 The Construction of SSDM

Shape Alignment. Magnetic Resonance (MR) images have higher a spatial resolution and signal-to-noise (SNR) ratio than RT3D echocardiography, and therefore are appropriate for building the SSDM. We acquired 32 sequences of electrocardiography (ECG)-gated canine short-axis MR images, with 16 temporal frames per sequence. The in-plane resolution was 1.6 mm, and the slice thickness was 5 mm. The ENDO and EPI surfaces were manually outlined by an experienced cardiologist using the BioImage Suite software [8]. We first extracted 153 landmarks on the ENDO surface and 109 landmarks on the EPI surface in the first frame of the first sequence. Then we propagated this set of landmarks to all frames in each sequence by mapping those frames to the first frame of the first sequence using inter- and intra-subject registrations, as shown in Figure 1. We used an affine transform to account for the global shape difference, as well as a shape-based non-rigid transform, as described in [9], to accommodate the detailed shape differences. Thus, we obtained 262 landmarks for each frame.

Shape Decomposition. In this paper, we use MPCA and MICA to decompose cardiac shapes (see [6,7] for an overview of MPCA and MICA). Here we denote the aligned cardiac shapes as third-order tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$, where $I = 32$ is the number of subjects, $J = 16$ is the number of frames within a sequence, and $K = 262 \times 3 = 786$ is the dimension of landmark vectors. By applying MPCA to tensor $\mathcal{S}$, we have

$$\mathcal{S} \approx \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} \quad (1)$$

where $\mathcal{Z} \in \mathbb{R}^{P \times Q \times R}$ is the core tensor, which represents the interaction of the subject, motion, and landmark subspaces. Matrices $\mathbf{U}^{\text{subject}} \in \mathbb{R}^{I \times P}$, $\mathbf{U}^{\text{motion}} \in \mathbb{R}^{J \times Q}$, and $\mathbf{U}^{\text{landmark}} \in \mathbb{R}^{K \times R}$ are the subject subspace, motion subspace, and landmark subspace respectively. Matrix $\mathbf{U}^{\text{subject}}$ contains row vectors $\mathbf{u}_i^{\text{subject}} \in \mathbb{R}^P$ ($1 \leq i \leq I$) of coefficients for each person i , and matrix $\mathbf{U}^{\text{motion}}$ contains row vectors $\mathbf{u}_j^{\text{motion}} \in \mathbb{R}^Q$ ($1 \leq j \leq J$) for frame j .

While it is reasonable to apply PCA in the subject subspace, it is inappropriate to use it in the motion subspace because the deformation of cardiac shapes does not have a Gaussian distribution. To handle this problem, we adopt ICA

in the motion subspace to obtain a set of independent modes in the motion subspace [10]. We rewrite Equation 1 as $\mathcal{S} \approx \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} = \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \mathbf{W}^T \mathbf{W}^{-T} \times_3 \mathbf{U}^{\text{landmark}} = \tilde{\mathcal{Z}} \times_1 \mathbf{U}^{\text{subject}} \times_2 \tilde{\mathbf{U}}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$, where the core tensor $\tilde{\mathcal{Z}} = \mathcal{Z} \times_2 \mathbf{W}^{-T}$, the column vectors of $\tilde{\mathbf{U}}^{\text{motion}}$ are independent components of the motion subspace $\tilde{\mathbf{U}}^{\text{motion}}$.

We have two steps to reduce dimensions. First, we select the complete eigenvectors in the motion subspace, i.e. $Q = J$, and perform MPCA in the subject and landmark subspace to find the optimal P and R such that the approximation keeps more than 98% of the original energy. Second, we fix P and R , and perform MICA in the motion subspace to find the modes that correspond to significant shape variations. In practice, we reduced $I = 32$ to $P = 5$ and $K = 786$ to $R = 11$ in the first step, and further reduced $J = 16$ to $Q = 3$ in the second step.

Dynamic Prediction. Given the segmentation of a new cardiac sequence from frame 1 to $t - 1$, we want to predict its segmentation in frame t . The idea is to first project the given segmentation from frame 1 to $t - 1$ to the subject subspace to identify the subject vector associated with this sequence, and then to use this subject vector to predict the LV shape at frame t .

Let $\mathbf{s}_{1:t-1} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{t-1}\}$ denote the observed segmentation of a new cardiac sequence. We predict the segmentation at frame t using two steps: projection and prediction. In the projection step, we estimate the subject vector associated with this new sequence as $\hat{\mathbf{u}}^{\text{subject}} = \mathbf{s}_{1:t-1} \cdot \mathbf{T}_{(1)}^{-1}$, where $\mathbf{T}_{(1)}$ is the mode-1 unfolding of tensor $\mathcal{T} = \tilde{\mathcal{Z}} \times_2 \tilde{\mathbf{u}}_{1:t-1}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$. In the prediction step, we use the subject vector estimated in the projection step to predict the segmentation at frame t as $\mathbf{s}_t^* = \tilde{\mathcal{Z}} \times_1 \hat{\mathbf{u}}^{\text{subject}} \times_2 \tilde{\mathbf{u}}_t^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$.

As mentioned above, we used MR sequences to build the SSDM, and then used this SSDM to predict the cardiac dynamics of a new RT3D echocardiographic sequence. However, this sequence may have a different length of cardiac cycle from the MR sequences. To handle this problem, we first align the end-diastolic

Fig. 1. The interaction of cardiac dynamics (intra-subject variability) and inter-subject variability

Fig. 2. Temporal interpolation to generate time frames from sequence 2 which correspond to those of sequence 1

and end-systolic frames using ECG signals, and then use linear interpolation to generate frames that correspond to the RT3D echocardiographic frames, as shown in Figure 2.

2.2 General Formulation

Assume that we are given a cardiac sequence $I_{1:t} = \{I_1, I_2, \dots, I_t\}$, and let $\mathbf{s}_t = \{\mathbf{s}_t^+, \mathbf{s}_t^-\}$ be the segmentation at frame t , where $\mathbf{s}_t^+$ is the ENDO surface, and $\mathbf{s}_t^-$ is the EPI surface. The problem of segmenting the current frame t can be addressed by maximizing the conditional probability

$$\begin{aligned} \mathcal{P}(\mathbf{s}_t|I_{1:t}) &\propto \mathcal{P}(I_t|\mathbf{s}_t, I_{1:t-1}) \mathcal{P}(\mathbf{s}_t|I_{1:t-1}) \\ &= \mathcal{P}(I_t|\mathbf{s}_t, I_{1:t-1}) \int \mathcal{P}(\mathbf{s}_t|\mathbf{s}_{1:t-1}) \mathcal{P}(\mathbf{s}_{1:t-1}|I_{1:t-1}) d\mathbf{s}_{1:t-1} \end{aligned} \quad (2)$$

In the following, we make two assumptions in order to lead to a computationally more feasible problem [11]. (1) The images $I_{1:t}$ are mutually independent, i.e. $\mathcal{P}(I_t|\mathbf{s}_t, I_{1:t-1}) = \mathcal{P}(I_t|\mathbf{s}_t)$. (2) The distributions of previous states are strongly peaked around the maxima of the respective distributions, i.e. $\mathcal{P}(\mathbf{s}_{1:t-1}|I_{1:t-1}) \approx \delta(\mathbf{s}_{1:t-1} - \hat{\mathbf{s}}_{1:t-1})$, where $\hat{\mathbf{s}}_i = \arg \max \mathcal{P}(\mathbf{s}_i|I_{1:i})$ and $\delta(\cdot)$ is Dirac delta function.

Thus, we have $\mathcal{P}(\mathbf{s}_t|I_{1:t}) \propto \underbrace{\mathcal{P}(I_t|\mathbf{s}_t)}_{\text{data adherence}} \underbrace{\mathcal{P}(\mathbf{s}_t|\hat{\mathbf{s}}_{1:t-1})}_{\text{dynamical shape prior}}$. It is a recursive

Bayesian formulation, where the ENDO and EPI contours are driven not only by the intensity information from the current frame, but also from the dynamical shape prior from the past frames.

2.3 Data Adherence

An entire cardiac image is partitioned by the ENDO and EPI contours into three regions: LV blood pool, LV myocardium, and background. The simplest intensity distribution for B-mode images is Rayleigh distribution [1]. However, Rayleigh distribution is only effective for fully-developed speckles. More complicated models, such as the Rice distribution and the K-distribution, have been proposed to account for the regular structure of scatters and low effective scatter density [1]. Unfortunately, the analytical complexity involved with these distributions is significant. In this paper, we utilize the Nakagami distribution [12], a simpler generalized distribution which can handle simultaneously the situations of regularly-spaced scatters and varying scatter densities. Thus, the intensity distribution for the LV blood pool and myocardium can be expressed as $\mathcal{P}_l(I; \mu_l, \omega_l) = \frac{2\mu_l^{\mu_l}}{\Gamma(\mu_l)\omega_l^{\mu_l}} I^{2\mu_l-1} \exp\left(-\frac{\mu_l}{\omega_l} I^2\right)$, where μ_l is the Nakagami parameter and ω_l is a scaling parameter. For $l = 1$, it models the intensity distribution in the LV blood pool. For $l = 2$, it models the intensity distribution in the LV myocardium.

Unlike the LV blood pool and myocardium, the background includes more than one tissue (e.g. RV blood pool, RV myocardium, and other tissues). Therefore, we use a mixture model and invoke the Expectation-Maximization (EM)

algorithm to fit the background histogram. Under the mixture model, the background distribution is given as $\mathcal{P}_3(I; \mu_3, \omega_3) = \sum_{k=1}^M \alpha_k \mathcal{P}_k(I; \mu_{3,k}, \omega_{3,k})$, where M is the number of components, α_k is the mixture proportion of component k that satisfies $\sum_{k=1}^M \alpha_k = 1$, $\mu_{3,k}$ and $\omega_{3,k}$ are the parameters of its component distributions. In the experiments, we set $M = 2$.

Let Ω_1 , Ω_2 , and Ω_3 denote three regions: LV blood pool, LV myocardium, and background, respectively. Then, the data adherence term can be defined as follows

$$\log \mathcal{P}(I|\mathbf{s}) = \sum_{l=1}^3 \int_{\Omega_l} \log \mathcal{P}_l(I; \mu_l, \omega_l) d\mathbf{x} \quad (3)$$

The maximization of Equation 3 can be interpreted as the propagation of $\mathbf{s} = \{\mathbf{s}^+, \mathbf{s}^-\}$ that maximizes the piecewise homogeneities.

2.4 Dynamical Shape Prior

As shown in Section 2.1, we predict the ENDO and EPI contours at frame t using the dynamic prediction algorithm. Thus, we define the dynamic prior as

$$\mathcal{P}(\mathbf{s}_t | \hat{\mathbf{s}}_{1:t-1}) \propto \exp \left\{ -\alpha \|\mathbf{s}_t - \mathbf{s}_t^*\|^2 / 2 \right\} \quad (4)$$

where α is a weighting parameter, and $\mathbf{s}_t^*$ is the predicted shape at frame t using the dynamic prediction algorithm described in Section 2.1. In the experiments, we found $1.5 \leq \alpha \leq 2.5$ is applicable to most of the data.

3 Results

Figure 3 represents the automatically segmented ENDO and EPI contours during ventricular systole. To further quantify the segmentation results, we asked two experts, blind to each other, to independently outline the ENDO and EPI contours of all of the frames of the image sequences. We then compared the manual results to the automatic results using three metrics: mean absolute distance (MAD), Hausdorff distance (HD), and the percentage of correctly segmented voxels (PTP). Let $A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$, $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\}$, we define $\text{MAD} = \frac{1}{2} \left\{ \frac{1}{n} \sum_{i=1}^n d(\mathbf{a}_i, B) + \frac{1}{m} \sum_{j=1}^m d(\mathbf{b}_j, A) \right\}$, $\text{HD} = \max \left(\max_i \{d(\mathbf{a}_i, B)\}, \max_j \{d(\mathbf{b}_j, A)\} \right)$, where $d(\mathbf{a}_i, B) = \min_j \|\mathbf{b}_j - \mathbf{a}_i\|$. Let Ω_a be the region enclosed by automatic segmentation, and Ω_m be the region enclosed by the manual segmentation, we define $\text{PTP} = \frac{\text{Volume}(\Omega_a \cap \Omega_m)}{\text{Volume}(\Omega_m)}$. While MAD represents the global disagreement between two contours, HD compares their local similarities.

Table 1 compares the segmentation results of ENDO and EPI contours from SM, GDM and SSDM. For ENDO contours, the automatic-manual MAD using

Fig. 3. The automatically segmented ENDO- and EPI contours during ventricular systole. Red: ENDO surface, Green: EPI surface.

Table 1. Comparison of automatic outline to two experts’ outline of ENDO and EPI boundaries

		MAD (mm)	HD (mm)	PTP (%)
ENDO	automatic-manual (SSDM)	1.41 ± 0.40	2.53 ± 0.75	95.9 ± 1.24
	automatic-manual (GDM)	1.52 ± 0.46	3.25 ± 0.98	94.8 ± 1.56
	automatic-manual (SM)	2.33 ± 0.67	4.31 ± 1.26	93.1 ± 1.51
	manual-manual	1.37 ± 0.36	2.38 ± 0.65	95.8 ± 1.48
EPI	automatic-manual (SSDM)	1.74 ± 0.39	2.79 ± 0.97	94.5 ± 1.74
	automatic-manual (GDM)	1.77 ± 0.41	2.91 ± 0.95	93.6 ± 1.78
	automatic-manual (SM)	1.81 ± 0.65	3.18 ± 1.23	92.3 ± 1.91
	manual-manual	1.73 ± 0.51	2.83 ± 1.50	94.5 ± 1.77

SM was 0.92 mm larger than that obtained using SSDM, although the automatic-manual MAD using GDM was similar to that obtained using SSDM. This implies that both SSDM and GDM are able to capture the global deformation of cardiac shapes, while SM has tendency to get stuck in local minima because it does not provides a prediction in time. We also observed that the automatic-manual HD using SSDM was 0.72 mm larger than that obtained using GDM, and 1.78 mm larger than that with SM. This suggests that while GDM produced globally correct results, it failed to capture local shape deformations. Moreover, we observed that the performance of the automatic segmentation using SSDM was comparable to that of manual segmentation because similar MAD, HD, and PTP were produced. For EPI contours, in comparison with SM, the GDM improved the MAD by 0.04 mm, the HD by 0.13mm, and the PTP by 1.3%. When the SSDM was applied, the MAD was further improved by 0.03 mm, the HD by 0.12 mm, and the PTP by 0.9%. The improvement of EPI segmentation was less pronounced than that of ENDO segmentation. This is because the EPI surface does not move as much as the ENDO surface. Furthermore, we observed that the variability of manual-manual segmentation was smaller for the ENDO boundary than for the EPI boundary. This is probably because the EPI boundaries are more ambiguous for observers to detect, which was also the reason we used two observers, instead of a single one.

4 Conclusion

In this paper, we have presented a subject-specific dynamical model that utilized MPCA and MICA to simultaneously decompose the cardiac shape and motion in different subspaces. We then used a dynamic prediction algorithm to sequentially predict the dynamics of a new RT3D echocardiographic sequence from the shapes observed in the past frames. Experiments on 15 sequences of echocardiographic data showed that automatic segmentation using the SSDM produced results that had an accuracy comparable to that obtained by manual segmentation. Future work would include the extension of this SSDM to human data and other modalities, and the development of an integrated framework that will combine cardiac segmentation and motion analysis for RT3D echocardiography.

References

1. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: A survey. *IEEE TMI* 25(8), 987–1010 (2006)
2. Bosch, J.G., Mitchell, S.C., Lelieveldt, B.P.F., Nijland, F., Kamp, O., Sonka, M., Reiber, J.H.C.: Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE TMI* 21(11), 1374–1383 (2002)
3. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 402–410. Springer, Heidelberg (2005)
4. Jacob, G., Noble, J.A., Behrenbruch, C., Kelion, A., Banning, A.: A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. *IEEE TMI* 21(3), 226–238 (2002)
5. Sun, W., Cetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.: Segmenting and tracking the left ventricle by learning the dynamics in cardiac images. In: *IPMI*, pp. 553–565 (2005)
6. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear analysis of image ensembles: TensorFaces. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2350, pp. 447–460. Springer, Heidelberg (2002)
7. Vasilescu, M., Terzopoulos, D.: Multilinear independent component analysis. In: *CVPR*, vol. 1, pp. 547–553 (2005)
8. Papademetris, X., Jackowski, M., Rajeevan, N., Okuda, H., Constable, R., Staib, L.: Bioimage suite: An integrated medical image analysis suite, section of bioimaging sciences, dept. of diagnostic radiology, yale university
9. Papademetris, X., Sinusas, A.J., Dione, D.P., Constable, R.T., Duncan, J.S.: Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *IEEE TMI* 21(7), 786–800 (2002)
10. Friedman, J., Hastie, T., Tibshirani, R.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York (2001)
11. Cremers, D.: Dynamical statistical shape priors for level set-based tracking. *IEEE TPAMI* 28(8), 1262–1273 (2006)
12. Shankar, P.M.: A general statistical model for ultrasonic backscattering from tissues. *UFFC* 47(3), 727–736 (2000)