An open, clinically-validated database of 3D+t cine-MR images of the left ventricle with associated manual and automated segmentations

Release 1.00

L. Najman[1], J. Cousty[1], M. Couprie[1], H. Talbot[1], S. Clément-Guinaudeau[2,3,4], T. Goissen[2,3,4] and J. Garot[2,3,4]

June 28, 2007

[1]Institut Gaspard-Monge, Laboratoire A2SI, Groupe ESIEE,
Cité Descartes, BP99, 93162 Noisy-le-Grand Cedex France
[2]Fédération de Cardiologie
[3] Unité INSERM 660, Faculté de Médecine de Créteil
[4]Mondor University Hospital, Assistance-Publique-Hôpitaux de Paris,
Créteil, France.

Abstract

In this paper, we describe a database of cine-MR (3D+t) images of the left ventricle. This database contains the voxel data, one automated and two manual segmentations for each sequence of images. The segmentations are validated from a clinical point of view. We detail how the images were obtained, as well as how the associated segmentations were performed. We also provide the data clinical validation process.

This database, including tools to compute quantitative measures and the software package used to obtain the automated segmentation, is freely available for research purposes.

The address of the site is `http://laurentnajman.org/heart`.

Contents

Key Words. Cardiac magnetic resonance, left ventricular function, left ventricular mass, myocardial infarction, image processing.

1 Introduction

The assessment of left ventricular (LV) function is performed routinely in the clinical field. It provides important prognostic information among patients with various cardiomyopathies. Cardiac magnetic resonance (CMR) can image the heart in arbitrary direction with excellent spatial and temporal resolution. This procedure yields full anatomic coverage and dynamic assessment of the heart throughout the cardiac cycle. Thanks to good image quality and contrast, as well as complete anatomic coverage of the LV, CMR has become a gold standard for the clinical assessment of LV function [3]. However, dynamic CMR in the cine mode yields a large amount of data. As a consequence, the clinical analysis of LV function from cine-MRI requires an interactive segmentation of adjacent 2D short-axis locations, as frequent manual corrections of myocardial contours are required. This is especially true in patients with segmental wall motion abnormality or deformed LV. Numerous authors (see for instance [9, 14, 10, 6, 8, 13, 7, 11]) have contributed in the development of accurate methods that aim to allow fully automated segmentation of the whole 3D cine-MRI dataset over time (i.e., 3D+t or 4D) for the assessment of LV function, volumes and mass.

One of the main problems in the assesment of those methods is the obtention of a database of cine-MRI and associated ground-truth segmentations that are also validated from a clinical point of view. Until now, to our best knowledge, no such base was available. This not only prevents the clinical validation of some of the pre-cited methods, but also precludes a fair comparison between the existing validated methods.

The goal of this paper is to propose a database freely available on the internet for research purposes. It contains cine-MR images of the LV, together with three associated segmentations: two hand-made segmentations – each one of them performed by an independent and blinded expert cardiologist – and one 4D automated segmentation. We also provide scripts and programs that compute each of the quantitative measures described in this paper, and the software package used to obtain the automated segmentation. The web-address of this database is `http://laurentnajman.org/heart`.

The outline of the paper is the following: we first describe the acquisition of the images. We then detail the methodologies for the three segmentations, concentrating on the manual ones. Finally we discuss the validity of the different segmentations, both quantitatively and qualitatively, by comparing them together, each one of them against the others.

2 MR images acquisition

Screened population patients referred to our Institutions for recent acute myocardial infarction (AMI) were prospectively screened regardless of the treatment received at the acute phase. To be included, patients had to

exhibit symptoms of AMI – i.e., chest pain, ST segment elevation of more than 1 mm in at least 2 contiguous leads of the ECG, and greater than double the normal elevation of creatine-kinase MB subfraction. In addition, the diagnosis of AMI was required to be confirmed by invasive coronary angiography with a clearly documented culprit epicardial coronary artery. If no contraindication to CMR was found, patients were scheduled to have CMR examination between day 2 and day 4 after AMI. The study protocol was approved by the Joint Committee of the Henri Mondor Medical Institutions and informed written consent was obtained from all patients.

In fine, 18 out of 25 patients from routine clinical practice were screened according to these guidelines. These patients had experienced a first AMI and had agreed to undergo subsequent CMR examination.

The patients were examined on a 1,5 T MR scanner (Magnetom Symphony®, Siemens, Erlangen, Germany) using 6-channel anterior and posterior phased array surface coil technology. Following a 3D fast gradient-echo localizer sequence, the long axis of the heart was located and dynamic cine-MR images of the heart were acquired in 2-chamber, 4-chamber, and 3-chamber views. From these, the short axis of the heart was located perpendicularly to the long axis of the LV. Contiguous short-axis slices of the LV were acquired from base to the apex encompassing the entire LV, through the use of repeated breath-held ECG-gated steady-state free precession sequence (SSFP) with typical imaging parameters as follow: 300-360 mm field of view, 2.1 ms TR, 1.6 ms TE, 60° flip angle, 6 mm slice thickness, no gap, image matrix 256x160, 30-40 ms temporal resolution.

The number of LV short-axis locations required to cover the entire LV by cine-CMR ranged from 9 to 14. The number of frames acquired during the entire cardiac cycle ranged from 22 to 37, depending on heart rate (49-91 bpm). The most basal slice included in the analysis was located just above the mitral valve within the LV cavity. To be included, the basal myocardium had to be visible in the entire circumference at end-systole. The most apical slice was chosen as the one with the smallest visible LV cavity at end-systole. Since the sequences are ECG-gated, the end-diastolic frame corresponds to the first image of the sequence.

3 Methodologies for segmentation

3.1 Preprocessing

For each patient, the cine MRI dataset consisted of a succession of contiguous (no gap) LV short-axis 2D planes that were successively imaged over time (2D+t). The sequences were registered to the heart-cycle, and could therefore be stacked in order to construct 3D sequences. Taken together, these different planes from base to apex were considered a 3D representation of the LV. The succession of these, over time, is a 3D+t representation of the LV. Before applying any segmentation procedure, the images were over-sampled in order to provide isotropic voxels. For each sequence of 3D+t images, a single mouse click on the center of the LV cavity at end-systolic time was recorded, and the images were cropped centered on the corresponding location. Typically, the size of each volume of the sequence represents $100 \times 100 \times 40$ voxels. When a misalignment of the different sections of a same volume was observed, a translation-only registration procedure was applied.

3.2 Automated 4D segmentation

The method for the 4D segmentation of the LV was recently developed and is described elsewhere [4] (a journal paper is in preparation). It automatically detects both the endocardial and epicardial borders of the LV for the whole 4D dataset, based on prior knowledge of the shape and appearance of the LV. The 4D

object isolated between the 2 borders is labeled as LV myocardium (LVM) and the 4D object isolated inside the endocardial border is labeled as LV cavity (LVC).

The computation time to automatically segment a whole 4D sequence ranged from 2 to 5 minutes on a conventional personal computer.

3.3 Manual segmentation

A conventional 2D manual segmentation of the LV myocardium was performed by two independent and blinded expert cardiologists. They used a software package that is well established for the post-processing of medical images (Analyze®, Biomedical Imaging Resource, Mayo Clinic Foundation, Rochester, MN). These two experts are called e_1 and e_2 in the sequel. The cine-MR dataset was analyzed as a succession of 2D LV short-axis planes. For each slice location, the experts manually overlaid the endocardial and epicardial contours both at end-diastolic and end-systolic times. During manual tracing, papillary muscles and LV trabeculae were included within the LV myocardium. Then, the segmented slices were stacked to rebuilt a 3D object for quantification.

The time spent by the experts to manually segment one volume of a 3D+t sequence ranged from 15 to 20 minutes. For this reason, a manual segmentation is not available at every time-step.

4 Validation of the segmentations

In this section, we discuss the quality of the segmentations of the two experts and of the automated method, both from a quantitative and a qualitative point of view.

4.1 Quantitative assessment

In order to characterize the accuracy of all methods, we used two different kinds of measures. The first measure is relative to the mean distance between the surfaces extracted from the automated segmentations and from the manual segmentations. The second characterises the false positive and false negative volume of the segmentations. Then, we assess the ability of the automated method to produce reliable characteristics of the LV function via the computation of the ejection fraction and of the myocardium mass.

Accuracy: point-to-surface measurement

Given two surfaces ∂X and ∂Y represented by two sets of polygons, the *point-to-surface measurement* (P2S) between ∂X and ∂Y estimates the mean distance between the vertices of ∂X and ∂Y (see [1]). A symmetrical measure is obtained by taking the maximum from the P2S between ∂X and ∂Y and the P2S between ∂Y and ∂X.

On our dataset, the endocardial and the epicardial borders were extracted from the segmentations by a marching cube algorithm[5]. The P2S was computed from the segmentations obtained by the automated method and the two experts. In order to evaluate the inter-observer variability the P2S between the two experts is also provided. Table 1 presents the mean and standard deviation of these measures at end-diastolic time and end-systolic time. We note that, in all cases in Table 1, the P2S is less than 1 voxel. The automated method achieved a mean P2S of 1.51mm $\pm$ 0.38 for the endocardial border and a mean P2S of 1.81mm $\pm$ 0.43

Table 1: Details of the point to surface measurements from the results of the various segmentation methods (mean point to surface measurements expressed in mm $\pm$ standard deviation).

	soft. vs. e_1	soft. vs. e_2	e_1 vs. e_2
End-diastolic time			
Endocardial border	1.52 ± 0.35	1.67 ± 0.43	1.37 ± 0.47
Epicardial border	2.04 ± 0.35	1.68 ± 0.39	1.23 ± 0.41
End-systolic time			
Endocardial border	1.50 ± 0.41	1.35 ± 0.34	1.15 ± 0.41
Epicardial border	1.90 ± 0.56	1.61 ± 0.41	1.31 ± 0.83

for the epicardial border. These results compare favourably with those obtained by other groups on their own datasets. Furthermore, the P2S between automated and manual segmentations is in the same range as the inter-observer P2S. This is a strong indication that the automated method produces as satisfying a segmentation as either manual one.

Although the accuracy of the methods is assessed by the P2S, these measures do not precisely describe the quality of the produced segmentations. In particular, the relative importance of the misclassified objects – a parameter which becomes crucial while quantifying the volume of the different objects – is not handled by the point-to-surface measurements.

Accuracy: False Negative/Positive Volume Fraction

In order to better characterise the accuracy of the segmentation methods, we also used the two following measures preconized by J. Udupa *et al.* in [12]. Let Y be a subset of the image voxels considered as the reference segmentation and let X be the segmentation which is to be evaluated. We set

$$FNVF(X,Y) = \frac{|Y \setminus X|}{|Y|} \quad \text{and} \quad FPVF(X,Y) = \frac{|X \setminus Y|}{|Y|}.$$

These measures are expressed as a fraction of the volume of "true" delineation. The $FNVF$ (False Negative Volume Fraction) indicates the fraction of tissue that was missed and $FPVF$ (False Positive Volume Fraction) denotes the amount of tissue falsely identified as a fraction of the total amount in the "true" delineation.

For each of the 18 patients, we computed $FNVF$ and $FPVF$ for e_2 against e_1, automated method against e_1, e_1 against e_2, and automated method against e_2.

These measures were computed for three objects: LVC, LVM and for $LVCM$ – the union of LCM and LVM. Table 2 presents the mean and standard deviation of these measures at end-diastolic and end-systolic times for the 18 datasets. We remark that error rates between the two experts and between experts and software are in the same range. In the task of segmenting LVC, $LVCM$ and LVM, the inter-expert errors are comparable with the errors between software and expert segmentations. However, we observe a tendency of the automated method to underestimate LVC and $LVCM$ with respect to the experts. Indeed, the false negatives are 1.5 to 4.5 times greater than the false positives. From a qualitative study described later, the two experts came to the conclusion that the automated contours were generally better localized than the manual ones. The apparent under-estimations of LVC and $LVCM$ can, thus, be seen as a side effect of the manual segmentation process. In Section 4.2, we explain some of the bias due to the manual segmentation process.

Table 2: Mean and standard deviation of FNVP, FPVF for all segmentations of *LVC*, *LVCM* and *LVM* at end-diastolic and end-systolic-time [see text].

End-diastolic time

		e_2 vs e_1	soft. vs e_1
LVC	FNVF	0.06 ± 0.03	0.13 ± 0.05
	FPVF	0.07 ± 0.08	0.03 ± 0.02
LVCM	FNVF	0.06 ± 0.01	0.09 ± 0.02
	FPVF	0.03 ± 0.06	0.03 ± 0.02
LVM	FNVF	0.20 ± 0.04	0.22 ± 0.04
	FPVF	0.10 ± 0.03	0.20 ± 0.06

		e_1 vs e_2	soft. vs e_2
LVC	FNVF	0.07 ± 0.05	0.14 ± 0.06
	FPVF	0.07 ± 0.02	0.03 ± 0.01
LVCM	FNVF	0.02 ± 0.05	0.06 ± 0.03
	FPVF	0.07 ± 0.01	0.04 ± 0.02
LVM	FNVF	0.11 ± 0.03	0.15 ± 0.04
	FPVF	0.22 ± 0.06	0.25 ± 0.10

End-systolic time

		e_2 vs e_1	soft. vs e_1
LVC	FNVF	0.10 ± 0.04	0.16 ± 0.06
	FPVF	0.06 ± 0.03	0.05 ± 0.03
LVCM	FNVF	0.07 ± 0.01	0.07 ± 0.03
	FPVF	0.02 ± 0.02	0.05 ± 0.02
LVM	FNVF	0.14 ± 0.04	0.14 ± 0.04
	FPVF	0.08 ± 0.03	0.16 ± 0.07

		e_1 vs e_2	soft. vs e_2
LVC	FNVF	0.06 ± 0.03	0.13 ± 0.05
	FPVF	0.11 ± 0.05	0.06 ± 0.03
LVCM	FNVF	0.02 ± 0.02	0.04 ± 0.02
	FPVF	0.07 ± 0.02	0.06 ± 0.02
LVM	FNVF	0.09 ± 0.03	0.09 ± 0.02
	FPVF	0.15 ± 0.05	0.18 ± 0.06

Assessment of critical parameters

Left ventricular ejection fraction (EF) and left ventricular myocardium mass (MM) are critical parameters for cardiac diagnosis and remodeling prevention. Their estimation is routinely used by cardiologists. The EF is the amount of blood ejected during a heart cycle expressed as a fraction of the tele-diastolic volume. In our dataset the EF (resp. MM) range was 20-75% (resp. 94-197 g).

From the segmented images, the EF can be computed by $(|LVC_{\max}| - |LVC_{\min}|)/|LVC_{\max}|$, where $|LVC_{\max}|$ (resp. $|LVC_{\min}|$) is the maximal (resp. minimal) volume of the left ventricular chamber along the heart cycle. Let X_p^o denote the measure of the parameter X performed by operator o for patient p, where $X \in \{EF, MM\}$, $o \in \{e_1, e_2, s\}$, and $p \in [1 \dots 18]$. We take $refX_p = (X_p^{e_1} + X_p^{e_2})/2$ as a reference value for the parameter X on patient p and we evaluate the deviation $\Delta X_p^o = |X_p^o - refX_p|/refX_p$. Notice that $\Delta X_p^{e_1} = \Delta X_p^{e_2}$. Over all 18 patients, the automated method achieved a mean deviation on the EF (resp. MM) of 0.032 (resp. 0.050) whereas the experts only achieved 0.055 (resp. 0.052). Furthermore, we observe that ΔEF_p^s (resp. ΔMM_p^s) is less than $\Delta EF_p^{e_1}$ (resp. $\Delta MM_p^{e_1}$) in 8 (resp. 9) of the 18 patients. In other words, the deviation on the EF (resp. MM) achieved by the automated method is less than the deviation achieved by the experts in 8 (resp. 9) of the 18 patients. We conclude that the automated tool produces reliable assessment of left ventricular functional parameters comparable to the experts'. A statistical analysis of these data – including linear regression and Bland-Altman plots – can be found on the web site http://laurentnajman.org/heart.

4.2 Qualitative assessment: manual segmentation and inter-expert variability

We now analyze some features of all segmentation methods which cannot be exhibited solely from a quantitative assessment. In Fig. 1, we present a view of the segmentations obtained by the two experts and by the automated method for a sample 3D image. We observe that the *LVM* segmented by e_1 is significantly thinner than the one segmented by e_2. We also remark that in location A, expert e_1 did not recognize a part

of the papillar muscle which was however segmented by e_2 and by the automated method. In location B, the myocardium segmented by e_1 presents a hole, which is not compatible with anatomy. We note that the automated method avoids such situations by construction and we observe that e_2's segmentation has no hole. This topological consideration has however no ill-effect on any of the quantitative measures.

Several explanations can be given. On the one hand, expert segmentations were realized exclusively on 2D images, which introduces a bias. Neither the spatial coherency between successive 2D sections of a 3D image, nor temporal coherency with the previous and next images of the sequence were taken into account. On the other hand, the precise delineation of a contour, pixel by pixel, is a very demanding task for human operators. It is well known that, in general, human operators can outperform computerized procedure in recognition tasks, whereas algorithms can often perform better than humans at delineation. Finally, we point out that there is no standardized procedure for manual segmentation of *LVM* in cardiac MR-images, contrary to what can be the case for other segmentation tasks in other modalities [2]. Therefore, the physicians who made this evaluation all believe that the automated method generally outperforms manual segmentation.

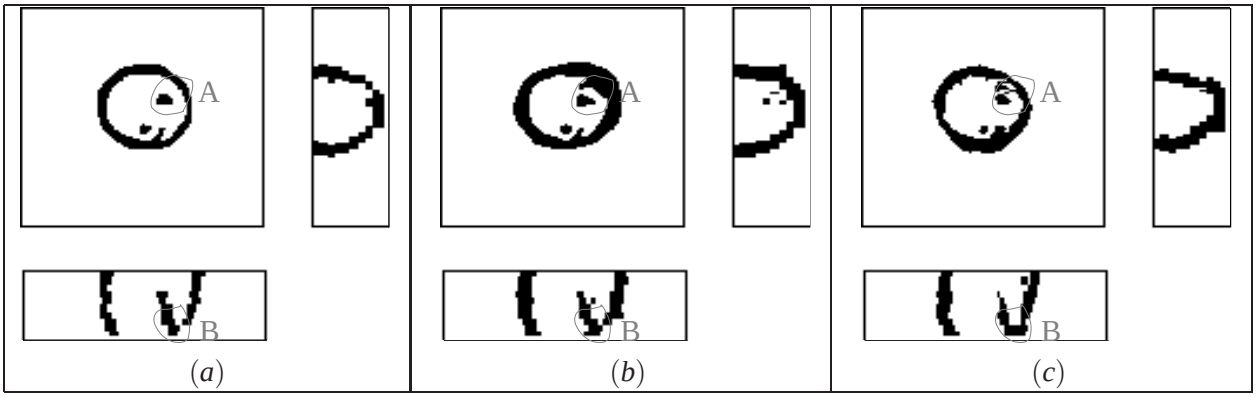

Figure 1: Examples of segmentations performed by e_1 (a), e_2 (b) and the automated method (c).

5 Conclusion

We describe in this paper the acquisition and analysis of a database of cine-MRI of the LV, with associated clinically-validated segmentations. The number of patients in the database may superficially appear to be small but, in actual fact, the number of contours processed by the tested techniques corresponds to 4752 endocardial and 4752 epicardial borders drawn from 216 LV short-axis slices, all performed by specialist cardiologists. The measured values obtained with the automated 4D analysis were found to be comparable with those obtain by the manual 2D+t methods. In this patient population, there was no other standard for comparison with true measurements of LV mass or volumes. The 4D analysis shows no significant difference in the clinical practice with values obtained from the interactive analysis of successive 2D locations.

This database is thus validated from a clinical point of view. Freely available for research purposes on http://laurentnajman.org/heart, it is a first step towards a fair evaluation and comparison of LV-segmentation methods. In the future, we plan to enrich this database with subsequent images and other modalities, as soon as these data are validated from a clinical point of view. We also plan to include anatomo-pathological animal data (using rabbits and pigs).

The authors would like to emphasize that such a work is only possible if cardiologists and computer scientists

are working together in close partnership.

References

[1] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. MESH: measuring errors between surfaces using the Hausdorff distance. In *IEEE International Conference in Multimedia and Expo*, volume 1, pages 705–708, 2002. 4.1

[2] Leonardo Bonilha, Eliane Kobayashi, Fernando Cendes, and Li Min-Li. Protocol for volumetric segmentation of temporal structures using high-resolution 3-d magnetic resonance imaging. *Human Brain Mapping*, 2004. 4.2

[3] P. T. Buser, W. Auffermann, W. W. Holt, S. Wagner, B. Kircher, C. Wolfe, and C. B. Higgins. Noninvasive evaluation of global left ventricular function with use of cine nuclear magnetic resonance. *Journal of the American College of Cardiology*, 2(13):1294–1300, 1989. 1

[4] J. Cousty, L. Najman, M. Couprie, S. Clément-Guinaudeau, T. Goissen, and J. Garot. Automated, accurate and fast segmentation of 4d cardiac mr images. In F.B. Sachse and G. Seemann, editors, *FIMH: Procs. of Functional Imaging an Modeling of the Heart*, volume 4466 of *LLNCS*, pages 474–483, 2007. 3.2

[5] X. Daragon, M. Couprie, and G. Bertrand. Discrete frontiers. In *Discrete geometry for computer imagery*, volume 2886 of *Lecture Notes in Computer Science*, pages 236–245. Springer, 2003. 4.1

[6] Michael R. Kaus, Jens von Berg, Jürgen Weese, Wiro Niessen, and Vladimir Pekar. Automated segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis*, 8:245–254, 2004. 1

[7] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elington, R. H. Mohiaddin, and D. Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 10:286–303, 2006. 1

[8] J. Lötjönen, S. Kivistö, J. Koikkalainen, D. Smutek, and K. Lauerma. Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images. *Medical Image Analysis*, 8:371–386, 2004. 1

[9] Tim McInerney and Demetri Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Computerized Medical Imaging and Graphics*, 19(1):69–83, 1995. 1

[10] Steven C. Mitchell, Boudewijn P. F. Lelieveldt, Rob J. van der Geest, Hans G. Bosch, Johan H. C. Reiber, and Milan Sonka. Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images. *IEEE Trans. on Medical Imaging*, 20:415–423, 1997. 1

[11] Johan Montagnat and Hervé Delingette. 4D deformable models with temporal constraints: application to 4D cardiac image segmentation. *Medical Image Analysis*, 9:87–100, 2005. 1

[12] J.K. Udupa, V.R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L.M. Currie, B.E. Hirsch, and J. Woodbrun. A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30:75–87, 2006. 4.1

[13] H. van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ordàs, J. J. M. Westenberg, J. H. C. Reiber, and B. P. F. Lelieveldt. SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, 10:286–303, 2006. 1

[14] Rob J. van der Geest, Vincent G. M. Buller, Eric Jansen, Hildo J. Lamb, Leo H. B. Baur, Ernst E. van der Wall, Albert de Roos, and Johan H. C. Reiber. Comparison between manual and semiautomated analysis of left ventricular parameters from short-axis MR images. *Journal of Computer Assisted Tomography*, 21:756–765, 1997. 1