Published in final edited form as:

Phys Med Biol. 2012 October 7; 57(19): 6125–6145. doi:10.1088/0031-9155/57/19/6125.

Registration-based segmentation of murine 4D cardiac micro-CT data using symmetric normalization

Darin Clark, Alexandra Badea, Yilin Liu, G. Allan Johnson, and Cristian T. Badea

Center for In Vivo Microscopy, Duke University Medical Center, Durham, NC 27710, USA

Abstract

Micro-CT can play an important role in preclinical studies of cardiovascular disease because of its high spatial and temporal resolution. Quantitative analysis of 4D cardiac images requires segmentation of the cardiac chambers at each time point, an extremely time consuming process if done manually. To improve throughput this study proposes a pipeline for registration-based segmentation and functional analysis of 4D cardiac micro-CT data in the mouse. Following optimization and validation using simulations, the pipeline was applied to in vivo cardiac micro-CT data corresponding to 10 cardiac phases acquired in C57BL/6 mice ($n = 5$). After edge-preserving smoothing with a novel adaptation of 4D bilateral filtration, one phase within each cardiac sequence was manually segmented. Deformable registration was used to propagate these labels to all other cardiac phases for segmentation. The volumes of each cardiac chamber were calculated and used to derive stroke volume, ejection fraction, cardiac output, and cardiac index. Dice coefficients and volume accuracies were used to compare manual segmentations of two additional phases with their corresponding propagated labels. Both measures were, on average, >0.90 for the left ventricle and >0.80 for the myocardium, the right ventricle, and the right atrium, consistent with trends in inter- and intra-segmenter variability. Segmentation of the left atrium was less reliable. On average, the functional metrics of interest were underestimated by 6.76% or more due to systematic label propagation errors around atrioventricular valves; however, execution of the pipeline was 80% faster than performing analogous manual segmentation of each phase.

Introduction

Non-invasive, in vivo preclinical imaging is critical to understanding the origin, progression, and treatment of cardiovascular disease because it allows longitudinal screening of morphological and functional phenotypes (Badea *et al.* 2007) as well as quantitative evaluation of therapeutic efficacy (Contag 2002). The mouse is a challenging model system for these studies because it is nearly 3000 times smaller than a human and its heart has an R-R interval 1/10th as long. Even so, in 2005 we successfully performed cardio-respiratory gated 4D micro-CT imaging of the mouse heart (Badea *et al.* 2005). Since then, several other groups have demonstrated 4D cardiac micro-CT with faster scanning times and lower dose, generally at the expense of spatial resolution (Drangova *et al.* 2007; Bartling *et al.* 2007). In spite of these promising results, MRI is considered the gold standard for small animal cardiac imaging (Brau *et al.* 2004; Epstein 2007). Our group has acquired some of the highest-resolution murine cardiac MR images (Bucholz *et al.* 2008, 2010); however, acquisition times for 4D micro-CT can be much faster (micro-CT: as little as 50 seconds,

Corresponding author: Cristian Badea, Center For In Vivo Microscopy, Box 3302 Duke University Medical Center, Durham, NC 27710, 919-684-7858, Cristian.Badea@duke.edu.

Supplemental material

Supplemental material related to figures 2 and 5 is available on our dedicated webpage: <http://www.civm.duhs.duke.edu/cardiac4D/index.html>.

Drangova *et al.* 2007; MR: 31 minutes, Bucholz *et al.* 2008) when isotropic spatial resolution is required, making CT an ideal alternative for high-throughput phenotyping applications.

Regardless of modality, once the data has been acquired, it must be segmented to extract functional information. Software packages such as Avizo (VSG, Burlington, MA), Vitrea (Vital Images, Inc., Minnetonka, MN), and ITK-SNAP (Yushkevich *et al.* 2006) provide tools to greatly expedite manual and semi-automated segmentation. Even so, it is not uncommon for automated segmentation to be one or more orders of magnitude faster than computer-aided manual segmentation (Heijman *et al.* 2008) while freeing up manpower, promoting consistency, and dramatically increasing scalability. These advantages have motivated the development of numerous automated segmentation algorithms, particularly for the left ventricle (Suri 2000, Frangi *et al.* 2001). Previously, we investigated segmentation of 4D cardiac micro-CT data using histogram and region sampling methods (Badea *et al.* 2008b); while these methods yielded consistent functional measurements for the left ventricle, they were not readily applicable to the myocardium and to the other chambers of the heart. Characterization of conditions such as right ventricle hypertrophy and myocardial infarction require assessment of metrics such as myocardial mass and wall motion, metrics which require characterization of changes in structure and function for the whole heart.

This study investigates segmentation of the whole mouse heart via propagation of anatomic labels over time using deformable registration. When successfully applied to 4D cardiac data sets, deformable registration yields not only functional measures but also creates a deformation map estimating the location, magnitude, and timing of the anatomic motion underlying these measures. Several papers report on the use of deformable registration for automated segmentation of phases of the human heart, particularly for the left ventricle. Such work can be further classified by the use of two popular mathematical deformation models: shape models and diffeomorphisms. Shape models generally restrict deformation based on prior information, trading flexibility for computational efficiency and reliability in solving a specific deformation problem (Cootes *et al.* 1999; Frangi *et al.* 2001, 2002; Lorenzo-Valdés *et al.* 2002; Lapp *et al.* 2004). A limited amount of work using MRI data has been performed to create such models of the mouse heart (Perperidis *et al.* 2011). Diffeomorphisms, a more flexible alternative, map a dense continuum of points in space between a reference volume and the data to be segmented (Peyrat *et al.* 2010). Diffeomorphisms generally operate under simplifying assumptions in the form of an iteratively optimized cost function which balances deformation smoothness against the value of a similarity metric (see derivations in Beg *et al.* 2005 and Avants *et al.* 2008), modeling the fact that more conservative deformations are often more anatomically accurate. Under ideal conditions, this model can create a homogenous map of both large and small-scale deformations, readily adapting to phenotypic variation. To the best of our knowledge, our study presents the first diffeomorphic registration of the mouse heart using gated 4D micro-CT data, with the ultimate goal of automated segmentation and analysis. Here we seek not only to fill this gap but to derive a pipeline flexible enough to characterize deformation of the atria, ventricles, and myocardium simultaneously.

To perform deformable registration, we use the open source, ITK-based toolkit called ANTs (Advanced Normalization Tools rev. 639; ITK rev. 3.20; <http://picsl.upenn.edu/ANTs/>; Avants *et al.* 2008, 2011a). In addition to the previously described cost function, ANTs implements a unique diffeomorphic model called symmetric normalization which requires a symmetric deformation response from both volumes involved in the deformation. This additional requirement has been highly successful in deformable registration of human brain MRI data and human lung CT data (see results and computation times in Klein *et al.* 2009

and Murphy *et al.* 2011, respectively). While ANTs' implementation of symmetric normalization does not explicitly solve for 4D deformation fields, the symmetric deformation requirement guarantees that the ending point of each deformation is the starting point of the next deformation, approximating temporal continuity while dividing the deformation problem into a series of smaller deformations which can be computed in parallel. Our hypothesis is that these characteristics make ANTs particularly well suited for 4D cardiac segmentation.

To evaluate the performance of ANTs and symmetric normalization for cardiac segmentation, the proposed segmentation pipeline is first optimized and validated in the 4D MOBY mouse phantom (Segars *et al.* 2004). Application of the pipeline to in vivo C57BL/6 mouse data sets is then presented and analyzed. To clarify notation, one 4D data set is defined as all 3D volumes (consecutive phases) collected during the cardiac cycle from one mouse.

Materials and methods

The goal of this pipeline is to segment the following anatomic structures within each data set for subsequent functional analysis: myocardium (Myo), left ventricle (LV), right ventricle (RV), left atrium (LA), and right atrium (RA). Based on the ECG cycle, we divided the R-R interval of the heart beat into 10 phases with phases 1 and 5 corresponding to ventricular diastole and systole, respectively. Figure 1 and the following subsections outline the segmentation pipeline required to process one 4D data set.

I. Animal preparation and data set acquisition

All animal studies were approved by the Duke University Institutional Animal Care and Use Committee. Five male C57BL/6 mice (Charles-River, Wilmington, MA; 12 weeks of age; 28.5 g average weight) were scanned. The animals were anesthetized with a mixture of 1.5% isoflurane, 50% oxygen, and 48.5% nitrogen delivered through a nose cone. The ECG signal was monitored with electrodes (Blue Sensor, Medicotest, UK) taped to the footpads, and body temperature was maintained using heat lamps, a rectal probe, and a feedback controller (Digi-Sense®, Cole Parmer, Chicago, IL). A pneumatic pillow on the thorax was used to monitor breathing. All physiological signals were processed with Coulbourn modules (Coulbourn Instruments, Allentown, PA) and recorded using a custom-written LabVIEW application (National Instruments, Austin, TX). An intravascular catheter (24 ga, BD Medical Systems, UT) was inserted in the lateral tail vein to inject a liposomal blood pool contrast agent (Mukundan *et al.* 2006, Ghaghada *et al.* 2011; 0.01 ml/g with 135 mg Iodine/ml; half-life ~40 hours). For imaging, the animal was placed in an acrylic cradle containing the nose cone, an upper incisor clamp, and the pneumatic pillow (Howles *et al.* 2009). The cradle assembly was seated vertically on the base of an angular stepper motor. X-ray projections were acquired with one chain of our dual tube/detector micro-CT system (described in Badea *et al.* 2008a). 342 projections (size: 1002×667, 88 µm pixels) were acquired per phase for a total scanning time of 40 minutes per mouse (4 minutes/phase, 10 phases). Note that this acquisition time can be reduced to approximately 5 minutes per mouse using a retrospective gating strategy (Badea *et al.* 2011), with some reduction in image quality. We chose to start with the highest possible quality data, providing a foundation against which future studies with retrospectively gated data might be judged. Scanning parameters included a peak voltage of 80 kVp, a current of 125 mA, and an exposure of 10 ms per projection. Imaging was gated prospectively at the coincidence of a 100 ms window defined as end-expiration and the appropriate 10% increment of the R-R interval (first 10% phase 1, second 10% phase 2, etc.; complete gating details: Badea *et al.* 2004, 2005). The maximum and average variabilities of the heart rate during the acquisition of a single phase were 9.0% and 2.4%, respectively, minimizing potential reconstruction

artifacts. Reconstruction was performed with the Feldkamp (FDK) algorithm (Feldkamp *et al.* 1984) and Parker weighting (Parker 1982), as implemented in the software package COBRA (Exxim Computing Corp., Livermore, CA). Each phase was reconstructed as a 512x512 matrix of isotropic, 88 μm voxels. The total radiation dose per mouse was 0.7 Gy.

II. Generation of a binary mask of the heart

All volumes were processed in the NIfTI file format (<http://nifti.nimh.nih.gov/>). As a pre-processing step, each phase volume was cropped to a rectangular volume of interest encompassing the lungs, the rib cage, and the spine. These prominent landmarks reinforced proper alignment of all phases during registration and guaranteed inclusion of the heart. Once the phases were cropped, a binary mask encompassing the heart was generated. The mask was then dilated to satisfy the static boundary conditions of the deformable registration operation. Given adequate dilation of the mask (~15 voxels), the cropping dimensions and mask generated for phase 1 were applicable to all other cardiac phases within the same data set.

III. Registration

The goal of registration is to create a continuum of transformations relating the anatomy at any one cardiac phase to the corresponding anatomy at every other phase within a given 4D data set. This is made possible by exploiting the correlation between anatomic features and their signature x-ray attenuation profiles (i.e. grayscale profiles) in space and time. Following a common paradigm, registration was conducted in two steps. Step 1, affine registration, co-localized prominent features between phases and was not restricted by the binary mask. Because the projections for a given phase were collected prospectively one phase at a time, affine registration was used to correct for small shifts in the animal's position over time (total shift ~30 μm over 10 phases). Step 2, deformable registration, estimated anatomic motion of the heart between phases within the binary heart mask by calculating a symmetric diffeomorphism. As implemented in ANTs, both steps of the registration process were iterative, optimizing the cross-correlation of intensities between phases:

$$\Pi_{(x)} = \frac{\sum_i ((I(x_i))^{-\mu_{I(x)}})(J(x_i))^{-\mu_{J(x)}})^2}{\sum_i (I(x_i))^{-\mu_{I(x)}})^2 \sum_i (J(x_i))^{-\mu_{J(x)}})^2} \quad (1)$$

where Π was the cross-correlation similarity metric, x was the position at the center of the cross-correlation kernel, i was the relative index within the kernel, I and J were the two volumes being registered, and where μ designated the mean value within the kernel for each respective volume. Iterations of deformable registration proceeded by gradient descent:

$$\phi_{I(x,0.5)} = \phi_{i(x,0.5) + (\delta K \otimes \nabla \Pi_{I(x,0.5)})_0} \phi_{I(x,0.5)} \quad (2)$$

where ϕ was the diffeomorphic map between any two phases (e.g. i = phase 1, 2), δ was the gradient step size, K was the Green's kernel, and where 0.5 was the midpoint (mean) between the two phases at which optimization occurred (for complete details, see Avants *et al.* 2011a).

ANTs was used with its multi-resolution scheme, performing the iterative optimization at each resolution level, starting from the lowest, to avoid convergence at a local minimum and to reduce computation time. Once complete, each registration step produced a transformation which was then used to transform one of the registered phases or a corresponding set of anatomic labels to the other phase. While it was possible to calculate

transforms directly between asynchronous phases (e.g. phases 2 and 7), for some degree of consistency in deformation magnitudes, label propagation was conducted between neighboring phases (e.g. phases 2 and 3). Given the linear nature of the affine transform, the symmetric nature of the deformable transform, and the ability to concatenate transformation files, ten two-step registrations (affine + deformable) between synchronous phases were enough to generate the desired continuum of transformations.

Performing accurate, consistent registrations with ANTs required the calibration of several parameters. For affine registration, we used cross-correlation as the similarity metric and 4 resolution levels with an unrestricted number of iterations allowed at each resolution level. For the deformable registration, we used cross-correlation as the similarity metric, a cross-correlation kernel radius of 3 (voxels), a symmetric normalization parameter of 1.4 (δ , gradient step size), and 4 resolution levels, again with an unrestricted number of iterations allowed at each resolution level. A brute force approach was used to optimize the cross-correlation radius against the symmetric normalization parameter using registrations between the 10 phases of one data set. In brief, using a cross-correlation radius of 4 more than doubled the computation time relative to a radius of 3 while improving the intensity correlation between registered phases by less than 0.02 (on a normalized scale from 0 to 1). Reducing the radius from 3 to 2 reduced the intensity correlation between registered phases by as much as 0.10 with no tangible reduction in computation time. Given a radius of 3, a symmetric normalization parameter of 1.4 was chosen, though parameters from 1 to 1.4 generally yielded similar results in terms of intensity correlation and computation time.

IV. 4D median-based bilateral filtration

To stabilize registration performance in the presence of noise and artifacts, ANTs' documentation recommends Gaussian smoothing of one or both components of the previously discussed cost function, the deformation field and the similarity (cross-correlation) gradient. Here, bilateral filtration (BF; Tomasi and Manduchi 1998), an edge-preserving smoothing algorithm, was used to preprocess the data in place of Gaussian smoothing during registration. Classic BF replaces each voxel in the original volume with the average intensity of surrounding voxels as weighted by geometric displacement (domain) and intensity deviation (range) from the voxel being corrected. As discussed in Clark *et al.* 2012, we combined two previous ideas, performing BF in 4D (Sawall *et al.* 2011) and centering the range weights on the median value within the kernel instead of the central value to improve denoising performance at edges (Francis and de Jager 2003). Here, as a novel, application-specific extension of the work presented in Clark *et al.* 2012, we combined these ideas with deformable registration, registering the two nearest neighbors of each phase to the current phase (e.g. to filter phase 3, phases 2 and 4 were first registered to phase 3). The domain component of this new filter gave equal weight to all voxels within the kernel across space and time. The range component was a Gaussian with a standard deviation roughly equal to the standard deviation of the noise in the data and a mean equal to the median intensity value within the kernel. Using simulations, the following parameters were found to work well: standard deviation of the range Gaussian, 79 HU; kernel size, $3 \times 3 \times 3$ (x, y, z, time).

V. Manual segmentation

To seed label propagation, the structures of interest (myocardium, LV, RV, LA, and RA) were manually segmented within the BF processed phase 3 of each data set using the segmentation tools in Avizo. BF smoothing was found to greatly speed up manual segmentation and to make segmentations more consistent between phases and data sets. Phase 3 was chosen to seed label propagation because of its morphological distance halfway between ventricular diastole and systole. Solely for evaluation purposes, the LV, RV,

LA, RA, and the portion of the aorta within the heart were aggregated into an additional label referred to as the blood pool. The blood pool label was significant because it indicated registration performance at blood-myocardium boundaries independent of the registration performance at the more poorly resolved atrioventricular valves.

VI. Label propagation

Before propagating the phase 3 manual labels, a second set of registration operations (affine + deformable) was performed between the filtered data sets to take advantage of the improved correlation between phases. This second registration executed 15–20% faster than the first set of registrations. During label propagation, ANTs provided two basic interpolation options: nearest neighbor and linear. Linear interpolation of the labels, while generally more accurate, introduced non-integer intensity values which were often invalid when rounded. To correct this problem, a $3 \times 3 \times 3$ kernel was used to round the results of linear interpolation to the nearest integer value present within the same neighborhood in the nearest neighbor interpolation, effectively smoothing the results of nearest neighbor interpolation. Because label propagation errors tended to concentrate at label boundaries and because these boundaries were compared with smooth, manually defined boundaries, correcting the interpolation results in this manner tended to increase the measured accuracy of label propagation. Also during propagation, concatenated transformations could have been used to propagate the manual labels directly from phase 3 to any other phase; in practice, however, independent propagations between contiguous phases appeared to yield more anatomically accurate labels. To minimize the potential accumulation of registration errors, label propagation was performed in the following order: phase 3 to phases 2, 1, 10, and then 9; phase 3 to phases 4, 5, 6, 7, and then 8.

VII. Cardiac function

Given the propagated label set, the voxel size of the acquired data, and the weight and average heart rate of the mouse from which the data was acquired, the following physiological metrics were readily derived:

$$\text{Stroke Volume (SV)} = \text{End Diastolic Volume (EDV)} - \text{End Systolic Volume (ESV)} \quad (3)$$

$$\text{Ejection Fraction (EF)} = \frac{\text{SV}}{\text{EDV}} * 100\% \quad (4)$$

$$\text{Cardiac Output (CO)} = \text{SV} * \text{Heart Rate} \quad (5)$$

$$\text{Cardiac Index (CI)} = \frac{\text{CO}}{\text{Body Weight}} \quad (6)$$

Percent error was used to compare these physiological metrics after deriving them by both manual segmentation (assumed to be the ground truth) and label propagation:

$$\text{Percent Error} = \frac{M_{\text{Propagated}} - M_{\text{Manual}}}{M_{\text{Manual}}} * 100\% \quad (7)$$

where M indicates the metric of interest, $M_{\text{Propagated}}$ was the value of the metric derived by label propagation, and M_{Manual} was the same metric derived by manual segmentation.

VIII. Simulations

To optimize and validate the segmentation pipeline, simulations were performed using the 4D MOBY mouse phantom. The intensity of the phantom's blood was modified to be ~300 HU greater than the intensity of soft tissue, similar to the enhancement obtained with in vivo blood pool contrast agents such as Fenestra VC (ART, Saint-Laurent, Quebec). Ten 3D volumes were generated corresponding to cyclic cardiac phases (phase 1, diastole; phase 5, systole). Each phase (320×320×301; 88 μm isotropic voxels) was used to create 342 virtual projections, replicating the logistics of our custom micro-CT system (Badea *et al.* 2008a). Reconstruction was performed with the Feldkamp algorithm (Feldkamp *et al.* 1984) using projections with and without Poisson noise. Poisson noise was added with MATLAB's *imnoise* function (The MathWorks, Natick, MA), resulting in a realistic noise level within the blood of the reconstructed data (standard deviation ~ 70 HU).

Using the phantom data set, three validation experiments were conducted: (1) 4D BF of the noisy phases, (2) label propagation with biased labels, and (3) label propagation with unbiased labels. The unbiased labels were defined as the labels generated with the original phantom. The bias in these unbiased labels was assumed minimal given the simplified nature of the phantom and the discrete intensities of its component materials. The biased labels were defined as the labels generated by manual segmentation of each phase after reconstruction with Poisson noise, registration, and then 4D BF, analogous to the processing of the in vivo data. Comparing the biased labels with the reference, unbiased labels allowed quantification of errors introduced during the execution of the segmentation pipeline.

In the first experiment, convex minimization of the RMSE relative to the original phantom was used to determine the optimal standard deviation for the Gaussian range weights (79 HU). Root mean square error (RMSE) was computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N_v} \sum_{i,j,k} (f(i,j,k) - b(i,j,k))^2} \quad (8)$$

where N_v was the total number of voxels in the reference phase, $f(i,j,k)$ was the intensity of voxels within the reference phase, and $b(i,j,k)$ was the intensity of voxels within the filtered phase.

In the second experiment, the biased phase 3 labels were propagated to all other phases using the proposed pipeline. Phase 3 was chosen to seed label propagation because its morphological positioning was expected to minimize average registration error. To confirm this choice, experiment 3 used the transformations from experiment 2 to propagate the unbiased labels for phases 1 (diastole), 3, and 5 (systole) to all other phases independently. Experiments 2 and 3 were evaluated using Dice coefficients and volume accuracies:

$$\text{Dice Coefficient} = \frac{2(V_1 \cap V_2)}{V_1 + V_2} \quad (9)$$

$$\text{Volume (Measurement) Accuracy} = 1 - \left(\frac{|V_1 - V_2|}{V_1} \right) \quad (10)$$

where V_1 and V_2 were the volumes of the manual and propagated labels of a given anatomic feature, respectively, and $(V_1 \cap V_2)$ was the volume of the overlap between the two labels. Both metrics take a maximum value of 1. In practice, high volume accuracies indicated accurate prediction of physiological trends while high Dice coefficients indicated sensitivity to changes in size and shape. Both quantities were proportionally more sensitive to features

of smaller volume. Acceptable Dice coefficients and volume accuracies for dynamic structures in the mouse heart remain to be determined.

IX. In vivo evaluation

Once the segmentation pipeline was optimized using the MOBY phantom, we applied it to five in vivo mouse data sets acquired as outlined in Section I. Although the pipeline only requires the manual segmentation of phase 3, phases 1 and 5 of each data set were also manually segmented allowing the calculation of Dice coefficients and volume accuracies. Furthermore, because phases 1 and 5 represent ventricular diastole and systole, respectively, manual segmentation of these phases yielded reference values for the SV, EF, CO, and CI. Significant differences in these metrics were determined using the paired t-test with a significance threshold of $p = 0.05$ and with paired measurements coming from the same animal.

To establish the accuracy of our validation strategy and the degree of uncertainty in manually segmenting each anatomic structure, in vivo segmentation variability was characterized by comparing the manual segmentations used to validate the segmentation pipeline with additional manual segmentations conducted by the same operator (the “primary” operator; intra-segmenter variability) and by one additional operator (the “secondary” operator; inter-segmenter variability). The secondary operator was trained by the primary operator while the primary operator manually segmented phases 1, 3, and 5 in one mouse for intra-segmenter assessment. The secondary operator was then responsible for manually segmenting phases 1, 3, and 5 in a second mouse without the aid of the primary operator for intersegmenter assessment. Intra- and inter-segmenter variability were assessed by comparing these additional segmentations with the original segmentations using Dice coefficients and volume accuracies between the two label sets. When computing volume accuracy, the original segmentations were considered the ground truth volumes (V_1). The segmentations used to assess intra- and inter-segmenter variability were not otherwise used during the validation of the segmentation pipeline.

Results

I. Simulations

Figure 2 illustrates the performance of 4D, median-based BF using the MOBY phantom. By subtracting phase 4 from phase 5 at each step, anatomic motion and edge preservation were evaluated quantitatively in HU. In the original data, this difference map indicated motion within the three visible chambers of the heart as well as between the myocardial wall and the lungs. Subtracting the noisy data from the filtered data at each phase (F) illustrated edge preservation in space with the mean of the displayed difference images being 0.49 ± 58 HU and 1.35 ± 58 HU (mean ± 1 standard deviation) for phases 4 and 5, respectively. Subtracting filtered phase 4 from filtered phase 5 (E, row 3) illustrated edge preservation in time. Looking at a line profile over edges which moved between phases (black arrow, A, row 3), 4D BF reduced the RMSE from 71 HU in the noisy data to 15 HU in the filtered data after accounting for the reconstruction error (C, row 3) in both cases. Over the entire heart mask, the RMSE was reduced from 72 HU to 10 HU. BF also preserved volume accuracy. For instance, the volume accuracy of blood pool label, as derived by intensity thresholding of phase 5, remained at 0.99 before and after filtration.

Figure 3 summarizes the results of combining the filter with registration-based segmentation. Without consulting the unbiased labels (shown in row A), the filtered version of phase 3 was manually segmented. These biased phase 3 labels compared very well with the unbiased phase 3 labels (see table 1). Propagating the biased blood pool label from phase

3 to all other filtered phases (row B) yielded an average Dice coefficient of 0.96 ± 0.01 and an average volume accuracy of 0.99 ± 0.00 when compared with the unbiased labels (all averages excluded the seed phase). Subdividing the blood pool into its anatomic components, these metrics remained at acceptable levels (figure 3D, “Row B” results); however, their standard deviations were generally more than twice the standard deviations for the aggregate blood pool, in part because the blood pool volume is ~ 2.3 times larger than the next largest component, the RV. The unbiased phase 3 labels were also propagated using the transformations generated between the filtered phases to isolate registration errors from human segmentation errors (figure 3, row C). The average statistics for the blood pool, LV, RV, LA, and RA were similar regardless of which labels were used to seed the propagation (figure 3D, “Row C” results). For the myocardium, however, there was a significant improvement in the average Dice coefficient and volume accuracy: biased labels, 0.71 ± 0.02 , 0.80 ± 0.02 ; unbiased labels, 0.87 ± 0.03 , 0.98 ± 0.01 .

Based on these results, phase 3 proved to be a valid seed phase for label propagation. To confirm that phase 3 was the best seed phase, phases 1 and 5 were also tested as seed phases (figure 4). Label propagation was performed with the appropriate unbiased labels using the transformations calculated between the filtered phases, again to isolate registration error. Starting from phase 3 yielded the highest average Dice coefficient for every anatomic feature (figure 4A). Starting from phase 3 also yielded the highest average volume accuracy for each feature except the blood pool (data not shown). Averaging the Dice coefficients for the myocardium, LV, RV, LA, and RA, figure 4B shows an accumulation of error with an increased number of label propagations from the seed phase. While phase 3 is again seen to be best starting phase on average, starting from phase 5 yielded a higher average Dice coefficient for phases 4 and 6.

II. In vivo data

Figure 5 illustrates the results of applying the segmentation pipeline to one of the in vivo data sets. Rows A and B show 2D slices through the long and short axis of the heart at phases 1, 3, and 5. Row C shows a 3D rendering of each label set. The middle column, phase 3, shows the manual segmentation of the seed phase. The labels on phases 1 and 5 were propagated from phase 3 using the segmentation pipeline. Integrating all 5 in vivo data sets, average Dice coefficients and volume accuracies are presented in figure 6. To generate these results the phase 3 manual labels from each data set were propagated to phases 1 and 5. The propagated phase 1 and 5 labels were then compared against the manual segmentations of phases 1 and 5. The results at each phase were then averaged across all 5 data sets. Propagation of the LV and blood pool labels was excellent with the average metrics at or above 0.93 and with standard deviations below 0.04 for both phases. Propagation of the myocardium, RV, and RA labels was good with average metrics at or above 0.85 for both phases but less consistent with a standard deviation approaching 0.07 for the RV Dice coefficient at phase 5 and approaching 0.11 for the RA volume accuracy at phase 5. Propagation of the LA labels proved to be the most difficult with highly variable volume accuracy at phase 1 (0.47 ± 0.25). These results are consistent with the measures of intra- and inter-segmenter variability presented in table 2, suggesting a positive correlation between the consistency of the manual segmentations at phases 1, 3, and 5, and the average metrics for each anatomic structure. Segmentations of the LV and blood pool, for example, were highly repeatable at each phase (individual metrics > 0.93), correlating with the average coefficients and accuracies over all 5 mice which were 0.93 ± 0.04 . Conversely, a high degree of variability in the segmentations of the LA at phase 1 relative to phases 3 and 5 (intra-segmenter volume accuracy: 0.64, phase 1 vs. 0.94, phase 3 and 0.98, phase 5) likely contributed to the variability in volume accuracy at phase 1 over all 5 mice (0.47 ± 0.25) because of the combined effect of manual segmentation errors and registration errors.

Comparing the metrics between figure 6 (in vivo results) and table 2 (segmenter variability) over all anatomic structures and over both segmenters yielded a correlation coefficient of $R = 0.75$ for Dice coefficients and of $R = 0.70$ for volume accuracies.

Extending the label propagation from phase 3 to all other phases in the previously described order, figure 7 details predicted volumes for the LV, RV, LA, RA, and the myocardium at every phase. As expected, the LV and RV have a maximum volume at phase 1 (diastole) and a minimum volume at phase 5 (systole). The volume of the myocardium was highly conserved, decreasing by only 6.84% on average between phases 1 and 10, consistent with a previous report of LV muscle volume conservation in the rat heart (Wise *et al.* 1998). The LA and RA volumes follow a trend nearly opposite to the ventricles with the maximum volume occurring at phase 5 and the minimum volume occurring at phase 9.

Looking at functional metrics, table 3 presents the average SV, EF, CO, and CI for the LV and RV derived for each mouse by manual segmentation of phases 1 and 5 and then averaged over all 5 mice. Table 4 presents the same quantities derived by label propagation from a manual segmentation of phase 3. The manually derived EF for the LV ($54.89 \pm 10.07\%$) is similar though more variable than our previous results in mice of the same strain and average weight, $50.9 \pm 2.8\%$ (Badea *et al.* 2005). The manually derived mean CO for the LV (22.87 ± 8.24 mL/min) and RV (21.11 ± 5.84 mL/min) agree with results from Wiesmann *et al.* (2002) of 21.7 ± 4.0 mL/min and 21.9 ± 4.3 mL/min, respectively. We note that the SV, EF, CO, and CI are similar between the LV and RV using the manual labels ($p > 0.05$). Compared with the manual results, the label propagation errors in these metrics are significant (LV: $p < 0.04$; RV: $p < 0.01$), with percent errors ranging between -6.76% (EF) and -12.19% (CI) for the LV and between -15.31% (EF) and -28.79% (CO) for the RV. Given lower average Dice coefficients (figure 6: LV, 0.94; RV, 0.86), larger errors are expected in the RV metrics derived by label propagation; however, plotting the manual EDVs and ESVs for all 5 mice against the analogous values derived by label propagation explains why the metrics for both chambers are underestimates of substantial magnitude (figure 8). The LV and RV volumes at diastole and systole are under and overestimated, respectively, conspiring to underestimate the SV and, therefore, the EF, CO, and CI as well. Performing a linear, least squares fit of the data reveals a high degree of correlation between the results of the two segmentation schemes (LV: $R = 1.00$; RV: $R = 0.99$), relating the systematic nature of the error over all 5 data sets.

III. Execution of the pipeline

Once established, the pipeline (figure 1) took approximately 10.5 hours to execute per 10-phase data set using a Macintosh workstation with two, 2.66 GHz quad-core Xeon processors and 16GB of RAM. Acquiring, reconstructing, cropping, and masking a data set took about two hours and generally required user guidance. Performing the first set of registrations (affine + deformable), 4D bilateral filtration, and then the second set of registrations, a fully automated process, generally completed within 3 hours when 5 phases were processed in parallel. Given sufficient computer hardware, all 10 phases could be processed in parallel with relatively little overhead. Label propagation and computation of the functional metrics, also fully automated, completed in less than 30 minutes. Complete manual segmentation of the LV, RV, LA, RA, and myocardium at phase 3 took about 5 hours per data set (~3 hours to segment the LV, RV, and complete myocardium; ~2 hours to segment the LA and RA). Assuming 2.5 hours for acquisition, reconstruction, and BF and 5 hours to segment each phase, analogous manual segmentation would require 52.5 hours per data set. Therefore, the pipeline represents a time savings of 42 hours per data set, a speed up of 80%.

Discussion

Deformable registration with symmetric normalization accurately propagated the blood pool and LV labels between phases in the MOBY phantom (figure 3D) and in the in vivo data (figure 6), with average Dice coefficients and volume accuracies above 0.90. Label propagation of the myocardium (unbiased labels), RV, and RA was acceptable in both types of data with average coefficients and accuracies above 0.80. For the LV and RV, one minus the average in vivo coefficients and accuracies (figure 6) is seen to approximate the percent error in the EF estimate and half of the percent error in the SV, CO, and CI (table 4). Intuitively, this makes sense because the EF is the SV normalized by the EDV. For future work, these observations suggest Dice coefficients and volume accuracies which are 0.95 (i.e. 5% error at each phase) to accurately characterize functional quantities from 4D cardiac data, similar to error tolerances in human cardiac CT studies by Haraokawa et al. (2004) but with additional technical challenges associated with preclinical imaging (e.g. high spatial and temporal resolution). Because this goal is on the order of intra-segmenter variability (table 2), further refinement of the segmentation pipeline and the methods used to assess its performance are required.

To improve registration performance in future work, we have identified several cases in which it is seen to underperform. As illustrated by dark gray arrows in figure 3 (row B), one such case occurs when propagating superimposed boundaries such as the atrioventricular valves in the MOBY phantom. The volume of the LV and RV are seen to be systematically under (diastole) and overestimated (systole). This effect is most pronounced in the standard deviations of the LA and RA volume accuracies because of their relatively small volumes (figure 3D, dark gray arrows). In vivo, we hypothesize that this is largely to blame for the systematic error detailed in table 4 and figure 8 as the atrioventricular valves are not well resolved within the in vivo data sets. The light gray arrows in figure 3 (row B) illustrate a closely related case in which it was difficult to manually segment the correct boundary of the myocardium in the filtered data given less than 10 HU of contrast between the myocardium and surrounding soft tissue. When the unbiased labels were propagated using the transforms from the filtered data (row C), some deformation of the myocardium was seen to occur. While this is an extreme example of low contrast deformation, it illustrates that deformable registration of labels from a reliable source can yield more accurate results than direct manual segmentation. It also underpins the role of algorithms such as 4D BF in improving segmentation accuracy and consistency.

Empirically, the deformable registration underperforms in two additional cases, the first of which partly explains why the average volume accuracy of the left atrium within the in vivo data was so low and highly variable at phase 1. Given inherent blurring and noise, the LA was extremely difficult to separate from the neighboring coronary artery when manually segmenting phase 1. While this is one reason phase 3 was chosen to seed label propagation, even propagation of labels from phase 3 to phase 1 would often mistake the coronary artery as part of the LA. While the introduction of 4D BF and the second registration step into the pipeline largely solved a similar problem between the RA and RV, figure 6 and table 2 illustrate additional room for improvement with the LA. The second case was observed at thin features and features with a small radius of curvature such as the tip of the RV, the outer edge of the RA, and generically in the LA, particularly when each feature approached its smallest volume. Modifying the registration parameters (e.g. reducing the cross-correlation radius) may have improved performance near these features; however, given the global optimization performed, any potential improvement near these features would come at the expense of overall registration performance and/or substantially increased computation time. Part of the reason why deformable registration of the LV was much more accurate than for

the RV (table 4) was because the LV has a much smoother, rounder shape, minimizing morphological dependencies.

To reduce the impact of these types of errors in future work and to provide an additional means to validate results, we will investigate the use of probabilistic atlases. A probabilistic atlas integrates intensity information and corresponding anatomic labels from multiple individuals, assigning a probability that any given voxel belongs to a particular anatomic feature. Segmentation of a new data set can then be performed by registering the atlas to the new data set, transferring the probabilistic labels, and then refining the labels to match the new data set using contextual information (Lorenzo-Valdés *et al.* 2004; Avants *et al.* 2011b). By creating such an atlas for multiple phases, the segmentation process could not only be fully automated but might better handle poorly resolved features such as the atrioventricular valves without sacrificing the flexibility of the diffeomorphic registration model.

Conclusions

The ability to perform 4D micro-CT of the mouse heart has generated exciting new possibilities for preclinical studies requiring high-throughput cardiac phenotyping. This paper has proposed and demonstrated the implementation of a pipeline which successfully quantifies functional metrics with relatively minimal user intervention. In its current form, the pipeline yields highly promising and reproducible results, especially for the blood pool and LV (Dice coefficients and volume accuracies > 0.90 at each phase and across segmenters), while reducing processing time by as much as 80% relative to manual segmentation. In future work, the use of a probabilistic atlas of each phase is expected to further improve registration accuracy for more highly variable structures such as the LA while increasing the pipeline's throughput, ultimately allowing characterization of mouse models of cardiac dysfunction and therapeutic efficacy.

Acknowledgments

All work was performed at the Duke Center for In Vivo Microscopy, an NIH/NCRR/NIBIB national Biomedical Technology Research Center (P41 EB015897), with additional support from NCI (U24 CA092656). Liposomal contrast agent was provided by Ketan Ghaghada and Ananth Annapragada (Texas Children's Hospital). We thank Sally Zimney for editorial assistance, Lucy Upchurch for assistance in preparing the supplemental material, and Brian Avants for helpful discussions regarding the use of ANTs.

References

- Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008; 12:26–41. [PubMed: 17659998]
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage.* 2011; 54:2033–2044. [PubMed: 20851191]
- Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics.* 2011
- Badea C, Hedlund LW, Johnson GA. Micro-CT with respiratory and cardiac gating. *Medical Physics.* 2004; 31:3324–3329. [PubMed: 15651615]
- Badea CT, Fubara B, Hedlund LW, Johnson GA. 4-D micro-CT of the mouse heart. *Mol Imaging.* 2005; 4:110–116. [PubMed: 16105509]
- Badea CT, Hedlund LW, Mackel J FB, Lan M, Rockman HA, Johnson GA. Cardiac micro-computed tomography for morphological and functional phenotyping of muscle LIM protein null mice. *Molecular Imaging.* 2007; 6:261–268. [PubMed: 17711781]

- Badea CT, Johnston S, Johnson B, Lin M, Hedlund LW, Johnson GA. A dual micro-CT system for small animal imaging. *SPIE*. 2008;691342-10.
- Badea CT, Johnston SM, Qi Y, Johnson GA. 4D micro-CT for cardiac and perfusion applications with view under sampling. *Phys Med Biol*. 2011; 56:3351–3369. [PubMed: 21558587]
- Badea CT, Wetzel AW, Mistry N, Pomerantz S, Nave D, Johnson GA. Left ventricle volume measurements in cardiac micro-CT: the impact of radiation dose and contrast agent. *Comput Med Imaging Graph*. 2008; 32:239–250. [PubMed: 18243656]
- Bartling SH, Stiller W, Grasruck M, Schmidt B, Peschke P, Semmler W, Kiessling F. Retrospective motion gating in small animal CT of mice and rats. *Invest Radiol*. 2007; 42:704–714. [PubMed: 17984768]
- Beg MF, Miller MI, Troune A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*. 2005; 61:139–157.
- Brau A CS, Hedlund LW, Johnson GA. Cine magnetic resonance microscopy of the rat heart using cardiorespiratory-synchronous projection reconstruction. *Journal of Magnetic Resonance Imaging*. 2004; 20:31–38. [PubMed: 15221806]
- Bucholz E, Ghaghada K, Qi Y, Mukundan S, Johnson GA. Four-dimensional MR microscopy of the mouse heart using radial acquisition and liposomal gadolinium contrast agent. *Magn Reson Med*. 2008; 60:111–118. [PubMed: 18581419]
- Bucholz E, Ghaghada K, Qi Y, Mukundan S, Rockman HA, Johnson GA. Cardiovascular phenotyping of the mouse heart using a 4D radial acquisition and liposomal Gd-DTPA-BMA. *Magn Reson Med*. 2010; 63:979–987. [PubMed: 20373399]
- Clark D, Johnson GA, Badea CT. Denoising of 4D Cardiac Micro-CT Data Using Median-Centric Bilateral Filtration. *Proc. SPIE*. 2012:8314–8142.
- Contag PR. Whole-animal cellular and molecular imaging to accelerate drug development. *Drug Discov Today*. 2002; 7:555–562. [PubMed: 12047855]
- Cootes TF, Beeston C, Edwards GJ, Taylor CJ. A Unified Framework for Atlas Matching Using Active Appearance Models. *Information Processing in Medical Imaging*. 1999:322–333.
- Drangova M, Ford NL, Detombe SA, Wheatley AR, Holdsworth DW. Fast retrospectively gated quantitative four-dimensional (4D) cardiac micro computed tomography imaging of free-breathing mice. *Invest Radiol*. 2007; 42:85–94. [PubMed: 17220726]
- Epstein FH. MRI of left ventricular function. *J Nucl Cardiol*. 2007; 14:729–744. [PubMed: 17826327]
- Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *Journal of The Optical Society of America A*. 1984; 1:612–619.
- Francis, JJ.; De Jager, G. The bilateral median filter. In: Nicolls, F., editor. *Proc of the 14th Annual Symposium of the Pattern Recognition Association of South Africa*. South Africa: Langebaan; 2003.
- Frangi AF, Niessen WJ, Viergever MA. Three-dimensional modeling for functional analysis of cardiac images: a review. *IEEE Trans Med Imaging*. 2001; 20:2–25. [PubMed: 11293688]
- Frangi AF, Rueckert D, Schnabel JA, Niessen WJ. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Trans Med Imaging*. 2002; 21:1151–1166. [PubMed: 12564883]
- Ghaghada KB, Badea CT, Karumbaiah L, Fettig N, Bellamkonda RV, Johnson GA, Annapragada A. Evaluation of tumor microenvironment in an animal model using a nanoparticle contrast agent in computed tomography imaging. *Acad Radiol*. 2011; 18:20–30. [PubMed: 21145026]
- Haraokawa, T.; Kanza, RE.; Kido, T.; Higashino, H.; Mochizuki, T.; Shen, Y. *RSNA*. Chicago: 2004. Accuracy of LV Volume Assessment Using Cardiac MSCT.
- Heijman E, Aben JP, Penners C, Niessen P, Guillaume R, Van Eys G, Nicolay K, Strijkers G J. Evaluation of manual and automatic segmentation of the mouse heart from CINE MR images. *J Magn Reson Imaging*. 2008; 27:86–93. [PubMed: 18050352]
- Howles GP, Nouns JC, Qi Y, Johnson GA. Rapid production of specialized animal handling devices using computer-aided design and solid freeform fabrication. *Journal of Magnetic Resonance Imaging*. 2009; 30:466–471. [PubMed: 19629999]
- Klein A, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*. 2009; 46:786–802. [PubMed: 19195496]

- Lapp RM, Lorenzo-Valdés M, Rueckert D. 3D/4D Cardiac Segmentation Using Active Appearance Models, Non-rigid Registration, and the Insight Toolkit. *Medical Image Computing and Computer-Assisted Intervention*. 2004:419–426.
- Lorenzo-Valdés M, Sanchez-Ortiz GI, Elkington AG, Mohiaddin RH, Rueckert D. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Computing and Computer-Assisted Intervention*. 2004:255–265.
- Lorenzo-Valdés M, Sanchez-Ortiz GI, Mohiaddin R, Rueckert D. Atlas-Based Segmentation and Tracking of 3D Cardiac MR Images Using Non-rigid Registration. *Medical Image Computing and Computer-Assisted Intervention*. 2002:642–650.
- Mukundan S Jr, Ghaghada KB, Badea CT, Kao CY, Hedlund LW, Provenzale JM, Johnson GA, Chen E, Bellamkonda RV, Annapragada A. A liposomal nanoscale contrast agent for preclinical CT in mice. *AJR Am J Roentgenol*. 2006; 186:300–307. [PubMed: 16423931]
- Murphy K, et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging*. 2011
- Parker DL. Optimal short scan convolution reconstruction for fanbeam CT. *Med Phys*. 1982; 9:254–257. [PubMed: 7087912]
- Perperidis D, Bucholz E, Johnson GA, Constantinides C. Morphological studies of the murine heart based on probabilistic and statistical atlases. *Comput Med Imaging Graph*. 2011
- Peyrat JM, Delingette H, Sermesant M, Xu C, Ayache N. Registration of 4D cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons. *IEEE Trans Med Imaging*. 2010; 29:1351–1368. [PubMed: 20304732]
- Sawall S, Bergner F, Lapp R, Mronz M, Karolczak M, Hess A, Kachelriess M. Low-dose cardio-respiratory phase-correlated cone-beam micro-CT of small animals. *Med Phys*. 2011; 38:1416–1424. [PubMed: 21520853]
- Segars WP, Tsui BMW, Frey EC, Johnson GA, Berr SS. Development of a 4-D digital mouse phantom for molecular imaging research. *Molecular Imaging and Biology*. 2004; 6:149–159. [PubMed: 15193249]
- Suri JS. Computer vision, pattern recognition and image processing in left ventricle segmentation: the last 50 years. *Pattern Analysis and Applications*. 2000; 3:209–242.
- Tomasi C, Manduchi R. Bilateral filtering for gray and color images. *Proc of the 1998 IEEE International Conference on Computer Vision*. 1998:839–846.
- Wiesmann F, Frydrychowicz A, Rautenberg J, Illinger R, Rommel E, Haase A, Neubauer S. Analysis of right ventricular function in healthy mice and a murine model of heart failure by in vivo MRI. *Am J Physiol Heart Circ Physiol*. 2002; 283:H1065–H1071. [PubMed: 12181136]
- Wise RG, Huang CL, Gresham GA, Al-Shafei AI, Carpenter TA, Hall LD. Magnetic resonance imaging analysis of left ventricular function in normal and spontaneously hypertensive rats. *J Physiol*. 1998; 513(Pt 3):873–887. [PubMed: 9824724]
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 2006; 31:1116–1128. [PubMed: 16545965]

Segmentation of 4D Cardiac Micro-CT Data

Figure 1.

Segmentation Pipeline: deformable registration and multi-dimensional bilateral filtration are used to map the motion of anatomic features between neighboring phases of the heartbeat, allowing a single set of user defined, anatomic labels to produce a complete characterization of the heart at all phases.

Figure 2.

Evaluation of Bilateral Filtration: (A) original MOBY phantom, phases 4 and 5, (B) phases as reconstructed without noise, (C) reconstruction error, (D) phases as reconstructed with Poisson noise, (E) results of applying bilateral filtration to (D), and (F) the difference before and after filtration. The black arrow in (A), row 3 points to a line profile referenced in text. The calibration bar at the bottom right denotes the window width and level in HU for the grayscale and difference images. A quantitative comparison of the bilateral filter's performance with competing filters can be found in the supplemental material: <http://www.civm.duhs.duke.edu/cardiac4D/index.html>.

Figure 3.

Evaluation of Registration-Based Segmentation: Anatomic segmentation (black contours) of representative phases of the MOBY phantom: 1 (ventricular diastole), 3, 5 (ventricular systole), and 7. (A) unbiased segmentation of all phases; (B) manual segmentation of filtered phase 3 and then propagation of the biased labels using the pipeline; (C) unbiased phase 3 labels propagated using the registrations from (B). All segmentation contours are shown over the original MOBY phantom phases for reference. (D) Average Dice coefficients and volume accuracies over all automatically segmented phases (averages exclude seed phase 3). White arrows indicate anatomic components included in the aggregate blood pool label. Dark gray arrows indicate systematic error at superimposed atrioventricular valves. Light

gray arrows highlight the propagation of the myocardium which could only be properly segmented in the original phantom. The calibration bar at the right of row A denotes the window width and level in HU for the grayscale images. Error bars = ± 1 standard deviation.

Figure 4.

The Best Starting Phase for Manual Segmentation: To confirm phase 3 as the best starting phase on which to perform manual segmentation within the framework of the pipeline, unbiased labels from phases 1, 3, and 5 of the MOBY phantom were independently propagated to all other phases. (A) Average Dice coefficients by starting phase and anatomy (starting phase excluded). (B) Average Dice coefficients (average includes the myocardium, LV, RV, LA, and RA coefficients) for all ten phases from each starting phase. Error bars = ± 1 standard deviation.

Figure 5. Registration Based Segmentation in Real Data: Slice views of the long axis (A) and the short axis (B) of the heart as manually segmented in phase 3 and as propagated to phases 1 and 5. Red contours indicate the boundary of the myocardium. (C) 3D rendering of the labels at each phase with matching colors by anatomy. The myocardium is semi-transparent. An animated version of this figure can be found in the supplemental material: <http://www.civm.duhs.duke.edu/cardiac4D/index.html>. The calibration bar at the right of row B denotes the window width and level in HU for the grayscale images before superimposing the labels.

Label Propagation from Phase 3: Average Dice Coefficients and Volume Accuracies

Figure 6.

Average Dice Coefficients and Volume Accuracies: After performing manual segmentation of phases 1, 3, and 5 within in vivo data sets from 5 mice, the phase 3 labels in each mouse were propagated to phases 1 and 5 within the same mouse using the pipeline. Dice coefficients and volume accuracies were calculated and averaged by comparing the manually segmented labels at these phases with the propagated labels. Error Bars: ± 1 standard deviation.

Figure 7.

Predicted Volume by Anatomy and Phase: After manually segmenting phase 3 within in vivo data sets for five mice, the pipeline was used to propagate the anatomic labels to all other phases. The results from each mouse were averaged. The final results are summarized above. (A) Predicted change in ventricular volume with phase. (B, left axis) Predicted change in atrial volume with phase. (B, right axis) Predicted change in myocardial volume with phase. Error bars: ± 1 standard deviation.

Ventricular Volume at Diastole and Systole by Segmentation Strategy

Figure 8.

Volume Measurement Bias: Ventricular volumes at diastole and systole as measured by manual segmentation and as predicted by label propagation from phase 3 using the segmentation pipeline. Using a linear fit (black line), label propagation is seen to underestimate ventricular volume at diastole and to overestimate it at systole (LV slope = 0.8709; RV slope = 0.7109) resulting in underestimation of SV, EF, CO, and CI. The manual and propagated volumes are strongly correlated (LV, $r = 1.00$; RV, $r = 0.99$) relating a high degree of consistency in segmentation performance for both methods.

Table 1

Comparison of biased and unbiased phase 3 labels.

Anatomy:	Blood	Myo	LV	RV	LA	RA	Blood Pool
Dice Coefficient	0.92	0.73	0.97	0.97	0.96	0.96	0.97
Volume Accuracy	0.96	0.81	1.00	0.98	0.99	0.96	1.00

Table 2

Intra- and inter-segmenter variability (3 Phases).

Phase:	Intra-Segmenter										Inter-Segmenter									
	Dice Coefficients					Volume Accuracies					Dice Coefficients					Volume Accuracies				
	1	3	5	mean	SD	1	3	5	mean	SD	1	3	5	mean	SD	1	3	5	mean	SD
Myo	0.92	0.94	0.92	0.93	0.01	0.99	0.95	0.96	0.97	0.02	0.91	0.92	0.92	0.92	0.00	0.89	0.97	0.99	0.95	0.05
LV	0.98	0.97	0.94	0.96	0.02	0.99	0.98	0.93	0.97	0.03	0.99	0.98	0.97	0.98	0.01	0.98	0.98	0.95	0.97	0.02
RV	0.97	0.95	0.90	0.94	0.04	0.99	0.98	0.86	0.95	0.07	0.94	0.96	0.95	0.95	0.01	0.88	0.96	0.99	0.94	0.05
LA	0.80	0.94	0.92	0.88	0.08	0.64	0.94	0.98	0.85	0.18	0.91	0.95	0.96	0.94	0.02	0.88	0.98	0.96	0.94	0.06
RA	0.93	0.96	0.93	0.94	0.02	0.99	0.98	0.92	0.96	0.04	0.97	0.94	0.95	0.95	0.02	0.99	0.91	1.00	0.97	0.05
Blood Pool	0.98	0.99	0.96	0.98	0.01	0.97	1.00	0.96	0.98	0.02	0.97	0.96	0.96	0.96	0.01	0.95	0.93	0.93	0.94	0.01

Table 3

Measures of cardiac function: manual labels.

	Stroke Volume (μ L)		Ejection Fraction (%)		Heart Rate (beats/min)	Body Weight (g)	Cardiac Output (mL/min)		Cardiac Index (mL/min/g)	
	LV	RV	LV	RV			LV	RV	LV	RV
Mean:	36.41	33.68	54.89	59.49	625.30	28.48	22.87	21.11	0.80	0.74
SD:	12.12	8.22	10.07	9.04	49.76	0.83	8.24	5.84	0.30	0.20

Measures of cardiac function: propagated labels.

Table 4

	Stroke Volume (μ L)		Ejection Fraction (%)		Heart Rate (beats/min)	Body Weight (g)	Cardiac Output (mL/min)		Cardiac Index (mL/min/g)	
	LV	RV	LV	RV			LV	RV	LV	RV
Mean:	32.08	24.07	51.18	50.39	625.30	28.48	20.08	15.03	0.71	0.53
SD:	9.95	6.55	8.89	11.83	49.76	0.83	6.61	4.31	0.24	0.15
% Error:	-11.89	-28.53	-6.76	-15.31	-	-	-12.18	-28.79	-12.19	-28.65