Unsupervised Free-View Groupwise Segmentation for M^3 Cardiac Images Using Synchronized Spectral Network

Yunliang Cai[1], Ali Islam[3], Mousumi Bhaduri[4], Ian Chan[4], and Shuo Li[2,1]

[1] Department of Medical Biophysics, University of Western Ontario, London ON, Canada
[2] GE Healthcare, London ON, Canada
[3] St. Joseph's Health Care, London ON, Canada
[4] London Health Sciences Centre, London ON, Canada

Abstract. Image-based diagnosis and population study on cardiac problems require automatic segmentation on increasingly large amount of data from different protocols, different views, and different patients. However, current algorithms are often limited to regulated settings such as fixed view and single image from one specific modality, where the supervised learning methods can be easily employed but with restricted usability. In this paper, we propose the unsupervised free-view groupwise M^3 segmentation: a simultaneous segmentation for a group of *M*ulti-modality, *M*ulti-chamber, from *M*ulti-subject images from an arbitrary imaging view. To achieve the segmentation, we particularly develop the Synchronized Spectral Network (SSN) model for the joint decomposing, synchronizing, and clustering the spectral representations of free-view M^3 cardiac images. The SSN model generates a set of synchronized superpixels where the corresponding chamber regions share the same superpixel label, which naturally provides simultaneous cardiac segmentation. The segmentation is quantitatively evaluated by more than 10000 images (MR and CT) from 93 subjects and high dice metric ($> 85\%$) is consistently achieved in validation. Our method provides a powerful segmentation tool for cardiac images under non-regulated imaging environment.

1 Introduction

The segmentation of large amount of M^3 (*M*ulti-modality, *M*ulti-chamber, *M*ulti-subject) cardiac image data is a clinical routine and notoriously known as tedious and time-consuming. However, most current cardiac segmentations are single modality methods for limited cardiac image views (short-axis/long-axis). These methods require a lot of manual efforts in pre-processing like ROI cropping and view alignments, and need exhaustive conversions in cross-protocol/modality comparison study. A unified segmentation for free-view M^3 images is strongly desired in quantitative and population study of cardiac problems which will play an increasingly important role in the practical diagnoses of heart diseases.

The segmentation for M^3 images under non-regulated settings, i.e., arbitrary field of views, is a challenging problem. Main difficulties arise from different

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 280–288, 2015.
DOI: 10.1007/978-3-319-24571-3_34

perspectives: **1)** Image view: cardiac shapes can be completely different in various views, shape-based segmentation could fail easily as unified shape model is difficult to define; **2)** Modality: intensity features in cardiac MR and CT are incompatible, segmentation method in one modality cannot be directly applied to the other; **3)** Joint chamber segmentation: some chamber regions (i.e., RV) are difficult to extract in non-regulated image views due to the edge fuzziness.

The current cardiac segmentation are often performed under regulated settings. As reviewed in [10], most algorithms focused on LV segmentation as RV has more complex motion patterns. Among the LV segmentation, in recent studies, Cousty *et al* [5] used watershed-cut algorithm and incorporated spatio-temporal representation for cardiac segmentation. Pednekar *et al* [9] proposed a intensity-based affinity estimation for LV region and perform segmentation by contour fitting. Deformable models are intensively applied in LV segmentation too. Ben Ayed *et al* [3] followed the level set approach and using overlap LV priors as global constrains to obtain robust segmentation results. In addition, strong shape priors are frequently used in LV-RV segmentation on single MR subject. Zhang et al [11] used a combined active shape and active appearance model (ASM+AAM) for segmentation of 4D MR images. Bai *et al* [1] used multi-atlas approach, applying local classifiers for segmentation LV segmentation. The popular methods in MR segmentation are also frequently used in CT segmentation. Zheng *et al* [12] applied marginal space learning techniques for warping prior control points in CT volumes to perform the model-based segmentation. Isgum *et al* [7] took the multi-atlas approach and perform cardiac and aortic CT segmentation using local label fusion. However, in the above mentioned methods, unregulated M^3 segmentation setting (views, modal) were rarely discussed.

To achieve segmentation for M^3 images, we propose a free-view groupwise approach which has the following advantages: **1)** Unregulated segmentation setting: our method can be directly applied on standard cardiac views (short-axis, long-axis) and unregulated cardiac views (coronal, axial). There are no shape limits and pre-training steps. **2)** Simultaneous segmentation: our method provides simultaneous segmentation for the set of images from the same arbitrary view. **3)** Multi-modality, Multi-subject, Multi-chamber (M^3): our method is independent of image modality, and provides implicit chamber matching between images from the same/different subjects. All chamber regions (i.e., LV/RV, LA+LV/RA+RV) from the input images are jointly extracted. **4)** Automatic ROI localization: our method can be directly apply on raw MR/CT scans without manual ROI cropping. Our method provides a general tool for cardiac segmentation problems.

2 Methodology

2.1 Overview

Our free-view groupwise segmentation is fully unsupervised, whereas the result is only driven by the input images. The segmentation can obtain multi-chamber

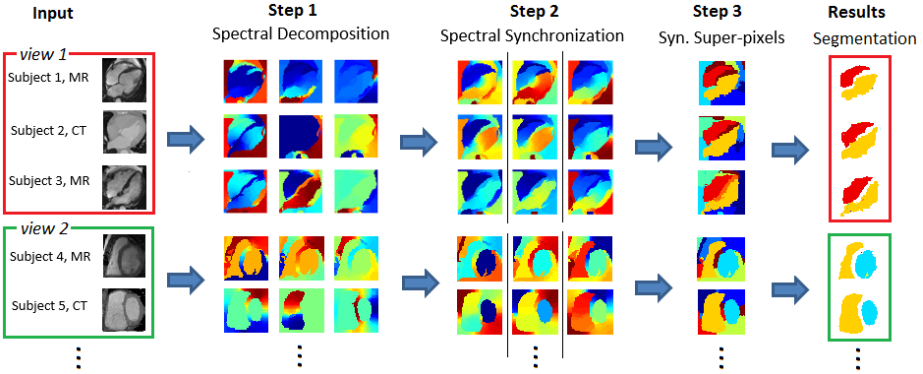

Fig. 1. The overview of our free-view groupwise segmentation. The M^3 images from different views are decomposed and then synchronized using our spectral synchronization network (SSN). The generated synchronized superpixels immediately provide multi-chamber segmentation.

segmentation simultaneously for a group of M^3 images from an unknown view. The overall segmentation process is illustrated in Fig. 1.

1. **Spectral Decomposition.** As Fig. 1 step 1 shows, the input images are decomposed into spectral bases representations. The spectral bases are a set of maps that represent the pairwise pixel similarities in an image, such that smooth regions are enhanced and separated by their in-region similarities. As the similarities only depends on the image itself, spectral bases are modality independent. The spectral bases model [2] is used in our method (see Sec 2.2.1).

2. **Spectral Synchronization.** As illustrated in Fig. 1 step 2, spectral synchronization forces the spectral bases of different images to have similar appearances, so that all spectral representations are jointly matched. *Synchronized Spectral Bases* are obtained by shuffling and recombination of the spectral bases, in which the corresponding regions of different images will have the same positive/negative responses. This provides an implicit registration for the input images without using any explicit alignments (see Sec 2.2.2).

3. **Synchronized Superpixels.** The synchronized spectral bases will serve as image features for clustering the image pixels into superpixels. As shown in Fig. 1 step 3, the synchronized spectral bases expand each pixel to a high-dim feature vector (3d vector in this case). These features are clustered by K-means, sorting all image pixels into a set of synchronized superpixels. These superpixels provide initial over-segmentation for the images, while corresponding image regions are now correlated with same superpixel labels.

4. **Segmentation.** The chamber regions of the input images can be easily identified from the synchronized superpixels of step 3 by their region sizes and the relative positions in images. They are then extracted as the final segmentation.

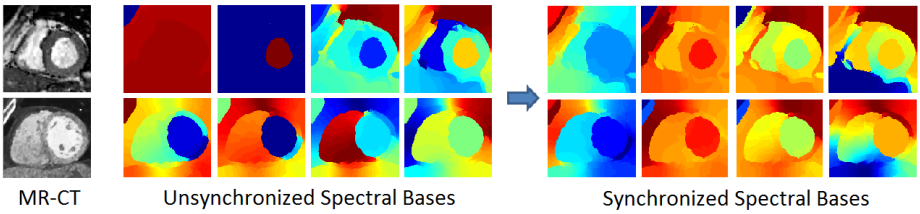

Fig. 2. The spectral bases versus the proposed synchronized spectral bases. Left to right: The MR-CT images, the unsynchronized spectral bases ([2]), and the synchronized spectral bases (obtained by Eqn (3)).

2.2 Spectral Decomposition and Synchronization Using SSN

The key to provide the spectral synchronization and synchronized superpixels is *Synchronized Spectral Network* (SSN), which is introduced as follows.

1. **Spectral Graph Decomposition.** For an image I, we construct a graph $\mathcal{G} = (V, E)$ such that each edge $e \in E$ that connecting two arbitrary pixels i, j is weighted by $W(i, j)$. The particular weight is determined by the intensities, and the contour interventions (not image gradient) between the two pixels:

$$W(i,j) = \exp\left(-||x_i - x_j||^2/\sigma_x - ||I_i - I_j||^2/\sigma_I - \max_{x \in line(i,j)} ||Edge(x)||^2/\sigma_E\right) \tag{1}$$

where x_i, x_j are the location of the pixels i, j and the I_i, I_j are their intensities respectively. $Edge(x)$ represents an edge detector in location x along line (i, j). $\sigma_x, \sigma_I, \sigma_E$ are constants that will be assigned empirically. In practice, $W(i, j)$ will only be computed in the set of k-nearest neighbors. Suppose image I contains N pixels, then W is a $N \times N$ sparse matrix. Let $\overline{W} = D^{-1/2} W D^{-1/2}$ be the normalized W where D is the diagonal matrix whose elements are the row summations of W. The eigen-decomposition of Laplacian $\mathcal{L} = Id - \overline{W}$ will generate the set of unsynchronized spectral bases $\{\xi_k(\mathcal{G})\}$ (Step 1, Fig. 1).

2. **Synchronization by Joint Diagonalization.** After the spectral decompositions of the original images, spectral synchronization is achieved by applying the joint Laplacian diagonalization [6] [8]. The goal of the diagonalization is to obtain a set of qausi-eigenvectors $Y_i = [y_{i,1} \dots y_{i,K}] \in \mathbb{R}^{N \times K}$ for image I_i:

$$\min_{Y_1,\dots,Y_M} \sum_{i \in \mathcal{I}} ||Y_i^T \mathcal{L}_i Y_i - \Lambda_i||_F^2 + \mu \sum_{i,j \in \mathcal{I}} ||F(Y_i) - F(Y_j)||^2, \tag{2}$$

where $\mathcal{I}$ is the image set and $\Lambda_i = \text{diag}(\lambda_1, \dots, \lambda_K)$ is the diagonal matrix for K-largest eigenvalues of $\mathcal{L}_i$. F is an arbitrary feature mapping that maps a spectral map to a fixed dimension feature vector. In other words, Y_i diagonalize the Laplacian $\mathcal{L}_i$ as the ordinary spectral bases do, and its columns are matched implicitly under the feature mapping F. Solving (2) can obtain the demanded synchronized spectral basis (Step 2, Fig. 1).

3. **Fast Computation**. In practice, each $y_{i,k}$ can be approximated by the linear combination of K unsynchronized bases $\{\xi_k(\mathcal{G}_i)\}_{k=1}^K$, which significantly resolves

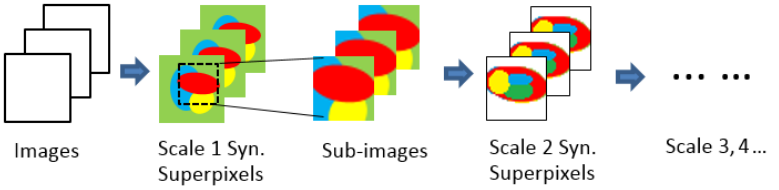

Fig. 3. The hierarchical SSN decomposition. The synchronized superpixels (scale 1) can be sub-decomposed into finer synchronized superpixels (scale 2), and even finer scales (scale 3, 4,...).

the ambiguity of Y_i. We let $Y_i = U_i A_i$ where A_i is a $K \times K'$ matrix variable for $K' \leq K$ and $U_i = [\xi_1(\mathcal{G}_i), \ldots, \xi_K(\mathcal{G}_i)]$, and let F be the matrix of Fourier bases for a Fourier coupling [6]. The optimization (2) is modified as:

$$\begin{aligned} &\min_{A_1,\ldots,A_M} \sum_{i\in\mathcal{I}} ||A_i^T \Lambda_i A_i - \Lambda_i||_F^2 + \mu \sum_{i,j\in\mathcal{I}} ||FU_iA_i - FU_jA_j||^2 \\ &\text{subject to}: A_i^T A_i = Id \quad \text{for all} \quad i \in \mathcal{I} \end{aligned} \tag{3}$$

Fig. 2 shows an example of spectral bases for a pair of MR-CT images versus their original spectral basis obtained by using [4].

2.3 Heart Localization Using Hierarchical SSN Decomposition

The SSN decomposition naturally provides the joint extraction of cardiac ROI sub-images from the raw scans (i.e., MR/CT upper body axial scans). This allows us to directly process the raw scans without any manual ROI cropping.

Hierarchical Decomposition. To achieve the ROI extraction, the spectral decomposition and synchronization are applied in multiple scales to perform the hierarchical decomposition as shown in Fig. 3. The subset of particularly identified synchronized superpixels in the Scale 1 decomposition can be extracted for another round of spectral decomposition and synchronization. The results are a new set of finer synchronized superpixels, which provide sub-segmentations of the local image regions and recover more finer details in the images.

Free-View Heart Localization. Fig. 4 shows an representative example of our heart localization. In this example, the two raw images from MR and CT respectively are first taken the SSN decomposition, generating 9 synchronized superpixels. Through the superpixels positions and the imaging protocol, the heart regions are simultaneously identified as the largest superpixels at the centers. The Scale 2 SSN decomposition can then be taken in the heart sub-images, generating a new set of synchronized superpixels for the cardiac segmentation. In practice, the actual number of superpixels of both scales will be determined experimentally. The number will be fixed for the same testing dataset.

3 Experiment

Our segmentation is tested on three M^3 datasets that contain more than 10000 MR+CT images in total from 93 subjects under various views. High dice metric

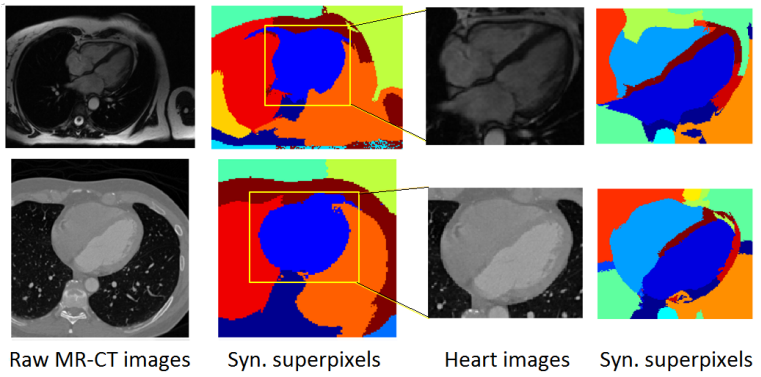

Fig. 4. The simultaneous heart localization in raw axial MR-CT images using hierarchical SSN decomposition. Left to right: the raw axial images; the synchronized superpixels of the raw images; the heart ROI identified from the synchronized superpixels; the scale 2 synchronized superpixels for the heart images.

(DM> 85%) is constantly achieved on both the tests for non-regulated cardiac view images (i.e., 3-chamber coronal view, axial view) and regulated cardiac view images (2/4-chamber view). The tests are arranged in two phases as follows.

Phase 1: Non-Regulated Free-View Evaluation. To evaluate our performance in free-view images, we test our method on a MR+CT dataset which contains 10 MR and 10 CT subjects with non-regulated cardiac views. The representative result of our segmentation on a MR+CT, multi-subject, two-chamber (M^3) image set is shown in Fig. 5 and Fig. 6, while the numerical results are summarized in Table 1. During the test, raw images from three views (axial, sagittal, coronal) of each subject are used. Particularly, 3 slices are sampled for each view around the volume center, generating 60 raw images in total for the subject. As the sampled slices are not accurately regulated along the short/long-axis of heart, their views are in fact arbitrary. Images from the same view are then randomly packed up in batches with 4 or 8 images per batch, generating 100 4-image and 100 8-image non-identical batches for testing.

As Fig. 5 shows, our method achieves simultaneous segmentation for the eight MR/CT images (different subjects) in the image batch. LV+LA/RV+RA regions are successfully extracted in the segmentation. Another two successful examples for the segmentation of 4-image batches under non-regulated sagittal and coronal cardiac views are shown in Fig. 6. As the complete evaluation presented in Table 1, the average dice metric on both modalities (MR+CT) for both chambers are 89.2% (4-image batch) and 85.2% (8-image batch) respectively.

Phase 2: Regulated Standard-View Evaluation. To evaluate our robustness performance on larger batches under regulated cardiac image views, we test two MR testsets: the York Cardiac MRI dataset[1] is denoted as testset 1 which contains 33 MR subjects (aligned in two-chamber view); our own collected MR

[1] `http://www.cse.yorku.ca/mridataset/`

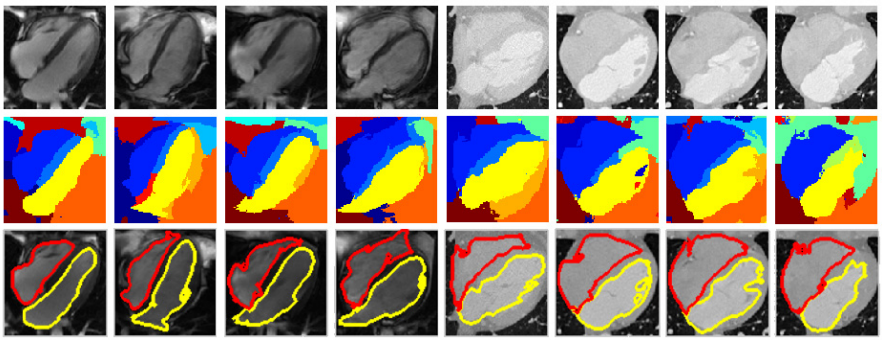

Fig. 5. Example of the unsupervised groupwise segmentation on a 8-image batch (4 MR subjects + 4 CT subjects) under a non-regulated axial view. Row 1: original images; Row 2: synchronized superpixels; Row 3: the LV+LA (**yellow**) and RV+RA (**red**) segmentation.

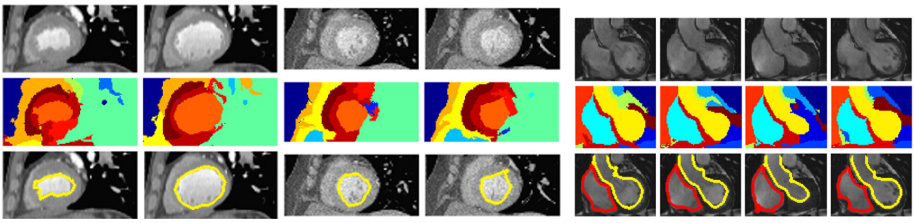

Fig. 6. Examples of the groupwise segmentation on non-regulated image views. Left: CT segmentation for two subjects (2 images/subject) in sagittal view with LV region extracted; Right: MR segmentation for one subject (4 frames) in coronal view with LV/RV (yellow/red, bottom) region extracted.

dataset is denoted as testset 2 which contains 10300 MR images (2740 four-chamber, 7560 two-chamber) from 40 MR subjects. Representative example is shown in Fig. 7 with numerical results are presented in Table 2. Particularly the images are sampled from complete cardiac cycles in order to quantitatively evaluate the robustness of groupwise segmentation. All images from the same cardiac cycles are already well-regulated in short/long-axis views.

As illustrated in Fig. 7, our method has successful simultaneous segmentation on the 20 MR images sampled from a cardiac cycle of one subject. As reported in Table 2, for testset 1, our method achieved average DM 89.1% for the LV segmentation in cardiac cycles under 2-chamber views, and at the same time obtained 85.2% in RV segmentation. For testset2, our segmentation obtain average DM 88.0% (LV) and 84.8% under 2-chamber views, and has 87.6% (LV+LA) and 87.0% (RV+RA) under 4-chamber views. The overall performance in testset 1 and 2 are close and insensitive to view changes.

Table 1. Dice metric (%) and standard derivation on the Free-view M^3 data.

	4-image batch			8-image batch		
	LV(+LA)	RV(+RA)	All Chambers	LV(+LA)	RV(+RA)	All Chambers
MR	91.8 ±3.2	90.2 ±3.1	90.4 ±3.3	88.2 ±4.2	87.8 ±4.5	86.8 ±4.2
CT	89.7 ±2.9	88.7 ±3.6	88.6 ±3.5	86.2 ±4.6	84.9 ±4.7	83.8 ±5.1
MR+CT	90.8 ±3.3	89.8 ±3.5	89.2 ±4.1	85.9 ±5.1	84.0 ±5.0	85.2 ±5.2

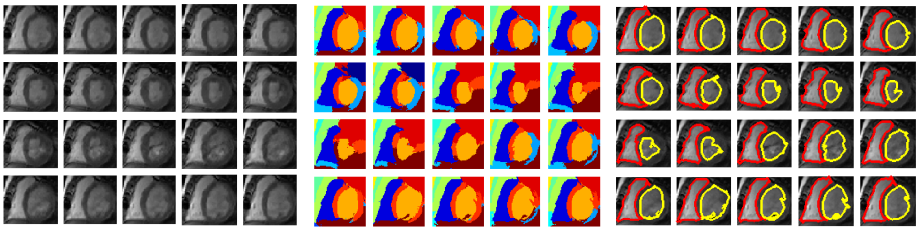

Fig. 7. Example of the unsupervised groupwise segmentation on 20 images under regulated 2-chamber view. Left to right: MR images from a cardiac cycle; synchronized superpixels; final LV-RV segmentation extracted from superpixels.

Table 2. Dice metric (%) evaluations of the regulated-view testset 1 and 2.

	testset 1			testset 2		
	LV(+LA)	RV(+RA)	All Chambers	LV(+LA)	RV(+RA)	All Chambers
2-Chamb. View	89.1 ±4.1	85.2 ±5.6	87.1 ±5.4	88.0 ±4.0	84.8 ±6.0	86.7 ±5.1
4-Chamb. View	/	/	/	87.6 ±4.6	87.0 ±6.2	87.4 ±3.6

4 Conclusions

In this paper, we proposed an unsupervised segmentation approach for M^3 cardiac images (Multi-modality/chamber/subject) under unregulated settings. We developed a synchronized spectral network (SSN) model to conduct hierarchical groupwise segmentation for the input images. The SSN model utilized the robust and modal-independent spectral features to achieve the M^3 segmentation. The dice metric of our method constantly larger than 85% in all test datasets.

References

1. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac mr images. Medical Image Analysis 19(1), 98–109 (2015)
2. Bart, E., Porteous, I., Perona, P., Welling, M.: Unsupervised learning of visual taxonomies. In: CVPR, pp. 1–8. IEEE (2008)
3. Ayed, I.B., Li, S., Ross, I.: Embedding overlap priors in variational left ventricle tracking. IEEE TMI 28(12), 1902–1913 (2009)
4. Cour, T., Benezit, F., Shi, J.: Spectral segmentation with multiscale graph decomposition. In: CVPR, vol. 2, pp. 1124–1131. IEEE (2005)

5. Cousty, J., Najman, L., Couprie, M., Clément-Guinaudeau, S., Goissen, T., Garot, J.: Segmentation of 4D cardiac mri: Automated method based on spatio-temporal watershed cuts. Image and Vision Computing 28(8), 1229–1243 (2010)
6. Eynard, D., Glashoff, K., Bronstein, M.M., Bronstein, A.M.: Multimodal diffusion geometry by joint diagnoalization of laplacians. arXiv:1209.2295 (2012)
7. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in ct scans. IEEE TMI 28(7), 1000–1010 (2009)
8. Kovnatsky, A., Bronstein, M.M., Bronstein, A.M., Glashoff, K., Kimmel, R.: Coupled quasi-harmonic bases. In: EUROGRAPHICS, pp. 439–448. Wiley (2013)
9. Pednekar, A., Kurkure, U., Muthupillai, R., Flamm, S., Kakadiaris, I.A.: Automated left ventricular segmentation in cardiac mri. IEEE Transactions on Biomedical Engineering 53(7), 1425–1428 (2006)
10. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis 15(2), 169–184 (2011)
11. Zhang, H., Wahle, A., Johnson, R.K., Scholz, T.D., Sonka, M.: 4-D cardiac mr image analysis: left and right ventricular morphology and function. IEEE TMI 29(2), 350–364 (2010)
12. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac ct volumes using marginal space learning and steerable features. IEEE TMI 27(11), 1668–1681 (2008)