Endocardial Segmentation using Structured Random Forests in 3D Echocardiography

João S. Domingos[1], Richard V. Stebbing[1], and J. Alison Noble[1]

Institute of Biomedical Engineering, University of Oxford, U.K.
domingos.domingos@eng.ox.ac.uk

Abstract. Segmentation of the left ventricle endocardium in 3D echocardiography is a critical step for the diagnosis of heart disease. Although recent work has shown effective endocardial edge detection, these techniques still preserve spurious anatomical edge responses that undermine overall ventricle segmentation. In this paper we propose a robust semi-automatic framework based on 2D structured learning that facilitates full 3D model-based endocardial segmentation. This method is evaluated on 30 publicly available datasets from different brands of ultrasound machines. Results show that the proposed method accurately finds the endocardium and effectively converges an explicit and continuous surface model to it.

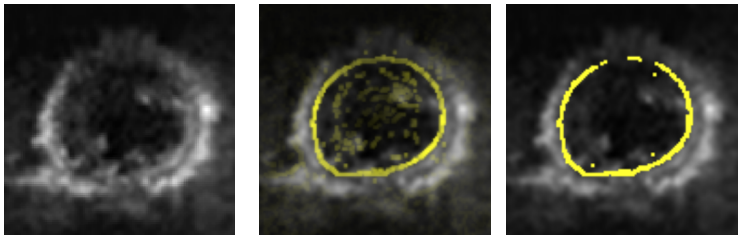
1 Introduction

In this paper we propose a two-step method to perform left ventricle (LV) endocardial segmentation from 3D echocardiography volumes.

LV endocardial boundary candidates are detected on each short-axis slice (2D approach) of each volume using a Structured Random Forest (SRF) [1, 2] to first regress the probability that each pixel belongs to the endocardium-blood boundary. The motivation for using a SRF for boundary detection is that it automatically learns to discard strong image edge responses which do *not* correspond to the endocardium-blood boundary [3]. These spurious edge responses — arising from the presence of papillary muscles, trabeculae, imaging artefacts, and speckle — are typically retained as boundary candidates by standard edge detectors.

Following detection, an *explicit* continuous surface model of the LV is deformed to a subset of the boundary candidates in each volume. The "surface fit" is defined as the distance of each boundary candidate to its *preimage* on the continuous surface [4, 5]. The motivation for using an explicit surface model instead of an *implicit* level set representation is that the explicit surface model has fewer model parameters and is amenable to non-linear optimisation methods more powerful and robust than gradient descent.

Recent and common approaches to endocardial segmentation include Active Contours [6] and Active Appearance Models [7] (*AC/AAM*). Our method is similar in spirit — we fit a parametric surface model to detected boundary



(a) (b) (c)

Fig. 1: **Endocardial boundary detection in 2D** (a) Short-axis slice. (b) The probability of a pixel belonging to the endocardium-blood boundary superimposed in yellow (non-maximum suppression applied to the output of the SRF). (c) Final boundary candidates.

candidates — but has two subtle and important differences. First, endocardial boundary candidates are detected separately from the model surface. This is similar to [8] where a parametric surface model is initialised and deformed *after* first classifying all edge responses as either endocardium, epicardium, or background boundary candidates. Second, we do *not* fix the boundary candidate preimages during the optimisation of the model surface. This is identical to [4] and [5] and it has the advantage that we do not have to make assumptions about how boundary candidates are distributed relative to the model surface. Specifically, we are not restricted to selecting only boundary candidates which are perpendicular to the model surface. Instead, we use all boundary candidates that are in the vicinity of the model surface to obtain an accurate final segmentation.

2 Method

2.1 Endocardial boundary detection

Given an input echocardiography volume, we produce a matrix of positions of boundary candidates which describe the LV endocardium-blood interface.

Processing each short-axis slice independently (Fig. 1a) we first use a SRF to predict the probability of each pixel belonging to the endocardium-blood interface from a given image patch. A SRF is similar to a standard Random Forest (RF) classifier used in image segmentation (e.g. [9]) except that we store edge maps at the leaf nodes, hence the resulting output is extended to an high dimensional structured output space $\mathcal{Y}$, and ultimately output *labels* are assigned to multiple pixels *jointly* instead of independently. That is, prediction is performed patch-by-patch instead of pixel-by-pixel and the output *structured labels* are vector quantities instead of scalars [2]. The motivating reason for using a SRF instead of a standard RF is that label structure present in training data is preserved when making a prediction.

In the context of multiclass classification, the standard information gain criterion is not well defined over structured labels, $y \in \mathcal{Y}$. To overcome this problem, the authors in [1] proposed a two-step mapping approach, $\mathcal{Y} \rightarrow \mathcal{Z}$ and $\mathcal{Z} \rightarrow \mathcal{C}$, with an intermediate space $\mathcal{Z}$ in which the Euclidean distance can be measured and the similarity over $\mathcal{Y}$ finally estimated. More specifically, because only an approximate measure of information gain is sufficient to train a random forest classifier, we map a set of structured labels $y \in \mathcal{Y}$ into a discrete set of labels $c \in \mathcal{C}$, so that labels with similar $\mathcal{Y}$ are assigned to the same discrete label c. In our method, these discrete labels are binary ($k = 2$) and the Gini impurity is used as the standard information gain measure. The discretisation step ($\mathcal{Z} \rightarrow \mathcal{C}$) yielding the discrete label set $\mathcal{C}$ given $\mathcal{Z}$ is computed independently when training each node and depends on the distribution of labels at each node. To do this, z is quantized based on the top $log_2(k)$ PCA dimensions, effectively assigning z a discrete label c according to the orthant into which z falls.

The final structured ensemble model merges a set of n labels $y_1 ... y_n \in \mathcal{Y}$ into a single prediction by effectively selecting the label y_k whose z_k is the medoid (the medoid z_k that minimizes $\sum_{ij}(z_{kj} - z_{ij})^2$). The resulting ensemble model is computationally efficient because it uses structured labels, capturing information for an entire image neighbourhood, thus reducing the number of decision trees T that need to be evaluated per pixel. Classical multiscale of the model is ultimately computed by averaging the result of three probability edge maps at the original, half and double resolution.

Given an input short-axis slice from an End-Diastolic or -Systolic (ED, ES) volume, the proposed Structured Edge Detector (SED) task is to label each pixel with a binary variable indicating whether it belongs to an edge or not. This was done by predicting a structured 16×16 segmentation patch within a larger 32×32 image patch. Each image patch was augmented with multiple gradient channels of information. We computed normalised gradient magnitude at 2 scales which were then split into 4 channels based on orientation (see [10] for more details). Let $X \in \mathbb{R}^{32 \times 32 \times K}$ denote the augmented image patch with K channels. The two types of features we used were *pixel lookup* — $I[r, c, k]$ — and *pairwise differences* — $I[r_1, c_1, k] - I[r_0, c_0, k]$, where lowercase variables are row, column and channel indices and constitute the feature parameters.

Parameter sweeps were run on the publicly available database (15 training + 15 testing datasets) provided by the 2014 MICCAI CETUS challenge. Importantly, the 15 patients with ground-truth segmentation surfaces were used to generate the required training data to train the SRF. Optimal settings for a subset of the SRF parameters (e.g. number of structured labels, patch dimensions, number of trees, maximum tree depth) were determined by applying the SRF to the *other* 15 patients, fitting the endocardial surface (§ 2.2) and using the MIDAS platform to measure distance error and clinical indices. Along with the parameters specified previously, a configuration of 2 structured labels and 8 trees evaluated on an alternating set of 4 trees — each to a maximum tree depth of 64 — achieved the minimum mean surface distance and modified Dice

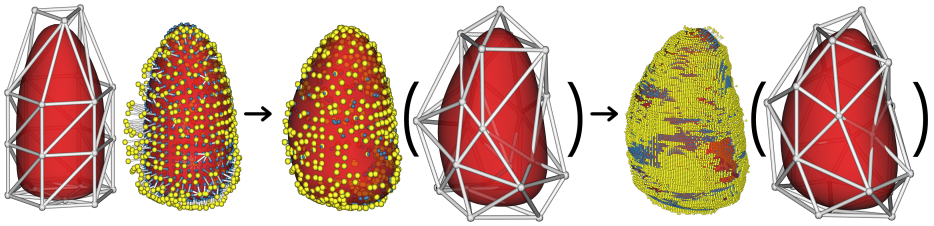


Fig. 2: **Endocardial surface fitting** (Left) The limit surface (red) is defined by the control mesh (gray). Surface points (blue) are initialised and boundary candidates are selected (yellow) ("model-to-data"). (Centre) The fitted surface with initial boundary candidates. (Right) The fitted surface with final boundary candidates ("data-to-model").

error for the 15 test patients. However, the total variation in error of the mean surface distance was only 0.7mm across all configurations tested.

Following the application of the SRF to each slice, we applied non-maximum suppression (Fig. 1b) by suppressing locations where edges are stronger in the orthogonal direction, and retained only the most probable 2% of candidates to produce the final matrix of boundary candidate positions (Fig. 1c).

2.2 Endocardial surface fitting

The explicit and continuous surface model of the LV is defined by a matrix of 24 control vertices $X \in \mathbb{R}^{3\times 24}$ (Fig. 2). Positions on the *limit surface* are defined by the function $\chi(\boldsymbol{u}, X)$ where $\boldsymbol{u} \in \Omega$ is a surface coordinate and $\Omega \subset \mathbb{R}^2$ is the surface domain. χ is defined using Loop subdivision [11] which generalises the definition of quadratic triangular B-splines to topologically irregular meshes [12, 5]. The important properties of χ are that it is non-linear in $\boldsymbol{u}$ but linear in X.

Let ϕ denote the matrix of *all* boundary candidates (§ 2.1, Fig. 1) and $\boldsymbol{l}$ denote the vector of *selected* boundary candidate indices (the exact specification of $\boldsymbol{l}$ is deferred until later). Also, let $N = |\boldsymbol{l}|$ denote the number of selected boundary candidates, $\boldsymbol{\phi}_l$ denote the l^{th} entry of ϕ, and l_i denote the i^{th} entry of $\boldsymbol{l}$. The surface "fit" is defined as:

$$E(X \mid \boldsymbol{l}) = \lambda \sum_{i=1}^{N} \min_{\boldsymbol{u}\in\Omega} \|\boldsymbol{\phi}_{l_i} - \chi(\boldsymbol{u}, X)\|^2 + \mathcal{R}(X) \tag{1}$$

where λ controls the weight of each error term. The min over $\boldsymbol{u}$ in (1) ensures that for each selected boundary candidate the distance to the surface is measured to its closest corresponding point on the surface — its *preimage* [4, 5]. $\mathcal{R}(X)$ is the thin-plate regulariser given by[1]:

$$\mathcal{R}(X) = \int_{\Omega_x} \|\chi_{xx}(\boldsymbol{x}, X)\|^2 + 2\|\chi_{xy}(\boldsymbol{x}, X)\|^2 + \|\chi_{yy}(\boldsymbol{x}, X)\|^2 d\boldsymbol{x} \tag{2}$$

[1] The reparameterisation of χ in $\boldsymbol{x} = (x, y)$ is required to obtain an isotropic definition [4].

Local continuous optimisation of E with respect to X is inhibited by the presence of the min in the summation and the integral in (2). However, E can be redefined with *all* preimages introduced as additional continuous unknown parameters:

$$\tilde{E}(X, \{\boldsymbol{u}_i\}_{i=1}^N \mid \boldsymbol{l}) = \lambda \sum_{i=1}^{N} \|\boldsymbol{\phi}_{l_i} - \chi(\boldsymbol{u}_i, X)\|^2 + \mathcal{R}(X) \tag{3}$$

so that minimisation of $\tilde{E}$ with respect to X *and* $\{\boldsymbol{u}\}_{i=1}^N$ is equivalent to minimisation of E with respect to X alone [5]. Furthermore, since χ is linear in X, the integral in (2) can be evaluated and reduced to a quadratic in sum-of-squares form. As a result, $\tilde{E}$ can be minimised using the Levenberg-Marquardt algorithm [13], taking care to handle the updates of each $\boldsymbol{u}_i$ so that preimages move smoothly across the limit surface [5]. Finally, to ensure that the surface fitting is robust to incorrectly selected boundary candidates, a Huber loss function[2] [13] is applied to each weighted square residual.

With reference to Fig. 2, the specification of the boundary candidate indices $\boldsymbol{l}$ and the complete optimisation schedule for X is now described. Starting with an initial[3] X, $\{\boldsymbol{u}_i\}$ was initialised by sampling a fixed number of points uniformly in the surface domain Ω. Each corresponding l was then set to its closest entry in ϕ — this is standard "model-to-data" boundary candidate selection. To allow for missing boundaries, all $\boldsymbol{u}_i$ with selected boundary candidates that had a distance to the surface exceeding 10^{-2}m were discarded. With $\{\boldsymbol{u}_i\}_{i=1}^N$ initialised and $\boldsymbol{l}$ set, X was then refined by minimising $\tilde{E}$ with $\lambda = 2^8$. Following the initial refinement, $\boldsymbol{l}$ was reset to include *all* entries in ϕ which were within a given radius (5mm) of the limit surface, and each corresponding $\boldsymbol{u}_i$ was initialised to the closest surface correspondence. The advantage of this "data-to-model" step is that it utilises as many boundary candidates as possible (typically in the order of 10^3). Finally, X was refined once more by minimising $\tilde{E}$ with $\lambda = 2^{15}/N$. The final segmentation surface was then produced by densely evaluating the limit surface.

3 Experimental Results

All 60 frames were processed in 960s (16s per volume) on 8 cores of an Intel Core i7-4930MX. The computational bottleneck was the second stage of "data-to-model" surface fitting. The results are presented in Table 1. Example slices from the worst performing patients are shown in Fig. 4. $SED1$ refers to the Structured Edge Detector which performed best on the training database. Similarly, $SED2$ refers to the endocardial boundary detector which performed best on the testing database. The detectors have slightly different configurations: SED1 uses 16 trees while SED2 uses 4.

[2] A scale of 0.025 is used in all experiments.

[3] This is provided by the user using a 3D interactive GUI and takes typically 5-15s.

	$d_{mean,ED}$ (μ,mm)	$d_{mean,ES}$ (μ,mm)	$d_{H,ED}$ (μ,mm)	$d_{H,ES}$ (μ,mm)	D^*_{ED}	D^*_{ES}	EF_{r^*}	EF_{bias} (μ,%)	EF_{LOA} ($\pm 1.96\sigma$,%)
$SED1_{onTraining}$	**0.97**	**1.17**	**5.79**	**6.65**	**0.039**	**0.055**	**0.986**	**2.7**	**[-2.76 ; 8.16]**
$SED1_{onTesting}$	1.98	2.46	7.32	9.36	0.099	0.139	0.811	10.65	[-1.7 ; 23]
$SED2_{onTraining}$	1.12	1.25	6.26	6.66	0.045	0.059	0.983	3.16	[-2.84 ; 9.16]
$SED2_{onTesting}$	**2**	**2.17**	**9.11**	**8.7**	**0.103**	**0.128**	**0.847**	**8.97**	**[-2.15 ; 20.08]**

	EDV_{r^*}	EDV_{bias} (μ,mL)	EDV_{LOA} ($\pm 1.96\sigma$,mL)	ESV_{r^*}	ESV_{bias} (μ,mL)	ESV_{LOA} ($\pm 1.96\sigma$,mL)
$SED1_{onTraining}$	**0.997**	**-3.51**	**[-18.34 ; 11.32]**	**0.997**	**-8.01**	**[-24.87 ; 8.85]**
$SED1_{onTesting}$	0.887	-2.94	[-47.71 ; 41.83]	0.962	-15.41	[-38.93 ; 8.11]
$SED2_{onTraining}$	0.997	-0.45	[-15.73 ; 14.83]	0.996	-6.45	[-23.69 ; 10.79]
$SED2_{onTesting}$	**0.892**	**1.03**	**[-41.51 ; 43.57]**	**0.972**	**-10.48**	**[-29.58 ; 8.62]**

Table 1: Summary of the distance errors and clinical indices: mean surface distance (d_{mean}), Hausdorff surface distance (d_H), modified Dice similarity index ($D^* = 1 - D$), Ejection Fraction (EF), End Diastolic Volume and End Systolic volume (ESV) measurements, each with modified Pearson product-moment correlation coefficient ($r^* = 1 - r$) and Bland-Altman bias and Limits of Agreement (LOA).

4 Discussion

Table 1 demonstrates that the proposed method achieves high endocardial border detection accuracy. Our method shows very strong correlation against the reference values for EF, EDV, and ESV, but poor correlation against SV on the Testing database ($SV_{r^*} = 0.322$). Bland-Altman analysis of the estimated indices shows that EF, SV and ESV bias are statistically significant at the 5% significance level (paired-sample t-test, $p < 0.05$) while EDV is not. Although the estimated EDV bias is close to zero, it exhibits high variance and wide limits of agreement (Fig. 3(c)). Finally, Fig. 3(a) and (d) demonstrate that the proposed framework is consistently underestimating EF and overestimating the ESV measurements.

Importantly, our current framework does not include a step to cut the segmentation at the basal region of the heart. Fig. 4 shows that this procedure is required.

5 Conclusion

A semi-automatic method based on 2D structured learning was developed to perform consistent endocardial segmentation in 3D echocardiography volumes from different brands of ultrasound machines.

Acknowledgements. This work was supported by the RCUK CDT in Healthcare Innovation, EPSRC grant EP/G030693/1, and the Rhodes Trust.

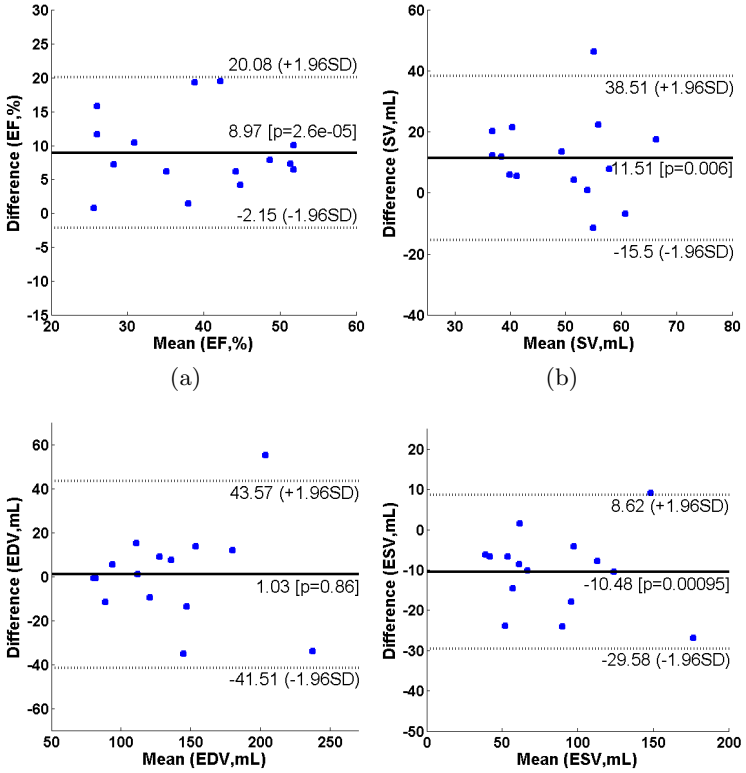
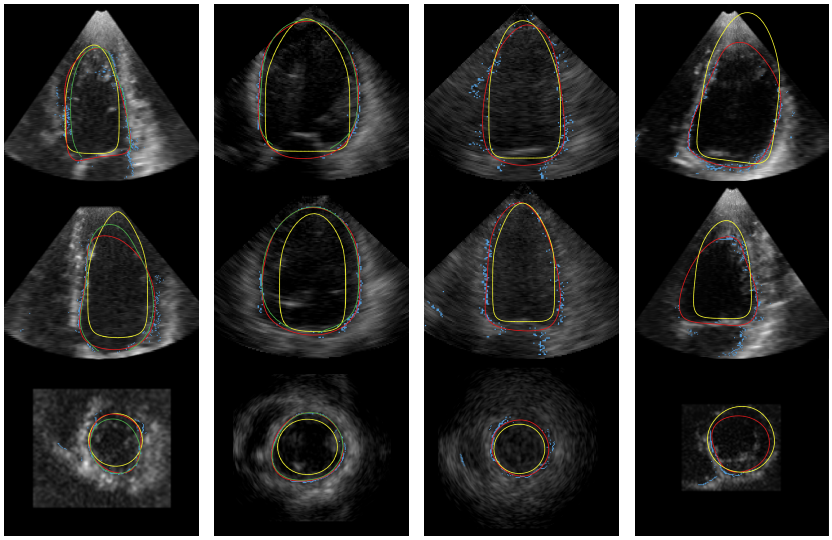


Fig. 3: Bland-Altman plots with bias and LOA for the $SED2_{onTesting}$ clinical indices measurements: (a) EF, (b) Stroke Volume (SV, $SV_{r^*} = 0.322$), (c) EDV and (d) ESV.

References

1. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: Computer Vision (ICCV), 2013 IEEE Conference on, IEEE (2013)
2. Nowozin, S., Lampert, C.H.: Structured learning and prediction in computer vision. Volume 6. Now Publishers Inc (2011)
3. Domingos, J., Stebbing, R., Leeson, P., Noble, J.: Structured random forests for myocardium delineation in 3d echocardiography. In: MICCAI Workshop on Machine Learning in Medical Imaging. (2014)
4. Cashman, T.J., Fitzgibbon, A.W.: What shape are dolphins? Building 3D morphable models from 2D images. IEEE TPAMI **35**(1) (2013) 232–244
5. Taylor, J., Stebbing, R.V., Ramakrishna, V., Keskin, C., Shotton, J., Izadi, S., Hertzmann, A., Fitzgibbon, A.W.: User-specific hand modelling from monocular depth sequences. In: CVPR, IEEE (2014)
6. Barbosa, D., Heyde, B., Cikes, M., Dietenbeck, T., Claus, P., Friboulet, D., Bernard, O., Dhooge, J.: Real-time 3d interactive segmentation of echocardio-



(a) (b) (c) (d)

Fig. 4: **SED2 worst results for the clinical indices** Example slices of the worst performing patients with selected boundary candidates (blue), initialisations (yellow), final segmentations (red) and reference segmentations (green) superimposed. (a) `Patient8_ES` (top and bottom) and `Patient8_ED` (middle), worst by EF, ESV and SV. (b) `Patient13_ES` (top and bottom) and `Patient13_ED` (middle), worst by EDV and ESV. (c) `Patient18_ED` (top and bottom) and `Patient18_ES` (middle), worst by EDV and ESV. (d) `Patient19_ED` (top and bottom) and `Patient19_ES` (middle), worst by EDV and ESV.

graphic data through user-based deformation of b-spline explicit active surfaces. Computerized medical imaging and graphics **38**(1) (2014) 57–67
7. Leung, K.E., van Stralen, M., van Burken, G., de Jong, N., Bosch, J.G.: Automatic active appearance model segmentation of 3d echocardiograms. In: Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on, IEEE (2010) 320–323
8. Stebbing, R.V., Noble, J.A.: Delineating anatomical boundaries using the boundary fragment model. Medical image analysis **17**(8) (2013) 1123–1136
9. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. Communications of the ACM **56**(1) (2013) 116–124
10. Lim, J.J., Zitnick, C.L., Dollár, P.: Sketch tokens: A learned mid-level representation for contour and object detection. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3158–3165
11. Loop, C.: Smooth subdivision surfaces based on triangles. (1987)
12. Stam, J.: Evaluation of Loop subdivision surfaces. (1998)
13. Agarwal, S., Mierle, K., Others: Ceres solver. `https://code.google.com/p/ceres-solver/` (2010–)