Segmentation of Multi-Center 3D Left Ventricular Echocardiograms by Active Appearance Models

Marijn van Stralen[1], Alexander Haak[2], K.Y. Esther Leung[3], Gerard van Burken[2] and Johan G. Bosch[2,*]

[1] Image Sciences Institute, University Medical Center Utrecht, Utrecht, Netherlands
[2] Thoraxcenter Biomedical Engineering, Erasmus MC, Rotterdam, Netherlands
[3] Albert Schweitzer Hospital, Dordrecht, Netherlands

Abstract. Segmentation of 3D echocardiograms (3DEs) is still a challenging task due to the low signal-to-noise ratio, the limited field of view, and typical ultrasound artifacts. We propose to segment the left ventricular endocardial surface by using Active Appearance Models (AAMs). Separate end-diastolic (ED) and end-systolic (ES) AAMs were built from presegmented 3DEs of the CETUS training data and 25 previously acquired 3DEs, imaged using various 3DE equipment. The AAMs fully automatically segmented the 15 training sets in a leave-one-out cross validation, comparing two training populations and various initialization strategies. All segmentations took about 15 seconds per patient.
The comparison on the CETUS training data shows that the AAM benefits from additional training data and more accurate initialization. The results on the CETUS training and testing data confirm good ED and ES segmentation accuracy on multi-center, multi-vendor, multi-pathology data, and corresponding EF estimation. Selection from different initialization strategies, based on the minimal residual error, and propagation of detected ED contours to initialize ES detection, contributed to more accurate segmentations in this heterogeneous population.

1 Introduction

3D echocardiography offers a widely accessible and quick qualitative assessment of cardiac function. However, image interpretation of this modality is subject to high intra- and interobserver variability. Therefore, its recent expanding popularity calls for objective, automated analysis techniques.

Fully automatic segmentation of the left ventricle in 3DE is a challenging task, due to the poor signal-to-noise, the limited field-of-view and typical image artifacts in transthoracic echocardiography, such as near-field noise and acoustic shadowing. Various approaches have been presented previously [1–4], but none of these approaches have been shown to be suitable for, or validated on multi-center, multi-vendor, multi-pathology data sets.

* Corresponding author: J.G. Bosch j.bosch@erasmusmc.nl

Active appearance models (AAM) were introduced by Cootes et al. [5] as an extension of the more widely used *active shape models* (ASM). ASMs model a collection of complex shapes by representing each example as a set of corresponding points, calculating average positions and finding principal modes of variation in shape over the collection by principal component analysis (PCA). This results in a *point distribution model* (PDM). AAMs add statistical modeling of the local image intensities, corresponding to the shape locations, to this formulation. Their ability to model typical shape and intensity variations across the target population, including typical artifacts, make them very suitable for the challenging segmentation of the left ventricle (LV) in echocardiography, as was previously shown in [7]. In this work we extend our previous work on segmentation of the LV in 3DE [10, 8] to explore the applicability of AAMs for multi-center, multi-vendor and multi-pathology LV segmentation in 3DE for assessment of the global LV function.

2 Methods

2.1 Data

AAMs use training data to model typical variations in shape and image intensities. The better these training data represent the target population, the better the AAM will be able to segment it. To extend the limited amount of training data available in the CETUS challenge (15 patients), we included 25 additional 3DE sequences of patients with varying pathologies. These were previously acquired at our center using a Philips Sonos 7500 with an X4 matrix transducer and semi-automatically segmented [6], albeit according to slightly different tracing conventions that were optimized for comparison with MRI.

To investigate the influence of the training population of our AAM and the applicability of the model built with varying compositions of data from various vendors and patients groups, we performed our experiments with the following training sets:

I. *CETUS15*: CETUS training data only,
II. *Extended40*: CETUS training data extended with 25 data sets previously acquired at our institute.

Both local and CETUS data were preprocessed to extract masks representing the valid ultrasound data. The CETUS training shapes were included in our PDM by alternatingly matching and projecting each shape onto the LV shape PCA model using iterative closest point (ICP) [9] matching, until convergence.

2.2 Modeling

Our AAM implementation for 3D echocardiography was developed earlier and extensively described in [8]. In this implementation, the left ventricular endocardial surface in 3D was represented as a fixed-topology grid (Fig. 1), consisting of 901 points (30 short-axis rings of 30 equally-spaced vertices, plus an apex point).

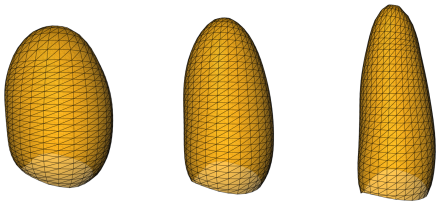

Fig. 1. Shape representation by 901 points for point distribution model. Variations along the first principal component of shape.

Pose variation (translation, rotation and isotropic scaling) was removed from the data before statistical analysis. The AAM texture was sampled along the normal vector of each vertex, from the inside out, extending from 0 to 150% of the local radius of the LV to capture contextual information, including the myocardium. The textures were normalized with an echo-specific intensity normalization [7]. Missing data (outside of the ultrasound image sector) was masked out during model generation, training and matching.

2.3 Training and matching

Active appearance models were generated from the selected training data. AAM matching aims at minimizing the dissimilarity between the model appearance $g_{model}(x, p)$, at position x and parameterized by p, and the target image $g_{im}(x)$. The dissimilarity is represented by the difference or residual image $r(x, p) = g_{im}(x) - g_{model}(x, p)$. We trained model matching by computing the first order approximation of the Jacobian, the partial derivatives of the residual images with respect to the model parameters [5],

$$J(x, p) = \frac{\partial r(x, p)}{\partial p} \tag{1}$$

The pseudo-inverse of the Jacobian,

$$U(x, p) = \left(J(x, p)^T J(x, p)\right)^{-1} J(x, p)^T \tag{2}$$

was used to match the AAM to new images by minimizing $r(x, p)$ through updating the AAM parameter vector p using $\delta p = -U(x, p)r(x, p)$.

2.4 Experiments

The matching experiments were carried out on the CETUS training and testing data sets with separate, independent models for ED and ES segmentation. As the iterative optimization of the AAM matching is sensitive to its initialization we experimented with different initialization stragtegies. We experimented with four strategies for ED and ES independently, varying the position (center of the image volume vs. mean of the training set) and scaling (100 vs. 80%) of the model. Additionally, one scenario coupled ED performance to ES detection. Summarizing, these 5 scenarios had the the mean appearance initialized at:

a. *Center100*: the center of the 3DE image,
b. *Center80*: the center of the 3DE image and 80% scaled in size,
c. *Mean100*: the mean position of the training shapes,
d. *Mean80*: the mean position of the training shapes, and 80% scaled in size.
e. *Selected & Propagated* (ES only): the position and orientation resulting from the best fully automatic ED detection (based on minimum $r(x, p)$), scaled at 80% of the ED size.

To allow further improvement of the matching and to avoid local minima, the matching was restarted once at the resulting position with mean the mean apearance. For the *Selected & Propagated* case this was not necessary, as the initialization was already given accurately by the ED detection.

We compared the *CETUS15* and the *Extended40* training sets and the various matching strategies by segmenting the CETUS training set only (15 patients). The performance and the possible introduction of a bias of the *Extended40* training set was assessed with respect to the *CETUS15* set. To avoid predisposition towards the target data set, a leave-one-out strategy was applied, removing each target from the training set, generating and training the model and matching it on the target patient.

For the final matching of the 15-patient CETUS *testing* data set, the model was generated and trained using the favourable training data set, employing the optimal matching strategy, following from the results on the previous experiments.

The matching was in all cases fully automated, no patient-specific model initialization or correction was applied. It was performed on an Intel Xeon X5570 2.93 GHz PC under Windows 7 (64-bit). The C++ code was single-threaded and used ITK [11], VTK [12], and QT (Digia plc, Helsinki, Finland).

The MIDAS evaluation formed the basis of our analysis. For comparison of the experiments with different training and matching strategies on the CETUS training data, we limited ourselves to a subset of the measures provided by MIDAS. Comparison was done based on statistics (mean and SD) for the following shape metrics: mean absolute difference (MAD), Hausdorff distance (HD) and modified dice (MD). The clinically more relevant correlations and biases of the ejection fraction (EF) and stroke volume (SV) were also investigated.

For the most favourable segmentation strategy, we evaluated the complete spectrum of geometrical and clinical measures, as provided by MIDAS.

3 Results

3.1 Comparison of methods

In a head-to-head comparison of the *CETUS15* and the *Extended40* training sets, the latter showed the added value of the extra training data on all relevant metrics for the *Center100* matching, as is shown in Table 1. As this case reflected the general performance of the two training sets, we focus our analysis on the *Extended40* set only in the remainder of this paper, although the detection was slightly biased by the different tracing conventions in the extended training set.

Table 1. Matching results for the CETUS training (15 patients), models generated using different training sets. All results are reported as mean ± standard deviation or correlation(bias). *Center100* matchings without reinitialization.

	MAD	**HD**	**MD**	**EF**	**SV**
CETUS15	3.5 ± 2.0	10.6 ± 5.5	0.15 ± 0.09	0.44(9.26)	0.69(−4.0)
Extended40	3.3 ± 1.8	10.3 ± 5.4	0.13 ± 0.09	0.54(9.26)	0.86(7.2)

Table 2. Matching results for the CETUS training set (15 patients), trained with the *Extended40* matched with different initialization scenarios. MAD, HD and MD results are reported as mean ± standard deviation, EF and SV as correlation (bias).

	MAD	**HD**	**MD**	**EF**	**SV**
Center100	3.4 ± 1.7	10.9 ± 5.4	0.14 ± 0.08	0.69(6.7)	0.88(3.0)
Center80	3.4 ± 1.8	11.2 ± 6.3	0.14 ± 0.10	0.67(2.8)	0.82(−3.3)
Mean100	3.1 ± 1.7	9.9 ± 4.9	0.13 ± 0.07	0.77(4.2)	0.88(−1.4)
Mean80	3.5 ± 1.9	11.4 ± 7.1	0.15 ± 0.10	0.63(10.9)	0.68(9.9)
Selected & Propagated	3.0 ± 1.3	9.4 ± 3.9	0.12 ± 0.07	0.84(4.0)	0.94(−0.3)

The results for the comparison of the different strategies on the CETUS training data are presented in Table 2. It shows that the performance was improved by more sophisticated initialization. Detection in ES improved by initialization with smaller shapes, boosting EF correlations. The most favorable method used *Selected & propagated* ED contours as initialization for ES detection. This method rendered segmentations with a mean MAD of approximately 3mm and good EF and SV correlations due to its loose coupling of ED and ES matchings.

Bias correction In general, matching using our *Extended40* model caused an overestimation of almost all ED and ES volumes (mean 12.3 ml in ED and 12.0 ml in ES). These overestimations were especially profound in the apical region. They were probably caused by the different tracing conventions that were employed on our own training sets. To compensate for these different tracing conventions, we downscaled our detected contours, relative to the mitral valve plane with 95%. corresponding to the relative bias that was found on the CETUS training data.

3.2 Challenge results on CETUS training and testing data

For the combined results on CETUS training and testing data, we demonstrated our optimal detection strategy, being training using the *Extended40* set and matching using the *Selected & Propagated* scenario, including bias correction Fig. 2 shows examples of detected contours on three of the training set images

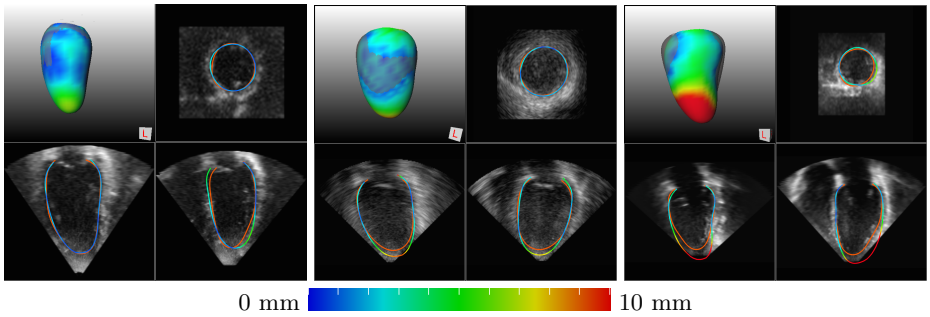

Fig. 2. Example of well (#09 ED), averagely (#15 ED) and poorly (#01 ED) detected contours (color coded) of the training set with the manual ground truth (orange) and the initial contour (white). Without applying a bias correction (as shown), overestimation at the apex is often seen.

Table 3. Results on the CETUS training and testing sets, with the *Extended40* training and *Selected & Propagated* matching.

	Training				**Testing**			
	MAD	**HD**	**mDice**	**Err**	**MAD**	**HD**	**mDice**	**Err**
ED	2.37	8.54	0.107	[0; 8.41]	2.38	8.06	0.119	[0; 8.02]
ES	2.53	8.77	0.132	[0; 8.77]	2.85	8.14	0.171	[0; 8.06]
	EDVol	**ESVol**	**EF**	**SV**	**EDVol**	**ESVol**	**EF**	**SV**
CC	0.981	0.982	0.844	0.941	0.952	0.948	0.790	0.828
Bias	−0.92	−3.83	3.98	2.92	−17.77	−14.18	3.61	−3.62
LOA	[−38; 36]	[−43; 36]	[−12; 20]	[−25; 30]	[−47; 11]	[−40; 11]	[−8.6; 16]	[−18; 11]

with the ground truth and the AAM initialization. The results are summarized in Table 3 following the MIDAS evaluation output format. With the bias correction based on the training set, the method still overestimates volumes in the testing set (negative bias in volume).

The matching computation time varied per data set, depending on the number of iterations, averaging at about 15 seconds.

4 Discussion

The current results (*Extended40, Selected & propagated*) on the training and testing data show that our AAM segmentation can provide good results in 3DE LV segmentation, especially for ED. In any case, the AAM provides a promising tool for clinical quantification of LV function in echocardiography, as the method is fast and fully automated.

The results on the training sets as well as on the test sets are comparable to what we reported earlier for our methods [8, 10]. This is a promising result, given

the heterogeneity of the CETUS data and the limited applicability of our current *Extended40* model. We took the majority of the training data from our own center, acquired with an older Philips system and a probe with different image characteristics. Furthermore, contours were drawn according to slightly different standards. The *Extended40* training set generally still gave better results than the native *CETUS15* dataset. Apparently, 15 data sets is hardly enough to create a stable model with sufficient generalization capability.

Even the *Extended40* training set might be small in order to model the variation in shapes (901 points) and especially appearances (including 45k voxels) of the left ventricles in 3DE. Therefore, the generalization of the model was suboptimal and matching easily converged into a local optimum. To this end, we combined multiple initialization strategies to accommodate for the wide variety in patients. A single reinitialization of the matching at the resulting position and orientation, with a clean mean appearance, contributed to a better performance. Furthermore, the selection based on the residual matching error improved the robustness of the segmentation with respect to the individual initialization strategies.

The considerable biases in detected volumes (-12.3 ml in ED and -12.0 ml in ES) may be linked to the 25 extra training sets. Therefore we applied a bias correction to our segmentations, based on the results on the CETUS training set. Compared to the CETUS tracings, our segmentations were generally too wide, which was most pronounced near the apex. As the level of the mitral valve plane was correctly matched, we implemented our bias correction as a scaling of the surfaces in the direction of the long axis with respect to the mitral valve plane. This correction was performed solely to comply with the CETUS tracing conventions and did, by design, not alter the clinically relevant ejection fractions. Including other data sets that are traced according to the CETUS standards would be preferred and would most probably decrease the detection bias.

Furthermore, improvement can be achieved by optimizing and finetuning the model generation and regression training settings of the AAM. For now, we used standard values that were previously obtained [8].

5 Conclusions

Our AAM for endocardial segmentation of 3D LV echocardiograms has been shown to provide fully automatic segmentation with good results for ED and reasonable results for ES over a database of different ultrasound machines, different pathologies and different centers. Selection from different initialization strategies, based on the minimal residual error, and propagation of detected ED contours to initialize ES detection, contributed to more accurate segmentations in this heterogeneous population.

6 Acknowledgments

We would like to thank the clinicians and sonographers of the Dept. of Echocardiography of the Thoraxcenter for providing the additional patient data used for the AAM model training. This research was supported by the Dutch Technology Foundation STW, which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs.

References

1. Leung KYE, Bosch JG. Automated border detection in three-dimensional echocardiography: principles and promises. Eur J Echocardiography 2010;11(2):97-108.
2. Angelini E, Laine A, Takuma S, Holmes J, Homma S. LV volume quantification via spatio-temporal analysis of real-time 3D echocardiography. IEEE Transactions on Medical Imaging 2001;20;457-469
3. Orderud F, Kiss G, Torp H. Automatic coupled segmentation of endo- and epicardial borders in 3D echocardiography. Proc IEEE Ultrasonics Symposium 2008
4. Barbosa D, Heyde B, Dietenbeck T, Houle H, Friboulet D, Bernard O, D'hooge J. Quantification of left ventricular volume and global function using a fast automated segmentation tool: validation in a clinical setting. Int J Cardiovasc Imaging 2013;29(2);309-316
5. Cootes TF, Edwards GJ, Taylor CJ Active Appearance Models. IEEE Trans Pattern Anal Mach Intell 2001;23(6):681-685
6. Stralen M van, Bosch JG, Voormolen MM, Burken G van, Krenning BJ, Lancee CT, Jong N de, Reiber JHC. Left ventricular volume estimation in cardiac 3D ultrasound: a semi-automatic border detection approach. Academic Radiology 2005;12:1241-1249.
7. Bosch JG, Mitchell SC, Lelieveldt BPF, Nijland F, Kamp O, Sonka M, Reiber JHC. Automatic segmentation of echocardiographic sequences by Active Appearance Motion Models. IEEE Trans Med Imaging 2002;21(11):1374-1383.
8. Leung KYE, Stralen M van, Voormolen MM, Jong N de, Steen AFW van der, Reiber JHC, Bosch JG. Improving 3D Active Appearance Model segmentation of the left ventricle with Jacobian tuning. Proc SPIE Medical Imaging 2008;6914(69143B):1-11.
9. Besl PJ, McKay ND. A Method for Registration of 3-D Shapes. IEEE Trans Pattern Anal Mach Intell 1992;14(2):239-256
10. Stralen M van, Leung KYE, Voormolen MM, Jong N de, Steen AFW van der, Reiber JHC, Bosch JG Automatic Segmentation of the Left Ventricle in 3D Echocardiography Using Active Appearance Models. IEEE Ultrasonics Symposium 2007;1480-1483
11. Yoo TS, Ackerman MJ, Lorensen WE, Schroeder W, Chalana V, Aylward S, Metaxas D, Whitaker R. Engineering and algorithm design for an image processing API: a technical report on ITK - the Insight Toolkit. Proc Medicine Meets Virtual Reality 2002;586-592
12. Schroeder W et al. The Visualization Toolkit. 3rd edition. Kitware Inc. 2003