

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

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

Biomedical Image Analysis Group, Imperial College London, UK

Abstract. As part of the CETUS challenge, we present a multi-atlas segmentation framework to delineate the left-ventricle endocardium in echocardiographic images. To increase the robustness of the registration step, we introduce a speckle reduction step and a new shape representation based on sparse coding and manifold approximation in dictionary space. The shape representation, unlike intensity values, provides consistent shape information across different images. The validation results on the test set show that registration based on our shape representation significantly improves the performance of multi-atlas segmentation compared to intensity based registration. To our knowledge it is the first time that multi-atlas segmentation achieves state-of-the-art results for echocardiographic images.

1 Introduction

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, availability, and high temporal resolution. However, the automatic delineation of the endocardial boundary remains a challenging task due to low image quality.

The existing work on echocardiography segmentation can be divided into two groups: model-based and data-driven approaches. The first category makes use of prior shape constraint, as in active shape models [14] or deformable models [13], which require a large training set. In the absence of large amounts of training data, they fail to cope with inter-subject anatomical variations. The second group of approaches, on the other hand, rely on image features or intensity distributions such as: edge-based level sets and Bayesian formulations [9]. They are less sensitive to anatomical differences, but have limitations when facing inconsistent intensity distribution and low signal-to-noise ratio (SNR).

This paper proposes a segmentation framework based on label propagation, similar to [15]. As an improvement, the proposed approach does not require image compounding. More importantly it does not employ hand-crafted image features for registration like local phase and image gradients, which are sensitive to noise level. Indeed, the main challenge when applying multi-atlas segmentation on echo images is the low SNR. To overcome this problem, we introduce

a novel data-driven feature learning algorithm. It eliminates noise and extracts the shape representation of image patches at atrium and ventricle boundaries. The new algorithm relies on sparse coding, spectral embedding, and manifold approximation. Compared to the other shape representations like the Laplacian images [11] or spectral embedding in the image domain [5], our method remains computationally efficient. The proposed learned shape representation is integrated in multi-atlas segmentation. During the evaluation, we demonstrate that the proposed approach increases the accuracy compared to the intensity images. Finally, it is a generic approach: Since it does not require any offline training, it can be applied on images with different modality, morphology and field of view.

2 Proposed method

In the proposed framework, the images from all patients are first pre-processed for speckle reduction, they are then converted to their shape representations by manifold learning. These representations are then used in the registration step of the multi-atlas segmentation. The details of each step are described as follows:

Speckle reduction: All echo sequences are preprocessed prior to segmentation. The speckle reduction is performed based on sparse representation of learned dictionary atoms. This approach has been successfully applied in image restoration [7], and it is also applicable to echo images assuming image patches are parameterizable and can be sparsely represented. In this way, speckle is removed while ventricle boundaries in the images are preserved. As proposed in [4], K-SVD algorithm is used to solve the cost function:

$$\min_{\mathbf{C}, \mathbf{X}} \sum_i \|\mathbf{x}_i\|_0 \quad \text{s.t.} \quad \forall i, \|\mathbf{y}_i - \mathbf{C}\mathbf{x}_i\|^2 \leq \epsilon \mid \epsilon \in \mathbb{R}_+ \quad (1)$$

which reconstructs the observations $\mathbf{y}_i \in \mathbb{R}^n$ as sparse combinations $\mathbf{x}_i \in \mathbb{R}^m$ of dictionary atoms in $\mathbf{C} \in \mathbb{R}^{n \times m}$. The K-SVD algorithm iteratively alternates between the dictionary update and the OMP sparse coding of \mathbf{X} . In contrast to other speckle reduction methods, such as non-local means [3], sparse coding of learned patches is not limited only to the intensity information in local neighbourhood, but the global information, learned from the image itself, is used as well. This advantage makes the dictionary based reconstruction more accurate in recovering the edge information in low SNR image areas.

Learning shape representation: Multi-atlas segmentation is performed on our novel shape representation extracted from echo image patches. In comparison to intensities, the proposed representation captures contextual information in the local spatial neighbourhood, which provides useful local shape information for the image registration in multi-atlas segmentation. As explained in [10,11], the representations are learned by extracting the underlying manifold structure of image patches and mapping them to a lower dimensional space. As this transformation preserves the locality, similar shaped patches are grouped together

in the low dimensional patch space while the variance of the patches is maximized. Learning a manifold requires the computation of an adjacency matrix $\mathbf{A} \in \mathbb{R}^{l \times l}$, where l is the total number of image patches. The edge weights of the connected graph are the pairwise similarities between the image patches: $A_{i,j} = \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2 / \gamma)$, where γ is the mean of squared patch distances. For echo images of large dimensions the computation of this matrix becomes computationally very expensive making this approach not directly applicable. To overcome this problem, Laplacian Eigenmaps [2] is applied on dictionary space instead of the image space. In that respect, the lowest K eigenmodes of the normalized Laplacian graph $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \tilde{\mathbf{A}} \mathbf{D}^{-1/2}$ are computed for the degree matrix $\mathbf{D} \in \mathbb{R}^{m \times m}$ and $\tilde{\mathbf{A}} \in \mathbb{R}^{m \times m}$. Afterwards, spectral coordinates of all patches are approximated by mapping each image patch to the manifold of dictionary atoms through locally constrained sparse coding as suggested in [12]. The cost function of the local linear coding algorithm is:

$$\min_{\mathbf{x}} \sum_i \|\mathbf{y}_i - \mathbf{C}\mathbf{x}_i\|^2 + \lambda \|\mathbf{b}_i \odot \mathbf{x}_i\|^2 \quad \text{s.t.} \quad \forall i, \mathbf{1}^\top \mathbf{x}_i = 1 \quad (2)$$

where \odot denotes the element-wise multiplication and $\lambda \in \mathbb{R}_+$. It computes the sparse code of each query patch by enforcing the locality constraint. Pairwise ℓ_2 norm distances $\mathbf{b}_i = \exp(\text{dist}(\mathbf{y}_i, \mathbf{C}) / \sigma)$ determine the locality weight of an atom. Fig. 1 illustrates a single component of approximated manifold, i.e. the shape quantization of echo image patches.

Multi-atlas segmentation: Similar to the framework in [1], the target image is segmented by propagating labels of a subset of atlases selected from the training dataset. The label propagation is based on linear and deformable image registration between the manually segmented subset of atlas images and target image. Different than the standard multi-atlas segmentation, in the proposed framework, the image similarity metric in registration is based on the shape representation of images instead of the intensity values.

The segmentation framework can be described as follows: Firstly, all atlases are linearly transformed to the target image space based on three manually selected landmarks (left ventricle apex, mid-ventricle, and mitral valve). A region of interest is defined by the overlap of propagated labels after linear transformation. Secondly, the most similar M atlases are selected based on normalized mutual information, which could alternatively be replaced by another global image similarity metric. Next, the target and selected atlases are converted into shape representations, and B-spline free-form deformation based registration [8] is employed to propagate the atlases (\mathcal{F}_A) to the target image (\mathcal{F}_Q). Finally, the segmentation is decided by majority voting of the propagated labels. The cost function for the registration is defined as:

$$\min_{\mathbf{u}} \sum_{k=1}^K \|\mathcal{F}_{A_k}(\mathbf{p} + \mathbf{u}) - \mathcal{F}_{Q_k}(\mathbf{p})\|^2 + \beta \mathcal{R}(\mathbf{u}) \quad (3)$$

where $\mathbf{p}, \mathbf{u} \in \mathbb{R}^d$, $\beta \in \mathbb{R}$ and $\mathcal{R}(\cdot)$ denote position, displacements, and bending energy. The cost function minimizes the sum of squared differences between

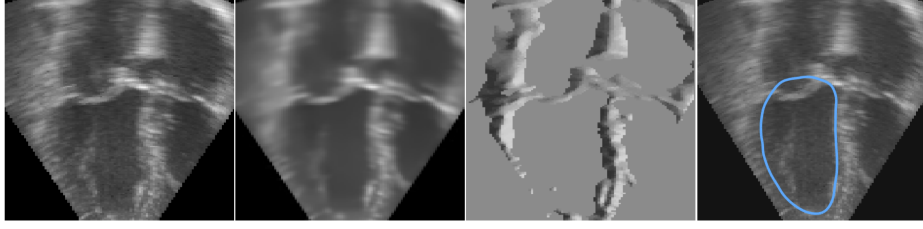


Fig. 1. From left to right: (1) input image, (2) proposed speckle reduction, (3) learned shape representation, and (4) multi-atlas segmentation of LV endocardium

the spectral coordinates of atlas and target images, where there are in total K separate spectral components in the low dimensional patch space.

3 Implementation and Results

In speckle reduction, dictionary and patch sizes are selected as $m = 850$, $n = 7 \times 7 \times 7$. The dictionary atoms \mathbf{c}_i are learned from zero-mean and unit variance image patches. Maximum residual error is limited to $\epsilon = 0.25$ for normalized images. In the shape representation, a single over-complete dictionary is learnt for all 15 training subjects, and the corresponding manifold of this dictionary is computed by finding 8 nearest neighbours of each atom. Then, image patches are mapped to the manifold using parameter values $\lambda = 0.3$, $K = 4$, and $\sigma = 0.2$. Lastly, the registrations are performed on a three level image pyramid for $\beta = 1$.

The evaluation is performed on both testing and training datasets, each consisting of 15 volumetric echo sequences. For both cases, the segmentations are performed on the end-diastolic (ED) and end-systolic (ES) frames. The images from these two phases are segmented individually using different subsets of atlases selected from the respective phases. This way of atlas selection is observed to be more accurate in the segmentation. The results are reported in Table 1 and 2. The accuracy of the LV delineation is measured in terms of surface distance and DICE metrics. For the training dataset, the results are obtained using a leave one out cross-validation.

The two tables show the multi-atlas segmentation results for three different types of image surrogates: (A) unprocessed images, (B) speckle reduced images, and (C) shape representation. The results show that the denoising improves the segmentation by 1.01 mm compared to the unprocessed images. In addition, the use of shape representations outperforms the intensity based approaches by providing a more robust registration. The best performance is obtained using 10 atlases with shape representations, for which the final result is reported as 2.32 mm mean error, 7.41 mm Hausdorff distance, and 0.87 Dice score. However, there is no statistical significant improvement by using 10 atlases instead of 5 atlases given that the dataset is small and contains large variations. For this

Table 1. Multi-atlas segmentation cross-validation results on the training dataset (Patient 1 to 15) based on (A) unprocessed images, (B) speckle reduced images, and (C) shape representations. The evaluation metrics are mean surface distance (MSD), Hausdorff distance (HD), Dice coefficient (DC), and t-test P value.

Number of atlases ($M = 5$)				
	MSD (mm)	HD (mm)	DC (%)	P -value
(A)	2.67 ± 0.86	8.81 ± 3.23	0.87 ± 0.06	0.47×10^{-3}
(B)	2.39 ± 0.62	8.55 ± 2.90	0.88 ± 0.05	0.36×10^{-3}
(C)	2.19 ± 0.56	7.63 ± 2.43	0.89 ± 0.04	-

Table 2. Multi-atlas segmentation results on the testing dataset (Patient 16 to 30) based on (A) unprocessed images, (B) speckle reduced images, and (C) shape representations. The accuracy is evaluated based on the surface overlap measure (Dice coefficient) and distance measurements (mean and Hausdorff distances).

	Number of atlases ($M = 5$)			Number of atlases ($M = 10$)		
	MSD (mm)	HD (mm)	DC (%)	MSD (mm)	HD (mm)	DC (%)
(A)	3.85 ± 2.06	12.24 ± 5.12	0.80 ± 0.11	3.12 ± 1.42	9.89 ± 3.21	0.84 ± 0.07
(B)	2.84 ± 1.07	10.00 ± 3.04	0.85 ± 0.06	3.07 ± 1.26	9.78 ± 3.05	0.84 ± 0.07
(C)	2.34 ± 0.85	8.15 ± 1.97	0.87 ± 0.04	2.32 ± 0.78	7.41 ± 1.84	0.87 ± 0.04

reason, a smaller number of atlases is preferred to reduce the computation time required for the segmentation.

Additionally, a paired t-test on the mean errors is performed to statistically verify the differences between the results; the P values are given in Table 1. Lastly, two clinical parameters, ejection fraction (EF) and stroke volume (SV), are evaluated for each patient and compared against their reference values. The comparison is done by calculating the Pearson’s correlation coefficient (PCC) and limit of agreement (LOA) between the estimated values and the ground-truth. Similarly, the same analysis is done for ES and ED volume measures. The accuracy of the derived clinical parameters is demonstrated in Table 3, which shows a close agreement between the reference and calculated values.

The reported results are obtained on a quad-core 3.00 GHz machine, and the computation time per image is recorded as: speckle reduction ≈ 3 min, shape representation ≈ 2.5 min, and deformable registrations ≈ 30 min for 5 atlases.

4 Discussion

The reported results indicate that multi-atlas segmentation can be used in left-ventricle boundary delineation in echocardiographic images. It performs as accurate as the state-of-the-art methods in estimating LV volume and clinical parameters. The main advantage of the atlas approach is that it does not require

Table 3. This table shows the accuracy of the derived clinical indices for the training (Patient 1 to 15) and testing datasets (Patient 16 to 30). Pearson’s correlation coefficient (PCC) and Bland-Altman’s limit of agreement ($\mu \pm 1.96\sigma$) values are given for the following indices: ejection fraction, stroke volume, end-systolic volume, and end-diastolic volume. Additionally, mean surface distance (MSD), Hausdorff distance (HD), and Dice coefficient (DC) values are computed separately for ED and ES volumes.

Testing dataset	MSD(mm)	HD(mm)	DC(%)	PCC	LOA ($\mu \pm 1.96\sigma$)
ED volume (ml)	2.32	7.17	0.885	0.926	12.81 \pm 33.77
ES volume (ml)	2.32	7.64	0.857	0.936	-7.77 \pm 28.27
Ejection fraction (%)	-	-	-	0.923	0.74 \pm 7.58
Stroke volume (ml)	-	-	-	0.832	-5.05 \pm 12.49

Training dataset	MSD(mm)	HD(mm)	DC(%)	PCC	LOA ($\mu \pm 1.96\sigma$)
ED volume (ml)	1.98	6.97	0.911	0.983	9.80 \pm 45.66
ES volume (ml)	2.46	8.53	0.870	0.961	11.21 \pm 56.91
Ejection fraction (%)	-	-	-	0.787	-1.07 \pm 18.52
Stroke volume (ml)	-	-	-	0.856	-1.38 \pm 27.08

any shape prior knowledge and model construction. For a given training data, a registration based approach has a higher degree of freedom in segmentation in comparison to the statistical shape model based approaches. The main difficulty of the application of the multi-atlas segmentation to echo images has been the lack of a reliable registration. As it is shown in the results section, the method is more reliable and accurate when the proposed shape representation is used for the characterization of images instead of the unprocessed intensity data.

The proposed framework does not require user interaction for the segmentation except for the selection of three landmark points for atlas initialization. In the experiments, these points are selected by a non-expert, so the sensitivity of the algorithm on the point selection is not very significant. In future work, the algorithm will be made fully-automatic by introducing a linear alignment stage prior to the shape representation based image registration.

Additionally, the accuracy of the segmentation method can be further improved by updating the shape representations at each level of the image registration algorithm. This can also be done in between the iterations of the registration algorithm. The main drawback of atlas based segmentation is probably the long computation time, which is measured in our experiments in the order of minutes. The primary reason for this computational load is the time required by the image registrations. As a solution to this problem, discrete optimization techniques can be used in registration algorithm, which is shown [6] to be very effective in reducing the computation time while preserving the registration accuracy.

5 Conclusion

In this paper, we demonstrated the applicability of multi-atlas segmentation for echocardiographic images using a novel shape representation. It captures the shape of the contours on endocardial surface. This advantage enables the algorithm to succeed even if tissue boundaries are disrupted by noise or are incomplete due to shadowing. The results show that by using the shape representation, the mean registration error can be reduced by 1.51 mm compared to the unprocessed image. Moreover, the proposed shape quantization method is computationally efficient and generic; such that it can be applied in other echo image applications. Future work will focus on faster registration algorithm and study the benefits of combining intensity data with shape representations for the improvement of multi-atlas segmentation in echocardiography.

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. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation* 15(6), 1373–96 (2003)
3. Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing* 18(10), 2221–2229 (2009)
4. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* 15(12), 3736–3745 (2006)
5. Estrada, F.J., Jepson, A.D., Chennubhotla, C.: Spectral embedding and min cut for image segmentation. In: *BMVC*. pp. 1–10 (2004)
6. Heinrich, M., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Transactions on Medical Imaging* 32(7), 1239–1248 (2013)
7. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. *IEEE Transactions on Image Processing* 17(1), 53–69 (2008)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8), 712–721 (1999)
9. Sarti, A., Corsi, C., Mazzini, E., Lamberti, C.: Maximum likelihood segmentation of ultrasound images with Rayleigh distribution. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 52(6), 947–60 (2005)
10. Shi, W., Lombaert, H., Bai, W., Ledig, C., Zhuang, X., Marvao, A., Dawes, T., O’Regan, D., Rueckert, D.: Multi-atlas spectral patchmatch: Application to cardiac image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer (2014)
11. Wachinger, C., Navab, N.: Entropy and Laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis* 16(1), 1–17 (2012)
12. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3360–3367. IEEE (2010)

13. Zagrodsky, V., Walimbe, V., Castro-Pareja, C.R., Qin, J.X., Song, J.M., Shekhar, R.: Registration-assisted segmentation of real-time 3-D echocardiographic data using deformable models. *IEEE Transactions on Medical Imaging* 24(9), 1089–1099 (2005)
14. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: A dynamical shape prior for LV segmentation from RT3D echocardiography. In: *Medical Image Computing and Computer-assisted Intervention—MICCAI 2009*, pp. 206–213. Springer (2009)
15. Zhuang, X., Yao, C., Ma, Y., Hawkes, D., Penney, G., Ourselin, S.: Registration-based propagation for whole heart segmentation from compounded 3D echocardiography. In: *IEEE International Symposium on Biomedical Imaging*. pp. 1093–1096. IEEE (2010)