Multistage Hybrid Active Appearance Model Matching: Segmentation of Left and Right Ventricles in Cardiac MR Images

Steven C. Mitchell, Boudewijn P. F. Lelieveldt, Rob J. van der Geest, Hans G. Bosch, Johan H. C. Reiber, and Milan Sonka*

***Abstract*—A fully automated approach to segmentation of the left and right cardiac ventricles from magnetic resonance (MR) images is reported. A novel multistage hybrid appearance model methodology is presented in which a hybrid active shape model/active appearance model (AAM) stage helps avoid local minima of the matching function. This yields an overall more favorable matching result. An automated initialization method is introduced making the approach fully automated.**

Our method was trained in a set of 102 MR images and tested in a separate set of 60 images. In all testing cases, the matching resulted in a visually plausible and accurate mapping of the model to the image data. Average signed border positioning errors did not exceed 0.3 mm in any of the three determined contours—left-ventricular (LV) epicardium, LV and right-ventricular (RV) endocardium. The area measurements derived from the three contours correlated well with the independent standard ($r = 0.96, 0.96, 0.90$), with slopes and intercepts of the regression lines close to one and zero, respectively. Testing the reproducibility of the method demonstrated an unbiased performance with small range of error as assessed via Bland–Altman statistic. In direct border positioning error comparison, the multistage method significantly outperformed the conventional AAM ($p < 0.001$). The developed method promises to facilitate fully automated quantitative analysis of LV and RV morphology and function in clinical setting.

***Index Terms*—Active appearance model, active shape model, cardiac segmentation, magnetic resonance image analysis.**

I. Introduction

CARDIOVASCULAR magnetic resonance imaging (MRI) is a highly flexible medical imaging modality suitable to assess cardiac function in a noninvasive manner. Particularly, multislice multiphase short-axis image views have shown to be highly useful to examine global and regional cardiac function. The heart is imaged at various stages of the cardiac cycle as a stack of slices perpendicular to the left-ventricular (LV) long axis. To quantitatively analyze global and regional cardiac function from such images, segmentation of the LV endocardial and epicardial borders is required, as well as segmentation of the right ventricle to assess right-ventricular (RV) functional defects. Due to the massive amount of data involved in a dynamic short-axis cardiac magnetic resonance (MR) examination (typically 200–300 images), robust automated segmentation of these structures is imperative for routine quantitative analysis of cardiac MRI.

Automated segmentation of cardiovascular MR images has shown to be a challenging task. Approaches dedicated to LV segmentation in MR and computed tomography (CT) volume data have been proposed, based on, among others, active contours and balloons [1]–[3], adaptive thresholding [4], dynamic programming [5], implicit wavefront propagation [6] or pixel/region classification [7], [8]. Though partially successful, three major problems are associated with many of the previously described contour detection strategies for cardiovascular structures:

- Because of the presence of noise and image acquisition artifacts in many routinely acquired cardiac MR images (especially in echo-planar imaging pulse sequences), image information can be ill defined, unreliable, or missing. In these cases a human observer is still capable of tracing a contour in the image data based on experience and prior knowledge, while many automated techniques fail. To overcome this problem, high-level knowledge about the image appearance, spatial organ embedding, characteristic organ shape, and its anatomical and pathological shape variations should form an integral part of a segmentation approach.
- A contour as drawn by an expert human observer may not always correspond to the location of the strongest local image evidence. In particular, in cardiovascular MR images, the papillary muscles and trabeculations pose a problem. Many experts draw the LV endocardial border as a convex hull around the blood pool, at a location somewhat outside of the strongest edge [9], [10]. A second example is the epicardial boundary, which can be embedded in fatty tissue, as a result of which the edge is strongest on the fat-air transitions. However, the contour should be drawn on the inside of this fatty layer, an intensity transition which is marked by only a faint edge. Therefore, a decision about the exact location of the contour cannot always be made based on the strongest

Manuscript received September 8, 2000; revised March 5, 2001. This work was supported in part by the Leiden University Medical Center during the honorary Boerhaave professorship of M. Sonka at the Leiden University in the spring of 1999 and by the Dutch Technology Foundation. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was A. Amini. *Asterisk indicates corresponding author.*

S. C. Mitchell is with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242 USA.

B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, and J. H. C. Reiber are with the Leiden University Medical Center, Department of Radiology, 2300 RC Leiden, The Netherlands.

*M. Sonka is with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242 USA (e-mail: milan-sonka@uiowa.edu).

Publisher Item Identifier S 0278-0062(01)04405-6.

0278–0062/01$10.00 ©2001 IEEE

image evidence, but should be learned from the examples provided by expert observers.

- Many segmentation techniques require user interaction to provide the initial conditions for a segmentation algorithm in the form of a seed point or initial boundary model. This may still be labor intensive, and is a potential source of interobserver and intraobserver variations in functional measurements.

Recently, Cootes and Taylor [11], [12] introduced a novel knowledge driven segmentation technique known as active appearance models (AAMs), which may solve these three problems mentioned above. The AAMs form an extension of the widely applied ASMs (ASMs) [13], [14]—statistical models of the distribution of a set of landmark points over a population of training samples. ASMs accommodate shape knowledge about object shape and its variations, augmented with local appearance knowledge in the vicinity of each landmark using separate appearance models for each landmark. In contrast, the AAMs represent the average object shape and its shape variations in combination with the appearance of a complete image patch in an integral statistical model. AAMs can be applied to segmentation by minimizing the error between a target image and a synthesized image patch generated from the model.

A major advantage of the AAMs over other segmentation methods is that the object shape and the underlying image appearance are derived via automated training from a set of segmentation examples. The AAMs are trained using shape examples manually drawn over the original image data. Consequently, they are able to capture the association between observer preferences and the underlying image evidence, making the AAMs highly suitable to model expert observer analysis behavior. Another advantage of the AAMs is the fact that multiple objects (in our case, the left and right cardiac ventricles) are modeled in their spatial embedding. The goal of this work was to develop a fully automated segmentation procedure for cardiac MR images by exploiting the clinical potential of AAMs.

The main contributions of this work as compared to Cootes' seminal work on AAMs are threefold. 1) A hybrid combination of ASMs and AAMs was developed. The model fitting procedure is driven by both ASM deformations and AAM deformations, whereas the error measure is based on AAMs. This decreases the risk of convergence toward a local minimum during the model matching. 2) A mechanism was developed to reliably initialize the segmentation process using the measure of model-image matching quality. 3) A method was developed allowing fully automated analysis of LV and RV morphology from MR image data.

II. Preliminaries

As stated earlier, AAMs are an extension of the well documented Point Distribution Models (PDMs) and ASMs introduced by Cootes and Taylor [13]. A brief overview of the PDM background is provided here. More detail is given elsewhere [5].

A. Modeling Shape Using ASMs

ASMs describe the shape and shape variation of a population by representing each shape sample as a set of n corresponding landmarks. By aligning the shape samples and applying a principal component analysis (PCA) on the sample distribution, any sample x within the distribution can be approximated by an average shape $\overline{x}$ with a linear combination of eigenvectors P superimposed

$$x \approx \overline{x} + Pb. \tag{1}$$

In two-dimensional images, the $2n$ eigenvectors P form the principal basis functions, while their corresponding eigenvalues provide a measure for the compactness of the distribution along each axis. By selecting the largest q eigenvalues, the number of eigenvectors can be reduced, where a proportion z of the total variance is described such that

$$\sum_{i=1}^{q} \lambda_i \geq z \cdot \text{Total} \quad \text{where Total} = \sum_{i=1}^{2n} \lambda_i. \tag{2}$$

Concatenation of the first q eigenvectors results in a compact shape model, which describes the average object shape in combination with its commonly occurring variation in a population.

B. Modeling Image Appearance

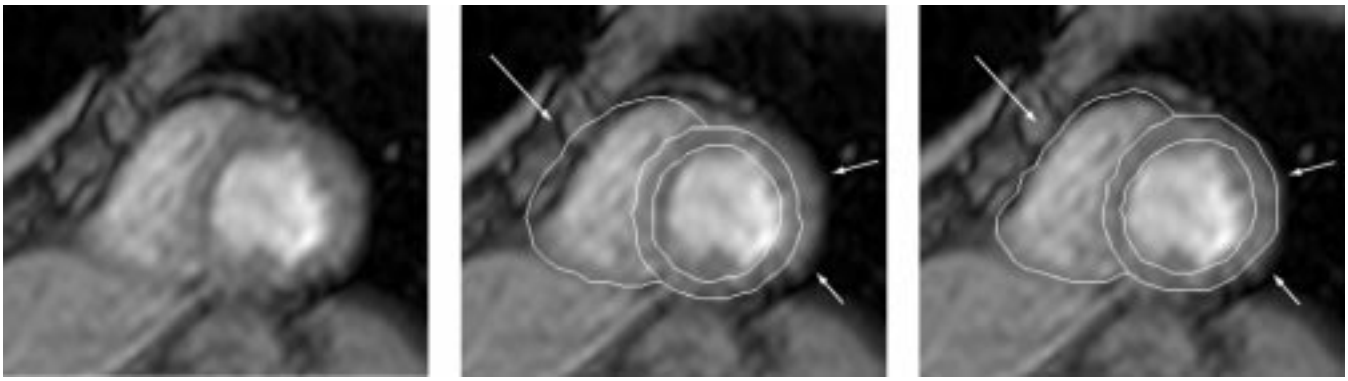
In the original ASMs, limited knowledge about image appearance is accommodated. For each landmark point, typically a scanline perpendicular to the shape is sampled, and for each landmark point, an intensity model is generated analogous to the generation of the shape models. These local appearance models serve to generate a proposed boundary location during image search of the ASM. Therefore only local appearance knowledge is modeled.

AAMs, however, describe the image appearance and the shape in an integral shape-appearance model of an image patch [15]. In the equations below, the subscript s corresponds to shape parameters while the subscript g represents appearance or gray-level parameters. An AAM is constructed in the following steps.

1) Compute an ASM and approximate each shape sample as a linear combination of eigenvectors, where $b_s = P_s^{\mathrm{T}}(x - \overline{x})$ represents the sample shape parameters (1).
2) Warp each image on the mean shape using a linear or nonlinear image interpolation.
3) Normalize each image to the average intensity and unit variance $\overline{g}$.
4) Perform a PCA on the normalized intensity images.
5) Express each intensity sample as a linear combination of eigenvectors, where $b_g = P_g^{\mathrm{T}}(g - \overline{g})$ represents the sample gray-level parameters.
6) Concatenate the shape coefficient vectors b_s and gray-level intensity coefficient vectors b_g in the following manner:

$$b = \begin{pmatrix} Wb_s \\ b_g \end{pmatrix} = \begin{pmatrix} WP_s^{\mathrm{T}}(x - \overline{x}) \\ P_g^{\mathrm{T}}(g - \overline{g}) \end{pmatrix} \tag{3}$$

where weighting matrix W is a diagonal matrix that relates the different units of shape and gray-level intensity coefficients.



(a) (b) (c)

Fig. 1. Comparison of conventional AAM and multistage hybrid ASM/AAM segmentation. (a) Original image. (b) Conventional AAM segmentation demonstrating a good gray level appearance fit but poor border positioning accuracy (arrows). (c) Multistage hybrid ASM/AAM approach result shows substantial improvement in border detection positioning (arrows).

7) Apply a PCA to the sample set of all b vectors, yielding the model

$$b = Qc \tag{4}$$

where Q is a matrix consisting of eigenvectors (from (3)) and c are the resulting appearance model coefficients.

C. Matching AAM to Images

Matching an appearance model to a target image involves finding an affine transformation, global intensity parameters, and appearance coefficients that minimize the root-mean-square (rms) difference between the appearance model instance and the target image. The method described by Cootes [15] suggests using a gradient descent method that relates model coefficients with the difference between a synthesized model image and the target image.

Let t represent the affine transformation parameters, and u the global intensity parameters. As shown above, shape x is derived in the target image from the appearance coefficients c and the affine transformation t. Then, the gray-level intensity vector g_s in the target image spanned by the shape x is extracted using image warping; g_s represents the image patch warped from the target image to the average shape. The model gray-level intensity vector g_m is derived from the appearance coefficients c and is later modified by the global intensity parameters u.

Employing reduced-rank multivariate linear regression on a set of known model parameters, the matrices R_c, R_t, and R_u are derived. Using a set of training images, their corresponding modeling parameters c, t, and u are randomly displaced, thus creating a difference in g_s and g_m. From the displacements and difference images, regression matrices are computed. Alternately, this is equivalent to finding the gradient of the image difference objective function.

The corresponding model correction steps can be computed as

$$\delta c = R_c (g_s - g_m) \tag{5}$$
$$\delta t = R_t (g_s - g_m) \tag{6}$$
$$\delta u = R_u (g_s - g_m). \tag{7}$$

Then, conventional AAM matching can be accomplished as follows.

1) Place an appearance model roughly on the object of interest using the parameters c, t, and u and compute the difference image $g_s - g_m$.
2) Compute the rms of the difference image, E.
3) Compute the model corrections δc, δt, and δu as derived above from the difference image (5)–(7).
4) Set $k = 1$.
5) Compute new model parameters as $c := c - k\delta c$, $t := t - k\delta t$, and $u := u - k\delta u$.
6) Based on these new parameters, recompute $g_s - g_m$ and find the rms.
7) If the rms is less than E, accept these parameters and go to Step 3.
8) Else set k to 1.5, 0.5, 0.25, etc. and go to Step 5. Repeat steps 5–8 until the error cannot be reduced any further.

III. Multistage Hybrid Active Shape and Active Appearance Model Segmentation

A. Arguments for a Hybrid Model

Although the conventional AAM approach performs well with locating objects, it is better suited for matching their appearance than it is for the segmentation of target images. The algorithm tends to produce plausible solutions, but the borders may be imprecise. This is because an AAM is optimized on global appearance and, thus, less sensitive to local structures and boundary information commonly used in other segmentation algorithms [Fig. 1(a) and (b)]. Conversely, ASMs tend to find local structures fairly well [15]. ASMs typically fit a shape-only model to a target image based upon the edges or edge patterns normal to the shape borders. ASM's strength lies upon a direct association between edge profiles and shape borders, however they are very sensitive to their initial placement and do not take advantage of overall gray level appearance information. Therefore, we concluded that ASM and AAM matching could be combined in a hybrid fashion during the final stages of optimization to improve the achieved segmentation with respect to the accuracy of object boundary placement.

In our hybrid ASM/AAM stage, the ASM and AAM matching is performed independently and the resulting parameters are combined after each iteration. This is done by extracting the shape model b_s from the AAM, performing a best fit using only the shape model and the pose parameters t and then transforming the refined shape model back into the AAM. At the same time, AAM refinement is applied to the image yielding its own set of coefficients for shape and pose. The two sets of shape and pose coefficients are combined using a weighted average.

B. Matching the Shape Model to the Target Using A Modified ASM

Matching only the shape model to the target image requires an ASM optimization. Nevertheless, the ASMs may be prone to noise because the strongest matches between individual points are not constrained by the locality of adjacent points. Since we assume that the hybrid refinement is used toward the later stages of optimization, the AAM has already converged to a close proximity to the target borders. Therefore, spurious edges may cause more harm than good during the ASM refinement process. This effect is partially resolved by projecting the new shape back into the shape model and constraining the coefficients. Further improvement of the ASM refinement strategy may come from employing a global boundary optimization step. We elected to use dynamic programming to globally optimize the resulting border. Our ASM refinement process, thus, involves the following.

1) Extracting the shape x from the shape model b_s.
2) Creating a cost matrix of normal profiles at each shape point x_i by determining the Mahalanobis distance function between model and target border profiles normal to the border at each landmark point. That is, a normalized mean profile $\overline{p}_i$ is created for each shape point x_i and compared to a set of new profiles p sampled along the normal of each shape point to create a cost matrix for dynamic programming. A simple Euclidean distance could be used, but Cootes suggests a Mahalanobis distance defined as

$$f(p) = (p - \overline{p}_i)^{\mathrm{T}} S_p^{-1} (p - \overline{p}_i) \tag{8}$$

where S_p is a covariance matrix of the profiles p_i sampled in the training set. Also suggested is that derivatives of the profiles normalized to a unit vector should be used instead of profile samples to reduce the effects of global intensities.
3) Finding the optimal lowest-cost path across the cost matrix using dynamic programming producing a new shape x'.
4) Transform x' into a new shape model instance defined by b_s and affine transformation t. This is done by aligning x' to the mean shape $\overline{x}$ yielding the new affine transformation t, and then computing b_s using

$$b_s = P_s^{\mathrm{T}} (x - \overline{x}). \tag{9}$$

5) Applying model constraints to the new shape model b_s so that the new shape is from the allowed shape-space. There are several ways to constraint b_s. The simplest is to apply a hard constraint to each coefficient of $\pm 3\lambda_i$. Another method is to scale the length of b_s until each coefficient is below $\pm 3\lambda_i$.
6) Repeat starting with Step 1 until there are no significant changes in x'.

C. Hybrid Model Matching

By combining the hybrid technique with conventional AAM matching, the segmentation is performed in the following way (in the hybrid stage, ASM parameters are denoted by single-primed terms, AAM parameters by double-primed terms)

1) The appearance model's current state with respect to the object of interest is described via the parameters c, t, and u. Compute the difference image $g_s - g_m$.
2) From the appearance parameter c (4), extract the shape parameters b_s, and refine b_s and pose t utilizing ASM matching as described above. This results in a new set of parameters b_s' and t'.
3) Compute the rms difference between the model and the target image, E.
4) Compute the model corrections δc, δt, and δu as derived above from the difference image (5)–(7).
5) Set $k = 1$.
6) Compute new model parameters as $c' = c - k\delta c$, $t'' = t - k\delta t$, and $u'' = u - k\delta u$.
7) Based on these new parameters, recompute $g_s - g_m$ and determine the new matching rms difference value.
8) If the new rms value is less than E, accept these parameters and go to Step 10.
9) Else set k to 1.5, 0.5, 0.25, etc.; go to Step 6.
10) From the new appearance coefficient c'', extract the shape model parameter b_s'' and the appearance parameter b_g''. With the ASM-refined shape coefficient b_s', recombine b_s', b_s'', and b_g'' into c using a weighted average on the shape coefficients ($\alpha = 0.5$ seems to produce the best results)

$$c = Q^{\mathrm{T}} \begin{pmatrix} W(\alpha b_s' + (1-\alpha) b_s'') \\ b_g'' \end{pmatrix}. \tag{10}$$

In addition, recombine the new ASM and AAM pose parameters, t' and t'', using a weighted average

$$t = \alpha t' + (1-\alpha) t''. \tag{11}$$

11) Repeat starting with Step 2 until there are no significant changes in the rms error.

D. Multistage Hybrid ASM/AAM in Fully Automated Cardiac MR Analysis

The complete active appearance segmentation method is a multistage process. First, an AAM is matched to the image until convergence, which results in defining the pose and general shape of the sought object. Note that the conventional AAM matching process ends at this point. However the model may be trapped in a local minimum where the edge boundaries are slightly off from the target borders, or the shape's position might

be significantly off due to an attraction to the wrong boundaries. In the second stage, our hybrid ASM/AAM process is repeated until convergence to help move the current solution out of the potential local minimum to which the first-stage AAM converged. After that, the third stage is invoked. The current gray-level appearance of the AAM is replaced with the mean appearance with the shape and pose kept unchanged by setting $b_g = 0$. The third-stage AAM is matched to the image using the conventional AAM matching process until convergence.

In the reported cardiac MR application, the approximate location of the left ventricle is determined fully automatically using a Hough transform approach that has been thoroughly validated and demonstrated to be highly reliable (part of MASS cardiac analysis package, [16]). For each slice in the study, the LV midpoint is detected, and a straight line is fitted to the estimated midpoints for all the slices. A point is, thus, determined in the approximate center of each slice of the LV cavity. The heart orientation however remains unknown. The orientation of the AAM is initially set to be horizontal, which was the most prevalent cardiac pose in the training set.

Throughout the three-stage matching process, the rms difference between the model and the target image is used to evaluate the quality of the final match. Five repeated matches are performed with initial orientation ranging from $-90°$ to $+90°$ in $45°$ intervals. If no initial positioning yields a good-enough rms error, the best of the five solutions is selected. Fig. 1(c) shows the final segmentation using the hybrid ASM/AAM approach in comparison with the conventional AAM approach [Fig. 1(b)].

Our fully automated multistage hybrid ASM/AAM segmentation can be summarized as follows.

1) Initialization: The approximate location of the LV midpoint is determined via Hough transform.
2) Early AAM Stage: The mean AAM is placed on the target image with its LV midpoint centered on the automatically determined midpoint and its orientation set at $-90°$ with respect to the horizontal orientation. The AAM is matched to the target image using the conventional AAM matching until convergence.
3) Hybrid ASM/AAM Stage: The hybrid ASM/AAM refinement is employed until convergence.
4) Final AAM Stage: The mean-appearance AAM ($b_g = 0$) is matched to the target image using the conventional AAM matching until convergence.
5) Best Match Selection: Reinitialize the AAM matching with the model rotated $+45°$ from the last initial position. Repeat steps 2–4 five times. Use the match with the lowest rms error as final.

Following this process, the model is fitted to the image data with a high likelihood to avoid local minima. Fig. 2 demonstrates a typical rms matching error as a function of the model matching iteration step. Note that each of the three sequential matching processes independently converges yielding an overall favorable rms matching error.

IV. Experimental Methods

The performance of our AAM segmentation methodology was tested in in vivo MR image data by comparison with a manually defined independent standard of contours drawn by a human expert.

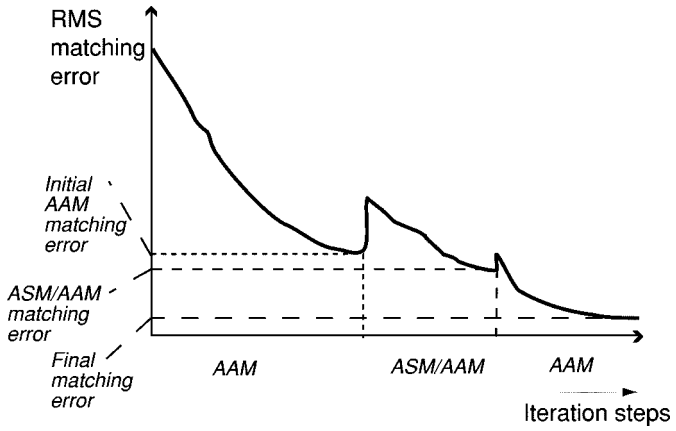


Fig. 2. Typical rms matching error as a function of the model matching stage and iteration step (rendering). The labels below the iteration step axis denote which model matching stage is employed.

A. Data

The model was trained and its performance tested in routinely acquired multislice multiphase short axis cardiac MR sets that were obtained from 34 healthy volunteers and 20 patients (Fig. 3). To obtain a balanced population, patients were selected that suffered from various common cardiac pathologies such as hypertrophic obstructive cardiomyopathy, myocardial infarction, or LV aneurysm. In each image set, three end-diastolic, mid-ventricular slices were selected for the total of 162 images of 256×256 pixels; field of view: 400–450 mm; pixel sizes: 1.56–1.75 mm.

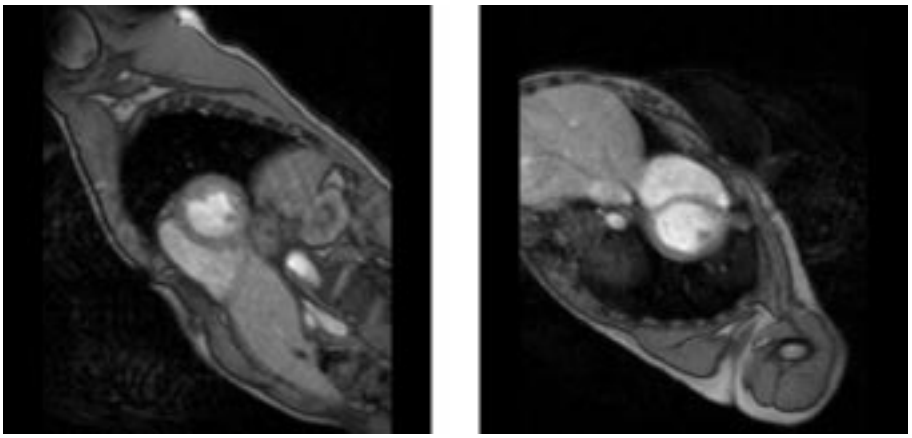
B. Training and Testing Data Sets

The available data were divided in a training set consisting of 102 images (from 11 patients and 23 normal subjects) and a testing set consisting of 60 images (from nine patients and 11 healthy subjects). The training and testing sets were completely disjoint—at the image as well as subject levels. The models were trained on the training set as described above. The method's performance was assessed in the testing set.

C. Independent Standard and Quantitative Assessment Indices

The three contours depicting the LV endocardium, RV endocardium, and LV epicardium were manually drawn by an expert observer in all images used in the reported study (training as well as testing) and served as an independent standard. Manually traced contours were defined using user-friendly tracing and editing tools of the cardiac analysis package MASS [16]. The observer was unaware of the results of the computer analysis and had unlimited time to define the independent standard to full satisfaction.

To quantitatively assess the accuracy of the automatically defined contours with respect to the independent standard, border positioning error measures comparing the automatically determined contours with the independent standard were determined. The average signed, unsigned, maximum, and



(a) (b)

Fig. 3. Example MR images used for validation. See Fig. 6 for the segmentation results.

rms border positioning errors were independently defined for all three borders by measuring the distances between corresponding border points. Corresponding points were defined as pairs of points, the first point being from the computer-detected border and the second one from the observer-determined border that were closest to each other using the Euclidean metric. Border positioning errors are expressed in millimeters as mean $\pm$ standard deviation. The negative sign of the signed error value means that the automatically determined border was inside of the observer-defined border.

Three clinically important area-dependent measures were calculated and used for performance assessment: LV cavity area, RV cavity area, and LV epicardial area. The individual indices were defined as the areas enclosed by the LV and RV endocardial and epicardial borders, respectively. Area measurements are expressed in centimeters squared. The information from the MR-image DICOM header was used for pixel size calibration.

D. Reproducibility

To assess reproducibility, the method was applied to the testing data set twice. In the first run, the initial positions of the AAM differed by 45°. The initial AAM orientations were rotated by $+22.5°$ for the second run. Thus, the two reproducibility runs were independent.

E. Statistical Assessment

To statistically assess the performance of the computer method, linear regression analysis was used to compare computer-detected and observer-defined area measurements. Regression equations were also compared to the line of identity using t-statistic for slope and intercept. A p value less than 0.05 was considered significant. Bland–Altman statistical analysis was used to assess the method's reproducibility by comparing the agreement between the area measurements in the two independently performed automated analysis studies. Since area errors increase proportionally with the increasing heart size, log–log Bland–Altman plots were used for reproducibility assessment [17]. To demonstrate that the introduced multistage hybrid approach yields significantly better results in comparison to the conventional AAM method, unsigned positioning errors obtained in the testing set were compared using a paired t-test.

TABLE I
BORDER POSITIONING ERRORS ASSESSING THE ACCURACY OF THE AUTOMATICALLY DEFINED BORDERS WITH RESPECT TO THE INDEPENDENT STANDARD

Errors (mm)	LV ENDO	LV EPI	RV
Avg. signed error	0.22±1.90	-0.01±1.92	-0.32±2.80
Maximum error	3.82±1.80	3.83±1.55	6.23±3.27
RMS error	1.71±0.82	1.75±0.83	2.46±1.39

Errors expressed as mean $\pm$ standard deviation.

V. RESULTS

The contours of the LV and RV endocardium, and LV epicardium were automatically detected in all images in the testing set by performing our multistage ASM/AAM segmentation as presented above. In all 60 cases, the matching resulted in a visually plausible mapping of the model to the image data. All borders were detected fully automatically and were not manually edited. The computer determined borders were accurate with the maximum and rms border positioning errors summarized in Table I. The results show minimal bias in border positioning and typical border placement within one to two pixels of the manually traced independent standard.

A good correlation was obtained between computer-detected and observer-defined LV and RV cavity areas as well as for the LV epicardial areas: ($r = 0.96$, $y = 1.04x - 0.55$ [cm^2]; $r = 0.90$, $y = 0.97x - 0.22$ [cm^2]; $r = 0.96$, $y = 0.90x + 4.01$ [cm^2], respectively; Fig. 4). The slopes and intercepts of the LV and RV cavity area regression lines did not differ significantly from one and zero, respectively ($p > 0.3$).

Assessing reproducibility, the Bland–Altman analysis of agreement revealed very small bias and good coherence between the two independently computer-determined epicardial, LV, and RV cavity areas (0.4%, 0.6%, and 0.2%, respectively). As shown in Fig. 5, 95% of the computer measurements of LV cavity areas can be expected to differ from repetitive measurements by less than 8% below and 8% above the mean

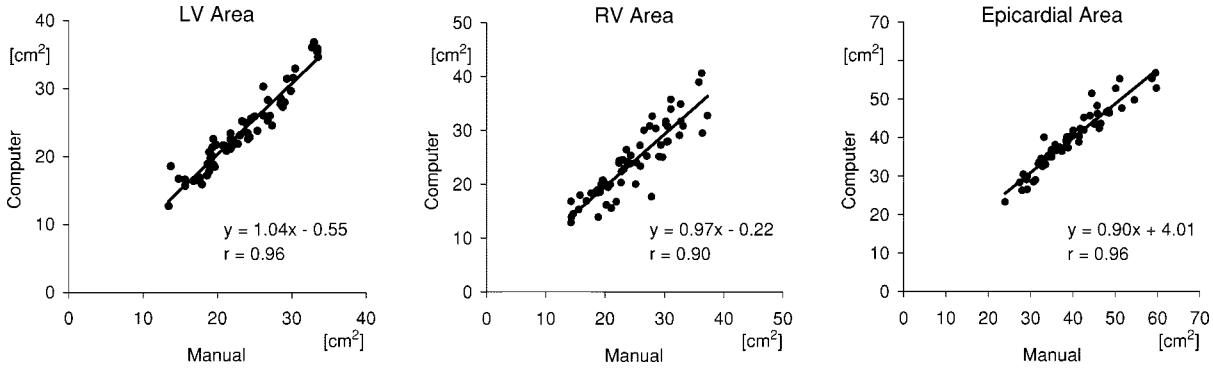


Fig. 4. Comparison of the observer-defined and computer-determined area measures in the 60 images of the testing set.

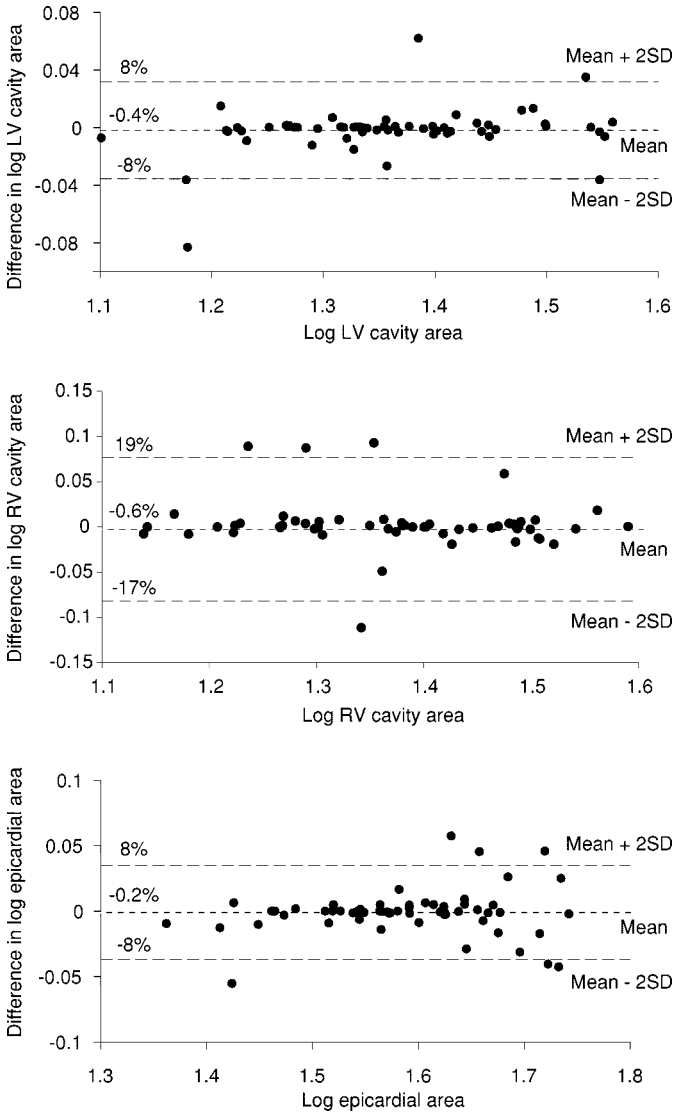


Fig. 5. Bland-Altman plots showing reproducibility of two independent computer-determined area measures in the 60 images of the testing set. (a) LV cavity areas. (b) RV cavity areas. (c) Epicardial areas.

of the two runs. For the computer measurements of the RV cavity area, 95% of cases can be expected to differ from repetitive measurements by less than 17% below and 19% above the mean. For the computer measurements of the epicardial area, 95% of cases can be expected to differ from repetitive measurements by less than 8% below and 8% above the mean.

Our multistage hybrid AAM method significantly outperformed the conventional AAM approach at the significance level of $p < 0.001$ as assessed by comparison of unsigned border positioning errors in the testing set.

Two examples of computer-detected segmentation achieved in MR images from the testing set are given in Fig. 6. The original image data and the manually defined borders are shown together with the results of the automated computer analysis.

VI. Discussion

The reported method solves several important problems of cardiac MR image segmentation. The method identifies structure boundaries using knowledge acquired during the training stage from expert-segmented examples. The approach consistently delivers high quality segmentations across a variety of disease conditions in not-always-perfect clinical-quality images. Fully automated performance facilitates utility in clinical care. Additionally, it is the first such method performing RV analysis. The discussion will concentrate on three topics—behavior of the reported multistage hybrid ASM/AAM method, its achieved performance, and applicability to clinical data.

A. Behavior of the Hybrid ASM/AAM Segmentation Method

The method is designed as a multistage process and the hybrid ASM/AAM model is used in the second stage. The overall strategy is to concatenate several independent matching phases in which the later stages help the model to emerge from a potential local minimum the earlier stages converge to. The matching sequence is AAM $\rightarrow$ hybrid ASM/AAM $\rightarrow$ AAM. There are several reasons for starting with a conventional AAM matching stage and not using the hybrid model exclusively. First, the ASM component in the hybrid model is not a suitable approach in the early model fitting stages. Initially, the search for the overall pose and location of the model is of primary importance and the ASM is not well suited for this task. Second, the higher computational demands of the hybrid ASM/AAM stage would cause a somewhat slower runtime. Our experiments have shown that the designed multistage approach results in excellent overall performance as far as speed, accuracy and stability are concerned. After the first stage of AAM matching finishes, chances are

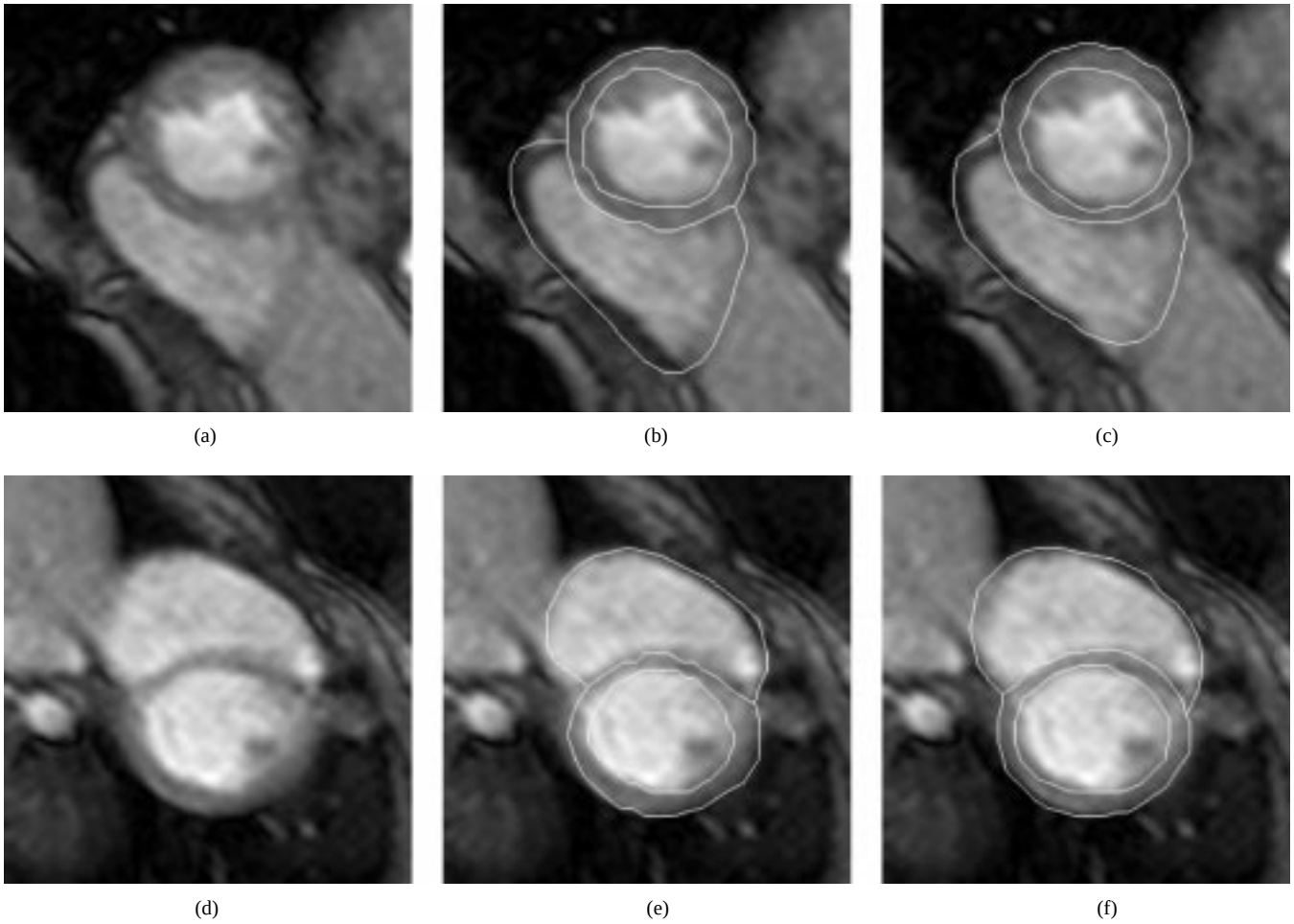


Fig. 6. LV and RV contours obtained by our multistage hybrid ASM/AAM segmentation method in cardiac MR images. Presented examples are from the testing set. (a) and (d) Show the originals, (b) and (e) the independent standard. (c) and (f) Show the results of our multistage segmentation method. See Fig. 3 for the full-size original images that were analyzed using our fully automated approach.

that the location represents a local minimum of the evaluation function. The competition introduced by the hybrid ASM/AAM together with the ASM's focus on local border properties effectively forces the AAM to move from such undesirable location. The third stage of the overall process starts with disregarding the gray-level appearance information in the current AAM which again aids in escaping a possible local minimum found during the ASM/AAM stage (Fig. 2). The final AAM optimization stage is then employed to correct any appearance errors introduced by the ASM. It is possible to continue this cycle of hybrid ASM/AAM and AAM repeatedly until convergence, but the small gains in improvement didn't justify the significantly slower runtime.

B. Performance

Performance of the multistage hybrid AAM method was found excellent by the variety of validation experiments performed. The method is highly reproducible and gives accurate border locations as well as area measurements.

The significant increase of border positioning accuracy in comparison with the conventional AAM method is mostly attributed to the cases in which the AAM does not converge to the final position due to a premature convergence to a local minimum. Consequently, the hybrid stage is most valuable in the images that may otherwise be considered failures. Indeed, the combination of multistage hybrid methodology and the automated initialization decreased the number of failures from five to zero in the testing set of MR images. We also compared the performance independently in the patient and normal-control subsets of the testing set. Not surprisingly, the results obtained in the normal data sets are consistently better, suggesting that a larger number of diseased hearts should be used for training. Still, the regression coefficients of all area-regression analyzes independently comparing LV, RV, and epicardial areas in normal or patient group with the independent standard were very good $(r > 0.95)$. The patient-group RV area was the only exception with a slightly lower correlation $r = 0.84$, clearly, reflecting higher difficulty of diseased RV segmentation as well as the limitations of the available independent standard.

The consistency of the independent standard is one of the key factors on the way to good performance. Misleading information, if present in the training set, will decrease the knowledge contained in the model. Similarly, local inaccuracies in the observer-defined borders may worsen the segmentation results. Our independent standards were carefully prepared by expert

cardiologists well aware of the requirements of computerized training. The RV borders are more difficult to identify consistently. Not surprisingly, the most frequent inconsistencies occur in the T-junction areas of the RV endocardial border and LV epicardium. Consequently, the local border positioning errors identified in the testing set frequently originate in these areas. Another less pronounced source of tracing ambiguities across the 102 images of the data set can be found in the areas where borders cut through papillary muscles. No physical border exists here that can be followed. The ambiguities in border positioning are reflected in the trained model subsequently used for segmentation. While it seems almost impossible to collect completely unambiguous and also large training and testing data sets, special care will be given in the future to the identified problem areas to further improve our unique and valuable independent standard data base and make it available for future studies.

C. Applicability to Clinical Data

Our fully automated method never failed in the analyzed testing set. The method was found excellently capable of capturing observer preference expressed in the training set tracings. Our method generated segmentation results that were not merely based on the strongest image evidence, but consistently took the preference of the expert observer into account—specifically, unwanted regions such as the papillary muscles or epicardial fat were consistently excluded. Accurate segmentations were achieved even in images where the image evidence was inconclusive or missing, for instance because of poor blood-muscle contrast or electrode artifacts. Due to the fact that AAM unifies shape and appearance knowledge in one integral model, the method works exceptionally robustly, even in low-fidelity, routinely acquired cardiac MR images from sometimes severely deformed patient hearts. However, the current model training has been limited to mid-ventricular, end-diastolic images. Further development is required to achieve similar results in apical and basal slices, as well as nondiastolic phases of the cardiac cycle. Together with the full automation and fast processing speed (5–10 s/image), and after additional work focusing on inherently three–dimensional and four-dimensional model training and model-based segmentation, the multistage hybrid ASM/AAM cardiac MR segmentation may be an ideal segmentation method for routine quantitative clinical analyzes of LV and RV morphology and function.

VII. Conclusion

A new highly reliable automated approach to segmentation of the left and right heart ventricles from magnetic resonance images was reported. The method incorporates a novel hybrid ASM/AAM strategy in a multistage fashion, which helps to avoid convergence to local minima. A new model matching initialization approach was developed making the method fully automated.

Comparing the results obtained in a testing set of 60 cardiac MR images with an observer-defined independent standard demonstrated good border positioning accuracy, as well as a good and unbiased agreement of area measurements. Methods' reproducibility was tested using Bland–Altman statistic with favorable results. Our new method was shown to significantly outperform the conventional AAM matching approach in the testing set.

The developed method is the first approach delivering automated segmentation of LV *and* RV borders. It promises to facilitate fully automated quantitative analysis of LV and RVmorphology and function in routinely acquired clinical MR images.

Acknowledgment

The authors would like to thank Dr. H.W.M. Kayser for providing the cardiac MR data and expert drawn contours.

References

[1] D. Geiger, A. Gupta, L. A. Costa, and J. Vlontzos, "Dynamic programming for detecting, tracking, and matching deformable contours," *IEEE Trans. Med. Imag.*, vol. 17, pp. 294–302, June 1995.

[2] L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable contour models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 1061–1075, Nov. 1992.

[3] ——, "Model-based deformable surface finding for medical images," *IEEE Trans. Med. Imag.*, vol. 15, pp. 720–731, Oct 1996.

[4] W. E. Higgins, M. Chung, and E. L. Ritman, "Extraction of left-ventricular chamber from 3-D CT images of the heart," *IEEE Trans. Med. Imag.*, vol. 9, pp. 384–395, Aug. 1990.

[5] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Pacific Grove, CA: PWS, 1998.

[6] W. J. Niessen, B. M. ter Haar Romeny, and M. A. Viergever, "Geodesic deformable models for medical image analysis," *IEEE Trans. Med. Imag.*, vol. 17, pp. 634–641, Aug. 1998.

[7] J. Weng, A. Singh, and M. Y. Chiu, "Learning-based ventricle detection from cardiac MR and CT images," *IEEE Trans. Med. Imag.*, vol. 16, pp. 378–391, Aug. 1997.

[8] M. R. Rezaee, P. M. J. van der Zwet, B. P. F. Lelieveldt, R. J. van der Geest, and J. H. C. Reiber, "A multiresolution segmentation technique based on pyramidal segmentation and fuzzy clustering," *IEEE Trans. Image Processing*, pp. 1238–1248, July 2000.

[9] P. M. T. Pattynama, H. J. Lamb, E. A. van der Velde, E. E. van der Wall, and A. de Roos, "Left ventricular measurements with cine and spin-echo MR imaging: A study of reproducibility with variance component analysis," *Radiology*, vol. 187, pp. 261–268, 1993.

[10] H. J. Lamb, J. Doornbos, E. A. van der Velde, M. C. Kruit, J. H. C. Reiber, and A. de Roos, "Echo-planar MRI of the heart on a standard system: Validation of measurement of left ventricular function and mass," *J. Comput. Assist. Tomogr.*, vol. 20, no. 6, pp. 942–949, 1996.

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. Eur. Conf. Computer Vision*, vol. 2, H. Burkhardt and B. Neumann, Eds., 1998, pp. 484–498.

[12] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor, "A unified framework for atlas matching using active appearance models," in *Information Processing in Medical Imaging*, A. Kuba and M. Samal, Eds. Berlin, Germany: Springer-Verlag, 1999, Lecture Notes in Computer Science, pp. 322–333.

[13] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models- their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[14] N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Trans. Med. Imag.*, vol. 17, pp. 1049–1062, Dec. 1998.

[15] T. F. Cootes. (1999) Statistical models of appearance for computer vision. [Online]. Available: http://www.isbe.man.ac.uk/~bim/Models/app_model.ps.gz

[16] R. J. van der Geest, V. G. M. Buller, E. Jansen, H. J. Lamb, L. H. B. Baur, E. E. van der Wall, A. de Roos, and J. H. C. Reiber, "Comparison between manual and semiautomated analysis of left ventricular volume parameters from short-axis MR images," *J. Comput. Assist. Tomogr.*, vol. 21, no. 5, pp. 756–765, 1997.

[17] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307–310, 1986.