Structured Random Forests for Myocardium Delineation in 3D Echocardiography

João S. Domingos[1], Richard V. Stebbing[1], Paul Leeson[2], and J. Alison Noble[1]

[1] Department of Engineering Science, University of Oxford, U.K.
[2] Department of Cardiovascular Medicine, John Radcliffe Hospital, Oxford, U.K.

Abstract. Delineation of myocardium borders from 3D echocardiography is a critical step for the diagnosis of heart disease. Following the approach of myocardium segmentation as a contour finding task, recent work has shown effective methods to interpret endocardial edge information in the left ventricle. Nevertheless, these methods are still prone to preserve irrelevant edge responses and would struggle to overcome chief ventricle anatomical challenges. In this paper we adapt Structured Random Forests, borrowed from computer vision, for fast and robust myocardium edge detection. This method is evaluated on a dataset composed of short-axis slices from 25 End-Diastolic echocardiography volumes. Results show that the proposed ensemble model outperforms standard intensity-based and local phase-based edge detectors, while removing or significantly suppressing irrelevant edges triggered by ultrasound image artefacts and blood pool anatomical structures.

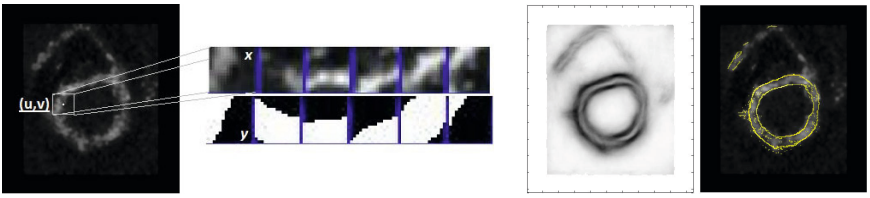
1 Introduction

In this paper we propose a fast and effective method to perform myocardial boundary detection in short-axis slices of 3D Echocardiography (3DE) volumes by integrating structural information of pixel neighbourhoods in classification random forests. These novel Structured Random Forests (SRFs) were introduced in [1] for fast edge detection in computer vision and were adapted here to demonstrate their value in the task of enhancing myocardial boundary. While a truly 3D analysis would be more consistent, slice-by-slice analysis does not lead to notably misaligned contours from observation. Any error has to be traded with the computational cost of a 3D implementation.

Delineation of myocardium borders is a critical step for accurate left ventricle (LV) segmentation and cardiac function quantification. Although LV border delineation has been a widely researched topic, it remains a challenging task mainly due to the anatomical presence of papillary muscles and trabeculae. In addition, there are 3DE image limitations such as speckle, low signal-to-noise ratio, low contrast images and stitching artefacts. In this context, development of computer aided ventricle delineation and segmentation frameworks aimed at improving volumetric analysis in 3DE is of particularly relevance.

Following the approach of myocardium segmentation as a contour finding task, it has previously been shown that intensity-invariant phase-based methods

G. Wu et al. (Eds.): MLMI 2014, LNCS 8679, pp. 215–222, 2014.



(a) (b)

Fig. 1. [**a**] Training data examples as used in our proposed SRFs. While standard random forests associate only the centre label at position (u,v) to an image patch $\mathbf{x}$, we incorporate the topology of the local label neighbourhood ($\mathbf{y}$) and hence learn relevant labelling transitions between myocardium and blood pool. A rich set of structured labels are then used by the ensemble model to select splits in the decision trees. [**b**] LV myocardial edge probability map (*left*) from the slice in [**a**], obtained from our SRFs, and its non-maximal suppression version (*right*).

offer a good alternative to underperforming intensity gradient-based ones in ultrasound images. In [2] a 3D edge detection method was proposed based on a local-phase Feature Asymmetry (FA) measure using the monogenic signal. Motivated by the principle that only the edges that contribute to the myocardium boundary are relevant for segmentation, a 3D Boundary Fragment Model-based method is proposed in [3] to perform anatomical heart boundary delineation. Nevertheless, when accurate myocardium delineation is required, these methods still preserve irrelevant edges. The proposed SRFs use the topological information in local image patches (Figure 1[**a**]) to selectively suppress spurious edge responses and learn only relevant local image neighbourhoods that encode the myocardial boundaries in a structured learning-based approach [4].

2 Methods

2.1 Structured Random Forests

Following a data-driven learning approach, we could firstly propose semantic myocardial boundary detection as a simple binary classification problem. The idea being that a given input image patch can be classified as a positive patch if its centre pixel contains an edge and negative otherwise. Nevertheless, this binary approach ignores valuable local structural information about edges. A multiclass classification approach could then be proposed by simply clustering label (Ground Truth, GT) patches into patch classes. Upon reaching a leaf node, a standard Random Forest (RF) classifier [5] could then directly predict, from a distribution over the labels, the most likely patch class correspondent to the input patch image. With the proposed SRFs, we directly predict local structure of a given image patch, at the cost of a high dimensional output space. As such, standard RFs need to be extended to arbitrary structured output spaces $\mathcal{Y}$.

In RFs the information stored at a leaf node can be arbitrary [1]: binary, multiclass or structured labels. Moreover, inference in SRFs is actually identical to inference in standard RFs, the only difference being is what information is stored at the leaf nodes and how it is used. For multiclass classification, the standard information gain criterion may also not be well defined over structured labels, $y \in \mathcal{Y}$, that encode the local image annotations of image patches $x \in \chi$. As a result of this, in [1] the authors propose a straightforward two-step mapping approach (defined below): firstly $\mathcal{Y} \rightarrow \mathcal{Z}$ and then $\mathcal{Z} \rightarrow \mathcal{C}$.

Intermediate Mapping and Information Gain Criterion. Given that our required information gain criterion depends on the similarity over $\mathcal{Y}$, we assume that for many structured output spaces, including for structured learning of myocardium edge detection, we can define a mapping Π of $\mathcal{Y}$ to an *intermediate* space $\mathcal{Z}$, $\Pi : \mathcal{Y} \rightarrow \mathcal{Z}$, in which the Euclidean distance in $\mathcal{Z}$ can be measured.

Considering that an approximate measure of information gain is sufficient to train an effective random forest classifier, our goal is to map a set of structured labels $y \in \mathcal{Y}$ into a discrete set of labels $c \in \mathcal{C}$, where $\mathcal{C} = \{1, ..., k\}$, in a way that labels with similar $\mathcal{Y}$ are assigned to the same discrete label c. Given that these discrete labels can be binary ($k = 2$) or multiclass ($k > 2$), we can use standard information gain measures such as Shannon entropy or Gini impurity [5]. The discretization 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 [1].

Because $\mathcal{Z}$ can be of high dimension and computationally expensive to deal with, and since an approximate distance measure is sufficient, we perform dimensionality reduction by sampling m dimensions of $\mathcal{Z}$ which yields a reduced mapping $\Pi_\phi : \mathcal{Y} \rightarrow \mathcal{Z}$ parametrised by ϕ. While training, we randomly generate and apply a unique mapping Π_ϕ to training labels y at each node. By sampling $\mathcal{Z}$, we not only make Π_ϕ faster to compute than Π, but also improve diversity of trees by injecting additional randomness into the learning process [1].

Ensemble Model. The structured ensemble model merges a set of n labels $y_1 ... y_n \in \mathcal{Y}$ into a single prediction both for training, upon association of labels with nodes, and testing i.e. merging of multiple predictions. After sampling a selected m dimensional mapping Π_ϕ and computing $z_i = \Pi_\phi(y_i)$ for each i, we finally select the label y_k whose z_k is the medoid i.e. the medoid z_k that minimizes $\sum_{ij}(z_{kj} - z_{ij})^2$. Because we only need an approximate distance measure to estimate the dissimilarity of y, by reducing $\mathcal{Z}$ dimensionality, the medoid only needs to be computed for small n, which means that an approximate distance metric is sufficient to select an effective element y_k. Notice that the ensemble model is incapable of synthesising new labels without added information about $\mathcal{Y}$. Hence, every prediction $y \in \mathcal{Y}$ must have been observed during training.

2.2 3D Echocardiogram Database

25 End-Diastolic (ED) 3D echocardiograms (224x208x208 voxels) of healthy volunteers (ranging from 19 to 26 years old) were recorded using a Philips iE33 xMATRIX System (X3-1 and X5-1 probes). LV myocardial boundary references (segmentation masks or GT shown in Figure 1[**a**]) were manually drawn for these. All the Structured Edge Detector (SED) models learned in this paper underwent 3-fold cross validation (CV) (divided as 8,8,9 randomly selected datasets). For example, these were trained on say 16 volumes on every 5^{th} short-axis slice of each volume and tested on the remaining 9 volumes, hence there was no correlation between training and testing volumes and slices.

2.3 Myocardium Boundary Detection

Given an input short-axis slice from an ED echocardiography volume, the proposed SED task is to label each pixel with a binary variable indicating whether it belongs to an edge or not. This is done by predicting a structured 24×24 segmentation patch from a larger 48×48 image patch (fixed for all experiments). This patch size was empirically determined to give best edge delineations, and is a result of the need to look at more global information, i.e. contribution of more neighbourhood pixel votes, in order to effectively avoid local irrelevant edge responses.

Regarding the input feature pool, each image patch was augmented with multiple channels of information yielding a feature vector $x \in \mathbb{Z}^{48\times48\times K}$ where K is the number of channels. Two types of features were used: pixel lookups $x(i, j, k)$ and pairwise differences $x(i_1, j_1, k) - x(i_2, j_2, k)$. A similar set of gradient channels used in [6] were implemented in this work. We computed the normalised gradient magnitude at 2 scales (original and half resolution) and each of these channels is then split into 4 channels based on orientation. The channels were blurred and then downsampled by a factor of 2. The resulting K consists of 11 channels (1 grayscale, 2 magnitude and 8 orientation channels). Pairwise difference features were obtained by sampling a blurred and downsampled (7×7) version of the previous candidate pairs, and computing their differences.

Upon training our SRF, and because the Euclidean distance over binary edge maps yields a weak distance measure, we define our mapping Π by sampling a pair of locations $j_1 \neq j_2$, where $1 \leq j \leq 256$ denote the j^{th} pixel of segmentation mask $y(j)$ (Figure 1[**a**]), and check if $y(j_1) = y(j_2)$. This defines $z = \Pi(y)$ as a large binary vector encoding $[y(j_1) = y(j_2)]$ for every distinct pair of indices $j_1 \neq j_2$. Hence, a subset of $m = 256$ dimensions of the high dimensional $\mathcal{Z}$, and $k = 2$, were found to effectively capture the similarity of segmentation masks.

Given that we can store edge maps (any arbitrary information) at the leaf nodes, we finally averaged these to compute a soft edge response. 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. The structured output was computed on the image with a stride of 2 pixels. Since both the

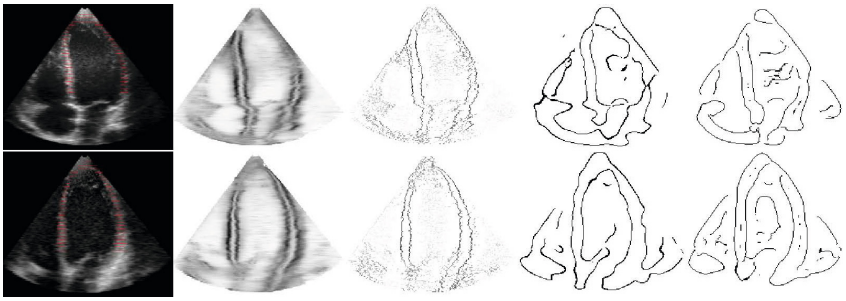
inputs and outputs of each tree overlapped, we trained T=8 trees and evaluated an alternating set of 4 trees at each adjacent location.

Motivated by [7], we finally performed classical multiscale of our SED by averaging the result of three probability edge maps at the original, half (robust but poor localisation), and double resolution (detail-preserving detection but sensitive to endocardial boundary artefacts) version of a given input image. Prior to evaluation, we performed standard non-maximal suppression on the resulting edge maps to obtain thinned edges.

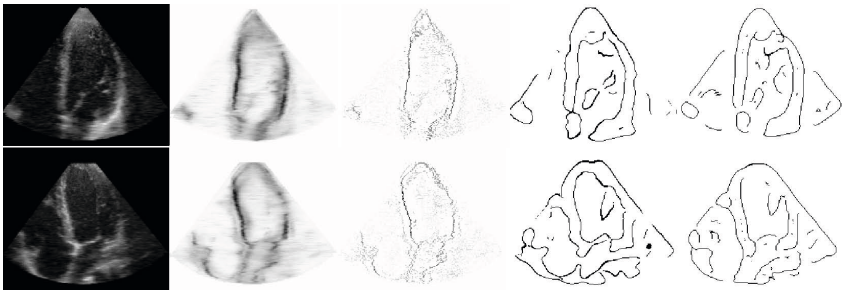
Finally, a SRF ensemble model, *SED1*, was trained on LV myocardial boundary references, and hand-optimized with the parameters previously discussed and a maximum depth of $D = 64$. In addition, a second model, *SED2*, was trained on the same volumes but on LV endocardial boundaries only. After 3-fold CV of results from both SEDs, we evaluated them qualitatively by comparing endocardial edge strength and enhancement against the best (hand-optimized) standard 2D and 3D local phase-based FA measure [2] (parameters: centre frequency: 0.25mm, and 2 octaves) and intensity-based Canny edge detector (magnitude of a Gaussian derivative operator) [8]. In the latter we used the "CannyEdgeDetectionImageFilter" from ITK (parameters: variance: 0.25, lower threshold: 0, upper threshold 1.0, maximum error: 0.0125). For the quantitative evaluation of *SED2*, we computed the Hausdorff distance between the detected endocardium boundaries (non-maximally suppressed) and their correspondent GT. The same was performed for the other two standard edge detectors. To compare these, we used a *masking procedure* in which we mask (GT contour filled and dilated) the (2D and 3D) FA and Canny edge responses to include all responses inside the GT and to explicitly exclude epicardium or other edge responses exterior to the myocardium that could have made the Hausdorff distances bogus.

3 Experimental Results and Discussion

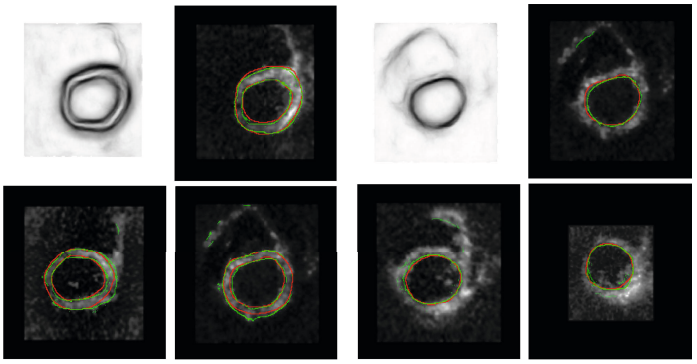
Qualitative Evaluation. Examples of unseen test cases with visible papillary muscles and trabeculation in the blood pool were selected for comparison between the proposed SED ensemble models and the best (thresholded) 3D FA measure. As depicted in Figure 2**[a]**, the proposed *SED1* significantly outperformed the best standard FA and Canny methods in the sense that where responses fade in the local phase-based measure (known to respond well to ultrasound images since they are intensity invariant), *SED1* yielded myocardium edges with high probability. Non-maximal suppression computation of these allowed to better delineate the myocardium. In our method, the stronger edge responses were derived from the topological knowledge gathered by the SRF from each edge pixel neighbourhood ($24 \times 24 = 576$ pixel votes), and therefore contribute to the completeness of the LV and RV blood pools. In addition, it is illustrated how our *SED1* was able to significantly suppress or, in most cases, completely remove any spurious or irrelevant edge responses that result from image artefacts or the presence of papillary muscles and trabeculations in the LV and RV blood pool. In the typical case where accurate myocardium delineation



(a)



(b)



(c)

Fig. 2. [**a**] From *left* to *right*: unseen testing examples with GT (red); LV myocardial edge probability maps from *SED1*; non-maximal suppressed versions of the previous; 3D FA-based edge maps; 3D Canny edge maps. [**b**] From *left* to *right*: unseen testing examples; LV endocardial edge probability maps from *SED2*; non-maximal suppressed versions of the previous; 3D FA-based edge maps; 3D Canny edge maps. [**c**] Unseen testing examples of LV myocardial (endocardium and epicardium) boundary detection from *SED1* and single endocardial boundary detection from *SED2* on short-axis slices. Where shown, GT (red) boundaries are superimposed on the detected ones (green) by the proposed SEDs.

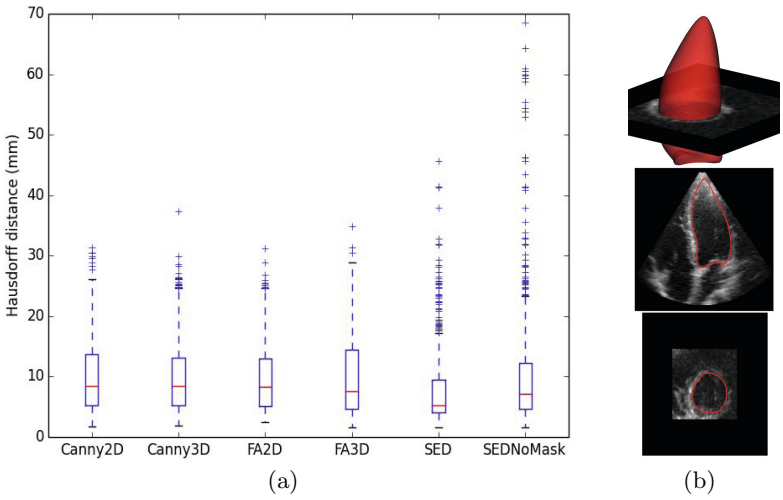


Fig. 3. [**a**] Comparison of the Hausdorff distance distribution between the detected myocardial boundaries (*SED2*) and the GT for different edge detectors. [**b**] Fitting a biquadratic B-spline surface [3] to the detected epicardial boundaries of an LV blood pool test example. This demonstrates that our *SED2* allows for fast and robust heart segmentation and volumetric quantification.

is required, both the FA and Canny methods preserve irrelevant edges which can be seen inside the blood pool of both ventricles in Figure 2**[a,b]**.

Unsurprisingly, when we fitted a deformable anatomical model to the LV blood pool, depicted in Figure 3**[b]**, we found that convergence of surfaces to detected endocardial boundaries was complete for our *SED2*, while in both FA and Canny methods, irrelevant edge responses will prevent deformable models from growing and converging to boundaries. A more extensive analysis of this method for LV volumetric quantification can be found in [9].

In Figure 2**[c]**, our SEDs demonstrated the ambiguity existent in the segmented masks (GT) since in some cases it is arguable that our method performed a better endocardial boundary detection than the GT, which could be due to the blurring and thus smoothing process occuring at the feature extraction level.

Quantitative Evaluation. Because the *masking procedure* preserves all the responses interior to the endocardium, the Hausdorff distance measures whether or not the different methods detect erroneous edges in the blood pool, which is the primary driver for our method. As depicted in Figure 3**[a]**, our *SED2* method ([4.1 5.3 9.5] mm) outperformed the standard FA (2D:[5.1 8.3 12.9] mm; 3D:[4.7 7.6 14.4] mm) and Canny (2D:[5.3 8.5 13.7] mm; 3D:[5.3 8.5 13.1] mm) methods at every percentile (25^{th}, 50^{th} and 75^{th}). More interestingly, even when not masked to exclude epicardium or other edge responses exterior to the myocardium, the proposed *SED2* ([4.7 7.1 12.2] mm) was still able to outperform the standard methods. Note that ultrasound images have been shown to respond

well to local phase-based methods, such as the 3D FA measure, and still our unmasked SED did slightly better when it comes to endocardial boundary detection and enhancement. The higher number of outliers in the *SEDNoMask* was related to some detected RV endocardial boundaries in short-axis slices where RV endocardial structure resembled the LV one.

Finally, at runtime, a 224x208x208 image volume took only 6.7s to generate the myocardial edge probability and orientation volumes on a single core of an Intel Mobile 4930MX (or 4.2s on 8 cores).

4 Conclusion

A novel structured learning approach borrowed from computer vision is shown to perform fast and robust myocardial edge detection. Qualitative and quantitative results demonstrate that our method outperforms standard edge detectors, effectively suppressing the prediction of irrelevant endocardial edge responses, and allowing deformable models and contour-based approaches to more stably converge to the detected myocardial boundaries, enabling computation of more accurate LV clinical indices. Future work will evaluate how accurate the proposed ensemble model is in performing wall thickness measurements.

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

References

1. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: ICCV (2013)
2. Rajpoot, K., Grau, V., Noble, J.: Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 783–786. IEEE (2009)
3. Stebbing, R.V., Noble, J.A.: Delineating anatomical boundaries using the boundary fragment model. Medical Image Analysis 17(8), 1123–1136 (2013)
4. Nowozin, S., Lampert, C.H.: Structured learning and prediction in computer vision, vol. 6. Now Publishers Inc. (2011)
5. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision 7(2-3), 81–227 (2012)
6. Lim, J.J., Zitnick, C.L., Dollár, P.: Sketch tokens: A learned mid-level representation for contour and object detection. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3158–3165. IEEE (2013)
7. Ren, X.: Multi-scale improves boundary detection in natural images. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 533–545. Springer, Heidelberg (2008)
8. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (6), 679–698 (1986)
9. Domingos, J., Stebbing, R., Noble, J.: Endocardial segmentation using structured random forests in 3D echocardiography. In: MICCAI Challenge on Endocardial Three-dimensional Ultrasound Segmentation (2014)