Left Ventricular Segmentation Challenge from Cardiac MRI: A Collation Study

Avan Suinesiaputra[1], Brett R. Cowan[1], J. Paul Finn[2], Carissa G. Fonseca[2], Alan H. Kadish[3], Daniel C. Lee[3], Pau Medrano-Gracia[1], Simon K. Warfield[4], Wenchao Tao[2], and Alistair A. Young[1]

[1] Auckland Bioengineering Institute, University of Auckland, New Zealand
[2] Department of Radiological Sciences, University of California Los Angeles, USA
[3] Division of Cardiology, Northwestern University, USA
[4] Computational Radiology Laboratory, Harvard Medical School, USA

Abstract. This paper presents collated results from the left ventricular (LV) cardiac MRI segmentation challenge as part of STACOM'11. Clinical cases from patients with myocardial infarction (100 test and 100 validation cases) were randomly selected from the Cardiac Atlas Project (CAP) database. Two independent sets of expert (manual) segmentation from different sources that are available from the CAP database were included in this study. Automated segmentations from five groups were contributed in the challenge. The total number of cases with segmentations from all seven raters was 18. For these cases, a ground truth "consensus" segmentation was estimated based on all raters using an Expectation-Maximization (EM) method (the STAPLE algorithm).

1 Introduction

In cardiac MRI, the LV segmentation is typically performed to derive important clinical indices such as LV mass and volume. The current clinical standard is manual contouring of the myocardial boundaries, a time consuming and error-prone process, requiring substantial training. The development of automated segmentation algorithms has been problematic due to the lack of "ground truth" in real clinical cases. Even expert manually drawn segmentations still suffer from inter- and intraobserver variability. This problem particularly applies in cardiac imaging, where the presence of papillary muscles, the heart dynamics, and soft tissue contrast variations are just some of the problematic areas in cardiac MRI.

In this segmentation challenge, we created a framework to solve this problem by providing the same data set to researchers to test their segmentation algorithms and also to estimate better set of ground truth segmentations at the same time. We applied the EM-based STAPLE method [6] to estimate the consensus ground truth segmentations. Therefore, the challenge was performed as a collaborative work rather than a competition. A large data set of clinical cardiac MRI cases was made available through the Cardiac Atlas Project[1] [2]. By using the

[1] `http://www.cardiacatlas.org`

O. Camara et al. (Eds.): STACOM 2011, LNCS 7085, pp. 88–97, 2012.

Table 1. Baseline characteristics of the data used in this challenge

	test set (N=100)	validation set (N=100)
EDV (ml)	193.86 (46.45)	199.44 (54.97)
ESV (ml)	113.41 (43.44)	123.80 (53.81)
LV mass (gr)	172.24 (42.57)	165.38 (40.30)
EF (%)	42.95 (10.88)	39.87 (11.25)
SV (ml)	80.41 (18.65)	75.59 (18.62)

EDV = endocardial volume at ED, ESV = endocadial volume at ES, EF = ejection fraction, and SV = stroke volume.

same data set, confounding difficulties to compare segmentation results between peers can therefore be eliminated.

2 Methods

2.1 Cardiac MRI Data

Cardiac MR images were randomly selected from the DETERMINE (Defibrillators To Reduce Risk by Magnetic Resonance Imaging Evaluation) cohort [4]. This study consists of patients with coronary artery disease and prior myocardial infarction. Two separate groups were defined as test (N=100) and validation (N=100) groups, by random selections (see Table 1). Cine MR images in short-axis and long-axis views were selected for this challenge. These MR images were acquired by using a Steady-State Free Precession (SSFP) pulse sequence. MRI parameters varied between cases, giving a heterogenous mix of scanner types and imaging parameters consistent with typical clinical cases.

2.2 Raters

Five automated raters (SCR, INR, DS, AO and EM) and two expert raters (AU and NU) participated in this study. Rater descriptions are given in Table 2. Two raters (SCR and INR) were fully automatic, although SCR required repositioning the center of LV segmentation in four cases. Three raters required some manual interactions, either by drawing initial contours (DS, EM and AO) or by having some parameter initialization (EM). One rater (INR) used the test dataset to train the algorithm, the others did not.

The expert NU rater was a manually drawn myocardial contour, traced by the DETERMINE MRI core laboratory using QMass software (Medis, Leiden, the Netherlands), while the AU rater was an expert-guided interactive customization of a finite element heart model using Cardiac Image Modeller (CIM) software (AMRG, Auckland, New Zealand). To generate the intersection between cardiac MRI with the 3D AU models and the image planes, the CAPClient software was used[2].

[2] The CAPClient is an open source software, available for download at `http://www.cardiacatlas.org/web/guest/tools`