Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography

Olivier Bernard*, Johan G. Bosch, Brecht Heyde, Martino Alessandrini, Daniel Barbosa, Sorina Camarasu-Pop, Frederic Cervenansky, Sébastien Valette, Oana Mirea, Michel Bernier, Pierre-Marc Jodoin, Jaime Santo Domingos, Richard V. Stebbing, Kevin Keraudren, Ozan Oktay, Jose Caballero, Wei Shi, Daniel Rueckert, Fausto Milletari, Seyed-Ahmad Ahmadi, Erik Smistad, Frank Lindseth, Maartje van Stralen, Chen Wang, Örjan Smedby, Erwan Donal, Mark Monaghan, Alex Papachristidis, Marcel L. Geleijnse, Elena Galli, and Jan D'hooge

***Abstract*—Real-time 3D Echocardiography (RT3DE) has been proven to be an accurate tool for left ventricular (LV) volume assessment. However, identification of the LV endocardium remains a challenging task, mainly because of the low tissue/blood contrast of the images combined with typical artifacts. Several semi and fully automatic algorithms have been proposed for segmenting the endocardium in RT3DE data in order to extract relevant clinical indices, but a systematic and fair comparison between such methods has so far been impossible due to the lack of a publicly available common database. Here, we introduce a standardized evaluation framework to reliably evaluate and compare the performance of the algorithms developed to segment the LV border in RT3DE. A database consisting of 45 multivendor cardiac ultrasound recordings acquired at different centers with corresponding reference measurements from three experts are made available. The algorithms from nine research groups were quantitatively evaluated and compared using the proposed online platform. The results showed that the best methods produce promising results with respect to the experts' measurements for the extraction of clinical indices, and that they offer good segmentation precision in terms of mean distance error in the context of the experts' variability range. The platform remains open for new submissions.**

***Index Terms*—Endocardium, left ventricle segmentation, real-time 3D echocardiography, standardized evaluation system.**

Manuscript received October 09, 2015; revised November 21, 2015; accepted November 23, 2015. Date of publication November 25, 2015; date of current version March 31, 2016. *Asterisk indicates corresponding author.*

O. Bernard, S. Camarasu-Pop, F. Cervenansky, and S. Valette are with the University of Lyon, CREATIS, CNRS UMR5220, Inserm U630, INSA-Lyon, University of Lyon 1, 69100 Villeurbanne, France (e-mail: olivier.bernard@creatis.insa-lyon.fr).

J. G. Bosch and M. L. Geleijnse are with the Thoraxcenter Biomedical Engineering, Erasmus MC, 3000 CA Rotterdam, The Netherlands.

B. Heyde, M. Alessandrini, O. Mirea and J. D'hooge are with the Lab on Cardiovascular Imaging and Dynamics, Department of Cardiovascular Sciences, KU Leuven, 3000 Leuven, Belgium.

D. Barbosa is with the Life and Health Sciences Research Institute (ICVS), University of Minho, 4704-553 Minho, Portugal.

M. Bernier and P. M. Jodoin are with the Université de Sherbrooke, Sherbrooke, QC, J1K 2R1 Canada.

J. S. Domingos and R.V. Stebbing are with the Institute of Biomedical Engineering, University of Oxford, OX1 2JD Oxford, U.K.

K. Keraudren, O. Oktay, J. Caballero, W. Shi, and D. Rueckert are with the Biomedical Image Analysis Group, Imperial College London, SW7 2AZ London, U.K.

F. Milletari is with the Computer Aided Medical Procedures, Technische Universität München, 80333 München, Germany.

S. A. Ahmadi is with the Department of Neurology, Klinikum Grosshadern, Ludwig-Maximilians-Universität München, 80539 München, Germany.

E. Smistad and F. Lindseth are with the Norwegian University of Science and Technology, 7491 Trondheim, Norway.

M. van Stralen is with Image Sciences Institute, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.

C. Wang and Ö. Smedby are with the Center for Medical Imaging Science and Visualization (CMIV), Department of Medical and Health Sciences (IMH), Linköping University, SE-58185 Linköping, Sweden and the School of Technology and Health, Royal Institute of Technology—KTH, SE-14152 Huddinge, Sweden.

E. Galli and E. Donal are with the Cardiology and CIC-IT, CHU Rennes, and LTSI—INSERM U 1099—University of Rennes-1, 35000 Rennes, France.

A. Papachristidis and M. Monaghan are with the King's College Hospital NHS Foundation Trust, SE5 9RS London, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2015.2503890

I. Introduction

LEFT ventricular geometry and function are important factors in terms of patient management, outcome, and long-term survival [1], [2]. Thus, the assessment of left ventricular volumes throughout the cardiac cycle and its associated indices derived from volume traces is a routine task in diagnostic cardiology. To this end, several imaging modalities such as echocardiography, cardiovascular magnetic resonance (CMR), cardiac computed tomography (CT) or positron emission computed tomography (PET) can be used.

While magnetic resonance imaging remains the gold standard for the assessment of cardiac morphology and function, several studies have shown that real-time 3D echocardiography can provide comparable results but with the distinct potential of performing the acquisition in real-time at high temporal resolution and bed-side, thereby giving it the potential to be the preferred modality in the near future [3], [4]. Indeed, over the last decade the assessment of cardiac morphology and function by ultrasound imaging has made a significant step forward by the introduction of 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 [3]. Unfortunately, due to the intrinsic physical limits of acoustical wave propagation, RT3DE currently suffers from a low spatial and temporal resolution compared to conventional 2D echocardiography and the potential presence of motion artifacts due to stitching strategies [5]. As a consequence, state-of-the-art commercial solutions towards LV segmentation still require some degree of user interaction both at the initialization step and after segmentation/tracking when corrections are

required [4]. Thus, the development of fully automatic and fast techniques for LV volumetric assessment is still an open issue and therefore an active field of research [6].

Although different solutions currently exist to segment the LV, the lack of a common database of recordings makes it difficult to evaluate and contrast their performance. The novelties introduced in this paper are the following[1]:

- provide a common set of RT3DE datasets together with reference measurements based on an established protocol to manually extract the endocardial surface in RT3DE;
- introduce an online open framework to consistently evaluate the performance of semi and fully automatic methods for the delineation of the LV border from RT3DE at end-diastolic and end-systolic phases;
- contrast state-of-the art segmentation algorithms in the field of ultrasound cardiac imaging using the aforementioned framework;
- identify and locate for each method the segmented regions with the highest errors in order to provide useful directions for further investigations.

The paper is organized as follows. In Section II, we discuss previous work on LV detection in RT3DE imaging. The evaluation framework is presented in Section III. Section IV provides a brief description of the nine evaluated methods. The results of the involved methods as produced by the online system are presented in Section V and discussed in Section VI. Finally, concluding remarks are given in Section VII.

II. Previous Work

Since LV segmentation in 3D echocardiography is a subject of research that has been investigated for more than ten years, a wide range of image processing methods are already available in this field [6]. Techniques successfully applied for LV segmentation in RT3DE can be categorized based on their underlying methodology: deformable models, surface fitting approaches, graph-cuts, multi-atlas and machine learning techniques. Deformable models constitute the most widely used technique to segment 3D ultrasound data because of its attractiveness in terms of mathematical formulation of the segmentation problem, its flexibility in terms of shape representation and its computational time efficiency (depending on the chosen representation) [8]–[11]. Surface fitting approaches are based on shape interpolation from a set of points that were previously identified as belonging to the endocardial surface. The most popular techniques that have been applied in RT3DE are active shape models (ASM) [12] and Doo-Sabin subdivision surfaces (DSS) [13], [14]. The main feature of ASM corresponds to the embedded representation of the shape based on a space built from a set of references. These methods are known to require a large number of training samples to be meaningful. The main asset of DSS is its capacity to represent smooth surfaces by a small number of parameters. Graph-cuts and multi-atlas segmentation are methods that have been recently applied in RT3DE [15], [16]. In graph-cut techniques, the underlying graph is usually expressed in a coordinate system adapted to the anatomy of the endocardial surface. The attractiveness of such approach resides in its capacity of expressing the segmentation problem as a graph and thus benefits from powerful graph-flow optimization techniques. Multi-atlas segmentation techniques are registration-based techniques that perform the segmentation of the endocardium from a set of pre-segmented volumes. Based on the selection of the best atlases after registration, merging techniques are usually used to enhance the quality of the segmented results. Finally, machine learning approaches have proven their value for LV segmentation in several studies [14], [17], [18]. Usually, these techniques are not designed to produce the final segmented surface but provide a fundamental intermediate step that guides the segmentation process. For example, methods based on marginal space learning have been specially designed to accurately locate the myocardial region [18]. Similarly, a boundary fragment model has been successfully applied for the identification of extracted contours [14]. Finally, random forest-based approaches have been applied as a discriminative classifier in order to characterize the affiliation of each voxel to the myocardium [17].

All these techniques could also be categorized into fully automatic [11], [18]–[21], [24] and semi-automatic [9], [23] approaches. For the latter case, the amount of user interaction for initialization could be considered either as low (few anatomical landmarks are needed) [9], [23] or time consuming (manual contouring at the end-diastolic frame is required) [22]. Since time is a key point in clinical ultrasound, it has to be noticed that most of the proposed techniques are fast, some even running in real-time (with processing times around 40 ms for the most efficient [19], [21]). Table I gives a summary of the accuracy of existing automatic 3D LV segmentation techniques when compared with manual references. Although this table provides some evidence of the relative performance of the different approaches, it is clear that a fair and true comparison is not feasible given the differences in patient population and image quality (due to different acquisition conditions and equipment). This points to the need for a publicly available common database.

III. Evaluation Framework

A. 3D Echocardiographic Data

1) Patient Selection: From November 2013 to August 2014, 3D images from 45 patients referred to three different hospitals (Rennes University Hospital—France, University Hospitals Leuven—Belgium and Thoraxcenter—Erasmus MC—Rotterdam—Netherlands) for 3D echocardiography were selected and included in this study (within the regulations set by the local ethical committee of each hospital). In order to provide a balanced and representative database of images with typical segmentation challenges, data was acquired in a population divided into 3 subgroups: 15 healthy subjects, 15 patients with a previous myocardial infarction at least 3 months prior to time of acquisition and 15 patients with dilated cardiomyopathy.

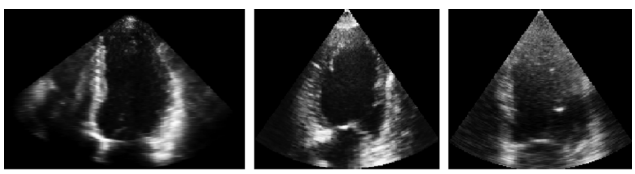
2) Acquisition Protocol: In order to avoid biasing the segmentation results toward the equipment of one vendor, RT3DE

[1]A preliminary version of this work appeared in [7]. While in [7] only the organization of the challenge is reported, the present paper evaluates in details the performance of nine state-of-the art segmentation algorithms on a large dataset of 45 patients with 3D reference meshes obtained by cardiologist consensus with known inter-observer variabilities. It is to the best of our knowledge the first time that algorithms in the field of ultrasound cardiac imaging could be accurately evaluated on such database.

TABLE I
STATE-OF-THE-ART ALGORITHMS FOR LV SEGMENTATION IN RT3DE

Study *	Algorithm		Volumes						Segmented Shapes		
	#	T	R			BA ($\mu \pm 2\sigma$)			MAD ($\mu \pm \sigma$)		
			EDV	ESV	EF	EDV (ml.)	ESV (ml.)	EF (%)	ED (mm)	ES (mm)	FC (mm)
Angelini *et al.* [9]	10		0.63	0.62	0.45	16.1±50	6.6±34	0.5±22			
Hansegard *et al.* [19]	21	0.04	0.91	0.91	0.74	-5.9±21	6.2±19	-7.7±12			2.2±0.6
Leung *et al.* [20]	99		0.95			-1.47±40			2.9±1.0		
Duan *et al.* [21]	35	0.033									4.0±3.2
Leung *et al.* [22]	35	6	0.982			1.9±14					1.35±0.5
Yang *et al.* [18]	67	1.5				1.32±12		1.0±10			1.28±1.1
Rajpoot *et al.* [23]	34					-5.0±49	1.2±26	-0.7±14	2.2±0.7	1.52±0.4	
Dikici *et al.* [24]	29	0.08							2.0±X	2.0±X	
Barbosa *et al.* [11]	24	1	0.97	0.97	0.91	-1.4±23	2±19	-1.0±10			

** The following symbols were used: EDV: End Diastolic Volume; ESV: End Systolic Volume; EF: Ejection Fraction; FC: full cycle; #: number of exams; T: average frame processing time (s) per volume; R: correlation coefficient; BA: Bland-Altman analysis; MAD: mean absolute surface distance.*



(a) (b) (c)

Fig. 1. Variability in the quality of the volumes acquired in clinical conditions (a) good (b) fair (c) poor.

TABLE II
CHARACTERISTICS OF THE TWO DATASETS. RESULTS AS MEAN ± STANDARD DEVIATION. *: AVERAGE SIGNIFICANTLY DIFFERENT FROM THE TRAINING SET (UNPAIRED T-TEST, $p < 0.01$)

Dataset	EDV	ESV	EF	Image Quality
	ml	ml	%	good/fair/poor
Training	213 ± 97	151 ± 91	33 ± 15	6 / 6 / 3
Testing	152 ± 62*	93 ± 54*	41 ± 11	8 / 10 / 12
Total	172 ± 80	113 ± 73	39 ± 13	14 / 16 / 15

exams were performed using machines from three different vendors: a GE Vivid E9, using a 4V probe, a Philips iE33, using either a X3–1 or a 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 five patients from each patient group, so that patient group, hospital and ultrasound systems were equally distributed. Only images of acceptable quality for clinical diagnosis were included. The following guidelines were followed during the acquisition and inclusion of the data. Image quality had to be sufficient for visual analysis by an expert. Optimization, e.g., choice for harmonics, spatial resolution or other system settings were up to the operator. Exclusion criteria were i) left bundle branch block (LBBB) or visually dyssynchronous LV; ii) frame rate below 16 volumes per second; iii) mitral plane out of image sector at ED; iv) significant stitching or other type of artifacts affecting the visibility of endocardium throughout the cardiac cycle; v) poor visualization of a LV wall or a wall out of the image sector to an extent that the image can no longer be manually analyzed with good confidence (since one important goal of this study is the assessment of clinical indices based on volume measurements). Given that datasets were acquired in a clinical setting, a considerable variability in image quality could be anticipated (Fig. 1).

3) Training and Testing Datasets: Fifteen of the 45 RT3DE recordings, together with reference meshes, are made available for training or tuning of the algorithms. The remaining 30 datasets are used for testing. Care was taken to ensure that the training and testing datasets 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. All data was converted to a general 4D image representation format (mhd/raw) without loss of resolution. ED and ES frames were identified based on ECG and valve opening/closure by a single expert. The training dataset is released with the associated reference meshes (saved in vtk format) obtained with the contouring protocol described in Section III-B.

Characteristics of the training and testing datasets are given in Table II. Image quality was assessed by one clinical expert as good, fair or poor. Image quality was slightly better in the training set. It can be seen that ED volumes as well as ES volumes are significantly different ($p < 0.01$) between the training and testing sets: the training set generally has higher volumes. These are unwanted effects of the fairly low amount of patients per subgroup. We refer the reader to [7] for more details on the acquisition setup.

B. Reference Segmentation

Establishing a well-defined ground truth segmentation was of utmost importance for this work. However, there are no clear guidelines for endocardial contouring in 3D echocardiography. Therefore, considerable effort was spent to define a consistent contouring method for manual segmentation of the 3D echocardiographic data. For the ground truth drawings, we aimed for a contour definition in line with clinical standards used in 2D echocardiography. A detailed contouring guideline was set up at the beginning of the study. This guideline was refined during the training phase (contouring of the first 15 patients) and also used to resolve conflicts during consensus discussions.

1) Contouring Protocol: We refer the reader to [25] for a complete description of the contouring procedure and protocol. In short, convention was defined for LV wall, mitral valve (MV) plane, trabeculations, papillary muscles and apex. Basic points were to a) include trabeculae and papillary muscles in the LV

cavity; b) keep tissue consistency between ED and ES instants; c) terminate the contours in the MV plane on the ventricular side of the bright ridge, at the points where the valve leaflets are hinging; d) partially exclude left ventricular outflow tract (LVOT) from the cavity by drawing from septal MV hinge point to the septal wall to create a smooth shape; e) draw the apex high up near the epicardium both in ED and ES taking into consideration that there should be little motion of the true apex point.

2) Tracing Protocol: Manual contouring of the endocardium at the ED and ES phases was performed independently by 3 expert cardiologists from 3 distinct institutions using a custom non-commercial contouring package for 3D echocardiograms (named Speqle3D), developed by the University of Leuven and tested in an earlier study [26]. In order to minimize the impact of the reference mesh design on the segmentation error measurement, the cardiologists were asked to manually select a dense number of 3D points that belong to the endocardial border. A protocol was defined that allowed extracting this dense point cloud in a reproducible manner. At first, all datasets were oriented by defining the LV long axis, LV apex, LV base and the right ventricle insertion point. Then, each expert independently traced the endocardial border in a set of predefined planes (short axis and longitudinal planes). For each longitudinal plane about 15 points were set at the endocardial border, starting from the mitral valve plane. In the short axis planes about 10 points were set using a similar process. Finally expert meshes were generated through spherical harmonic least square fitting of the 3D point cloud. By doing so, the 3D endocardial surfaces were defined by each expert from more than 110 3D points manually annotated, which we considered to be a good trade-off between the degree of smoothness of the final endocardial shape, the accuracy of the extracted border (determined by the resolution and quality of the ultrasound equipment) and the time required for manual contouring. Since for the clinical study the volumes are computed directly from the reference meshes and since all the evaluated methods perform segmentation directly in 3D, we also decided to use the interpolated reference meshes for the computation of the segmentation accuracy.

3) Evaluation of Correspondence and Consensus: Consensus between experts was checked from pairwise differences in LV volume and EF and Hausdorff distances. To qualify for consensus between all operators the following criteria had to be met: relative difference in LV volume $\leq$ 10%, absolute difference in LVEF $\leq$ 5% and Hausdorff distances $\leq$ 5 mm. All experts were asked to review the contouring of sets that did not meet criteria in the first round and suggest modifications, and one or more experts would retrace. Differences were then checked against (slightly relaxed) criteria: the average of the three pairwise observer differences was evaluated, and Hausdorff distances $\leq$ 7 mm were accepted (the same initial criteria relative to the volumes and the EF having been used). In only two cases, the three operators did not agree within the consensus criteria. These drawings were then accepted, in the context of persistent observer interpretation difference. From the final contours, a mean mesh was constructed. This mean mesh (named reference mesh in the sequel) was used in the online evaluation platform as the experts' ground truth.

C. Evaluation Metrics

The goal of the online platform is twofold: measuring the degree of accuracy of the detected endocardial surface against the ground truth (image processing side) and comparing global LV morphological and functional indices (clinical side). Two sets of complementary measurements have thus been deployed as described hereunder. It is however worth to point out that a good segmentation method will not necessary obtain the equivalent good scores in terms of clinical indices extraction (and vice versa), emphasizing the need to analyze the results in a separate way.

1) Segmentation Accuracy: To measure the degree of accuracy of the extracted endocardial surface S_{user} obtained from a semi or fully automatic segmentation method, three standard metrics are used.

Mean Surface Distance (d_m): The mean surface distance, d_m, between the surface S_{user} and the corresponding reference surface S_{ref} is defined as the mean of the Euclidean distances between every point in s_{user} and the closest surface point in S_{ref}. This is carried out efficiently using the Proximity Query Package (PQP) [27], Which we have slightly modified to compute point to triangle distances. One interest of the mean surface distance metric is to avoid situations where an evaluated method obtains good segmentation results by over and under-estimating different local regions of the reference surface with the same amount of error.

Hausdorff Surface Distance (d_H): The Hausdorff distance, d_H, measures the local maximum distance between the two surfaces S_{user} and S_{ref}. This criterion was also computed using PQP. Note that in order to minimize the difference between sampling densities of S_{user} and S_{ref}, We apply a linear subdivision operator to the surface containing the lowest number of vertices.

Modified Dice Similarity Index: The modified dice similarity index, $D^* = 1-2(|V \cap V_{ref}|)/(|V|+|V_{ref}|)$, is computed as a measure of overlap between the segmented volume (V) extracted from a semi or fully automatic method and the corresponding reference volume (V_{ref}), Giving A Measurement Value Between 0 (full Overlap) and 1 (no Overlap).

These three metrics are computed for both ED and ES and averaged over all patients, e.g., notated as $d_{H,ED}$.

2) Clinical Performance: To measure the ability of the algorithms in extracting relevant clinical indices, modified correlation ($corr^* = 1 - corr$), bias and standard deviation (std) values are computed from the end-diastolic volumes (EDV, expressed in ml), end-systolic volumes (ESV, expressed in ml) and ejection fractions ($EF = 100 * (EDV - ESV)/EDV$, expressed in percent) measurements. The combination of the bias and standard deviation also provides useful information on the corresponding limit of agreement values. The following notations are used in the rest of this paper:

- $\boldsymbol{EDV_{corr^*}}(\boldsymbol{ESV_{corr^*}}, \boldsymbol{EF_{corr^*}})$: modified correlation computed from EDV (resp. ESV, EF) measures;
- $\boldsymbol{EDV_{bias}}(\boldsymbol{ESV_{bias}}, \boldsymbol{EF_{bias}})$: bias computed from EDV (resp. ESV, EF) measures;
- $\boldsymbol{EDV_{std}}(\boldsymbol{ESV_{std}}, \boldsymbol{EF_{std}})$: the standard deviation computed from EDV (resp. ESV, EF) error measures.

TABLE III
LEFT VENTRICLE SEGMENTATION ACCURACY OF THE NINE EVALUATED METHODS ON THE FULL TESTING DATASET (30 PATIENTS). THE VALUES IN BOLD CORRESPOND TO THE BEST PERFORMANCE FOR EACH MEASURE. CPU VALUES REPRESENT THE TIME NEEDED FOR THE SEGMENTATION OF ONE VOLUME

Method			ED						ES					
			d_m		d_H		D^*		d_m		d_H		D^*	
			mean	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*
Participants	category	cpu	mm		mm		val.		mm		mm		val.	
Inter-observer			1.39	0.40	4.70	1.27	0.069	0.021	1.34	0.35	4.70	1.15	0.080	0.021
Bernier *et al.* [28]	Semi-auto.	2 *sec.*	2.37	**0.60**	9.41	2.62	0.118	**0.029**	2.64	**0.60**	9.34	**2.08**	0.163	**0.047**
Domingos *et al.* [29]	Semi-auto.	16 *sec.*	**2.09**	0.68	9.31	3.89	**0.106**	0.038	**2.20**	0.72	**8.35**	2.67	**0.129**	0.050
Oktay *et al.* [31]	Semi-auto.	32 *min.*	2.18	0.70	**7.55**	**1.77**	**0.106**	0.033	2.47	0.74	8.57	2.96	0.151	0.049
Wang *et al.* [33]	Semi-auto.	90 *sec.*	2.54	0.99	9.04	3.58	0.125	0.042	2.68	1.11	9.14	3.33	0.159	0.057
Barbosa *et al.* [36]	Fully-auto.	5 *sec.*	2.26	0.73	**8.10**	**2.66**	**0.106**	0.041	**2.43**	0.91	**8.13**	3.08	**0.144**	0.057
Keraudren *et al.* [38]	Fully-auto.	90 *sec.*	2.44	0.95	8.98	3.09	0.130	0.048	2.54	**0.75**	9.15	3.24	0.158	0.057
Milletari *et al.* [40]	Fully-auto.	40 *sec.*	**2.14**	**0.68**	8.25	3.87	0.107	**0.031**	2.91	1.01	8.53	**2.30**	0.162	0.062
Smistad *et al.* [42]	Fully-auto.	65 *msec.*	2.62	0.95	8.26	2.98	0.115	0.038	2.92	0.93	8.99	2.98	0.156	**0.050**
van Stralen *et al.* [44]	Fully-auto.	60 *sec.*	2.44	0.91	8.45	3.50	0.121	0.054	2.79	1.24	8.65	2.85	0.165	0.079

The code used to compute the different evaluation metrics is publicly-accessible at the web address used for the presentation of the online platform: https://miccai.creatis.insa-lyon.fr/miccai/.

D. Method Categories

Depending on the amount of user-interaction, we distinguish two categories of algorithms on the proposed platform:

- Fully automatic: fully automatic methods segment the endocardial surface without any user-interaction. The RT3DE is the only input used by the method.
- Semi-automatic: semi-automatic methods are allowed having a small number of manual steps in order to initialize the algorithm. Adjustments of the resulting contours after segmentation are not allowed.

E. Online Evaluation Platform

The registration of the teams and the computation of the different error measures of the involved methods are done through a dedicated Midas[2] online platform[3] which was specifically installed and customized for the purpose of this study. This platform is now available for new submissions. It will be maintained and kept open as long as the data remains relevant to clinical practice. The code corresponding to the evaluation platform is publicly-accessible at https://miccai.creatis.insa-lyon.fr/miccai/.

F. MICCAI 2014 Workshop

The evaluation framework was launched during the "Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS)" workshop which was organized during the 17th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), on September 14th, 2014 in Boston, USA [7]. Around 50 groups from academia and industry were invited by e-mail to participate in the segmentation challenge. Twenty-one teams created an account on the dedicated website. Nine teams that managed to achieve meaningful results within the allotted time for training participated in the challenge workshop.

[2]http://www.midasplatform.org

[3]https://miccai.creatis.insa-lyon.fr/miccai/

IV. EVALUATED ALGORITHMS

The nine evaluated algorithms are described in this Section. More details can be found in the corresponding proceedings papers of the workshop which have been published in the open-access MIDAS journal (http://www.midasjournal.org/browse/journal/76). Among these methods, four belong to the semi-automatic sub-category and five to the fully-automatic sub-category (see Table III).

1) Bernier et al. (Semi-Automatic Method): [28] proposed a fast semi-automatic method based on graph cut and an implicit U-shape prior. In a first step, the user was required to place a T-bar target (three point clicks) at the base of the LV and extends it all the way to the apex. From this target, the 3D volume was sampled and projected to a spherical-cylindrical coordinate system. Then, a graph was made in which each node was associated to a voxel. The capacity of each edge was assigned a gradient-based energy term, some of which being infinite to make sure the resulting volume passes through key anatomical points. Then, a graph-cut procedure provided delineation of the endocardial surface which, in turn, was projected back to the original Euclidean space.

2) Domingos et al. (Semi-Automatic Method): [29] proposed a semi-automatic two-step method. Manual initialization was performed through means of a 3D interactive GUI. Processing each short-axis slice independently, a Structured Random Forest (SRF) [30] was used to predict the probability of each pixel belonging to the endocardium-blood interface from a given image patch. Following detection of an endocardial surface, an explicit continuous LV surface model based on Doo-Sabin subdivision surfaces [13] was deformed to a subset of the boundary candidates in each volume, firstly in a standard "model-to-data" approach, and finally in a "data-to-model" step where all detected boundary candidates were used to fit the surface.

3) Oktay et al. (Semi-Automatic Method): [31] applied a multi-atlas label propagation method. To initialize the algorithm, the user was required to place three points (apex,

mid-ventricle, and mitral valve location) by doing a rigid alignment between the target and the atlas images before performing the image registration. In order to improve correspondences between ultrasound images, a spectral representation expressed through a dictionary space was used [32]. This image feature was shown to be effective in aligning ultrasound images by achieving state-of-the-art segmentation accuracy with a multi-atlas approach while being computationally efficient.

4) Wang et al. (Semi-Automatic Method): [33] proposed a semi-automatic method, which utilizes a multi-scale quadrature filter method [34] to enhance the 3D volume, followed by a model-based level-set method. As initialization, four points were indicated in order to define the long and short axis of the left ventricle. Moreover, the short axis was used to estimate the initial orientation of the model and thus had to point to the right ventricle. In addition to image enhancement, the integrated phase map of the multi-scale filters was used to weigh the model fitting process. To speed up the proposed segmentation method, a fast level-set method using coherent propagation was implemented [35].

5) Barbosa et al. (Fully-Automatic Method): [36] proposed the following workflow. The segmentation of ED frame started from an ellipsoid surface automatically fitted to the contours of the left ventricle [11]. This ellipsoid approximation of the LV anatomy was then further refined using B-spline Explicit Active Surfaces [37]. The cardiac motion during the cardiac cycle was estimated using a tracking-oriented approach, using both a global optical flow-based tracker and local block matching. The core novelty of the proposed algorithm relied on the recursive formulation of the block-matching problem, which introduced temporal consistency on the patterns being tracked.

6) Keraudren et al. (Fully-Automatic Method): [38] segmented the LV endocardial surface iteratively using a succession of Random Forest classifiers, based on the autocontext framework [39]. The image features used were selected as: (1) pixel intensities, (2) pixel labels obtained in previous classification steps, and (3) geodesic distance of pixels to the centers of the LV, mitral valve and myocardium. By using the labels computed in the preceding steps, each classifier gained contextual information and the resulting classification pipeline implicitly learned the shape of the LV and the spatial layout of the mitral valve and the myocardium.

7) Milletari et al. (Fully-Automatic Method): [40], proposed a method based on Hough votes incorporating shape and appearance priors into a decision forest [41]. During training, annotated volumes were used to obtain features, votes, segmentation and intensity patches. The votes encoded relative displacements between the LV center and the current position of the data-point. Segmentations were obtained by discriminating data-points. Each of the data-points belonging to the narrow band around the contour was enabled to cast votes in order to localize the center of the LV in the image. Once the LV center was determined, all the votes that accumulated around its location back-projected the associated segmentation patches to the position of the data-points that had cast them.

8) Smistad et al. (Fully-Automatic Method): [42] proposed a method based on the use of a mesh model of the LV that was transformed using global and local deformation. The deformation parameters were predicted and estimated for each frame using an extended Kalman filter. This method was based on the approach of [13]. However, the proposed method used mean value coordinates instead of subdivision surfaces to deform the mesh model locally. This enabled deformation of a complex model with few parameters. Edge detection along the normal of each vertex in the mesh model were used as observations for the Kalman filter. Edges were detected using the STEP model [43].

9) Van Stralen et al. (Fully-Automatic Method): [44] proposed to segment the LV endocardial surface by using 3D Active Appearance Models. Separate ED and ES AAMs were built from the reference meshes provided through the training database, extended with 25 previously acquired 3D echocardiographic volumes, imaged using various equipment. Multiple automatic initialization strategies were used and the best match was selected based on the AAM matching residual. The ED segmentation served as initialization for ES matching after scaling. A bias correction was applied by scaling segmentations along the long axis with 95% to compensate for the wider drawings in the training data.

Finally, it is worth to note that the broad range of methods evaluated in this study covers the current solutions implemented in the standard echocardiographic equipment (e.g., machine learning [18], statistical shape model [45] and Kalman filtering [19].)

V. Results

The results given in this section are based on the nine algorithms described above. Moreover, the result tables provide mean (or standard deviation) expert inter-observer variability values for each measure. This allows us to compare the best scores obtained by the algorithms relatively to the agreement reached by the experts. Finally, for each algorithm, the time needed to perform the segmentation of one single volume is given. Obviously, these values are only indicative since both the programming language and hardware configuration are different for each method.

A. Left Ventricle Segmentation Accuracy

Table III shows the segmentation accuracy of the 9 algorithms (4 semi-automatic, 5 fully automatic) and the corresponding expert inter-observer variabilities at ED and ES. Each of the measures corresponds either to the mean or the standard deviation of the different metrics computed at ED and ES on the full testing dataset for each participant. From these results, the following observations can be made. In the semi-automatic category, in terms of mean value, the method of Domingos *et al.* outperforms all the other methods for all but one of the measures. Additionally, this method showed competitive behavior in terms of standard deviation of the distance metrics. In the fully automated category, the same conclusions can be made for the method of Barbosa *et al.*.

For this dataset, the overall best algorithms produced segmentation results whose accuracy was close (1.6 times bigger on average) to the inter-expert variability range after consensus agreement. Although there still exists room for improvement, these results are very encouraging regarding the capacity of the

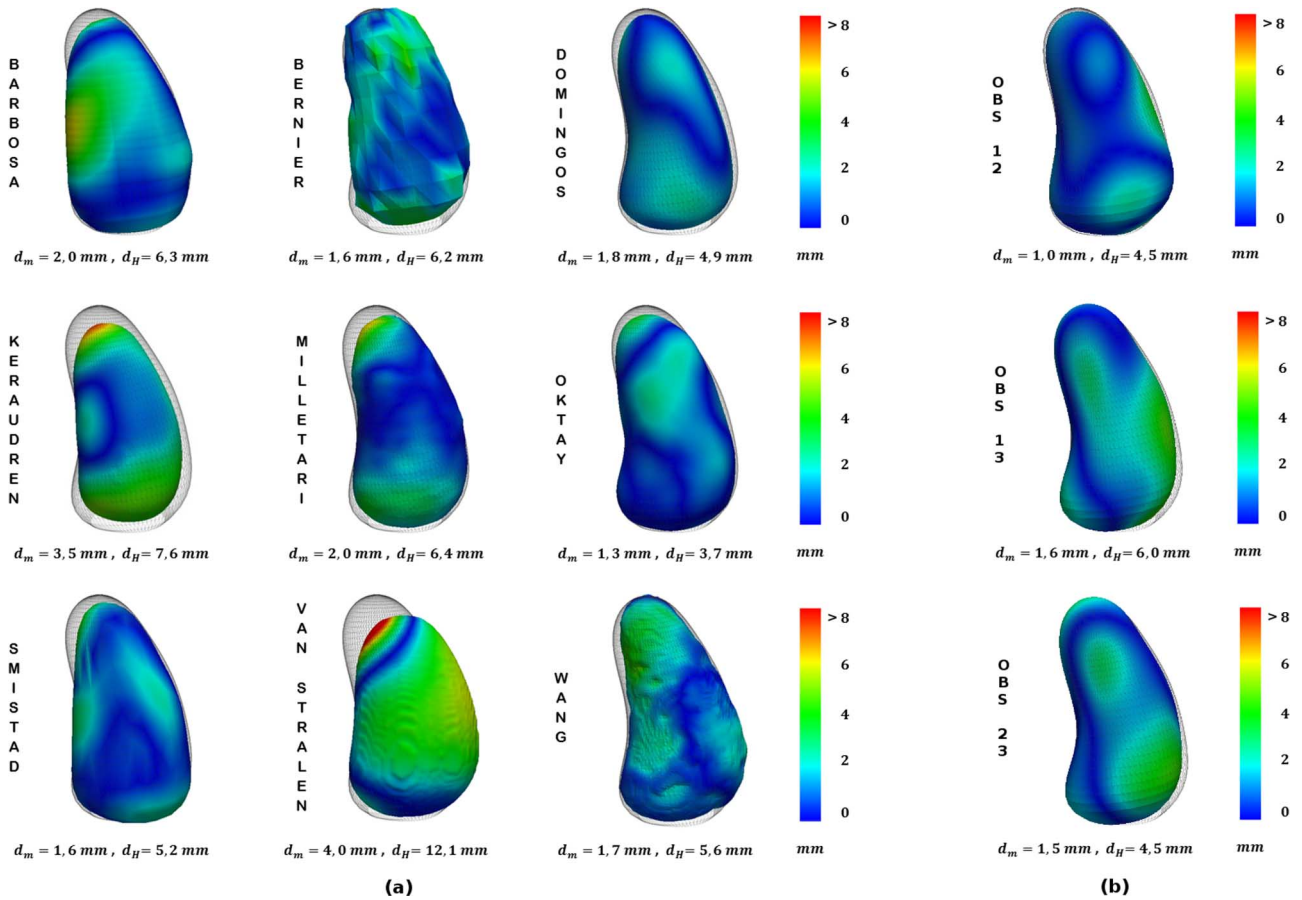


Fig. 2. Example of LV segmentation results obtained from the same particular subject at ED. The wire-frame mesh in gray corresponds to the reference standard while the colored mesh corresponds either to the segmentation result obtained by each algorithm (a) or to the surface manually drawn by each expert (b). The color map shows the distance error d_m (in mm) computed either between the participants' mesh and the reference one (a) or computed between a pair of expert meshes (b).

semi/fully automatic methods in performing LV segmentation with a high level of accuracy. Finally, it is interesting to note that the fully automatic method of Barbosa *et al.* produces among the best results in terms of mean values for $\boldsymbol{d_m}$, $\boldsymbol{d_H}$ and $\boldsymbol{D^*}$ at both ED and ES when compare to the semi-automatic results. This illustrates the strong potential for fully automatic approaches in performing competitive LV segmentation without the need of user interventions.

An example of the spatial distribution of the endocardial segmentation accuracy at ED is shown in Fig. 2[4]. The wire-frame mesh in gray corresponds to the reference standard while the colored mesh corresponds to either the segmentation result obtained by each algorithm (Fig. 2(a)) or to the surface manually drawn by each expert (Fig. 2(b)). The color map shows the distance error d_m (expressed in mm) computed either between the participants' mesh and the reference mesh (Fig. 2(a)) or between a pair of expert meshes (Fig. 2(b)). For the majority of the participants, the error is not uniformly distributed but is concentrated on some parts of the meshes. This suggests that local drops of signal or the presence of spurious structures such as papillary muscles or trabeculae which can involve strong edges, can wrongly influence algorithms locally. In this particular example, the result of Oktay *et al.* and Domingos *et al.* provide competitive results in the context of the experts' variability range both in terms of error measurement and geometries.

In order to further explore this, we provide in Fig. 3(b) the bullseye plots of the mean distance error d_m between the experts' meshes, expressed in mm, computed per AHA segment (representation of the LV which divides its anatomy into 17 predefined segments and averaged over the full testing dataset at ED (figure on the top) and ES (figure at the bottom). These plots reveal the different locations of the endocardium that are currently difficult to contour, even for expert cardiologists. From this figure, we can observe that the regions with the highest disagreement correspond mostly to the apical cap (segment 17) for both ED and ES and the anterior and anterolateral wall from the base to the apex (segments 1, 7, 13 and 6, 12, 16 respectively) at ED. This could be explained by two main reasons: i) the anterolateral papillary muscle in these segments makes endocardial definition more difficult and ii) the anterior wall is the most difficult to visualize with echo because of its unfavorable orientation with respect to the ultrasound beam and its proximity to the lung tissue.

Fig. 3(a) provides the bullseye plots of distance error computed between the reference meshes and the participants'

[4]This kind of information is provided to each user of the evaluation platform as described in the video tutorial at https://miccai.creatis.insa-lyon.fr/miccai/core/tutorial1.swf

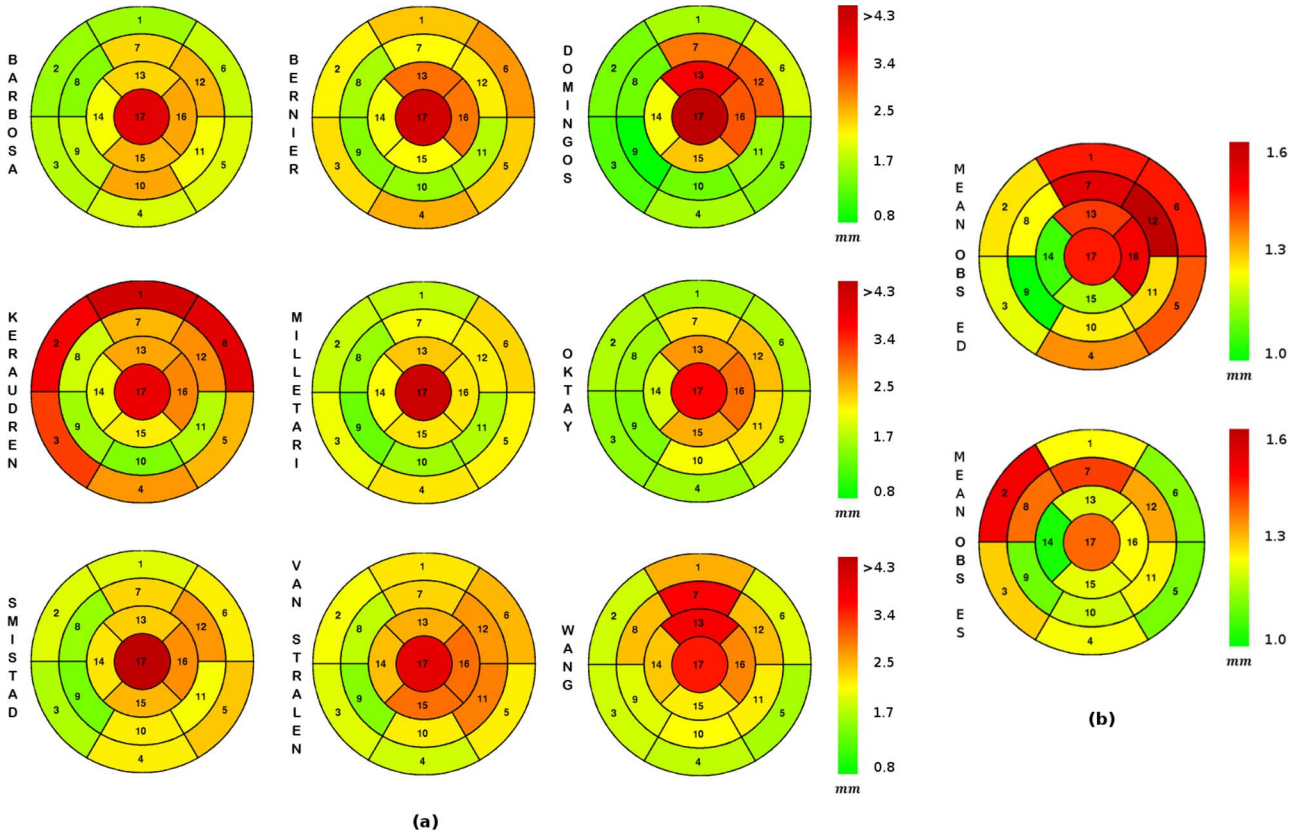


Fig. 3. Bullseye plots of distance error d_m (in mm) computed for the full testing dataset between a) the reference meshes and the participants' meshes at ED instant and b) the experts (designed as observers) at ED and ES. The scales of the colorbar in fig. a) and b) have been adjusted to the data range.

meshes averaged over the full testing dataset at ED. It should be noted that the scale of the colorbar for these plots has been adjusted to the data, so the same set of colors now covers a wider range of values. The obtained results are consistent with the observations made for Fig. 3(b). Indeed, the regions involving the highest variability between experts are the ones where the different algorithms share difficulties in performing proper segmentation, in particular near the apical region (segment 13 to 17) and the anterior and anterolateral wall at the mid-level (segments 7, 12). Finally, some methods (Keraudren *et al.* and Bernier *et al.*) make substantial segmentation errors at the basal regions (segment 1 to 6), which could be attributed to the presence of the valves and the difficulties in dealing with the open nature of the heart in this region of the volume. The same study has been performed for ES and is reported in the supplementary materials attached to this paper.

B. Clinical Indices Extraction Performance

Table IV presents an overview on the performance of the 9 methods in their ability to estimate accurate clinical indices. Results are reported on the full testing dataset at ED and ES, along with the corresponding mean inter-observer variabilities.For the clinical indices, the algorithms of Bernier *et al.* and Barbosa *et al.* globally outperform the other methods in their respective category. In particular, the method of Bernier *et al.* obtained almost the overall best results for both EDV and ESV indices while the algorithm of Barbosa *et al.* achieved the best scores for all the measures involving the EF index. In addition, the best algorithms gave results that come close to inter-observer variability values for all clinical indices. This suggests that (semi) automatic methods perform already quite similar to human observers, although a well-defined observer consensus still provides more reliable values. The differences seem larger for ES and EF than for ED. This observation can be partially explained by the high level of agreement obtained by the three cardiologists on this dataset for the ESV index which is 2 times lower than for the EDV from the same patients.

VI. Discussion

A. Evaluation Framework

In the proposed framework, 45 datasets were acquired at three different medical centers with ultrasound machines from three different vendors, ensuring that algorithms would not be biased towards specific equipment. Radio-frequency data would also have been of interest for research purposes, however the clinical equipment that we used do not offer the possibility to acquire such signal. The publicly available database[5] is thus only composed of conventional B-mode volumes. The pathologies were carefully selected in order to be as representative as possible for a majority of clinical scenarios necessitating an endocardial

[5]The patient data were provided by hospitals under conditions of restricted scientifc usage in the context of this study. The accessibility to the data is thus subject to a license agreement available at https://www.creatis.insa-lyon.fr/EvaluationPlatform/CETUS/rules.html

TABLE IV
CLINICAL INDEX ACCURACY OF THE NINE EVALUATED METHODS FOR THE FULL TESTING DATASET (30 PATIENTS). THE VALUES IN BOLD CORRESPOND TO THE BEST PERFORMANCE FOR EACH MEASURE

Method		EDV			ESV			EF		
		corr*	**bias**	**std**	**corr***	**bias**	**std**	**corr***	**bias**	**std**
Participants	category	val.	ml.	ml.	val.	ml.	ml.	val.	val.	val.
Inter-observer Mean		0.015	-3.0	11.1	0.007	-1.9	6.5	0.048	-0.1	3.3
Bernier *et al.* [28]	Semi-auto.	**0.021**	2.7	**13.9**	**0.032**	2.2	**13.7**	0.189	**0.1**	7.8
Domingos *et al.* [29]	Semi-auto.	0.083	8.7	25.0	0.044	-5.2	15.9	0.181	8.3	7.2
Oktay *et al.* [31]	Semi-auto.	0.055	-6.0	20.8	0.076	**-0.4**	20.6	0.220	-1.5	6.9
Wang *et al.* [33]	Semi-auto.	0.073	**2.0**	23.8	0.044	-3.9	16.1	**0.119**	3.5	**5.2**
Barbosa *et al.* [36]	Fully-auto.	0.035	**-5.0**	17.7	**0.033**	-6.8	**13.9**	**0.111**	**2.9**	**5.2**
Keraudren *et al.* [38]	Fully-auto.	0.079	15.9	24.6	0.048	**-6.2**	16.6	0.281	12.1	10.6
Milletari *et al.* [40]	Fully-auto.	0.047	5.1	19.0	0.040	-16.8	15.2	0.255	15.2	7.6
Smistad *et al.* [42]	Fully-auto.	0.049	-10.1	19.4	0.036	-11.3	14.6	0.121	3.7	**5.2**
van Stralen *et al.* [44]	Fully-auto.	**0.034**	-15.4	**16.0**	0.036	-13.2	14.4	0.389	3.7	8.8

segmentation (thus showing significant variation in volume and shape).

The interest of an evaluation framework critically relies on the quality of the reference. Creating reference delineations with multiple observers is a tedious and complex task. For this purpose, a dedicated manual contouring protocol for LV delineation has been specifically defined with close involvement of three expert cardiologists. Although the same guidelines were given to the experts, their annotations were not always consistent. The largest differences were usually located near the mitral valve hinge points and at the apex. Strict distance criteria were used to ensure coherence between manual contouring. A second contouring round had to be put in place for most of the treated volumes (42 of the 45 patients) to reach consensus between the experts.

B. Evaluation Results

The algorithm of Domingos *et al.* (machine learning technique) produces four of the six best segmentation scores in terms of mean value while the algorithm of Bernier *et al.* (graph cut model) obtained five of the nine best scores for the extraction of relevant clinical indices. These results are very encouraging since these techniques have only recently been applied for segmentation of RT3DE, which shows that the application of novel theory in the field of RT3DE can still be a source of improvement. For a long term perspective, this means that significant improvements are still achievable in this domain. In terms of image information extraction, it is interesting to note that one third of the methods in competition exploit machine learning techniques [29], [31], [40], among them the method of Domingos *et al.* obtained particular good results in the semi-automatic sub-category for LV segmentation. This interesting aspect is even more striking given the limited size of the training dataset (15 patients). Indeed, given that machine learning techniques generally require a large amount of cases during the training stage, one can assume that the current results obtained by this group of methods were still sub-optimal and can potentially be further improved.

The analysis of the results given in Tables III and IV confirms the necessity of evaluating segmentation accuracy and the ability to retrieve relevant clinical indices in a complementary way. For instance, the algorithm of Domingos *et al.* which provides the overall best segmentation scores in the semi-automatic category did not obtain the same best tendency in terms of clinical indices extraction. This observation can be partially explained by the fact that this method obtained the worst scores in terms of bias in its relative category which reveals a tendency of the algorithm to over-segment (respectively under-segment) the encodardial volume at ED (respectively ES). The analysis of the distribution of the errors (Fig. 3) reveals a tendency for all the methods to produce higher segmentation errors on LV regions which are associated with local drops of signal, *i.e.,* near the apical region and the anterior and anterolateral wall at the mid-level. Moreover, some methods shared difficulties in performing proper segmentation at the basal regions, which could be explained by the presence of valvular leaflets and the lack of a blood/tissue boundary in the mitral annulus.

Since echocardiography is a real-time imaging modality, the CPU time required by the different algorithms is also worth noting even though this measure was not taken into account for the challenge ranking (because of the underlying technical difficulties). In this context, the algorithm of Smistad *et al.* is the fastest method with an average runtime per segmented volume of 65 milliseconds. The chosen strategy of representing the surface by a dedicated model involving few parameters combined with an extended Kalman filtering to guide the surface deformation [42] allowed the authors to achieve near real time processing with competitive segmentation accuracy. The rest of the algorithms could then be categorized into three groups: methods with a processing time around few seconds [28], [29], [36], methods with a processing time around a minute [33], [38], [40], [44] or tens of minutes [31].

Based on the score values provided in Tables III and IV, it appears that the fully automatic algorithm of Barbosa *et al.*

(explicit deformable model) produced very competitive results for both LV segmentation and clinical indices extraction, illustrating the strong potential of fully automatic method to be deployed in a clinical environment. Although the current best semi/fully automated algorithms produce promising results when compared to the expert's variability, there still exists room for improvement. Indeed, performance of the best algorithms for the segmentation task is close to but still below the agreement of the expert cardiologists (about 1.6 times worse in average) while the same conclusions can be made for the clinical indices (e.g., 1.2, 2.1 and 2.3 times worse for EDV, ESV and EF standard deviation measures, respectively). Finally, the evaluation of the 9 different algorithms with the standardized evaluation framework provides also useful directions for further investigations. For example, it may be interesting to investigate the robustness of the segmentation methods with respect to their manual/automatic initialization. Also, it would be interesting to investigate whether a combination of algorithms outperforms the current best scores.

C. Study Limitations

The quality of an evaluation framework critically relies on the number of datasets that are made available for training and testing. In this study, 45 datasets were acquired at three different medical centers. Although this limited number of patients can be seen as an issue, it has to be pointed out that it required already more than 10 months to generate the manual references from 3 different cardiologists including consensus revisions. To the best of our knowledge, this is the first time such an echocardiographic database is made publicly available to the community.

The segmentation quality has been investigated in this study through the use of standard distance measurements, *i.e.,* mean surface distance, Hausdorff surface distance and Modified Dice similarity index. Although other indices could have been used, the chosen metrics allow a direct comparison of the results presented in this work with most of the previous studies performed in the same field, as shown in Table I.

Finally, as detailed in Section III-A2, we have defined and used a restrictive acquisition protocol. In particular images have been excluded because of poor visualization of a LV wall or wall out of the image sector to an extent that the image can no longer be manually analyzed with good confidence. Although dropout occurs regularly in ultrasound images and segmentation strategies that are robust to it are of research interest, the diagnostic value of such poor quality images is questionable. More importantly, the purpose of the evaluation platform is to compare semi/fully automatic segmentation results with manual references for clinical indices extraction based on volume. Because the accuracy of the manually extracted volume is directly linked to the visibility of the LV cavity in the image, we decided to restrict the data to those images which allow assessing those measures with enough accuracy.

VII. Conclusions

A publicly available standardized evaluation framework to compare the performance of endocardial segmentation techniques in RT3DE was presented in this article. The results showed that the current best algorithms produce promising results with respect to the experts' measurements for the extraction of clinical indices, and that they offer good segmentation precision in terms of mean distance error in the context of the experts' variability range after consensus agreement. Although these results are very encouraging, they also reveal that there still exists room for improvement. The evaluation framework remains open for new submissions at https://www.creatis.insa-lyon.fr/EvaluationPlatform/CETUS/.

References

[1] H. D. White *et al.*, "Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction," *Circulation*, vol. 76, pp. 44–51, 1987.
[2] R. Norris, H. White, D. Cross, C. Wild, and R. Whitlock, "Prognosis after recovery from myocardial infarction: The relative importance of cardiac dilatation and coronary stenoses," *Eur. Heart J.*, vol. 13, pp. 1611–1618, 1992.
[3] V. Mor-Avi *et al.*, "Real-time 3-dimensional echocardiographic quantification of left ventricular volumes: Multicenter study for validation with magnetic resonance imaging and investigation of sources of error," *J. Am. College Cardiol.*, vol. 1, no. 4, pp. 413–423, 2008.
[4] D. Muraru *et al.*, "Validation of a novel automated border-detection algorithm for rapid and accurate quantitation of left ventricular volumes based on three-dimensional echocardiography," *Eur. J. Echocardiogr.*, vol. 11, no. 4, pp. 359–368, 2010.
[5] R. M. Lang *et al.*, "EAE/ASE recommendations for image acquisition and display using three-dimensional echocardiography," *Eur. Heart J.—Cardiovasc. Imag.*, vol. 13, no. 1, pp. 1–46, Jan. 2012.
[6] K. Leung and J. Bosch, "Automated border detection in three-dimensional echocardiography: Principles and promises," *Eur. J. Echocardiogr.*, vol. 11, no. 2, pp. 97–108, 2010.
[7] O. Bernard *et al.*, "Challenge on endocardial three-dimensional ultrasound segmentation (CETUS)," *Proc. MICCAI CETUS*, pp. 1–8, 2014.
[8] C. Corsi, G. Saracino, A. Sarti, and C. Lamberti, "Left ventricular volume estimation for real-time three-dimensional echocardiography," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1202–1208, Sep. 2002.
[9] E. Angelini, S. Homma, G. Pearson, J. Holmes, and A. Laine, "Segmentation of real-time three-dimensional ultrasound for quantification of ventricular function: A clinical study on right and left ventricles," *Ultrasound Med. Biol.*, vol. 31, no. 9, pp. 1143–1158, 2005.
[10] Q. Duan *et al.*, "Region-based endocardium tracking on real-time three-dimensional ultrasound," *Ultrasound Med. Biol.*, vol. 35, no. 2, pp. 256–265, 2009.
[11] D. Barbosa *et al.*, "Fast and fully automatic 3D echocardiographic segmentation using b-spline explicit active surfaces: Feasibility study and validation in a clinical setting," *Ultrasound Med. Biol.*, vol. 39, no. 1, pp. 89–101, 2013.
[12] M. van Stralen *et al.*, "Automatic segmentation of the left ventricle in 3d echocardiography using active appearance models," in *IEEE Ultrason. Symp.*, 2007, pp. 1480–1483.
[13] F. Orderud and S. Rabben, "Real-time 3D segmentation of the left ventricle using deformable subdivision surfaces," in *Comput. Vis. Pattern Recognit.*, 2008.
[14] R. V. Stebbing and J. A. Noble, "Delineating anatomical boundaries using the boundary fragment model," *Med. Image Anal.*, vol. 17, no. 8, pp. 1123–1136, 2013.
[15] R. Juang, E. McVeigh, B. Hoffmann, D. Yuh, and P. Burlina, "Automatic segmentation of the left-ventricular cavity and atrium in 3D ultrasound using graph cuts and the radial symmetry transform," in *IEEE Int. Symp. Biomed. Imag.*, 2011, pp. 606–609.
[16] X. Zhuang *et al.*, "Registration-based propagation for whole heart segmentation from compounded 3D echocardiography," in *IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 1093–1096.
[17] V. Lempitsky, M. Verhoek, J. Noble, and A. Blake, "Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography," in *Funct. Imag. Model. Heart*, 2009, pp. 447–456.
[18] L. Yang *et al.*, "Prediction based collaborative trackers (PCT): A robust and accurate approach toward 3D medical object tracking," *IEEE Trans. Med. Imag.*, vol. 30, no. 11, pp. 1921–1932, Nov. 2011.
[19] J. Hansegard, F. Orderud, and S. Rabben, "Real-time active shape models for segmentation of 3D cardiac ultrasound," in *Comput. Anal. Images Patterns*, 2007, pp. 157–164.

[20] K. Y. Leung, M. van Stralen, G. van Burken, N. de Jong, and J. G. Bosch, "Automatic active appearance model segmentation of 3D echocardiograms," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 320–323.

[21] Q. Duan, E. D. Angelini, and A. F. Laine, "Real-time segmentation by active geometric functions," *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 223–230, 2010.

[22] K. Y. Leung *et al.*, "Left ventricular border tracking using cardiac motion models and optical flow," *Ultrasound Med. Biol.*, vol. 37, no. 4, pp. 605–616, 2011.

[23] K. Rajpoot, V. Grau, J. A. Noble, H. Becher, and C. Szmigielski, "The evaluation of single-view and multi-view fusion 3D echocardiography using image-driven segmentation and tracking," *Med. Image Anal.*, vol. 15, no. 4, pp. 514–528, 2011.

[24] E. Dikici and F. Orderud, "Generalized step criterion edge detectors for Kalman filter based left ventricle tracking in 3D + T echocardiography," *Stat. Atlases Computat. Models Heart Imag. Model. Challenges*, pp. 261–269, 2012.

[25] A. Papachristidis *et al.*, "Clinical expert delineation of 3d left ventricular echocardiograms for the CETUS segmentation challenge," in *Proc. MICCAI CETUS*, Sep. 14, 2014, pp. 9–16.

[26] B. Heyde, D. Barbosa, P. Claus, F. Maes, and J. D'hooge, "Three-dimensional cardiac motion estimation based on non-rigid image registration using a novel transformation model adapted to the heart," in *Proc STACOM*, 2012, vol. 7746, LNCS.

[27] S. Gottschalk, M. C. Lin, and D. Manocha, "Obbtree: A hierarchical structure for rapid interference detection," in *Proc. of the 23rd Annu. Conf. Comput. Graph. Interactive Tech.*, 1996, pp. 171–180.

[28] M. Bernier, P. Jodoin, and A. Lalande, "Automatized evaluation of the left ventricular ejection fraction from echocardiographic images using graph cut," *Proc. MICCAI CETUS*, pp. 25–32, 2014.

[29] J. Domingos, R. Stebbing, and J. Noble, "Endocardial segmentation using structured random forests in 3D echocardiography," in *Proc. MICCAI CETUS*, 2014, pp. 33–40.

[30] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *Int. Conf. Comput. Vis.*, 2013, pp. 1841–1848.

[31] O. Oktay, W. Shi, K. Keraudren, J. Caballero, and D. Rueckert, "Learning shape representations for multi-atlas endocardium segmentation in 3D echo images," in *Proc. MICCAI CETUS.*, 2014, pp. 57–64.

[32] O. Oktay, W. Shi, J. Caballero, K. Keraudren, and D. Rueckert, "Sparsity based spectral embedding: Application to multi-atlas echocardiography segmentation," in *Proc. 2nd MICCAI Workshop Sparsity Tech. Med. Imag.*, 2014.

[33] C. Wang and O. Smedby, "Model-based left ventricle segmentation in 3d ultrasound using phase image," in *Proc. MICCAI CETUS*, 2014, pp. 81–88.

[34] G. Lathen, J. Jonasson, and M. Borga, "Blood vessel segmentation using multi-scale quadrature filtering," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 762–767, 2010.

[35] C. Wang and O. Smedby, "Automatic multi-organ segmentation in non-enhanced CT datasets using hierarchical shape priors," presented at the Int. Conf. Pattern Recognit., Stockholm, 2014.

[36] D. Barbosa, D. Friboulet, J. D'hooge, and O. Bernard, "Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching," in *Proc. MICCAI CETUS*, 2014, pp. 17–24.

[37] D. Barbosa *et al.*, "B-spline explicit active surfaces: An efficient framework for real-time 3D region-based segmentation," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 241–251, 2012.

[38] K. Keraudren, O. Oktay, W. Shi, J. Hajnal, and D. Rueckert, "Endocardial 3d ultrasound segmentation using autocontext random forests," in *Proc. MICCAI CETUS*, 2014, pp. 41–48.

[39] Z. Tu, "Auto-context and its application to high-level vision tasks," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[40] F. Milletari, M. Yigitsoy, N. Navab, and S. Ahmadi, "Left ventricle segmentation in cardiac ultrasound using Hough-forests with implicit shape and appearance priors," in *Proc. MICCAI CETUS*, 2014, pp. 49–56.

[41] K. Rematas and B. Leibe, "Efficient object detection and segmentation with a cascaded Hough forest ISM," in *Int. Conf. Comput. Vis.*, 2011.

[42] E. Smistad and F. Lindseth, "Real-time tracking of the left ventricle in 3d ultrasound using Kalman filter and mean value coordinates," in *Proc. MICCAI CETUS*, 2014, pp. 65–72.

[43] S. I. Rabben *et al.*, "Semiautomatic contour detection in ultrasound m-mode images," *Ultrasound Med. Biol.*, vol. 26, no. 2, pp. 287–296, 2000.

[44] M. van Stralen, A. Haak, K. Leung, G. van Burken, and J. Bosch, "Segmentation of multi-center 3d left ventricular echocardiograms by active appearance models," in *Proc. MICCAI CETUS*, 2014, pp. 73–80.

[45] O. Ecabert *et al.*, "Automatic model-based segmentation of the heart in CT images," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1189–1201, Sep. 2008.