

# Unsupervised Freeview Groupwise Cardiac Segmentation using Synchronized Spectral Network

Yunliang Cai, *Member, IEEE*, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li\*

**Abstract**—The diagnosis, comparative and population study of cardiac radiology data require heart segmentation on increasingly large amount of images from different modalities/chambers/patients under various imaging views. Most existing automatic cardiac segmentation methods are often limited to single image segmentation with regulated modality/region settings or well-cropped ROI areas, which is impossible for large datasets due to enormous device protocols and institutional differences. A pure data-driven unsupervised segmentation without regulated setting requirements is crucial in this scenario, and will significantly automate the manual work and adopt the various changes of modality, subject or view. In this paper, we propose a general unsupervised groupwise segmentation: a direct simultaneous segmentation for a group of multi-modality, multi-chamber, multi-subject ( $M^3$ ) cardiac images from a freely chosen imaging view. The segmentation can directly perform not only on regulated two/four-chamber images, but also on non-regulated uncropped raw MR/CT scans. A new Synchronized Spectral Network (SSN) is developed for the simultaneous decomposing, synchronizing, and clustering the spectral features of free-view  $M^3$  cardiac images. The SSN-based groupwise analysis of image spectral bases immediately leads to groupwise segmentation of  $M^3$  freeview images. The segmentation is quantitatively evaluated by three datasets (MR and CT mixed) with more than 200 subjects. High dice metric ( $> 85\%$ ) is consistently achieved in validation. Our method provides a powerful tool for medical images under general imaging environment.

**Index Terms**—Cardiac segmentation, spectral synchronization, spectral network, heart localization.

## I. INTRODUCTION

The large amount of  $M^3$  (Multi-modality, Multi-chamber, Multi-subject) cardiac radiology data [1] is widely and intensively used in today's clinical comparative or population analysis.  $M^3$  is popular in describing the diagnosis analysis across modalities/chambers/subjects, especially in the comparative study. It directly correspond to three important comparative measurement: multi-modal image comparison, multi-chamber dynamic pattern (volume changes) comparison, and multi-object shape/region comparison. In addition to  $M^3$ , multiple image views from the traditional short axis/four-chamber views to non-aligned even non-regulated sagittal/axial/coronal

views are frequently switched to provide integrated diagnosis, prognostic and theragnostic for patients. The clinical cardiac image analysis desires increasingly large number of segmentation tasks on  $M^3$  and freeview data.

Despite its popularity, there is a lack of segmentation methods for large freeview  $M^3$  data. Existing segmentation methods fail mostly in a  $M^3$  scenario due to the incompatibilities of modalities/views/subjects and result in exhaustive manual labeling. Most of existing methods are model/atlas-based and supervised learning approaches, which require the input images to have regulated settings: fixed view, fixed modality, well-cropped ROIs. However, due to the differences of device configurations and institutional protocols, uniformly regulated settings are almost impossible in large dataset, especially dataset with unlabeled raw image scans. Tedious manual alignment and cropping are needed. Also, in population study, the number of images are even larger and will involve more sophisticated image appearances, views, and modalities. Image atlas and training samples are often insufficient to foresee all the environment changes.

To overcome all these limitations, a data-driven unsupervised groupwise segmentation for general imaging setting:  $M^3$  and freely chosen views, is strongly desired. The unsupervised+groupwise approach can enable automatic adaptive segmentation regardless the various modality/region/subject/view changes, and can significantly reduce the manual work. This approach is highly demanded in practice and is increasingly important in quantitative or population study of cardiac radiology data.

The unsupervised groupwise segmentation for general cardiac  $M^3$  images under non-regulated settings is a challenging problem: **1) No prior models:** the increasingly changes of  $M^3$  scenarios can hardly be described by a fixed trained model, unsupervised segmentation should perform without strict hypotheses. **2) Incompatible views:** image view determines the 2D projected shapes of the hearts, so that features for a specific view cannot be used for another views. **3) Incompatible modality:** intensity features in MR and CT for the same heart regions are often not similar, segmentation in one modality cannot be directly applied to the other.

We propose a fully unsupervised approach that enables cardiac segmentation over large  $M^3$  data. Testing report of regulated  $M^3$  datasets were presented in [2]. Our new contributions are:

- Fully data-driven unsupervised groupwise segmentation: our method extracts and coordinates the shared features directly and automatically from input images, no prior training or manual labeling are needed.

Yunliang Cai is with the Department of Medical Biophysics, Western University, London, ON, Canada e-mail: (ycai82@uwo.ca).

Ali Islam is with St.Joseph's Health Care London (SJHC), London ON, Canada. (e-mail: Ali.Islam@sjhc.london.on.ca)

Mousumi Bhaduri is with London Health Sciences Center (LHSC), London ON, Canada. (e-mail: Mousumi.Bhaduri@lhsc.on.ca)

Ian Chan is with London Health Sciences Center (LHSC), London ON, Canada. (e-mail: Ian.Chan@lhsc.on.ca)

\*Shuo Li is with the Digital Imaging Group of London, Western University, London, ON, Canada. (e-mail: slishuo@gmail.com)

(\*) indicates corresponding author.

Manuscript received July 29, 2015; revised Jan 10, 2016.

- Unlimited  $M^3$  segmentation: our method can perform segmentation across multi-modality/chamber/subject.
- Free-view segmentation: the segmentation is not limited to specific regulated image views, arbitrary image views can be accepted for the segmentation.
- Automatic heart localization: Our groupwise segmentation method naturally provides a heart region localization. It can be directly performed on raw image scans without manually cropping the heart ROIs.
- Incremental segmentation: Our new groupwise segmentation can propagate the  $M^3$  segmentation from a small image subset to the whole large image set.

## II. RELATED WORK

The existing cardiac segmentation are often performed under regulated settings. Images from freeviews and/or other non-regulated  $M^3$  settings can easily fail in these methods.

**Unsupervised v.s. Supervised Segmentation.** Supervised learning methods are intensively used in most existing cardiac segmentation methods. For example, level set with prior shape knowledge [3] [4] [5], Marginal Space Learning (MSL) [6] for learning the heart shapes, multi-atlas approaches [7] [8] [9] [10] for transferring reference segmentation labels to target images, are popular supervised methods in cardiac segmentation. However, for large cardiac datasets, pre-selected training sets will be insufficient to cover all the changes of modality/regions/views.

**Groupwise v.s. Single Segmentation.** Most existing segmentation methods focus on single image segmentation other than joint segmentation over a set of images. Despite their success in LV/RV segmentation, single image segmentation is not efficient for batch analysis, i.e. a group of images from a cardiac cycle. Instead, groupwise approach conducts the segmentation simultaneously to all input images and automatically coordinates their segment labels. A groupwise segmentation significantly simplify the labeling/coordinating work and provide joint analysis to the segmentations, and can be easily incorporated with existing groupwise registration or groupwise analysis (i.e., [11], [12]) methods.

**Regulated Modality Segmentation.** Due to the differences between the intensity features of MR and CT, the segmentation of MR/CT were separately done in the previous studies. Zheng *et al* [6] applied marginal space learning techniques for warping prior control points specific for CT volumes to perform the model-based segmentation. Isgum *et al* [8] took the multi-atlas approach and perform cardiac and aortic CT segmentation using local label fusion. Jolly [13] proposed a separated MR and CT heart localization steps and used a unified deformation step to combine the two modality results. Ecabert *et al* [14] developed a deformable model-based multi-chamber segmentation for 2D CT images, and the similar method was shown successful in MR by Peters *et al* [15].

**Regulated Chamber Segmentation.** As reviewed in [16], many cardiac segmentation algorithms focused on LV segmentation only. Among the LV segmentation, in recent studies, Cousty *et al* [17] used watershed-cut algorithm and incorporated spatio-temporal representation for cardiac segmentation.

Pednekar *et al* [18] proposed an intensity-based affinity estimation for LV region and perform segmentation by contour fitting. Ben Ayed *et al* [19] followed the level set approach and using overlap LV priors as global constrains to obtain LV segmentation. Strong shape priors are also frequently used in LV or RV segmentation. Zhang *et al* [20] used a combined active shape and appearance model (ASM+AAM) for segmentation of 4D MR images. Mahapatra *et al* [21] utilized the geometric relation between LV and RV learned from training sets to obtain segmentation of both regions.

**Regulated Image Views.** Besides the regulated segmentation target regions, cardiac segmentation are often restricted on the four-chamber long axis view and two-chamber short axis view. 3D models are intensively used for synthesizing multiple image views, but each view still requires a view-dependent segmentation algorithm. Lelieveldt *et al* [22] trained each standard view separately with an Active Appearance models (AAM) to obtain multi-view cardiac segmentation. Lotjonen *et al* [23] constructed a 3D statistical shape model of heart using images from both long-axis and short-axis views. Sermesant *et al* [24] proposed a deformable model approach for 4D segmentation by fitting a manually created tetrahedral mesh heart model to the MRI/SPECT images. Lorenzo-Valdes *et al* [7] used the stack of MR images from short axis view to construct the 3D heart model and then provided the 4D segmentation model via model-fitting. These 3D models should be constructed from regulated two/four-chamber views and should be associated to the same subject.

## III. METHOD OVERVIEW

Our segmentation is fully data-driven and unsupervised. As illustrated in Fig. 1, groups of  $M^3$  images from an arbitrary view are segmented automatically and simultaneously by following the four steps below:

1. **Spectral Decomposition** (Fig. 1 Step 1, Sec.IV). Each input image is decomposed into *Spectral Bases* in this step, to obtain a set of robust, modality independent, and characteristic-preserving feature representations. The spectral bases are the principal components (major eigenvectors) of a *Spectral Matrix*, whose elements are the pairwise pixel similarities of the images (a  $N$ -pixel image will have a  $N \times N$  matrix). The spectral matrix constitutes a graph structure over the image, called *Spectral Graph* ([25], [26]). The spectral base representation depends on pairwise pixel similarities other than absolute values, thus are independent of the intensity changes across modalities. As Fig. 1 and 2 show, smooth regions and boundaries in images are extracted in the spectral bases (converted to 2D) regardless of the original modalities, which simplifies the segmentation as direct clustering.

2. **Spectral Synchronization** (Fig. 1 Step 2, Sec.V). Spectral bases from a group of images are correlated in this step, obtaining a uniform feature representation. Similar heart features (i.e., chamber regions) are thus uniformly matched across images which enables a groupwise analysis over all images. The unification is done by shuffling and recombining the spectral bases under a minimization (*Spectral Synchronization*), thus the new *Synchronized Spectral Bases* are obtained

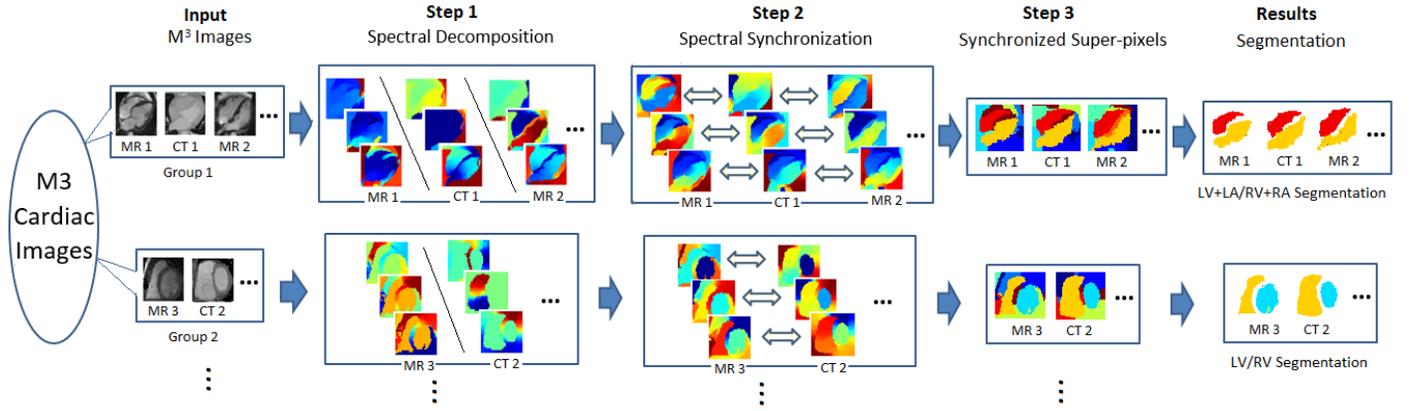


Fig. 1. The overview of our free-view groupwise segmentation. The  $M^3$  images from different views are decomposed and then synchronized using our spectral synchronization network (SSN). The generated synchronized superpixels immediately provide multi-chamber segmentation.

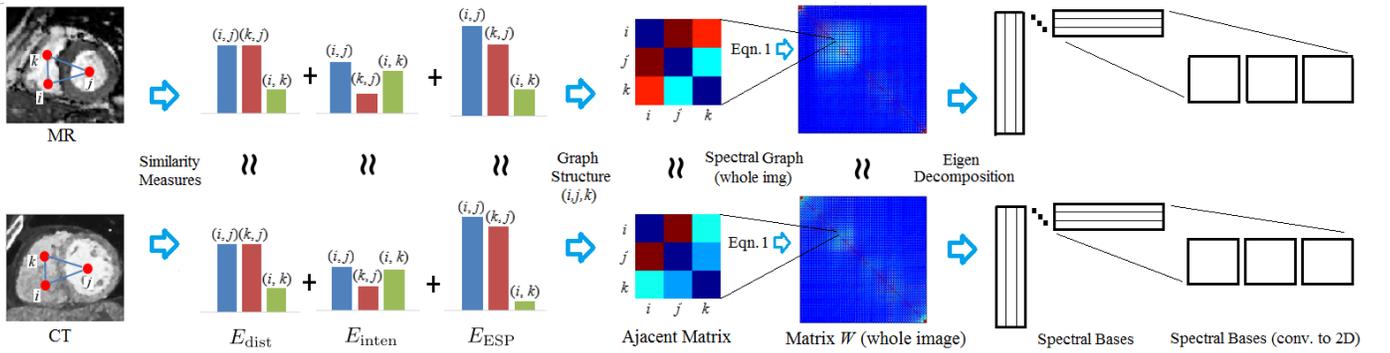


Fig. 2. Invariant spectral graphs and their spectral bases that are shared across MR and CT images. The values of  $E_{\text{dist}}$ ,  $E_{\text{inten}}$ ,  $E_{\text{ESP}}$  show different modality images can have similar self-similarity measures. Thus adjacent matrixes of both graphs are similar which can be used for cross-modality image matching. The full matrixes  $W$  shown at last are constructed by the same self-similarity over the whole MR/CT image.

with minimum 2D appearance differences. This provides an implicit matching for all images without explicit registration.

3. **Synchronized Superpixels** (Fig. 1 Step 3, Sec.VI A and B). The group of synchronized spectral bases are clustered into superpixels in this step to achieve groupwise oversegmentation with superpixel-superpixel correspondences across all images. Unlike traditional superpixels which are obtained by spectral clustering of one image (e.g. [26]), our synchronized superpixel representation is obtained by groupwise clustering over the synchronized spectral bases from all images. Synchronized superpixels not only enhance the clustering robustness, but also correlate all labeled regions across images as Fig. 6 shows.

4. **Segmentation** (Sec.VI C). The chamber regions of the input images can be easily identified from the synchronized superpixels. This is done by stacking the superpixel maps and locating the largest superpixels around the centers. The identified superpixels are extracted as the segmentation result.

#### IV. SPECTRAL DECOMPOSITION FOR IMAGES

Unsupervised segmentation of  $M^3$  cardiac images requires a modality independent feature representation for all images that capture their common image structural information. Image spectral bases only depend on image self-similarities (pairwise pixel similarity) and can be used for modality independent representation. These bases are obtained by the eigen-decomposition for the *Spectral Graph* over each image.

Spectral decomposition is the building block of our SSN model.

##### A. Spectral Graph by Image Self-similarity

The spectral graph is a graph structure for the self-similarities among all pixels within an image. The graph is represented as a *Spectral Matrix*, whereas each element is a similarity value for a pixel pair. The same image structures from different modalities will have similar spectral matrixes so that cross modality matching becomes possible.

For an image  $I$  with total  $N$  pixels, we construct spectral graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  such that  $\mathcal{V}$  ( $N = |\mathcal{V}|$ ) is the pixel set and each edge  $e \in \mathcal{E}$  connects two arbitrary pixels  $i, j$  in the image. Each  $e$  for  $i, j$  is weighted by  $W(i, j)$ , and  $W$  is a  $N \times N$  spectral matrix for  $I$ .  $W(i, j)$  is the combination of three self-similarity measures:

$$W(i, j) = \exp\left(-\frac{E_{\text{dist}}^2(i, j)}{\sigma_x} - \frac{E_{\text{inten}}^2(i, j)}{\sigma_I} - \frac{E_{\text{ESP}}^2(i, j)}{\sigma_E}\right) \quad (1)$$

where the terms  $E_{\text{dist}}$ ,  $E_{\text{inten}}$ ,  $E_{\text{ESP}}$  are the Euclidean distance, intensity difference, and the *Edge Stopping Penalty* (ESP, see Fig. 3, Fig. 4) respectively which are defined as:

$$\begin{aligned} E_{\text{dist}}(i, j) &= \|x_i - x_j\| \\ E_{\text{inten}}(i, j) &= \|I_i - I_j\| \\ E_{\text{ESP}}(i, j) &= \max_{x \in \text{line}(i, j)} \|Edge(x)\| \end{aligned} \quad (2)$$

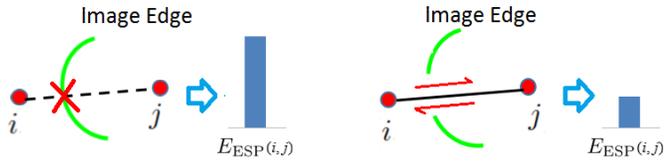


Fig. 3. Typical cases of edge-stopping penalty  $E_{ESP}$ . (Left) the  $line(i, j)$  is intersected by the image contour (Green) versus (Right) no intersection of contours. The similarity between pixel  $i, j$  on left is lower than that on right.

$x_i, x_j$  are the location of the pixels  $i, j$  and the  $I_i, I_j$  are their intensities respectively.  $Edge(x)$  represents an edge detector (i.e., Canny detector) in location  $x$ .  $\sigma_x, \sigma_I, \sigma_E$  are constants that will be assigned empirically. In practice,  $W(i, j)$  will only be computed in  $k$ -nearest neighbors, thus  $W$  is a sparse matrix.

Fig. 2 shows an illustrative example of a spectral graph which contains three pixels over a MR and a CT image. The weight values of each modality invariant self-similarity measure ( $E_{dist}, E_{inten}, E_{ESP}$ ) are shown as normalized histograms. It is easy to see from Fig. 2 that the spectral graph structure is invariant to MR/CT modality.

A self-explained example of edge-stopping penalty is illustrated in Fig. 3. Term  $E_{ESP}$  in (2) is a penalty cost function developed in [26] for measuring the contour intervention between two pixels. Unlike other measures such as intensity similarity (the second term in (2)), or gradient similarity (e.g.,  $\|\nabla I_i - \nabla I_j\|$ ), the edge-stopping penalty only consider the intensity changes along line segment  $line(i, j)$  of pixel  $i, j$ .

### B. Spectral Bases on Graph

Spectral bases are the eigenvectors of the matrix representation  $W$  for the spectral graph. We obtain the eigen-vectors from the graph Laplacian instead of direct decomposition of  $W$ . Let  $D$  be the diagonal matrix whose elements are the row summations of  $W$ . We can have the Laplacian matrix

$$\mathcal{L} = Id - D^{-1/2}WD^{-1/2}. \quad (3)$$

The eigenvectors of  $\mathcal{L}$  are unsynchronized spectral bases such that  $\mathcal{L}$  can be approximated by the  $K$ -smallest eigenvectors:

$$\mathcal{L} \approx \sum_{k=1}^K \lambda_k \xi_k \xi_k^T \quad (4)$$

where  $\lambda_k$  and  $\xi_k$  are the eigenvalue and its associated eigen-vector respectively. These spectral bases preserve the self-similarity information described by (2). As spectral bases are obtained by independent eigen-decompositions, bases from different images are to be synchronized for a more uniform representation.

## V. SPECTRAL SYNCHRONIZATION ACROSS IMAGES

A groupwise synchronization is applied to unify similar chamber regions/boundaries representations among the spectral bases so that these image parts can be jointly correlated across images. Group of spectral bases are synchronized via an optimization computation to obtain the new Synchronized Spectral Bases. The spectral graphs are then correlated by these new bases, forming a *Synchronized Spectral Network*

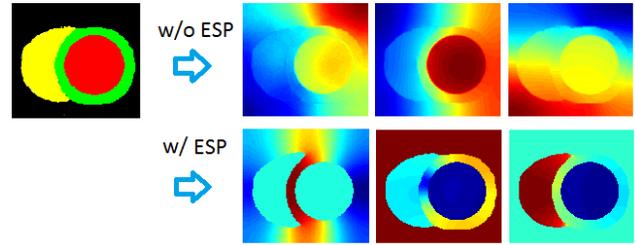


Fig. 4. Spectral bases with/without edge-stopping effect. The artificial image (top left) is decomposed into spectral bases under: (first row) without  $E_{ESP}$  in Eqn (2); (second row) with ESP term. The spectral bases obtained via ESP have sharper boundaries than those without it.

(SSN) which enables the groupwise analysis of all images. Fig. 5 and Fig. 6 show the overall idea of constructing synchronized spectral network using spectral synchronizations.

### A. Spectral Synchronization for Graphs

Spectral synchronization is the searching of a new set of spectral bases for the input images so that the common 2D spectral features are enhanced and correlated across images. Once the spectral bases are synchronized, the related image features (regions/edges) they represented are matched accordingly and can be simultaneously segmented.

Our synchronization are applied to the whole group of 2D spectral maps, using Joint Laplacian Diagonalization with Fourier coupling [27] [2]. Suppose  $\mathcal{L}_m$  is the Laplacian matrixes for  $I_m \in \mathcal{I}$ , we aim to obtain a new set of generalized quasi-eigenvectors  $Y_m = [y_{m,1} \dots y_{m,K}] \in \mathbb{R}^{N \times K}$  which satisfies

$$\sum_{m \in \mathcal{I}} \|Y_m^T \mathcal{L}_m Y_m - \Lambda_m\|_F^2 < \epsilon$$

for  $\Lambda_m = \text{diag}(\lambda_1(\mathcal{G}_m), \dots, \lambda_K(\mathcal{G}_m))$  and  $\epsilon > 0$  is small. In addition, different quasi-eigenvectors  $Y_m$  and  $Y_l$  should be matched in a feature space, such that for a linear/non-linear feature mapping  $F: \mathbb{R}^N \rightarrow \mathbb{R}^{N_F}$  we have

$$\sum_{m, l \in \mathcal{I}, m \neq l} \|F(Y_m) - F(Y_l)\|^2 < \epsilon \quad (5)$$

where mapping  $F$  is determined according to different applications. The complete joint diagonalization problem can be formulated by the following optimization problem:

$$\min_{Y_1, \dots, Y_M} \sum_{m \in \mathcal{I}} \|Y_m^T \mathcal{L}_m Y_m - \Lambda_m\|_F^2 + \mu \sum_{m, l \in \mathcal{I}} \|F(Y_m) - F(Y_l)\|^2 \quad (6)$$

The optimized results  $Y_1^*, \dots, Y_M^*$  are the demanded synchronized spectral basis. The resulting vectors not only serve as ordinary eigenvectors of each spectral graph, but are also matched in pairwise fashion under feature transform  $F$ .

In practice, each quasi-eigenvector  $y_{m,k}$  can be considered as the linear combination of  $\{\xi_k(\mathcal{G}_m)\}_{k=1}^K$ . This assumption resolves the ambiguity of  $Y_m$  and simplifies the optimization. We first let  $Y_m = U_m A_m$  where  $A_m$  is a  $K \times K'$  matrix variable for  $K' \leq K$  and  $U_m = [\xi_1(\mathcal{G}_m), \dots, \xi_K(\mathcal{G}_m)]$ . We also adopt the Fourier coupling [27] in diagonalization, and let  $F$  be the matrix of discrete Fourier bases: a  $N' \times N$  matrix

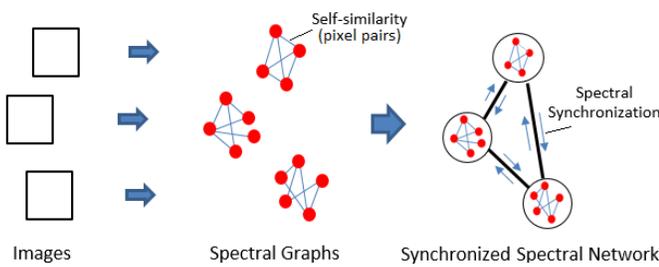


Fig. 5. Uniform feature correlation from spectral graphs to synchronized spectral network. Spectral graphs from independent images are correlated by spectral synchronization. Spectral graph correlates the pixels pairs while spectral synchronization correlates the graph bases.

which contains  $N'$  vectorized discrete 2D Fourier bases. Let  $\tilde{\Lambda}_m = \text{diag}(\lambda_1(\mathcal{G}_m), \dots, \lambda_{K'}(\mathcal{G}_m))$ , Eqn. (6) is modified as:

$$\begin{aligned} \min_{A_1, \dots, A_M} \sum_{m \in \mathcal{I}} \|A_m^T \Lambda_m A_m - \tilde{\Lambda}_m\|_F^2 \\ + \mu \sum_{m, l \in \mathcal{I}} \|FU_m A_m - FU_l A_l\|^2 \\ \text{subject to: } A_m^T A_m = Id \quad \text{for all } m \in \mathcal{I}. \end{aligned} \quad (7)$$

The above optimization problem can be effectively computed by the constrained programming solvers in Matlab. Fig. 6 shows a toy example of spectral synchronization using (7) over two images with different modalities. In this example, the corresponding anatomic structures in different images are forced to match by their similar spectral features.

### B. Incremental Spectral Synchronization

We propose the incremental spectral synchronization for reducing the complexity of large groupwise synchronization and avoiding frequent re-synchronization for newly added data. As defined in Eqn. 7, the spectral synchronization needs to synchronize all images in the group, which will not be efficient for large dataset or frequent data updates. Instead of complete re-synchronization for the data, the incremental method adaptively adopt the previously synchronized bases to build the new spectral bases obtained from the new images.

Similar to (6), for a new input image  $I_w$ , we construct Laplacian  $\mathcal{L}_w$  and the re-synchronized spectral bases  $Y_w$  by:

$$\min_{Y_w} \|Y_w^T \mathcal{L}_w Y_w - \tilde{\Lambda}_w\|_F^2 + \mu \sum_{m \in \mathcal{I}} \|F(Y_m) - F(Y_w)\|^2 \quad (8)$$

where  $\mathcal{I} = \{I_1, \dots, I_M\}$  is the set of images already correlated by SSN.  $\mathcal{I}$  will be fixed during the optimization computation. In other words,  $\mathcal{I}$  serves as the training set for supervising the spectral synchronization of  $I_w$ . The resulting  $Y_w$  can be calculated using the same fast computation method in (7). For a large dataset, the computation complexity of (8) will remain linear as  $Y_w$  can be computed one by one for each  $w$  in the dataset.

The incremental method also provide great flexibility for practical applications. For example in four-chamber segmentation, chambers in the long axis view are not naturally separated in the images. By using our incremental region correspondences, users can still easily separate these chambers

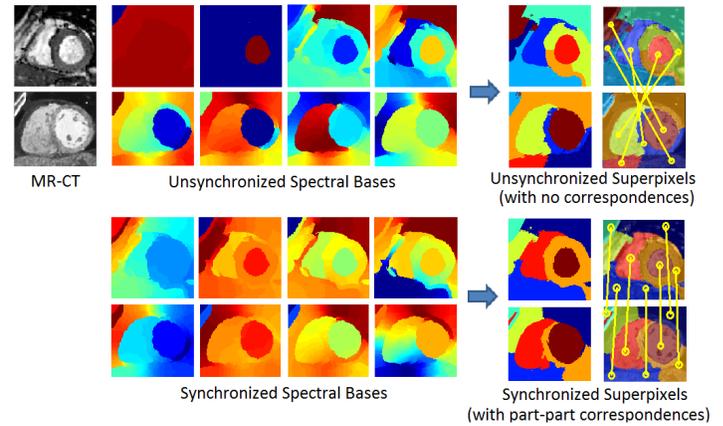


Fig. 6. The comparison between unsynchronized and synchronized spectral bases, the unsynchronized and the synchronized superpixels. The MR-CT images are expanded to the unsynchronized spectral bases using the self-similarity defined in Eqn. 2. The synchronized spectral bases are obtained by Eqn. 7 with the unsynchronized bases. The superpixels obtained accordingly from both types of bases show that the synchronized one can provide part-part correspondences between images.

in long axis view with only one or two simple drawings. The drawings on one image will propagate to all image group under incremental synchronization thus this separation can be achieved simultaneously for the whole image group. This significantly improves the existing manual segmentation methods on four-chamber images.

## VI. SYNCHRONIZED SUPERPIXELS FOR HEART LOCALIZATION AND SEGMENTATION

The groupwise clustering of all synchronized spectral bases immediately leads to a set of synchronized superpixels. Chamber regions are segmented by these superpixels within each image and superpixel-superpixel correspondences are obtained across all images in the group. As heart regions of the images are groupwisely correlated, heart locations can thus be jointly identified from the correlated central superpixels.

### A. Single-scale Synchronized Superpixels for Segmentation

The synchronized superpixels are directly obtained from the K-means clustering of synchronized spectral bases. Unlike the classical single image superpixel approach, we conduct K-means clustering for all synchronized spectral bases in the group to obtain uniform superpixel labeling. The resulting superpixels not only provide oversegmentation for each image, but also provide the superpixel-superpixel correspondences across all images which enables correlated groupwise segmentation. Let  $S_n(m)$  denotes the  $n$ th superpixel on image  $I_m$ , the synchronized superpixels  $\{S_n(1), \dots, S_n(M)\}$  are correlated and assigned to the same label. Corresponding image regions in  $I_1, \dots, I_M$  are matched accordingly. The generation of single-scale SSP can be summarized as Algorithm 1.

In practice, additional contour refinement is applied for each extracted segmentation regions. The refinement is done by first applying the automatic shape adjustment [28] [29] over the extracted heart region superpixels, then smoothing the superpixel contours by classical Fourier descriptor method.

**Algorithm 1:** Single-scale Synchronized Superpixels Generation (SSSG)

- Input** : # of SSP  $N$ , # of spectral bases  $K, K'$ , image set  $\{I_1, \dots, I_M\}$
- Output:** Superpixels  $\{S_1(m), \dots, S_n(m)\}_{m=1}^M$
- 1) Construct spectral graph  $\mathcal{G}_m$  for each  $I_m$ ;
  - 2) Set  $\Lambda_m = \text{diag}(\lambda_1(\mathcal{G}_m), \dots, \lambda_K(\mathcal{G}_m))$ ;  
 $\tilde{\Lambda}_m = \text{diag}(\lambda_1(\mathcal{G}_m), \dots, \lambda_{K'}(\mathcal{G}_m))$ ;  
 $U_m = [\xi_1(\mathcal{G}_m) \cdots \xi_K(\mathcal{G}_m)]$  for all  $m = 1, \dots, M$ ;
  - 3) Compute optimization (7) using  $\{\Lambda_m, \tilde{\Lambda}_m, U_m\}_{m=1}^M$ , obtaining  $Y_m \leftarrow U_m A_m$  for  $m = 1, \dots, M$ ;
  - 4) Stack matrixes  $Y \leftarrow [Y_1 Y_2 \cdots Y_M]$ ;
  - 5) Cluster rows of  $Y$  into  $N$  clusters using K-means;
  - 6) For all  $m = 1, \dots, M$ , partition  $I_m$  to  $\{S_1(m), \dots, S_n(m)\}$  according to the row clusters of  $Y$ ;

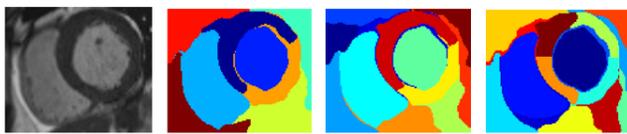


Fig. 7. The effect of superpixel granularity. The increase of superpixel number ( $N$  in Algorithm 1) does not affect the flat regions (eg., LV, RV), it only increases the split of non-flat regions.

**Synchronized v.s. Unsynchronized Superpixels.** Synchronized superpixels are more robust for within-image segmentation and allow cross image correlations. As shown in Fig. 6, traditional superpixels are independently obtained for each image by K-means clustering of its spectral bases. No cross image correspondences are available from the resulting superpixels. In contrast, synchronized superpixels are groupwisely determined by clustering of all synchronized spectral bases. Regions with common spectral representations are enhanced and uniformly labeled in the groupwise clustering. This labeling provides part-part correspondences across images.

**Superpixel Granularity.** The increase of number  $N$  in Algorithm 1 will lead to smaller superpixels. Because of the edge-stopping effect in Eqn (2), the smooth regions and major boundaries will be preserved even if the number increases. The additional superpixels generated will appear in non-flat regions of the images. As Fig. 7 shows, the flat regions (LV, RV regions) are insensitive to the increase of superpixel number.

**B. Multi-scale Synchronized Superpixels for Segmentation**

Coarse-to-fine segmentation is required in the analysis of the multiscale image structures in input data. Our spectral synchronization network can be conducted iteratively for the input images to obtain the multiscale hierarchical decomposition. In the multi-scale synchronization, images are iteratively correlated and segmented, obtaining the multiscale synchronized superpixels. The multi-scale SSPs are then progressively extracted as sub-images/sub-sub-images for finer segmentation.

As illustrated in Fig. 8, we first decompose and synchronize the spectral bases of the input images (Scale 1), generating the set of synchronized spectral bases and subsequently the

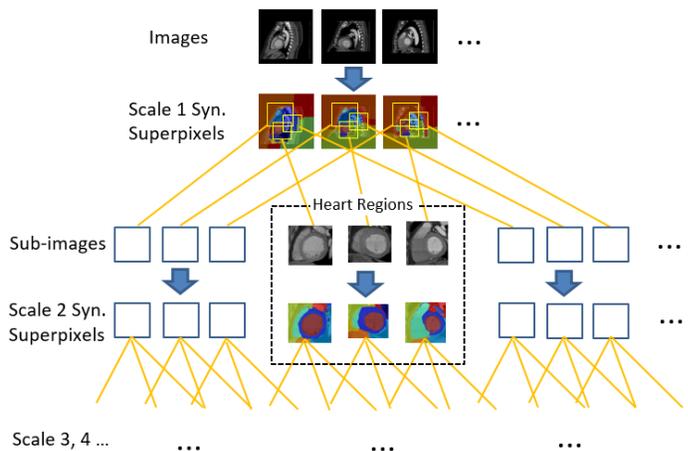


Fig. 8. The multi-scale hierarchical SSN decomposition. The synchronized superpixels (scale 1) can be sub-decomposed into finer synchronized superpixels (scale 2), and even finer scales (scale 3, 4,...).

synchronized superpixels. For each synchronized superpixel  $S$ , the corresponding image regions forms a set of subgraphs out of the original input images. We then apply the spectral decomposition and synchronization to these subgraphs, obtaining a new set of synchronized superpixels (Scale 2). The sub-decomposition can be carried out to even finer scales (Scale 3, 4, etc.) to obtain smaller synchronized superpixels for refined part-part correspondences. At each scale, the single-scale SSP algorithm (Algorithm 1) can be directly used for the generation of the SSP at that scale. The algorithmic form of the complete multi-scale SSP generation is presented as Algorithm 2.

**Algorithm 2:** Multi-scale Synchronized Superpixels Generation (MSSG)

- Input** : Scale  $L$ , # of SSP at each scale  $N_1, \dots, N_L$ , Image set  $\{I_1, \dots, I_M\}$
- Output:** Multiscale Superpixels  $\{S_{n_1, \dots, n_l}^l(m)\}_{l \in \{1, \dots, L\}, m \in \{1, \dots, M\}, n \in \{1, \dots, N\}}$
- 1) Initialize  $N_0 = 1, l = 0$ , let  $S^0(m) = I_m$  for all  $m \in \{1, \dots, M\}$ ;
  - 2) For each  $n \in 1, \dots, N_l$ , collect superpixels in group  $S_{n, n_{l-1}, \dots, n_1}^l = \{S_{n, n_{l-1}, \dots, n_1}^l(1), \dots, S_{n, n_{l-1}, \dots, n_1}^l(M)\}$  for all  $n_{l-1} \in \{1, \dots, N_{l-1}\}, \dots, n_1 \in \{1, \dots, N_1\}$ ;
  - 3) For each  $S_{n, n_{l-1}, \dots, n_1}^l$ , generate  $N_{l+1}$  superpixels  $\{S_{1, n, n_{l-1}, \dots, n_1}^{l+1}(1), \dots, S_{N_l, n, n_{l-1}, \dots, n_1}^{l+1}(1)\}, \dots, \{S_{1, n, n_{l-1}, \dots, n_1}^{l+1}(M), \dots, S_{N_l, n, n_{l-1}, \dots, n_1}^{l+1}(M)\}$  by Algorithm 1;
  - 4)  $l \leftarrow l + 1$  and go to step 2 if  $l < L$ ;

**C. Groupwise Heart Localization using Multiscale Synchronized Superpixels**

The synchronized superpixels naturally provide superpixel-superpixel correspondences for all images in the group, which can be used for simultaneous localization for similar objects among the images. This property can be employed for locating the heart regions from the raw cardiac scans.

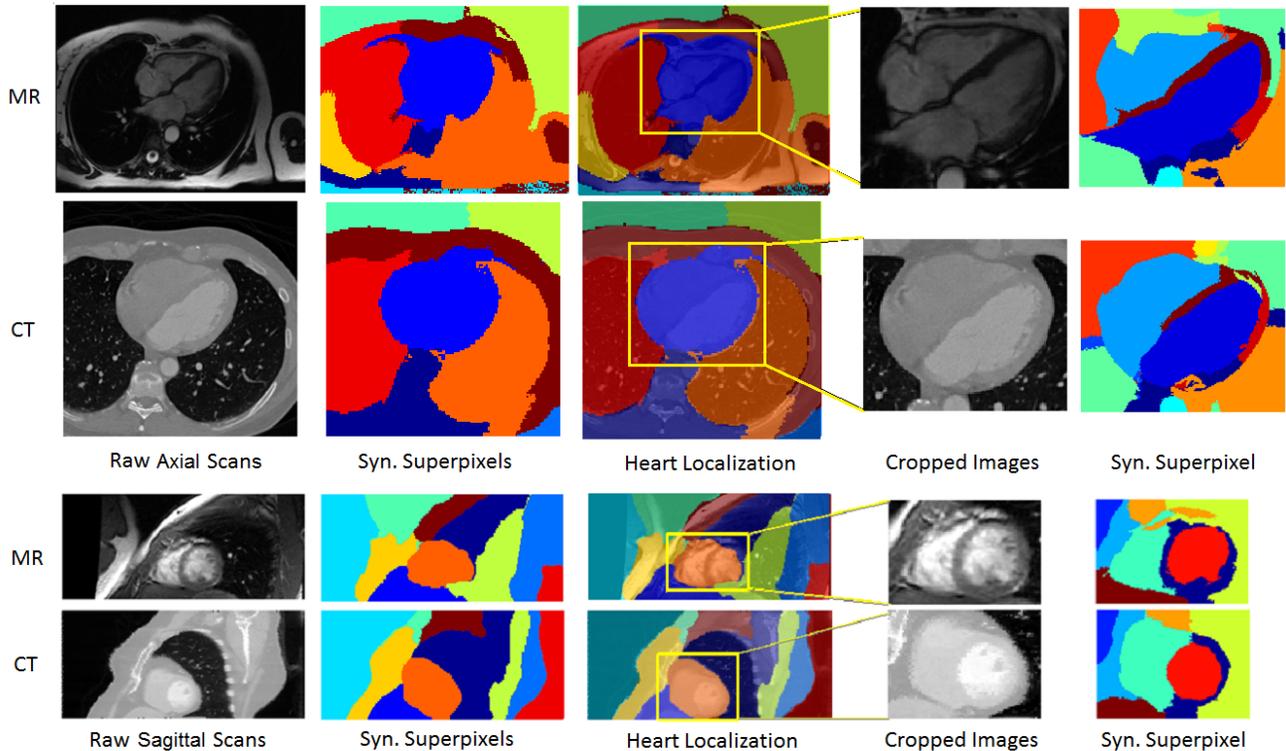


Fig. 10. Multiscale synchronized superpixels (two scales) for simultaneous heart localization in raw MR-CT images. Two examples are shown in this figure. Top: the example on multiscale decomposition and heart localization in raw axial MR and CT images. Bottom: the example on sagittal MR and CT images. The heart regions in these examples are simultaneously identified as the largest synchronized superpixels around the centers.

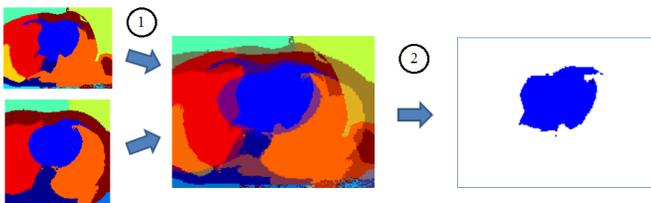


Fig. 9. Heart localization by synchronized superpixel maps. The heart regions are identified by: 1) stitching the two maps together and finding the overlapping regions that are sharing the same label; 2) locating the region stack near image center as the heart region stack.

Fig. 9 shows the heart region localization from two synchronized superpixels. The joint localization can be obtained by stacking the two superpixel maps (with rescaling) and finding the overlapped superpixels at the centers. Some existing heart localization methods such as maximum discrimination [13], circle fitting [30], Hough transform [31], template matching [32], can provide automatic LV detection but are restricted to short axis view only. Our method does not rely on a particular image view and can be applied to freeview scans.

For a group of raw cardiac scans (MR/CT, or both), the spectral synchronization and synchronized superpixels correlate the regions among the raw scans. As shown in Fig. 10, the synchronized superpixels at scale 1 provide part-part correspondences for regions in a MR image and a CT image. The correlated regions include the regions of the hearts, spine, and lung areas. Similarly, as shown in previous examples, the synchronized superpixels at scale 2 provide the correspondences between chamber regions. All heart regions

are simultaneously identified by stacking superpixels together and detecting the highest overlapping regions near the center.

## VII. EXPERIMENTS

Our segmentation is tested on one open and two self-collected  $M^3$  cardiac datasets covering MR/CT modalities, short-axis/long-axis (two/four-chamber) views, and various heart sizes/poses/field-of-view conditions. High dice metric ( $DM > 85\%$ ) is constantly achieved in almost all tests, which shows our segmentation is a reliable tool for analyzing the general  $M^3$  cardiac images.

### A. Datasets

The three datasets are organized as follows.

Dataset 1 ([2]): collected 93 subjects with 10300 short axis view (two-chamber) and horizontal long axis view (four-chamber) MR heart images, all manually cropped by experts according to their experience. All short axis view images are with resolution  $80 \times 80$ , obtained from 60 frames out of three different slice levels from the short axis stacks in a cardiac cycle. The four-chamber images are of  $100 \times 100$ , manually sampled from 25 frames in a cardiac cycle along the long axis.

Dataset 2: 64 subjects (32 MR + 32 CT) collected by our supportive hospitals. The MR images are all uncropped, with resolution  $192 \times 192$  and pixel spacing  $1.51 \times 1.51$  mm. The CT images are uncropped, with resolution  $512 \times 512$  and pixel spacing  $0.98 \times 0.98$  mm. The MR/CT in this set are extracted as upper body raw axial slices.

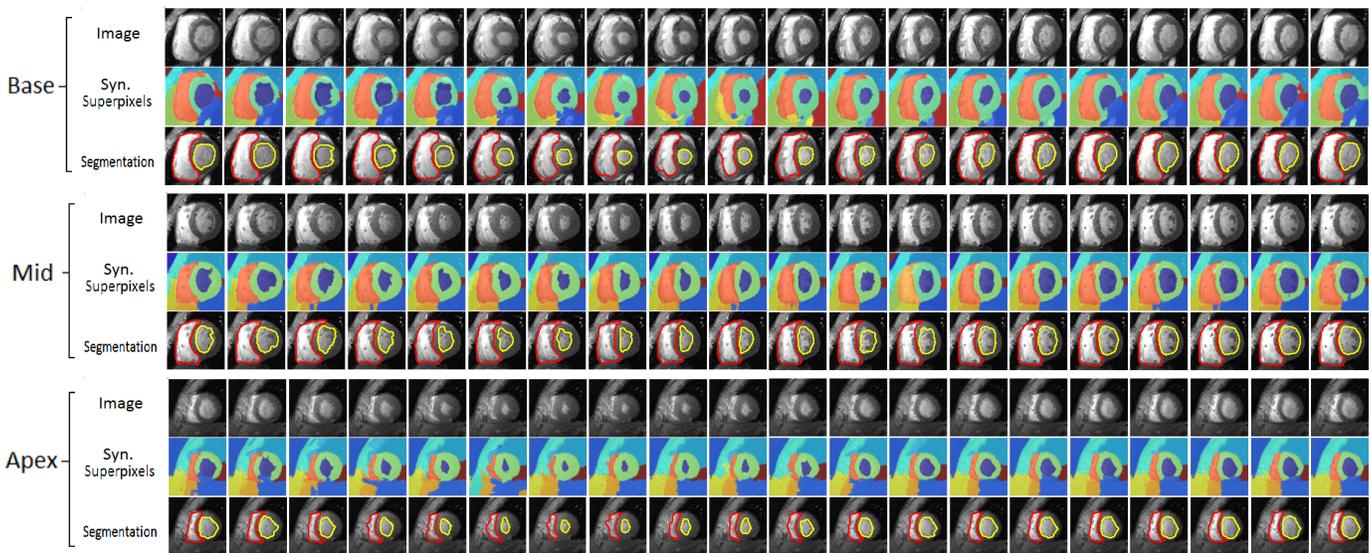


Fig. 11. (Spatial-Temporal Test) Groupwise segmentation for 60 regulated cropped short axis view (two-chamber) MR images from three slice levels (base, mid, and apex) of a whole cardiac cycle. The synchronized superpixels automatically correlate chamber regions among the images. The LV (**yellow contour**) and RV (**red contour**) are automatically identified. The total 60 images are jointly segmented using the incremental synchronization (Sec. V-B).

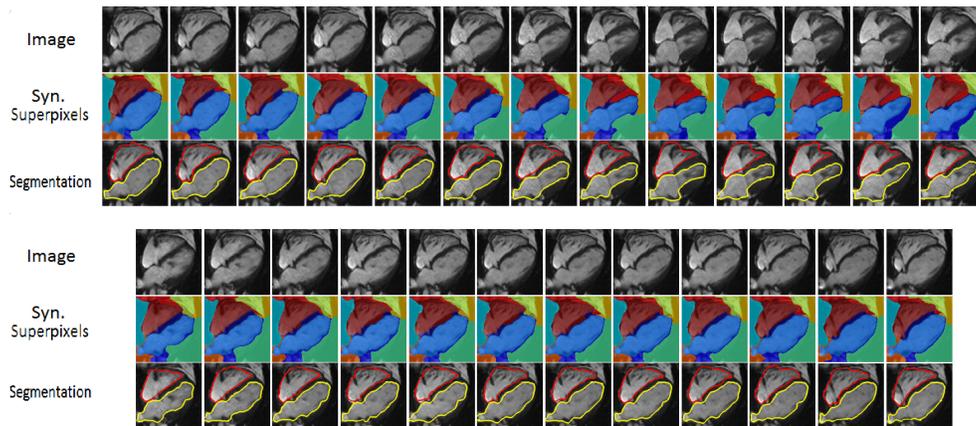


Fig. 12. Groupwise segmentation for 25 regularly cropped horizontal long axis view (four-chamber) MR images of a cardiac cycle. The LV+LA (**yellow contour**) and RV+RA (**red contour**) are automatically identified. The total 25 images are jointly segmented using the incremental synchronization (Sec. V-B).

Dataset 3: 64 subjects (32 MR + 32 CT). The MR data in this dataset is from York Cardiac MRI dataset <sup>1</sup> while the CT data is collected by our group. The MR images are with resolution  $256 \times 256$  and pixel spacing 0.93 to 1.65 mm, while the CT images are extracted from the  $512 \times 512 \times 456$  volumes. The LV groundtruth of the York MR Dataset is provided by the original dataset, and the RV groundtruth is manually obtained by our supportive expert. The MR/CT in this set are extracted as upper body sagittal slices.

### B. Regularly Cropped Short/Long-Axis View Images

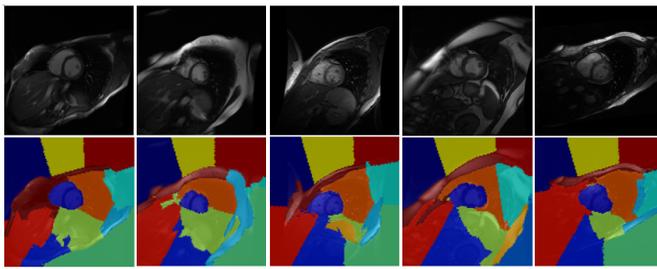
We first test the segmentation in a mostly used spatial-temporal/temporal diagnostic environment: segmentation for images from one full cardiac cycle in either two or four chamber view. In this scenario, we use the regularly cropped short-axis/long-axis view (two/four-chamber) MR images from different cardiac cycles (Dataset 1) to form segmentation groups.

The accuracy (avg. DM) and robustness (std) of our segmentation are evaluated. The resulting Table I shows that our segmentation can achieve 88% highest DM for cardiac cycle data of 93 subjects. This shows the unsupervised SSN segmentation is an accurate and convenient measurement tool for multi-view cardiac analysis.

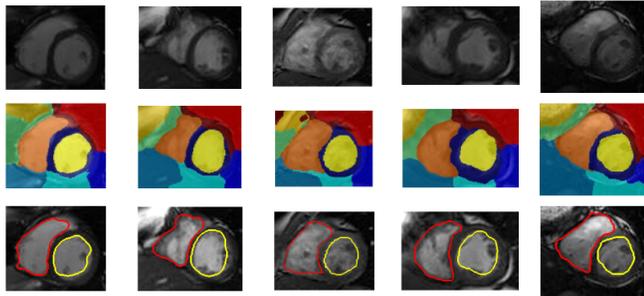
As reported in Table I, our segmentation obtain average DM 88.0% (LV) and 84.8% (RV) under short-axis two-chamber views, and has 87.6% (LV+LA) and 87.0% (RV+RA) under long-axis four-chamber views. During the test, for each subject, all 60 short-axis view (two-chamber) images are segmented using the incremental segmentation (Sec.V-B). Similarly, 25 four-chamber images for each subject are simultaneously segmented too. The overall performance in both two/four-chamber views are close. The difference of DM between both views is less than 3%, which proves that our freeview segmentation is insensitive to view changes.

A representative example is shown in Fig. 11, which shows a test for groupwise segmentation for 60 short axis view two-chamber images (one cardiac cycle of temporal resolution

<sup>1</sup><http://www.cse.yorku.ca/mridataset/>



(a) Scale 1 groupwise segmentation for raw MR scans.



(b) Scale 2 segmentation for automatically identified sub-images from (a).

Fig. 13. (Single Modality Test) Groupwise segmentation for raw MR scans with different heart poses/sizes. The scale 1 SSPs successfully segment the heart images, while scale 2 SSPs immediately provide heart region segmentation: LV (yellow contour), RV (red contour).

	LV(+LA)	RV(+RA)	All Chambers
Short Axis View	88.0±4.0	84.8±6.0	86.7±5.1
Horizontal Long Axis View	87.6±4.6	87.0±6.2	87.4±5.6

TABLE I

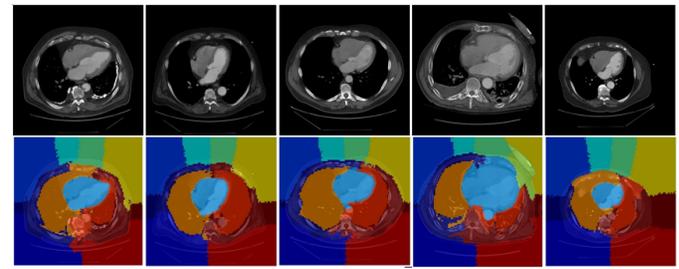
DICE METRIC (%) EVALUATIONS OF THE REGULATED-VIEW IMAGES (93 SUBJECTS IN DATASET 1).

20, with three slice levels: base, mid, and apex). All images in the test are already manually cropped to short/long-axis views to simulate the regulated clinical condition. The LV and RV regions are directly obtained from the superpixels and quantitative volume estimation is immediately available from Fig. 11. Similarly, Fig. 12 shows a test for groupwise segmentation for 25 images four-chamber images from a cardiac cycle (same subject). The synchronized superpixels directly provide the cardiac segmentation and volume estimation in regulated clinical condition.

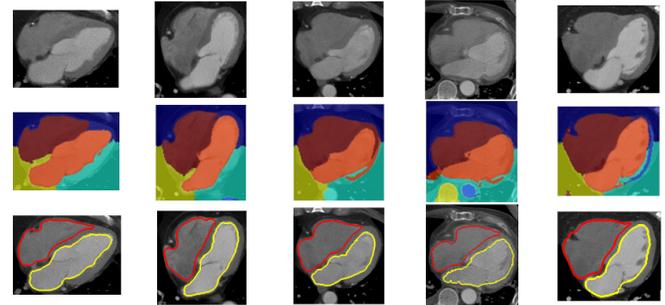
The same regulated configuration can be used to test the quantitative performance different segmentation properties. We test the improvement of the contour refinement of the proposed method in this paper (see Sec.VI A), and that of the pure groupwise segmentation (the same presented in [2]). The test is conducted on the York MR Cardiac Dataset (now part of the Dataset 3), and the results are summarized in Table II. As shown in the table, the SSP+refinement has higher accuracy than the pure SSP segmentation. The refinement is applied for all following tests unless stated otherwise.

### C. Non-regulated Raw Images: Single Modality

We then test the segmentation performance under a multi-subject comparative study environment. Our segmentation is tested on groups of single modality raw scans from Dataset 2 and 3. In total 30 upper body axial scans (15 MR + 15 CT,



(a) Scale 1 groupwise segmentation for raw CT scans.



(b) Scale 2 segmentation for automatically identified sub-images from (a).

Fig. 14. (Single Modality Test) Groupwise segmentation for raw axial cardiac CT scans with different heart poses/sizes. The scale 1 SSPs successfully segment the heart images, while scale 2 SSPs immediately provide heart region segmentation: LV (yellow contour), RV (red contour).

Representation	LV	RV	All Chambers
Syn. Superpixels (SSP) [2]	89.1±4.1	85.2±5.6	87.1±5.4
SSP+Refinement	91.6±3.6	87.6±5.1	90.4±4.7

TABLE II

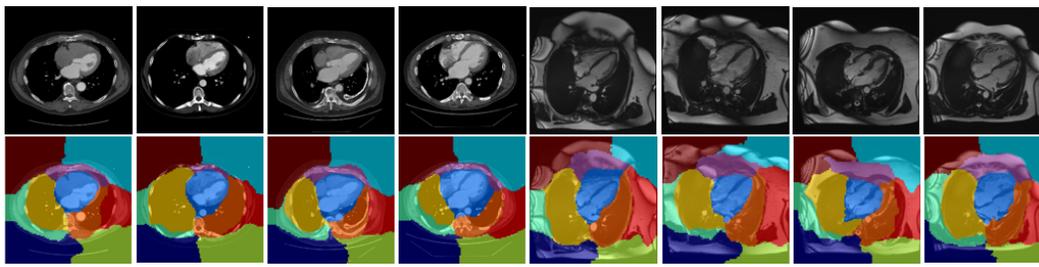
QUANTITATIVE COMPARISON (DM) OF DIFFERENT SEGMENTATION REPRESENTATION.

different subjects) from Dataset 2 and 30 upper body sagittal scans (15 MR + 15 CT, different subjects) from Dataset 3 are used. Those images are with different heart sizes, poses, institutional protocols and even different field of views. During this test, we rearrange the data into four sets: 15 MR axial scans, 15 CT axial scans, 15 MR sagittal scans, and 15 CT sagittal scans. Each set is further divided into three 5-image groups and processed by the groupwise segmentation. The overall results for LV(+LA) and RV(+RA) segmentations are reported in Table III and Table IV, which shows our segmentation can obtain average DM larger than 88%. The multiscale groupwise segmentation provides a convenient coarse-to-fine shape/volume comparative analysis across subjects directly from raw scans.

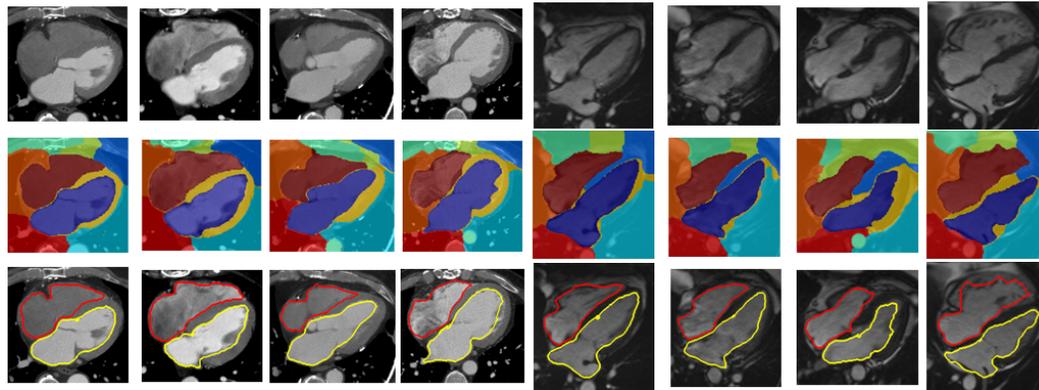
Fig. 13 shows a representative example of a 5-MR image group from Dataset 3, which contains 5 subjects with different sizes/poses. Using our two-scale groupwise segmentation (Algorithm 2), the LV-RV regions of the five MR images are automatically extracted, achieving average DM 89.2%. Similarly, Fig. 14 shows another groupwise segmentation of a 5-CT image group, with DM 90.1%. The segmentation performance of both examples are insensitive to the changes of subject poses and sizes.

### D. Non-regulated Raw Images: Cross-Modality

The segmentation is next tested under the multi-subject + multi-modality comparative study environment. We perform

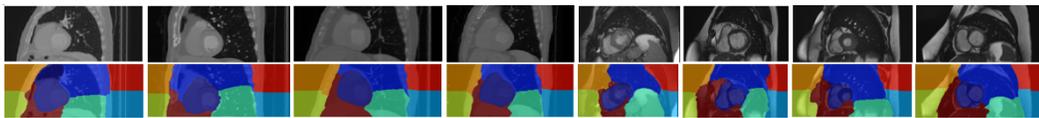


(a) Scale 1 groupwise segmentation for MR-CT mixed axial scans.

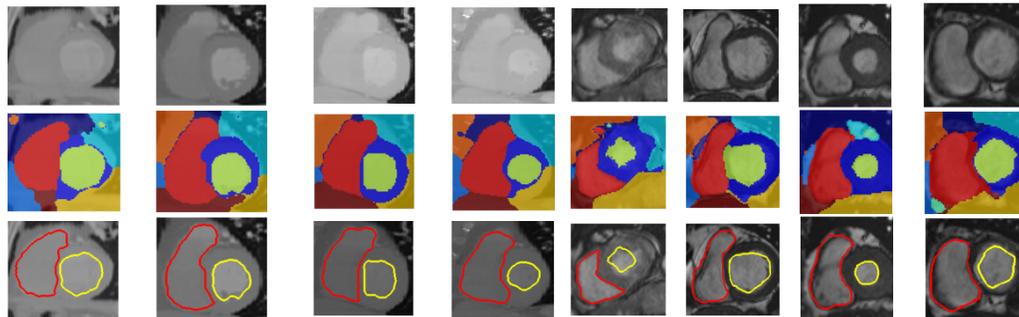


(b) Scale 2 groupwise segmentation for the MR-CT mixed heart sub-images automatically identified from (a).

Fig. 15. (Cross Modality Test) Two-scale groupwise segmentation from MR-CT mixed raw axial slices. (a) Successful groupwise segmentation for 4 CT + 4 MR different subjects under a non-regulated axial view; (b) groupwise segmentation for heart regions extracted from (a). The LV+LA (**yellow contour**) and RV+RA (**red contour**) regions are jointly identified after segmentation.



(a) Scale 1 groupwise segmentation for MR-CT mixed sagittal scans.



(b) Scale 2 groupwise segmentation for the MR-CT mixed heart sub-images automatically identified from (a).

Fig. 16. (Cross Modality Test) Two-scale groupwise segmentation from MR-CT mixed raw sagittal slices. (a) Successful groupwise segmentation for 4 CT + 4 MR different subjects under a non-regulated sagittal view; (b) groupwise segmentation for heart regions extracted from (a). The LV (**yellow contour**) and RV (**red contour**) regions are jointly identified after segmentation.

groupwise segmentation on uncropped images from Dataset 2 and 3, with varying degrees of MR-CT mixtures to simulate the cross-modality conditions. As shown in Table V and Table VI, the overall segmentation results (LV, RV, and LV+RV) remain constantly above DM 85% for most of the cases, except some rare cases in RV region segmentation lower results (83%) due to the region fuzziness. Thanks to the spectral features, our  $M^3$  segmentation is insensitive to the changes of modality/subject.

Particularly, all subjects from Dataset 2 and 3 are involved

in the tests to increase the data variety. Uncropped images are directly arranged in two originally acquired imaging views: the axial and sagittal views. For each view, the images are divided in four groups: 2 MR + 2 CT, 4 MR + 4 CT, 16 MR + 16 CT, and 32 MR + 32 CT to simulate different levels of modality mixtures. The MR dataset contains a variety of pathology problems including cardiomyopathy, aortic regurgitation, enlarged ventricles and ischemia as mentioned in [33]. However, no significant segmentation defeats are found during the test which shows our method is insensitive to structural

	4-image group			8-image group			16-image group			32-image group		
	LV+LA	RV+RA	All Cham.	LV+LA	RV+RA	All Cham.	LV+LA	RV+RA	All Cham.	LV+LA	RV+RA	All Cham.
MR	92.8 ±2.8	91.4 ±3.0	91.4 ±2.8	89.8 ±3.9	90.4 ±3.2	89.9 ±3.2	87.8 ±3.8	88.0 ±3.9	87.7 ±4.0	86.2 ±4.8	85.1 ±4.7	85.5 ±5.0
CT	88.7 ±3.9	86.5 ±3.3	87.7 ±3.6	87.0 ±3.9	87.9 ±3.8	87.1 ±4.0	87.1 ±4.3	85.2 ±4.9	86.9 ±4.0	84.2 ±5.7	82.9 ±5.2	84.0 ±5.3
MR+CT	91.0 ±3.7	89.1 ±3.5	89.3 ±3.1	87.9 ±4.0	88.1 ±3.7	87.2 ±3.5	87.5 ±3.9	87.1 ±4.9	87.0 ±4.1	85.7 ±5.1	84.1 ±5.0	84.5 ±5.0

TABLE V  
AVERAGE DM (%) AND STANDARD DEVIATION ON THE MR-CT MIXED RAW AXIAL SCANS (SEE EXAMPLES IN FIG. 15).

	4-image group			8-image group			16-image group			32-image group		
	LV	RV	All Cham.	LV	RV	All Cham.	LV	RV	All Cham.	LV	RV	All Cham.
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	86.8 ±3.8	85.9 ±4.8	86.4 ±4.5	85.2 ±4.9	85.8 ±4.5	86.0 ±4.7
CT	89.7 ±2.9	87.7 ±3.6	88.2 ±3.5	86.2 ±4.6	84.9 ±4.7	85.8 ±5.1	86.7 ±4.9	85.0 ±5.0	85.9 ±4.5	86.2 ±4.7	83.1 ±5.1	83.8 ±5.1
MR+CT	90.8 ±3.3	89.8 ±3.5	89.2 ±3.2	87.9 ±4.0	89.1 ±3.7	87.2 ±3.9	87.4 ±4.3	85.1 ±4.9	86.2 ±4.1	85.9 ±5.1	84.6 ±5.0	84.2 ±5.2

TABLE VI  
AVERAGE DM (%) AND STANDARD DEVIATION ON THE MR-CT MIXED RAW SAGITTAL SCANS (SEE EXAMPLES IN FIG. 16).

	LV	RV	All Chambers
MR (15 subjects)	91.5±2.8	89.0±3.2	90.2±3.1
CT (15 subjects)	90.1±2.7	88.3±4.0	89.1±3.8

TABLE III  
AVERAGE DM (%) ON SHORT-AXIS SINGLE MODALITY DATA.

	LV+LA	RV+RA	All Chambers
MR (15 subjects)	90.2±3.2	88.0±3.7	89.2±3.3
CT (15 subjects)	89.1±3.7	87.3±4.1	88.2±3.8

TABLE IV  
AVERAGE DM (%) ON LONG-AXIS SINGLE MODALITY DATA.

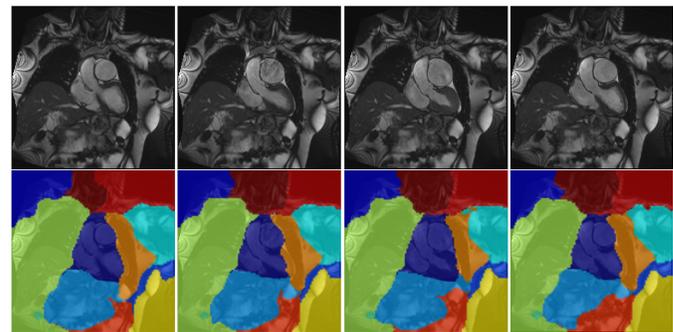
abnormalities. Note that one randomly selected subject from York MR Dataset (now Dataset 3) is removed for the equal pair-up of the MR-CT joint segmentation. The final results are calculated by taking the average DM over all images in the same group.

A representative example for this test is presented in Fig. 15. This example shows a 4 MR + 4 CT mixed image group with different heart poses/sizes under the non-cardiac axial imaging view. Our unsupervised groupwise segmentation has results in successful extraction of the LV+LA and RV+RA regions in the eight images, achieving average DM 88.2%. Similarly, Fig. 16 presents a successful segmentation on 8 MR-CT mixed raw sagittal scans. The full numerical results on axial scans are reported in Table V and Table VI. The overall results for axial and sagittal scans are insensitive to view and modality changes. This proves the high versatility of our method in different clinical conditions.

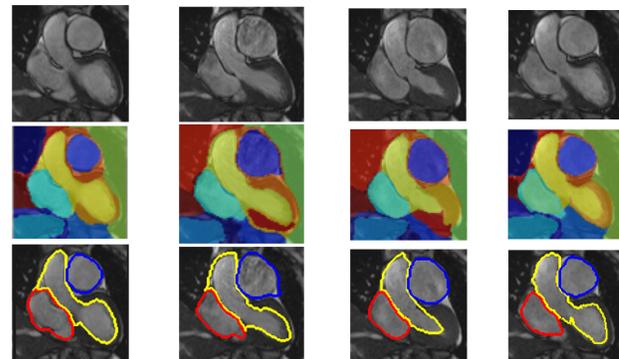
### E. Non-regulated Raw Images: Non-diagnostic View

We finally test the segmentation on even less regulated, non-diagnostic view images, to evaluate the performance under extreme conditions. The test images are from Dataset 2 and 3 under a non-diagnostic view. A representative example of this test is shown in Fig. 17. Our unsupervised segmentation can achieve average DM 89.0% in this case, which provides a highly flexible tool for arbitrary customized scenarios.

As Fig. 17 shows, four MR coronal view upper body scans including the most significant frames of end-diastolic and end-systolic stages are used in the test. Our scale 1 groupwise segmentation successfully synchronizes the MR scans and



(a) Scale 1 groupwise segmentation for coronal images of a cardiac cycle.



(b) Scale 2 groupwise segmentation for identified sub-images from (a).

Fig. 17. Groupwise segmentation for non-diagnostic, freely chosen coronal MR scans. (a): The scale 1 SSPs successfully segmented and identified the hearts. (b): The LV+AO (yellow), RV (red), and PA (blue) are jointly obtained by scale 2 SSPs.

generates 10 synchronized superpixels. The heart locations are jointly identified, generating 8 new synchronized superpixels for the cropped images. Then in the scale 2 groupwise segmentation, the LV, RV, and PA regions are jointly identified from the synchronized superpixels. Segmentation on other views are immediately available by following the same process.

### F. Implementation Details and Computation Time

The segmentation is implemented in Matlab, and performed on Intel Core i7 CPU PC for small image group and also the SHARCNET platform (<http://www.sharcnet.ca>) for large image group. The optimization computation (Eqn. 7 and Eqn. 8) are implemented by the optimization toolbox in Matlab. Our

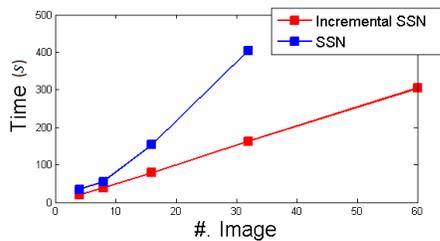


Fig. 18. The comparison between the computation time of SSN synchronization (Sec. V-A) and incremental SSN synchronization (Sec. V-B).

unoptimized implementation of SSN synchronization is very efficient, as most of the cases will converge in less than 100 iterations with default settings.

**Implementation Comparison.** We also test the implementation of SSN synchronization (Eqn. 7) and the incremental synchronization (Eqn. 8) in Fig. 18. For a randomly selected group of 60 images sampled from Dataset 1, the incremental synchronization is initially synchronized with a 4-image set in either two-chamber short axis or four-chamber long axis images. The DM accuracies on both implementation are very close ( $\pm 3\%$ ). For the same 60-image group, the SSN non-incremental synchronization will take longer than 40 min to converge, while incremental synchronization only need less than 5.2s per image using our matlab implementation. Incremental method has linear time growth and thus more suitable for large image set synchronization and segmentation.

## VIII. DISCUSSION

Our spectral synchronization method has many new properties that improve the performances and functionalities of traditional spectral-based segmentation methods and existing cardiac segmentation methods.

### A. SSN v.s. Traditional Spectral Segmentations

**Enhanced Feature Clustering.** Traditional spectral-based methods (or referred as random-walk methods in some literatures) [26] [38] utilize the principal decomposition of spectral graph matrix to provide robust image feature clustering and image segmentation. Compared to these traditional methods, our SSN-based method not only preserves the robust spectral clustering of within-image features like contours/textures, but automatically builds up the correspondences of these features across images. These new correspondences are established by borrowing the spectral clustering of the whole image group. The SSN-correlated spectral clustering, in turn, enhances the feature clustering of each image and finally improves the segmentation quality.

**Co-segmentation without Explicit Matching.** In addition, compared to other spectral groupwise segmentation methods (or referred as co-segmentation) [43], SSN-based image correspondences are modality independent and do not require exact matching such as alignments of SIFT/intensity features between images. This enables our SSN model to perform groupwise segmentation across different modalities that have diverse image intensities. The cross image correspondence also constitutes explicit correlations at multiple superpixel scales, which can be considered as a groupwise registration

at multiple scales. Specific superpixels can thus be jointly identified and extracted, which provides automatic groupwise analysis for these target regions.

### B. SSN Groupwise Segmentation v.s. Existing Cardiac Segmentations

**Success/Failure of Groupwise Segmentation.** For all the tests we conducted, no fail cases are found in the groupwise segmentations. However, one can imagine that when the groupwise assumption is not met the proposed approach would fail. One extreme case would be the mixture of the test image group with images from incompatible different views, i.e. a short axis view image group mixed with long axis view images. Nevertheless, in practice, it has no application to put different view images in correspondence thus this failure can be avoided.

**M<sup>3</sup> Segmentation.** Compared to the existing cardiac segmentation methods [33] [6] [14] [19] [20] [34] [35] [36] [37] [10], our approach is the first to provide groupwise segmentation with region correspondences in multi-modality, multi-subject, multi-chamber situation for different regular/non-regular views. The overall comparison is presented in Table VII. It is clear from the table that, after testing the largest datasets (221 subjects) so far, not only our performance in LV/RV segmentation is very competitive to existing major segmentation methods, but we also provide a more general and versatile solution for almost all clinical conditions. For the traditional “single modality + single chamber” segmentation tasks, our groupwise segmentation is still among the top performance methods. In Table VIII, we present the quantitative comparative results of our method and the top-ranked methods on MICCAI LV Segmentation Challenge Dataset <sup>2</sup>. The groupwise setting in this case is similar to those in Sec.VII-B and Sec.VII-C, where a 5-image set out of 20 cardiac cycle images is first randomly chosen for groupwise segmentation then incrementally propagates to the whole cycle. Our method has the best all phases performance (Endo+Epi) in training set, and is tight to the best method in validation set. Illustrative examples of this dataset are presented in Fig. 19, which shows a groupwise segmentation with 93.2% DM accuracy.

**Freewiew Segmentation.** As can be observed from Table VII, unlike our segmentation method, all methods except [8] are restricted to the segmentation of well-cropped short-axis/long-axis (two/four-chamber) views. However, the method proposed in [8] still requires the manual building of image atlas and has not shown its ability to handle LV/RV chamber segmentation. Our cross-modality whole heart segmentation can be done without atlas/trainings. It makes the complete cardiac segmentation process fully automatic, and is more suitable for the segmentation tasks of today’s large scale M<sup>3</sup> image sets.

**Region-to-Region Correspondences.** We propose the novel unsupervised region-to-region correspondences between MR and CT images in cardiac segmentation in this paper. Although some multi-atlas-based methods (i.e., [8] [10]) claimed they can work on both modalities, the image correspondences

<sup>2</sup>[http://smial.sri.utoronto.ca/LV\\_Challenge/](http://smial.sri.utoronto.ca/LV_Challenge/)

Method	Modality	View	Chamber	Dataset	Performance (metric / result)
Andreopoulos <i>et al</i> [33] (2008)	MR	Short axis	LV	33 subjects	Endo volumetric error / $1.43 \pm 0.49$ mm Epi volumetric error / $1.51 \pm 0.48$ mm
Zheng <i>et al</i> [6] (2008)	CT	Horizontal long axis	LV, LA, RV, RA	137 subjects	LA point-to-mesh / $1.13 \pm 0.55$ mm RA point-to-mesh / $1.57 \pm 0.48$ mm LV-Endo point-to-mesh / $0.98 \pm 1.32$ mm LV-Epi point-to-mesh / $0.82 \pm 1.07$ mm RV point-to-mesh / $0.84 \pm 0.94$ mm
Ecabert <i>et al</i> [14] (2008)	CT	Horizontal long axis	LV, LA, RV, RA	108 subjects	LA point-to-surface / $0.71 \pm 0.88$ mm RA point-to-surface / $0.89 \pm 0.96$ mm LV-Endo point-to-surface / $0.98 \pm 1.32$ mm LV-Epi point-to-surface / $0.82 \pm 1.07$ mm RV point-to-surface / $0.84 \pm 0.94$ mm
Ben Ayed <i>et al</i> [19] (2009)	MR	Short axis	LV	10 subjects	Cavity DM / $88 \pm 9$ % Myocardium DM / $81 \pm 10$ %
Isgum <i>et al</i> [8] (2009)	CT	Axial view	Whole heart, Aortic	29 subjects	Tanimoto Coefficient $0.8847 \pm 0.0331$
Zhang <i>et al</i> [20] (2010)	MR	Short axis	LV, RV	50 subjects	LV-Endo point-to-surface $1.67 \pm 0.3$ mm LV-Epi point-to-surface / $1.81 \pm 0.4$ mm RV-Endo point-to-surface / $2.13 \pm 0.39$ mm
Ben Ayed <i>et al</i> [34] (2012)	MR	Short axis	LV	20 subjects	Cavity DM / $92 \pm 3.1$ % Myocardium DM / $82 \pm 6.1$ %
Nambakhsh <i>et al</i> [35] (2013)	MR	Short axis	LV	20 subjects	Cavity DM / $92 \pm 7$ % Myocardium DM / $80 \pm 10$ %
Queiros <i>et al</i> [36] (2014)	MR	Short axis	LV	45 subjects	Endo DM $93 \pm 3$ % Epi DM $94 \pm 2$ %
Ringenberg <i>et al</i> [37] (2014)	MR	Short axis	RV	16 subjects	Endo DM $88 \pm 11$ % Epi DM $90 \pm 8$ %
Bai <i>et al</i> [10] (2015)	MR	Short axis	LV	83 subjects	Method 1 DM $81.2$ % Method 2 DM $81.5$ %
<b>Our Method</b>	<b>MR, CT</b>	<b>Freeview</b> (short/long-axis, axial/sagittal views...)	<b>LV(+LA), RV(+RA)</b> <b>Whole heart</b>	<b>221 subjects</b> (157 MR, 64 CT)	<b>DM &gt; 85 % for almost all situations</b> (See Table I to VI)

TABLE VII  
COMPARISON WITH MAJOR CARDIAC SEGMENTATION METHODS DEVELOPED IN RECENT YEARS.

	Huang <i>et al</i> [30]	Lu <i>et al</i> [39]	O'Brien <i>et al</i> [41]	Jolly [40]	Constantinides <i>et al</i> [42]	Casta <i>et al</i> [32]	Wijnhout <i>et al</i> [31]	Ours
Training (15 cases)	90±4 (Endo) 93±2 (Epi)	\	80 (Endo) 91.3 (Epi)	88.1±5.7 (Endo) 93.5±1.4 (Epi)	88.4±10.2 (Endo) 92.9±6.5 (Epi)	\	\	<b>90.8±2.6 (All)</b>
Validation (15 cases)	89±4 (Endo) 94±1 (Epi)	89±3 (All)	(same as above)	87.9±3.2 (Endo) 93.3±1.8 (Epi)	92.3±6.1 (Endo) 92.2±5.0 (Epi)	92.7 (All)	89±3 (All)	<b>92.3±3.8 (All)</b>

TABLE VIII  
QUANTITATIVE COMPARISON OF AVERAGE DM (%) AND STANDARD DEVIATION ON THE MICCAI LV CHALLENGE DATASET.

cannot be automatically obtained from the manually built atlas. Instead, our method can automatically build up the correspondences, and an accurate MR-CT registration can also be done by simply aligning the identified chamber regions from our synchronized superpixels. For single subject study, this enables the cross-modality comparative measurement/diagnosis for different cardiac problems. For multiple subject study, particularly in the big data environment, this enables a more comprehensive and non-biased statistical analysis for cardiac data obtained from different modalities and protocols.

## IX. CONCLUSIONS

We proposed an unsupervised groupwise segmentation for general cardiac images. Our method can provide simultaneous segmentation for a group of cardiac images with multiple modalities, multi-chamber, and multiple subjects images. The segmentation is based on the Synchronized Spectral Network (SSN) model, which correlates the modality independent spectral features across images and clusters the features into synchronized superpixels. The heart regions from different

images can be identified simultaneously using synchronized superpixels and chambers can then be extracted as the segmentation results. Our segmentation has accurate and robust results (DM > 85%) for uncropped scans, regulated short-axis/long-axis (four-chamber) images, even non-regulated images. It provides a general algorithmic framework for today's cardiac segmentation tasks.

## REFERENCES

- [1] Z. Wang, X. Zhen, K. Tay, S. Osman, W. Romano, and S. Li, "Regression segmentation for m3 spinal images," *IEEE TMI*, 2015.
- [2] Y. Cai, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Unsupervised freeview groupwise segmentation for m3 cardiac imaged using synchronized spectral network," in *MICCAI*. Springer, 2015.
- [3] N. Paragios, "A variational approach for the segmentation of the left ventricle in cardiac image analysis," *IJCV*, vol. 50, pp. 345–362, 2002.
- [4] M. Lynch, O. Ghita, and P. F. Whelan, "Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge," *Comput. Medical Imaging and Graphics*, vol. 30, no. 4, pp. 255–262, 2006.
- [5] —, "Segmentation of the left ventricle of the heart in 3-d+ t mri data using an optimized nonrigid temporal model," *IEEE TMI*, vol. 27, no. 2, pp. 195–203, 2008.

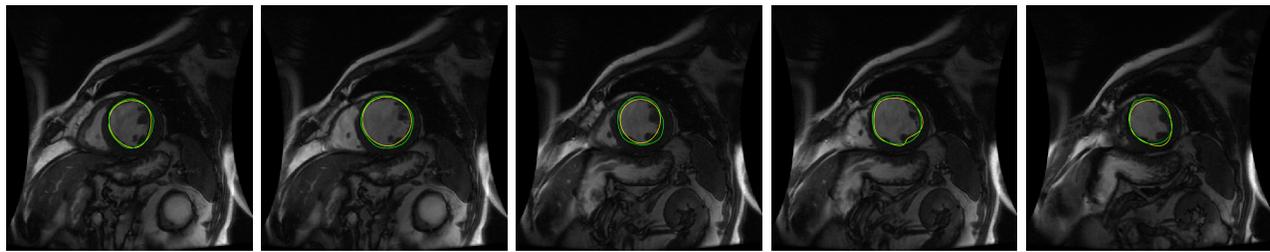


Fig. 19. Examples of groupwise segmentation results (5-image set) in MICCAI LV Challenge Dataset. Our final contour results (**yellow contours**) are very close to the ground truths (**green contours**). The average DM is 93.2% for this image set.

[6] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuring, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features," *IEEE TMI*, vol. 27, no. 11, pp. 1668–1681, 2008.

[7] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert, "Segmentation of 4d cardiac mr images using a probabilistic atlas and the em algorithm," *MedIA*, vol. 8, no. 3, pp. 255–265, 2004.

[8] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion application to cardiac and aortic segmentation in ct scans," *IEEE TMI*, vol. 28, no. 7, pp. 1000–1010, 2009.

[9] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A. Viergever, J. P. Pluim, and B. van Ginneken, "Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus," *MedIA*, vol. 14, no. 1, pp. 39–49, 2010.

[10] W. Bai, W. Shi, C. Ledig, and D. Rueckert, "Multi-atlas segmentation with augmented features for cardiac mr images," *MedIA*, vol. 19, no. 1, pp. 98–109, 2015.

[11] D. Mahapatra, Z. Li, P. Schöffler, F. Vos, and J. Buhmann, "Joint segmentation and groupwise registration of cardiac dce mri using sparse data representations," *In Proc. IEEE ISBI*, pp. 1312–1315, 2015.

[12] K. Mcleod, C. Seiler, M. Sermesant, and X. Pennec, "Spatio-temporal dimension reduction of cardiac motion for group-wise analysis and statistical testing," in *MICCAI*. Springer, 2013, pp. 501–508.

[13] M.-P. Jolly, "Automatic segmentation of the left ventricle in cardiac mr and ct images," *IJCV*, vol. 70, no. 2, pp. 151–163, 2006.

[14] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. Von Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanian, G. Lavi *et al.*, "Automatic model-based segmentation of the heart in ct images," *IEEE TMI*, vol. 27, no. 9, pp. 1189–1201, 2008.

[15] J. Peters, O. Ecabert, C. Meyer, H. Schramm, R. Kneser, A. Groth, and J. Weese, "Automatic whole heart segmentation in static magnetic resonance image volumes," in *MICCAI*. Springer, 2007, pp. 402–410.

[16] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *MedIA*, vol. 15, no. 2, pp. 169–184, 2011.

[17] J. Cousty, L. Najman, M. Couprie, S. Clément-Guinaudeau, T. Goissen, and J. Garot, "Segmentation of 4d cardiac mri: Automated method based on spatio-temporal watershed cuts," *Image and Vision Computing*, vol. 28, no. 8, pp. 1229–1243, 2010.

[18] A. Pednekar, U. Kurkure, R. Muthupillai, S. Flamm, and I. A. Kakadiaris, "Automated left ventricular segmentation in cardiac mri," *IEEE Trans. on Biomed. Eng.*, vol. 53, no. 7, pp. 1425–1428, 2006.

[19] I. Ben Ayed, S. Li, and I. Ross, "Embedding overlap priors in variational left ventricle tracking," *IEEE TMI*, vol. 28, no. 12, pp. 1902–1913, 2009.

[20] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-d cardiac mr image analysis: left and right ventricular morphology and function," *IEEE TMI*, vol. 29, no. 2, pp. 350–364, 2010.

[21] D. Mahapatra and J. M. Buhmann, "Cardiac lv and rv segmentation using mutual context information," in *Machine Learning in Medical Imaging*. Springer, 2012, pp. 201–209.

[22] B. P. Lelieveldt, M. Üzümcü, R. J. van der Geest, J. H. Reiber, and M. Sonka, "Multi-view active appearance models for consistent segmentation of multiple standard views: application to long-and short-axis cardiac mr images," in *International Congress Series*, vol. 1256. Elsevier, 2003, pp. 1141–1146.

[23] J. Lötjönen, S. Kivistö, J. Koikkalainen, D. Smutek, and K. Lauerma, "Statistical shape model of atria, ventricles and epicardium from short-and long-axis mr images," *MedIA*, vol. 8, no. 3, pp. 371–386, 2004.

[24] M. Sermesant, C. Forest, X. Pennec, H. Delingette, and N. Ayache, "Deformable biomechanical models: Application to 4d cardiac image analysis," *MedIA*, vol. 7, no. 4, pp. 475–488, 2003.

[25] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*, 2001.

[26] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *CVPR*, vol. 2. IEEE, 2005, pp. 1124–1131.

[27] D. Eynard, K. Glashoff, M. M. Bronstein, and A. M. Bronstein, "Multimodal diffusion geometry by joint diagonalization of laplacians," *IEEE TPAMI*, 2015.

[28] A. Tsai, A. Yezzi Jr, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE TMI*, vol. 22, no. 2, pp. 137–154, 2003.

[29] J. Shen, S. H. Kang, and T. F. Chan, "Euler's elastica and curvature-based inpainting," *SIAM Journal on Applied Mathematics*, vol. 63, no. 2, pp. 564–592, 2003.

[30] S. Huang, J. Liu, L. Lee, S. Venkatesh, L. Teo, C. Au, and W. Nowinski, "Segmentation of the left ventricle from cine mr images using a comprehensive approach," *MIDAS J-Card MR Left Ventricle Segmentation Challenge*, 2009.

[31] J. Wijnhout, D. Hendriksen, H. Assen, and R. der Geest, "Lv challenge lkeb contribution: Fully automated myocardial contour detection," *The MIDAS Journal*, vol. 43, p. 2, 2009.

[32] C. Casta, P. Clarysse, J. Schaerer, and J. Pousin, "Evaluation of the dynamic deformable elastic template model for the segmentation of the heart in mri sequences," *MIDAS J-Card MR Left Ventricle Segmentation Challenge*, 2009.

[33] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

[34] I. B. Ayed, H.-m. Chen, K. Punithakumar, I. Ross, and S. Li, "Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure," *Medical image analysis*, vol. 16, no. 1, pp. 87–100, 2012.

[35] C. M. Nambakhsh, J. Yuan, K. Punithakumar, A. Goela, M. Rajchl, T. M. Peters, and I. B. Ayed, "Left ventricle segmentation in mri via convex relaxed distribution matching," *Medical image analysis*, vol. 17, no. 8, pp. 1010–1024, 2013.

[36] S. Queirós, D. Barbosa, B. Heyde, P. Morais, J. L. Vilaça, D. Friboulet, O. Bernard, and J. Dhooge, "Fast automatic myocardial segmentation in 4d cine cmr datasets," *Medical image analysis*, vol. 18, no. 7, pp. 1115–1131, 2014.

[37] J. Ringenber, M. Deo, V. Devabhaktuni, O. Berenfeld, P. Boyers, and J. Gold, "Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac mri," *Computerized Medical Imaging and Graphics*, vol. 38, no. 3, pp. 190–201, 2014.

[38] L. Grady, "Random walks for image segmentation," *TPAMI*, vol. 28, no. 11, pp. 1768–1783, 2006.

[39] Y. Lu, P. Radau, K. Connelly, A. Dick, and G. Wright, "Automatic image-driven segmentation of left ventricle in cardiac cine mri," *The MIDAS Journal*, vol. 49, p. 2, 2009.

[40] M. Jolly, "Fully automatic left ventricle segmentation in cardiac cine mr images using registration and minimum surfaces," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 4, 2009.

[41] S. O'Brien, O. Ghita, and P. F. Whelan, "Segmenting the left ventricle in 3d using a coupled asm and a learned non-rigid spatial model," 2009.

[42] C. Constantinides, Y. Chenoune, N. Kachenoura, E. Rouillot, E. Mousseaux, A. Herment, and F. Frouin, "Semi-automated cardiac segmentation on cine magnetic resonance images using gvf-snake deformable models," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, 2009.

[43] M. D. Collins, J. Xu, L. Grady, and V. Singh, "Random walks based multi-image segmentation: Quasiconvexity results and gpu-based solutions," in *CVPR*. IEEE, 2012, pp. 1656–1663.