IEEE Transactions on Medical Imaging

Detailed Evaluation of Five 3D Speckle Tracking Algorithms using Synthetic Echocardiographic Recordings

Journal:	*Transactions on Medical Imaging*
Manuscript ID	Draft
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Alessandrini, Martino; KULeuven, Cardiovascular Imaging and Dynamics Heyde, Brecht; University of Leuven, Cardiovascular Imaging and Dynamics Queirós, Sandro; University of Minho, ICVS/3Bs - PT Government Associate Laboratory; University of Minho, Algoritmi Center; KU Leuven, Cardiovascular Imaging and Dynamics Cygan, Szymon; University of Technology, Warsaw, Institute for Metrology and Biomedical Engineering Zontak, Maria; University of Washington, Department of Bioengineering Somphone, Oudom; Philips Paris, Medisys Bernard, Olivier; Institut National des Sciences Appliquées, Lyon, CREATIS-LRMN De Craene, Mathieu; Philips, Research O'Donnell, Matthew; University of Washington, Bioengineering D'hooge, Jan; Catholic University of Leuven, Imaging and Cardiovascular Dynamics
Keywords:	Ultrasound < Imaging modalities, Heart < Object of interest, Biomechanical modeling < General methodology, Evaluation and Performance < General methodology, Motion compensation and analysis < General methodology, Quantification and estimation < General methodology, Validation < General methodology
Specialty/Area of Expertise:	Echocardiography, Deformation Imaging, Speckle Tracking, Cardiac Strain, Ultrasound Simulation, Cardiac Modeling

Jan 27th 2016

Dear Editor and Reviewers,

We would like to thank you for your thorough review and the useful comments and suggestions formulated about our paper “Detailed Evaluation of Five 3D Speckle Tracking Algorithms using Synthetic Echocardiographic Recordings.” submitted to IEEE Transactions on Medical Imaging.

The remainder of this letter contains our answers to your reviews. A revised version of the manuscript follows, where modified parts appear in blue for a better identification. The authors hope this will simplify the work of the editor and reviewers.

We thank you again for your valuable review, and we remain available for any further information you may need.

The detailed point-to-point response letter to the reviewers’ comments follows.

Sincerely yours,

the authors.

Reviewer 1

This manuscript describes a detailed comparison of 5 different types of algorithms for assessing cardiac motion estimation. The authors have exerted great effort to introduce a variety of different methods straight from the source rather than making their own implementations. This is attractive because it somewhat voids the critique of failing to implement the algorithms correctly. Additionally, the authors implement their selected methods on realistic simulated data, which provides a real truth value for robust comparison. On the whole, this manuscript is well written and logical, and I have few real critiques. My only significant critique is the real relevance of the work. The authors briefly cite one of their review papers to indicate that these methods have some clinical diagnostic potential, but in reality these methods have been around for a long time in both MR and US and still fail to be used clinically. Despite the authors' extensive efforts, it's still unclear whether this manuscript contributes anything that will really drive the field forward. Based on the existing literature combined with current clinical utilization, arguments for and against cardiac strain based metrics can still be made either way. (In this case by clinical utilization, I refer to clinical practice rather than large patient focused studies.) I think it would be useful for the field if the authors' spent at least a little more space justifying their extensive well-considered efforts.

General Comments:

***Q1**: My primary concern is expressed above. It would be useful to make a stronger argument for the work.*

Answer: We have expanded the introduction by presenting the echocardiographic strain technology more extensively. In particular, we tried to make clearer where we stand in terms of use of STE in the clinical practice. In particular, we emphasize 2 things that show that the strain imaging has matured into a modality that is used in clinical routine (and not only in clinical studies/publications):

1. The guidelines on monitoring the cardiotoxic effect of chemotherapy include strain measurements.
2. Global longitudinal strain (GLS) has now been demonstrated by multiple studies to be more prognostic than ejection fraction for future cardiac events in different patients populations, as shown in the recent meta-analysis done by Kalam et al. (ref. [4] in the manuscript). In the conclusion, the authors explicitly point out that GLS outprforms EF but that it will take time to have this migrate to clinical practice for historical reasons.

***Q2**: The manuscript focuses on using data that can be obtained using standard clinical technology. In this regard, it's unclear why RF based methods are included since it seems that these methods cannot be applied to data acquired on current clinical systems, and RF methods are otherwise doomed to fail on the frame rates used in this study. The authors do mention this known shortcoming of RF methods, but specifically, I question the relevance of RF methods in this manuscript at all given the stated motivation of the study. What would be more compelling is if RF methods were employed with the necessary frame* *rates so it could be determined whether RF method are still worth pursuing for the particular problem of cardiac strain imaging.*

Answer: We agree this is a natural concern. As we discussed in the paper, RF poses additional methodological and technical challenges as compared to B-mode tracking. Therefore, it remains unclear whether the technique can be applied to 3D echo data at clinical frame rates. Yet, there are strong ongoing efforts in this directions. In [1] RF tracking was applied to 3D echo data acquired on children at 38-51 Hz. In [2] and [3], it was applied to open-chest dog's scans at 51-56 Hz. In this sense, it is also noteworthy mentioning that, when presenting our simulation pipeline in conferences [4], we were asked several times to make RF data available. Finally, the fact that a well respected team with longstanding expertise in the field of cardiac strain imaging accepted to participate to this study, is in itself a measure of the interest and the expectations behind the RF tracking technology. As such, RF tracking is an important actor in the present research scenario which, we believe, deserves to be represented in this study.

[1] Richard G.P. Lopata, Maartje M. Nillesen, Johan M. Thijssen, L. Kapusta, Chris L. de Korte, Three-Dimensional Cardiac Strain Imaging in Healthy Children Using RF-Data, Ultrasound in Medicine & Biology, Volume 37, Issue 9, September 2011, Pages 1399-1408;

[2] Compas, C.B.; Wong, E.Y.; Xiaojie Huang; Sampath, S.; Lin, B.A.; Pal, P.; Papademetris, X.; Thiele, K.; Dione, D.P.; Stacy, M.; Staib, L.H.; Sinusas, A.J.; O'Donnell, M.; Duncan, J.S., "Radial Basis Functions for Combining Shape and Speckle Tracking in 4D Echocardiography," in Medical Imaging, IEEE Transactions on , vol.33, no.6, pp.1275-1289, June 2014;

[3] Congxian Jia; Kolias, T.J.; Rubin, J.M.; Ping Yan; Sinusas, A.J.; Dione, D.P.; Duncan, J.S.; Qifeng Wei; Thiele, K.; Lingyun Huang; Sheng-Wen Huang; O'Donnell, M., "3D elasticity imaging on an open-chest dog heart," in Ultrasonics Symposium (IUS), 2009 IEEE International , vol., no., pp.155-158, 20-23 Sept. 2009

[4] M. Alessandrini, M. De Craene, O. Bernard, S. Giffard-Roisin, P. Allain, I. Waechter-Stehle, J. Weese, E. Saloux, H. Delingette, M. Sermesant, and J. D'hooge, "A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database," Medical Imaging, IEEE Transactions on, vol. 34, no. 7, pp. 1436–1451, July 2015.

***Q3**: In a number of places the authors use citations as nouns. Under most style conventions this is considered inappropriate, but I do see this in a lot of papers submitted to transactions on medical imaging so it may be considered acceptable for this journal. The authors' might check with the journal's style guide or one of the editors.*

Answer: We thank the reviewer for bringing our attention on this. Yet, we haven't found any explicit guidelines against citing papers by the name of the first author. In this paper, we cite by name only those papers describing the algorithms involved in the comparison. This is done to make immediately clear to the reader which author contributed which algorithm.

Specific Comments:

Q3: *External data: It would be useful if the external data included some kind of readme file.*

Answer: As recommended, we added a readme.txt to illustrate the content of the folder. The new supporting material can now be downloaded at http://bit.ly/1QnUEKm

Q4: *Page 1, Line 48, 2nd column: Without a citation it's unclear that this statement about commercial systems is true. In this reviewer's experience at least some (maybe 2 or3) of the top commercial vendors do not use block matching, or at least use a hybrid technique that relies significantly on non block matching data. I think a citation would be important to really make this claim.*

Answer: This is true. We have smoothed that statement and added a few references (cf. pp. 3 first column, text in blue).

Q5: *Page 4, Line 17, 1st column (***now pp. 4, second col., text in blue***): This particularly error metric doesn't seem adequately justified. Typically, without some additional justification mean squared error (or root mean squared error) is a less informative (meant in the theoretical sense) error metric due to the fact that assuming Gaussian statistics is a theoretically less informative decision than imposing error metrics associated with other error distributions.*

Answer: The expression the reviewer is referring to measures the relative distance between two curves, i.e. the computed and the reference strain. We normalize it by the reference in order to make it less dependent on i) the strain amplitude variation over time and ii) the strain direction. We believe this is an appropriate way to show how accurate is an algorithm in average. We used similar measurements previously in [1, 2] below. We also agree that the average error could be complemented, e.g., by the maximum distance between the two curves. Anyway we don't believe this would bring additional insights on one's algorithm performance. Moreover, we note that this kind of 'punctual' information is also provided by the time-to-peak measurements (cf. Fig. 10). Time-to-peak differences have been proposed as a clinical marker to assess dyssynchrony.

[1] Heyde, B.; Alessandrini, M.; Hermans, J.; Barbosa, D.; Claus, P.; D'hooge, J., "Anatomical image registration using volume conservation to assess cardiac deformation from 3D ultrasound recordings," in Medical Imaging, IEEE Transactions on , in press.

[2] M. Alessandrini, A. Basarab, L. Boussel, X. Guo, A. Serusclat, D. Friboulet, D. Kouame, O. Bernard, and H. Liebgott, "A new technique for the estimation of cardiac motion in echocardiography based on transverse oscillations: A preliminary evaluation in silico and a feasibility demonstration in vivo," Medical Imaging, IEEE Transactions on, vol. 33, no. 5, pp. 1148–1162, May 2014.

Q6: *Figure 3: This figure does not contain a dependent axis label, and based on the caption it looks like a subplot might be missing.*

Answer: The y-axis has been added and the caption has been updated.

Q7: *Figure 4: It seems like the title would be more appropriate as the dependent axis label, which is missing.*

Answer: The correction has been made.

Q8: *Figure 5 (***now Fig. 6***): Could you please clarify what is mean by 'outliers' numbers reduced of a factor of 10'. Does this mean the outlier values were divided by 10 or the population of outliers was decimated.*

Answer: We made it now clear that the population was decimated.

Q9: *Figure 6 and corresponding results: It is strictly speaking considered incorrect or at least unnecessary to use a Bland-Altman plot/analysis when there is an actual truth value. (I realize that in some literature it is used this way anyway.) Effectively though it doesn't show anything different from what is shown in the correlation plots.*

Answer: We agree. We keep on reporting bias and limits of agreement but we don't refer to Bland Altman analysis anymore. The old 'Bland Altman' plots were therefore removed.

Reviewer 2

The manuscript "Detailed evaluation of five 3D speckle tracking algorithms using synthetic echocardiographic recordings" by Alessandrini and coll. proposes a shared database of realistic US synthetic cardiac phantoms for benchmarking 3D speckle tracking techniques based on the most used methods for cardiac deformation quantification. Five different speckle tracking techniques were tested on 7 US sequences corresponding to different pathological conditions. The performances were evaluated considering tracking and strain accuracy. Results for tracking accuracy show AFFD, AAOF and S-Demon perform similarly and better than BM and RFBM. AFFD confirmed high accuracy for strain measurements, irrespective of strain directions.

Overall the paper is clearly organized and well written. The description of the considered algorithm is short but they are well known and, for each one, the parameter values are reported. Statistics for algorithm comparison seems appropriate and the discussion of the results is comprehensive.

My main concern regards methodological and innovative contents of the paper. The paper does not provide any innovative methodological contribution and it overlaps in some parts with previously published paper (ref.s 15-17) therefore publication on IEEE TMI journal is questionable. However, a shared realistic synthetic dataset available for benchmarking new algorithms is of great interest and a direct comparison of new algorithms on the same dataset is also a common need. In this regards, this paper is a step forward towards a common platform for new algorithm testing. In addition, STE represents a promising technique in the echocardiographic field and a deeper insight to better understand the performance of the state-of-art 3D STE techniques is of great interest.

Q1: *I suggest to include some realistic data corresponding to physiological conditions in the synthetic database;*

Answer: We've followed this suggestion and added a physiological simulation to the synthetic dataset. Correspondingly, the results section has been updated by including the tracking results on the healthy sequence. The bull's eye plots for the healthy sequence have been added in Fig. 9(a).

Q2: *considering many commercial packages based on the BM approach include spatial/temporal smoothing, my suggestion is to add the fully regularized block matching algorithm in the present paper;*

Answer: with respect to commercial software packages, it was not feasible to include them in this study. Yet, a comparison between commercial products is the topic of an ongoing study. We also agree the lack of smoothing in the BM algorithm was one of the main limitations of the study.

After this revision, we thus asked the BM contributing team to possibly include a spatio-temporal smoothing step. As mentioned in Sec. III-D BM updated instead their original submission by adding both spatial and temporal smoothing. Nonetheless, after benchmarking spatial smoothing was not found beneficial and was therefore switched off in the final submission.

Concerning RFBM, we did not mention in the paper that the original submission already included a spatial regularization of the displacement estimates. This is now made explicit in Sec. III-E and in Table I. Moreover, as mentioned in the discussion (Sect. VI-D), RFBM would introduce smoothing at the strain computation step by fitting an affine transform to the displacement estimates. However we decided to maintain strain computation consistent between participants in order to avoid further sources of variability, besides being consistent with the recent clinical guidelines.

Q3: *the tracking accuracy should be evaluated also on a regional segment basis.*

Answer: We thank the reviewer for this suggestion. We have added a regional evaluation of tracking accuracy, where we measure variations with respect to i) myocardial layer (i.e. endocardium vs. mid myocardium vs. epicardium), ii) myocardial level (i.e. basal segments vs. middle segments vs. apical segments) and iii) functional region (i.e. healthy segments vs. ischemic segments). The new analysis is presented in Sect. IV.A.2.

Reviewer 3

General comment*: This is overall a well-defined and well-written study that compares the accuracy of several methods for 3D tracking of myocardial deformation using relevant metrics and statistical analysis for the given context. The methods included come from different groups which is a strength in itself, but one can also speculate whether equal effort was put into the optimization of each approach. This is especially apparent for the block matching approaches which seem relatively crude. The end results are thus partly a bit difficult to assess in absolute terms, but still of high value and interest. The manuscript deserves to be published in my opinion, but I do have some comments that I would like to see addressed first.*

DATA:

__Q1__: There is very little discussion about the potential limitations of the simulation approach. How does it differ from reality and how could this limit the evaluation?

Answer: The discussions in the first submission were indeed shrunk to the essential in order to meet the 10 pages limit. We have expanded the discussions in the revised manuscript and addressed to our best the important points raised by the reviewer. In particular, we address explicitly limitations related to the ultrasound data, the motion model, the error metrics used and the algorithms considered (cf. Sect. VI).

__Q2__: Using a model implies simplifications. Could the model used favor some of the methods presented? For instance, could it be the case that the underlying simplifications of the model itself (tissue model / motion model) fits very well with the underlying assumptions of the global regularization routines included in some of the methods? Would this differ in real-life?

Answer: This point is now addressed in the discussions (Sect. VI-B).

***Q3:** Some more discussion about data limitations would improve the manuscript.*

Answer: Please see answer to Q1 above.

METHODS:

***Q4:** The methods presented involve different similarity metrics and different kinds and amounts of regularization / smoothing. An improved presentation of the different methods would strengthen the manuscript, for instance by including a more detailed taxonomy with regards to similarity measure, variational (global) vs local approach, regularization / smoothing, etc.*

Answer: As recommended, we have worked out a table summarizing the principal features of each algorithm (cf. Table 1).

***Q5:** The methods are said to be state of the art, but this should be further explained. What makes them state of the art? Given that the commercial implementations are not in the comparison study this is perhaps a bit premature, even though some of the algorithms compared may be more sophisticated and have proven valuable in other settings such as image registration.*

Answer: We have expanded on the state of development/evaluation of each technique in Section III (cf. text in blue). We have also eliminated the term ‘state-of-the-art’. We stress that, although these are not commercial implementations, at least 4 of the techniques are the output of long-lasting research project from well-established teams with leading expertise in the field of cardiac strain imaging. We acknowledge the only exception could be for the BM algorithm. Indeed, the considered implementation was not evaluated previously, with the exception of the small preliminary study in [1]. Nonetheless, BM is a very well established principle and the implementation considered followed a thorough literature overview. The quality of the implementation is also proved by the results presented in this study.

[1] Alessandrini, M.; Heyde, B.; Cygan, S.; Sermesant, M.; Delingette, H.; Bernard, O.; De Craene, M.; D'hooge, J., "Elastic registration vs. block matching for quantification of cardiac function with 3D

ultrasound: Initial results of a direct comparison in silico based on a new evaluation pipeline," in Ultrasonics Symposium (IUS), 2014 IEEE International , vol., no., pp.608-611, 3-6 Sept. 2014.

Q6: *Related to the pros and cons of the different approaches, I would for instance look more into the trade-off between regularization and accuracy in time and space. Smoothing / overfitting is a concern for 3D strain.*

Answer: This is true, we have elaborated further on this point in the discussions, where this is explicitly stated as a limitation of the study (Sect. VI-B). Looking into this compromise would imply including more datasets with different degree of ischemia and performing a sensitivity analysis of the hyper-parameters balancing the regularization. This could be the topic for future studies.

Q7: *What about the influence of frame rate? Given the model it would be interesting to see when methods break down for lower frame rates and improve differently for higher frame rates. At the minimum a paragraph in the discussion should be included.*

Answer: We realize the information on the frame rate was missing in the previous submission. This has been added now (cf. Sec. II.A), in particular all datasets were simulated at a temporal resolution of 34 frames/s. Temporal resolution is easily adjustable in the e/m model. The value chosen was considered a reasonable average of what available clinically. We agree evaluating the influence of frame rate is important and we commented on it in the discussion. Yet, this kind of evaluation remained beyond the scope of this specific study. We have initiated an effort in this direction in [1]

[1] Alessandrini, M.; Heyde, B.; Ling Tong; Bernard, O.; D'hooge, J., "Tracking quality in plane-wave versus conventional cardiac ultrasound: A preliminary evaluation in-silico based on a state-of-the-art simulation pipeline," in Ultrasonics Symposium (IUS), 2015 IEEE International , vol., no., pp.1-4, 21-24 Oct. 2015.

RESULTS:

Q8*: The results from block matching was actually quite encouraging considering that no spatial or temporal smoothing was involved. Computationally, this algorithm is further highly parallelizable. If possible I would also include a commercial implementation, this could for instance be a vendor independent software like the one from Tomtec. But I understand if this is for further work.*

Answer: We agree block matching was a very close competitor of methodologically more elaborated techniques. In the revised manuscript, BM was further improved by including temporal and spatial smoothing (see answer 2 to Reviewer 2). Nonetheless, the implementation of spatial smoothing was not found beneficial after benchmarking and, therefore, it was switched off in the final submission. Yet. Temporal smoothing help improving accuracy further (cf. result section). Concerning parallelizability, please see the answer to Q11.

Benchmarking commercial 3D STE solutions is a priority in our ongoing efforts. Indeed, they were left out of this study since:

- Including commercial solutions would require taking care of several practical issues such as the access to the software package, data conversion to an appropriate (possibly proprietary) file format, making sure the software is used by expert users;
- Given the amount of new solutions for cardiac strain imaging which are constantly published, we believe providing a (yet limited) snapshot of what available in research was still relevant and could deserve a dedicated publication;

In this regard, it's also noteworthy mentioning we are currently in the process of setting up a comparison study between commercial packages for what concerns 2D strain, by using an extension of the pipeline considered in this paper. This has been added to the paper (cf. end of Sect. VI).

Q10*: There are a lot of results, if needed I would reduce the amount of results to make room for more analysis / discussion as mentioned previously.*

Answer: As recommend, and due, we have expanded the discussion section trying to address all the points raised. Moreover, as suggested in Reviewer 1 (Q9), we have dropped the Bland-Altman analysis.

Q11: *With regards to computation time, it is quite possible to evaluate this from an algorithm standpoint (e.g. complexity / flops etc.). And in this setting it would also be interesting to note particular implementation details such as parallelization. Some techniques such as block matching are local and can easily be implemented in parallel for execution of GPUs. Global variational approaches based on iterative optimization solvers are in this sense typically less parallelizable.*

Answer: we expanded the section dedicated to computational complexity (cf. Sect. IV.C) and added a paragraph on computational complexity to the conclusions (cf. Sect. VII).

Discussion / conclusion

Q12: *I agree with the conclusion, but I find it is too bad that the block matching algorithms are not at the level of sophistication with regards to its implementation as the other techniques.*

Answer: We added some more detail on the level of development/evaluation of each algorithm in Sec. III. For what concerns the degree of sophistication, we agree the lack of regularization represents a gap with respect to what available commercially. For what concerns RFBM we don't believe this comes from poor sophistication. The technique is indeed the output of several years of research in the developing team and the same implementation was recently used in [1]. One of the reasons why the technique suffered in this comparison is the way strain was computed. As mentioned in the discussions (Sec. VI.3), the technique as implemented would introduce smoothing at the strain computation stage by using affine least squares fitting of the displacement fields. For what concerns B-mode BM we agree the absence of smoothing was a limitation of the study. We asked therefore the contributing team to include a (spatial and temporal) regularization step. The revised manuscript considers therefore an updated BM implementation inclusive of regularization (cf. Sec. III). Of note, spatial smoothing was not found beneficial after benchmarking and was therefore switched off in the final submission.

[1] Congxian Jia; Kolias, T.J.; Rubin, J.M.; Ping Yan; Sinusas, A.J.; Dione, D.P.; Duncan, J.S.; Qifeng Wei; Thiele, K.; Lingyun Huang; Sheng-Wen Huang; O'Donnell, M., "3D elasticity imaging on an open-chest dog heart," in Ultrasonics Symposium (IUS), 2009 IEEE International , vol., no., pp.155-158, 20-23 Sept. 2009.

Detailed Evaluation of Five 3D Speckle Tracking Algorithms using Synthetic Echocardiographic Recordings

Martino Alessandrini*, Brecht Heyde, Sandro Queirós, Szymon Cygan, Maria Zontak, Oudom Somphone, Olivier Bernard, Maxime Sermesant, Hervé Delingette, Daniel Barbosa, Mathieu De Craene, Matthew O'Donnell and Jan D'hooge

***Abstract*—A plethora of techniques for cardiac deformation imaging with 3D ultrasound, typically referred to as 3D speckle tracking techniques, are available from academia and industry. Although the benefits of single methods over alternative ones have been reported in separate publications, the intrinsic differences in the data and definitions used makes it hard to compare the relative performance of different solutions. To address this issue, we have recently proposed a framework to simulate realistic 3D echocardiographic recordings and used it to generate a common set of ground-truth data for 3D speckle tracking algorithms, which was made available online.**

The aim of this study was therefore to use the newly developed database to contrast non-commercial speckle tracking solutions from research groups with leading expertise in the field. The five techniques involved cover the most representative families of existing approaches, namely block-matching, radio-frequency tracking, optical flow and elastic image registration. The techniques were contrasted in terms of tracking and strain accuracy. The feasibility of the obtained strain measurements to diagnose pathology was also tested for ischemia and dyssynchrony.

***Index Terms*—3D echocardiography, speckle tracking, cardiac strain, standardization, quality assurance, synthetic datasets.**

I. INTRODUCTION

A. *Echocardiographic Strain Imaging*

Echocardiographic strain imaging has emerged as a powerful tool to quantify local cardiac mechanics non-invasively [1]. Although strain heavily depends on extrinsic conditions (size, preload, and afterload), it decreases when contractility is affected [2]. Strain is less sensitive to tethering from surrounding segments than velocity or displacement measurements, and is therefore preferred for detecting abnormal segments [2].

Clinical feasibility of echocardiographic strain has been shown in a multitude of studies: strain has been used to diagnose myocardial ischaemia; it has been proposed as a tool to predict infarct size after coronary reperfusion; it is recommended as routine measurement in patients undergoing chemotherapy; it has been proposed as predictor of risk of ventricular arrhythmias; it may be applied to guide placement of the pacing lead in patients receiving cardiac resynchronization therapy [3]. Strain is more sensitive than ejection fraction as a measure of systolic function [3] and has better prognostic value [4]. Although strain imaging still remains a technology under development whose employment is limited to centers with expertise in deformation imaging, there are strong ongoing efforts aimed to standardize its use and foster its adoption in the clinical routine [5].

Recently, real time 3D ultrasound has become technically feasible. Volumetric imaging solves intrinsic shortcomings of the standard 2D modality, such as the presence of out-of-plane motion, problems in the optimal slice selection and the need of geometrical assumptions. This has motivated natural attempts from academia and industry to translate the strain imaging technology to 3D. Yet, the reduced temporal resolution of today's 3D equipment and the increased data size make accurate and time effective estimation of cardiac strain in 3D still challenging. As such, 3D strain technology remains so far mainly confined to the research arena [6], [7].

B. *3D Speckle Tracking*

To prepare for the clinical translation of the 3D strain technology, a plethora of solutions have been constantly proposed, generally referred to as 3D speckle tracking (3D STE). Based on the way the displacement is computed from the image sequence, they can be grouped in three main categories.

The work of Brecht Heyde was supported by the Research Foundation-Flanders (FWO-Vlaanderen, Belgium) under grant 12O7515N. The work of S. Queirós was funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, in the scope of the PhD grant SFRH/BD/93443/2013. Work performed by Szymon Cygan has been supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme, realized by Center for Advanced Studies. The research leading to these results has also received funding from the EU FP7 for research, technological development and demonstration under grant agreement VP2HF (no 611823).

Martino Alessandrini*, Brecht Heyde and Jan D'hooge are with the Laboratory on Cardiovascular Imaging and Dynamics, University of Leuven (KU Leuven), 3000 Leuven, Belgium (email: martino.alessandrini@gmail.com). *Asterisk indicates corresponding author.*

Oudom Somphone and Mathieu De Craene are with Philips Research, Medisys, Suresnes, France

Maria Zontak and Matthew O'Donnell are with the Department of Bioengineering, University of Washington, Seattle, United States of America.

Szymon Cygan is with the Institute for Metrology and Biomedical Engineering, Warsaw University of Technology, Warsaw, Poland.

Sandro Queirós is with the ICVS/3Bs - PT Government Associate Laboratory and the Algoritmi Center, University of Minho, Portugal, and also with the Department of Cardiovascular Sciences, Laboratory of Cardiovascular Imaging and Dynamics, KU Leuven, Leuven, Belgium.

Daniel Barbosa is with ICVS/3Bs - PT Government Associate Laboratory, Braga/Guimarães, Portugal.

Olivier Bernard is with Université de Lyon; CREATIS; CNRS UMR5220; Inserm U1044; INSA-Lyon; Université Lyon 1; Lyon 69622, France.

Maxime Sermesant and , Hervé Delingette are with the Inria-Asclepios Project, Sophia Antipolis, France.

One possibility is to extend the original *block matching* technique, initially developed for 2D ultrasound [8] to 3D [9], [10]. Hereto, a set of 3D image patches are tracked independently from one another over the cardiac cycle. Their position in the following frame is updated by looking for the most similar patch over a predefined search window. Similarity can be defined in several ways, *e.g.* by normalized cross correlation [9]. Smoothness of the displacement field is typically imposed at a post-processing stage by interpolation and filtering. Block matching represents the conventional implementation on many commercial systems. The same concept, with appropriate modifications, can be applied to raw radio-frequency (RF) data (*RF tracking*). RF tracking is receiving an increased interest given the proven benefits when assessing small deformations [11], [12]. Nonetheless, due to the low temporal resolution of commercial systems and therefore the associated large deformations between subsequent frames, the application of RF tracking to 3D echocardiographic clinical data remains an issue. Although an initialization technique can be employed in this case, *e.g.* based on the B-mode [11], the bias introduced by the latter might overbalance the theoretical benefits of the RF-based estimation. Thus, it still remains unclear whether RF tracking has to be preferred to B-mode tracking for this particular task.

One alternative is to use the *optical flow* principle, which assumes conservation of pixel brightness over time [13]. Optical flow is typically coupled with the assumption of small displacements and either solved locally on independent image patches [14] or globally by adding a smoothing term [13]. Large deformations can be coped with by recursively applying the estimator in a pyramidal refinement scheme [14]. The Demons algorithm represents a computationally efficient simplification of the optical flow problem [15], [16]. Unlike block matching, spatial smoothness can be more naturally included as a constraint term in the optimization problem.

Finally, *elastic image registration* provides an alternative framework to compute a smooth dense deformation field. Hereto, cardiac deformation is parameterized by using its decomposition into a set of basis functions, typically B-splines [17], [18]. Several kinds of constraints can be easily included as additive penalty terms to the global cost function, such as smoothness [18] and incompressibility [19], [20].

C. Motivation

With the number of available solutions rapidly growing, assessing and comparing their performance in a reliable and reproducible way becomes of primary importance. For what concerns cardiac deformation imaging, magnetic resonance imaging (MRI) is commonly considered as the “gold standard” [6]. Nonetheless, comparison against a different modality introduces additional difficulties to the benchmarking process, such as the need to co-register the two datasets in space and time. One alternative is to make use of physical phantoms or animal preparations with sonomicrometry as a reference measurement [6]. As a drawback, all aforementioned setups are costly, complex to implement and scarcely reproducible between different research groups, thus making the reported performance hard to compare. In this regard, the use of synthetic datasets represents a more feasible alternative for a preliminary evaluation [6], [17], [21] since *i)* it is a simple setup requiring only a personal computer and *ii)* testing data can be shared electronically. Nonetheless, the poor realism of the current generation of synthetic datasets represents a limited application scope for the evaluation studies reported so far (cf. Fig. 1(a)).

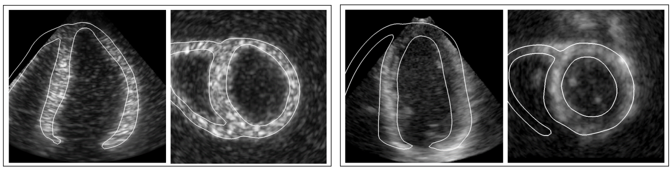

(a) synthetic dataset in [21]   (b) synthetic dataset in this study

Fig. 1. Comparison between the synthetic datasets used in [21] and the present study.

In this context, we have been actively investing in the development of more realistic synthetic datasets for benchmarking 3D STE algorithms. In particular, our efforts went on combining an accurate cardiac motion model [21] with a realistic ultrasound speckle texture [22]. As such, in our most recent contribution [23], state-of-the-art solutions in the fields of electromechanical modeling (E/M) [24] and ultrasound simulation [25] were integrated in an original framework that exploits a real ultrasound recording to learn and simulate realistic speckle textures. The obtained synthetic sequences are visually realistic, *i.e.* fairly reproduce all major elements that make motion tracking challenging, yet fully synthetic, in particular the reference values of deformation and strain are available voxelwise from the electromechanical model. The pipeline was used to generate an online testing set for 3D STE techniques[1]. The aim of this study was therefore to use this pipeline to contrast the performance of non-commercial 3D STE solutions.

D. Statement of the Contribution

We asked groups with leading expertise in the field of 3D US deformation imaging to contribute to this comparison study. Our principal effort was ensuring a maximal methodological diversity of the techniques represented. For practical reasons, the call was restricted to groups within the pre-existing network of our laboratory. Ultimately, five techniques from the following centers were contrasted: University of Leuven, University of Minho, Philips Research, Warsaw University of Technology and University of Washington. At least one technique from each of the methodological families previously identified was present, namely: one technique based on elastic registration, one implementation of RF tracking, one implementation of B-mode block matching and two techniques based on optical flow. The comparison was focused on both displacement and strain accuracy.

Of note, with respect to the comparison study between 3D STE solutions recently reported in [21], the work presented here has substantial differences:

---

[1] https://team.inria.fr/asclepios/data/straus/

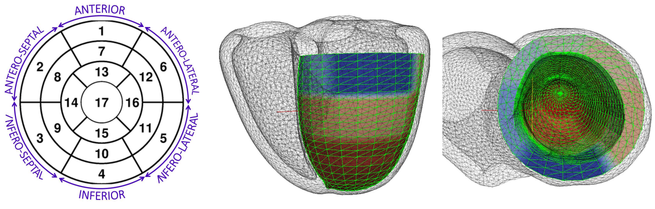
- The synthetic dataset employed here is considerably more realistic (hence representative of a real clinical setting) than the one in [21], both in terms of image properties and motion model (cf. Fig. 1). Most importantly, the new sequences embed the major elements making wall motion estimation challenging, such as *i)* the motion of surrounding structures like papillary muscles, valves and trabeculations which may hinder/bias the tracking accuracy near endo- and epicardium [26] and *ii)* the spatially variant contrast-to-noise-ratio along the the myocardial wall [19]. Moreover, the E/M simulations were improved as compared to [21] in order to obtain more representative motion patterns, with particular attention to ejection fraction and global longitudinal, circumferential and radial strain values [23];
- This study offers a more comprehensive and representative view of the state of the research in the field of 3D US cardiac deformation imaging. Specifically, in [21], all considered techniques were based on elastic registration or optical flow. Differently, we aimed to include at least one technique from each of the main families that cluster the current research scenario. In particular, including block matching is especially relevant given that several commercial implementations of this method are available [6], [27], [28]. Moreover, as mentioned, although RF tracking is gaining popularity, a thorough comparison against B-mode based techniques is currently still missing.

A very preliminary version of this report was presented in [29]. Nonetheless i) elastic registration and block matching were considered only, ii) the implementations of the two algorithms were different (*i.e.* less optimized) from the ones considered here, iii) 5 synthetic sequences were considered only and iv) the performance analysis was less thorough. The paper proceeds as follows. The setup of the comparison study is described in Sect. II. Sect. III briefly describes the considered algorithms, while the results are presented in Sect. IV. Discussion and limitations of the study are presented in Sect. V and Sect. VI, respectively. Conclusions are left to Sect. VII.

II. SETUP OF THE COMPARISON STUDY

A. Synthetic Data

The synthetic evaluation database was generated with the pipeline described in [23] and is available online at[1]. The synthetic sequences appear similar to real ultrasound recordings, yet, the myocardial motion is fully controlled by the E/M model in [24]. By varying the parameters of the E/M model, we generated 8 sequences corresponding to different patho-physiological conditions, namely: one healthy sequence; four ischemic cases, corresponding to occlusion of the proximal or distal parts of the left anterior descending coronary artery (LADprox and LADdist, respectively), of the left circumflex coronary artery (LCX) and of the right coronary artery (RCA); three simulations of dilated cardiomiopathy, of which one with a synchronous activation pattern (sync) and two dyssynchronous due to left branch bundle block (LBBBsmall and LBBBlarge), characterized by a progressively longer delay in the activation of septum and lateral wall.



(a) (b)

Fig. 2. (a) AHA segments. (b) In black the tetrahedral used by the E/M simulator, in colors the anatomical mesh used to compute displacement and strain indices. Colors denote different AHA segments.

Each E/M simulation returned a time series of tetrahedral meshes defining the instantaneous position of the myocardium over the full simulated cycle. Indices of cardiac deformation/strain are more conveniently expressed in anatomical coordinates, namely radial (R), longitudinal (L) and circumferential (C) [6]. The anatomical coordinates were therefore used to re-sample the original tetrahedral meshes regularly along L (30 points), C (25 points) and R (3 points). All details are provided in [30]. The final result of the re-meshing operation is illustrated in Fig. 2(b). For regional analysis, the LV was split into 17 segments according to the standard AHA subdivision (cf. Fig. 2). The re-sampled meshes were used as ground-truth in the comparison study, as explained in Sect. II-C, therefore being referred to as "ground truth meshes".

The simulated ultrasound volumes were obtained by applying the fast US simulator COLE [25]. Parameters were tuned so to match as close as possible to the current state-of-the-art 3D ultrasound systems in clinical practice. In particular, we implemented a phased array with center frequency of 3.3MHz with a relative bandwidth at -6dB of 65%. Sampling frequency was 50MHz. The US pyramid consisted of 107 beams in azimuth and 80 beams in elevation direction over an angle of 76 degrees in both directions. As such, the size of RF volumes was 107×13637×80 voxels. After scan conversion, the size of B-mode volumes was 224×176×208 voxels, for a voxel size of 0.7×0.9×0.6 mm^3. For all datasets, time resolution was 34 frames/s. For a dynamic visualization of the synthetic recordings, we refer to our project's web page[1] and to the interactive web interface[2].

B. Processing done by the participants

For each of the 8 simulations the following was available:

- the raw RF images before envelope detection and scan conversion;
- the B-mode voxel images, *i.e.* after envelope detection, log compression and scan conversion;
- the "ground truth meshes", *i.e.* after the re-sampling of Fig. 2(b);

Additionally, a set of MATLAB scripts to compute tracking and strain errors, hence to be used for parameters tuning, were also distributed. The dataset was split between a training and a

---

[2] https://desk.creatis.insa-lyon.fr/straus/

TABLE I
CONSIDERED ALGORITHMS. REFER TO THE TEXT FOR A DETAILED DESCRIPTION.

Algorithm	*Family*	*Image Data Processed*	*Global/ Local*	*Cost function*	*Spatial Regularization*	*Temporal Regularization*
AFFD [20]	*Elastic Image Registration*	*B-mode*	*Global*	*Sum of Squared Differences*	*Bending + Incompressibility*	*Forward + backward tracking*
AAOF	*Optical flow*	*B-mode*	*Local*	*Brightness constancy*	*Local Affine model fitting*	*Forward + backward tracking*
S-Demons [16]	*Optical flow*	*B-mode*	*Global, sparse*	*Sum of Squared Differences*	*Gaussian Smoothing*	*Forward + backward tracking*
BM	*Block Matching*	*B-mode*	*Local*	*Normalized cross correlation*	*None*	*Pointwise Gaussian smoothing*
RFBM	*Block Matching*	*RF*	*Local*	*Complex Normalized cross correlation*	*local NCC averaging + spatial smoothing*	*None*

testing set. The training set consisted of the full data package (*i.e.* RF data, B-mode images and ground truth meshes) for the LADprox and LBBBlarge sequences. For the remaining sequences (*i.e.* the testing set) we provided all simulated ultrasound data (*i.e.* RF data and B-mode images) plus the ground truth mesh for the first frame only. The remaining ground truth meshes were kept by the organizers[3].

Participants ran their specific tracking solution on each simulated dataset and used it to propagate the nodes of the first ground truth mesh over the full cycle. Performance was then assessed by comparing the ground truth positions against the tracking result, as explained in Sect. II-C.

C. Performance Assessment

1) Tracking Accuracy: Tracking error vector was defined as $\mathbf{e}^i(k) = \mathbf{x}^i(k) - \hat{\mathbf{x}}^i(k)$, with $\mathbf{x}^i(k)$ representing the 3D position of the i-th node of the ground-truth mesh at frame k and $\hat{\mathbf{x}}^i(k)$ the position of the same node when using the tracking result. Besides error amplitude, we also evaluated separately its projections along the R, L and C anatomical directions.

2) Strain Accuracy: Radial, longitudinal and circumferential strains (ϵ_R, ϵ_L and ϵ_C, respectively) were measured by the relative change in distance between two neighboring mesh nodes. Namely, $\epsilon_n(k) = \ell_n(k)/\ell_n(0) - 1$ with $\ell_n(k)$ the distance between two consecutive nodes along direction $n \in \{R, L, C\}$ at time k. Segmental strain values were obtained by averaging point measurements per cardiac segment.

For the ischemic dataset, we focused the error analysis to end-systolic (ES) strain, given its clinical interest as a "techno-marker" for ischemic heart disease [6]. Hereto, we measured the Pearsons's correlation coefficient ρ and the slope α of the regression line, as well as the bias μ and the limits of agreement (LOA=1.96σ). Moreover, we tested the accuracy of the computed ES strain values in detecting ischemia by measuring the area under the curve (AUC) of the receiver operating characteristic (ROC) curves. The ROC curves were computed from a progressive threshold by assuming ES (absolute) strain values below the threshold as indicative of ischemia. Segmental strain values (*i.e.* point estimates averaged per segment) were considered in the analysis.

For the dyssynchronous dataset, the analysis of the full strain profiles was considered. Accuracy in matching full strain profiles was therefore measured by the relative error $\sum_t |estimated\_strain(t) - reference\_strain(t)| / \sum_t |reference\_strain(t)|$. The error measurement was restricted to the septum (segments # 8 and 9) and lateral wall (segments # 11 and 12), cf. Fig. 2(a). In particular, time-to-peak measurements were extracted as clinical markers to quantify a contraction delay. Time-to-peak values were computed using a continuous B-spline interpolation of the strain profiles. As such, non-integer values were allowed. For all the experiments, only the mesh nodes falling inside the field of view were considered.

3) Statistical Tests: For each reported ρ value, Fisher's transformation was used to test the hypothesis of no correlation. The slope of the regression line was assessed with a F-test. Strain bias μ was instead evaluated with a t-test. The statistical significance of the reported AUC values (*i.e.* of AUC > 0.5) was tested with the non-parametric technique described in [31]. We also tested the statistical significance of differences between all possible pairs of algorithms. Hereto, the Fisher's transformation was used for correlation coefficients, a paired t-test was employed for regression slopes and biases while limits of agreement were compared with an F-test. All values were considered statistically significant when $p < 0.05$.

III. CONSIDERED ALGORITHMS

The main features of the algorithms considered in the comparison are summarized in Table I. A detailed description follows.

A. Anatomical Free Form Deformation (AFFD)

The group from KU Leuven contributed with an anatomical free-form deformation technique (AFFD). In this model, the

---

[3] Note that public release of the ground truth data on our web-page was postponed until the completion of this study.

displacement field is represented using a tensor-product B-spline and the control points' grid is locally oriented along the cardiac R, L and C directions of the endocardial surface. The nodes's weights are optimized with a limited memory Broyden-Fletcher-Goldfarb-Shanno optimizer with bound constraints (L-BFGS-B). The initial 3D AFFD technique was presented by Heyde *et al.* in [32], and uses an energy consisting of a sum-of-squared difference image metric representing the data term and a bending energy to enforce smoothness. The implementation used in this comparison evolves from the initial one by further including a penalty for deformations deviating from myocardial volume conservation [20]. The hyper parameters balancing bending energy and volume conservation were optimized in the training phase.

The AFFD technique and all intermediate steps leading to it have been thoroughly evaluated in-silico, in-vitro and in-vivo (cf. [20] and the references therein). In particular, AFFD was the technique of choice for strain imaging in a recent open challenge [21].

B. Anatomical Affine Optical Flow (AAOF)

The group from the University of Minho contributed with a localized anatomically-constrained affine optical flow (AAOF) algorithm. Their algorithm is the 3D extension of the one proposed in Queirós *et al.* [33] for fast left ventricle tracking in cine cardiac MRI datasets. The principle is to estimate the motion between adjacent cardiac phases using optical flow. In order to reduce its computational burden and avoid the influence of surrounding tissues, the motion estimation step is anatomically-constrained by only considering a region of interest around the tracked surface. Subsequently, the estimated motion is integrated into a local affine motion model, in which each surface point considers the motion of its neighbor points and their relative distances. By including an iterative displacement refinement scheme, the algorithm is able to accurately capture large displacements. In the present case, the reference mesh is initially divided into three surfaces, namely endo-, mid- and epicardial surfaces. Each surface is then propagated independently by estimating its motion throughout the cardiac sequence.

The AAOF algorithm extends the recent BEAS framework for LV endocardial tracking [34]. BEAS was evaluated clinically and showed competitive against non-commercial solutions (cf. [34] and the references therein). Moreover, it was among the most accurate techniques in a recent open challenge on 3D LV segmentation [35]. Although this is the first application of AAOF to 3D US, its global counterpart for cine MRI was evaluated extensively [36].

C. Sparse Demons (S-Demons)

Philips contributed with a sparse implementation of the Demons algorithm [15], called Sparse-Demons (S-Demons), which was previously presented in Somphone *et al.* [16]. Briefly, the idea is to find a dense, non-rigid displacement field by minimizing an energy functional defined only on a finite number of points of interest. A fluid-like regularization of the displacement is adopted, which can be approximated by Gaussian filtering [16]. The crucial parameters to be adjusted are the width of the Gaussian kernel σ (*i.e.* the larger sigma, the smoother the estimated displacement field) and the number and location of the points of interest. For more details we refer to [16].

S-Demons was previously evaluated on synthetic 3D recordings in [16] and found competitive with the best performing algorithms considered in [21].

D. B-mode Block Matching (BM)

The group from the Warsaw University of Technology contributed with an implementation of B-mode block matching. Three dimensional normalized cross correlation (NCC) was used as a similarity function. Prior to localizing its maximum, NCC was interpolated in 3D by using cubic B-splines in order to achieve a resolution of 1/16 of a voxel. The search range for each block was set to cover a maximum physiological velocity of 12 cm/s. Kernel size was chosen experimentally by assessing the average and median displacement errors on the training set. As such, final estimations were carried out using a kernel of $10.7\times13.6\times9.3$ mm (*i.e.* $16\times16\times16$ voxels). All the mesh nodes were tracked through the whole deformation cycle. The method implemented spatial smoothing based on gaussian filtering. This smoothing was executed at each frame, before estimating positions of mesh nodes at next time step. Temporal smoothing for each point trajectory was also implemented. This was carried out after the whole cycle of displacements was estimated. Assessment of the method's performance showed that, due to the sparsity of tracking points, spatial smoothing caused estimation errors to increase. For this reason spatial smoothing was omitted. For temporal smoothing, Gaussian filter with size of 7 time points was chosen.

E. RF block matching (RFBM)

The group from the University of Washington contributed with a RF block matching implementation based on Chen *et al.* [12]. Block matching was applied to RF data in polar coordinates, where the axial dimension was sub-sampled with a factor 4 for computational efficiency. Block size was $21\times7\times3$ voxels and computed based on the expected speckle size, as measured by the average width of the auto-correlation function of the RF signal. The search region was $51\times3\times3$ voxels and chosen based on the average displacement measured from the ground truth. Sub-pixel axial displacements were computed with the phase zero-crossing method and were totally unconstrained based on any model assumptions of expected displacement/strain values. A second-order polynomial approximation was employed in the lateral and elevational directions. A tracking quality check was performed based on comparing the result of incremental (*i.e.* between time $t-1$ and t) and direct (*i.e.* between time 0 and t) tracking. Motion patterns of adjacent speckles in the myocardium have innate coherence due to their physical proximity. We achieve spatial coherence by averaging the similarity (NCC) maps of adjacent voxels. Moreover, each voxel participates in multiple blocks, hence has several estimates, which might be further averaged

TABLE II
GLOBAL TRACKING ERROR IN MM [5-TH PERCENTILE, MEDIAN, 95-TH PERCENTILE]

Algorithm	Full database	Normal Geometry	Dilated Geometry
AAOF	[0.09; **0.38**; 1.51]	[0.09; **0.38**; 1.35]	[0.08; **0.39**; 1.88]
AFFD	[0.14; **0.47**; 1.28]	[0.14; **0.48**; 1.30]	[0.13; **0.44**; 1.23]
S-Demons	[0.14; **0.49**; 1.41]	[0.15; **0.49**; 1.41]	[0.13; **0.50**; 1.41]
BM	[0.26; **0.90**; 2.46]	[0.26; **0.86**; 2.11]	[0.25; **0.99**; 3.19]
RFBM	[0.09; **0.72**; 4.33]	[0.10; **0.72**; 4.40]	[0.07; **0.73**; 4.23]

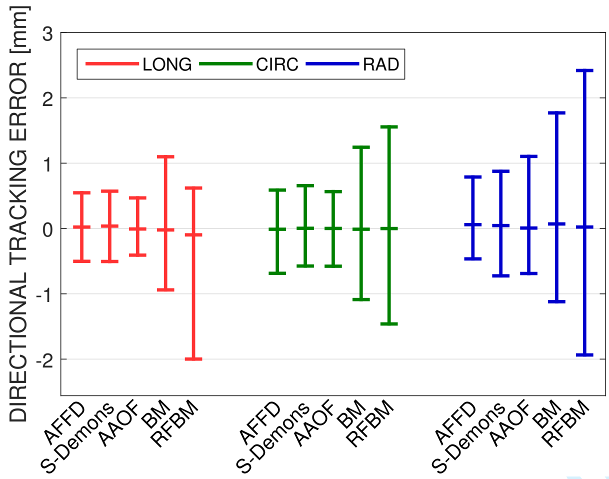


Fig. 3. Box-plot of tracking error along the 3 anatomical directions. The three markers represents 5-th, 50-th and 95-th percentile, respectively.

to reject non-consistent estimates. The algorithm was used to find a dense displacement field on the whole volume. The mesh nodes were then propagated by linearly interpolating the displacement field.

RFBM was previously evaluated in-vitro on cardiac phantoms and in-vivo on open chest animal preparations (cf. [37] and the references therein).

IV. RESULTS

For synthesis, we provide in this report only a, yet representative, subset of the obtained results. A supporting document is made available containing the full set of error plots. For reproducibility, the ground-truth meshes and the associated tracking result from all participants are made available at our project web-page[1], along with the MATLAB scripts used to compute tracking and strain errors **** note for the reviewers: All supporting material will be made public at the time of publication. At this stage, it can be downloaded at http://bit.ly/1QnUEKm ****. Of note, all algorithms (with a slight difference for RFBM, as mentioned in Sect. III) performed the tracking by accumulating incremental estimates between couples of consecutive frames. Moreover, in order to minimize error accumulation, AFFD, S-Demons and AAOF implemented a bi-directional tracking strategy combining forward and backward estimates (cf. Table I).

A. Tracking Accuracy

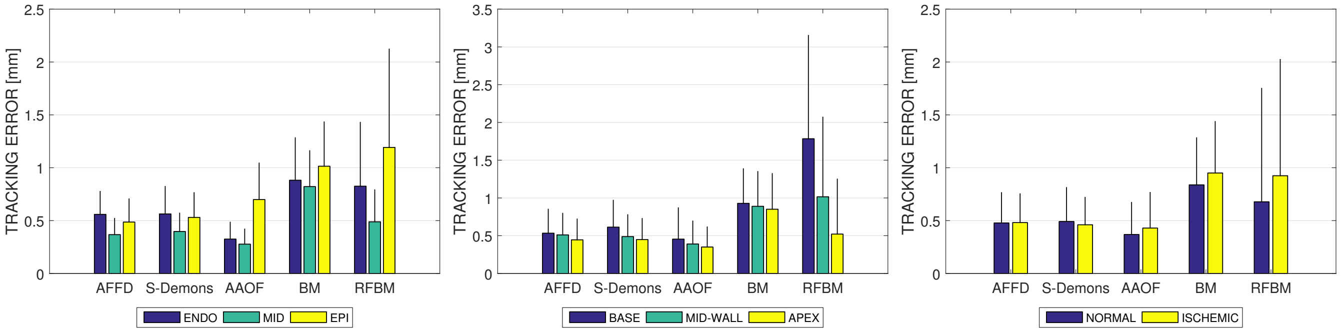
1) Global analysis: Global tracking errors (*i.e.* inclusive of all nodes, time instants and sequences) are summarized in Table II. To evaluate the influence of the geometry on the tracking quality, error statistics are also reported separately for sequences obtained from a healthy geometry (*i.e.* the healthy simulation and the four ischemic ones) and sequences obtained from a dilated geometry. AAOF returned the smallest mean error on both ischemic and dilated datasets. AFFD and S-Demons returned slightly larger average errors while error dispersion was instead smaller as compared to AAOF. BM and RFBM had a larger error bias and wider dispersion ranges as compared to non-block matching solutions. As a reference, the average ground-truth displacement at end-of-systole was 4.64 mm.

Fig. 3(b) presents the 3 anatomical components of the tracking error. For all algorithms, radial errors had the largest variance and a slight positive bias (*i.e.* tendency to underestimate radial motion). Among non-block matching solutions, the relative loss in accuracy in the radial direction was particularly relevant for AAOF. Longitudinal and circumferential errors did not exhibit clear biases, except for RFBM, which tended to underestimate longitudinal displacements. Such longitudinal errors are explained by the difficulty of RF tracking in coping with large motions. As an example, note that, with the US parameters employed, a displacement of 1 mm leads to a shift of ~ 65 RF samples.

2) Regional analysis: Error dependency on the myocardial layer was analyzed first (cf. Fig. 4(a)). All algorithms returned more accurate estimates at the mid-myocardial level, *i.e.* where boundary effects are less important. Moreover, errors were generally higher on the epicardial surface. This is a consequence of the poor delineation of the epicardial border typical of cardiac ultrasound [38]. As expected, error increase was the highest for the solutions which did not apply an explicit constraint in the radial direction (AAOF and RFBM in particular). This effect was particularly important for RFBM, as also illustrated by the bull's eye plot in Fig. 5.

Error variations per left ventricular level (*i.e.* basal vs. mid vs. apical segments) and per functional region (*i.e.* normal vs. ischemic segments) were also analyzed, cf. Fig. 4(b, c). The first experiment showed a progressive error increase when moving from apex to base, cf. Fig. 4(b). This is partly explained by the lower image quality when moving away from the probe and by the progressively larger absolute displacements in the apex to base direction. This effect was particularly important for RFBM. Again, this can be due to the difficulty in coping with large displacements. Errors were in contrast more uniformly distributed with respect to variations in the mechanical function of the segments. Yet, except for S-Demons and AFFD, there was a tendency in returning more accurate tracking on normal segments as compared to ischemic ones, Fig. 4(c). This might be a consequence of the tuning strategy chosen. In particular, tuning with respect to global error will naturally privilege normal regions due to their larger relative extent.

Fig. 6 displays the tracking error over time for the LADprox training sequence. The vertical blue lines denote aortic valve opening, aortic valve closure and opening of the mitral valve, respectively. All algorithms made the highest errors at ES, *i.e.* when the displacement is the largest. Techniques combining



(a) (b) (c)

Fig. 4. Box-plots of tracking error per trans-mural layer (a), left ventricular level (b) per functional region (c). The main bar represents the median error while the black segment denotes the mean absolute deviation.

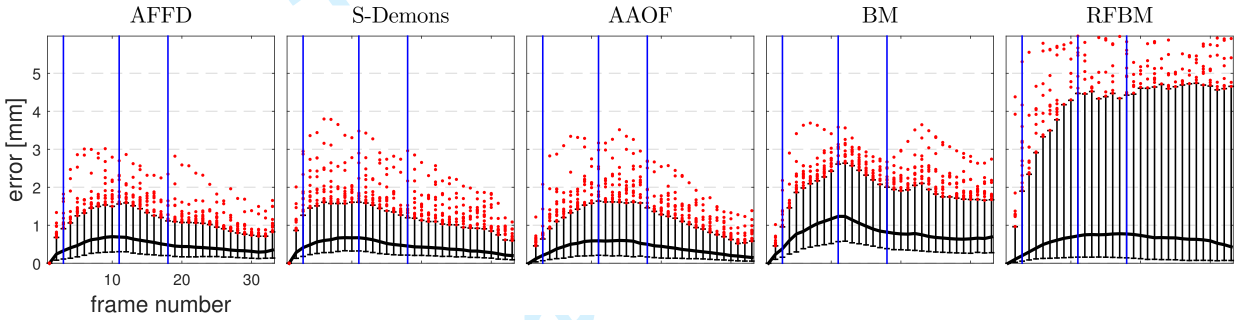
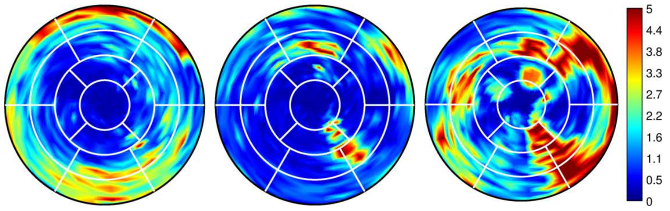


Fig. 6. Boxplots of tracking error over time on the LADprox training sequence (5-th, 50-th and 95-th percentile values, outliers as red dots). The vertical blue lines denote aortic valve opening, aortic valve closure and opening of the mitral valve, respectively. For all techniques, the population of outliers was decimated for better visualization. For the same reason, the error plot of the RFBM was partly cropped.



(a) (b) (c)

Fig. 5. Bull's eye plot of tracking error for the RFBM sequence evaluated at ES on the LADprox sequence independently on the 3 trans-mural layers.

forward and backward tracking (AFFD, AAOF and S-Demons) effectively reduced error accumulation as compared to BM and RFBM, as evidenced by the lower end-diastolic error values. In this sense, AFFD had a slightly higher error drift at the end of the cardiac cycle as compared to AAOF and S-Demons. In general, median errors had comparable behavior and were mostly below 1 mm. Among the considered solutions, RFBM suffered the most from the presence of outliers, as shown by the larger error dispersion.

B. Strain Accuracy

1) Detection of Ischemia: The obtained numerical values of correlation coefficient ρ, regression slope α, bias μ and limits of agreement LOA for ES strain are summarized in Table III. For each metric, the best values are denoted in bold font. In these experiments, the 4 ischemic sequences were considered only.

AFFD was the only algorithm to preserve high accuracy irrespective of the strain direction. In particular, it was the only technique for which the bias was never statistically significant. Moreover, it had the highest ρ for C- and R-strain and the lowest μ and LOA for R-strain. Remaining algorithms had instead problems in retrieving R-strain accurately. AAOF returned highly accurate estimates of L-strain, with the highest ρ, the narrowest LOA and a non-significant bias. Compared to AFFD and S-Demons, AAOF had the slight tendency to under-estimate C-strain, as shown by the lower α. Moreover, AAOF returned a statistically significant positive bias for R-strain. S-Demons had high accuracy for L- (highest α) and C-strain (highest ρ and narrowest LOA), while it had the tendency to under-estimate R-strain substantially, as explained by the statistically significant negative μ and the low α.

Algorithms based on block-matching were in general less accurate than algorithms based on elastic registration or optical flow. In particular, they mainly suffered from the noisy displacement estimates, which explain the larger LOA. The gap in performance was the smallest for BM, in particular for L- and C-strain. Yet, in both cases, BM had the tendency to underestimate strain values as shown by the relatively low α. RFBM suffered the most from the presence of outliers as shown by the larger LOA.

For each accuracy metric and each possible pair of algorithms, we tested the statistical significance of the difference between the values obtained, as explained in Sect. II-C. The results of the test have been left out of this manuscript

TABLE III

CORRELATION COEFFICIENT ρ, SLOPE OF THE REGRESSION LINE α, BIAS μ, LIMITS OF AGREEMENT LOA, AND AREA UNDER THE ROC CURVE (AUC) FOR THE 5 ALGORITHMS. AN ASTERISK DENOTES STATISTICAL SIGNIFICANCE OF THE REPORTED MEASUREMENTS, NAMELY OF $\rho > 0$, $\alpha > 0$, $\mu > 0$ AND AUC > 0.5. BEST VALUES ARE REPORTED IN BOLD FONT. VALUES WERE MEASURED ON THE ISCHEMIC DATASETS ONLY.

	ϵ_L					ϵ_C					ϵ_R				
	ρ	α	μ	LOA	AUC	ρ	α	μ	LOA	AUC	ρ	α	μ	LOA	AUC
AFFD	0.84*	0.89*	-0.25	3.95	0.90*	**0.98***	0.88*	0.16	4.84	**0.99***	**0.93***	0.83*	**-0.06**	**10.00**	**0.98***
S-Demons	0.88*	**0.96***	0.42	3.65	0.93*	**0.98***	0.89*	0.43	**4.52**	**0.99***	0.83*	0.62*	-12.58*	15.55	0.95*
AAOF	**0.95***	0.88*	-0.35	**2.20**	**0.96***	**0.98***	0.81*	**-0.09**	5.63	**0.99***	0.83*	**0.92***	9.25*	17.21	0.96*
BM	0.75*	0.61*	**-0.06**	4.53	0.89*	0.89*	0.59*	1.71	11.14	0.97*	0.47*	0.22*	-12.64*	24.25	0.81*
RFBM	0.36	1.43*	14.98	25.34	0.67	0.87*	**0.98***	8.22*	11.95	0.95*	0.35*	0.46*	-5.04	36.36	0.80*

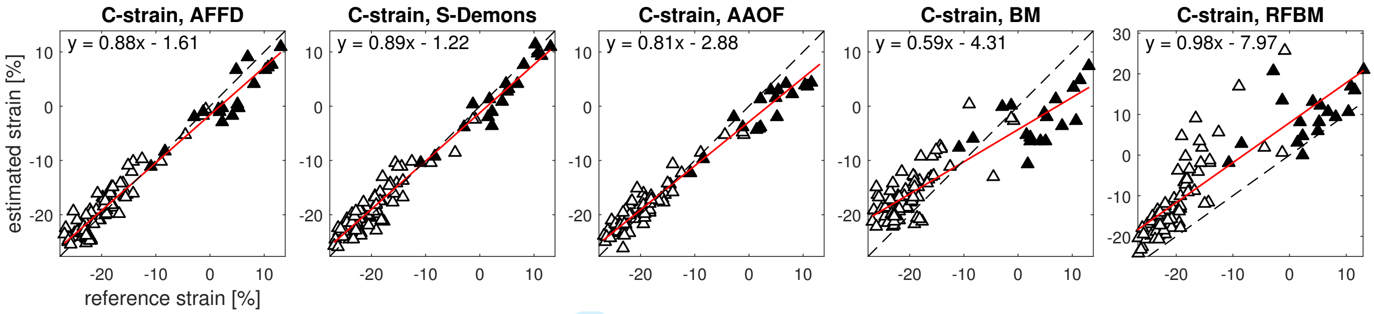


Fig. 7. Correlation plot for end-systolic circumferential strain. Empty and full markers denote healthy and ischemic segments respectively. The dashed line in represents $y = x$ (*i.e.* perfect correlation), while the regression line, whose equation is reported inside the plot, is displayed in red. Note the different error scale used in both plots for RFBM.

for the sake of brevity and are provided in the supporting document. The main points arising from the statistical test are the following.

For what concerns non block-matching algorithms (AFFD, AAOF and S-Demons), differences in L- and C-strain were, for most of the metrics, not statistically significant. The accuracy of AFFD in R-strain was statistically higher with respect to any competing solution. For what concerns block-matching solutions (BM and RFBM), BM accuracy was statistically higher with respect to all metrics for L- and C-strain. Differences in R-strain were instead not significant for most metrics.

When comparing the two groups of algorithms against each other, BM had non-significant differences to at least one of the non block-matching based solutions in L-strain (ρ, α, μ and LOA, *i.e.* all assessed metrics), C-strain (ρ, α, μ and LOA) and R-strain (μ). Instead, error differences between non-block-matching algorithms and RFBM were in most cases significant.

Correlation plots for ES C-strain are reported in Fig. 7, where empty and full markers denote healthy and ischemic segments, respectively. The plots in Fig. 7 allow to make several additional considerations. At first, BM and, to a lesser extent, AAOF seemed to be biased towards healthy segments. The overall lower α seemed therefore mainly explained by the tendency to underestimate strain in ischemic regions. This trend was instead less apparent for L- and R-strain. The plots also show that C-strain errors of RFBM were mainly localized in few segments, while L- and R-strain errors were instead more uniformly distributed.

The bull's eye plots in Fig. 9 allow to gain further insights on the algorithms' performance. In particular, they correspond to C-strain estimates on the normal and LCX sequences. Ischemic segments, as available from the ground truth, are denoted by the bold black contours. To reduce boundary effects, strain was measured at the mid-myocardial layer. All algorithms returned strain maps qualitatively similar to the ground truth and ischemic segments were identified by lower (absolute) strain values. The plots confirm the tendency of BM and, to a lesser extent, AAOF to return low C-strain estimates. The bull's eyes also show that regularized solutions tended to smear the interface between healthy and ischemic regions out. Differently, block-matching algorithms returned strain maps that were overall more noisy but also more sensitive to rapid strain variations. This is particularly evident for RFBM.

To evaluate more quantitatively the capability to discriminate ischemic from normal segments, we measured, for each algorithm, the area (AUC) under the ROC curve. The obtained results are reported in Table III, while the ROC curves are plotted in Fig. 8. The AUC values measured on the ground-truth were AUC $= 0.98$ for R-strain, AUC $= 0.95$ for L-strain and AUC $= 0.99$ for C-strain. In Fig. 8, the ground-truth ROC curve is represented by the gray shaded region. All algorithms achieved a good separation (AUC > 0.9) when using C-strain, with S-Demons reaching the ground truth value. Interestingly, although less accurate than for the other components, R-strain estimates allowed a good discrimination for all algorithms. Ground truth values of L-strain were intrinsically the less discriminative and this explains the lower area values which were obtained in average. Yet, all algorithms achieved AUC$\sim$0.9, except for RFBM where L-strain was deteriorated due to the low SNR of the tracking results.

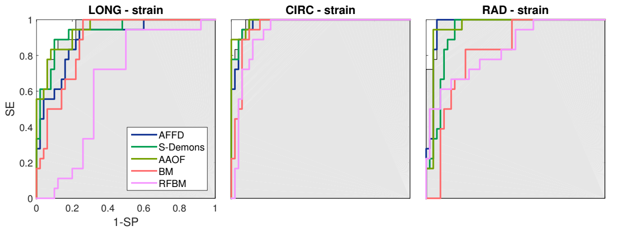


Fig. 8. ROC curves for detection of ischemic segments based on thresholding of ES strain. The grey-shaded region represents the ground-truth curve.

TABLE IV
RELATIVE ERROR (IN %) BETWEEN ESTIMATED AND REFERENCE STRAIN CURVES (MEAN ± STANDARD DEVIATION OVER THE DYSSYNCHRONOUS DATASET).

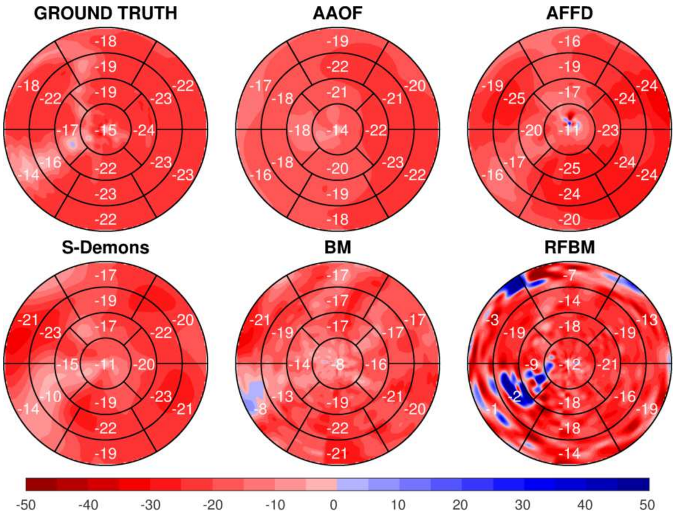
	C-strain	L-strain	R-strain
AFFD	0.16 ± 0.44	1.20 ± 3.58	**0.48 ± 0.93**
S-Demons	**0.12 ± 0.15**	**1.10 ± 4.42**	1.37 ± 2.46
AAOF	0.18 ± 0.76	1.26 ± 5.60	2.04 ± 1.96
BM	0.62 ± 1.94	9.27 ± 46.08	0.73 ± 1.16
RFBM	0.45 ± 0.38	15.00 ± 89.40	2.54 ± 4.01

2) Detection of Dyssynchrony: Table IV reports the relative errors between estimated and ground-truth strain curves, as defined in Sect. II-C, where the best performance is denoted by the bold font. S-Demons returned the most accurate L- and C-strain estimates while, as in the ischemic dataset, it tended to underestimate R-strain. AFFD estimates were instead more constant with respect to the strain direction. AAOF tended to overestimate R-strain. Fig. 10 compares longitudinal strain profiles measured from the septum and lateral wall. The higher error drift for BM and RFBM is partly explained by the fact that neither technique compensated for error accumulation. The time-to-peak values computed from the estimated strain profiles are represented by the markers in Fig. 10. Fig. 10 shows that, for all algorithms, the computed timing differences were accurate enough to detect a dyssynchronous contraction.

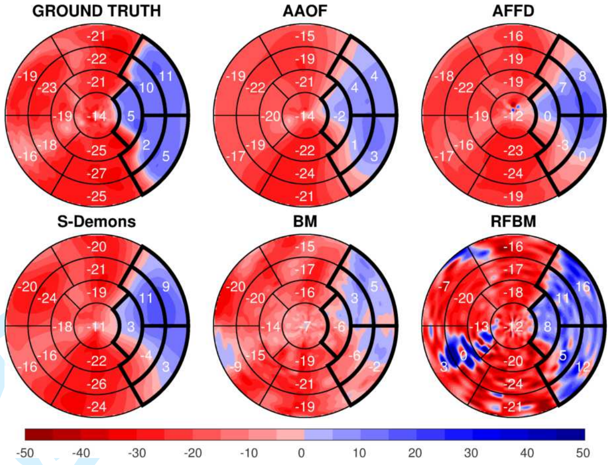
C. Computation Time

Computation times cannot be compared directly because of *i)* the different implementations/hardware employed *ii)* conceptual differences between algorithms computing the displacement voxelwise on the full image domain (AFFD, RFBM) versus algorithms localizing the computation to a sparse set of points only (AAOF, S-Demons, BM) and *iii)* differences in the kind of data processed (*i.e.* RF tracking is intrinsically computationally more demanding given the bigger data size). Moreover, for the purpose of this comparison, algorithms were optimized for accuracy and not execution time. Nonetheless, computational complexity remains an important constraint towards clinical translation and is therefore reported here for indicative purposes.

AFFD was implemented in C++ and required ∼10 min/frame on a Linux cluster (1.8GHz CPU, 256GB RAM). The AAOF algorithm was implemented in MATLAB and took 3.8 s/frame on a 4-core 3.60 GHz processor. The S-Demons algorithm required 0.08 s/frame on an Intel Core i7-4800MQ CPU @ 2.70GHz with 16GB RAM. The RFBM algorithm required 50 min/frame on a stand-alone Intel(R) Core i7 with 3.6GHz CPU and 32GB RAM. BM was coded in MATLAB and took 129 s/frame on a Intel i7, 6GB RAM notebook. For computational efficiency, a C implementation of the NCX function was wrapped in the MATLAB code.



(a) normal



(b) LCX

Fig. 9. Bull's eye plots of end-systolic circumferential strain for the normal (a) and LCX (b) sequence. The ground truth is at the top left. Ischemic segments are identified by the bold black contours. The text in white reports the average strain in %, rounded to the closest integer value, measured in the segment.

Assuming a reference temporal resolution of 30 frames/s for 3D US, S-Demons and AAOF were therefore the best candidates for online processing. Moreover, although this was not the case for the considered implementation, we note that block-matching can be easily optimized for fast execution on parallel architectures such as GPU's. AFFD is instead based on iterative optimization solvers and is therefore typically less parallelizable. Improving computational efficiency of the AFFD technique is topic of active research within the developing team [39].

V. DISCUSSION

We contrasted five non-commercial solutions for 3D deformation imaging in cardiac ultrasound by using a simulation

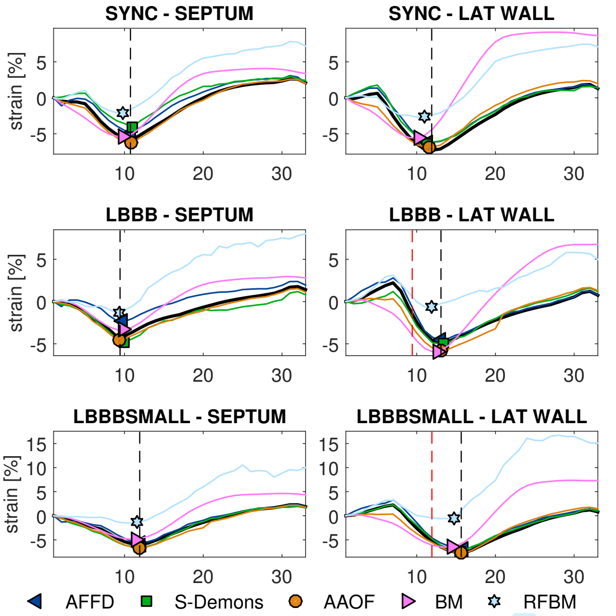


Fig. 10. L-strain profiles for septum and lateral wall measured on the dyssynchronous dataset. The bold black curve represents the ground truth. The markers indicate the time-to-peak values computed from the strain curves. Note that drift corrections was performed before extracting the time-to-peak values. The the black dashed vertical lines denote the reference time-to-peak value. The red dashed vertical line denotes the reference time-to-peak value in the septum.

pipeline we have recently developed [23].

Techniques based on elastic registration (AFFD) and optical flow (AAOF and S-Demons) had similar tracking accuracy. In particular, AAOF returned the smallest mean error while error dispersion was slightly larger as compared to AFFD and S-Demons. L-strain estimates were accurate for all three solutions with correlation coefficients close to 0.9 and non-significant biases. This is noteworthy given the relevance of L-strain in clinical diagnostics [6]. AFFD and S-Demons returned accurate C-strain estimates while AAOF had the slight tendency to underestimate circumferential deformations. Yet, C-strain error differences where found not to be statistically significant. AFFD was the only solution to preserve high accuracy for R-strain while performance dropped for AAOF and S-Demons. This is particularly relevant given that achieving accurate estimates of radial deformation appears to be an open challenge of the 3D STE technology [21], [40]. The higher accuracy of AFFD can be related to the presence of a volume conservation penalty, which constrains R-strain estimates to the, typically more robust, L- and C-strain estimates. R-strain errors for AAOF could be due to the fact that the three trans-mural surfaces were tracked independently. Hence, the technique could benefit from a spatial constraint in the radial direction.

Techniques based on block-matching (BM and RFBM) were in general less accurate as due to the more noisy displacement (hence, strain) estimates. Overall, BM was more accurate than RFBM. In particular, its accuracy for L- and C-strain was not statistically different from the one of non block-matching solutions. RFBM suffered the most from the presence of outliers, as due to the large inter-frame displacements in RF space and to the boundary effects at endo- and epicardium. In this regard, we note that RF-based tracking is typically coupled with high frame rate imaging where displacements are intrinsically smaller [41]. Hereto, an initialization based on B-mode images could help cope with large displacements [11]. Moreover, the high sensitivity of RF tracking can be preserved even for 3-D tracking at low frame rates when RFBM results are integrated with shape-based methods to constrain displacement values, especially near boundaries where RFBM methods exhibit high variance but shape-based methods perform well [42].

Despite the different levels of accuracy, the measured AUC values showed that all algorithms could detect ischemia with a good accuracy, especially when using C-strain (AUC $>$ 0.9). This result is also partly explained by the large strain differences between healthy and ischemic segments. This point is addressed further in the following. Strain accuracy on the dilated dataset reflected qualitatively what observed on the ischemic one. Nonetheless, time-to-peak measurements were appropriate to detect dyssynchrony for all five algorithms. Along with the AUC value, this result points to the fact that absolute strain accuracy and the ability to derive clinically relevant features from the strain estimates are problems which deserve separate attention.

VI. LIMITATIONS OF THE STUDY

A. *Synthetic ultrasound data*

As discussed in [23], the synthetic sequences exhibit a sharp transition in the speckle properties, *i.e.* speckle motion is coherent in the myocardium, while it is fully random in the surrounding regions. This could make boundary effects more important as compared to real recordings, in particular for the un-regularized solutions. To account for this, all seed points considered in the evaluation were placed well inside the myocardium, *i.e.* far from the interface between the two regions (cf. Fig. 2).

Second, stationary ultrasound artifacts such as dropout and reverberations are not included in the evaluation dataset. Robustness to these effects is an important design criterion for effective STE solutions. Including such artifacts and evaluating their influence on the strain estimates is priority in our ongoing research.

B. *Motion model*

Concerning the healthy simulation, peak systolic L-strain remains low as compared to reported normality ranges [23]. This is a common limitation of existing heart models and improving this aspect is the topic of ongoing research.

Moreover, in the ischemic dataset, ischemic areas were large and characterized by a substantial strain reduction. This prevented from evaluating more systematically the compromise, intrinsic to regularized solutions, between the enforced smoothness of the displacement field, the spatial resolution of the strain estimates and the sensitivity to subtle strain changes. A sensitivity analysis for those methods fell beyond the scope of this study, nonetheless it remains a necessary further step towards the clinical translation of these techniques. In this perspective, segmental strain values should be complemented by more detailed error maps. In this sense, the bull's eye plots in Fig. 9 showed that regularized solutions tended to over-smooth the interface of the ischemic region, while block matching estimates, although more noisy, identified the interface with better contrast. Enlarging the synthetic dataset by including more localized and progressive levels of ischemia is thus part of our ongoing research.

Moreover, post-systolic shortening is an important additional marker of ischemic dysfunction. To our knowledge, this effect is not reproduced by existing cardiac models and, in particular, it is not present in our synthetic dataset. For what concerns the detection of dyssynchrony, other strain markers, such as of septal flash, can be used to complement time-to-peak measurements [43]. Including these effects and evaluating the accuracy in their detection is therefore the topic of ongoing investigation.

A further point is that strain estimates are known to be dependent on the temporal resolution of the imaging system [44]. In this study, temporal resolution was equal to 34 frames/s for all dataset. The value was chosen as a representative average of what is available clinically. Modifying frame rate is a straightforward adjustment of the E/M model. Evaluating the impact of frame rate on tracking and strain accuracy fell beyond the scope of this paper. Yet, this represents one of our ongoing topics of research and a preliminary report can be found in [45].

Overall, the employment of a motion model, by definition, introduces simplifications and, therefore, a bias in the results. In particular, numerical stability aspects may favor (temporally and spatially) smooth deformation fields and, hence, regularized solutions. In this regard, we note that the bio-mechanical model employed is among the most advanced available: it is the output of a long-lasting research activity and several ongoing studies are preparing its clinical translation [46], [47]. Moreover, we also stress that none of the regularized solutions employed the bio-mechanical model as a constraint to the computed displacement estimates which, obviously, would introduce an unacceptable bias.

C. Sample Size

The small sample size, along with the limited diversity of the motion patterns and image qualities mentioned above, prevents from generalizing the obtained results as well as extrapolating them to the clinical practice. Enriching the synthetic database is therefore the topic of ongoing research.

D. Performance Metrics Used

For all participants, strain was computed in the same way from the tracked positions of the mesh nodes. Nonetheless, we note that different algorithms could benefit from different strain computation strategies: *e.g.* regularized solutions could compute strain analytically from the continuous deformation field, while, for block-matching solutions, robust parametric (*e.g.* affine) fitting of the displacement estimates could be used to improve robustness to outliers. We kept strain computation uniform in order to rule out additional sources of variability in the comparison. Moreover, achieving a common set of definitions is crucial towards the standardization of the 3D STE technology [5], [6]. In particular, our strain definition is based on extending to 3D the recent recommendations for computation of global longitudinal strain with 2D US [5].

We also note that the definition of directional strain employed is insensitive to shear strain. In this perspective, the concept of *principal strain* could be adopted to compute deformation indices independent on the coordinate system used [37].

E. Algorithm implementation

This study considered an in-house implementation of block matching. As such, there is necessarily a disparity with respect to what is available in commercial packages. Although the block matching principle is in itself very well established, there are several possible degrees of freedom, most importantly in the way the displacement estimates are regularized, which can affect the final output considerably. Involving commercial packages is therefore our priority in the near future.

With these limitations in mind, this remains to our knowledge the first study where an heterogeneous set of well established non-commercial 3D STE techniques is directly contrasted on a set of realistic echocardiographic recordings. In silico testing is obviously not sufficient for clinical translation, nonetheless, it is recognized as a necessary preliminary step in this direction [48]. Extending the comparison to commercial solutions for 2D strain by using a 2D extension of the evaluation pipeline presented in this study [49] is the topic of ongoing research.

VII. CONCLUSIONS

We contrasted five 3D STE algorithms from leading teams in the field by using an evaluation pipeline we have recently developed. To our knowledge, this was the first time a diverse range of techniques were contrasted directly on a realistic set of ground-truth data.

Overall, non block matching based solutions returned closely accurate L- and C-strain estimates. AFFD was instead the only technique to preserve good accuracy for R-strain while performance dropped for competing solutions. Block matching algorithms (BM and RFBM) were in average less accurate. Yet, L- and C-strain errors for BM were mostly not statistically different from the ones of the previous three methods. RFBM suffered instead from the low temporal resolution (hence the large displacement) of the used dataset. Concerning computational complexity, S-Demons and AAOF were close to real-time processing while competing solutions (AFFD and RFBM in particular) were computationally more demanding. Moreover, BM could benefit from a substantial speed-up due to the intrinsic parallelizability.

Whether the reported findings can be extrapolated to the clinical practice remains to be established.

Our effort is to allow for a more solid and objective assessment of the state-of-the-art of the 3D STE technology, thereby promoting a more organized and effective development of the latter.

REFERENCES

[1] J. D'hooge, B. Bijnens, J. Thoen, F. Van de Werf, G. Sutherland, and P. Suetens, "Echocardiographic strain and strain-rate imaging: a new tool to study regional myocardial function," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 9, pp. 1022–1030, Sept 2002.

[2] G. R. Sutherland, G. Di Salvo, P. Claus, J. D'hooge, and B. Bijnens, "Strain and strain rate imaging: a new clinical approach to quantifying regional myocardial function," *J Am Soc Echocardiogr*, vol. 17, no. 7, pp. 788–802, 2004.

[3] O. A. Smiseth, H. Torp, A. Opdahl, K. H. Haugaa, and S. Urheim, "Myocardial strain imaging: how useful is it in clinical decision making?" *European Heart Journal*, 2015.

[4] K. Kalam, P. Otahal, and T. H. Marwick, "Prognostic implications of global LV dysfunction: a systematic review and meta-analysis of global longitudinal strain and ejection fraction," *Heart*, 2014.

[5] J.-U. Voigt, G. Pedrizzetti, P. Lysyansky, T. H. Marwick, H. Houle, R. Baumann, S. Pedri, Y. Ito, Y. Abe, S. Metz, J. H. Song, J. Hamilton, P. P. Sengupta, T. J. Kolias, J. D'hooge, G. P. Aurigemma, J. D. Thomas, and L. P. Badano, "Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the eacvi/ase/industry task force to standardize deformation imaging," *European Heart Journal - Cardiovascular Imaging*, 2014.

[6] R. Jasaityte, B. Heyde, and J. D'hooge, "Current state of three-dimensional myocardial strain estimation using echocardiography," *J Am Soc Echocardiogr*, vol. 26, no. 1, pp. 15 – 28, 2013.

[7] V. Mor-Avi, R. M. Lang, L. P. Badano, M. Belohlavek, N. M. Cardim, G. Derumeaux, M. Galderisi, T. Marwick, S. F. Nagueh, P. P. Sengupta, R. Sicari, O. A. Smiseth, B. Smulevitz, M. Takeuchi, J. D. Thomas, M. Vannan, J.-U. Voigt, and J. L. Zamorano, "Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: ASE/EAE consensus statement on methodology and indications: Endorsed by the japanese society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 24, no. 3, pp. 277 – 313, 2011.

[8] L. Bohs and G. Trahey, "A novel method for angle independent ultrasonic imaging of blood flow and tissue motion," *Biomedical Engineering, IEEE Transactions on*, vol. 38, no. 3, pp. 280–286, March 1991.

[9] Q. Duan, E. D. Angelini, S. L. Herz, C. M. Ingrassia, K. D. Costa, J. W. Holmes, S. Homma, and A. F. Laine, "Region-based endocardium tracking on real-time three-dimensional ultrasound," *Ultrasound in Medicine & Biology*, vol. 35, no. 2, pp. 256 – 265, 2009.

[10] J. Crosby, B. H. Amundsen, T. Hergum, E. W. Remme, S. Langeland, and H. Torp, "3-D speckle tracking for assessment of regional left ventricular function," *Ultrasound in Medicine & Biology*, vol. 35, no. 3, pp. 458 – 471, 2009.

[11] R. G. Lopata, M. M. Nillesen, J. M. Thijssen, L. Kapusta, and C. L. de Korte, "Three-dimensional cardiac strain imaging in healthy children using RF-data," *Ultrasound in Medicine & Biology*, vol. 37, no. 9, pp. 1399 – 1408, 2011.

[12] X. Chen, H. Xie, R. Erkamp, K. Kim, C. Jia, J. M. Rubin, and M. O'Donnell, "3-D correlation-based speckle tracking," *Ultrasonic Imaging*, vol. 27, no. 1, pp. 21–36, 2005.

[13] R. Mukherjee, C. Sprouse, A. Pinheiro, T. Abraham, and P. Burlina, "Computing myocardial motion in 4-dimensional echocardiography," *Ultrasound in Medicine & Biology*, vol. 38, no. 7, pp. 1284 – 1297, 2012.

[14] M. Alessandrini, H. Liebgott, D. Barbosa, and O. Bernard, "Monogenic phase based optical flow computation for myocardial motion analysis in 3D echocardiography," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, ser. Lecture Notes in Computer Science, O. Camara, T. Mansi, M. Pop, K. Rhode, M. Sermesant, and A. Young, Eds. Springer Berlin Heidelberg, 2013, vol. 7746, pp. 159–168.

[15] J.-P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243 – 260, 1998.

[16] O. Somphone, M. De Craene, R. Ardon, B. Mory, P. Allain, H. Gao, J. D'hooge, S. Marchesseau, M. Sermesant, H. Delingette, and E. Saloux, "Fast myocardial motion and strain estimation in 3D cardiac ultrasound with sparse demons," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, April 2013, pp. 1182–1185.

[17] A. Elen, H. F. Choi, D. Loeckx, H. Gao, P. Claus, P. Suetens, F. Maes, and J. D'hooge, "Three-dimensional cardiac strain estimation using spatio;temporal elastic registration of ultrasound images: A feasibility study," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 11, pp. 1580–1591, Nov 2008.

[18] B. Heyde, S. Bouchez, S. Thieren, M. Vandenheuvel, R. Jasaityte, D. Barbosa, P. Claus, F. Maes, P. Wouters, and J. D'hooge, "Elastic image registration to quantify 3-D regional myocardial deformation from volumetric ultrasound: Experimental validation in an animal model," *Ultrasound in Medicine & Biology*, vol. 39, no. 9, pp. 1688 – 1697, 2013.

[19] Y. Zhu, X. Papademetris, A. J. Sinusas, and J. S. Duncan, "A coupled deformable model for tracking myocardial borders from real-time echocardiography using an incompressibility constraint," *Medical Image Analysis*, vol. 14, no. 3, pp. 429 – 448, 2010.

[20] B. Heyde, M. Alessandrini, J. Hermans, D. Barbosa, P. Claus, and J. D'hooge, "Anatomical image registration using volume conservation to assess cardiac deformation from 3D ultrasound recordings," *Medical Imaging, IEEE Transactions on*, Epub. ahead of print.

[21] M. De Craene, S. Marchesseau, B. Heyde, H. Gao, M. Alessandrini, O. Bernard, G. Piella, A. Porras, L. Tautz, A. Hennemuth, A. Prakosa, H. Liebgott, O. Somphone, P. Allain, S. Makram Ebeid, H. Delingette, M. Sermesant, J. D'hooge, and E. Saloux, "3D strain assessment in ultrasound (straus): A synthetic comparison of five tracking methodologies," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 9, pp. 1632–1646, 2013.

[22] M. Alessandrini, H. Liebgott, D. Friboulet, and O. Bernard, "Simulation of realistic echocardiographic sequences for ground-truth validation of motion estimation," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 2329–2332.

[23] M. Alessandrini, M. De Craene, O. Bernard, S. Giffard-Roisin, P. Allain, I. Waechter-Stehle, J. Weese, E. Saloux, H. Delingette, M. Sermesant, and J. D'hooge, "A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 7, pp. 1436–1451, July 2015.

[24] S. Marchesseau, H. Delingette, M. Sermesant, M. Sorine, K. Rhode, S. Duckett, C. Rinaldi, R. Razavi, and N. Ayache, "Preliminary specificity study of the Bestel-Clement-Sorine electromechanical model of the heart using parameter calibration from medical images," *J Mech Behav Biomed Mater.*, vol. 20, pp. 259–271, 2013.

[25] H. Gao, H. F. Choi, P. Claus, S. Boonen, S. Jaecques, G. Van Lenthe, G. Van der Perre, W. Lauriks, and J. D'hooge, "A fast convolution-based methodology to simulate 2D/3D cardiac ultrasound images," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 56, no. 2, pp. 404–409, 2009.

[26] K. Negishi, T. Negishi, K. Kurosawa, K. Hristova, B. A. Popescu, D. Vinereanu, S. Yuda, and T. H. Marwick, "Practical guidance in echocardiographic assessment of global longitudinal strain," *JACC: Cardiovascular Imaging*, vol. 8, no. 4, pp. 489 – 492, 2015.

[27] P. Reant, L. Barbot, C. Touche, M. Dijos, F. Arsac, X. Pillois, M. Landelle, R. Roudaut, and S. Lafitte, "Evaluation of global left ventricular systolic function using three-dimensional echocardiography speckle-tracking strain parameters," *Journal of the American Society of Echocardiography*, vol. 25, no. 1, pp. 68 – 79, 2012.

[28] S. A. Kleijn, M. F. Aly, C. B. Terwee, A. C. van Rossum, and O. Kamp, "Reliability of left ventricular volumes and function measurements using three-dimensional speckle tracking echocardiography," *European Heart Journal - Cardiovascular Imaging*, vol. 13, no. 2, pp. 159–168, 2012.

[29] M. Alessandrini, B. Heyde, S. Cygan, M. Sermesant, H. Delingette, O. Bernard, M. De Craene, and J. D'hooge, "Elastic registration vs. block matching for quantification of cardiac function with 3D ultrasound: Initial results of a direct comparison in silico based on a new evaluation pipeline," in *Ultrasonics Symposium (IUS), 2014 IEEE International*, Sept 2014, pp. 608–611.

[30] Y. Zhou, O. Bernard, E. Saloux, A. Manrique, P. Allain, S. Makram-Ebeid, and M. D. Craene, "3D harmonic phase tracking with anatomical regularization," *Medical Image Analysis*, 2015, in press.

[31] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.

[32] B. Heyde, D. Barbosa, P. Claus, F. Maes, and J. Dhooge, "Three-dimensional cardiac motion estimation based on non-rigid image registration using a novel transformation model adapted to the heart," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, ser. Lecture Notes in Computer Science, O. Camara, T. Mansi, M. Pop, K. Rhode, M. Sermesant, and A. Young, Eds. Springer Berlin Heidelberg, 2013, vol. 7746, pp. 142–150.

[33] S. Queirós, J. L. Vilaça, P. Morais, J. C. Fonseca, J. D'hooge, and D. Barbosa, "Fast left ventricle tracking in CMR images using localized anatomical affine optical flow," in *Proc. SPIE Medical Imaging*, vol. 9413, 2015, pp. 941 306–941 306–7.

[34] D. Barbosa, T. Dietenbeck, B. Heyde, H. Houle, D. Friboulet, J. Dhooge, and O. Bernard, "Fast and fully automatic 3-d echocardiographic segmentation using b-spline explicit active surfaces: Feasibility study and validation in a clinical setting," *Ultrasound in Medicine & Biology*, vol. 39, no. 1, pp. 89 – 101, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301562912004723

[35] O. Bernard, J. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier, P. Jodoin, J. Domingos, R. Stebbing, K. Keraudren, O. Oktay, J. Caballero, W. Shi, D. Rueckert, F. Milletari, S. Ahmadi, E. Smistad, F. Lindseth, M. van Stralen, C. Wang, O. Smedby, E. Donal, M. Monaghan, A. Papachristidis, M. Geleijnse, E. Galli, and J. Dhooge, "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography," *Medical Imaging, IEEE Transactions on*, 2015, Epub. ahead of print.

[36] S. Queirós, D. Barbosa, J. Engvall, T. Ebbers, E. Nagel, S. I. Sarvari, P. Claus, J. C. Fonseca, J. L. Vilaça, and J. D'hooge, "Multi-centre validation of an automatic algorithm for fast 4D myocardial segmentation in cine cmr datasets," *European Heart Journal - Cardiovascular Imaging*, 2015.

[37] C. Jia, T. Kolias, J. Rubin, P. Yan, A. Sinusas, D. Dione, J. Duncan, Q. Wei, K. Thiele, L. Huang, S.-W. Huang, and M. O'Donnell, "3D elasticity imaging on an open-chest dog heart," in *Ultrasonics Symposium (IUS), 2009 IEEE International*, Sept 2009, pp. 155–158.

[38] J. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 8, pp. 987–1010, Aug 2006.

[39] B. Chakraborty, B. Heyde, M. Alessandrini, and J. D'hooge, "Fast myocardial strain estimation from 3D ultrasound through elastic image registration with analytic regularization," in *Proc. SPIE Medical Imaging*, accepted.

[40] M. Alessandrini, A. Basarab, L. Boussel, X. Guo, A. Serusclat, D. Friboulet, D. Kouame, O. Bernard, and H. Liebgott, "A new technique for the estimation of cardiac motion in echocardiography based on transverse oscillations: A preliminary evaluation in silico and a feasibility demonstration in vivo," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 5, pp. 1148–1162, May 2014.

[41] E. E. Konofagou and J. Provost, "Electromechanical wave imaging for noninvasive mapping of the 3D electrical activation sequence in canines and humans in vivo," *Journal of Biomechanics*, vol. 45, no. 5, pp. 856 – 864, 2012, special Issue on Cardiovascular Solid Mechanics.

[42] C. Compas, E. Wong, X. Huang, S. Sampath, B. Lin, P. Pal, X. Papademetris, K. Thiele, D. Dione, M. Stacy, L. Staib, A. Sinusas, M. O'Donnell, and J. Duncan, "Radial basis functions for combining shape and speckle tracking in 4D echocardiography," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 6, pp. 1275–1289, June 2014.

[43] C. Parsai, B. Bijnens, G. R. Sutherland, A. Baltabaeva, P. Claus, M. Marciniak, V. Paul, M. Scheffer, E. Donal, G. Derumeaux, and L. Anderson, "Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms," *European Heart Journal*, vol. 30, no. 8, pp. 940–949, 2009.

[44] C. Yodwut, L. Weinert, B. Klas, R. M. Lang, and V. Mor-Avi, "Effects of frame rate on three-dimensional speckle-trackingbased measurements of myocardial deformation," *Journal of the American Society of Echocardiography*, vol. 25, no. 9, pp. 978 – 985, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0894731712004518

[45] M. Alessandrini, B. Heyde, L. Tong, O. Bernard, and J. D'hooge, "Tracking quality in plane-wave versus conventional cardiac ultrasound: A preliminary evaluation in-silico based on a state-of-the-art simulation pipeline," in *Ultrasonics Symposium (IUS), 2015 IEEE International*, Oct 2015, pp. 1–4.

[46] M. Sermesant, R. Chabiniok, P. Chinchapatnam, T. Mansi, F. Billet, P. Moireau, J. Peyrat, K. Wong, J. Relan, K. Rhode, M. Ginks, P. Lambiase, H. Delingette, M. Sorine, C. Rinaldi, D. Chapelle, R. Razavi, and N. Ayache, "Patient-specific electromechanical models of the heart for the prediction of pacing acute effects in crt: A preliminary clinical validation," *Medical Image Analysis*, vol. 16, no. 1, pp. 201 – 215, 2012.

[47] S. Marchesseau, H. Delingette, M. Sermesant, M. Sorine, K. Rhode, S. Duckett, C. Rinaldi, R. Razavi, and N. Ayache, "Preliminary specificity study of the bestelclmentsorine electromechanical model of the heart using parameter calibration from medical images," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 20, pp. 259 – 271, 2013.

[48] J. D'hooge, D. Barbosa, H. Gao, P. Claus, D. Prater, J. Hamilton, P. Lysyansky, Y. Abe, Y. Ito, H. Houle, S. Pedri, R. Baumann, J. Thomas, and L. P. Badano, "Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data," *European Heart Journal - Cardiovascular Imaging*, 2015.

[49] M. Alessandrini, B. Heyde, S. Giffard-Roisin, H. Delingette, M. Sermesant, P. Allain, O. Bernard, M. De Craene, and J. D'hooge, "Generation of ultra-realistic synthetic echocardiographic sequences to facilitate standardization of deformation imaging," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, April 2015, pp. 756–759.