Optimizing the automatic segmentation of the left ventricle in magnetic resonance images

E. Angelié, a) P. J. H. de Koning, M. G. Danilouchkine, H. C. van Assen, G. Koning, R. J. van der Geest, and J. H. C. Reiber
Leiden University Medical Center, Department of Radiology, C2S Division of Image Processing, P.O. Box 9600, 2300 RC Leiden, The Netherlands

(Received 9 September 2004; revised 5 November 2004; accepted for publication 5 November 2004; published 14 January 2005)

Automatic segmentation of the left ventricular (LV) myocardial borders in cardiovascular MR (CMR) images allows a significant speed-up of the procedure of quantifying LV function, and improves its reproducibility. The automated boundary delineation is usually based on a set of parameters that define the algorithms. Since the automatic segmentation algorithms are usually sensitive to the image quality and frequently depend heavily on the acquisition protocol, optimizing the parameters of the algorithm for such different protocols may be necessary to obtain optimal results. In other words, using a default set of parameters may be far from optimal for different scanners or protocols. For the MASS-software, for example, this means that a total of 14 parameters need to be optimized. This optimization is a difficult and labor-intensive process. To be able to more consistently and rapidly tune the parameters, an automated optimization system would be extremely desirable. In this paper we propose such an approach, which is based on genetic algorithms (GAs).

The GA is an unsupervised iterative tool that generates new sets of parameters and converges toward an optimal set. We implemented and compared two different types of the genetic algorithms: a simple GA (SGA) and a steady state GA (2SGA). The difference between these two algorithms lies in the characteristics of the generated populations: “nonoverlapping populations” and “overlapping populations,” respectively. “Nonoverlapping” population means that the two populations are disjoint, and “overlapping” means that the best parameters found in the previous generation are included in the present population. The performance of both algorithms was evaluated on twenty routinely obtained short-axis examinations (eleven examinations acquired with a steady-state free precession pulse sequence, and nine examinations with a gradient echo pulse sequence). The optimal parameters obtained with the GAs were used for the LV myocardial border delineation. Finally, the automatically outlined contours were compared to the gold standard—manually drawn contours by experts. The result of the comparison was expressed as a degree of similarity after a processing time of less than 72 h to a 59.5% of degree of similarity for SGA and a 66.7% of degree of similarity for 2SGA. In conclusion, genetic algorithms are very suitable to automatically tune the parameters of a border detection algorithm. Based on our data, the 2SGA was more suitable than the SGA method. This approach can be generalized to other optimization problems in medical image processing. © 2005 American Association of Physicists in Medicine. [DOI: 10.1118/1.1842912]

Key words: optimization, automatic segmentation, MR cardiac images, genetic algorithm

I. INTRODUCTION

Cardiac magnetic resonance (MR) imaging is playing an increasingly important role for anatomical and functional assessment of the cardiovascular system. An accurate delineation of the endocardial and epicardial boundaries is important to quantify the left ventricular (LV) dimensions. Manual segmentation requires expert knowledge and is a time-consuming procedure, which limits the routine clinical use of cardiovascular MR. Moreover, the manual segmentation is observer dependent and therefore is associated with considerable inter- and intra-observer variability. Automated and semi-automated contour detection techniques have been developed in order to overcome the disadvantages of manual contour drawing, but the robustness of the currently available contour detection algorithms is still not optimal for routine use. At our laboratory, we have developed a cardiac MRI analysis software package, MASS, which includes automated contour detection.1 A major challenge when designing and implementing a reliable automated contour detection algorithm is to deal with the large variations in image characteristics due to differences in MR pulse sequences used, the usage of different receiver coils and differences between MR scanners from different vendors. Consequently, for optimal performance, the automated segmentation method needs to be optimized for a specific type of acquisition procedure.

To date, the commonly used method of optimizing an automatic segmentation algorithm is to manually tune the parameters of the segmentation algorithm until accurate results are reached. Generally, numerous iterations are required to obtain the optimal segmentation result, and there is no guar-
Antee that the optimal result will be achieved. To deduce the optimal value for a specific parameter, the underlying physical or geometrical constraints must be taken into consideration. But some parameters are less intuitive and therefore hard to find by means of ad-hoc search. Moreover, segmentation algorithms are often complex and the segmentation parameters are dependent upon each other (all the parameters have to be optimized together and not by means of a step-by-step analysis) and as a consequence hard to optimize manually on a trial and error basis.2

To automate the optimization, a brute-force approach could be utilized. By such approach, all possible permutations of parameters would be taken into consideration. Taking into consideration the complexity of segmentation algorithms and the number of parameters to be optimized, this approach appears to be computationally and time expensive, thus infeasible.

The aim of this work was to design an automated optimization system based on unsupervised tools that can generate new parameter sets in order to automatically tune an automated contour detection algorithm. Previous papers presented an adaptation of the genetic algorithm (GA),3 the parallel genetic algorithm (PGA), as an intelligent parameter optimization tool.4 The PGA has already proved to be an effective method to optimize dynamic model-based contours5 and to improve marker placement.6 Consequently, the main issue raised in this work was to study the ability of the GA to solve the optimization problem faced in our border detection algorithms. A number of optimization techniques exist already and differ in their capability to find a global optimum, which is often hidden among local optima, and by their speed of convergence toward the global optimum.7–9 These algorithms are employed for finding a global optimal solution in a high-dimensional search space. Finding a global optimal solution is often difficult due to noise or perturbations, and GA-based methods have shown to be capable of finding a robust solution.8 Therefore, GAs appear to be valuable candidates for the implementation of an optimization method.3–10

Because the segmentation performance varies according to the pulse sequence used, we assumed that the optimized parameter setting for the segmentation of images acquired with different pulse sequences should be different too. In order to test the ability of a GA to solve the optimization problem faced, we compared the optimized parameter settings found using a GA on different images acquired with different pulse sequences.

The number of GAs existing in the literature makes the choice of the best optimization algorithm difficult. Thus, the second issue raised in this paper was therefore which GA should be used in the optimization faced. Since most of the GAs can be divided into two different classes, two GAs were tested. The one using overlapping populations, the steady state GA (2SGA)11 and the other one using nonoverlapping populations, the so-called simple GA (SGA).

II. MATERIALS AND METHODS

A. Material

For the reason of limiting the computation time required, and because this work aimed at examining the ability of the GA to solve our optimization problem, the study population was limited to 20 patients with a history of cardiovascular disease [with a wide range of ejection fraction (4%–78%)]. The study material consisted of eleven MR steady-state free processing (SSFP, TR=3.2 ms, TE=1.60 ms, flip angle =60°) and nine MR gradient echo (GRE, TR=50 ms, TE =4.80 ms, flip angle=20°) (3 females and 17 males with a mean age of 63 years) short axis examinations of the cardiac LV. MR images were acquired on a 1.5T MR system (Siemens Sonata). A temporal resolution or phase interval varying between 46 and 50 ms was chosen to ensure that the end-systolic phase was properly defined. The images were acquired with a FOV of 400 mm and reconstructed to a 256×256 image matrix. Slice thickness varied between 8 and 10 mm and the number of slices varied between 10 and 19; the number of cardiac phases between 13 and 26 depending on the heart rate. Image analysis was limited to the end-diastolic (ED) and end-systolic (ES) phases. Endocardial contours were drawn by an expert using MASS version 5.0 following a predefined protocol as described previously.1

(1) Window and level settings were standardized and kept unchanged for the entire study;
(2) ED and ES phases were chosen after viewing the images in the movie mode;
(3) The uppermost slice showing more than 50% of the circumference of the ventricular myocardium was defined as the basal slice;
(4) Papillary muscle and trabeculations were treated as being a part of the blood pool to be able to assess the wall thickness and wall thickening of the cardiovascular system.

B. Method

In this paper we tested whether the GA was suitable to find the optimal setting of the segmentation parameters used in MASS software package. We focused only on the optimization of the algorithm for automated detection of the endocardial contours of the cardiac LV.

The GA is an unsupervised tool that can generate new settings of segmentation parameters based on a sample of initial parameters sets. Each parameter set is called an individual. The GA starts with randomly chosen individuals, called population, including a fixed number of individuals. Next, the GA creates a new population of individuals by mutating and mating the best individuals from the previous population and producing new offspring. The value of each parameter is encoded in a binary string and the mutation operator flips the bits with a given probability \( p_{\text{mut}} \).11 For instance, for a probability of 0.2, 20% of the zeroes and ones in the binary string will be mutated in ones and zeroes, re-
spectively. Then the crossover operator (with probability $p_{\text{cross}}$) is applied to the string population for generating new individual from two binary strings.\textsuperscript{11,12}

The first step of this study was to evaluate whether or not this tool could solve the optimization of the segmentation algorithm used in MASS. Next, we compared the optimization result of two different GAs in order to define the optimization tool to use in the scope of our study.

1. GA’s capability to solve the MASS optimization problem

a. Segmentation parameters. The automated contour detection algorithm implemented in MASS was described previously.\textsuperscript{1} It follows a three-step procedure as illustrated in Fig. 1. To determine the middle point of the left ventricle, a Hough transform is applied to the images, resulting in an identically sized parameter image, with high values near the center point of the LV having a radius within a certain range defined by two parameters: minimum and maximum radius. Then, the model determination step consists of a first segmentation of the blood pool area using an adaptive thresholding technique. Two different thresholding techniques were implemented: The first one consists of generating radial scan lines emanating from the detected center and collecting for each scan the gray value of the pixel with the highest edge value within the minimum and maximum radius. The second one is based on generating iso-intensity contours surrounding the center point of the LV and considering the one with the highest edge strength. The choice of which thresholding method to use is made by one parameter. The mean gray value of these edge pixels weighted by a percentage parameter is then defined as the threshold value. Next, a smooth convex hull surrounding the blood pool area is determined. A total of four parameters (the knowledge about the LV geometry, the choice of the thresholding technique to use and a weighting percentage) are required during the model determination. In a third step, the final contour is obtained by refining the model contour using two successive minimal cost edge detection techniques.\textsuperscript{13} Given the previous model contour, the image data in a certain neighborhood around the model is spatially transformed into a rectangular array (so-called scan array), the width of which can vary. Then all the points in this array are transformed into a cost matrix. The cost matrix is obtained using a derivative operation equivalent in a computational point of view to a convolution with a kernel that has to be optimized. The cost values included in the matrix reflect the probability that the corresponding point in the image is part of the desired contour. All possible paths through the matrix are evaluated for their total accumulated cost. Some constraints, such as the distance to look for a neighbor, or the distance toward the center of the matrix, or the orientation of the path are added to weigh the calculation of the minimal cost path. Five parameters had to be optimized for each minimal cost algorithm, resulting in a total of ten parameters optimizing in this latest step.

Since the output of each step is input for the following step in the algorithm, the optimal parameter settings for an individual step are dependent upon the parameter setting of the preceding steps. All the segmentation parameters are dependent upon the image characteristics. The differences in the pulse sequences cause discrepancies in, e.g., the brightness of the blood pool and the sharpness of the contours, which are crucial for the segmentation. Therefore, it is necessary to optimize these segmentation parameters for each specific MR pulse-sequence.

b. Criterion to be optimized. Optimizing the automated contour detection algorithm requires a criterion to be optimized. The criterion considered in the optimization is called the fitness value and is an indicator of the match between the automatically detected contour and a manually drawn contour accepted as a gold standard. The aim of the fitness function is to quantitatively represent the performance of the algorithm, when using a set of segmentation parameters, called individual. We defined the fitness value as the degree of similarity $S$ between the automatically detected contour and the manually drawn contour. The degree of similarity is defined as the percentage of points that is similar between two contours\textsuperscript{14,15} in the following:

$$ S = \sum_{n=1}^{N} p_n(d), $$

where

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig1}
\caption{Flow chart illustrating the contour detection process and the inter-dependency of different steps and thus the segmentation parameters to optimize (A: endocardial model obtained on a SSFP MR image, B: final contour).}
\end{figure}
is more appropriate to our optimization process. We considered the 2SGA belonging to the “overlapping population” GA family. We expected that using this process, the optimization would converge faster than using a “nonoverlapping populations” type of GA. The output of MASS was used to assess the average degree of similarity \( \bar{S} \) between automatically detected contours corresponding to one individual and manually drawn contours. The best \( \bar{S} \) defined the best individual per generation. Each individual is tested on the entire set of exams. 20 individuals were tested per generation on two different collections of cardiac MR images, one consisting of eleven sets acquired with a SSFP sequence and the other consisting of nine sets acquired with a GRE sequence. Ten optimization runs were performed on the two collections of examinations using \( G_{\text{max}} = 100 \). The average and standard deviation of the optimal parameter sets and the best fitness value \( \bar{S} \) found before and after the optimization were compared for the two collections of images.

2. Choice of the GA to use in the MASS optimization process

a. Fitness study. In this work, we tested two GA implementations (SGA and 2SGA\(^{11}\)) to determine which of them is more appropriate for optimizing the segmentation algorithm used. The SGA uses nonoverlapping populations of individuals. At each generation the algorithm creates an entirely new population as it was described before. Contrary to SGA, 2SGA uses overlapping populations. In each generation the algorithm creates a temporary population of individuals, adds these to the previous population, then removes the worst individuals in order to reduce the population to its original size. The amount of overlap between generations can be specified by a replacement probability \( p_{\text{replace}} \). It corresponds to the percentage of the population that will be replaced at each new generation. The probability of mutation \( p_{\text{mut}} = 0.6 \), combined with a crossover probability \( p_{\text{cross}} = 0.05 \) was used with the SGA. For the 2SGA, a probability of mutation \( p_{\text{mut}} = 0.6 \), a crossover probability \( p_{\text{cross}} = 0.05 \), combined with a replacement probability \( p_{\text{replace}} = 0.9 \) was fixed.\(^{11}\) To test which of the two algorithms is more suitable to our optimization process, we analyzed the convergence of

\[ p_{\text{mut}}(d) = \begin{cases} 1 & \text{if } d \leq T \\ 0 & \text{if } d > T, \end{cases} \]

where \( d \) is the distance between each pair of corresponding points on the manually drawn contour and the automatically detected contour, \( N \) is the number of point per contour \( (N = 100) \), and \( T \) is a distance threshold; distances are calculated along the entire contour.\(^{16}\) Pairs of corresponding points are assumed to be similar if the distance does not exceed a certain threshold value \( T \) \((T = 2 \text{ mm})\), which approximately corresponds to the average distance \( \bar{d}_{\text{intra-observer}} \) between two contours drawn by the same observer.

c. Optimization procedure using the GA. The GA library used in the optimization was previously described by Wall.\(^{11}\) The optimization of the detection algorithm involves fourteen dependent parameters that constitute an individual. The optimization procedure was executed as a series of iterations, called generations. Each generation includes several individuals (called population) and can be described as follows:

Step 1: automatic run of MASS package software with one individual at a time on a set of MR examinations.

Step 2: calculate the average degree of similarity \( \bar{S} \) between manual and automatically detected contours corresponding to the particular individual.

Step 3: repeat steps 1 and 2 with all individuals included in the first population, thus creating the first generation.

Step 4: The GA creates a new population of individuals by mating the best individuals from the current population producing new offspring (Fig. 2). In each generation, the individuals are evaluated by calculating \( \bar{S} \).

Step 5: Determine whether the number of generations exceeds the preset maximum \( G_{\text{max}} \) \((G_{\text{max}} = 100 \text{ generations})\). If so, stop the iteration; otherwise go to Step 3.

As this process continues, the population converges toward better individuals defined by a higher \( \bar{S} \).

d. Performance analysis: Optimization procedure for different MR pulse sequences. The GA family includes two different types of algorithm: GAs using “nonoverlapping populations” and GAs using “overlapping populations.” To study the ability of GAs to solve the optimization problem faced, only one GA needs to be tested. Only if this first step is fulfilled, we can go further in this work and test which GA
the fitness value $\bar{S}$ for both optimization processes on a set of 11 SSFP MR examinations using an arbitrarily chosen stop criterion ($G_{\text{max}}=100$) for 50 segmentation parameter sets or individuals. The optimal fitness values found with the SGA and the 2SGA algorithms, respectively, were compared and the algorithm with the highest fitness value was concluded to be the most suitable to solve the optimization problem under consideration.

b. Convergence speed. The most difficult part of a probabilistic search method is to predict when the process is completed. Each individual run is different because of the probabilistic nature of the algorithm. In Bevilacqua et al.12 the evolution process stopped when the average of the six highest fitness values in a generation reached a plateau. In other words, the optimization process ends, when the fitness values and the standard deviation (S.D.) of the fitness values in a generation converge toward an asymptotic value. To compare the convergence speed of the two GAs (SGA and 2SGA), we analyzed the variation of the standard deviation between six highest fitness values per generation over the optimization processes.

C. Results

1. GA’s capability to solve the MASS optimization problem

The optimization of the automatic segmentation on different pulse sequence (SSFP and GRE, respectively) resulted in different segmentation parameter sets (Fig. 3). The average degree of similarity between automatically detected and manually drawn contours ($\bar{S}$) found with the segmentation algorithm before and after the optimization increased from $\bar{S}=44.8\%$ in an earlier study on MASS15 to $\bar{S}=59.3\%$ in the set of images acquired with GRE pulse sequence and from $\bar{S}=58.5\%$ to $\bar{S}=66.7\%$ in the set of images acquired with SSFP pulse sequence.

2. Choice of the GA to use in MASS optimization process

a. Fitness study. In order to test the performance of the algorithm we focused first on the optimal fitness value found after 100 generations on the set of SSFP examinations. With 2SGA ($p_{\text{mut}}=0.6$, $p_{\text{cross}}=0.05$, $p_{\text{replace}}=0.9$) the optimal fitness function reached a value of $\bar{S}=66.7\%$, whereas with the SGA ($p_{\text{mut}}=0.6$, $p_{\text{cross}}=0.05$) the optimal fitness value found was $\bar{S}=59.5\%$ (Fig. 4).

b. Test regarding the stop criteria. The performance of GAs is crucially dependent upon the loop termination criteria. We analyzed the evolution of the distribution of the six highest fitness values within a generation with the two different algorithms. Figure 5 clearly shows that the S.D. of the six highest fitness values increases when using the SGA method while no general trend of the S.D. is noticed when using the 2SGA.

III. DISCUSSION AND CONCLUSION

The goal of this work was to find a procedure to automatically optimize the segmentation parameters for the contour detection procedure in the MASS analytical software package. The optimization requires a total of 14 dependent segmentation parameters to be adjusted simultaneously. The evaluation of one parameter takes approximately 36 s. With the assumption that these parameters are allowed to take only five different values, a brute-force method would require testing $5^{14}$ sets of segmentation parameters (or individuals). Due to time limitations, a brute-force method for tuning the segmentation algorithm used in MASS would be entirely im-
The intra-observer study is considered as a gold standard for assessment of variability according to clinical analysis. In this study, the intra-observer degree of similarity between manual endocardial contours was found to be 71% on GRE image data sets and 77% for SSFP image data sets. After application of the 2SGA-based algorithm, the degree of similarity of the automatic segmentation algorithm is comparable to the intra-observer degree of similarity. Thus, the accuracy of the optimized segmentation algorithm is now comparable to the accuracy of manual segmentation. Therefore, we conclude that the optimization method developed using 2SGA is a promising method to automate the procedure of finding the optimal parameter setting for LV endocardial segmentation of MR images acquired with SSFP and GRE pulse sequences in MASS.

In this study, the need for automatic optimization of the segmentation algorithm used in the MASS software package...
has been demonstrated. The GA methods appeared to be suitable to fulfill the optimization task. This work demonstrated that the use of the 2SGA as an optimization method drastically increased the accuracy of the endocardial segmentation algorithm used in MASS. The MASS software package includes endocardial and epicardial contour detection algorithms. Since both of them need an initial set of parameters to run, the 2SGA is also an interesting tool to optimize the epicardial contour detection.

\[\text{Electronic mail: e.angelie@lumc.nl}\]


