Learning Shape Representations for Multi-Atlas Endocardium Segmentation in 3D Echo Images

O. Oktay, W. Shi, K. Keraudren, J. Caballero, and D. Rueckert
Biomedical Image Analysis group, Imperial College London, UK

Abstract

The analysis of left ventricular (LV) mass and function are of particular interest to clinicians for diagnostic and therapeutic purposes, commonly relying on 3D echocardiography due to its low cost and high temporal resolution. However, the automatic delineation of the endocardial boundary remains a challenging task due to low image quality.

We present a multi-atlas segmentation framework to delineate the LV endocardium in echocardiographic images. To increase the robustness of the registration step, a speckle reduction step and a new spectral representation are introduced, which are based on sparse coding and manifold learning. The spectral representation, unlike intensity values, provides consistent structural information across different images.

Speckle Reduction

Dictionary learning and sparse coding based speckle denoising technique is used to increase signal-to-noise ration in echo images, particularly K-SVD algorithm is used.

Image patches $\mathbf{y}_n \in \mathbb{R}^P$ are approximated as sparse combinations $\mathbf{x}_n \in \mathbb{R}^M$ of atoms from an over-complete dictionary $\mathbf{C} \in \mathbb{R}^{P\times M}$, namely solving:

$$\min_{\mathbf{C},\mathbf{X}} \sum_{n=1}^{N} \|\mathbf{x}_n\|_0 \quad \forall n\,, \|\mathbf{y}_n - \mathbf{C}\mathbf{x}_n\|^2 \leq \epsilon \in \mathbb{R}_+ .$$

Spectral Image Representation

Spectral Representation

- ⇒ Dense local image features.
- ⇒ Structural information is captured.
- ⇒ Features are computed with Laplacian Eigenmaps.
- ⇒ Particularly useful in multi-modal image registration[1] and ultrasound image representation where intensity encoding does not provide enough information for registration.

Dictionary Based Spectral Representation

- ⇒ Computationally much more effective.
  - No need for computation of large Laplacian matrix ($10^6 \times 10^6$ elements)
- ⇒ Single dictionary can be used to map all images to the same lower dimensional embedding space.
- ⇒ Mapping to the manifold space is done by computing linear codes of image patches, particularly locality constrained sparse coding[2] is used:

$$\min_{\mathbf{X}} \sum_n \|\mathbf{y}_n - \mathbf{C}\tilde{\mathbf{x}}_n\|^2 + \lambda \|\mathbf{b}_n \odot \tilde{\mathbf{x}}_n\|^2$$

$$\text{s.t.} \quad \forall n\,, \mathbf{1}^\top \tilde{\mathbf{x}}_n = 1$$

$$b_{(n,m)} = \exp\left(\|(\mathbf{y}_n - \mathbf{c}_m\|^2 / \sigma \right)$$

Spectral embedding of image patches is an unsupervised learning method to generate a structural representation from medical images. For a given input image $I \in \mathbb{R}^n$, a spectral representation $\mathcal{S} \in \mathbb{R}^{(n+1)}$ is generated by performing non-linear dimensionality reduction on image patches.

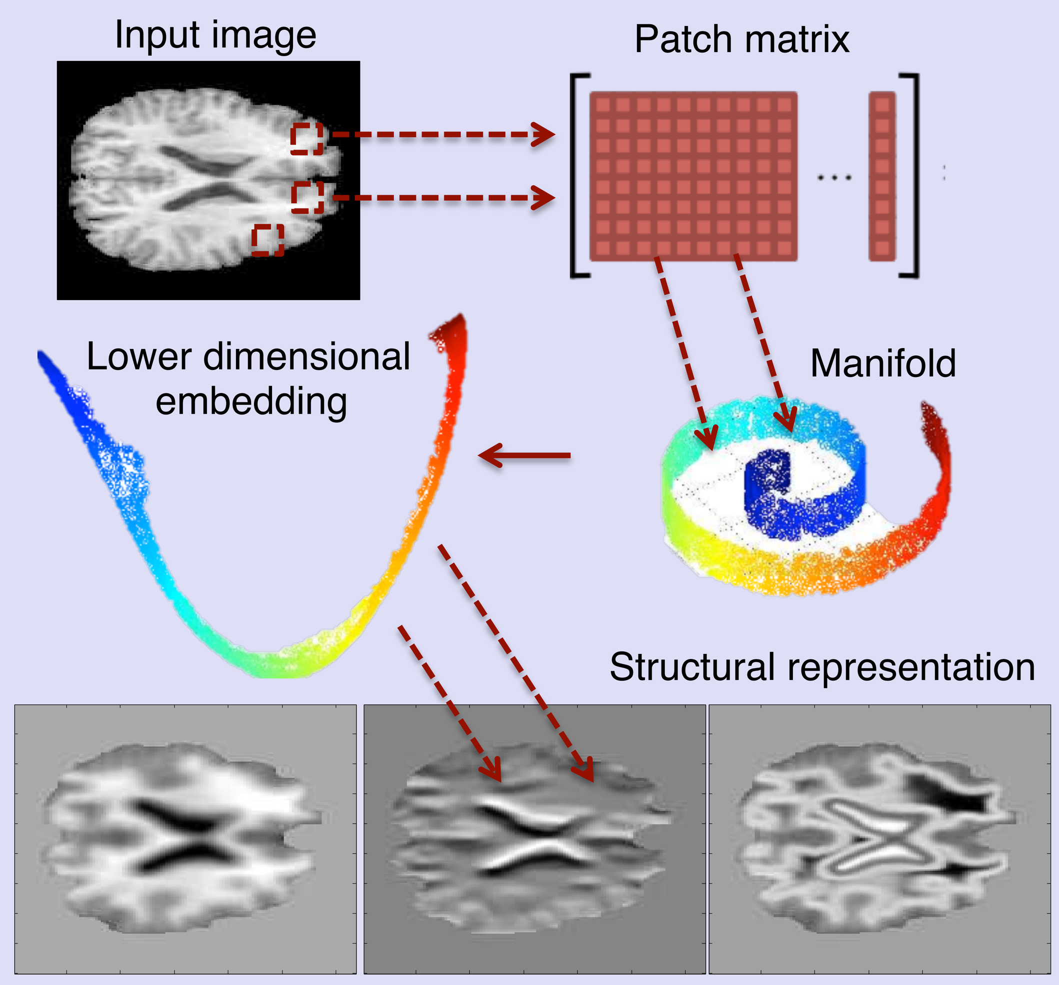


Figure 1: Spectral embedding of image patches

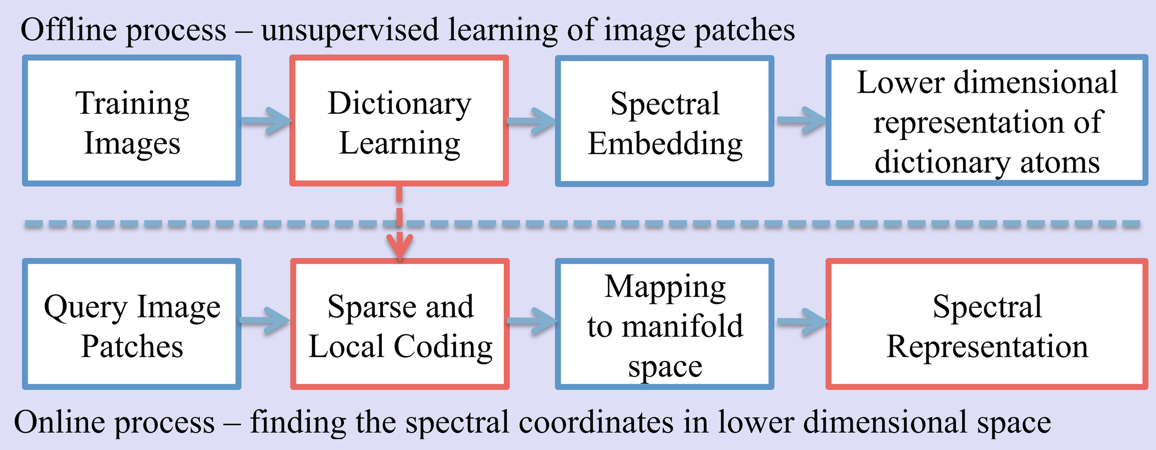


Figure 2: Block diagram of the proposed spectral representation.

Multi-Atlas Segmentation Framework

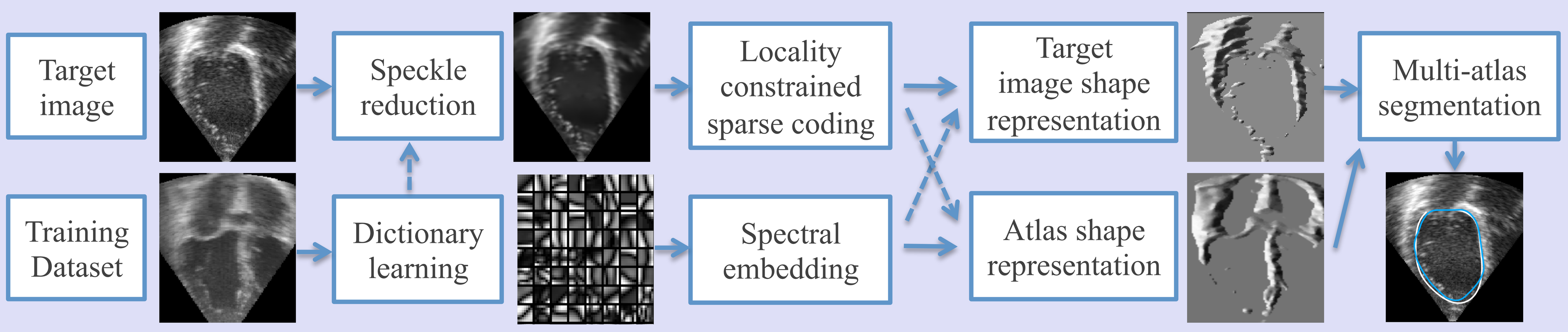
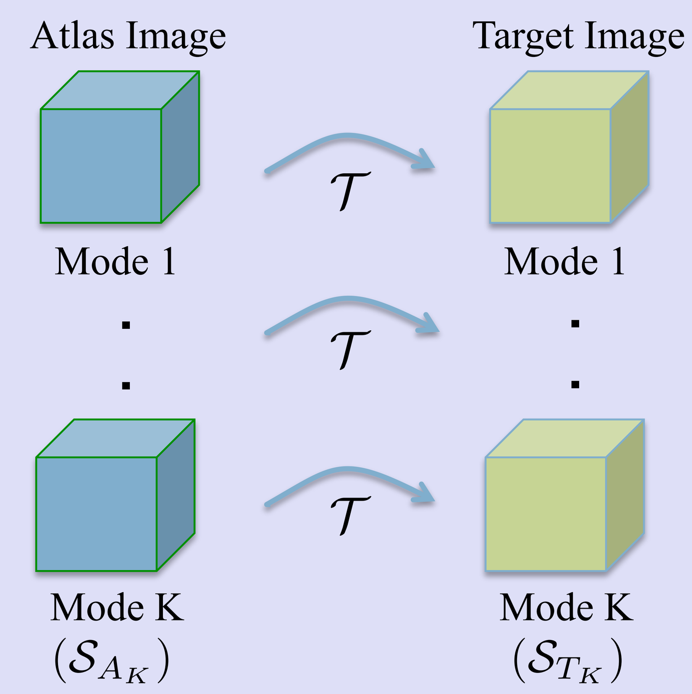


Figure 3: Block diagram of the proposed multi-atlas segmentation framework.

A single over-complete dictionary is learnt for all 15 training subjects, and the corresponding manifold of this dictionary is computed. Then, image patches obtained from target image are mapped to the manifold using local sparse coding. The corresponding atlases ($M = 5$) selected based on image similarity criterion are also mapped to the same manifold space. After registering images in the feature space, the labels are propagated and final segmentation is decided by majority voting.

Registration Strategy



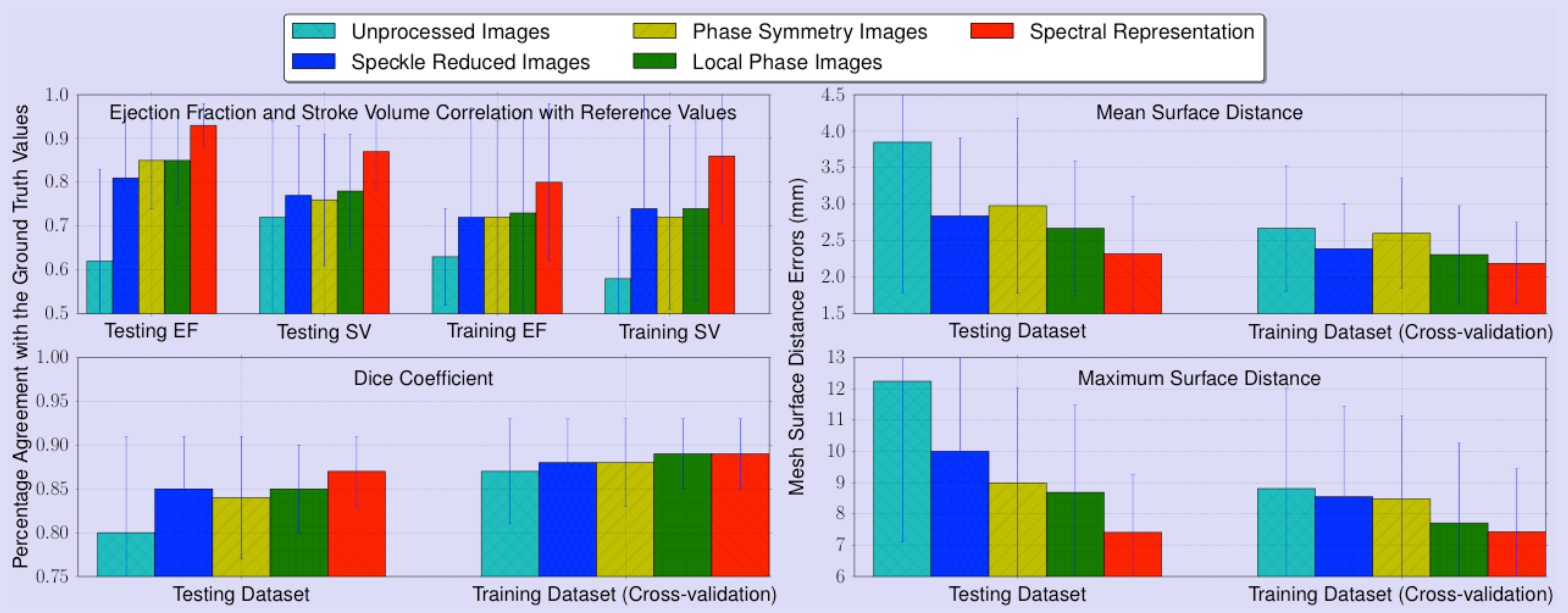
Different than the standard multi-atlas approach[3], the image similarity metric in the proposed registration algorithm is based on image descriptors. In that respect, images are aligned to each other by minimizing sum-of-squared differences between their spectral coordinates. The cost function is defined as:

$$\min_{\mathcal{T}} \sum_{k=1}^{K} \|\mathcal{S}_{A_k}(\mathcal{T}(\mathbf{p})) - \mathcal{S}_{T_k}(\mathbf{p})\|^2 + \beta\,\mathcal{R}(\mathcal{T})$$

B-spline FFD based registration is employed to propagate atlases $(\mathcal{S}_{\mathcal{A}})$ to target image $(\mathcal{S}_{T})$.

Results

The proposed segmentation framework is validated on a dataset of 3D echo cardiac image sequences acquired from 30 subjects. LV segmentation is done only for the end-diastolic and systolic frames. The accuracy of computed segmentations and clinical indices are used as criteria to evaluate the proposed method and compared against phase and intensity images. The bar plots show the measured surface distances and the accuracy in predicting clinical indices: ejection fraction and stroke volume.



References

[1] C. Wachinger and N. Navab, "Entropy and Laplacian images: Structural representations for multi-modal registration," *Medical Image Analysis*, 2012.

[2] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *CVPR*, 2010.

[3] P. Aljabar *et al.*, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, 2009.