

Left Ventricle Segmentation in Cardiac Ultrasound Using Hough-Forests With Implicit Shape and Appearance Priors

Fausto Milletari¹, Mehmet Yigitsoy¹, Nassir Navab¹, and Seyed-Ahmad Ahmadi²

¹ Computer Aided Medical Procedures, Technische Universität München, Germany
`fausto.milletari@tum.de`,

² Department of Neurology, Klinikum Grosshadern, Ludwig-Maximilians-Universität München, Germany.

Abstract. We propose a learning based approach to perform automatic segmentation of the left ventricle in 3D cardiac ultrasound images. The segmentation contour is estimated through the use of a variant of Hough forests whose object localization capabilities are coupled with a patch-wise, appearance driven, contour estimation strategy. The performance of the proposed method is evaluated on a dataset of 30 images acquired from 15 patients using different equipment and settings.

1 Introduction and related work

The MICCAI 2014 Endocardial Three-Dimensional Ultrasound Segmentation Challenge provides the community with a challenging multi-site, multi-scanner echocardiography dataset of 45 subjects. Their heart is scanned in end-diastolic (ED) and end-systolic (ES) phase instances, with the aim of automatic left ventricular (LV) volume assessment. The challenging nature of 3D realtime echocardiography images, which are affected by a large amount of noise and artefacts such as coarse speckle patterns and dropout regions, makes an automatic segmentation algorithm highly desirable.

In our contribution to the MICCAI echocardiography segmentation challenge, we propose cascaded Hough forests, a special application of Random Forests [2], for simultaneous object detection and segmentation.

The localization of the left ventricle in the images is obtained in a fully automatic manner, using the Hough forest voting strategy. The contour is estimated by making use of a code-book of surface patches associated with the votes. Each contribution to the contour is weighted considering local appearances.

Compared to active surface methods, which are often regulated by statistical shape models (SSM) [1, 4], Hough-forest-based methods neither rely on complex construction of a SSM, nor is it necessary to initialize the pose of the SSM into the image. Among others, the distinctive power of Random Forests was demonstrated in localization and bounding box detection for multiple organs simultaneously, for example in CT [2] or in MRI Dixon sequences [8]. Fast and accurate

voxel-classification for multi-organ segmentation was achieved by Montillo et al. [6], through entangled Decision Forests. Compared to that, Riemenschneider et al. propose a joint single-object localization and segmentation using Hough Regions and Bayesian labeling of a random field [10], implicitly modeling object shape. Rematas and Leibe [9] refined this approach and proposed a unified Hough Forest framework predicting object location and fuzzy segmentation in a streamlined manner. Random Forest lesion detection in 3D transcranial ultrasound, a modality with similar properties as echocardiography, was demonstrated in [7]. Ionasec et al. combine several Probabilistic Boosting Trees and Marginal Space Learning to fit a complex aortic-mitral valve model to 4D Cardiac CT and 4D transesophageal echocardiography data [5].

Our approach builds upon [9]. As our main contribution, in order to deal with the characteristics of the US images, we additionally incorporate an appearance prior enhancing the implicit shape model with an implicit appearance model.

2 Method

We propose a fully automatic method to segment the left heart ventricle in 3D ultrasound scans.

Preprocessing. Our approach comprises a training phase and a test phase. We normalize the spatial resolutions of the 3D-US volumes and then equalize their intensity histograms. In ultrasound images, high contrasts are produced locally by tissue interfaces such as the ventricular wall. Therefore, we define the foreground region as a narrow band around the annotated contour boundary and the rest of the image as background, represented with the class labels $c \in C = \{fg, bg\}$.

Hough forests differ from simple random forests because, beside the classification outcome, they provide means to localize object instances through a voting strategy. In our approach, we couple the voting strategy with a code-base of segmentation patches and associated intensity patches, enabling the forest to natively estimate a segmentation contour [9].

Training stage. During the training phase, the 3D-US volumes are discretized into a regular grid with uniform spacing Δd . At each position of the grid, one data-point p and its associated features f_p are extracted. Each data-point is also associated to a segmentation patch S_p , an intensity patch I_p and a vote $\mathbf{v}_p$, encoding its displacement vector from point p to the estimated left ventricle center LV_C (cf. Fig. 1).

The segmentation and intensity patches S_p and I_p , respectively, are pairs of equal-sized 3D volumetric patches of dimension $(p_x \times p_y \times p_z)$ around the center voxel p . The patches S_p are extracted from the expert labeling volumes of the training dataset, and the patches I_p from the associated intensity volumes.

The N -dimensional feature vectors $f_p \in \mathbb{R}^N$ associated with each data-point are extracted applying a bank of N box filters within a radius R_f around each sampling position, obtaining mean intensities over randomly displaced, asymmetric cuboidal regions similar to Criminisi et al. [3] (see Fig. 1 left). The set

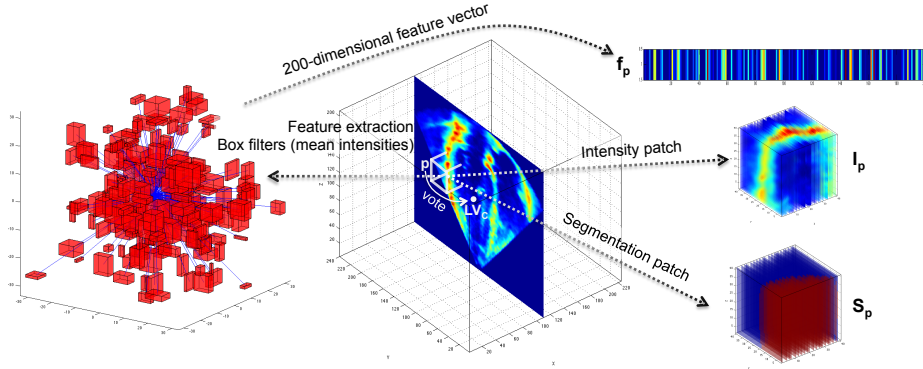


Fig. 1. Schematic representation of the information carried by each training data-point.

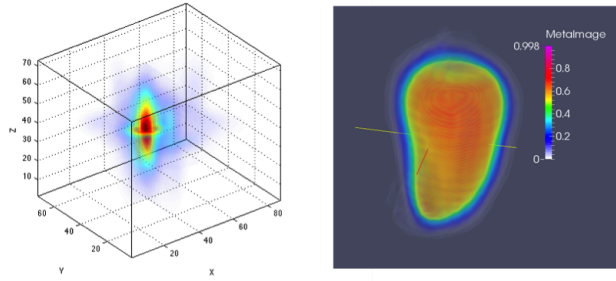


Fig. 2. Left: votes cast by data-points classified as foreground during testing. Right: probabilistic segmentation contour.

of all the training data-points constitutes the training set T , which is used to train the forest. Each tree is trained with a random subset of T (circa 70%), which is recursively split in each node until a termination criterion is met and leaf nodes are established. The axis-aligned splits are performed according to a threshold θ defined on one randomly selected feature. The criterion for the threshold selection is randomly chosen among the following two:

$$IG(T, \theta) = H(T) - \sum_{i \in \{left, right\}} \frac{|T_i|}{|T|} H(T_i) \quad (1)$$

$$VU(T, \theta) = \sum_{i \in \{left, right\}} \sum_{j: c_i \neq bg} dist(\mathbf{v}_{ij}, \bar{\mathbf{v}}_i). \quad (2)$$

The first criterion aims at maximizing the info-gain (IG), which is a measure expressing the change of Shannon entropy $H(T) = -\sum_{c \in \{fg, bg\}} p(c) \log(p(c))$ of the data labels after the split. The second criterion, vote uniformity (VU), aims

at maximizing the similarity of the votes $\mathbf{v}$ associated with the training data, such that the most similar, in terms of Euclidean distance, are grouped together after a split.

The splitting procedure is terminated and a leaf node is instantiated if either the maximal height of the tree or the minimal number of training examples is reached at a certain node, leaving no further possibility of splitting. A leaf of a tree stores not only the class distribution, but also the segmentation and the intensity patches of the training data-points falling in that leaf as well as the displacement vectors between the coordinates of the left ventricle (LV) centroid and each stored data-point belonging to the foreground.

Test stage. During the test phase, feature vectors extracted from unlabeled data-points are sent down each tree, where, at each node, a test stored during training is applied to each feature vector to send it either to the left or to the right child node until a leaf is reached. The points that were classified as foreground with an average confidence across the forest greater than 0.5, are allowed to cast votes to determine the LV centroid in the image. The votes are obtained according to the displacement vectors stored in the leaves of the trees during training. Each vote is weighted by the classification confidence and is accumulated in a volume having the same dimension of the test volume. The location holding the maximum number of votes is chosen to be the LV centroid, LV_C . All the votes in a radius R_{LV_C} from the estimated centroid are re-associated with the data-points that have casted them. Assuming that a number of M votes fell within R_{LV_C} , for each of these M votes, the corresponding intensity and segmentation patches, I_{p_i} and S_{p_i} , respectively, are retrieved from the code-base obtained during the training phase.

The intensity patches I_{p_i} are compared with the image content around each data-point using Truncated Normalized Cross Correlation (TNCC) to obtain weights $w_i \in [0, 1]$. We obtain a probabilistic segmentation $\mathcal{S} = \sum_{i=1}^M w_i \cdot S_{p_i}$, where $w_i = TNCC_i / \sum_{i=1}^M TNCC_i$, as the result of the weighted back-projection of the patches S_{p_i} around the corresponding data-points. The final LV contour is obtained by truncating the probabilistic segmentation at 0.5 (cf. Fig. 2).

Using this strategy not only we obtain a fully automatic method that does not rely on manual initialization, but we can limit the influence of false positives. In contrast with methods that perform segmentation relying on pixel-wise classification, our approach can easily discard misclassified data-points since they are unlikely to cast votes that accumulate around the actual object's center. Additionally, the intensity patches associated to each vote model the expected appearance in correspondence of the segmentation patch stored in the code-book, further helping in rejecting false-positive segmentation votes.

3 Results

We evaluated our algorithm on the MICCAI echocardiography segmentation challenge datasets. Our method was mainly implemented in Matlab®, while some parts, like the segmentation stage, were implemented in C++. We trained

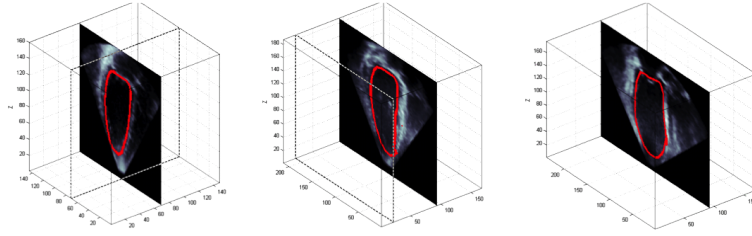


Fig. 3. Typical contour estimates. In these pictures just one slice of the contour is depicted although the contour is estimated in 3D.

a forest containing 20 Hough trees, having a maximal height of 16 and a minimum leaf population of 10 data-points. We performed two experiments, where in the first we perform a leave-one-out-cross validation on the training set, and in the second we do the training on the entire training set, and testing on the test volumes one-by-one.

In the first experiment, in order to maintain computational efficiency, the data-points were sampled from a coarsely spaced grid, where the data-points were located at every $\Delta d = 8mm$ circa. The segmentation and intensity patches S_p and I_p stored in the code base had a size $(p_x \times p_y \times p_z)$ of approximately $18mm \times 16mm \times 13mm$. The features were extracted using a bank of $N = 200$ random box-filters that could be displaced at most $R_f = 15mm$ from the grid-points. All the votes in a radial neighborhood of $R_{LV_C} = 4mm$ from the detected centroid were considered for back-projection. The results of the cross validation, in terms of modified Dice coefficient, are shown graphically in Fig. 4 (blue bars) while more detailed numerical results are provided in Table 1. In this experiment, training required circa 5 minutes while segmenting each volume took circa 6 seconds.

In the second experiment, during testing, since we needed to train the forest only once using the entire training set, we sampled the volumes more densely, reducing the grid point distance to $\Delta d = 6mm$ circa. All other parameters remained constant. The results are depicted graphically in Fig. 4 (red bars), while more accurate numerical results are provided in Table 2. The training stage required 15 minutes and one testing image could be segmented in circa 30 seconds.

4 Discussion

As reported in the results section, our method is able to segment a considerable number of images accurately and in a fully automatic manner. Nevertheless, in a few cases the contour differed significantly from the ground truth. In some cases this was due to the fact that the LV was not fully contained in the 3D volume, while most fail cases were due to the fact that during end systolic phase the mitral valve was not very distinguishable and therefore the contour was placed

Table 1. Evaluation of our approach on training data via leave-one-out cross validation. In the upper half of the table we present the statistics in terms of mean absolute distance, Hausdorff distance, modified dice coefficient and minimum surface error, respectively. In the second half of the table we report the correlation coefficient, bias and limit of agreement that we achieved with respect to the clinical indices; ED volume, ES volume, ejection fraction and stroke volume.

	MaD	HD	Mod. Dice	Min Err.
Mean ED	2.66 mm	9.01 mm	0.129	0.09 mm
Mean ES	2.43 mm	8.09 mm	0.131	0 mm
	ED Volume	ES Volume	Ejection Fraction	
Corr. Coeff	0.986	0.974	0.861	
Bias	38.42	-2.57	19.33	

Table 2. Evaluation of our approach on test data. In the upper half of the table we present the statistics in terms of mean absolute distance, Hausdorff distance, modified dice coefficient and minimum surface error. In the second half of the table we report the correlation coefficient, bias and limit of agreement that we achieved with respect to the clinical indices; ED volume, ES volume, ejection fraction and stroke volume.

	MaD	HD	Mod. Dice	Min Err.
Mean ED	2.09 mm	8.3 mm	0.109	0 mm
Mean ES	2.77mm	8.48 mm	0.16	0 mm
	ED Volume	ES Volume	Ejection Fraction	
Corr. Coeff	0.913	0.968	0.849	
Bias	1.05	-16.52	13.21	

past this point. We observed that ES images are more difficult to segment with our method than ED images, since the LV appears smaller and our implicit statistical shape model might allow the contour to be placed past real boundaries due to misclassified data-points occurring very close to the correct boundary. The clinical indices can be strongly negatively affected by this, despite accurate ED volume segmentation, since the stroke volume and the ejection fraction rely on the correct estimation of *both* ES and ED volumes.

Nevertheless, it is important to note that our method is largely capable of automatically discarding mis-classified data-points due to the presence of the voting strategy. In contrast to methods based on pixel-wise classification, our approach is more robust since it requires a group of data-points to agree on the location of the object and discards all those data-points that do not agree with the prediction. When a data-point that belongs to the background is classified as foreground, it is very unlikely to cast a vote that is coherent with the others, and even if the vote will be accepted, its contribution in terms of segmentation will be averaged out by the other votes. On the other hand, a false negative data-point will have a limited impact on the final contour since its contribution is not essential in presence of strong agreement by other data-points. Furthermore, compared to previous Hough-Forest based segmentation techniques, the novel inclusion of an appearance prior in our method further helped in reducing false-

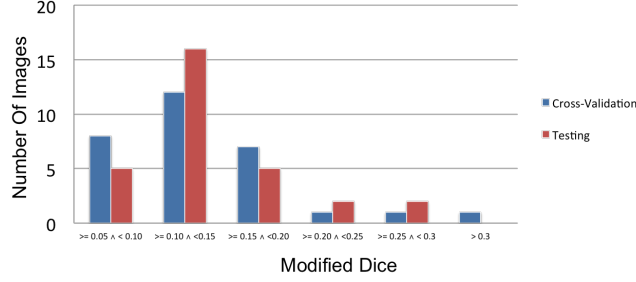


Fig. 4. Histogram of number of images vs. modified Dice coefficient. The results obtained during cross-validation and test are shown respectively in blue and red.

positive foreground classifications, which in turn smoothed the segmentation contour and increased its accuracy.

In the future, we will focus on the integration of more intensity information during the estimation of the contour and more robust features to avoid those outliers that are usually not discarded automatically by our system. It is worth mentioning that our method is able to detect and segment the LV in isolated ES and ED states, while we will investigate the usage of in-between stages of the cardiac cycle, as provided in the CETUS dataset. Furthermore, the inclusion of additional anatomic regions such as mitral valves and other ventricles, as demonstrated in [5], could further help in delineation of the LV contour. Through multi-class detection and segmentation, our framework could easily incorporate such additional anatomical context.

5 Conclusion

In this work, we have proposed a novel, fully automatic 3D segmentation technique built on top of the Hough forest framework. The proposed method differs from the previous approaches since we utilize an appearance code-book coupled with a segmentation code-book to obtain a robust detection of object boundaries. Experiments conducted on a challenging multi-patient, multi-scanner echocardiographic ultrasound dataset demonstrate the robustness of the proposed approach as well as its potential use in similar scenarios.

6 Acknowledgement

We would like to thank Christoph Hennersperger for his support through the development of this method. This work was funded by the DFG grant number BO 1895/4-1.

References

1. Cootes, T., Ionita, M., Lindner, C., Sauer, P.: Robust and accurate shape model fitting using random forest regression voting. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *Computer Vision – ECCV 2012*, Lecture Notes in Computer Science, vol. 7578, pp. 278–291. Springer Berlin Heidelberg (2012)
2. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Tech. rep., Microsoft Research Cambridge, UK (2011)
3. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in ct studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, Lecture Notes in Computer Science, vol. 6533, pp. 106–117. Springer Berlin Heidelberg (2011)
4. Heimann, T., Meinzer, H.: Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis* 13(4), 543–563 (2009)
5. Ionasec, R., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Vega-Higuera, F., Navab, N., Comaniciu, D.: Patient-specific modeling and quantification of the aortic and mitral valves from 4-d cardiac ct and tee. *Medical Imaging, IEEE Transactions on* 29(9), 1636–1651 (Sept 2010)
6. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of ct images. In: *IPMI* (2011)
7. Pauly, O., Ahmadi, S.A., Plate, A., Boetzel, K., Navab, N.: Detection of substantia nigra echogenicities in 3d transcranial ultrasound for early diagnosis of parkinson disease. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, Lecture Notes in Computer Science, vol. 7512, pp. 443–450. Springer Berlin Heidelberg (2012), http://dx.doi.org/10.1007/978-3-642-33454-2_55
8. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Martinez-Moeller, A., Nekolla, S., Navab, N.: Fast multiple organ detection and localization in whole-body mr dixon sequences. In: *Proc. MICCAI* (2011)
9. Rematas, K., Leibe, B.: Efficient object detection and segmentation with a cascaded hough forest ism. In: *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference on. pp. 966–973 (Nov 2011)
10. Riemenschneider, H., Sternig, S., Donoser, M., Roth, P., Bischof, H.: Hough regions for joining instance localization and segmentation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *Computer Vision – ECCV 2012*, Lecture Notes in Computer Science, vol. 7574, pp. 258–271. Springer Berlin Heidelberg (2012)