Left Atrial Segmentation Challenge: A Unified Benchmarking Framework

Catalina Tobon-Gomez[1], Jochen Peters[2], Juergen Weese[2], Karen Pinto[1], Rashed Karim[1], Tobias Schaeffter[1], Reza Razavi[1,3], and Kawal S. Rhode[1]

[1] Division of Imaging Sciences & Biomedical Engineering, King's College London, London, UK*

[2] Philips Technologie GmbH, Innovative Technologies, Forschungslaboratorien, Hamburg, DE

[3] Department of Cardiology, Guys and St. Thomas NHS Foundation Trust, London, UK*

Abstract. The knowledge of left atrial (LA) anatomy is important for atrial fibrillation ablation guidance. More recently, LA anatomical models have been used for cardiac biophysical modelling. Segmentation of the LA from Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images is a complex problem. We aimed at evaluating current algorithms that address this problem by creating a unified benchmarking framework through the mechanism of a challenge, the *Left Atrial Segmentation Challenge 2013* (LASC'13). Thirty MRI and thirty CT datasets were provided to participants for segmentation. Ten data sets for each modality were provided with expert manual segmentations for algorithm training. The other 20 data sets per modality were used for evaluation. The datasets were provided by King's College London and Philips Technologie GmbH. Each participant segmented the LA including a short part of the LA appendage trunk plus the proximal parts of the pulmonary veins. Details on the evaluation framework and the results obtained in this challenge are presented in this manuscript. The results showed that methodologies combining statistical models with region growing approaches were the most appropriate to handle the proposed task.

1 Introduction

Atrial fibrillation (AF) is the most common cardiac electrical disorder which doubles the mortality rate of patients. It has been shown that ectopic beats from within the pulmonary veins (PVs) commonly initiate AF [1]. Therefore, catheter ablation strategies attempt to electrically isolate the PVs from the left atrial (LA) body. Knowing the LA anatomy is crucial for the success of the

* This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

O. Camara et al. (Eds.): STACOM 2013, LNCS 8330, pp. 1–13, 2014.

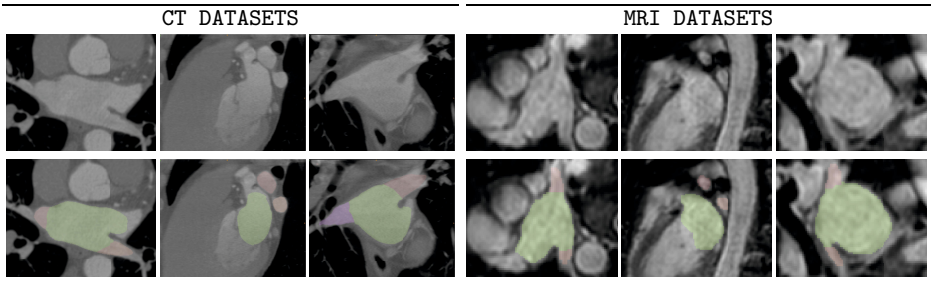

Fig. 1. Example datasets as provided for the challenge. Colour overlay shows the manual ground-truth, where green areas represent the LA body, and magenta areas represent the PVs (for details see Sec. 2).

intervention, since it enables accurate planning of ablation lines and guidance during the procedure [2]. More recently, LA anatomical models have been employed for cardiac biophysical modelling [3]. These models aim at understanding the mechanisms of AF and, eventually, at predicting optimal therapy.

Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are commonly used for imaging the heart. There are several topological variants of the LA and this means that segmentation of the endocardial boundary and PVs is a non-trivial task [4]. Therefore, we aimed at evaluating current algorithms that address this problem by creating a unified benchmarking framework through the mechanism of a challenge, the *Left Atrial Segmentation Challenge 2013* (LASC'13). This challenge was held at the MICCAI 2013 Workshop on Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges (STACOM'13). Each participant segmented the LA including a short part of the LA appendage trunk plus the proximal parts of the PVs, from 3D whole heart MRI and/or CT modalities. Details on the evaluation framework and the results obtained in this challenge are presented in this manuscript.

2 The Challenge

Thirty MRI and thirty CT datasets were provided to participants for segmentation. Ten data sets for each modality were provided with expert manual segmentations for algorithm training (see Fig. 1). The other 20 data sets per modality were used for evaluation. Datasets were limited to the most common topological variant showing four PVs. The datasets were provided by King's College London and Philips Technologie GmbH.

Participants were expected to segment the LA including a short part of the LA appendage (i. e. trunk) plus the proximal parts of the PVs (i. e. up to the first branching point or after 10 mm from the vein ostium). The LA body should have extended at least up to the mitral valve (MV) (i. e. reach into the funnel of the MV). Results were submitted as a single-valued binary mask covering all these structures in NifTI format[1].

[1] `http://nifti.nimh.nih.gov/`

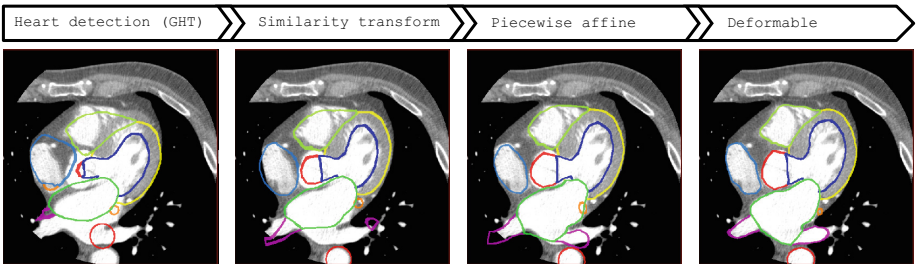

Fig. 2. Automatic segmentation pipeline on a CT image for ground truth generation. Different colours represent different parts of the deformable model. Green and magenta regions correspond to the LA and the PVs, respectively (for details see Sec. 3).

MRI Datasets. MRI acquisition was performed on a 1.5 T Achieva scanner (Philips Healthcare, The Netherlands). A 3D whole heart (3DWH) image was acquired using a 3D balanced steady-state free precession acquisition [5]. The sequence acquires a non-angulated volume covering the whole-heart with voxel resolution of $1.25 \times 1.25 \times 2.7\,\text{mm}^3$. Images were acquired during free breathing with respiratory gating and at end-diastole with ECG gating. Main acquisition parameters include: TR/TE=4.4/2.4 ms, flip angle=90°, cardiac phases=1. Typical acquisition time for a complete volume is 10 min.

CT Datasets. Retrospectively ECG-gated cardiac multi-slice CT images were acquired with Philips 16-, 40-, 64- and 256-slice scanners (Brilliance CT and Brilliance iCT, Philips Healthcare, Cleveland OH, USA) at different cardiac phases. All images are reconstructed using a 512×512 matrix with an in-plane voxel resolution ranging from 0.30×0.30 to $0.78 \times 0.78\,\text{mm}^2$ and with slice thickness ranging from 0.33 to 1.00 mm. All scans were acquired after injection of ca. 40–100 ml contrast media (density 320–370 mg iodine/ml), depending on the exact purpose of the study (assessment of coronary arteries or cardiac valves). Contrast levels vary widely over the images provided for this challenge. Acquisition times for a complete CT volume ranged from 3-5 sec on modern iCT scanners to 20 sec for the older 16-slice scanners.

3 Ground-Truth Generation

In order to obtain a set of ground-truth (GT) segmentations consistent across modalities, we started by performing an automatic model-based segmentation with a method which is optimised for both CT and MRI modalities. After the automatic segmentation, manual corrections were performed. Details are provided next.

Automatic Segmentation. The automatic segmentation used in this study was described in [7–9]. The segmentation uses shape-constrained deformable models. These are based on a mesh representation of surfaces of cardiac chambers and the attached great vessels. These meshes have a complex topology with T-junctions

where different structures meet. The automatic adaptation starts by a localisation step using the Generalized Hough Transform [10] to place the mesh model close to the targeted organ. Thereafter, several adaptation steps with increasing degrees of freedom refine the model pose and shape. Each step uses trained boundary detectors that enable a robust and accurate detection of the wanted organ boundaries in the image. These detectors are trained individually per mesh triangle and can capture the varying appearance of organ boundaries in the images. Using the detected boundaries, a first step adjusts the global pose of the complete model by performing a rigid adaptation with scaling that minimises the squared distances of the model surface to the detected boundaries. Subsequent steps add more degrees of freedom by subdividing the model into mesh regions (such as cardiac chambers or short parts of the tubular vessels) and adapts these parts via individual affine transformations. Finally, a deformable adaptation step leads to a locally accurate segmentation where each mesh vertex is free to move under the image forces that pull the mesh triangles to the detected boundaries while internal forces regularise the adaptation and penalise strong deformations of the model shape. After adaptation of the model is complete, the regions enclosed by the surfaces are converted into a label image with region-specific labels. Labels not covering the LA and the PVs were discarded (see Fig. 2).

Manual Correction Criteria. Each automatic segmentation was manually corrected by an experienced observer to obtain the final GT segmentation. Manual corrections were performed using ITK-SNAP [11] for MRI datasets and Philips in-house editing tools for CT datasets. PVs were followed distally to the LA body ensuring at least 10 mm coverage. They were truncated at the branching point when there was no clear main PV to follow. This *early truncation* mainly happened in MRI, either due to image artefacts or low signal-to-noise ratio. Each obtained GT segmentation consists of five labels: one label for LA body and LA appendage trunk, and, four labels for each of the PVs. These labels were used for standardisation purposes (see Sec. 4).

4 Standardisation Framework

Even for a human observer, defining certain regions of the LA is difficult. One of these regions is the boundary between atrium and ventricle. Since the MV leaflets can be at different levels of *opening/closure*, the definition of a MV plane can be arbitrary. Unless an exact segmentation of the mitral annulus is available which, however, may be non-planar. Another one of these regions is the PVs. For these structures both the start points (i. e. ostia) and end points can be arbitrary. Finally, the LA appendage (LAA) varies greatly among the population which makes it difficult to segment. Since for most applications the actual shape of the LAA is not relevant, we opted for removing it from the LA body. We only retain the region most proximal to the LA cavity (i. e. trunk).

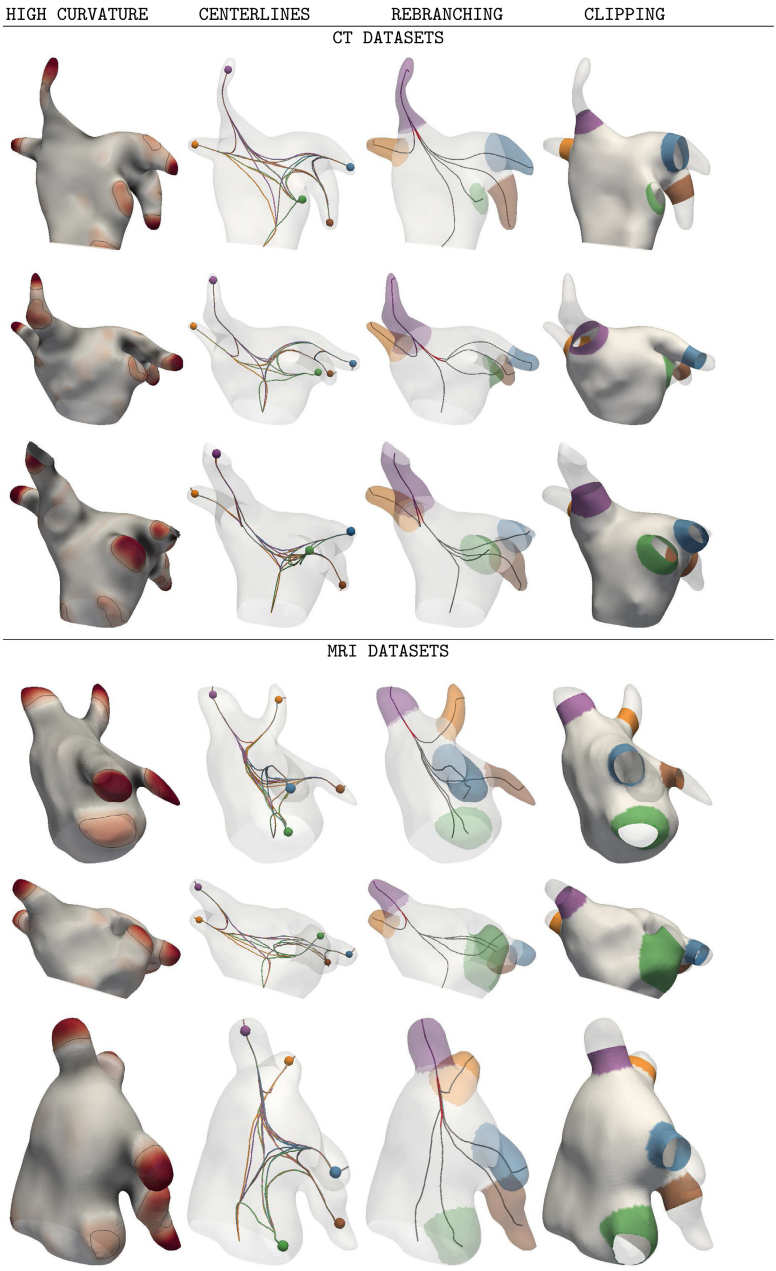

Fig. 3. Standardisation framework. Distal end points of the PVs and LAA were extracted by computing areas of maximum curvature. The centroids of the high-curvature areas were used as seed points for centerline extraction. The surface mesh was branched based on maximum inscribed sphere radius [6]. For PVs and LAA clipping: we computed a plane that is normal to the centerline and located 10 mm away from the ostia (for details see Sec. 4).

To ensure that the calculated metrics are not negatively affected by these regions, we standardised all submitted segmentations. The framework was implemented using the Visualization Toolkit (VTK), the Vascular Modeling Toolkit (VMTK), and MATLAB Toolbox Graph[2].

Mitral Valve. Given the predefined labels of our GT, we could compute certain anatomical landmarks in an automatic manner. For MV plane computation, we extracted a surface mesh representing the LA body and the PVs separately. We computed Principal Components Analysis on the LA body. The clip point was set along the main axis (or the average of the two main axes for more spherical bodies), at a distance of 35% $\times$ the maximum body length below the centroid.

Pulmonary Veins and LA Appendage. Along this clipped surface we computed the Gauss curvature [12, 13]. We then normalised the curvature values to obtain a unified range of $(-2, 2)$. We thresholded the highest curvature values (>0.5). These patches of high curvature were used as candidate positions for PV and LAA end points (black contours in Fig. 3-`HIGH CURVATURE`). For each PV we selected the patch furthest from the ostium. The patches belonging to the body were discarded based on two criteria: *(1)* small surface area ($< 0.5 \times$ largest patch area); or *(2)* vicinity to the MV or the PVs ostia. The remaining body patches belonged to the LAA. From the selected patches we computed the centroids and stored them as *seed points* for centerline extraction.

We calculated the centerlines that connected each seed point to all remaining seed points plus the centroid of the MV edge, as displayed in Fig. 3-`CENTERLINES`. Using the approach of Antiga et al. [6], we computed bifurcation regions in the centerlines corresponding to each seed. From the most distal bifurcation point we defined a new splitting point located 0.75 $\times$ the maximum inscribed sphere radius, similarly to the approach used by Piccinelli et al. [14] to define the neck of cerebral aneurysms (red section of centerlines in Fig. 3-`REBRANCHING`).

Next, we labeled the surface based on the branched centerlines. These automatically computed labels proved to be more consistent among the GT population than the arbitrarily defined manual labels. Therefore, we used them as a final definition of each anatomical region of the LA. Using the labels, we isolated each PV and clipped it with a plane perpendicular to its corresponding centerline and located 10 mm from the PV ostium. The LAA was clipped at 80% $\times$ the maximum length of the labelled LAA surface (see Fig. 3-`CLIPPING`).

Automatic Segmentations Standardisation. For all submitted segmentations (binary masks), we performed a close filling operation to ensure a single connected region. From it we generated a surface mesh using marching cubes followed by volume preserving smoothing. Next, we clipped the mesh with the MV plane generated from the GT mesh (Sec. 4), discarding unconnected regions. Then, we transferred the automatic branch labels of the GT mesh to their closest points in the automatically segmented mesh. For each label, we ensured a

[2] `www.vtk.org www.vmtk.org www.ceremade.dauphine.fr/~peyre/matlab/graph/content.html`

single connected region to avoid transferring PV labels to neighbouring areas. Finally, using the labels we isolated each PV and LAA and clipped them using the planes computed automatically from the GT.

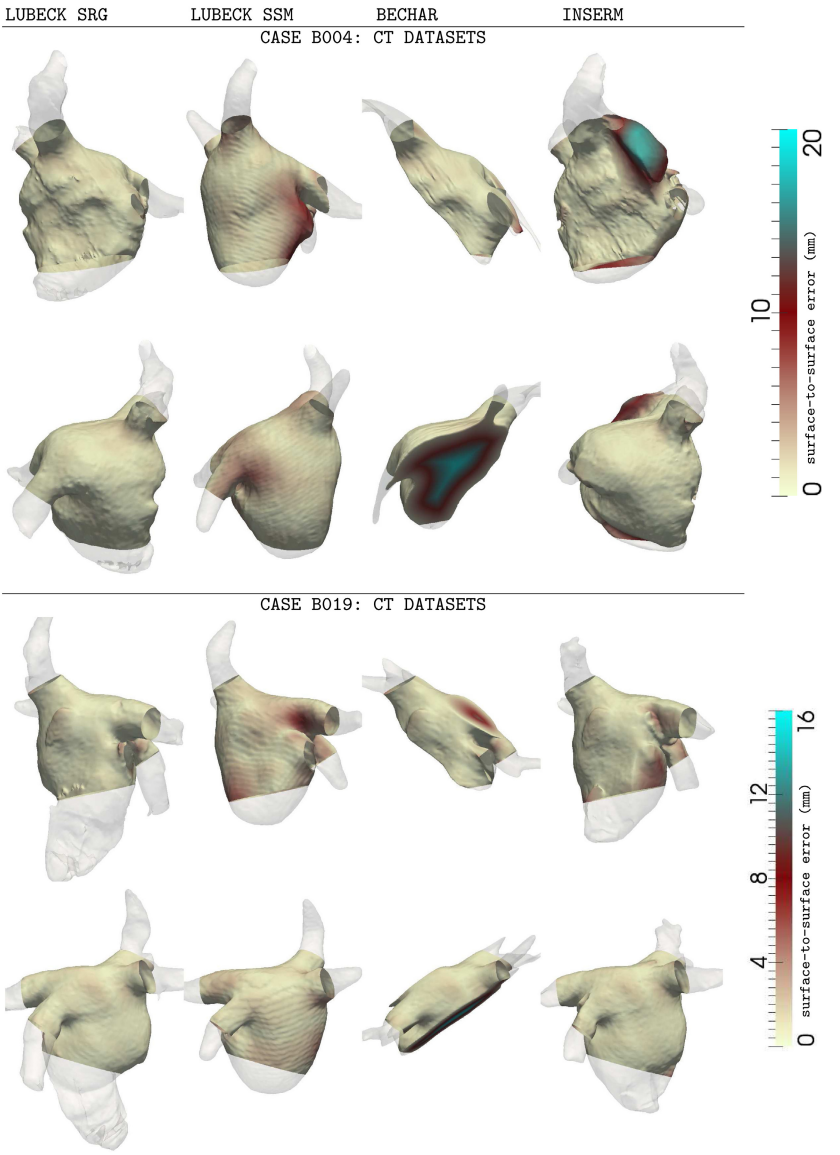

Fig. 4. Results for each participant for the CT datasets. The original meshes are displayed with transparency. The standardised meshes are colour mapped with surface-to-surface errors. Note that for visualisation purposes, only the automatic-to-GT errors are displayed in this figure. The symmetric surface-to-surface errors are summarised in Table 1 (for details see Sec. 5).

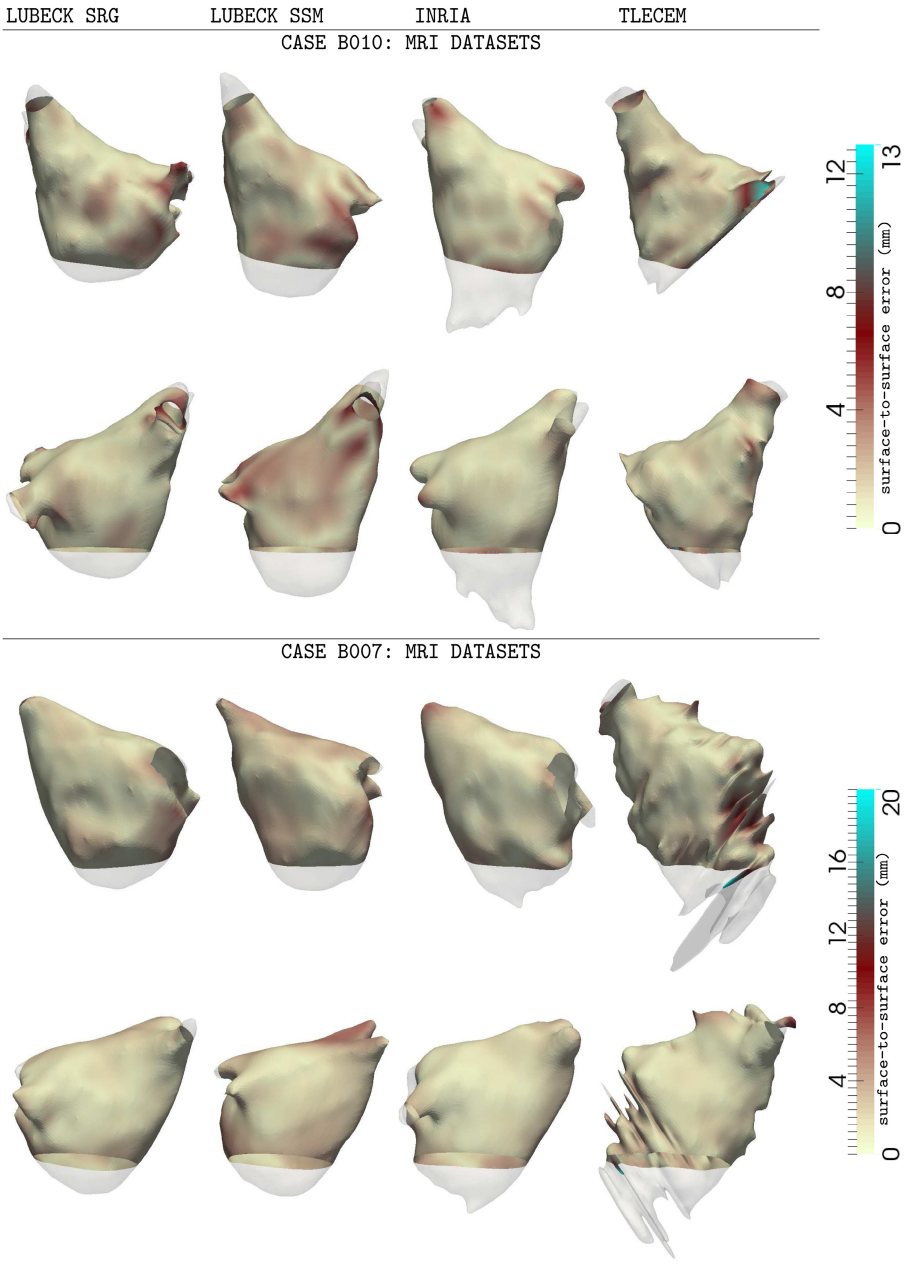

Fig. 5. Results for each participant for the MRI datasets. The original meshes are displayed with transparency. The standardised meshes are colour mapped with surface-to-surface errors. Note that for visualisation purposes, only the automatic-to-GT errors are displayed in this figure. The symmetric surface-to-surface errors are summarised in Table 1 (for details see Sec. 5).

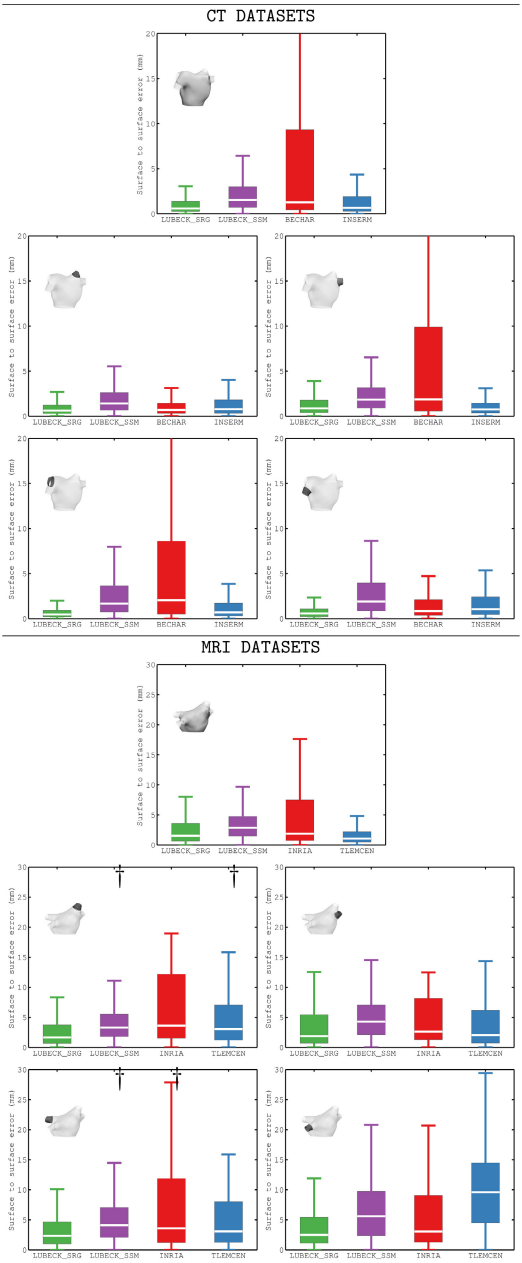

Fig. 6. Box-plots of segmentation errors for each participant for each anatomical region. The corresponding region is represented in the vignette on the upper left corner. Maximum whisker corresponds to approximately 99.3% coverage if the data were normally distributed. Pair of samples that yielded statistically non-significant differences ($p < 0.001$) are marked on the plot (†).

5 Discussion

Participants. We received submissions from 5 groups. University of Lubeck, Germany, processed CT and MRI datasets with two methodological approaches: one based on statistical shape models (`LUBECK SSM`), and, another one combining statistical shape models and region growing (`LUBECK SRG`). University of Bechar, Algeria, processed the CT datasets (19 out of 20) with a combination of region growing and gradient vector flow snakes (`BECHAR`). INSERM Rennes, France, processed the CT datasets with a multi-atlas, multi-voting and region growing approach (`INSERM`). Inria, Sophia-Antipolis, France, processed the MRI datasets using decision forests (`INRIA`). University of Tlemcen, Algeria, processed the MRI datasets with a combination of threshold localisation and circularity shape descriptors (`TLEMCEN`). For details on each methodology refer to the articles in these proceedings. Examples of the submitted segmentations before and after standardisation are shown in Fig. 4 and Fig. 5.

Results. To test segmentation accuracy, symmetric surface-to-surface error (S2S) and Dice metric were computed for all the standardised segmentations. The median and standard deviation of both metrics (i. e. LA body, left superior PV, left inferior PV, right superior PV and right inferior PV) are summarised in Table 1. Fig. 6 shows the box-plots of the S2S errors.

Results showed that statistical shape approaches combined with region growing obtained the best accuracy (LUBECK SRG and INSERM). It must be noted that this type of methodology often leaks into the left ventricle, the aorta and sometimes into the right atrium. Some of these segmentation errors were removed by our standardisation process, hence they were not penalised in by the evaluation metrics. However, to be implemented as a feasible clinical tool, the region growing should be somewhat constrained. For instance, limiting the region growing process to the PVs areas and/or to the surroundings of the initial surface. The statistical shape approach in itself (LUBECK SSM) although highly robust (i. e. valid shape instances of the LA) obtained lower accuracy than its corresponding extension with region growing. However, for certain applications this level accuracy could be sufficient.

BECHAR's approach shows potential since the slices that were processed obtained good accuracy. Unfortunately, due to the large amount of missing slices (specially on the lower part of the LA) the performance metrics were low in most anatomical regions. A 3D extension of the approach able to handle the whole span of the LA would increase the feasibility of the methodology. This was suggested by the authors as part of their future work.

INRIA's approach even though it makes few assumptions, often obtained good results. However, when then segmentation failed the errors where rather large resulting on poor average performance. A possible improvement of this approach would be to split the segmentation in two tasks: one for the body only and one for the PVs. Another improvement would be to impose shape constrains on the raw output of the decision forrest.

Table 1. Summary of error metrics: all structures measured after standardisation

		PARTICIPANTS															
		LUBECK SRG				LUBECK SSM				BECHAR				INSERM			
		s2s		dice		s2s		dice		s2s		dice		s2s		dice	
		m	std	m	std	m	std	m	std	m	std	m	std	m	std	m	std
CT	BODY	0.58	4.25	0.94	0.09	1.53	2.34	0.88	0.08	1.27	8.53	0.55	0.17	0.67	3.18	0.89	0.10
	LSPV	0.62	2.10	0.88	0.21	1.42	2.29	0.78	0.21	0.69	3.16	0.83	0.24	0.78	3.34	0.74	0.27
	LIPV	0.87	3.06	0.86	0.13	1.84	2.35	0.68	0.25	1.87	7.10	0.15	0.36	0.76	2.68	0.86	0.33
	RSPV	0.48	1.04	0.89	0.15	1.67	3.87	0.53	0.28	2.05	6.98	0.13	0.34	0.70	1.46	0.79	0.27
	RIPV	0.57	1.08	0.89	0.14	1.90	3.05	0.53	0.33	0.84	2.64	0.71	0.29	1.05	1.99	0.76	0.33
		LUBECK SRG				LUBECK SSM				INRIA				TLECEM			
		s2s		dice		s2s		dice		s2s		dice		s2s		dice	
		m	std	m	std	m	std	m	std	m	std	m	std	m	std	m	std
MRI	BODY	1.54	3.39	0.91	0.09	2.86	2.99	0.83	0.08	1.87	9.17	0.88	0.30	1.07	3.79	0.90	0.07
	LSPV	1.69	3.33	0.72	0.28	3.31	2.48	0.39	0.24	3.63	15.85	0.35	0.30	3.06	8.17	0.08	0.36
	LIPV	1.89	3.36	0.76	0.38	4.29	4.25	0.23	0.20	2.64	14.37	0.42	0.36	2.06	5.63	0.50	0.41
	RSPV	2.34	6.27	0.35	0.28	4.10	5.65	0.08	0.22	3.59	8.53	0.12	0.33	3.09	8.06	0.36	0.28
	RIPV	2.49	6.35	0.46	0.29	5.60	5.59	0.04	0.20	3.07	7.01	0.34	0.31	9.59	7.63	0.00	0.18

s2s = surface-to-surface error (mm); dice = Dice metric; m = median; std = standard deviation; LSPV = left superior PV; LIPV = left inferior PV; RSPV = right superior PV; RIPV = right inferior PV; BODY = LA body without LA appendage.

TLECEM's approach is based on circular shape descriptors from the sagittal plane. Thus it obtained good accuracy on the middle of the LA body (lowest error). However, the PVs were often missing and the lower part of the LA body (closer to the MV) was often over segmented. Similarly to the approach of BECHAR, a 3D extension could improve its feasibility for a clinical application.

6 Conclusions

This manuscript presents a unified benchmarking framework for current algorithms for segmentation of the left atrium from MRI and CT datasets. Strong effort was dedicated to implement a standardisation framework for the ground-truth and the automatic segmentations.

The results showed that methodologies combining statistical models with region growing approaches were the most appropriate to handle the proposed task. Visual results showed that an approach with good performance according to the error metrics (low surface-to-surface and high dice) does not always provide the best overall 3D structural result. This has pointed us to believe that it is important to explore other complementary metrics that better reflect the similarities in shape between the desired ground-truth surface and the automatic segmented surface.

As a follow-up work, we will submit a journal publication with the benchmarking framework presented in this workshop. In other to include other methodologies, we will make a second call for participants. In this follow-up work, we plan to evaluate the influence of image quality on each segmentation algorithm, include a measure of inter-observer variability, and, extend the performance metrics. Examples of these metrics include: centerline-to-centerline distance of the pulmonary veins, leakage metric to reflect the effect of a failed region growing, and, more advanced statistical measures of the shape differences between the surfaces.

Acknowledgements. The authors would like to thank C. Butakoff, O. Camara and A.J. Geers for their very useful suggestions for the automatisation of the evaluation framework.

References

1. Haïssaguerre, M., Jaïs, P., Shah, D.C., Takahashi, A., Hocini, M., Quiniou, G., Garrigue, S., Le Mouroux, A., Le Métayer, P., Clémenty, J.: Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. N. Engl. J. Med. 339(10), 659–666 (1998)
2. Calkins, H., Kuck, K.H., Cappato, R., Brugada, J., Camm, A.J., Chen, S.A., Crijns, H.J.G., Damiano, J.R.J., Davies, D.W., DiMarco, J., Edgerton, J., Ellenbogen, K., Ezekowitz, M.D., Haines, D.E., Haissaguerre, M., Hindricks, G., Iesaka, Y., Jackman, W., Jalife, J., Jais, P., Kalman, J., Keane, D., Kim, Y.H., Kirchhof, P., Klein, G., Kottkamp, H., Kumagai, K., Lindsay, B.D., Mansour, M., Marchlinski, F.E., McCarthy, P.M., Mont, J.L., Morady, F., Nademanee, K., Nakagawa, H., Natale, A., Nattel, S., Packer, D.L., Pappone, C., Prystowsky, E., Raviele, A., Reddy, V., Ruskin, J.N., Shemin, R.J., Tsao, H.M., Wilber, D.: 2012 hrs/ehra/ecas expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. Europace 14(4), 528–606 (2012)
3. Aslanidi, O.V., Colman, M.A., Stott, J., Dobrzynski, H., Boyett, M.R., Holden, A.V., Zhang, H.: 3d virtual human atria: A computational platform for studying clinical atrial fibrillation. Prog. Biophys. Mol. Biol. 107(1), 156–168 (2011)
4. Kato, R., Lickfett, L., Meininger, G., Dickfeld, T., Wu, R., Juang, G., Angkeow, P., LaCorte, J., Bluemke, D., Berger, R., Halperin, H.R., Calkins, H.: Pulmonary vein anatomy in patients undergoing catheter ablation of atrial fibrillation: lessons learned by use of magnetic resonance imaging. Circulation 107(15), 2004–2010 (2003)
5. Uribe, S., Muthurangu, V., Boubertakh, R., Schaeffter, T., Razavi, R., Hill, D.L.G., Hansen, M.S.: Whole-heart cine mri using real-time respiratory self-gating. Magn. Reson. Med. 57(3), 606–613 (2007)
6. Antiga, L., Steinman, D.A.: Robust and objective decomposition and mapping of bifurcating vessels. IEEE Trans. Med. Imaging 23(6), 704–713 (2004)
7. Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A., Weese, J.: Automatic whole heart segmentation in static magnetic resonance image volumes. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 402–410. Springer, Heidelberg (2007)
8. Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M.J., Vembar, M., Olszewski, M.E., Subramanyan, K., Lavi, G., Weese, J.: Automatic model-based segmentation of the heart in CT images. IEEE Trans. Med. Imaging 27(9), 1189–1201 (2008)
9. Ecabert, O., Peters, J., Walker, M.J., Ivanc, T., Lorenz, C., von Berg, J., Lessick, J., Vembar, M., Weese, J.: Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. Med. Image Anal. 15(6), 863–876 (2011)
10. Ballard, D.H.: Generalizing the Hough transform to detect arbitrary shapes. Pattern Recogn. 13(2), 111–122 (1981)

11. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006)
12. Alliez, P., Cohen-Steiner, D., Devillers, O., Lévy, B., Desbrun, M.: Anisotropic polygonal remeshing. ACM Transactions on Graphics 22, 485–493 (2003); SIGGRAPH 2003 Conference Proceedings
13. Cohen-Steiner, D., Morvan, J.M.: Restricted delaunay triangulations and normal cycle. In: 19th Annual Symposium on Computational Geometry, pp. 237–246 (2003)
14. Piccinelli, M., Veneziani, A., Steinman, D.A., Remuzzi, A., Antiga, L.: A framework for geometric analysis of vascular structures: application to cerebral aneurysms. IEEE Trans. Med. Imaging 28(8), 1141–1155 (2009)