

Probabilistic Edge Map (PEM) for 3D Ultrasound Image Registration and Multi-atlas Left Ventricle Segmentation

Ozan Oktay¹(✉), Alberto Gomez², Kevin Keraudren¹, Andreas Schuh¹,
Wenjia Bai¹, Wenzhe Shi¹, Graeme Penney², and Daniel Rueckert¹

¹ Biomedical Image Analysis Group, Imperial College London, London, UK
o.oktay13@imperial.ac.uk

² Division of Imaging Sciences and Biomedical Engineering,
King's College London, London, UK

Abstract. Automated left ventricle (LV) segmentation in 3D ultrasound (3D-US) remains a challenging research problem due to variable image quality and limited field-of-view. Modern segmentation approaches (shape, appearance and contour model based surface fitting) require an accurate initialization and good image boundary features to obtain reliable and consistent results. They are therefore not well suited for this problem. The proposed method overcomes those limitations with a novel and generic 3D-US image boundary representation technique: Probabilistic Edge Map (PEM). This new representation captures regularized and complete edge responses from standard 3D-US images. PEM is utilized in a multi-atlas LV segmentation framework to spatially align target and atlas images. Experiments on data from the MICCAI CETUS challenge show that the proposed approach is better suited for LV segmentation than the active contour, appearance and voxel classification approaches, achieving lower surface distance errors and better LV volume estimates.

Keywords: Structured decision forest · Probabilistic edge map · Multi-atlas label fusion · Left ventricle segmentation · Ultrasound image analysis

1 Introduction

Cardiac ultrasound remains the primary imaging modality in the assessment of left ventricular systolic function, mass and volume to assess the morphology and function of the heart. Automated tools to analyse three-dimensional ultrasound (3D-US) images are important to ensure reproducibility as well as consistency of segmentations and to reduce the workload of clinicians. The development of such tools is still an ongoing research problem due to limitations posed by low image quality, restricted field-of-view and anatomical variations. For these reasons, accurate and generic image analysis techniques are crucial.

Related Work: Automated left ventricle (LV) segmentation techniques can be broadly categorized into two groups: (1) image-driven and (2) model-driven

approaches. Level-set approaches such as phase asymmetry [13] are part of the first category. They calculate 3D LV surfaces with weak or no shape constraints and do not require the fitting of a model to a large number of images. Also the B-spline active surface approach proposed in [4] does not require model fitting. Instead, the surface is initialized with an ellipsoid and B-splines are used to regularize the deformation of the surface model. Approaches in the second group use additional a-priori information by analyzing intensity patterns in training samples and manually traced contours. This includes approaches such as appearance models (AAM) [15] and semantic labelling of voxels using a classifier such as a decision forest [9]. Another method proposed in [10] uses labeled atlases and image registration to segment the LV volume. It does not require the training of a shape model, but makes an implicit use of such model through the atlases.

Research Motivation and Method Proposal: Active contour and level-set approaches require an accurate estimate of LV shape and position for initialization. This is because final segmentation results are sensitive to initializations obtained either manually [7, 10] or through ad-hoc solutions such as Hough transform of edges [4] or through selection of image center points [15]. Such approaches depend on the acquisition field-of-view and cannot be generalized to acquisitions from different acoustic windows such as apical and parasternal views together.

Similarly, these approaches [4, 13, 15] make use of intensity and phase based features to delineate ventricle borders. Since phase features rely on the agreement of phases between different Fourier components (and are therefore insensitive to contrast), less importance is given to local energy information. This causes these features to be sensitive to noise. Likewise, intensity based approaches are sensitive to low image quality, shadowing, speckle and clutter.

This paper proposes a fully automatic multi-atlas LV segmentation framework for US images. Additionally, a novel robust 3D boundary representation method, Probabilistic Edge Map (PEM), is presented and utilized within this framework to address the challenges outlined above. PEMs delineate object boundaries in the input images by using a trained structured decision forest (SDF) classifier [6]. With this method, we are extending the structural representation proposed in [7], applied on 2D cardiac short-axis slices, to a 3D structural analysis together with the use of US related image features. In this way, discontinuous and spurious edge responses in through plane direction can be eliminated, while achieving smooth and regularized tissue boundaries, as shown in Fig. 1.

In the proposed multi-atlas LV segmentation framework (PEM-MA), the PEMs are used in robust affine registration [11] and non-rigid registration [14] to spatially align multiple atlas images to the target. PEM based US image registration provides more reliable initialization between target and atlas images, and achieves better atlas selection [1] and LV segmentation performance. The proposed segmentation framework is evaluated on a benchmark dataset used in the MICCAI 2014 CETUS segmentation challenge. The results collected from the online evaluation platform show that PEM-MA achieves state-of-the-art LV segmentation accuracy in both surface distance and volumetric measure metrics, while outperforming all other challenge participants [3, 7, 15] in terms of the used evaluation criteria.

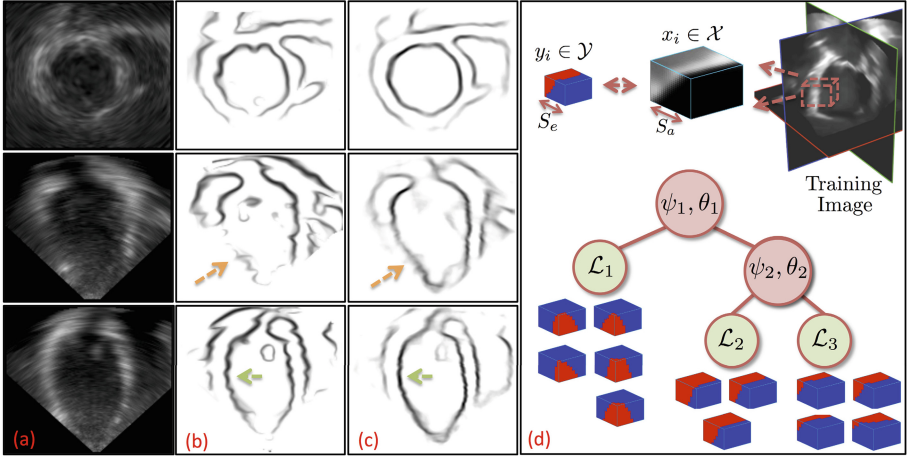


Fig. 1. (a) 3D cardiac US image, (b) phase congruency [13], and (c) PEM which captures missing structures (orange arrows) and provides smoother edge response (green arrows). In (d) SDF training is illustrated, where the label patches (y_i) are clustered at each node split, and the weak learners (ψ_i) search for the optimal threshold value (θ_i) and feature (x_i) to separate the two clusters (Color figure online).

2 Methodology

2.1 Probabilistic Edge Map (PEM) Representation

In cardiac imaging, 3D-US images outline an anatomical representation of the heart chambers. Further image analysis typically requires an accurate and smooth object boundary delineation. Data driven approaches may fail due to severe intensity artefacts and missing boundaries. A machine learning approach such as a structured decision forest (SDF) [6] can cope with these difficulties as the training data guides the boundary extraction. This is shown in Fig. 1, where the proposed PEM captures the missing boundaries and delineates them accurately.

The US images are initially resampled to isotropic voxel size. Furthermore, speckle noise is reduced using a sparse coding approach: The K-SVD algorithm [8] is used to learn an over-complete dictionary from US image patches. After the learning stage, the image is reconstructed from a sparse combination of the learned dictionary atoms to remove speckle patterns. Finally, a SDF classifier for the PEM is trained from the preprocessed images. While SDFs are similar to decision forests, they possess several unique properties and advantages.

In the tree structure of SDF, the output space (\mathcal{Y}) is assumed to be structured. In our case, this means that the output labels ($y_i \in \mathcal{Y}$) of size $(S_e)^3$ represent the edge labelling for image patches. In general, any type of multi-dimensional output can be stored at each tree leaf node, as long as labels can be clustered into two or more subsets by determining the optimal splitting function (ψ) at each tree branch, as shown in Fig. 1(d). In the PEM classifier training,

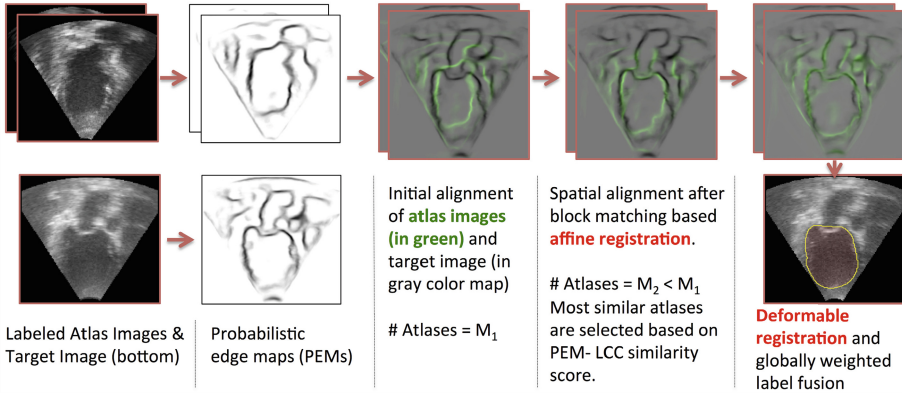


Fig. 2. A block diagram of the proposed multi-atlas segmentation framework.

this is achieved by mapping each image patch label to an intermediate space ($\Theta : \mathcal{Y} \rightarrow \mathcal{Z}$) where label clusters can be generated based on the Euclidean distance in \mathcal{Z} (cf. [6]). Similar to decision forests, SDFs operate on standard input feature space which is defined by the high dimensional appearance features ($x_i \in \mathcal{X}$) extracted from image patches of fixed size (S_a)³. These features are computed in a multi-scale fashion and correspond to image intensities, gradient magnitudes, soft-binning based histogram of oriented gradients, and local phase features. Weak classifiers $\psi(x_i, \theta)$, e.g., 1D and 2D decision stumps, are trained by maximizing the entropy based information gain criterion at each tree node with one of the selected image features. The parameter vector θ contains the stump threshold value and selected feature identifier. At testing time, each target image voxel is voted for (S_e)³ $\times N_t$ times by N_t number of trees and these votes are aggregated by averaging all the predictions. Multiple and overlapping patch label predictions are the main advantage of PEMs, as these result in smooth, regularized and complete delineations of the cardiac chambers.

2.2 Multi-atlas Left Ventricle Segmentation

Next, we detail our proposed multi-atlas LV segmentation framework as outlined in Fig. 2, employing the generated edge maps. Initial affine alignment, atlas selection and deformable registration between target (I) and atlas images (J_i) are performed based on the PEMs (P^I, P_i^J) generated from the US images. A dataset consisting of a number of manually annotated US images is used in the atlas formation. The annotations for these atlases contain only the LV endocardial labels. The composite spatial transformations transfer the atlas labels to the target, followed by a globally weighted label fusion based on PEM similarity.

Global Alignment: The PEMs from both target image and atlases are first aligned using a block matching technique [11] which maximizes the normalized correlation coefficient between image blocks. The set of vectors defined by the

displacement of each block is regularized before finding the global affine transformation A_i . A least trimmed squared regression based regularization (cf. [11]) removes the influence of displacements for the atlas blocks which have no target block correspondence due to missing features in the images. For this reason, this approach is robust to shadowing and anatomical variations and can provide an accurate spatial alignment for atlas selection and good initial segmentation.

Atlas Selection: It was shown in multi-atlas brain segmentation [1], that a selection of most similar atlases is beneficial. Therefore, after affine registration, all M_1 atlases are ranked according to their average local correlation coefficient [5] score, $LCC(P^I, P_i^J \circ A_i)$, and the $M_2 < M_1$ top scoring atlases in the upper quartile are selected. The LCC similarity metric is defined in (1), where Ω denotes the target voxels within a region defined by the dilated LV mask.

$$LCC(P^I, P^J) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \frac{|\langle P^I, P^J \rangle_x|}{\sqrt{\langle P^I, P^I \rangle_x \langle P^J, P^J \rangle_x}} \quad (1)$$

A Gaussian window G_σ with variance σ^2 locally weights the PEMs and $\langle P^I, P^J \rangle_x = G_\sigma * (P^I \cdot P^J)[x] - (G_\sigma * P^I)[x](G_\sigma * P^J)[x]$, where \cdot denotes the Hadamard product, and $*$ the convolution. As the SDF classifier makes use of image intensities in node splits ψ , local intensity changes in the input images can influence the edge probabilities in PEMs. For this reason, LCC is a more suitable similarity measure for PEMs than global metrics such as sum of squared differences.

Local Alignment: To correct for residual misalignment, a registration based on free-form deformations (FFDs) [14] follows the atlas selection. The total energy $E(T_i) = -LCC(P^I, P_i^J \circ T_i \circ A_i) + \lambda BE(T_i)$ is minimised in a multi-resolution scheme, where BE is the bending energy of the cubic B-spline FFD T_i and λ defines the trade-off between local PEM alignment and deformation smoothness.

Label Fusion: Finally, the transferred atlas labels are fused using a globally weighted voting¹ [2] based on the dissimilarity $m_i = 1 - LCC(P^I, P_i^J \circ T_i \circ A_i)$. The LV segmentation of the target image is then given by the labelling function $S^I(x) = \arg \max_{l \in \{0,1\}} \sum_{i=1}^{M_2} w_i \cdot \delta(S_i^J(x) - l)$, where δ is the Dirac delta function and global weights $w_i = \exp(-m_i / \frac{1}{M_2} \sum_{j=1}^{M_2} m_j)$. In this fusion strategy, atlases more similar (higher LCC score) to the target image have a stronger influence on the final segmentation and those with a relatively lower score are downgraded.

3 Algorithm Evaluation

The proposed segmentation framework is evaluated on a benchmark dataset used in the MICCAI 2014 CETUS challenge [12]. It consists of 4D echo sequences acquired from an apical window in healthy volunteers and patients with myocardial infarction and dilative cardiomyopathy. The dataset is divided into 15 training and 30 testing image sequences. Contours of the heart chambers were outlined

¹ Locally weighted and majority voting fusion methods were also evaluated in the experiments, and the best results were obtained with the global fusion method.

Table 1. LV segmentation results on 30 subjects (CETUS challenge testing dataset [12]). Mean distance (MD [mm]), Hausdorff distance (HD [mm]) and Dice coefficient (DC [%]) results are listed separately for ED and ES frames.

	MD_{ED}	MD_{ES}	HD_{ED}	HD_{ES}	DC_{ED}	DC_{ES}
Manual [12]	$1.01 \pm .30$	$1.01 \pm .38$	$3.37 \pm .87$	$3.30 \pm .94$	$0.949 \pm .15$	0.938 ± 0.21
AAM [15]	$2.44 \pm .xx$	$2.79 \pm .xx$	$8.45 \pm x.xx$	$8.65 \pm x.xx$	$0.879 \pm .xx$	$0.835 \pm .xx$
BEAS [3,4]	$2.26 \pm .xx$	$2.43 \pm .xx$	$8.10 \pm x.xx$	$8.13 \pm x.xx$	$0.894 \pm .xx$	$0.856 \pm .xx$
SE-MA [10]	$2.18 \pm .70$	$2.47 \pm .74$	7.55 ± 1.76	8.57 ± 2.96	$0.894 \pm .03$	$0.849 \pm .04$
SDF-LS [7]	$2.09 \pm .xx$	$2.20 \pm .xx$	$9.31 \pm x.xx$	$8.35 \pm x.xx$	$0.894 \pm .xx$	$0.871 \pm .xx$
PEM-MA	$1.94 \pm .55$	$2.23 \pm .60$	7.00 ± 1.99	7.53 ± 2.23	$0.904 \pm .02$	$0.874 \pm .04$

by three experts, but only those of the training set are publicly available. Therefore, the CETUS web site² is used for evaluation. Submissions are automatically evaluated based on surface distance errors and clinical LV volumetric indices.

In all experiments, segmentations are computed only for end-diastolic (ED) and end-systolic (ES) phases. Table 1 lists the surface distance errors obtained in the first experiment. The proposed PEM-MA framework achieves better results than the challenge top performing algorithms: AAM [15] (active appearance model), BEAS [3,4] (B-spline active contours), SDF-LS (structured decision forest followed by level-set segmentation), and SE-MA [10] (spectral embedding multi-atlas method). The inter-observer manual segmentation [12] variations are reported for comparison. We can conclude that PEMs provide a better boundary representation than spectral features [10] based on mean ($p < 0.01$) and Hausdorff distance ($p < 0.01$). Moreover, the proposed approach does not require landmark selection [10] or manual affine alignment of LV surface template to initialize the segmentation [7].

The difference in segmentation accuracy between PEM-MA and model based surface fitting methods (AAM, BEAS) can be explained as follows. The proposed approach employs affinely aligned atlas labels as shape priors which are selected based on LCC similarity of PEMs, whereas the other methods use less data specific priors such as mean LV shape [15] and ellipsoid [4] shape assumption. Similarly, in PEM-MA, the LV segmentation is initialized with position priors obtained through a robust affine block matching of PEMs. This delineates the left ventricle position in the image more accurately than Hough transform [4] and the mean LV position of the training images [15].

In the second experiment, clinical indices, such as ejection fraction (EF), ED and ES volume values, are computed from the proposed segmentation approach. The obtained results are compared against their reference values using the aforementioned web site. The results in Table 2 show that PEM-MA achieves a better agreement with the ground truth compared to the other methods. As PEM-MA delineates LV boundaries more accurately, better volume estimates are obtained. Additionally, we observe that PEM-MA displays a consistent performance in

² <https://miccai.creatis.insa-lyon.fr/miccai/community/1>.

Table 2. Segmentation results on 30 images (CETUS testing data [12]). Pearson’s correlation coefficient (corr) and Bland-Altman ($\mu \pm 1.96\sigma$) limit of agreement (LOA) between ground-truth and estimated LV volume values are reported.

	ED_{corr}	ED_{LOA}	ES_{corr}	ES_{LOA}	EF_{corr}	EF_{LOA}
Manual [12]	0.981	-0.636 ± 18.2	0.987	-0.50 ± 14.4	0.959	0.13 ± 6.07
AAM [15]	0.966	-15.42 ± 32.1	0.964	-13.2 ± 28.9	0.611	3.69 ± 17.58
SE-MA [10]	0.945	-6.02 ± 41.6	0.924	-0.42 ± 41.2	0.780	-1.55 ± 13.88
BEAS [3, 4]	0.965	-4.99 ± 35.3	0.967	-6.78 ± 27.7	0.889	2.88 ± 10.48
SDF-LS [7]	0.917	8.73 ± 49.9	0.956	-5.16 ± 31.7	0.819	8.33 ± 14.46
PEM-MA	0.961	-4.14 ± 34.0	0.973	-3.47 ± 26.7	0.892	0.48 ± 10.78

both LV surface fitting and volume estimation in contrast to SDF-LS. The performance difference between the two can be linked to the improved structural representation and the choice of different surface fitting algorithm.

All experiments were carried out on a 3.00 GHz quad-core machine. The average computation time per image pair was 74s for non-rigid registration, 16s for affine alignment and 20s to compute each PEM. The training of the SDF (70m per tree) and atlas PEM computation were performed offline prior to target segmentations. The segmentation of the LV takes in total 16 m per image. The proposed approach is computationally more complex than the methods in [4, 7] due to the multitude of registrations. However, a parallel implementation of these registrations significantly reduces the total runtime.

Implementation Details: In total $N_t = 8$ PEM decision trees are trained using 20 US sequences plus rotated versions of these images. PEM quality was not improved further by including more trees. Patch sizes for training features and ground-truth edges are chosen as $S_a = 20$ and $S_e = 10$ per dimension. For global alignment, blocks of size 5^3 voxels were used with search radius equal to the block size as in [11]. A multi-scale optimization strategy was employed to capture large displacements and to improve convergence. A total of $M_1 = 30$ ED and ES atlases were aligned to each subject. Of these, on average $M_2 = 6.3$ were selected based on their LCC score, with a standard deviation of the Gaussian $\sigma = 7$ voxels in each dimension.

4 Conclusion

We presented a novel US image representation (PEM) which achieves state-of-the-art cardiac US image registration and LV segmentation results within a multi-atlas framework. The proposed framework outperforms all other methods participating in the MICCAI CETUS challenge based on the obtained surface mesh evaluation criteria. The main contributions of the paper are: (1) highly accurate 3D edge map representation for cardiac US images, and (2) block-matching based robust and accurate initialization technique for automatic LV segmentation. The proposed PEM representation is generic and modular. It has

the potential of being applied to echo images acquired from other organs and does not make assumptions on the acquisition window and image orientation. Additionally, the multi-atlas segmentation framework is shown to be applicable for clinical routine as it can estimate functional indices very accurately.

References

1. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage* **46**(3), 726–738 (2009)
2. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solórzano, C.: Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans. Med. Imag.* **28**, 1266–1277 (2009)
3. Barbosa, D., Friboulet, D., D’hooge, J., Bernard, O.: Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching. In: *Proceedings of MICCAI CETUS Challenge* (2014)
4. 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 Med. Biol.* **39**(1), 89–101 (2013)
5. Cachier, P., Pennec, X.: 3D non-rigid registration by gradient descent on a Gaussian windowed similarity measure using convolutions. In: *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 182–189 (2000)
6. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: *ICCV*, pp. 1841–1848. *IEEE* (2013)
7. Domingos, J.S., Stebbing, R.V., Leeson, P., Noble, J.A.: Structured random forests for myocardium delineation in 3D echocardiography. In: Wu, G., Zhang, D., Zhou, L. (eds.) *MLMI 2014*. LNCS, vol. 8679, pp. 215–222. Springer, Heidelberg (2014)
8. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006)
9. Lempitsky, V., Verhoek, M., Noble, J.A., Blake, A.: Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) *FIMH 2009*. LNCS, vol. 5528, pp. 447–456. Springer, Heidelberg (2009)
10. Oktay, O., Shi, W., Caballero, J., Keraudren, K., Rueckert, D.: Sparsity based spectral embedding: application to multi-atlas echocardiography segmentation. In: *Proceedings of MICCAI STMI Workshop* (2014)
11. Ourselin, S., Roche, A., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. *Image Vis. Comput.* **19**(1), 25–31 (2001)
12. Papachristidis, A., et al.: Clinical expert delineation of 3D left ventricular echocardiograms for the CETUS segmentation challenge. In: *Proceedings of MICCAI CETUS Challenge*, pp. 9–16 (2014)
13. 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. *MedIA* **15**(4), 514–528 (2011)
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 Trans. Med. Imag.* **18**(8), 712–721 (1999)
15. Stralen, M.V., Haak, A., Leung, K., Burken, G.V., Bosch, J.: Segmentation of multi-center 3D left ventricular echocardiograms by active appearance models. In: *Proceedings of MICCAI CETUS Challenge*, pp. 73–80 (2014)