Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS)

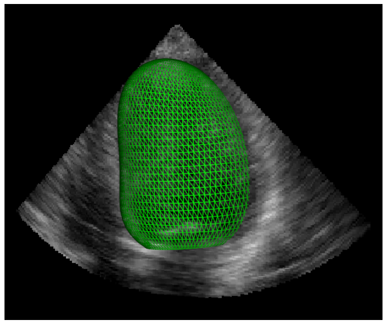
Olivier Bernard[1], Brecht Heyde[2], Martino Alessandrini[2], Daniel Barbosa[3], Sorina Camarasu-Pop[1], Frederic Cervenansky[1], Sebastien Valette[1], Oana Mirea[2], Elena Galli[4], Marcel Geleijnse[5], Alexandros Papachristidis[6], Johan G. Bosch[5], and Jan D'hooge[2]

[1] Université de Lyon, CREATIS; CNRS UMR5220; INSERM U630; Université Lyon 1; INSA-LYON, France
[2] Lab on Cardiovascular Imaging & Dynamics, Department of Cardiovascular Sciences, KU Leuven, Belgium
[3] Life and Health Sciences Research Institute (ICVS), University of Minho, Portugal
[4] University of Rennes, France
[5] Thoraxcenter, Erasmus MC, Rotterdam, Netherlands
[6] King's College London, United Kingdom

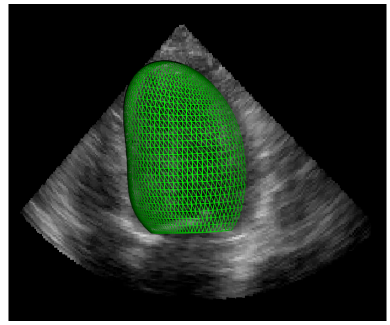
Abstract. Real-time 3D echocardiography has already been shown to be an accurate tool for left ventricular (LV) volume assessment. However, LV border identification remains a challenging task, mainly because of the low contrast of the images combined with drop-out artifacts. Many (semi-)automatic algorithms have been proposed to segment the LV border, but a systematic and fair comparison between such methods has so far been impossible due to a lack of publicly available common database. The aim of this MICCAI challenge was to gather researchers around the field of LV segmentation in 3D cardiac ultrasound by providing a common database to compare algorithms directly. The proposed platform will allow a consistent evaluation and ranking of the current state-of-the-art segmentation solutions and will contribute to a faster clinical translation of groundbreaking technical advances. The purpose of this paper is to describe the technical aspects of the generation of the database, give an overview of the ranking strategy and the outline of the challenge itself.

1 Introduction

Echocardiography is a widely used clinical imaging technique to examine myocardial function in patients with known or suspected heart pathologies. Over the last decade, the assessment of cardiac morphology and function by ultrasound imaging has made a significant step forward by the introduction of real-time 3D echocardiography (RT3DE), as it allows a truly 3D visualization of the heart avoiding some of the problems intrinsically associated with 2D imaging such as foreshortening, out-of-plane motion and the need of geometric assumptions for volume estimation [4]. Unfortunately, due to the intrinsic physical limits of acoustical wave propagation, 3D ultrasound imaging requires advanced beamforming techniques at the cost of the image quality. Indeed, RT3DE currently



(a) End Diastolic (ED) phase



(b) End Systolic (ES) phase

Fig. 1. Illustration of the purpose of the challenge: a competition on segmenting the endocardial surface at both ED (a) and ES (b) time instances from RT3DE

suffers from both a decrease of spatial and temporal resolutions (compared to conventional 2D echo) and the presence of motion artifacts due to stitching of sub-volumes [1]. In this context the development of fully automatic and fast techniques for LV volumetric assessment is still an open issue and thus an active field of research[3,2,6].

Although 3D ultrasonic imaging systems have been widely introduced into the clinical practice over the past decade, and although different (semi-)automatic solutions currently exist to segment the LV, the lack of a common database makes it difficult to evaluate and compare their performance. The aim of the Challenge on Endocardial Three-dimensional Ultrasound Segmentation[1] (CETUS) held during the MICCAI 2014 conference was thus to propose the first common evaluation open platform, that includes 3D echocardiographic acquisitions, reference segmentations obtained from a consensus between 3 expert cardiologists and standard evaluation measures performed in a fully automatic manner thanks to an online system. More precisely, the goal of this challenge was to evaluate and compare the performance of several approaches for the delineation of the LV endocardial border from RT3DE at the end-diastolic (ED) and end-systolic (ES) phases, as illustrated in Figure 1.

The remainder of the paper is as follows. The acquisition protocol and the generation of the reference segmentations are described in Section 2. The evaluation measures and the associated ranking strategy are given in Section 3. Finally the challenge outline is described in Section 4.

[1] http://www.creatis.insa-lyon.fr/Challenge/CETUS/

2 Cardiac data and manual reference

2.1 3D echocardiographic data

Patients - From November 2013 to August 2014, 45 patients referred to three different hospitals (Rennes University Hospital - France, KU Leuven - Belgium and Thoraxcenter - Erasmus MC - Rotterdam - Netherlands) with a clinical indication for 3D echocardiography were included in this study. In order to provide a balanced and representative database of images typically obtained within a clinical context, patients were stratified into 3 groups: 15 healthy subjects, 15 patients with previous myocardial infarction at least 3 months priors to the scan and 15 patients with dilated cardiomyopathy.

Acquisition protocol - In order to avoid biasing the segmentation results toward the equipment of one vendor, RT3DE exams were performed using machines from three different vendors: a GE Vivid E9, using a 4V probe, a Philips iE33, using either an X3-1 or an X5-1 probe, and a Siemens SC2000, using a 4Z1c probe. Moreover, all three hospitals acquired with two different ultrasound systems and were asked to acquire a certain number of patients from each patient group, so that patient group, hospital and ultrasound machine were equally distributed. The following guidelines were followed during the acquisition of the data: 1) the image quality should be as good as possible; choice for harmonics, spatial resolution or other settings were up to the operator; 2) depending on the heart rate, a frame rate of at least 16 volumes per second was targeted (using stitching if needed); 3) stitching artifacts were avoided as much as possible; 4) the coverage of the left ventricle was maximized as much as possible; 5) at end diastole, the mitral should be inside the acquired volume. The 45 patients were equally divided over three batches: Training, Testing 1 and Testing 2, for the different parts of the challenge. Each batch had a similar distribution of pathologies, hospitals and ultrasound machines. Acquired data were fully anonymized and handled within the regulations set by the local ethical committees of each hospital. Given that these images were acquired in clinical practice, one can observed variability in the image quality, as illustrated in Figure 2

(a) Good quality image (b) Mid quality image (c) Poor quality image

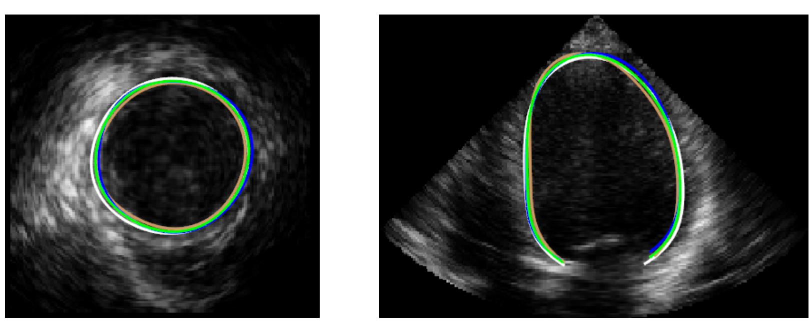
Fig. 2. Variability in the quality of the volumes acquired in clinical conditions

2.2 Generation of the manual reference

Manual contouring of the endocardium at ED and ES was performed independently by 3 expert cardiologists from 3 distinct institutions (King's college of London - UK, Rennes hospital - France, Thoraxcenter - Rotterdam - Netherlands) using an in-house software package developed at KU Leuven termed Speqle3D. If the contours or their clinical parameters differed by more than a predefined level, the tracings were compared and the experts would reach a consensus interpretation on the best segmentation. One or more experts would then adapt their tracings. For more details on the manual contouring protocol, please refer to [5]. In order to uniformize the contouring process and ease the comparison, all volumes were pre-oriented prior to distribution by defining LV long axis, LV apex, LV base and the right ventricle (RV) insertion point. Moreover, each cardiologist was asked to delineate the endocardial border in 4 longitudinal planes through the long axis under 45 degrees angles and five transversal (short-axis) planes divided equally along the long axis between base and apex. These planes were automatically generated by the software. In each long axis view, contours stopped at the mitral hinges. The papillary muscles and trabeculations were included in the LV volume. From this set of 2D contours, a 3D surface triangular mesh was automatically generated by a spherical harmonics interpolation. Given that each cardiologist contoured in the same planes, the three generated meshes were all defined in the same reference system, making the derivation of the mean mesh straightforward, as illustrated in Figure 3. The mean mesh for each patient was then used as the reference when computing the evaluation metrics of the different segmentation solutions.

3 Evaluation metrics and ranking strategy

The performance of the methods involved in the challenge were analyzed by measuring the degree of accuracy of the detected endocardial surface against the ground truth (segmentation accuracy), and by comparing global LV morphological and functional indices (clinical applicability).



(a) Short Axis (SAX) view

(b) Long Axis (LAX) view

Fig. 3. Illustration of the mean mesh result (green contour) computed from the three meshes manually contoured by the experts (blue, brown and white)

3.1 Segmentation accuracy

To measure the degree of accuracy of the extracted endocardial border, three standard metrics were used.

***Mean surface distance* -** The mean surface distance, d_m, between the surface (S) extracted using a (semi-)automatic segmentation method and the corresponding reference surface (S_{ref}) defined as:

$$d_m = \frac{1}{2}\left[\overline{d}(S, S_{ref}) + \overline{d}(S_{ref}, S)\right] \tag{1}$$

where $\overline{d}(S1, S2)$ is the mean of the euclidean distances between every mesh point in $S1$ and the closest surface point in $S2$.

***Hausdorff surface distance* -** The Hausdorff distance, d_H, measures the local maximum distance between the two surfaces S and S_{ref} and is defined as:

$$d_H = \max(\max(d(S, S_{ref})), \max(d(S_{ref}, S))) \tag{2}$$

***Modified Dice similarity index* -** The modified Dice similarity index, D^*, is computed as a measure of overlap between the volume (V) extracted from a (semi-)automatic method and the corresponding reference volume (V_{ref}), giving a measurement value between 0 (full overlap) and 1 (no overlap):

$$D^* = 1 - \frac{2\,(V \cap V_{ref})}{V + V_{ref}} \tag{3}$$

These three metrics were computed for both end-diastole ($d_{m,ED}$, $d_{H,ED}$, D^*_{ED}) and end-systole ($d_{m,ES}$, $d_{H,ES}$, D^*_{ES}). The following notations are introduced for the remainder of this paper:

- $\boldsymbol{d_{m,ED}}$: d_m averaged over all ED images;
- $\boldsymbol{d_{H,ED}}$: d_H averaged over all ED images;
- $\boldsymbol{D^*_{ED}}$: D^* averaged over all ED images;

3.2 Clinical applicability

To measure the ability of the algorithms in extracting relevant clinical indices, modified correlation ($corr^* = 1 - corr$), bias and standard deviation (std) values were computed from the end-diastolic volumes (EDV, expressed in ml), end-systolic volumes (ESV, expressed in ml) and ejection fraction ($EF = 100 * (EDV - ESV)/EDV$, expressed in percentage) measurements. The following notations are introduced for the remainder of this paper:

- $\boldsymbol{EDV_{corr^*}}$: modified correlation computed from EDV measures;
- $\boldsymbol{EDV_{bias}}$: bias computed from EDV measures;
- $\boldsymbol{EDV_{std}}$: the standard deviation computed from EDV measures.

3.3 Ranking procedure

The different methods involved in the challenge were ranked according to the set of measures defined in sections 3.1 and 3.2. First, each individual measure was normalized by the maximum value of the corresponding measures among all participants. By doing so, each measure was normalized between 0 (i.e. the best score obtained if the result perfectly fits the reference mesh) and 1 (i.e. the worst result among all participants). A global score S was then computed for each participant as follows

$$S = \frac{1}{2}\left(M_D + M_C\right). \tag{4}$$

where M_D and M_C correspond to the technical and clinical errors defined as the following weighted average of the constituting error metrics:

$$M_D = \frac{1}{6}\left(d_{m,ED} + d_{m,ES} + d_{H,ED} + d_{H,ES} + D^*_{ED} + D^*_{ES}\right) \tag{5}$$

$$M_C = \frac{1}{9}\left(EDV_{corr^*} + EDV_{bias} + EDV_{std} + ESV_{corr^*} + ESV_{bias} + ESV_{std} + EF_{corr^*} + EF_{bias} + EF_{std}\right) \tag{6}$$

A single ranking was then performed using the global scores for each participant as defined by Eq. (4).

3.4 Midas online evaluation platform

The participant registration, the computation of the different error measures and the ranking of the involved methods were done automatically through a dedicated Midas[2] online platform [3] which was specifically designed for this challenge. A screen capture of the corresponding interface is given in Figure 4.

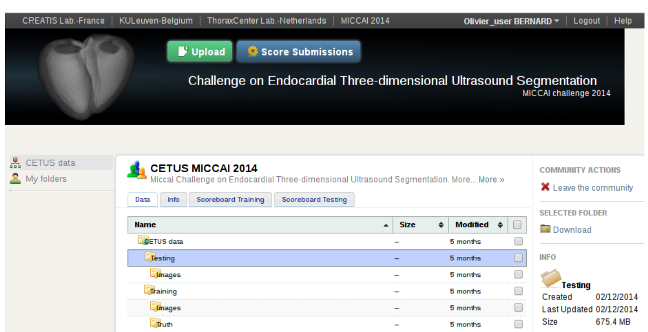


Fig. 4. Screen capture of the Midas online platform designed for this challenge

[2] http://www.midasplatform.org
[3] https://miccai.creatis.insa-lyon.fr/miccai/

4 Challenge outline

The CETUS challenge officially started in March 2014 with the invitation of a large number of researchers working on cardiac ultrasound segmentation to visit the website and to participate in the challenge. Twenty-one teams initially registered to the challenge and 10 of them submitted a paper in June 2014. The CETUS event went through three consecutive dataset releases, each of them being equally distributed among vendors.

4.1 Dataset distribution

Release 1 (Training dataset) - Early March 2014, participants were given a dataset consisting of 15 patients (5 healthy subjects, 5 patients with previous myocardial infarction and 5 patients with dilated cardiomyopathy). For each patient of this dataset, the following data was released: a sequence of volumes (saved in raw/mhd format), the associated reference meshes (saved in vtk format) and the positions in the sequence of the volumes corresponding to ED and ES.

Release 2 (Testing1 dataset) - Mid May 2014, participants were given a dataset consisting of 15 new patients (5 healthy subjects, 5 patients with previous myocardial infarction and 5 patients with dilated cardiomyopathy). For each patient of the dataset, the following data was released: a sequence of volumes (saved in raw/mhd format) and the positions in the sequence of the volumes corresponding to ED and ES. From the 1st of August, a public web-page displayed in real-time the current ranking of the challengers using the segmentation results of this dataset. By doing so, the participants were encouraged to keep on improving their method until the day of the challenge.

Release 3 (Testing2 dataset) - On the day of the challenge (September 14th 2014, at MICCAI), participants were given a new dataset consisting of 15 new patients (5 healthy subjects, 5 patients with previous myocardial infarction and 5 patients with dilated cardiomyopathy). A three-and-a-half hours-time-slot was dedicated to the on-site competition.

4.2 Method categories

This challenge accepted both automatic and semi-automatic segmentation solutions. An automatic method does not require any landmarks. Moreover the same settings should be used to process the different dataset. A semi-automatic method was allowed to have a small number of manual steps in order to initialize the algorithm. Adjustments of the resulting contours after segmentation were not allowed. Among the three semi-automatic methods that compete in the CETUS challenge, an average of 4 points has been used for the manual initialization step.

4.3 Day of the challenge

During the on-site workshop, the testing2 dataset was processed by the participating teams. The different methods were ranked within their sub-categories: the semi-automatic and the fully automatic approaches. The final score was calculated as the mean value of the scores obtained from each testing dataset (average value of the global score S obtained from the testing1 and testing2 dataset). For each sub-category, a diploma was awarded to the challenger that obtained the best global score. Finally, the team that obtained the best score from all the participants (whatever the sub-category) was rewarded by a life-size human heart model, sponsored by Materialise inc. (`http://www.materialise.com/`).

4.4 After the challenge

The results of this evaluation framework will be reported in a collaborative technical paper collating the automatic and semi-automatic segmentation results. Moreover, the protocol of the manual contouring of the endocardial border was specifically developed during this challenge and will be submitted to a clinical journal.

References

1. Dorosz, J., Lezotte, D., Weitzenkamp, D., Allen, L., Salcedo, E.: Performance of 3-dimensional echocardiography in measuring left ventricular volumes and ejection fraction. J Am Coll Cardiol 59(20), 1799–1808 (2012)
2. Leung, E., Danilouchkine, M., van Stralen, M., de Jong, N., van der Steen, A., Bosch, J.: Left ventricular border tracking using cardiac motion models and optical flow. Ultrasound Med Biol 37(4), 605 – 616 (2011), `http://www.sciencedirect.com/science/article/pii/S0301562911000536`
3. Leung, K., Bosch, J.: Automated border detection in three-dimensional echocardiography: principles and promises. Eur J Echocardiogr 11(2), 97–108 (2010)
4. Mor-Avi, V., Jenkins, C., Kuhl, H., Nesser, H., Marwick, T., Franke, A., Ebner, C., Freed, B., Steringer-Mascherbauer, R., H.Pollard, Weinert, L., Niel, J., Sugeng, L., Lang, R.: Real-time 3-dimensional echocardiographic quantification of left ventricular volumes: Multicenter study for validation with magnetic resonance imaging and investigation of sources of error. JACC: Cardiovasc Im 1(4), 413 – 423 (2008)
5. Papachristidis, A., Geleijnse, M., Galli, E., Heyde, B., Alessandrini, M., Barbosa, D., Donal, E., Monaghan, M., Bernard, O., D'hooge, J., Bosch, J.G.: Clinical expert delineation of 3d left ventricular echocardiograms for the cetus segmentation challenge. In: Proc. of MICCAI Workshop: Challenge on Endocardial Three-dimensional Ultrasound Segmentation. (2014)
6. Yang, L., Georgescu, B., Zheng, Y., Wang, Y., Meer, P., Comaniciu, D.: Prediction based collaborative trackers (PCT): A robust and accurate approach toward 3D medical object tracking. IEEE T Med Imaging 30(11), 1921 –1932 (2011)