Sparsity Based Spectral Embedding: Application to Multi-Atlas Echocardiography Segmentation

Ozan Oktay, Wenzhe Shi, Jose Caballero,
Kevin Keraudren, and Daniel Rueckert

Biomedical Image Analysis Group, Imperial College London, UK
o.oktay13@imperial.ac.uk

Abstract. Echocardiography is one of the primary imaging modalities used in the diagnosis of cardiovascular diseases. It is commonly used to extract cardiac functional indices including the left ventricular (LV) volume, mass, and motion. The relevant echocardiography analysis methods, including motion tracking, anatomical segmentation, and registration, conventionally use the intensity values and/or phase images, which are highly sensitive to noise and do not encode contextual information and tissue properties directly. To achieve more accurate assessment, we propose a novel spectral representation for echo images to capture structural information from tissue boundaries. It is computationally very efficient as it relies on manifold learning of image patches, which is approximated using sparse representations of dictionary atoms. The advantage of the proposed representation over intensity and phase images is demonstrated in a multi-atlas LV segmentation framework, where the proposed spectral representation is directly used in deformable registration. The results suggest that the proposed spectral representation can provide more accurate registration between images. This in turn provides a more accurate LV segmentation. Finally, it is the first time that a multi-atlas approach achieves state-of-the-art results in echo image segmentation.

1 Introduction

In the diagnosis of cardiovascular diseases, echocardiography is still the most important and widely used tool due to its high availability and ease of use. It has been used to extract functional and quantitative indices like left ventricular (LV) mass, volume and motion. The accuracy of these measurements depends on the correct delineation of endocardial boundary; thus, automated segmentation tools are more desirable for analysis as manual tracing is subject to inter-observer variability and human error. However, volumetric segmentation is still a challenging task for echo images due to image artefacts and low image quality.

The existing approaches to echocardiography segmentation can be divided into model-based and data-driven. Deformable surface models [2] and active-shape models [5] are two examples of the former category, which require a good model initialization or training to learn shape prior information to subsequently segment the ventricle boundary in target images. Although these approaches

achieve state-of-the-art segmentation results, they are limited by inter-subject anatomical variations due to the extensive training needed to cope with large shape variations. On the other hand, the second category relies on intensity distributions instead of a trained model. Thus, they are less sensitive to inter-subject variability, but are highly susceptible to the noise level and inconsistencies in the intensity distribution. Two common examples are edge-based level sets [12] and multi-atlas segmentation [20]. Particularly, although atlas based approaches have been successful for MRI segmentation [1], large registration errors on echo images prevent them from being effective for echo images.

Indeed, intensity and phase images are not representative enough to guide image registration because they do not directly reflect properties of the tissues or their contextual information. In this paper, to address this problem we propose a novel spectral representation for echo images, through which we extend and outperform the multi-atlas segmentation framework proposed in [20]. The new image representation captures structural information and guides the deformable registration to obtain a better tissue alignment. It also reduces the noise sensitivity and removes the need for image compounding, and ultimately achieving higher segmentation accuracies.

Spectral embedding is employed to compute the proposed representation, which has been successfully applied in min-cut segmentation [8], multi-modal image registration [16], and large deformation estimation problems [10]. Nevertheless, spectral embedding is not directly applicable to 3D echo images due to the large number of image patches, resulting in long computation times and intensive memory usage. We therefore propose, as an additional contribution, a more efficient embedding that exploits the redundant nature of echo image patches. The underlying manifold structure is learned only for atoms from a trained dictionary that sparsely represent the image patches. A single over-complete dictionary is assumed to be representative enough for all echo image patches to approximate the low dimensional space and each image patch is mapped to the underlying manifold space as a sparse linear combination of atoms yielding a set of spectral coordinates. To preserve the geodesics and local structure, sparse selection in coding is achieved by enforcing the locality constraint [17], which implies both sparsity and locality as explained in [18].

In the context of this paper, the proposed image patch embedding is referred as spectral representation. The paper is structured as follows: In section 2 of this paper, we introduce the relevant theory of dictionary learning and sparse coding for the manifold approximation. Section 3 presents validation results on the CETUS challenge data [4], which shows significant improvement in segmentation accuracy using the proposed representation over phase and intensity images. In the last section, the paper concludes by a brief discussion of the results.

2 Methodology

In the proposed segmentation framework, echo images are first sparsely reconstructed with dictionary atoms for speckle reduction. Secondly, a spectral repre-

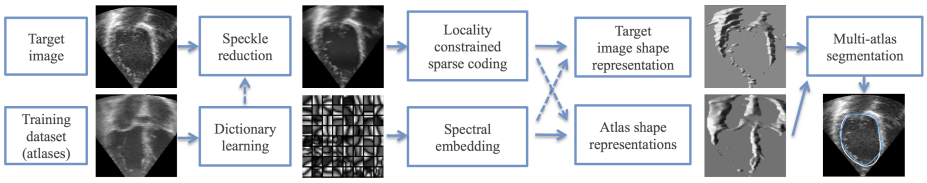

Fig. 1. Block diagram of the proposed multi-atlas segmentation framework

sentation is extracted from the processed images by mapping image patches to the manifold space of the dictionary atoms. Then, atlas labels are propagated to the target image by deformable registration using the spectral representation. The framework is shown in Fig. 1 and detailed below.

Speckle reduction: Target echo sequences are preprocessed prior to segmentation to increase signal-to-noise ratio. Instead of relying on standard speckle reduction techniques [6], images are denoised using dictionary learning and sparse coding similar to the image denoising application in [7]. On top of achieving state-of-the-art denoising, dictionary learning provides global patch analysis by building a set of atoms from training data that sparsely represent image patches. For echo images, these atoms have characteristic edge patterns. We use the K-SVD algorithm [7] to approximate image patches $\mathbf{y}_n \in \mathbb{R}^P$ as sparse combinations $\mathbf{x}_n \in \mathbb{R}^M$ of atoms from an over-complete dictionary $\mathbf{C} \in \mathbb{R}^{P \times M}$ with a precision bounded by ϵ, namely solving: $\min_{\mathbf{C},\mathbf{X}} \sum_{n=1}^{N} \|\mathbf{x}_n\|_0 \;\; \text{s.t.} \;\; \forall n \;, \|\mathbf{y}_n - \mathbf{C}\mathbf{x}_n\|^2 \leq \epsilon \mid \epsilon \in \mathbb{R}_+$. Patches are overlapping and wrap around image boundaries, meaning there are N patches for an image of N pixels.

Spectral representation: As shown in previous works [10,16], spectral coordinates can be computed using non-linear dimensionality reduction of image patches; this paper particularly focuses on Laplacian Eigenmaps (LE) [3]. The algorithm computes the Laplacian graph $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ using the adjacency and degree matrices $\mathbf{A}, \mathbf{D} \in \mathbb{R}^{N \times N}$ corresponding to all image patches. Then, spectral coordinates are obtained by finding the lowest K eigenvectors of the matrix $\mathbf{L}$. This representation is suited for small datasets such as small stacks of MRI slices, but is prohibitive for 3D echo volumes due to the large amounts of voxels in the image that result in a very large adjacency matrix. Furthermore, finding a fixed low-dimensional space for all images is also challenging and is usually solved by point-matching algorithms.

To overcome these problems, we propose to perform manifold learning on dictionary atoms ($\mathbf{c}_m \in \mathbb{R}^P$) and then the spectral coordinates are approximated by sparse linear combinations of dictionary atoms. For this approximation, two main assumptions are made: (1) echo image patches can be sparsely represented by dictionary atoms and (2) these patches can be expressed in a manifold [11] that groups atoms with similar edge patterns. In that respect, the learned dictionary atoms that are the byproduct from the speckle reduction step are transformed into spectral coordinates with LE. Dictionary atoms of similar shape are

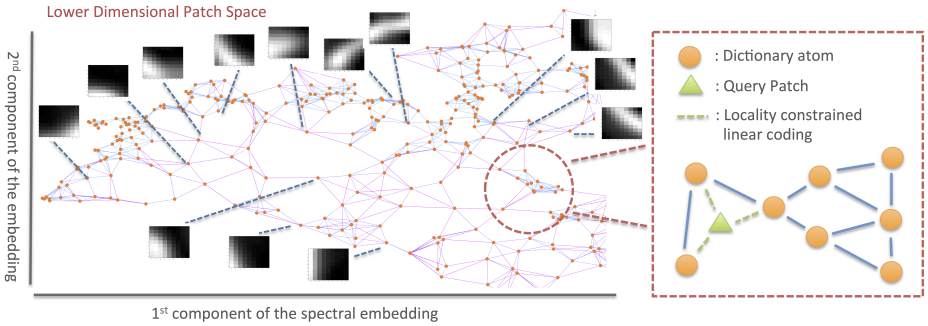

Fig. 2. The lowest two spectral coordinates of the dictionary atoms (left), locality constrained linear coding of a query patch to map it to manifold space (right).

grouped together in the spectral coordinates and the variance of the patches is maximized, as shown in Fig. 2. With the embedding learned, each image patch ($\mathbf{y}_n$) is mapped to the lower dimensional space through linear combinations ($\tilde{\mathbf{x}}_n$) of spectral coordinates corresponding to dictionary atoms ($\mathbf{s_c} \in \mathbb{R}^{M\times K}$) solving: $\mathbf{s}_{y_n} = \tilde{\mathbf{x}}_n^\top \mathbf{s_c}$. The linear codes are found by minimizing the cost function $\min_{\mathbf{X}} \sum_n \|\mathbf{y}_n - \mathbf{C}\tilde{\mathbf{x}}_n\|^2 + \lambda \|\mathbf{b}_n \odot \tilde{\mathbf{x}}_n\|^2$ s.t. $\forall n$, $\mathbf{1}^\top \tilde{\mathbf{x}}_n = 1$, where $\odot$ denotes the element-wise multiplication and $\lambda \in \mathbb{R}_+$. This formulation enforces a locality constraint [17] based on pair-wise distances $\mathbf{b}_n = [b_{(n,1)}, \ldots, b_{(n,M)}]$ where $b_{(n,m)} = \exp\left(\|(\mathbf{y}_n - \mathbf{c}_m\|^2 / \sigma \right)$ and σ is the variance term. The penalty term assigns higher weights to dictionary atoms $\mathbf{c}_m$ that are close to the patch $\mathbf{y}_n$.

As explained in [18], the locality constraint also implies sparsity, thus the solution can be considered as a sparse weighting of the dictionary atoms. The sparse codes computed in Euclidean space can be applied in manifold space as long as the locality constraint is applied. A single component of approximated spectral coordinates is displayed in Fig. 3.

Multi-atlas segmentation: The proposed spectral representation is used in image registration to perform multi-atlas segmentation on echocardiographic images. Different than the standard multi-atlas approach [1], 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 (SSD) between their spectral coordinates instead of image intensity values.

The proposed segmentation framework is described as follows: All atlases collected from the training dataset are linearly aligned to a target image using manually selected three landmarks (left ventricle apex, mid-ventricle, and mitral valve). Similar to the approach in [1], a subset of atlases is selected by computing normalized mutual information (NMI) over a region of interest defined by the atlas labels and target image. The most similar L atlases are then selected and spectral represented atlases ($\mathcal{S}_A$) are deformable registered to the target spectral image ($\mathcal{S}_T$) with B-spline FFD [14]. The following cost function is minimized:

$\sum_{k=1}^{K} \|\mathcal{S}_{A_k}(\mathbf{p}+\mathbf{u}) - \mathcal{S}_{T_k}(\mathbf{p})\|^2 + \beta\,\mathcal{R}(\mathbf{u})$, where $\mathbf{p}$, $\mathbf{u}$ and k denote position, displacement and spectral coordinate dimension. The regularization $\mathcal{R}$ is defined as bending energy and weighted by $\beta \in \mathbb{R}_+$. The algorithm estimates a single common displacement field between volumetric spectral image pairs (in total K) while minimizing the cost function. Lastly, the segmentation is decided by majority voting of the propagated atlas labels.

Local phase images: The work in [20] on multi-atlas echocardiograhy segmentation uses local phase images to register atlases to target images. To demonstrate the contribution of the proposed spectral representation, local phase images are evaluated in the same segmentation framework. Images are first converted to a band-pass signal with Laplacian of Poisson filter [19]: $\mathcal{F}\{LOP\}(\mathbf{w}) = -8\pi^3|\mathbf{w}|^2 \exp(-2\pi|\mathbf{w}|\rho)$, where $\mathbf{w} \in \mathbb{R}^3$ is the position vector in the frequency domain and ρ controls the central frequency. In our experiments, this filter selection achieved better results compared to Gaussian derivative filter. Afterwards, an analytic signal is obtained by filtering with an isotropic Riesz filter, and the phase image is characterized by the angle between real and imaginary components as explained in [20]. Additionally, local-phase based boundary images [13] are evaluated in the same framework, which are computed in multi-scale ($\boldsymbol{\rho} \in \mathbb{R}^3$) using a monogenic signal. An example of the computed phase and boundary images is displayed in Fig. 3. In the registration step of multi-atlas segmentation, the cost function is defined as $\mathcal{C} = \omega_1 \,.\, \mathrm{NMI}(I_A, I_T) + \omega_2 \,.\, \mathrm{NMI}(\phi_A, \phi_T)$, where ϕ and I denote phase and intensity images. A similar formulation is used for boundary images by replacing phase images in the cost function.

3 Implementation and Results

Validation dataset: The proposed segmentation framework is validated on the dataset provided by the MICCAI 2014 CETUS challenge [4], consisting of a set of 3D echo cardiac image sequences acquired from 30 subjects and separated into training (15 subjects) and testing (15 subjects) datasets. As ground truth segmentations for the testing set are not provided, the validation is performed blindly using the evaluation system provided by the CETUS.

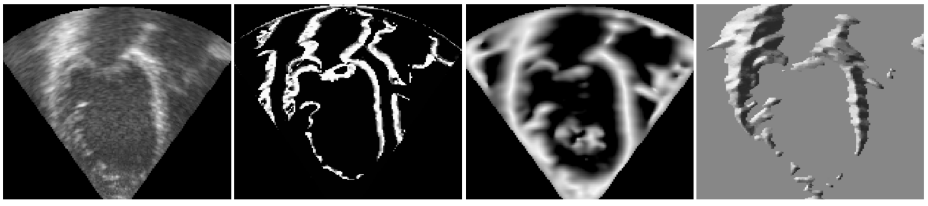

Fig. 3. Left to right: (1) input image, (2) phase-based boundary detection, (3) local phase image, (4) proposed spectral representation (single component of embedding)

Table 1. Multi-atlas segmentation results on training (cross-validation) and testing datasets based on (A) Unprocessed images, (B) Speckle reduced images, (C) Local phase-based boundary images [13], (D) Local phase images [20], and (E) Proposed spectral representation. EF and SV values are reported based on the percentage error measure. Surface distances are given in terms of mean and maximum values.

		Mean (mm)	Max (mm)	Dice Score	Ejection fraction	Stroke volume
Testing	(A)	3.85±2.06	12.24±5.12	0.80±0.11	0.62±0.21	0.72±0.22
	(B)	2.84±1.07	10.00±3.04	0.85±0.06	0.81±0.14	0.77±0.16
	(C)	2.98±1.20	8.99±3.05	0.84±0.07	0.85±0.11	0.76±0.15
	(D)	2.67±0.92	8.69±2.78	0.85±0.05	0.85±0.10	0.78±0.13
	(E)	2.32±0.78	7.41±1.84	0.87±0.04	0.93±0.05	0.87±0.09
Training	(A)	2.67±0.86	8.81±3.23	0.87±0.06	0.63±0.11	0.58±0.14
	(B)	2.39±0.62	8.55±2.90	0.88±0.05	0.72±0.25	0.74±0.34
	(C)	2.60±0.75	8.48±2.66	0.88±0.05	0.72±0.22	0.72±0.21
	(D)	2.31±0.67	7.71±2.55	0.89±0.04	0.73±0.23	0.74±0.21
	(E)	2.19±0.56	7.63±2.43	0.89±0.04	0.80±0.18	0.86±0.15

Implementation details: The patch and dictionary sizes selected for dictionary learning are $P = 7\text{x}7\text{x}7$ and $M = 850$. The adjacency graph is constructed by linking each dictionary atom to its 8 most similar neighbours in terms of ℓ_2 norm distance. In total $K = 4$ spectral components are selected for spectral representation. In locality constraint coding, parameters are set to $\lambda = 0.3$ and $\sigma = 0.2$ for normalized images. The bandpass filter parameter for the computation of phase images is chosen to be $\rho = 4.5$ and $\boldsymbol{\rho} = [3.5, 5.0, 7.0]$ for the boundary images. To conclude, the regularization weights in the registration cost functions are defined as $\omega_1 = \omega_2 = \beta = 1$, and $L = 5$ atlases are selected from the training dataset for label propagation.

Validation strategy and results: In the validation, 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. In that respect, multi-atlas segmentation is performed on 5 different types of image surrogates, which are provided in Table 1. The evaluation is performed separately for testing and training datasets (cross-validation). As shown, the best result for the testing dataset is obtained using spectral representation based multi-atlas segmentation, which achieves 2.32 mm mean error and 0.87 Dice score. In comparison to intensity and phase images, an improvement of 1.53 and 0.35 mm is observed for the mean surface distance and a similar figure of merit is seen for Dice score results. This suggests that the proposed representations provide more useful information to guide the registration algorithm. Moreover, the results for the cross-validation on the training dataset also demonstrate that spectral representation outperforms segmentation based on speckle reduced intensity images.

Table 2. Comparison of the proposed multi-atlas approach (E) against the state-of-the-art echocardiogaphy segmentation: active surfaces [2] and active shape model [5]. Estimated ejection fraction (EF) and end-diastolic volume (EDV) are compared against their reference values. The correlation accuracy is reported in terms of Pearson's coefficient (R) and Bland-Altman's limit of agreement (BA).

	Mean (mm)	R_{EF}	BA_{EF} $(\mu \pm 2\sigma)$	R_{EDV}	BA_{EDV}	# of Patients
(E)	2.32±0.78	0.923	-0.74±6.26	0.926	12.88±35.71	15
[2]	-	0.907	-1.0±9.8	0.971	-1.4±23.2	24
[5]	1.84±1.86	-	0±19	-	3.06±46.86	10

To assess the clinical usefulness of the proposed framework, ejection fraction (EF_c) and stroke volume (SV_c) are evaluated for each patient and compared against their reference values (EF_r). The comparison is done by the percentage error measure $\mathcal{C}_{EF} = 1 - |EF_c - EF_r| / EF_r$, similarly for the stroke volume $\mathcal{C}_{SV}$. The mean value of the percentage errors given in Table 1 demonstrate that the values computed using the spectral representation show a closer agreement with reference values compared to the other representations. Furthermore, a qualitative comparison of the segmentations obtained with the spectral representation and phase image is given in Fig. 4. Finally, the validation results are compared against two state-of-the-art methods in echocardiography segmentation, shown in Table 2. Although the given results are obtained for different datasets, the comparison demonstrates that multi-atlas segmentation can be as successful as the best-performing methods in estimation of important clinical parameter values.

It is observed that the use of a larger number of atlases does not increase the accuracy significantly because the dataset is small and contains large variations, so limiting the number of atlases to 5 in all experiments reduced computation time. Experiments were carried out on a 3.00 GHz quad-core machine, and the approximate computation time per image was 3 min for denoising, 2.5 min for spectral representation, and 30 min for deformable registrations with 5 atlases.

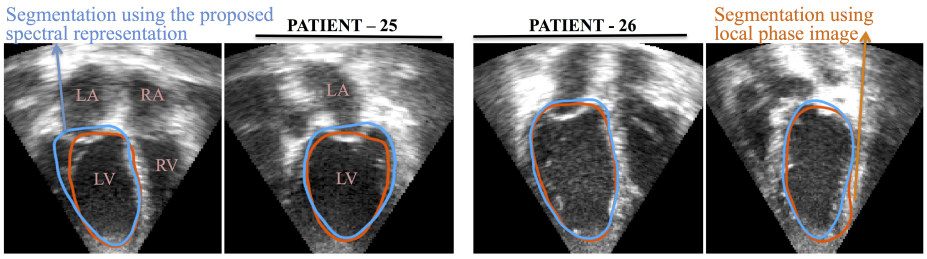

Fig. 4. Left ventricle segmentation of two different subjects. Segmentation obtained with the spectral representation (in blue) delineates the endocardium more accurately than the local phased multi-atlas segmentation (in orange).

4 Discussion and Conclusion

In this paper, we presented a new spectral representation for echocardiograhy images based on sparse reconstruction of dictionary atom spectral embeddings. The advantages of this representation are both quantitatively and qualitatively demonstrated in a multi-atlas LV segmentation framework. The results show that it outperforms the local phase and boundary representations in terms of segmentation accuracy. This finding can be related to an improved noise robustness and the explicit use of encoded contextual information, the lack of which in intensity and phase images is an important limitation for guiding the deformable registration in multi-atlas segmentation.

In addition, the proposed representation is computationally efficient and does not require image feature design by hand-crafting as in phase images since distinctive spectral representation is learned from the data itself. Another interesting realization is that spectral representation based multi-atlas segmentation can achieve state-of-the-art results in echocardiography LV segmentation, without requiring any shape prior models. Previous attempts on multi-atlas segmentation required image compounding, and they were not as successful as the proposed framework due to inaccuracies in registration algorithm.

The proposed image descriptors could alternatively be replaced by the sparse coding coefficients without a need for spectral embedding, as proposed in [15] for image segmentation and in [9] for image registration. One particular example would be the use of histogram of sparse codes to represent the image patches. However, the spectral embedding has two main advantages in comparison to the coefficient based representation: Spectral coordinates obtained from different image patches are comparable, and they allow the use of globally smooth distance metrics which are necessary for the registration stage. On the other hand, sparse codes in an overcomplete dictionary, unlike in an orthonormal basis, are non-unique and are therefore unsuitable for image patch comparison. Locality constrained coding yields a locally smooth sparse selection of atoms, meaning that a distance metric could potentially be defined locally at best, but not globally. The other advantage of spectral embedding is the reduced computation time. The spectral coordinates provide a compact and rich representation of patches with a few components, whereas histogram of coefficients would require a large number of computations for comparison and a large vector to represent an image patch.

In conclusion, sparse and parametrizable characteristics of echo images enable us to develop a consistent spectral representation that contains rich structural information. The proposed representation is generic, and can be applied to other ultrasound image applications which require image registration.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46(3), 726–38 (2009)

2. Barbosa, D., et al.: Fast and fully automatic 3-D echocardiographic segmentation using B-spline explicit active surfaces: Feasibility study and validation in a clinical setting. Ultrasound in medicine & biology 39(1), 89–101 (2013)
3. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15(6), 1373–96 (2003)
4. Bernard, O., et al.: Challenge on endocardial three-dimensional ultrasound segmentation (CETUS) (2014)
5. Butakoff, C., Sukno, F., Doltra, A., Silva, E., Sitges, M., Frangi, A.F.: Order statistic based cardiac boundary detection in 3D+T echocardiograms. In: Functional Imaging and Modeling of the Heart, pp. 359–366. Springer (2011)
6. Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. IEEE T Image Process pp. 2221–29 (2009)
7. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE T Image Process 15(12), 3736–45 (2006)
8. Estrada, F.J., Jepson, A.D., Chennubhotla, C.: Spectral embedding and min cut for image segmentation. In: BMVC. pp. 1–10 (2004)
9. Guerrero, R., Rueckert, D.: Data-specific feature point descriptor matching using dictionary learning and graphical models. In: SPIE Medical Imaging. International Society for Optics and Photonics (2013)
10. Lombaert, H., Grady, L., Pennec, X., Ayache, N.: Spectral log-demons: Diffeomorphic image registration with very large deformations. IJCV pp. 1–18 (2013)
11. Peyré, G.: Manifold models for signals and images. Computer Vision and Image Understanding 113(2), 249–260 (2009)
12. Rajpoot, K., Grau, V., Alison Noble, J., Becher, H., Szmigielski, C.: The evaluation of single-view and multi-view fusion 3D echocardiography using image-driven segmentation and tracking. Medical image analysis 15(4), 514–528 (2011)
13. Rajpoot, K., Grau, V., Noble, J.A.: Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In: ISBI. pp. 783–786. IEEE (2009)
14. Rueckert, D., Sonoda, L., Hayes, C., Hill, D.L., Leach, M., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE TMI 18(8), 712–21 (1999)
15. Tong, T., Wolz, R., Coupé, P., Hajnal, J.V., Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage 76, 11–23 (2013)
16. Wachinger, C., Navab, N.: Entropy and Laplacian images: Structural representations for multi-modal registration. Medical Image Analysis 16(1), 1–17 (2012)
17. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR. pp. 3360–67. IEEE (2010)
18. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: NIPS. vol. 9, p. 1 (2009)
19. Zhang, L., Zhang, D., Guo, Z.: Monogenic-LBP: a new approach for rotation invariant texture classification. In: ICIP. pp. 2677–2680. IEEE (2010)
20. Zhuang, X., Yao, C., Ma, Y., Hawkes, D., Penney, G., Ourselin, S.: Registration-based propagation for whole heart segmentation from compounded 3D echocardiography. In: ISBI. pp. 1093–1096. IEEE (2010)