Published in final edited form as:

IEEE Trans Med Imaging. 2010 March ; 29(3): 669–687. doi:10.1109/TMI.2009.2031063.

Segmentation of the Left Ventricle from Cardiac MR Images Using a Subject-Specific Dynamical Model

Yun Zhu[Student Member, IEEE] and Xenophon Papademetris[Member, IEEE]

Department of Biomedical Engineering, Yale University, New Haven, CT, 06520 USA

yun.zhu@yale.edu

Albert J. Sinusas

Department of Diagnostic Radiology with Yale University

James S. Duncan[Fellow, IEEE]

Department of Biomedical Engineering, Yale University, New Haven, CT, 06520 USA

Abstract

Statistical models have shown considerable promise as a basis for segmenting and interpreting cardiac images. While a variety of statistical models have been proposed to improve the segmentation results, most of them are either static models (SM), which neglect the temporal dynamics of a cardiac sequence or generic dynamical models (GDM), which are homogeneous in time and neglect the inter-subject variability in cardiac shape and deformation. In this paper, we develop a subject-specific dynamical model (SSDM) that simultaneously handles temporal dynamics (intra-subject variability) and inter-subject variability. We also propose a dynamic prediction algorithm that can progressively identify the specific motion patterns of a new cardiac sequence based on the shapes observed in past frames. The incorporation of this SSDM into the segmentation framework is formulated in a recursive Bayesian framework. It starts with a manual segmentation of the first frame, and then segments each frame according to intensity information from the current frame as well as the prediction from past frames. In addition, to reduce error propagation in sequential segmentation, we take into account the periodic nature of cardiac motion and perform segmentation in both forward and backward directions. We perform “Leave-one-out” test on 32 canine sequences and 22 human sequences, and compare the experimental results with those from SM, GDM, and Active Appearance Motion Model (AAMM). Quantitative analysis of the experimental results shows that SSDM outperforms SM, GDM, and AAMM by having better global and local consistencies with manual segmentation. Moreover, we compare the segmentation results from forward and forward-backward segmentation. Quantitative evaluation shows that forward-backward segmentation suppresses the propagation of segmentation errors.

Keywords

Cardiac Segmentation; Statistical Shape Model; Dynamical Model; Bayesian Method

I. Introduction

Segmentation of the left ventricle (LV) from cardiac magnetic resonance (MR) images is a prerequisite for further quantitative analysis of cardiac function, such as estimation of

ejection fraction and myocardial motion analysis. In clinical studies, the segmentation task, particularly the delineation of LV endocardium and epicardium, is often performed manually. This process, however, is tedious and time consuming. To expedite this process, an automatic and robust approach to segmenting cardiac MR sequences is highly desirable. Nonetheless, there are a number of challenges in automatic segmentation of the LV. First, the presence of noise and artifacts makes image information misleading physically corrupted, and sometimes incomplete. Second, the papillary muscles and other anatomy parts, such as the liver and LV/right ventricle (RV) junctures, pose additional challenges for the segmentation. Third, the low contrast between the myocardium and lung air makes segmentation of the epicardium especially difficult.

To overcome the misleading visual support, two related classes of models, i.e. shape and temporal models, have been proposed to regularize the segmentation process. The main idea behind a shape model is to establish, from a set of training shapes, patterns of variation in shapes and spatial relationships of structures for shapes. These patterns of variations thus offer a compact and appealing way to code prior knowledge about the shapes to be identified (see [13] for an overview). The most classical shape model is probably the Active Shape Model (ASM) [8], which applies principal components analysis (PCA) to a set of training shapes in order to extract a certain number of modes that represent shape variations. Cootes *et al.* generalized ASM to Active Appearance Model (AAM) by combining both shape and appearance variations into a unified model [9]. Mitchell *et al.* introduced AAM to medical image analysis by applying a multistage AAM to the segmentation of the LV and RV from cardiac images [21]. They later extended their work to segmenting three-dimensional (3-D) images [20]. However, both ASM and AAM are static models (SM), as they build a prior just for shape/appearance, but not for the motion of that shape.

While shape models provide information in the spatial domain, temporal models can provide knowledge in the time domain. A weak temporal constraint is to enforce the temporal consistency throughout the entire cycle. In its simplest form, this can be achieved using an average function $s_t = (s_t + s_{t-1}) / 2$, where s_t is the shape at time t [22]. A more complicated form is to add a penalty term enforcing the temporal consistency of image intensity. For example, Lorenzo-Valdes proposed encoding temporal coherence using a spatial-temporal Markov Random Field, whose initial parameters were estimated from a probabilistic atlas of temporally aligned sequences using Expectation-Maximization (EM) algorithm [18]. In contrast to weak constraints, which simply enforce temporal consistency, a strong temporal constraint models cardiac dynamics, such that it provides a prior knowledge about motion trajectories in the segmentation process. For example, Lynch *et al.* proposed approximating temporal cardiac movements by using a parametric model (Gaussian function), whose parameters were further refined as the parallel segmentation was being performed on 3-D volumes [19].

To unify shape and temporal information into a single model, some researchers proposed spatial-temporal shape models. For example, Bosch *et al.* extended two-dimensional (2-D) AAM to 2-D Active Appearance Motion Model (AAMM) by concatenating the shape points from all frames in a sequence into a single shape vector before PCA was applied [5]. The AAMM was then applied to match a whole 2-D image sequence, instead of a single image, in a way similar to conventional AAM. Perperidis *et al.* proposed a 4-D (3-D+ t) statistical model by building two separate models to account for the inter-subject variability and cardiac dynamics (intra-subject variability), respectively [25]. While these two models were successfully applied to the classification of cardiac images from normal volunteers and patients of hypertrophic cardiomyopathy, they are not related, making them inappropriate for LV segmentation.

One interesting type of spatio-temporal shape model is dynamical shape model, which sequentially predicts the shape of each frame based on the shapes observed in previous frames, using the temporal dynamics learnt from the training sequences. In the computer vision community, Blake *et al.* used an autoregressive model to track objects in a filtering framework [2]. Cremers *et al.* used a second-order autoregressive model to describe the silhouettes of a walking person, and then incorporated this autoregressive model into a level set-based segmentation framework as a shape prior [10]. Senegas *et al.* introduced dynamical shape model to the segmentation of cardiac images in order to model cardiac dynamics [28]. Sun *et al.* proposed learning the cardiac dynamics using a second-order nonlinear model [31]. While dynamical shape models are superior to SM, they are homogeneous in time¹ and are therefore insufficient to describe complex shape deformations, such as cardiac dynamics. Also, by applying a uniform model to all sequences, they ignore the variability of motion patterns between subjects. Therefore, they are generic dynamical models (GDM).

This paper is an extended version of our two conference papers [38], [37]. It presents a subject-specific dynamical model (SSDM) which simultaneously accounts for two factors that cause cardiac shape variability. One factor is inter-subject variability, and the other factor is the temporal dynamics caused by cardiac deformation during the cardiac cycle, as shown in Figure 1. These two factors are interactive and cannot be separated into two independent statistical models. Unfortunately, conventional PCA and independent component analysis (ICA) can only account for one factor at one time, and are therefore inappropriate for our dynamical model.

To address this issue, we extend conventional PCA and ICA to higher orders, and use multilinear PCA (MPCA) and multilinear ICA (MICA) [35], [36] to build a dynamical model that decomposes the training shapes in order to simultaneously take into account the temporal dynamics (intra-subject variability) and inter-subject variability. In addition, we introduce a dynamic prediction algorithm that can progressively identify the subject vector associated with a new cardiac sequence, and use this subject vector to predict subject-specific dynamics from shapes observed in previous frames.

The integration of this SSDM into the segmentation process is formulated in a recursive Bayesian framework for deformable model-based cardiac sequence segmentation. The evolution of the myocardial surfaces is driven both by the intensity information of the current frame as well as by the SSDM, which uses the dynamical prediction algorithm to predict the shape in the current frame based on the segmentation obtained in the preceding frames, as shown in Figure 2. To further reduce the propagation of segmentation errors, we utilize the periodicity property of cardiac dynamics, and perform bi-directional cardiac sequence segmentation. Starting from the first frame of a sequence, we simultaneously perform forward and backward segmentation, and combine them into a statistical framework, as shown in Figure 3.

The remainder of this paper is organized as follows. In Section II, we introduce the cardiac shape decomposition algorithm and dynamical shape prediction algorithm. In Section III, we introduce a Bayesian formulation for the cardiac sequence segmentation that integrates the intensity information and SSDM. In Section IV, we describe the data set, training process, and quantitative measures that are used to evaluate the segmentation results. In Section V, we provide experimental results using SSDM in forward and forward-backward directions. We also compare the segmentation results with those obtained with SM, GDM, and AAMM.

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

In Section VI, we provide a discussion of our approach. Finally, Section VII provides the conclusion to this paper. In Appendix A and B, we review the mathematical foundations of multilinear algebra, MPCA, and MICA. In Appendix C, we review the fundamentals of thin-plate spline (TPS) which is used in the segmentation process.

II. The Construction of SSDM

A. Shape Alignment

To build the SSDM, we need to find a set of correspondent landmarks defined over all the frames of each sequence with equal length of cycle. While this can be achieved in various ways [25], [16], we used a two-stage procedure in this work. In the first stage, we extracted a set of correspondent landmarks over all the frames of each sequence. However, the cycle length may differ from sequence to sequence. In the second stage, we used a temporal resampling and interpolation step to generate a set of correspondent landmarks with equal length of cycle.

In the first stage, we used an approach similar to the one in [14] to perform a three-step shape alignment and landmark extraction.

- 1) We used Electrocardiography (ECG) signals to align the end-diastole (ED) frames of each sequence, and used surface-based inter-subject and intra-subject registrations to align all frames to the first frame of the first sequence, as shown in Figure 4. Specifically, we used an affine transform to account for the global shape difference, followed by a shape-based non-rigid registration, which minimizes the differences in principal curvatures, to accommodate detailed shape differences [29], [24]. The accuracy of shape-based non-rigid transform was validated using implanted markers in [24]. After this step, we obtained dense surface correspondence over the training shapes;
- 2) We asked our cardiology collaborator to manually select a set of representative landmarks on the endocardial (ENDO) and epicardial (EPI) surfaces of the first frame of the first sequence. To faithfully represent the shapes of the ENDO and EPI surfaces, we placed more landmarks on highly-curved regions and fewer landmarks on smooth regions. Also, we progressively refined the selection of landmarks by reconstructing the cardiac shape and comparing it with the original shape until fully satisfied;
- 3) We propagated this set of landmarks from the first frame of the first sequence to all frames of each sequence using the dense intra- and inter-subject surface correspondence found in the first step.

The whole shape alignment and landmark extraction process is illustrated in Figure 5.

After the shape alignment and landmark extraction process, we obtained a set of correspondent landmarks defined over all frames of each sequence. However, the cycle length may differ from sequence to sequence. To generate a set of correspondent landmarks with equal cycle length, we used a temporal resampling and interpolation step, as described below.

- 1) We selected a reference sequence (i.e. the first sequence in the data set);
- 2) We used ECG signals, which have now become standard in cardiac imaging, to align the ED frames of all sequences;
- 3) We used linear interpolation to generate frames from other sequences that correspond to the reference frame, as illustrated in Figure 6.

After the temporal resampling and interpolation process, we obtained a set of correspondent landmarks defined over all frames of each sequence, with equal cycle length.

B. Shape Decomposition

In the following discussion, we use MPCA and MICA (see Appendix A for basic mathematics of multilinear algebra and Appendix B for fundamentals of MPCA and MICA) to perform shape decomposition. We use a third-order tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$ to represent the aligned shapes (represented as a set of landmarks), where I is number of subjects, J is the number of time frames within a sequence, and K is the dimension of the landmark vectors. By applying MPCA to tensor $\mathcal{S}$, we obtain

$$\mathcal{S} \simeq \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$$

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

While it reasonable to apply PCA in the subject subspace, it is not appropriate to use it in the motion subspace, because the deformation of cardiac shapes does not have a Gaussian distribution. Decomposition of the motion subspace using PCA would result in a set of mixed modes, as shown in Figure 7. To handle this problem, we adopt ICA in the motion subspace to obtain a set of independent modes in the motion subspace.

$$\begin{aligned} \mathcal{S} &\simeq \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} \\ &= \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \mathbf{W}^T \mathbf{W}^{-T} \times_3 \mathbf{U}^{\text{landmark}} \\ &= \tilde{\mathcal{Z}} \times_1 \mathbf{U}^{\text{subject}} \times_2 \tilde{\mathbf{U}}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} \end{aligned} \quad (1)$$

where the core tensor $\tilde{\mathcal{Z}} = \mathcal{Z} \times_2 \mathbf{W}^{-T}$, and the column vector of $\tilde{\mathbf{U}}^{\text{motion}}$ are the independent components of the motion subspace $\tilde{\mathbf{U}}^{\text{motion}}$.

To reduce dimensions, we 1) select the complete eigenvectors in the motion subspace, i.e. $J' = J$, and perform MPCA in the subject and landmark subspaces to find the optimal I' and K' such that the approximation keeps more than 98% of the original energy; and 2) fix I' and K' , and perform MICA in the motion subspace and chose J' according to the following criteria:

- 1) All selected modes correspond to significant variations;
- 2) After decomposition, more than 96% of the original energy is retained.

When MICA is used in the second step, we have three significant modes of shape variations: short-axial contraction, twisting of the heart, and long-axial contraction, as shown in Figure 8.

C. Dynamic Prediction

In Section II-B, we decomposed the training shapes into the subject subspace $\mathbf{U}^{\text{subject}}$, motion subspace $\tilde{\mathbf{U}}^{\text{motion}}$, and landmark subspace $\mathbf{U}^{\text{landmark}}$. Thus, we can represent a new cardiac shape using

$$s = \tilde{Z} \times_1 \mathbf{u}^{subject} \times_2 \tilde{u}^{motion} \times_3 \mathbf{U}^{landmark}$$

where $\mathbf{u}^{subject}$ is the subject vector and $\tilde{u}^{motion}$ is the motion vector. Given the shapes of a new sequence from frame 1 to $t-1$, we want to predict its shape at frame t . The idea is to first project the given shapes to the subject subspace to identify the subject vector associated with this sequence, and then use this subject vector to reconstruct the LV shape in frame t .

Let $s_{1:t-1} = \{s_1, s_2, \dots, s_{t-1}\}$ denote the observed shapes of a new sequence, we can predict its shape at frame t as follows:

- **Projection** The subject vector associated with this sequence can be represented as

$$\hat{u}^{subject} = \mathcal{B} \times_1 s_{1:t-1} \quad (2)$$

where the “projection tensor” $\mathcal{B}$ is obtained by retensorizing the matrix $\mathbf{B}_{(1)} = \mathbf{T}_{(1)}^{-T}$. (The matrix $\mathbf{T}_{(1)}$ is the 1-mode unfolding of tensor $\mathcal{T} = \tilde{\mathcal{Z}} \times_2 \tilde{u}^{motion} \times_3 \mathbf{U}^{landmark}$).

- **Prediction** With this subject vector, we can predict the shape in frame t as

$$s_t^* = \tilde{\mathcal{Z}} \times_1 \hat{u}^{subject} \times_2 \tilde{u}_t^{motion} \times_3 \mathbf{U}^{landmark} \quad (3)$$

III. Cardiac Segmentation

A. General Framework

In this section, we first describe a recursive Bayesian framework that uses SSDM to segment a cardiac sequence in forward direction. Then, we discuss the segmentation in a backward direction. Finally, we combine the forward and backward segmentation into a unified framework.

Assume that we are given a cardiac sequence $I_{1:t} = \{I_1, I_2, \dots, I_t\}$, where I_t is the image at frame t . Also let s_t be the cardiac shape at frame t . Using Bayesian formula, the problem of segmenting the current frame t can be addressed by maximizing the conditional probability

$$\mathcal{P}(s_t | I_{1:t}) \propto \mathcal{P}(I_t | s_t, I_{1:t-1}) \mathcal{P}(s_t | I_{1:t-1}) \quad (4)$$

We further represent the second term on the right-hand side of Equation 4 as a maximization of the posterior over all states up to time $t-1$ given all image frames up to frame $t-1$:

$$\begin{aligned} \mathcal{P}(s_t | I_{1:t-1}) &= \int \mathcal{P}(s_t, s_{1:t-1} | I_{1:t-1}) ds_{1:t-1} \\ &= \int \mathcal{P}(s_t | s_{1:t-1}) \mathcal{P}(s_{1:t-1} | I_{1:t-1}) ds_{1:t-1} \end{aligned} \quad (5)$$

Combining Equations 4 and 5, we have

$$\begin{aligned} \mathcal{P}(s_t | I_{1:t}) &\propto \mathcal{P}(I_t | s_t, I_{1:t-1}) \int \mathcal{P}(s_t | s_{1:t-1}) \\ &\quad \mathcal{P}(s_{1:t-1} | I_{1:t-1}) ds_{1:t-1} \\ &\stackrel{(a)}{=} \underbrace{\mathcal{P}(I_t | s_t)}_{\text{data adherence}} \underbrace{\mathcal{P}(s_t | s_{1:t-1})}_{\text{forward dynamics}} \end{aligned} \quad (6)$$

>

As explained in [10], we make two assumptions in step (a), which lead to a computationally more feasible problem:

- We assume that $I_{1:t}$ are mutually independent, i.e. $\mathcal{P}(I_t|s_t, I_{1:t-1}) = \mathcal{P}(I_t|s_t)$;
- We assume that the distribution of previous states is strongly peaked around the maxima of the respective distributions, i.e. $\mathcal{P}(s_{1:t-1}|I_{1:t-1}) \simeq \delta(s_{1:t-1} - \widehat{s}_{1:t-1}^+)$, where $\widehat{s}_i^+ = \arg \max \mathcal{P}(s_i|I_{1:i})$ are the estimates of myocardial contours in the past, and $\delta(\cdot)$ denotes the Dirac Delta function.

Thus, Equation 6 defines a recursive Bayesian formula to implement sequential cardiac segmentation in which the segmentation of each particular time frame is based not only on the image data at that instant, but also on predictions from previous frames.

Similarly, we can express the backward segmentation as the marginalization of the posterior of all states from $t+1$ to N , given the image frames from $t+1$ to N :

$$\mathcal{P}(s_t|I_{t+1:N}) = \int \mathcal{P}(s_t|s_{t+1:N}) \mathcal{P}(s_{t+1:N}|I_{t+1:N}) ds_{t+1:N} \quad (7)$$

Similar to the forward segmentation, we assume that $\mathcal{P}(s_{t+1:N}|I_{t+1:N}) \simeq \delta(s_{t+1:N} - \widehat{s}_{t+1:N}^-)$ where $\widehat{s}_i^- = \arg \max \mathcal{P}(s_i|I_{i:N})$ are the segmentations of previous frames in a backward direction. Thus, we can rewrite Equation 7 as follows

$$\mathcal{P}(s_t|I_{1:N}) = \mathcal{P}(s_t|\widehat{s}_{t+1:N}^-) \quad (8)$$

Now assume that we are given the entire cardiac sequence $I_{1:N}$, the problem of segmenting frame t can be formulated as maximizing the conditional probability $\mathcal{P}(s_t|I_{1:N})$. As explained in [12], the conditional probability can be rewritten as

$$\mathcal{P}(s_t|I_{1:N}) \propto \mathcal{P}(s_t|I_{1:t}) \mathcal{P}(s_t|I_{t+1:N}) \quad (9)$$

By combining Equations 6, 8, and 9, we arrive at the following conditional probability for forward-backward segmentation

$$\begin{aligned} \widehat{s}_t &= \arg \max_{s_t} \mathcal{P}(s_t|I_{1:N}) \\ &= \arg \max_{s_t} \mathcal{P}(I_t|s_t) \mathcal{P}(s_t|\widehat{s}_{1:t-1}^+) \mathcal{P}(s_t|\widehat{s}_{t+1:N}^-) \\ &= \arg \max_{s_t} \left\{ \underbrace{\log \mathcal{P}(I_t|s_t)}_{\text{data adherence}} + \underbrace{\log \mathcal{P}(s_t|\widehat{s}_{1:t-1}^+)}_{\text{forward dynamics}} + \underbrace{\log \mathcal{P}(s_t|\widehat{s}_{t+1:N}^-)}_{\text{backward dynamics}} \right\} \end{aligned} \quad (10)$$

Equation 10 defines a forward-backward framework that starts from the manual segmentation in the first frame and uses shape observed in the previous frames to predict the shape in the current frame.

B. Data Adherence

In this paper, we take advantage of the statistical intensity distribution of cardiac MR images and adopt a region-based segmentation strategy, which is robust for images with weak

boundaries [7]. The main idea behind region-based deformable model is to partition an image into several regions with visually consistent and mutually distinct intensity distributions. In this paper, we use a three-phase region-based deformable model, which partitions an entire cardiac image into: LV blood pool (inside the ENDO contour), LV myocardium (between the ENDO and EPI contours), and background (outside the EPI contour).

Let $\mathcal{P}_1(I)$, $\mathcal{P}_2(I)$, and $\mathcal{P}_3(I)$ denote the intensity distributions of the LV blood pool, LV myocardium, and background, respectively. The LV blood pool and myocardium are homogeneous, and can therefore be modeled with a single probability density function (pdf). The most common distribution for MR images is Gaussian distribution

$$\begin{aligned}\mathcal{P}_1(I) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(I-\mu_1)^2}{2\sigma_1^2}\right\} \text{ LV blood pool} \\ \mathcal{P}_2(I) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(I-\mu_2)^2}{2\sigma_2^2}\right\} \text{ LV myocardium}\end{aligned}$$

where μ_1 and μ_2 are the means of the Gaussian distribution, and σ_1 and σ_2 are their deviations.

Unlike the LV blood pool and myocardium, the background is inhomogeneous, because it includes more than one tissue (RV blood pool, RV myocardium, lung air, and etc). Therefore, it is insufficient to model it with a single distribution. Here we use a mixture model and invoke EM algorithm to fit the background histogram. The pdf-fitted histograms of each region are shown in Figure 9.

$$\mathcal{P}_3(I) = \sum_{k=1}^M \alpha_k \mathcal{G}_{3,k}(I; \mu_{3,k}, \sigma_{3,k})$$

where $\mathcal{G}_{3,k}(I; \mu_{3,k}, \sigma_{3,k})$ is the component Gaussian distribution, M is the number of components, and α_k is the mixture proportion of component k that satisfies $\sum_{k=1}^M \alpha_k = 1$.

Let Ω_1 , Ω_2 , and Ω_3 denote LV blood pool, LV myocardium, and background, respectively. Also, let $\mathcal{L}(x;s)$ be a TPS warping (an overview of TPS is given in Appendix C) which is controlled by the boundary landmarks of the images. We can define the segmentation by finding a smooth TPS warping (i.e. finding an optimal set of s) such that the resulting projection maximizes the piecewise homogeneities of cardiac images [32]. Mathematically, it can be written as follows:

$$\log \mathcal{P}(I|s) = \sum_{l=1}^3 \int_{\Omega_l} \log \mathcal{P}_l(I(\mathcal{L}(x;s))) dx \quad (11)$$

where s is a set of boundary landmarks, and $\mathcal{L}(x;s)$ is the TPS warping controlled by landmarks s . The maximization of Equation 11 can be achieved by finding an optimal location s that maximizes the piecewise homogeneities of a cardiac image.

C. Dynamic Prior

As shown in Section II-C, we can predict the shape at frame t using Equations 2 and 3. Thus, we define the dynamic prior in forward and backward directions as follows:

$$\mathcal{P}(s_t | \widetilde{s}_{1:t-1}^+) \propto \exp \left\{ \alpha^+ \| s_t - s_t^{*+} \|^2 / 2 \right\} \quad (12)$$

$$\mathcal{P}(s_t | \widetilde{s}_{t+1:N}^-) \propto \exp \left\{ \alpha^- \| s_t - s_t^{*-} \|^2 / 2 \right\} \quad (13)$$

where α^+ and α^- are weighting parameters, and s_t^{*+} and s_t^{*-} are the predicted shapes in forward and backward directions from the dynamic prediction algorithm described in Section II-C. We found that $0.5 \leq \alpha^+, \alpha^- \leq 2$ is applicable to most of the data sets.

D. Optimization and Implementation

The maximization of Equation 10 can be identified by taking the gradient descent leading to the following differential equation:

$$\frac{ds_t}{d\tau} = \frac{\partial \log \mathcal{P}(I_t | s_t)}{\partial s_t} + \frac{\partial \log \mathcal{P}(s_t | \widetilde{s}_{1:t-1}^+)}{\partial s_t} + \frac{\partial \log \mathcal{P}(s_t | \widetilde{s}_{t+1:N}^-)}{\partial s_t}$$

where $\mathcal{T}$ is the numerical step. From Equations 11, 12, and 13, we have

$$\begin{aligned} \frac{\partial \log \mathcal{P}(I_t | s_t)}{\partial s_t} &= \sum_{l=1}^3 \int_{\Omega_l} \frac{\partial \log(\mathcal{P}_l \circ I)}{\partial x'} \frac{\partial \mathcal{L}}{\partial s_t} dx \\ \frac{\partial \log \mathcal{P}(s_t | \widetilde{s}_{1:t-1}^+)}{\partial s_t} &= -\alpha^+ (s_t - s_t^{*+}) \\ \frac{\partial \log \mathcal{P}(s_t | \widetilde{s}_{t+1:N}^-)}{\partial s_t} &= -\alpha^- (s_t - s_t^{*-}) \end{aligned} \quad (14)$$

where $x' = \mathcal{L}(x; s_t)$. Due to the linear relationship between transformed coordinate x' and landmarks s_t , $\frac{\partial \mathcal{L}(x; s_t)}{\partial s_t}$ (i.e. $\frac{\partial x'}{\partial s_t}$) can be computed very efficiently, as shown in Appendix C.

Also, TPS offers great flexibility in image warping, and thus needs to be taken care of to avoid unrealistic warpings. In this paper, we use stiff-to-flexible strategy suggested in [17] by starting at the first iteration with more rigid warping and progressively reducing the rigidity of warping in subsequent iterations to accommodate more detailed non-rigid transforms. The details of how to control the rigidity of TPS warpings are explained in Appendix C.

An appropriate selection of step size for gradient descent method is of great importance. The values of the objective function might start to oscillate (i.e. not monotonically decrease) when the search is close to the minimum. This is caused by the constant step size in the gradient descent method. To address this issue, we use the Armijo rule [1] to adaptively select a step size at each iteration. In practice, we found that the oscillation problem was largely eliminated when the Armijo rule was applied.

In forward segmentation, errors might accumulate in the forward direction, i.e. the segmentation at the beginning of the cycle is likely to be more accurate than the segmentations at the end of the cycle. Therefore, it makes sense to give α^+ more weight at the beginning of the cycle, and less weight at the end of the cycle. In practice, we reduced α^+ linearly from 0.8 to 0.6 from the first to the last frame in the cardiac cycle. Similarly, we reduced α^- linearly from 0.8 to 0.6 in a backward direction.

The whole segmentation scheme is initialized with the manual segmentation in the first frame, and then uses the observed shapes from the previous frames to predict the shape in

the current frame. Also, we use the results from previous frames as an initialization for the current frame. Due to the similarity between two adjacent frames, it normally takes less than 30 seconds to segment a 3-D volume on a 2.8 GHz Pentium D workstation with a 2.0 gigabyte memory.

IV. Experimental Setup

A. Data

Canine—ECG-gated, breath-hold cine MR imaging on canine hearts was performed on a GE signa 1.5-T scanner. Short-axis images through the LV were obtained with gradient echo technique. A total of 16 contiguous 5-mm-thick slices were collected per frame. This sequence provides images with an in-plane resolution of 1.6×1.6 mm for a 256×256 matrix and a 5-mm resolution perpendicular to the imaging plane. This sequence also provides a temporal resolution of 16 frames/cardiac cycle, which is equivalent to 40 ms/frame.

We acquired a total of 32 sequences of normal canine MR images, each with 16 temporal frames per cycle. We asked our cardiology collaborator to manually outline the ENDO and EPI surfaces using the 4-D surface Editor in the Bioimage Suite software [23]. We then used surface triangulation method to represent the ENDO and EPI surfaces [24].

Human—A group of 22 healthy volunteers were examined on a 1.5-T scanner with a phased array coil. The entire heart was imaged in the short-axis orientation with ECG-triggered breath-hold steady-state free precession (SSFP) sequence. This sequence provides images with an in-plane resolution of 1.7 mm and a 6-mm resolution perpendicular to the imaging plane. The resulting 3-D image set consists of 22–25 temporal frames per cardiac cycle, which is equivalent to around 40 ms/frame.

As was done for the canine data, we manually outlined the ENDO and EPI surfaces, and represented these using triangulation method.

B. Training

We built two separate SSDMs: one for canine data and the other for human data. To maximize the effective size of the training set, we utilized leave-one-out approach which alternatively leaves out one sequence for validation and uses the rest of the sequences to build the SSDM.

Canine—We used 31 sequences to build the SSDM. In the shape alignment process (as described in Section II-A), we extracted 153 landmarks on the ENDO surface and 109 landmarks on the EPI surface of the reference frame (i.e. the first frame of the first sequence), and then propagated this set of landmarks to all frames of each sequence using the dense displacement obtained from the intra- and inter-subject registrations. We ultimately obtained 262 landmarks for all frames of each sequence. As the canine sequences had equal cycle length, we did not perform temporal resampling and interpolation for the canine data. In the shape decomposition procedure, we used a third-order tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$ to represent the aligned shapes, where $I=31$ is the number of subjects, $J=16$ is the number of time frames within a sequence, and $K=263 \times 3=786$ is the dimension of landmark vectors. In the two-step dimension reduction procedure (as described in Section II-B), we reduced $I=31$ to $I'=5$ and $K=786$ to $K'=11$ in the first step, and further reduced $J=16$ to $J'=3$ in the second step. We kept around 98% of the original energy in the first step and 97.1% of the original energy in the second step.

Human—We used 21 sequences to build the SSDM. In the shape alignment process, we extracted 165 landmarks on the ENDO surface and 127 landmarks on the EPI surface of the reference frame (i.e. the first frame of the first sequence), and propagated this set of landmarks to all the frames of each sequence. Unlike the canine data, however, the cycle length of human sequences varied from subject to subject. Therefore, we used an additional temporal resampling and interpolation step to generate sequences with 22 frames per cycle. We then used a third-order tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$ to represent the aligned shapes, where $I = 21$ is the number of subjects, $J = 22$ is the number of time frames within each sequence, and $K = 292 \times 3 = 876$ is the dimension of landmark vector. In the dimension reduction step, we reduced $I = 21$ to $I' = 4$, and $K = 876$ to $K' = 17$ in the first step, and further reduced $J = 16$ to $J' = 3$ in the second step. We kept around 98.3% of the original energy in the first step and 97.4% of the original energy in the second step.

C. Quantitative Validation

To quantify the accuracy of our approach, we used two distance error metrics, namely mean absolute distance (MAD) and Hausdorff distance (HD), and two area error metrics, namely percentage of true positives (PTP) and percentage of false positives (PFP).

Let A and B be two surfaces to be compared, and supposing that they are represented as point sets, i.e. $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_m\}$, we can define MAD and HD as follows

$$\begin{aligned} \text{MAD}(A, B) &= \frac{1}{2} \left\{ \frac{1}{n} \sum_{i=1}^n d(a_i, B) + \frac{1}{m} \sum_{j=1}^m d(b_j, A) \right\} \\ \text{HD}(A, B) &= \max \left\{ \max_i d(a_i, B), \max_j d(b_j, A) \right\} \end{aligned}$$

where $d(a_i, B) = \min_j |b_j - a_i|$. While MAD is a global measure of the match of two surfaces, the HD reflects their local similarities.

Let Ω_A and Ω_B be two regions enclosed by surfaces A and B , respectively, we can define PTP and PFP as follows

$$\begin{aligned} \text{PTP} &= \frac{\text{Volume}(\Omega_A \cap \Omega_B)}{\text{Volume}(\Omega_B)} \\ \text{PFP} &= \frac{\text{Volume}(\Omega_B) - \text{Volume}(\Omega_A \cap \Omega_B)}{\text{Volume}(\Omega_A)} \end{aligned}$$

V. Experimental Results

In this section, we present the segmentation results from our segmentation algorithm. In Section V-A, we compare forward segmentation results using subject-specific dynamical model (SSDM) with those using static model (SM), generic dynamical model (GDM), and Active Appearance Motion Model (AAMM). In Section V-B, we compare the results from forward segmentation with those from forward-backward segmentation. Finally in Section V-C, we investigate the sensitivity of SSDM to the manual segmentation in the first frame.

A. SSDM versus SM, GDM, and AAMM

In the following, we compare the forward segmentation results using SSDM against those using SM, GDM, and AAMM. The SM was built by performing conventional PCA (regardless of their temporal orders) on the training shapes. Unlike the SM, GDM took into

account temporal shape dynamics by recursively estimating the shape at the current frame based on the shapes observed in past frames [31]. Unlike GDM which performed sequential segmentation, the AAMM concatenated the shape/appearance vectors from all time frames in the order of their frame numbers, and treated them as a single shape/appearance vector [5]. Conventional PCA was then performed on the set of training sequences to generate mean and eigenvectors associated with the shape and appearance variations over the complete cycle.

Figure 10 compares the segmentation results of the ENDO surface using SM, GDM, AAMM, and SSDM. The yellow arrows in the first row of Figure 10 indicate the imaging artifacts caused by an implanted marker in the cardiac image. While implanted markers were used as a golden standard for LV motion analysis [24], they were undesirable for the segmentation purposes because they may cause imaging artifacts leading to local minima in the optimization process. We observed, as seen in the first row of Figure 10, that SM had a tendency to be stuck in the local minimum because it did not exploit temporal shape information. In comparison with SM, GDM had a better performance by offering globally correct segmentation. Also, GDM successfully overcame the local minimum caused by the implanted marker. However, GDM is inaccurate in describing local shape variations. The second row of Figure 10 shows the performance of GDM in the presence of papillary muscles (see white arrows), which have significant geometric features. Since papillary muscles have similar intensities as myocardium, it is always controversial whether to include them in the myocardium or in the blood pool. The answer to this question depends on the specific application. While some researchers have taken papillary muscles to be part of the blood pool, we include them in the myocardium because their local geometric properties are important for shape-based cardiac motion tracking algorithm [24]. We observed from the second row of Figure 10 that the segmented ENDO contours using GDM were over-smoothed, and failed to describe the curvedness of the papillary muscle. The third row of Figure 10 shows the segmentation results using AAMM. Similar to GDM, AAMM provides a globally correct and generally consistent segmentation during cardiac systole, although it was slightly less accurate at end-systole (ES). The last row of Figure 10 shows the segmentation results using SSDM. We observed that SSDM correctly described the local shape variations of papillary muscles, and therefore was more feasible for some applications, such as shape-based motion tracking.

Compared with the ENDO surfaces, the EPI surfaces are more difficult to detect due to the low intensity contrast between the myocardium and lung air, as shown in Figure 11. In Figure 11, we compare the segmentation of EPI surfaces without shape prior, using SM, GDM, AAMM, or SSDM. We observed that when no shape prior was added, the evolution of the EPI surface encountered a leakage problem. The lateral portion of the EPI contour evolved out to segment the lung air because the edge information between the myocardium and lung air was too weak. Figure 11 shows that SM, GDM, AAMM, and SSDM all correctly segmented the lateral portion of the EPI contour and that their differences were not significant. This is probably because the EPI surface does not move as significantly as the ENDO surface.

Another challenge in EPI segmentation is the presence of the LV/RV juncture. The LV and RV myocardium share similar intensity values and there is no clear edge at their junctures. We observed that the algorithm without shape prior produced “leaked” results (see blue arrow in Figure 11 (a)) while the algorithm using SM slightly overestimated the EPI contour at the LV/RV juncture (see yellow arrows in Figure 11 (b)). The GDM, AAMM, and SSDM produced more accurate results because all of them correctly separated the myocardium from the background and none overestimated the EPI contours at the LV/RV junctures, as shown in Figure 11 (c), (d), and (e).

Tables I and II use MAD, HD, PTP, and PFP to quantify the ENDO segmentation errors for the canine and human hearts, respectively. We observed that when SM was applied, the MADs were 2.31 mm and 2.15 mm for the canine and human hearts, respectively. When GDM was applied, they decreased by 1.21 mm and 0.94 mm, respectively. Since MAD reflects the global similarity, the reduction in MAD implies that SM was biased in the global sense. When SSDM was applied, the MAD decreased somewhat further, by 0.46 mm and 0.52 mm for the canine and human hearts, respectively. The HD decreased by 1.99 mm and 2.01 mm for the canine and human hearts, respectively. This implies that SSDM had a better local segmentation accuracy than the GDM.

We also observed that AAMM yielded smaller MAD than the SM (AAMM: MAD= 1.23 mm and 1.42 mm for the canine and human hearts, respectively; SM: MAD= 2.31 mm and 2.15 mm for the canine and human hearts, respectively). This is because AAMM takes into account the temporal behavior of the heart, while SM does not. In comparison with SSDM, the AAMM yielded generally comparable results, although the MAD from AAMM was slightly larger. This implies that the results from AAMM were slightly inconsistent with cardiac dynamics. A possible explanation for this could be that while AAMM models cardiac dynamics by concatenating the shape/appearance vectors from all time frames into a single vector, this modeling is not explicit because after PCA decomposition, there is no eigenvector associated with temporal motion of the heart. In other words, AAMM does not perform dimension reduction in the temporal domain to reduce correlation within a cycle. This often leads to the high-dimensionality of AAMM, causing it to require a large training set to effectively cover the variations in cardiac motion patterns. In contrast, SSDM performs MPCA and MICA in both subject subspace and motion subspace, and therefore provides a more compact representation of cardiac dynamics. For the canine hearts, when SSDM was applied, the MAD decreased by 0.59 mm, the HD decreased by 2.05 mm, the PTP increased by 2.2%, and the PFP decreased by 2.09%. For human data, the MAD decreased by 0.73 mm, the HD decreased by 1.96 mm, the PTP increased by 3.5%, and the PFP decreased by 2.91%.

Tables III and IV quantify the EPI segmentation errors for the canine and human hearts, respectively. When SM was applied, the MADs were 3.21 mm and 3.24 mm for the canine and human hearts, respectively. They decreased by 1.90 mm and 1.84 mm, respectively, when GDM was applied. When SSDM was applied, they decreased somewhat further by 0.08 mm and 0.13 mm for the canine and human hearts, respectively, while the HD decreased by 0.97 mm and 0.95 mm, respectively. The SSDM improved ENDO segmentation more than EPI segmentation because the EPI surface was smoother than the ENDO surface. Also, AAMM segmented the EPI contours with similar accuracy to the SSDM, although the MAD from SSDM was slightly smaller. Specifically, for the canine hearts, when SSDM was applied, the MAD decreased by 0.18 mm, HD decreased by 1.29 mm, the PTP increased by 2.3%, and the PFP decreased by 2.2%. For the human hearts, the MAD decreased by 0.24 mm, the HD decreased by 1.28 mm, the PTP increased by 2.8%, and the PFP decreased by 3.0%.

To further investigate the distribution of local geometric errors, we used a 17-segment LV model [6], as shown in Figure 12. In this model, the LV is divided into equal thirds perpendicular to the long axis of the heart generating basal, mid-cavity, and apical slices. The basal and mid-cavity slices are then divided into six segments, according to the anterior and posterior inter-ventricular grooves serving as landmarks for determining these segments. Similarly, the apical slices are partitioned into four segments, including the septal sector and three equal subsectors for the lateral sector.

We computed the MADs of the ENDO and EPI surfaces for each segment, and these MADs reflected the local geometric errors for those segments (see Tables V and VI). We observed that when GDM was applied, the ENDO MADs of the mid-cavity anterolateral, inferolateral, and inferior segments were significantly larger (2.01 mm, 1.87 mm, and 1.91 mm, respectively) than those of other segments. This was due to the existence of papillary muscles in the mid-cavity regions, which were over-smoothed by GDM. When AAMM was applied, the ENDO MADs were slightly larger than those of the SSDM, indicating that the results from AAMM were slightly inconsistent with cardiac dynamics. Also, we observed that while the segments with papillary muscles still had larger errors than other segments, the difference was not as significant as with the GDM. This suggests that AAMM captures local shape better than the GDM. When SSDM was applied, the MADs of mid-cavity anterolateral, inferolateral, and inferior segments were 1.12 mm, 1.14, and 1.19 mm smaller than those from GDM, and 0.86 mm, 1.11 mm, and 1.05 mm smaller than those from AAMM. For the EPI contours, we did not observe significant differences among the results from GDM, AAMM, and SSDM, although the local MADs from SSDM were slightly smaller. This was because the EPI surface was smoother than the ENDO surface. Moreover, we found that the errors from free wall (anterior, inferior, and lateral segments) were slightly larger than those for the septum. This was probably due to the low contrast between the myocardium and lung air/liver.

To find the upper bound of our approach, we studied the sequences with the worst MAD, HD, PTP, and PFP, respectively. For each sequence, we computed the mean MAD of all the frames in a cycle. We then defined the worst MAD as the maximal mean MAD of 54 sequences. Similarly, we defined the worst HD, PTP, and PFP. Table VII reports the worst MAD, HD, PTP, and PFP of 54 sequences.

For the ENDO contour, the worst MAD, PTP, and PFP appeared in one sequence (we denote it as sequence A), while the worst HD appeared in another sequence (we denote it as sequence B). For sequence A, the MAD started from 0.49 mm in early systole (at the beginning of the cycle), increased to 1.34 mm in ES (in the middle of the cycle), and reached 1.81 mm in ED (at the end of the cycle). The PTP started at 97.5% in early systole, decreased to 92.4% in ES, and reached 88.4% in ED. The PFP started at 2.1% in early systole, increased to 5.1% in ES, and reached 9.6% in ED. The bad MAD, PTP, and PFP are mainly caused by the fuzziness of the inferior and inferolateral segments of the ENDO contours, as shown in Figure 13 (a) and (d). We observed a smooth intensity transition from the LV blood pool to the LV myocardium in the inferior and inferolateral segments without a clear ENDO boundary. For sequence B, the anterolateral papillary muscle is ambiguous, although the posteromedial papillary muscle is clear, as shown in Figure 13 (b) and (e). This leads to a local disagreement between the manual and automatic contours, and increases the HD of the results.

For the EPI contour, the worst MAD, HD, PTP, and PFP appeared in one sequence (we denote it as sequence C). This is caused by the partially missing inferior segments of the myocardium (as shown in Figure 13 (c) and (f)), which misleads the EPI contour to converge to an incorrect position.

Moreover, we calculated the MAD and HD between the predicted contour and manually segmented contour for each frame during ventricular systole (see Figure 14). For comparison purposes, we took the mean shape of SM as its predicted contour. We observed that the SSDM progressively improved the predictive accuracy during systole. In contrast, SM was most accurate in mid systole, while GDM produced an almost constant error.

In addition, we performed Bland-Altman analysis [3] to compare the volume of LV blood pool from manual segmentation and automatic segmentation. Let M and A be the volume of LV blood pool from manual segmentation and automatic segmentation, respectively. The x-axis of the Bland-Altman plot is the average of M and A , i.e. $\text{Mean}(M, A) = \frac{M+A}{2}$, while the y-axis is the difference between M and A , i.e. $\text{Diff}(M, A) = M - A$. To take into account interobserver variability, we asked an additional observer to independently outline the ENDO and EPI contours at early systole, mid systole, and ES, and computed the interobserver variability, as shown in Figure 15.

Figure 16 shows that the blood pool volume from SM has a positive bias (bias= 16.3 ml and 15.8 ml compared to observers 1 and 2, respectively) at early systole, reduces to around 4.5 ml at mid systole, and becomes negative (bias= -7.4 ml and -7.7 ml comparing to observers 1 and 2, respectively) at ES. This is because SM provides too flexible a shape model by ignoring the temporal cardiac dynamics. In comparison, the Bland-Altman plots for SSDM (Figure 17) reveal smaller biases (bias= -2.3 ml, -1.7 ml, and 0.5 ml) at early systole, mid systole, and ES, respectively) between the manual and automatic segmentation. To further study this bias reduction, we performed a paired t-test [30] to assess the statistical significance of the difference between the biases from SM and SSDM. We observed that when SSDM was applied, the reduction of bias was significant at early systole and ES ($p \ll 0.001$), although less significant at mid systole ($p = 0.008$ and 0.012 when observers 1 and 2 were taken as ground truth, respectively).

Figure 18 shows the Bland-Altman plots comparing the volume of blood pool from manual segmentation and automatic segmentation using GDM. We observed that the biases from GDM (comparing to observer 1: bias= 1.7 ml, -1.8 ml, and 1.1 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= 1.2 ml, -1.6 ml, and 0.9 ml at early systole, mid systole, and ES, respectively) were similar to those from SSDM (comparing to observer 1: bias= -1.8 ml, -1.9 ml, and 0.8 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= -2.3 ml, -1.7 ml, and 0.5 ml at early systole, mid systole, and ES, respectively). Also, we did not detect significant differences between the biases from the GDM and SSDM at early systole ($p = 0.41$ and 0.28 when observers 1 and 2 were taken as ground truth, respectively), mid systole ($p = 0.22$ and 0.17 when observers 1 and 2 were taken as ground truth, respectively), and ES ($p = 0.21$ and 0.39 when observers 1 and 2 were taken as ground truth, respectively).

Figure 19 shows the Bland-Altman plots comparing the volume of blood pool from manual segmentation and automatic segmentation using AAMM. We did not observe significant difference between the biases from AAMM (comparing to observer 1: bias= 2.8 ml, -1.7 ml, and -1.8 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= 2.2 ml, 1.5 ml, and -2.0 ml at early systole, mid systole, and ES, respectively) and SSDM (comparing to observer 1: bias= -1.8 ml, -1.9 ml, and 0.8 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= -2.3 ml, -1.7 ml, and 0.5 ml at early systole, mid systole, and ES, respectively). We also did not identify significant differences between the biases from AAMM and SSDM at early systole ($p = 0.25$ and 0.14 , when observers 1 and 2 were taken as ground truth, respectively) and mid systole ($p = 0.11$ and 0.16 , when observers 1 and 2 were taken as ground truth, respectively), although the significance value was slightly lower at ES ($p = 0.04$ and 0.06 , when observers 1 and 2 were taken as ground truth, respectively). This is due to the curse of dimensionality dilemma and small sample size problem as discussed above.

Similarly, we performed Bland-Altman analysis to compare the myocardial volume obtained from manual and automatic segmentation, as shown in Figures 20, 21, 22, and 23. We observed from Figure 20 that when SM was applied, the myocardial volume was biased at

early systole (bias= -9.6 ml and -12.3 ml comparing to observers 1 and 2, respectively) and ES (bias= 12.0 ml and 9.1 ml comparing to observers 1 and 2, respectively). This is due to globally incorrect positioning of the ENDO and EPI contours at early systole and ES, as explained previously. When SSDM was applied, however, the bias at early systole (bias= 1.8 ml and 1.1 ml comparing to observers 1 and 2, respectively) and ES (bias= 2.3 ml and -0.6 ml comparing to observers 1 and 2, respectively) was improved, as shown in Figure 21. A paired t-test shows that the improvement was significant at early systole and ES ($p \leq 0.0001$). In addition, from Figures 22 and 23, we observed that the GDM (comparing to observer 1: bias= 1.8 ml, 1.3 ml, and -1.3 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= -1.0 ml, -1.2 ml, and -4.3 ml at early systole, mid systole, and ES, respectively) and AAMM (comparing to observer 1: bias= 2.9 ml, -2.6 ml, 3.4 ml at early systole, mid systole, and ES; comparing to observer 2: bias= 0.2 ml, -5.1 ml, and 0.4 ml at early systole, mid systole, and ES, respectively) produced similar accurate measurement of myocardial volume as the SSDM (comparing to observer 1: bias= 1.8 ml, 1.9 ml, and 2.3 ml at early systole, mid systole, and ES, respectively; comparing to observer 2: bias= 1.1 ml, -0.7 ml, and -0.6 ml at early systole, mid systole, and ES, respectively). A paired t-test did not detect statistically significant difference between the biases from GDM and SSDM at early systole ($p = 0.18$ and 0.21 , when observers 1 and 2 were taken as ground truth, respectively), mid systole ($p = 0.27$ and 0.14 , when observers 1 and 2 were taken as ground truth, respectively), and ES ($p = 0.15$ and 0.17 , when observers 1 and 2 were taken as ground truth, respectively). Neither did we detect significant difference between the biases from AAMM and SSDM at early systole ($p = 0.11$ and 0.19 , when observers 1 and 2 were taken as ground truth, respectively) and mid systole ($p = 0.19$ and 0.16 , when observers 1 and 2 were taken as ground truth, respectively), although the significance value was slightly lower at ES ($p = 0.07$ and 0.02 , when observers 1 and 2 were taken as ground truth, respectively).

B. Forward versus Forward-Backward Segmentation

In Section V-A, we compared the segmentation results using SSDM with those using SM, GDM, and AAMM in a forward direction. In practice, however, we observed that the segmentation errors from previous frames may propagate to subsequent frames in a sequence. To reduce the propagation of segmentation errors, we performed segmentation in both forward and backward directions and combined them to achieve a more accurate result, as described in Section III-A.

Figure 24 compares the segmentation results using SSDM in forward and forward-backward directions. We observed that when segmentation was performed only in the forward direction, segmentation errors from previous frames propagated into the subsequent frames, and the maximal error appeared at the end of the cycle, as shown in the upper row of Figure 24. When backward segmentation was added, however, error propagation was reduced and the maximal error appeared in the middle of the cycle (i.e. the ES frame), as shown in the lower row of Figure 24. The 3-D results from forward-backward segmentation are shown in Figure 25.

Figure 26 compares the quantitative results of ENDO segmentation over an entire cycle with and without the backward strategy. We observed that when backward segmentation was added, the MAD improved by 0.82 mm, the HD improved by 1.92 mm, the PTP improved by 4.0%, and the PFP improved by 4.2% at the end of the cycle. Figure 27 compares the quantitative results of EPI segmentation over the entire cycle with and without backward strategy. We observed that when backward segmentation was added, the MAD improved by 1.34 mm, the HD improved by 1.98 mm, the PTP improved by 7.8%, and the PFP improved by 8.8% at the end of the cycle. We also observed that the forward-backward segmentation

produced a more consistent delineation of LV boundaries over the full cardiac cycle, despite a slight increase in error in the middle of the cycle (i.e. close to the ES frame).

In addition, we investigated the forward-backward segmentation results of the worst cases discussed in Section V-A. When forward-backward segmentation was performed, the maximal error appeared approximately in the middle of the cycle, instead of at the end of the cycle when only forward segmentation was performed. For sequence A (which has the maximal MAD, PTP, and PFP for the ENDO contour), when forward-backward segmentation was performed, the MAD decreased by 0.21 mm at ES and 1.3 mm at the end of the cycle, the PTP increased by 1.5% at ES and 8.1% at the end of the cycle, the PFP decreased by 0.8% at ES and 7.6% at the end of the cycle. For sequence B (which has the maximal HD for the ENDO contour), the HD decreased by 0.6 mm at ES and 2.6 mm at the end of the cycle. For sequence C (which has the maximal MAD, HD, PTP, and PFP for the EPI contour), when forward-backward segmentation was performed, the MAD decreased by 0.31 mm at ES and 2.2 mm at the end of the cycle, the HD decreased by 0.79 mm at ES and 3.56 mm at the end of the cycle, the PTP increased by 1.4% at ES and 11.1% at the end of the cycle, and the PFP decreased by 1.6% at ES and 9.5% at the end of the cycle.

C. Sensitivity Analysis

As mentioned in Section I, our approach starts with a manual segmentation of the first frame. To determine the sensitivity of automatic results to the accuracy of the first manual contour, we performed sensitivity analysis by varying the first manual contour and assessing the performance of automatic algorithm with varied versions of the first contour.

We used the following three steps to generate varied versions of the first manual contour:

- 1) We built a SM for the first frame (i.e. the ED frame);
- 2) We projected the first manual contour onto the shape space, such that it could be represented as $s_{ED} \approx \bar{s}_{ED} + \Phi_{ED} \mathbf{b}_{ED}$, where $\bar{s}_{ED}$ was the mean shape, Φ_{ED} was the eigenshape matrix, and $\mathbf{b}_{ED}$ was the shape coefficient;
- 3) We generated a set of new ED contours using the formula $s_{ED}^{new} = \bar{s}_{ED} + \alpha \Phi_{ED} \mathbf{b}_{ED}$. We changed α from 0 to 1, such that the generated shapes varied from the mean shape to the one closest to the “true” manual contour.

In practice, we randomly selected two sequences (one canine sequence and one human sequence) as test sequences. We define MAD_1 as the MAD between the “true” manual segmentation and its varied version. Also, we define MAD_2 as the MAD over the entire sequence from automatic algorithm. For each α , we have two MAD_1 's (one for the ENDO contour and the other for the EPI contour) and two MAD_2 's, as shown in Figure 28.

We observed that for $\alpha = 0$ (i.e. the mean shape), the MAD_2 (canine data: $MAD_2 = 1.44$ mm and 1.63 mm for the ENDO and EPI contours, respectively; human data: $MAD_2 = 1.74$ mm and 2.06 mm for the ENDO and EPI contours, respectively) was smaller than the MAD_1 (canine data: $MAD_1 = 1.98$ mm and 2.14 mm for the ENDO and EPI contours, respectively; human data: $MAD_1 = 2.62$ mm and 2.87 mm for the ENDO and EPI contours, respectively). This implies that when the first manual contour is inaccurate, SSDM can improve the segmentation accuracy as it gathers information from multiple frames in the sequence. We also observed that for the ENDO contour, when $\alpha > 0.4$, the MAD_2 ($MAD_2 = 0.74$ mm and 0.82 mm for the canine and human data, respectively) was close to the MAD obtained with “true” first manual contour ($MAD = 0.60$ mm and 0.71 mm for the canine and human data, respectively). For the EPI contour, when $\alpha > 0.2$, the MAD_2 ($MAD_2 = 1.38$ mm and 1.66 mm for the canine and human data, respectively) was close to the MAD

obtained with “true” first manual contour (MAD=1.21 mm and 1.42 mm for the canine and human data, respectively). This implies that SSDM only needs a reasonably accurate ($\alpha > 0.4$ for the ENDO contour and $\alpha > 0.2$ for the EPI contour), rather than a perfect ($\alpha = 1$) manual segmentation in the first frame. This makes it possible to automate the first segmentation step, as will be discussed in Section VI.

VI. Discussion

In Section II-B, we used ICA to perform shape decomposition in the motion subspace because the motion subspace does not have a Gaussian distribution. An alternative is to use non-linear PCA, such as kernel PCA, in the motion subspace. The main idea of kernel PCA is to first map the original data to a high-dimensional space $\mathbf{F}$, such that it can model a non-linear manifold of the original data. Like other non-linear dimension methods [27], [33], however, kernel PCA requires a large training set to effectively characterize the complex non-linear manifold in the high-dimensional space $\mathbf{F}$. Also, it is more computationally expensive than linear PCA and ICA. The possibility of using non-linear PCA in the motion subspace could be a future direction of work.

In Section IV-A, we used normal canine and human data to build two separate SSDMs. However, when applied to segment pathological data, it is possible that these two SSDMs might have a tendency to prefer “normal” shape and motion patterns. There are several potential solutions to this problem. One is to use mixed normal and pathological data as training samples, as was done in [5], [34], [26]. Another possible solution is to introduce uncertainty estimation [32] to the segmentation process. The main idea behind uncertainty estimation is that the weight of the shape prior can be dynamically adjusted based on the regional prediction accuracy from the SSDM. The incorporation of uncertainty estimation into the segmentation process is also a future direction of research.

Our approach requires manual segmentation of the first frame to initiate the whole segmentation process. As shown in Section V-C, this only requires a reasonably accurate, rather than a perfect, manual segmentation. One possible way to automate the first segmentation step would be to build a SM for the ED frame (rather than for the whole sequence), and to use a similar approach to that described in this paper, except that we replace the SSDM with SM, to segment the first frame.

VII. Conclusion

In this paper, we have proposed a SSDM that simultaneously handles temporal dynamics (intra-subject variability) and inter-subject variability in order to predict the specific cardiac dynamics of a new sequence based on the shapes observed in past frames.

In the training phase, we decomposed the aligned cardiac shapes into subject subspace, motion subspace, and landmark subspace using MPCA and MICA. In the prediction phase, we first projected the shapes observed in the past frames onto the subject subspace in order to identify the subject vector associated with this sequence. Next, we used this subject vector to predict the shapes in the future frames, such that we can progressively identify the motion patterns of a new sequence based on the observations in the past frames. We incorporated this SSDM into the segmentation process in a sequential Bayesian framework, which segments each frame according to both the intensity information from the current frame and the prediction from the past frames. In addition, we performed segmentation in both forward and backward directions in order to reduce error propagation.

We performed “Leave-one-out” test on 32 canine sequences and 22 human sequences, and quantitatively compared the experimental results with those obtained from SM, GDM, and AAMM using MAD, HD, PTP, and PFP. We observed that SM had a tendency to be trapped in local minima and was biased in the global sense. This is because SM does not incorporate temporal information which can help to predict the motion of cardiac shapes. Both GDM and AAMM performed better than SM because they are not globally biased. However, GDM is stationary in time and neglects the inter-subject variability of shape and deformation, and therefore missed out on local shape properties that were important for shape-based motion tracking [24]. AAMM does not explicitly model cardiac dynamics because it does not have eigenvectors directly associated with cardiac motion patterns. Also, it does not perform dimension reduction in the temporal domain, making it very high-dimensional and ineffective for covering all possible variations of cardiac dynamics with a limited amount of training data. SSDM outperforms both GDM and AAMM because it decomposes the training shapes in both subject and motion subspaces, and uses a dynamic prediction algorithm to predict subject-specific cardiac shape and dynamics. In addition, we observed a reduction in error propagation when segmentation was performed in both forward and backward directions.

Future work includes in the application of this SSDM to the segmentation of echocardiographic data and the integration of cardiac segmentation and motion analysis into a unified framework.

Appendix A Multilinear Algebra

A tensor, also known as multi-dimensional array, is a higherorder generalization of a vector (first-order tensor) and a matrix (second-order tensor). An element of $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_n}$ is denoted as $\mathcal{A}_{i_1 \dots i_n}$, where $1 \leq i_n \leq I_n$. The mode- n fibers of tensor $\mathcal{A}$ are the I_n -dimensional vectors obtained from $\mathcal{A}$ by varying the index i_n while keeping other indices fixed. The mode- n unfolding of tensor $\mathcal{A}$ is defined by stacking its mode- n fibers in a particular order, as shown in Figure 29. We denote the n -mode unfolding of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_n}$ as $\mathcal{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}$. The n -mode product of tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N)$ -tensor of which the entries are given by

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} \mathcal{A}_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} \mathbf{U}_{j_n i_n} \quad (15)$$

Appendix B Multilinear PCA and ICA

A matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ is a second-order tensor that can be decomposed using singular value decomposition (SVD) as $\mathbf{A} = \mathbf{U}^{(1)} \mathbf{S} \mathbf{U}^{(2)T}$, where $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ are orthonormal matrices. In terms of the n -mode tensor products defined above, this matrix product can be rewritten as $\mathbf{A} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}$. By extension, an M th-order tensor can be decomposed using multilinear SVD (MSVD) [11], [35] as follows

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \dots \times_n \mathbf{U}^{(n)} \dots \times_N \mathbf{U}^{(N)} \quad (16)$$

where tensor $\mathcal{S}$ is the core tensor governing the interaction between mode matrices $\mathbf{U}^{(n)}$ containing orthonormal vectors in the column space of $\mathbf{U}^{(n)}$.

In PCA, we can approximate matrix A by deleting the eigenvectors associated with the smaller singular values.

$$\mathbf{A} \simeq \mathbf{A}' = \mathbf{S}' \times_1 \mathbf{U}'^{(1)} \times_2 \mathbf{U}'^{(2)} \quad (17)$$

Dimension reduction in MPCA is optimized in the least square sense using the Alternating Least Squares (ALS) method [11]. Analogously to Equation 17, we can reconstruct tensor $\mathcal{S}$ as

$$\mathcal{A} \simeq \mathcal{A}' = \mathcal{S}' \times_1 \mathbf{U}'^{(1)} \dots \times_n \mathbf{U}'^{(n)} \dots \times_N \mathbf{U}'^{(N)} \quad (18)$$

where $\mathcal{S}' \in \mathbb{R}^{J^1 \times J^2 \times \dots \times J_N}$ and $\mathbf{U}'^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for all $n = 1, \dots, N$.

Unlike PCA, ICA looks for a linear combination of a set of statistically independent components whose non-Gaussianity is maximized. ICA can be computed from a PCA solution through the rotation of the principal components, such that they become independent [15]:

$$\mathbf{A}^T = \mathbf{U}^{(2)} \mathbf{S} \mathbf{U}^{(1)T} = (\mathbf{U}^{(2)} \mathbf{S} \mathbf{W}^{-1}) (\mathbf{W} \mathbf{U}^{(1)T}) = \mathbf{K}^T \mathbf{C}^T \quad (19)$$

where $\mathbf{W}$ is an invertible transformation matrix computed by the ICA algorithm, $\mathbf{C} = \mathbf{U}^{(1)} \mathbf{W}^T$ are the independent components, and $\mathbf{K} = \mathbf{W}^{-T} \mathbf{S} \mathbf{U}^{(2)T}$ are the coefficients.

Analogous to Equation 19, MICA is obtained using the relationship between SVD and ICA as shown above [36]:

$$\begin{aligned} \mathcal{A} &\simeq \mathcal{S} \times_1 \mathbf{U}^{(1)} \dots \times_N \mathbf{U}^{(N)} \\ &= \mathcal{S} \times_1 \mathbf{U}^{(1)} \mathbf{W}_1^T \mathbf{W}_1^{-T} \dots \times_N \mathbf{U}^{(N)} \mathbf{W}_N^T \mathbf{W}_N^{-T} \\ &= \mathcal{S} \times_1 \tilde{\mathbf{U}}^{(1)} \mathbf{W}_1^{-T} \dots \times_N \tilde{\mathbf{U}}^{(N)} \mathbf{W}_N^{-T} \\ &= (\mathcal{S} \times_1 \mathbf{W}_1^{-T} \dots \times_N \mathbf{W}_N^{-T}) \times_1 \tilde{\mathbf{U}}^{(1)} \dots \times_N \tilde{\mathbf{U}}^{(N)} \\ &= \tilde{\mathcal{S}} \times_1 \tilde{\mathbf{U}}^{(1)} \dots \times_N \tilde{\mathbf{U}}^{(N)} \end{aligned} \quad (20)$$

where the core tensor $\tilde{\mathcal{S}} = \mathcal{S} \times_1 \mathbf{W}_1^{-T} \dots \times_N \mathbf{W}_N^{-T}$.

Appendix C Thin-Plate Spline

Given a set of $\mathbf{q} = \{q^1, q^2, \dots, q^N\}$, the TPS [4] can be written as a composition of an affine transformation and a non-rigid warping:

$$\mathcal{L}(\mathbf{x}; \mathbf{q}) = [\mathbf{A} \quad \mathbf{T} \quad \mathbf{M}] \begin{bmatrix} \mathbf{x} \\ 1 \\ V(\|\mathbf{x} - \mathbf{q}\|) \end{bmatrix} \quad (21)$$

where $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ and $\mathbf{T} \in \mathbb{R}^{3 \times 1}$ define the affine transform, while $\mathbf{M} \in \mathbb{R}^{3 \times N}$ is the weight matrix of the non-rigid warping. $V(r) = -|r|$ is the radial basis function of the spline in 3-D. Also, to ensure that the spline has finite energy, we have the following constraint $\mathbf{M} [\mathbf{q}^T \mathbf{1}] = \mathbf{M} \cdot \mathbf{E} = \mathbf{0}$, where $\mathbf{1} \in \mathbb{R}^{N \times 1}$ and $\mathbf{0} \in \mathbb{R}^{3 \times 4}$. Given the transformed landmarks $\mathbf{s} = \{s^1, s^2, \dots, s^N\}$, the composition of $\mathbf{A}$, $\mathbf{T}$, and $\mathbf{M}$ is obtained as the solution of the following linear equation:

$$\begin{bmatrix} \mathbf{s}^T & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{T} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{E}^T & 0 \\ \mathbf{V} & \mathbf{E} \end{bmatrix} \quad (22)$$

where $\mathbf{V}_{ij} = V(\|\mathbf{q}_i - \mathbf{q}_j\|)$ and $0 \in \mathbb{R}^{4 \times 4}$. Combining 21 and 22, we can parameterize TPS only in terms of transformed landmarks $\mathbf{q}$, via a linear mapping [17]:

$$\mathcal{L}(\mathbf{x}; \mathbf{q}) = \begin{bmatrix} \mathbf{s} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}^T & 0 \\ \mathbf{V} & \mathbf{E} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ 1 \\ V(\|\mathbf{x} - \mathbf{q}\|) \end{bmatrix} \quad (23)$$

The linear mapping makes extremely efficient in the computation of the Jacobian of the warp with respect to the transform landmarks $\frac{\partial \mathcal{L}(\mathbf{x}; \mathbf{s})}{\partial \mathbf{s}}$ in Equation 14. To prevent noninvertible transforms from appearing during the minimization process, we use the approach suggested in [17], consisting in replacing matrix $\mathbf{V}$ with $\mathbf{V} + \lambda \mathbf{I}$ with decreasing λ during the optimization process.

REFERENCES

- [1]. Bertsekas, DP. Nonlinear Programming. 2nd ed.. Athena Scientific; 1999.
- [2]. Blake, A.; Isard, M. Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. Springer; 2000.
- [3]. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. i
- [4]. Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE TPAMI. 1989; 11(no. 6):567–585.
- [5]. Bosch JG, Mitchell SC, Lelieveldt BPF, Nijland F, Kamp O, Sonka M, Reiber JHC. Automatic segmentation of echocardiographic sequences by active appearance motion models. IEEE TMI. 2002; 21(no. 11):1374–1383.
- [6]. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, Pennell DJ, Rumberger JA, Ryan T, Verani MS. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. Circulation. 2002; 105:539–542. [PubMed: 11815441]
- [7]. Chan TF, Vese LA. Active contours without edges. IEEE TIP. 2001; 10(no. 2)
- [8]. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models - their training and application. Computer Vision and Image Understanding. 2001; 61(no. 1):38–59.
- [9]. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE TPAMI. 2001; 23(no. 6): 681–685.
- [10]. Cremers D. Dynamical statistical shape priors for level set-based tracking. IEEE TPAMI. 2006; 28(no. 8):1262–1273.
- [11]. de Lathauwer L, de Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 2000; 21(no. 4):1253–1278.
- [12]. Forsyth, DA.; Ponce, J. Computer Vision: A Modern Approach. Prentice Hall; 2002.
- [13]. Frangi AF, Niessen WJ, Viergever MA. Three-dimensional modeling for functional analysis of cardiac images: A review. IEEE TMI. 2001; 20(no. 1):2–25.
- [14]. Frangi AF, Rueckert D, Schnabel JA, Niessen WJ. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. IEEE TMI. 2002; 21(no. 9)
- [15]. Hastie, T.; Tibshirani, R.; Friedman, JH. The Elements of Statistical Learning. Springer;

- [16]. Kachelriess M, Sennst D, Maxlmoser W, Kalender W. Kymogram detection and kymogram-correlated image reconstruction from subsecond spiral computed tomography scans of the heart. *Medical Physics*. 2002; 29(no. 7):1489–1503. [PubMed: 12148730]
- [17]. Lim J, Yang M-H. A direct method for modeling non-rigid motion with thin plate spline. *CVPR*. 2005; 1:1196–1202.
- [18]. Lorenzo-Valdes M, Sanchez-Ortiz GI, Elkington AG, Mohiaddin RH, Rueckert D. Segmentation of 4d cardiac mr images using a probabilistic atlas and the em algorithm. *MeDIA*. 2004; 8:255–265.
- [19]. Lynch M, Ghita O, Whelan PF. Segmentation of the left ventricle of the heart in 3-d+t mri data using an optimized nonrigid temporal model. *IEEE TMI*. 2008; 27(no. 2):195–203.
- [20]. Mitchell SC, Bosch JG, Lelieveldt BPF, van der Geest RJ, Reiber JH, Sonka M. 3-d active appearance models: Segmentation of cardiac mr and ultrasound images. *IEEE TMI*. 2002; 21(no. 9):1167–1178.
- [21]. Mitchell SC, Lelieveldt BPF, van der Geest RJ, Bosch HG, Reiber JH, Sonka M. Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac mr images. *IEEE TMI*. 2001; 20(no. 5):415–423.
- [22]. Montagnat J, Delingette H. 4d deformable models with temporal constraints: Application to 4d cardiac image segmentation. *MeDIA*. 2005; 9:87–100.
- [23]. Papademetris X, Jackowski M, Rajeevan N, Okuda H, Constable RT, Staib LH. Bioimage suite: An integrated medical image analysis suite. section of Bioimaging Sciences, Dept. of Diagnostic Radiology, Yale School of Medicine. <http://www.bioimagesuite.org>.
- [24]. Papademetris X, Sinusas A, Dione DP, Constable RT, Duncan JS. Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *IEEE TMI*. 2002; 21(no. 7)
- [25]. Perperidis D, Mohiaddin R, Rueckert D. Construction of a 4d statistical atlas of the cardiac anatomy and its use in classification. *MICCAI*. 2005:402–410. [PubMed: 16685985]
- [26]. Qian Z, Metaxas DN, Axel L. Boosting and nonparametric based tracking of tagged mri cardiac boundaries. *MICCAI*. 2006:636–644. [PubMed: 17354944]
- [27]. Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000; 290(no. 5500):2323–2326. [PubMed: 11125150]
- [28]. Senegas J, Netsch T, Cocosco CA, Lund G, Stork A. Segmentation of medical images with a shape and motion model: A bayesian perspective. *MMBIA*. 2004:157–168.
- [29]. Shi P, Sinusas AJ, Constable RT, Ritman E, Duncan JS. Point-tracked quantitative analysis of left ventricular surface motion from 3-d image sequences. *IEEE TMI*. 2000; 19(no. 1)
- [30]. Siegel, S.; N. J. C.. *Nonparametric Statistics for The Behavior Sciences*. 2nd ed.. McGraw-Hill Humanities/Social Sciences/Languages; 1988.
- [31]. Sun W, Cetin M, Chan R, Reddy V, Holmvang G, Chandar V, Willsky A. Segmentation and tracking the left ventricle by learning the dynamics in cardiac images. *IPMI*. 2005:553–565.
- [32]. Taron M, Paragios N, Jolly M-P. Registration with uncertainties and statistical modeling of shapes with variable metric kernels. *IEEE TPAMI*. 2009; 31:99–113.
- [33]. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000; 290(no. 5500):2319–2323. [PubMed: 11125149]
- [34]. van Assen HC, Danilouchkine MG, Dirksen MS, Reiber JH, Lelieveldt BP. A 3-d active shape model driven by fuzzy inference: Application to cardiac ct and mr. *IEEE Transactions on Information Technology and Biomedicine*. 2008; 12(no. 5):595–605.
- [35]. Vasilescu MAO, Terzopoulos D. Multilinear subspace analysis of image emsembles. *CVPR*. 2003; 2:93–99.
- [36]. Vasilescu MAO, Terzopoulos D. Multilinear independent components analysis. *CVPR*. 2005; 1:547–553.
- [37]. Zhu Y, Papademetris X, Sinusas AJ, Duncan JS. Bidirectional segmentation of three-dimensional cardiac mr images using a subject-specific dynamical model. *MICCAI*. 2008; 2:450–457. [PubMed: 18982636]

- [38]. Zhu Y, Papademetris X, Sinusas AJ, Duncan JS. Segmentation of left ventricle from 3d cardiac mr image sequences using a subject-specific dynamical model. CVPR. 2008

Fig. 1.
The interaction of cardiac temporal dynamics and inter-subject shape variability.

Fig. 2.
LV sequence segmentation using SSDM.

Fig. 3.
Forward-backward segmentation of a cardiac sequence.

Fig. 4.
The inter-subject and intra-subject registration.

Fig. 5.

Illustration of shape alignment and landmark extraction. The goal is to find a set of correspondent landmarks on shape 1 and 2. Step 1: we found the dense surface correspondence between surface 1 and 2; Step 2: we asked our cardiology collaborator to manually select a set of representative landmarks on shape 1 and 2; Step 3: we propagated this set of landmarks to shape 2 using the dense surface correspondence found in Step 1.

Fig. 6. Temporal resampling and interpolation. Assume that sequence 1 is the reference sequence. We first aligned the ED frames of two sequences, and then used linear interpolation to generate a sub-frame (between the second and third frames, in gray) from sequence 2, which correspond to the second frame in sequence 1.

Fig. 7.

The mixed-mode decomposition in the motion subspace obtained with PCA. All three modes contain short-axial contraction. The dense displacement fields are obtained using the shape-based tracking algorithm described in [24].

Fig. 8. The significant modes of variations in the motion subspace obtained with ICA. Mode 1: short-axial contraction; Mode 2: twisting of myocardium; Mode 3: long-axial contraction.

Fig. 9.

The histograms of LV blood pool, LV myocardium, and background with their fitted distributions. (a) LV blood pool; (b) LV myocardium; (c) background.

Fig. 10.

Forward segmentation of ENDO surface from a cardiac sequence at early systole (close to the ED frame) (left), mid systole (middle), and ES (right). First row: SM. Second row: GDM. Third row: AAMM. Fourth row: SSDM. Yellow arrow: implanted marker. White arrow: papillary muscle. Green arrow: slight mismatch between automatic and manual contours. Red line: automatic segmentation. Blue dotted line: manual segmentation.

Fig. 11. Forward segmentation of EPI surface (a) without shape prior, (b) using SM, (c) GDM, (d) AAMM, (e) SSDM. Yellow arrow: LV/RV juncture. Blue arrow: lateral portion of EPI surface.

Fig. 12.
The 17-segment LV model [6].

Fig. 13.

Worst results from our approach. (a) and (d) The worst MAD of the ENDO contours. The black arrows indicate the fuzzy inferior and inferolateral segments of the ENDO boundary; (b) and (e) The worst HD of the ENDO contours. The white arrows indicate two papillary muscles. While the posteromedial papillary muscle is clear, the anterolateral papillary muscle is ambiguous, leading to the regional disagreement between the automatic and manual contours; (c) and (f) The worst MAD and HD of the EPI contours. The white arrows indicate the image artifact caused by the partially missing myocardium. Red solid line: automatic segmentation. Green dotted line: manual segmentation.

Fig. 14.

Comparison the prediction errors from SSDM with SM and GDM. The values of the horizontal axis are expressed as relative to 0 at ED and 1 at ES.

Fig. 15. Bland-Altman plots of interobserver variability. (a) blood pool, (b) myocardium.

Fig. 16.

Bland-Altman plots comparing the volume of blood pool between manual segmentation and automatic segmentation using SM. (a) manual 1-SM at early systole, (b) manual 2-SM at early systole, (c) manual 1-SM at mid systole, (d) manual 2-SM at mid systole, (e) manual 1-SM at ES, (f) manual 2-SM at ES.

Fig. 17.

Bland-Altman plots comparing the volume of blood pool between manual segmentation and automatic segmentation using SSDM. (a) manual 1-SSDM at early systole, (b) manual 2-SSDM at early systole, (c) manual 1-SSDM at mid systole, (d) manual 2-SSDM at mid systole, (e) manual 1-SSDM at ES, (f) manual 2-SSDM at ES.

Fig. 18.

Bland-Altman plots comparing the volume of blood pool between manual segmentation and automatic segmentation using GDM. (a) manual 1-GDM at early systole, (b) manual 2-GDM at early systole, (c) manual 1-GDM at mid systole, (d) manual 2-GDM at mid systole, (e) manual 1-GDM at ES, (f) manual 2-GDM at ES.

Fig. 19.

Bland-Altman plots comparing the volume of blood pool between manual segmentation and automatic segmentation using AAMM. (a) manual 1-AAMM at early systole, (b) manual 2-AAMM at early systole, (c) manual 1-AAMM at mid systole, (d) manual 2-AAMM at mid systole, (e) manual 1-AAMM at ES, (f) manual 2-AAMM at ES.

Fig. 20.

Bland-Altman plots comparing the volume of myocardium from manual and automatic segmentation using SM. (a) manual 1-SM at early systole, (b) manual 2-SM at early systole, (c) manual 1-SM at mid systole, (d) manual 2-SM at mid systole, (e) manual 1-SM at ES, (f) manual 2-SM at ES.

Fig. 21.

Bland-Altman plots comparing the volume of myocardium from manual and automatic segmentation using SSDM. (a) manual 1-SSDM at early systole, (b) manual 2-SSDM at early systole, (c) manual 1-SSDM at mid systole, (d) manual 2-SSDM at mid systole, (e) manual 1-SSDM at ES, (f) manual 2-SSDM at ES.

Fig. 22.

Bland-Altman plots comparing the volume of myocardium from manual and automatic segmentation using GDM. (a) manual 1-GDM at early systole, (b) manual 2-GDM at early systole, (c) manual 1-GDM at mid systole, (d) manual 2-GDM at mid systole, (e) manual 1-GDM at ES, (f) manual 2-GDM at ES.

Fig. 23.

Bland-Altman plots comparing the volume of myocardium from manual and automatic segmentation using AAMM. (a) manual 1-AAMM at early systole, (b) manual 2-AAMM at early systole, (c) manual 1-AAMM at mid systole, (d) manual 2-AAMM at mid systole, (e) manual 1-AAMM at ES, (f) manual 2-AAMM at ES.

Fig. 24.

Comparing forward and forward-backward segmentation using SSDM at 1/8, 2/8, 3/8, 4/8, 5/8, 6/8, and 7/8 of the time during a cardiac cycle. Upper: forward segmentation; Lower: forward-backward segmentation. Red line: automatic contour, Blue dotted line: manual segmentation. Arrow: maximal segmentation error.

Fig. 25.
Results of 3-D segmentation of a cardiac sequence at $1/8$, $2/8$, $3/8$, $4/8$, $5/8$, $6/8$, and $7/8$ of the time during a cardiac cycle.

Fig. 26.

Comparing the segmentation results for the ENDO contours using SSDM in forward direction and forward-backward direction. (a) MAD, (b) HD, (c) PTP, (d) PFP.

Fig. 27.

Comparing the segmentation results for EPI contours using SSDM in forward direction and forward-backward direction. (a) MAD, (b) HD, (c) PTP, (d) PFP.

Fig. 28. Sensitivity analysis with different manual segmentations in the first frame. (a) canine data, (b) human data.

Fig. 29.

Unfolding of the $(I_1 \times I_2 \times I_3)$ -tensor $\mathcal{A}$ to the $(I_1 \times I_2 I_3)$ -matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$ -matrix $\mathbf{A}_{(2)}$, and the $(I_3 \times I_1 I_2)$ -matrix $\mathbf{A}_{(3)}$. Left: Mode-1, 2, and 3 fibers; Right: Mode 1, 2, and 3 unfolded matrices.

TABLE I

Comparison with SM, GDM, and AAMM for the segmentation of ENDO surface from canine images.

	SM	GDM	AAMM	SSDM
MAD(mm)	2.31 ± 0.13	1.10 ± 0.12	1.23 ± 0.37	0.64 ± 0.08
HD(mm)	3.55 ± 0.13	3.51 ± 0.15	3.57 ± 0.90	1.52 ± 0.11
PTP(%)	87.4 ± 2.43	95.8 ± 0.98	96.3 ± 2.34	98.5 ± 0.53
PFP(%)	12.6 ± 2.34	4.2 ± 0.90	3.6 ± 1.90	1.51 ± 0.50

Error expressed as mean $\pm$ standard deviation.

TABLE II

Comparison with SM, GDM, AAMM for the segmentation of ENDO surface from human images.

	SM	GDM	AAMM	SSDM
MAD(mm)	2.15 ± 0.17	1.21 ± 0.17	1.42 ± 0.47	0.69 ± 0.13
HD(mm)	3.52 ± 0.25	3.48 ± 0.24	3.43 ± 1.01	1.47 ± 0.16
PTP(%)	88.4 ± 2.85	96.8 ± 1.21	95.4 ± 2.51	98.9 ± 0.59
PFP(%)	11.7 ± 2.73	3.4 ± 1.31	4.2 ± 2.41	1.29 ± 0.70

Error expressed as mean $\pm$ standard deviation.

TABLE III

Comparison with SM, GDM, AAMM for the segmentation of EPI surface from canine images.

	SM	GDM	AAMM	SSDM
MAD(mm)	3.21 ± 0.13	1.31 ± 0.11	1.41 ± 0.45	1.23 ± 0.10
HD(mm)	4.34 ± 0.21	2.69 ± 0.15	3.01 ± 0.93	1.72 ± 0.15
PTP(%)	84.1 ± 2.39	93.1 ± 0.98	94.3 ± 2.24	96.6 ± 0.89
PFP(%)	15.9 ± 2.40	6.6 ± 0.81	5.7 ± 2.01	3.5 ± 0.77

Error expressed as mean $\pm$ standard deviation.

TABLE IV

Comparison with SM, GDM, AAMM for the segmentation of EPI surface from human images.

	SM	GDM	AAMM	SSDM
MAD(mm)	3.24 ± 0.19	1.40 ± 0.14	1.51 ± 0.67	1.27 ± 0.18
HD(mm)	4.39 ± 0.28	2.56 ± 0.18	2.89 ± 0.91	1.61 ± 0.26
PTP(%)	84.7 ± 2.45	92.9 ± 1.01	94.1 ± 2.41	96.9 ± 0.91
PFP(%)	14.7 ± 2.58	7.1 ± 1.03	5.9 ± 2.37	2.9 ± 0.89

Error expressed as mean $\pm$ standard deviation.

TABLE V

Comparison with GDM and AAMM for the local MADs of ENDO surface.

		GDM (mm)	AAMM (mm)	SSDM (mm)
Base	anterior	0.60 ± 0.09	1.02 ± 0.12	0.47 ± 0.08
	anterolateral	0.93 ± 0.14	1.13 ± 0.22	0.59 ± 0.09
	inferolateral	0.94 ± 0.10	1.12 ± 0.15	0.62 ± 0.07
	inferior	0.92 ± 0.11	1.18 ± 0.25	0.59 ± 0.10
	inferoseptal	0.87 ± 0.13	1.19 ± 0.19	0.50 ± 0.08
	anteroseptal	0.84 ± 0.11	1.17 ± 0.13	0.62 ± 0.05
Mid	anterior	0.99 ± 0.10	1.13 ± 0.12	0.53 ± 0.04
	anterolateral	2.01 ± 0.22	1.75 ± 0.47	0.89 ± 0.13
	inferolateral	1.87 ± 0.10	1.84 ± 0.26	0.73 ± 0.11
	inferior	1.91 ± 0.23	1.77 ± 0.57	0.72 ± 0.15
	inferoseptal	1.33 ± 0.11	1.17 ± 0.24	0.74 ± 0.12
	anteroseptal	0.97 ± 0.10	1.13 ± 0.17	0.62 ± 0.09
Apex	anterior	1.12 ± 0.09	1.23 ± 0.18	0.76 ± 0.06
	lateral	1.09 ± 0.12	1.24 ± 0.16	0.78 ± 0.05
	inferior	1.11 ± 0.11	1.37 ± 0.15	0.82 ± 0.06
	septal	1.08 ± 0.10	1.24 ± 0.12	0.80 ± 0.08

Error expressed as mean $\pm$ standard deviation.

TABLE VI

Comparison with GDM and AAMM for the local MADs of EPI surface.

		GDM (mm)	AAMM (mm)	SSDM (mm)
Base	anterior	1.35 ± 0.14	1.54 ± 0.43	1.26 ± 0.13
	anterolateral	1.27 ± 0.07	1.56 ± 0.45	1.21 ± 0.19
	inferolateral	1.29 ± 0.09	1.53 ± 0.54	1.22 ± 0.17
	inferior	1.23 ± 0.08	1.47 ± 0.47	1.27 ± 0.13
	inferoseptal	1.25 ± 0.11	1.28 ± 0.35	1.15 ± 0.17
	anteroseptal	1.18 ± 0.07	1.32 ± 0.43	1.22 ± 0.09
Mid	anterior	1.44 ± 0.10	1.49 ± 0.44	1.34 ± 0.12
	anterolateral	1.38 ± 0.09	1.52 ± 0.46	1.21 ± 0.06
	inferolateral	1.32 ± 0.10	1.47 ± 0.60	1.24 ± 0.07
	inferior	1.29 ± 0.11	1.47 ± 0.57	1.29 ± 0.09
	inferoseptal	1.26 ± 0.13	1.25 ± 0.42	1.14 ± 0.09
	anteroseptal	1.18 ± 0.08	1.23 ± 0.33	1.11 ± 0.08
Apex	anterior	1.53 ± 0.13	1.48 ± 0.51	1.27 ± 0.13
	lateral	1.52 ± 0.11	1.49 ± 0.56	1.31 ± 0.10
	inferior	1.48 ± 0.12	1.56 ± 0.67	1.33 ± 0.13
	septal	1.50 ± 0.11	1.48 ± 0.47	1.32 ± 0.12

Error expressed as mean $\pm$ standard deviation.

TABLE VII

The worst MAD, HD, PTP, and PFP from our approach

	MAD (mm)	HD (mm)	PTP (%)	PFP (%)
ENDO	1.37	3.78	92.1	5.50
EPI	2.11	3.45	89.7	9.81