A new methodology for local evaluation of cardiac MR segmentation algorithms accuracy

D. Nguyen[1], J-P. Vallée[1]

[1]Radiology, Geneva Universitary Hospital, Geneva, Geneva, Switzerland

A. INTRODUCTION / MOTIVATION: In the field of cardiac MR images segmentation, active contour models or *snakes*, have been extensively used, owing to promising results and to numerous extensions proposed to improve their performance. The objective of this work was to investigate two aspects that have been largely neglected in this field: the first one concerns the method of evaluation of such algorithms, which are usually based on averaged parameters (ejection fraction (EF), end-systolic-volume (ESV), end-diastolic volume (EDV) and left ventricular (LV) mass) and for which, local segmentation errors may be masked. The second aspect concerns the relative merits of the various external forces extensions proposed, since no attempt has been made to compare them in a systematic way. This work proposes a new metric for evaluating cardiac MR images segmentation algorithms, by assessing local discrepancy between computer-generated and observer's hand-outlined boundaries. This metric was applied on various external forces extensions, in order to compare their segmentation performances.

B. MATERIAL AND METHODS:

B.1. Data acquisition: Data from 6 patients were analyzed. MR imaging was performed on a 1.5T Intera MR system (Philips Medical System, Best, Netherlands). Images were collected during breath-hold with ECG-gated balanced steady-state free precession (b-SSFP) sequences. For each patient, there were 8 short-axis slices along the LV long-axis, at both end-systolic and end-diastolic images, resulting in 96 short-axis views.

B.2. Image Analysis: Segmentation performances were established for a GVF snake (GVF), a pressure force-based snake (PF), and a guided pressure force-based snake (guided-PF) in comparison to the traditional snake (Trad) formulation. For the traditional model, the use of a pre-treatment with and without non-linear anisotropic filtering (Trad/Aniso) was also compared. For each subject, endocardial and epicardial contours have been traced manually by one observer in end-diastolic and end-systolic phases. All images segmentation have been performed by a single trained observer with the use of a self-made application written in Java. The initial contour was placed only on the first slice of each phase as an oval region-of-interest (ROI) and was the same for every segmentation scheme. For each segmentation model, parameters such as internal and external forces weights has been tuned in order to get the best segmentation results.

B.3. Regional error metric (REM): Segmentation accuracy of the various segmentation schemes, has been evaluated in comparison to manually traced contours with our REM for both endocardial and epicardial borders. The proposed metric is based on a method that allows to evaluate an average curve from various curves (1) and to establish one-to-one correspondences between the points of the curves to be averaged. The process is iterative and the average curve is found by computing the centroid of *N* corresponding points, which are initially the closest points between the curves to be averaged, and then become the intersection points with normals to the average curve. The process is iterated until the average curve does not change anymore, which is usually done in less than five iterations. The distance has been computed as the mean of the absolute Euclidian distance of corresponding points, which have been divided into 6 radial segments, as shown in figure 1. The same division was kept the same along the LV axis by propagating the heart orientation with polar coordinates.

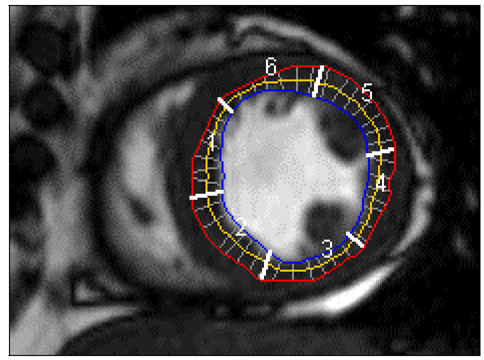

Fig 1: Local evaluation of segmentation accuracy with REM: average curve (yellow), manually traced contour (blue), semi-automated contour (red) and LV myocardium division (1-6).

C. RESULTS:

The use of REM permitted to confirm objectively the key difficulties in cardiac MR segmentation. For the endocardial border, the metric showed clearly that segmentation results were less accurate in myocardial lateral and anterior regions due to the presence of papillary muscles. For the epicardial border, the segmentation performances of every model were less accurate and there were no clear difference between myocardial segments as shown in the plots of figure 2 and 3. One-way analysis of variance (ANOVA) on REM results between the various segmentation schemes, did not reveal any statistically significant difference of performance ($P>0.05$) for the endocardial border. Concerning the epicardial border, statistically significant differences were found ($P<0.05$), with worst results with the use of GVF and with application of non-linear anisotropic filtering technique. Agreement between manual and semi-automated methods for global derived parameters was assessed with a two-tailed paired t-test. A good agreement was found for the EF estimation ($P>0.05$) for every segmentation schemes. The LV mass measurement was systematically overestimated ($P<0.05$), excepted with the use of non-linear anisotropic filtering pre-treatment. This finding was in disagreement with REM results, for which non linear anisotropic filtering gave the worst results for the epicardial border segmentation.

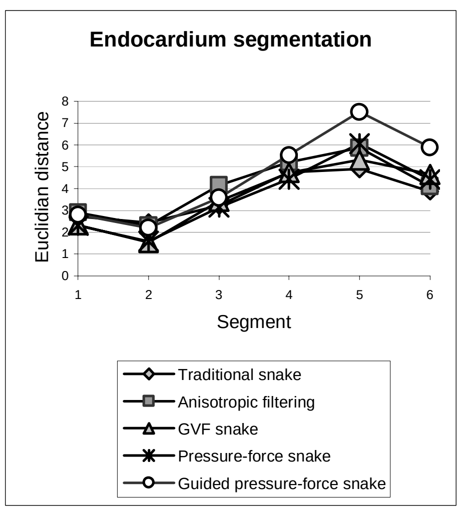

Fig 2: Results of REM application for endocardial border.

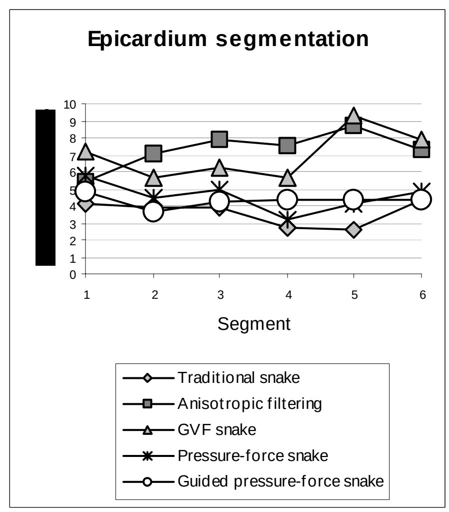

Fig 3: Results of REM application for epicardial border.

D: CONCLUSION:

We proposed a new methodology for assessing local performances of segmentation algorithms for cardiac MR images, which fitted more to subjective evaluation of their performances, and which showed that using derived parameters based on geometric assumptions could give wrong estimation on the real performances of the segmentation algorithm. While previous studies reported promising results through a good agreement between manual and semi-automated methods for the cardiac EF estimation, local assessment of segmentation accuracy showed that there was still a number of segmentation difficulties at both endocardial and epicardial levels, despite of the numerous proposed extensions of the traditional snake formulation, and particularly, despite of the good agreement found with the EF estimation. The use of such metric is an important step toward the acceptance and clinical use of cardiac segmentation algorithms.

REFERENCE: (1) Chalana V, Kim Y. A methodology for evaluation of boundary detection algorithms on medical images. IEEE Trans Med Imag 1997.

ACKNOWLEDGEMENT: This work was supported by the Swiss National Science Foundation.