



**Universiteit
Leiden**
The Netherlands

3D active shape modeling for cardiac MR and CT image segmentation

Assen, Hans Christiaan van

Citation

Assen, H. C. van. (2006, May 10). *3D active shape modeling for cardiac MR and CT image segmentation*. Retrieved from <https://hdl.handle.net/1887/4460>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4460>

Note: To cite this publication please use the final published version (if applicable).

*'We are shaped by our thoughts; we become what we think.
When the mind is pure, joy follows like a shadow
that never leaves.'*

Buddha (563–483 BC)

A 3D-Active Shape Model driven by Fuzzy Inference: Application to Cardiac CT and MR

This chapter was adapted from:

A 3D-Active Shape Model driven by Fuzzy Inference: Application to Cardiac CT and MR
H.C. van Assen, M.G. Danilouchkine, M.S. Dirksen, J.H.C. Reiber, and B.P.F. Lelieveldt
Submitted

Abstract

Manual quantitative analysis of cardiac left ventricular function using Multi-Slice CT and MR is arduous because of the large data volume. In this paper, we present a 3D-Active Shape Model for automatic segmentation of cardiac CT and MR volumes, without the requirement of retraining the underlying statistical shape model. A Fuzzy-C-Means based fuzzy inference system was incorporated into the model. Thus, relative gray-level differences instead of absolute gray values were used for classification of 3D ROIs, removing the necessity of training different models for different modalities/acquisition protocols.

The 3D-ASM was evaluated using 25 CT and 15 MR data sets. Automatically generated contours were compared to expert contours in 100 locations. For CT, 82.4% of epicardial contours and 74.1% of endocardial contours had a maximum error of 5 mm along 95% of the contour arc length. For MR, those numbers were 93.2% (epicardium) and 91.4% (endocardium). Volume regression analysis revealed good linear correlations between manual and automatic volumes, $r^2 \geq 0.98$.

This study shows that the fuzzy inference 3D-ASM is a robust promising instrument for automatic cardiac left ventricle segmentation. Without retraining its statistical shape component, it is applicable to routinely acquired CT and MR studies.

4.1 Introduction

In cardiac imaging, multi-slice multi-phase CT and MR are increasingly used for cardiac left ventricle (LV) function analysis, yielding large amounts of three-dimensional image data. Segmentation of such dynamic medical image data is a necessary step for (automatic) quantification of global and regional left-ventricular function in terms of stroke volume, ejection fraction, wall thickness and wall thickening. Manual segmentation of these data sets is a tedious, labor-intensive and practically infeasible task. Therefore, an automatic and intrinsically three-dimensional segmentation method is highly desired.

Region and edge-based low-level image processing techniques are generally not robust enough for application in clinical practice. Such low-level methods expect clear well-defined image features, while in medical image data fuzzy edges, inhomogeneity and noise are commonly observed, causing segmentation failures. Because of this, three-dimensional medical image segmentation tasks require a-priori knowledge that the classical segmentation methods lack, e.g., knowledge about organ shape and intensity properties.

In recent years, much work on knowledge- and model driven segmentation has been described. Examples of such methods are deformable statistical models, which enable incorporation of a-priori knowledge about statistically plausible organ shapes. These models are built using a training set of expert segmentations of an organ. Examples of statistical models are:

- Point Distribution Models (PDMs) [11]

PDMs are derived by modeling the distribution of a set of corresponding landmark points from a set of example shapes. After Procrustes alignment of the shape samples, the remaining shape variation is modeled by a mean shape and a set of eigendeformations derived by applying Principal Component Analysis on the sample point covariance matrix. PDMs have been explored extensively

in 2D, but also in 3D, mainly for the analysis of complex 3D shapes, e.g., the cortical sulci [15], the lumbar vertebrae [16], the femoral head and the lateral ventricles of the brain [17], and the heart [18].

- **Active Shape Models (ASMs) [19]**
ASMs consist of a PDM, extended with a matching scheme driven by information from the target image data. The original ASM formulation uses a gray-level model - derived from the training data - of scan lines perpendicular to the model contour or surface to estimate new update positions for the model sample points. The differences between the cloud of candidate sample points and the model points are used for model alignment and deformation in each iteration. The intrinsic shape deformation description restricts the search space within statistically plausible limits.
- **Active Appearance Models (AAMs) [28]**
AAMs are an extension of ASMs with a statistical intensity model of a complete image patch [62] or volume [31], as opposed to scan lines in the ASM matching. AAM matching is based on minimizing a criterion expressing the difference between model and target image. This enables a rapid search for the correct model location during the matching stage of AAMs, while utilizing precalculated parameter derivative images. The sum of squares of the difference between the model generated patch and the underlying image serves as a criterion for model convergence.
- **Spherical harmonics**
Kelemen *et al.* [35] propose a 3D model that uses a parametric shape representation instead of a PDM. Using spherical harmonics, they automatically generate surface meshes with homogeneous node distributions. They calculate eigenvariations in the shape parameter space, whereas Cootes *et al.* [19] derive shape statistics from sample point coordinates. Matching performance was tested with a left hippocampus model; results were compared to manual segmentation.
- **Constrained level sets**
Tsai *et al.* [37] introduced a method for segmentation of medical imagery by incorporating a-priori knowledge into the level set framework. They derived a model-based curve evolution technique and applied it in 3D to a prostate gland segmentation example from pelvic MRI. Like Kelemen *et al.*, Tsai *et al.* do not derive shape statistics from sample point coordinates, but use the signed distance function as a shape representation, and extract modes of shape variation by applying PCA to those distance functions.
- **Deformation statistics**
Rueckert *et al.* [38] present an algorithm for the construction of 3D anatomical models of the brain from MR images. They use a nonrigid registration algorithm based on free-form deformations and normalized mutual information. Thus, dense correspondences between subjects within a target population are derived from the deformations between them. These statistical deformation models have been applied to cardiac modeling by, e.g., Chandrashekhara *et al.* [39] and Lötjönen *et al.* [32], who recently presented a statistical deformation model built from combined long-axis and short-axis views of the cardiac ventricles and atria.

Amongst others, active shape- and appearance models have proven to be highly effective in clinical medical image segmentation applications. The combination of expert shape knowledge from a training set and image information extracted during matching makes this segmentation technique very robust. As a consequence, both ASMs and AAMs have found widespread application in medical image segmentation. AAMs have been extended to 3D [31], and have also been described combining multiple 2D views of a single subject [62, 63]. However, most applications for Active Shape Models (modeling + matching) have been limited to 2D [19] and 2D + time [23]. The application of 3D ASMs to image segmentation is still largely unexplored territory.

A more elaborate introduction on segmentation methods, including considerations on their usefulness and limitations with respect to medical image segmentation can be found in Van Ginneken *et al.* [21]. A review on 3D cardiovascular image analysis can be found in Frangi *et al.* [64].

The main contribution of this work is the development of a three-dimensional Active Shape Model that performs comparably on MR and CT data sets, without retraining the model (i.e., collecting expert contours in a large amount of new data sets). Therefore, we required a method that minimizes dependence on image modality: this excludes the incorporation of a gray level model, as in the "classic" ASM and AAM definitions by Cootes *et al.* [19, 28]. Gray value information should not be part of the statistical expert knowledge included in the model; hence, we only include shape knowledge. The method presented here builds on previous work, which was presented as a proof of concept [65]. While in [65], a small and qualitative evaluation study involving only CT data sets was presented, this paper demonstrates modality independence in larger quantitative evaluation studies on MR and CT data.

The main novelty of the method presented here, a 3D-ASM, is a Fuzzy Inference system that generates model updates during matching based on *relative* intensity differences, thus removing the requirement of model retraining for different modalities. We chose to adopt fuzzy inference (FI) for generation of the model update steps for two reasons:

- by applying a classification approach to image patches instead of pixel scan lines, sensitivity to local disturbances and noise can be greatly reduced,
- the model should be applicable to both MR and CT. This can be realized by selecting fuzzy clustering for classification, since it is based on relative intensity differences between tissues.

In addition, we addressed a number of matching problems specific to a 3D-ASM, e.g., 3D point correspondence, 3D model alignment and transformation of 2D vectors from 2D image slices to 3D space.

This paper is further structured as follows. In the next section, a brief background on the deformable statistical models introduced by Cootes *et al.* is given. Following that, a more extensive explanation of the 3D-ASM, will be given, including model training and matching details. In Section IV, we describe the experimental setup to evaluate our model, followed by experimental results in Section V. These are discussed in Section VI, followed by a conclusion.

4.2 Background

Active Shape Models consist of two parts, the statistical shape model and the model-matching algorithm.

4.2.1 Shape Modeling

To generate a point distribution model, the shape of interest is described using a set of sample points. Shape differences can only be modeled if a point correspondence has been established, according to which the sample point coordinates are inserted into a shape vector

$$\mathbf{x} = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)^T. \quad (4.1)$$

For this, landmark points on expert drawn contours are determined, being easily identifiable positions in the image data. Resampling of the contours between the landmark points yields the sample points that build the shape vector.

From all shapes \mathbf{x}_i in the training set (number of samples n_s) a mean shape vector

$$\bar{\mathbf{x}} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{x}_i \quad (4.2)$$

and a covariance matrix

$$S = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (4.3)$$

are calculated, followed by Principal Component Analysis (PCA) on S to compute the eigenvectors and eigenvalues of the training set. Depending on the amount of variation in the training set represented by the model, the eigenvectors ϕ_i corresponding to the t largest eigenvalues are retained, represented by matrix $\Phi = (\phi_1 \mid \phi_2 \mid \dots \mid \phi_t)$. From this matrix, a shape can be approximated

$$\mathbf{x} \approx \bar{\mathbf{x}} + \Phi \mathbf{b}, \quad (4.4)$$

with \mathbf{b} a t -dimensional vector containing the model parameters

$$\mathbf{b} = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}). \quad (4.5)$$

More elaborate background information on active shape models can be found in [11, 19].

In case of few anatomical landmarks, a more dense shape description can be obtained by adding pseudo-landmarks yielding a more dense correspondence. The definition of point correspondence is an active field of research. Apart from a manual definition of point correspondence, different automatic methods have been explored. In [18], a registration-based method for automatic landmark generation is described. From the training set, an atlas is constructed for which a set of pseudo-landmarks is generated. The pseudo-landmarks are propagated to the individual shapes in the training set through the inverse transformation of the non-rigid registration alignment used in the atlas construction. Styner *et al.* [17] presented an evaluation of four 3D correspondence methods. They found, that a Minimum Description Length (MDL) approach [56]

and minimization of the determinant of the covariance (DetCov [66]) outperformed SPHARM [67] and Manually initialized Subdivision Surfaces (MSS). However, since the focus of this paper is the development of a modality independent 3D matching strategy, we did not strive for an optimal point correspondence. Hence, an application specific point correspondence is applied at this stage.

4.2.2 Model matching

Usually, the matching algorithm (in ASMs and AAMs) is an iterative approach, in which the model is aligned and deformed. The pose and shape of the model is repeatedly adjusted to image evidence, collected from the surroundings of the actual model state. The image evidence is obtained using scan lines perpendicular to the model contour (2D) or surface (3D). Methods for extracting information from the model environment can vary between applications. The classic ASM by Cootes et al [19] uses a statistical Gray Level Model (GLM) containing knowledge about image appearance along the scan lines mentioned above. For each scan line, the local GLM is fitted to the sampled profile, yielding a candidate position for each scan line.

During matching, two shape vectors exist, one representing the mean shape and one representing the proposed shape. The proposed shape contains the positions of the sample points after fitting the GLM. From the mean and proposed shapes, the parameter vector \mathbf{b} controlling model deformation is calculated (from (4.5)):

$$\mathbf{b} = \Phi^T (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (4.6)$$

with \mathbf{x}_i the vector representing the proposed shape, and $\bar{\mathbf{x}}$ the vector representing the mean shape of the shape model. The allowed shape instances are limited by the statistical shape description from the training set.

Alternative gray level matching strategies have been investigated. In [21] e.g., instead of using the normalized first order derivative, an extensive set of differential features and moments is extracted, and an optimal feature subset is selected for each model scan line. These features are input for a classifier determining the probability that a pixel is inside of the object. From this, the candidate boundary position is inferred. In [68], the classic ASM was extended with three additional features, one of which was to overcome variance related to uniform gray-level scaling and gray-level offset. Gray-level derivatives along profiles perpendicular to the model surface, normalized, and averaged over the training set, were used to ensure invariance with respect to the uniform gray-level features mentioned. Both these methods however, would require extensive retraining to a new population of training samples (including the collection of manual contours) when applied to a different modality.

4.3 3D-ASM

This section describes our approach to 3D-ASM model construction. With regard to the matching algorithm, the FI-based edge detection technique and the model update scheme are presented.

4.3.1 Model generation

To perform statistical analysis on 3D-shapes, three important issues were addressed:

- point correspondence for 3-dimensional cardiac left-ventricular (LV) shapes,
- shape alignment in 3D,
- statistical modeling of 3D shape and shape variation.

Our 3D-ASM was constructed from manual segmentations in cardiac short-axis image slices. As mentioned above, for the definition of point correspondence, we did not use an automatic landmark generation method. Expert defined landmarks were used as an alternative [18]. In our model, a single manually indicated landmark was used, positioned at the posterior junction of the right- and left-ventricular myocardium in a mid-ventricular slice. Since in the mid-ventricular section of the heart the position of the posterior junction does not vary significantly between slices, it is not crucial to use the exact mid-ventricular slice for landmark placement. To define the parameterization representing a particular shape instance, starting at the angle derived from the manually defined landmark, each contour is sampled at equidistant angles with respect to the LV long-axis, thus generating additional (pseudo-) landmarks. The resulting point sets from different subjects do not necessarily contain an equal number of slices. To resolve this, the resulting point sets were resampled to all represent the same number of slices, and concurrently the same number of landmarks. Each shape sample is then expressed as a $3n$ element vector \mathbf{x} containing n concatenated 3D landmark points of a shape instance (see (4.1)). The endocardial contours are sampled counter clockwise and from the apical to the basal slice, filling half the shape vector. The other half of the vector is filled by sampling the epicardial contours counter clockwise, from the basal slice towards the apical slice. This yields a line-parameterization with a specific point-to-point ordering, shown in Figure 4.1.

Using this parameterization, both the LV endo- and epicardial surface are represented in the same vector, and thus modeled together. This application-specific point correspondence has also been used in a 3D AAM presented earlier [31] and in our earlier work regarding the 3D-ASM [65].

After all shape samples have been resampled, they are aligned. For alignment of the training samples, Procrustes analysis [58] is applied. We adopted Besl and McKay’s method for registration of corresponding 3D point sets [59]. The remaining differences between the training samples are solely shape related, and thus the effects of trivial variations in pose and scale are eliminated. Shape modeling is analogous to the 2D case, except that 3D coordinates are used. To extract the modes of variation of the training set, the eigenvectors of the covariance matrix are calculated using PCA.

4.3.2 Model matching

3D model updates from 2D image features

The model was extended with a matching algorithm. The key idea behind our matching approach is applicability to data acquired with arbitrary image slice orientations, from different modalities, and even to sparsely sampled data with arbitrary image slice orientations. In sparsely sampled data containing large undersampled regions where no information from the sampled subject is available, interpolation of the data to build a (dense) 3D data volume is not feasible. Moreover, interpolation between image slices with different orientations (e.g., a radial stack of cardiac long-axis views) is ill-posed.

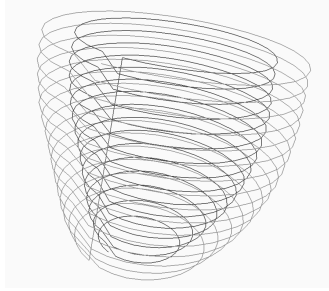


Figure 4.1: Line parameterization, defining the point correspondence for the 3D cardiac left ventricle.

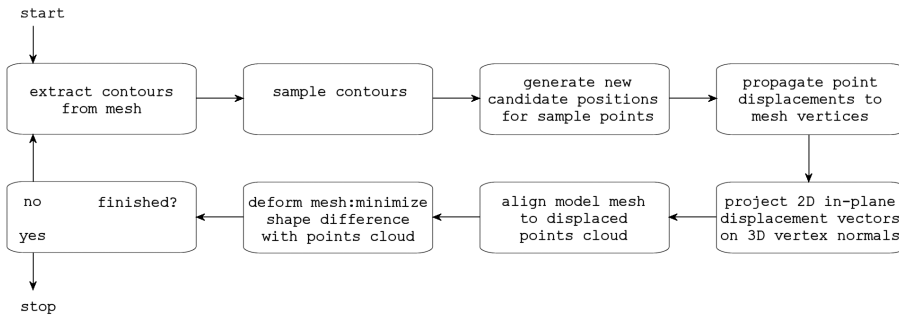


Figure 4.2: Schematic representation of the matching procedure used in this 3D-ASM.

The design requirement of applicability to arbitrarily oriented (and sparsely sampled) image slices implicates that only 2D image data be used for updating our 3D model. To realize this, a 3D triangular mesh was constructed from the sample points. During the matching, this mesh is intersected by the image planes, thus generating 2D contours spanned by the intersections of the mesh triangles. To remove dependencies on image orientation or limited through-plane resolution, model update information is represented by 2D point-displacement vectors. The 2D update vectors located at the intersections of the mesh with the image slices are first propagated to the nodes of the mesh, and to exploit the 3D character of the model, projected to the local surface normals. Multiple contributions from different mesh intersections to a single mesh node are averaged to yield a single 3D update vector per node.

Shape parameter vector updates

Similar to the training stage, scaling, rotation, and translation differences between the current state of the model and the point cloud representing the candidate updates are eliminated by alignment. The current mesh state is aligned with the candidate model state using the method of Besl and McKay [59] mentioned earlier. Successively, the parameter vector \mathbf{b} controlling model deformation is calculated. Different from (4.6), we calculate an adjustment to \mathbf{b} with respect to the previous iteration, using both

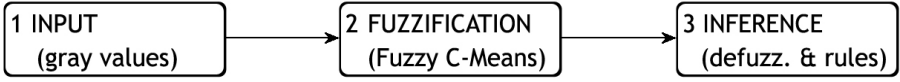


Figure 4.3: Takagi-Sugeno-based fuzzy inference system.

$\hat{\mathbf{x}}_{n+1}$ (the proposed model shape for the next iteration) and \mathbf{x}_n (the model shape of the current iteration)

$$\hat{\mathbf{b}}_{n+1} = \mathbf{b}_n + \Delta \mathbf{b} = \mathbf{b}_n + \Phi^T(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_n) \quad (4.7)$$

with \mathbf{x}_n representing the aligned current state of the mesh (see Fig. 4.2), and \mathbf{b}_n representing the parameter vector describing the current shape of the model within the statistical context. The vector $\hat{\mathbf{b}}_{n+1}$ is the shape parameter vector before statistical constraints have been applied.

4.3.3 Edge detection using Fuzzy Inference

In the classic ASM [19], model updates were generated using a (multi-resolution) statistical gray level model in each sample point: this requires a modality dependent training stage. For generating model updates without the use of a GLM, we developed an alternative decision scheme based on a Fuzzy Inference System (FIS) [61, 69] and using relative intensity differences. The basic FIS consists of three parts:

- a rule base which contains a selection of fuzzy rules,
- a dictionary which holds the definitions of the membership functions, and
- a reasoning mechanism that infers a conclusion based on given evidence.

Inputs to the basic FIS can be either fuzzy or crisp, while outputs are always fuzzy sets. If crisp outputs are desired, defuzzification needs to be performed on the fuzzy output sets. The mapping from input to output is performed by a set of fuzzy if-then rules, where the antecedent part defines a fuzzy region in the input space, and the consequence part defines a fuzzy region in the output space. A widely applied fuzzy system is the Mamdani model [69], which has a fuzzy set as output. As an alternative to the Mamdani model, the Takagi-Sugeno Fuzzy Inference System (TSFIS) or Sugeno model [61] was developed. The difference between the Mamdani and the Sugeno models is in the output. The output of the Mamdani model is a fuzzy set, that of the TSFIS is a crisp value. Therefore, the TSFIS does not need an additional defuzzification step, while the Mamdani model does, if a crisp output is desired. The output of a TSFIS is always a linear combination of the inputs, a typical fuzzy rule of the TSFIS is of the form "if input $x = A$ and input $y = B$; then output $z = aA + bB + c$ ". In this paper, a zero-order TSFIS is used, for which the output corresponds with a zero-order polynomial, i.e., a constant ($a = b = 0$).

The TSFIS in our 3D-ASM is implemented as follows (see Fig. 4.3):

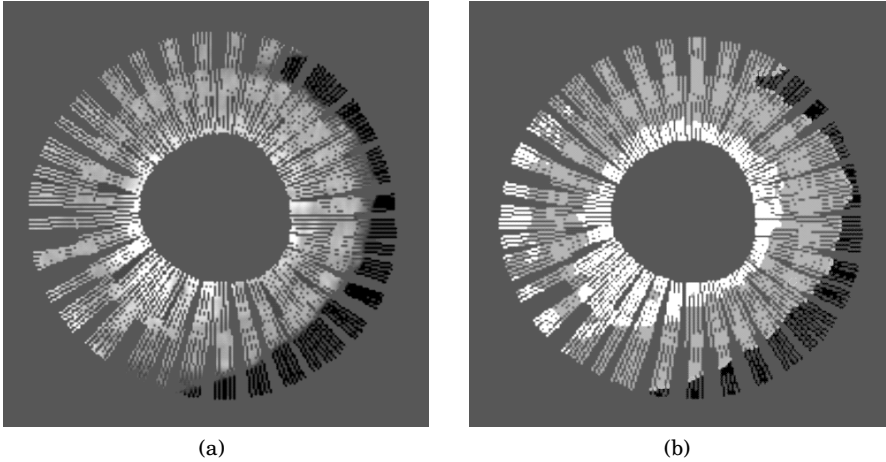


Figure 4.4: Ring of image patches from a cardiac short-axis image slice. (a) The true pixel gray values. (b) A ring of image patches after classification by the Takagi-Sugeno-based fuzzy inference decision scheme.

1. input

For each intersection point between the mesh and each of the 2D images, an image patch, centered on this point, is considered. Thus, a large number of patches from the stack of images are sampled (see Fig. 4.4). Patch size was selected such that myocardium pixels, pixels from the blood pool and/or air in the lung are included in the patch.

2. fuzzification

To determine the location of the transition from blood to myocardium or from myocardium to air, gray values are first classified. To ensure modality independence, only relative gray value differences between blood, myocardium and air are used. This makes Fuzzy C-Means (FCM) [70] clustering a suitable algorithm to distinguish between three classes, and use the class transitions as borders. Fuzzy C-Means is an unsupervised clustering algorithm that has previously been applied to segmentation of short-axis cardiac MR data sets. In Rezaee et al. [71], FCM is used in two different segmentation approaches, a stand-alone application of FCM, and a multi-resolution technique based on pyramidal segmentation where FCM is used in the merging of multiple clusters. In our application, pixels are labeled by minimizing a generalized within group sum of squared error objective function

$$J_m(U, V; X) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m * (\|x_k - v_i\|_I)^2 \quad (4.8)$$

where $u_{ik} : X \rightarrow [0, 1]$ defines the membership degree of a pixel x_k to the i -th cluster center v_i . The parameter m ranges from 1 to ∞ and controls the fuzziness of the resulting partition, the matrix I defines the norm (Euclidean). The number of classes used in the FCM was set to three: bright, dark, and medium bright, which represent blood pool, air and myocardium respectively.

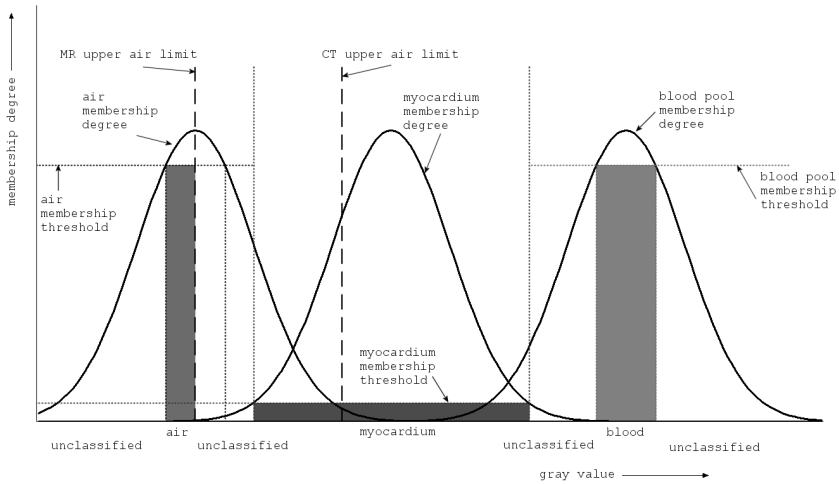


Figure 4.5: Defuzzification in the Takagi-Sugeno-based fuzzy inference system, schematically.

FCM was applied to all image patches extracted at the intersection rings from all images simultaneously.

3. inference (see Figs. 4.5 and 4.6)

For each pixel, three fuzzy membership degrees (FMDs) result from the fuzzification, above. Based on the FMDs, the inference step looks like:

- *for each pixel*
if (gray value is bright) then pixel is blood pool
if (gray value is medium) then pixel is myocardium
if (gray value is dark) then pixel is air
- *for each line (i.e., horizontal line in Fig. 4.6)*
if (majority of pixels is class i) then line is class i
else line is unclassified
- *transition*
endocardial border *from outside to inside, find first transition from myocardium to blood pool*
epicardial border
 - *at the septum*
from inside to outside, find first transition from myocardium to blood pool
 - *at the lung, anterior and posterior border*
from inside to outside, find first transition from myocardium to another tissue

Defuzzification can be performed, e.g., by assignment of pixels to the tissue class for which the pixel has the highest FMD. Thus, no parameters need tuning. However, in

our data, hyper- and hypo-intense regions in the data required a more complex Fuzzy Inference System. These regions attribute to the ill-balanced input population to the FIS. The severity of the balance problem depends heavily on the location of the 3D-ASM in the target data set. E.g., closer to the base of the LV, more blood is present in the images, whereas closer to the apex, myocardium and surrounding tissues dominate the spectrum.

Fuzzy C-Means clustering produces a balanced partitioning of the input population into a number of classes. However, a balanced partitioning does not always correspond to the reality in the input population, which is not necessarily balanced with respect to the classes present. In step 3 above, this can be tackled by application of minimum membership thresholds for the different classes to favor classification to one class with respect to other classes in the defuzzification. Another solution is the definition of a gray value threshold below or above which certain tissue types should not occur. In this application, this threshold is defined as the gray value at a given proportion p between the class centers of the dark and the medium tissue class (resulting from step 2, fuzzification)

$$g^{v_t} = g^{v_{dark}} + p * (g^{v_{medium}} - g^{v_{dark}}) \quad (4.9)$$

where g^{v_t} denotes the gray value of the threshold, $g^{v_{dark}}$ and $g^{v_{medium}}$ denote the gray values of the class centers of the dark and medium classes respectively, and p is the proportion that is used to move the threshold. Below the threshold, all is dark, above it, dark cannot occur. If the pixel is not dark, both for MR and CT, it is assigned to the tissue class for which it has the highest FMD, provided this FMD is above the tissue's membership threshold. If no tissue can be assigned using the above rules, the pixel remains unclassified. This defuzzification is shown schematically for MR in Fig. 4.5. Besides the classification for MR, it shows the air tissue cut-off value for both MR and CT, according to (4.9).

In pilot experiments, classification results were shown masked with a manual indication of the tissues, thus high-lighting the differences. These pilot experiments revealed that if pixels were assigned to the tissue class for which the pixel has the highest FMD, the blood pool and air tissues were too "greedy" for the MR case, i.e., claiming too many pixels with respect to the myocardium class. The opposite was observed for CT. This is a consequence of the ill-balanced data in the input population to the FCM. For a better class separation, both a minimum membership degree threshold and a gray value threshold were implemented in our defuzzification scheme. Depending on a visual evaluation of the masked classification differences, the combination of minimum membership degree thresholds and the gray value threshold that gave the best overall performance was selected.

4.3.4 Robust update selection

Pixels are only classified if they clearly belong to one tissue class. If a pixel does not have the required "minimum membership" value for any tissue class it is not classified and not considered for inference (see thick arrow in Fig. 4.6). To assert that model updates are solely based on reliable edge information, a three-stage algorithm was developed. The first step was the classification using FCM on image patches of all cardiac slices together (3D volume), rather than to single pixel strips, as explained above. Secondly, in the inference step, a majority voting along each patch line (horizontal line

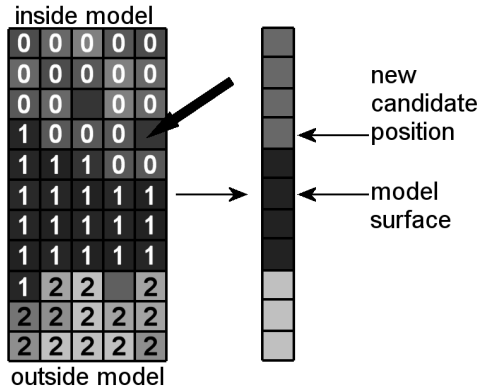


Figure 4.6: On the left, an image patch with classified pixels is shown. The pixels are classified to blood (0), myocardium (1), and air (2) or none (empty) (part a of the inference step). On the right, a single strip is left after majority voting (according to number 2 of the inference step), each square representing the label of the corresponding horizontal line on the left. The squares below are air, the squares in the middle are myocardium, and the top squares are blood pool. On the right, the current position and the candidate position of the model surface are indicated by arrows.

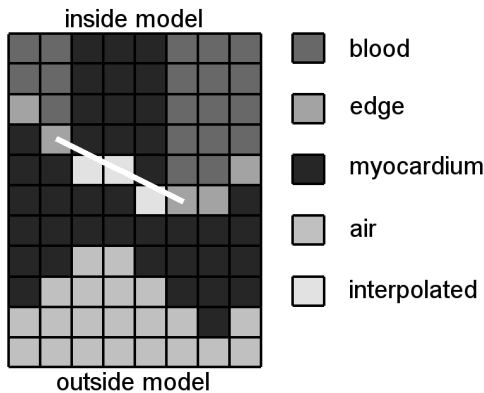


Figure 4.7: Candidate position interpolation. Eight classified image patches (after majority voting) aligned and positioned next to each other. In the third, fourth, and fifth patch from the left, no transition between myocardium and blood could be detected, because the myocardium tissue extends to the end of the patch on the inside of the model. Candidate positions in these patches are generated by interpolation (white line) between the detected tissue transitions in the second and sixth patches from the left, in which candidates *had* been generated by the fuzzy inference system.

in the patch on the left in Fig. 4.6) is performed to derive a single tissue class per patch line, compressing the patch to a single line (see Fig. 4.6). Thus, local contextual information is used to improve robustness. The new candidate point resulting from this decision scheme is located at the transition found in the last step of the scheme above ("transition"). If no reliable update can be deduced in the neighborhood of a mesh intersection (i.e. in an image patch), information from a larger neighborhood around the problem region can generate an update location for this region. In that case, the edge position was defined by interpolation between update positions from the nearest reliably processed image patches (see Fig. 4.7).

4.4 Experimental setup

4.4.1 Training and testing data

For model training, expert drawn contours of a group of 53 patients and normal subjects from short-axis 3D MR data were used. During the manual delineation, the papillary muscles were considered part of the blood pool. The apical end of each data set was defined as the first slice that exhibits both a complete endocardial and epicardial border. The basal end of each data set was marked by the last slice that showed a complete circumferential endocardial contour. Thus, the mitral valve plane and the true apex were excluded, leaving both ends open. The shape parameterization presented in Section 4.3.1 was applied (see Fig. 4.1), where each sample was resampled to 16 slices, each containing 32 points for the epicardial contour and 32 points for the endocardial contour. To reduce model dimensionality, the model was restricted to represent 99% of the shape variation present in the training data, resulting in 33 modes of variation.

To test the performance of the ASM, two clinical evaluation studies for two different modalities were performed. One data set consisted of multi slice CT cardiac left ventricle short-axis reconstructions (25 patients); the other consisted of MR cardiac left ventricle short-axis acquisitions (15 healthy subjects). Both data sets involved the end-diastolic phase of the cardiac cycle. The MR data set was acquired with the balanced FFE-protocol using the Q-Body coil (12 subjects) or a dedicated cardiac coil (3 subjects) on a Philips Gyroscan NT Intera, 1.5T scanner (Philips Medical Systems, Best, Netherlands). Images had in-plane resolution of approximately 1,5 mm/pixel and 6 mm slice thickness plus 4 mm slice spacing and, typically requiring 10 to 11 slices to cover the entire cardiac LV. The CT data sets were acquired with multi-slice CT scanners from two different vendors, with slice thickness on the order of 2 mm, and in-plane resolution of approximately 0.5 mm/pixel. All CT data sets were reformatted to yield short-axis image slices with slice thickness of approximately 5 mm and in-plane resolution of approximately 0.4-0.5 mm/pixel. The resulting data sets typically involved 16 to 28 image slices containing LV data.

4.4.2 Matching parameters

Prior to matching, the model pose was initialized manually. First, the origin of the model (which is its center of mass) was positioned in the center of mass of the image data stack. Following that, a manual offset in model position and orientation was ap-

plied to the model. Model position was initialized inside the cardiac LV, more towards the LV base rather than the LV apex. Model scale was chosen so that the model was not larger than the actual LV in the image data. Model shape was initialized to cover a large number of slices (typically 7-11 for MRI, 14-21 for CT) from the start. For all CT and MR studies the model shape was initialized identically (i.e., having the same initial b-vector). The class centers of the tissue classes used by FCM were initialized identically for each data set (but different between CT and MRI) and for each iteration of the model matching process. The stop criterion of FCM $\varepsilon = 0.1$. This means that FCM iteration stopped when the difference between two class centers from successive iterations was smaller than ε for all tissue classes.

The fuzzification and defuzzification (part 3 of the TSFIS) use slightly different settings for CT and MR studies. For CT, fuzzification is performed using four tissue classes (one for blood, one for myocardium and two classes for air and other darker structures, which are combined before defuzzification), while for MR five classes are used (three for blood (which are combined before defuzzification), one for myocardium and one for air and other darker structures).

The values for the minimal membership degrees and the gray value threshold were deduced by performing several pilot experiments. Taking equal minimum membership degree thresholds of 0.5 for each tissue class, the findings from the pilot experiments implied raising the threshold for the blood and air classes in MR, and for the myocardium class for CT. Consequently, for CT, the minimum membership degree for a pixel to be effectively classified is 0.5 for blood and air and 0.6 for myocardium, whereas for MR, these values are 0.7 for blood and air, and 0.5 for myocardium. The gray value threshold settings in (4.9) were for MR, $p = 0.0$, and for CT, $p = 0.75$.

During model matching, deformation was limited by constraining each component of the model deformation parameter vector between -3σ and $+3\sigma$. The parameters of our model with the values as we used them in our experiments (and their ranges, if applicable) are presented in Table 4.1.

In ASM, there is no built-in criterion that expresses how well the model fits the ground truth. A stop criterion based on such a fitness factor, is therefore not feasible. In this work, we adopted the approach described by Van Ginneken *et al.* [21], setting a fixed number of iterations. Van Ginneken used $N_{max} = 40$ (N_{max} is the maximum number of iterations) in a multi-resolution approach to a 2D segmentation task ($N_{max} = 10$ for each of four resolutions). Our 3D-ASM was set to iterate for a fixed number of iterations of $N_{max} = 100$. All analyses were performed using the contours from the last iteration.

4.4.3 Quantitative assessment indices

For both evaluation studies, four types of analysis were performed:

- Point-to-point distances

For each image slice, a fixed number ($N=100$) of radial point-to-point distances between automatic and manually drawn contours were measured. During manual delineation, the papillary muscles were considered part of the blood pool, as they were in the delineation process in the training stage. From all distances per subject, and separately for the endocardium and epicardium, average and maximum distances were calculated to give an idea of the global fit of the model to the data.

Table 4.1: Parameters of the 3D-ASM and their values (ranges between brackets)

ASM		
Number of modes of variation	33	
Allowed variation per mode	3σ	
Image patches		
	CT	MRI
Width	5 pixels	5 pixels
Length	14.5 mm	19.5 mm
FCM		
Maximum number of iterations	15	
Stop criterion ε	0.1	
Initial membership degrees u_{ik}	0.1 (0 - 1)	
m	2 (1 - ∞)	
Initial tissue class center gray value		
	CT	MRI
Blood pool 1	1250	150
Blood pool 2	-	140
Blood pool 3	-	130
Myocardium	1000	40
Air 1	600	20
Air 2	0	-
Defuzzification		
	CT	MRI
Air cut-off proportion (p in (4.9))	0.75	0.00
Blood pool membership threshold	0.50 (0.0 - 1.0)	0.70 (0.0 - 1.0)
Myocardium membership threshold	0.60 (0.0 - 1.0)	0.05 (0.0 - 1.0)
Air membership threshold	0.50 (0.0 - 1.0)	0.70 (0.0 - 1.0)

- Clinical contour quality analysis

The clinical contour quality represents a measure for the actual workload for a physician. It is based on the idea that the physician may set a certain maximum distance threshold between expert and automatic contours. Automatic contours are accepted if the distances to manual contours remain below the chosen threshold. Among those distances, a number of outliers (exceeding the threshold) were allowed, varying between 0%, 5%, and 10% outliers. The distance threshold was varied between 2 mm and 10 mm. Single contour quality was expressed as the percentage of accepted automatic contours from a data set. The percentage of accepted contours represents a trade-off between the desired accuracy of the 2D contours and the amount of work for the physician to manually correct the contours. This trade-off is to be made by the physician. The contour quality is evaluated because it relates closely to the clinical applicability of the method.

- Volume regression analysis

Volumes inside the automatic contours were plotted against volumes inside the manually drawn contours, and a linear regression formula was calculated, to-

gether with a correlation coefficient r . Volumes are calculated using Simpson's rule, accumulating areas (in cm^2) over all contours and multiplying with the slice thickness. Separate analyses for blood pool volume (endocardial contours) and LV volume (epicardial contours) were performed.

- Bland-Altman analysis

Bland-Altman analyses were performed on the volumes derived from the expert drawn contours and those from automatic contours, and 95%-confidence-intervals were calculated. Plots were made to evaluate whether a trend is present in the observed volume errors of the automatic volumes with respect to manual volumes.

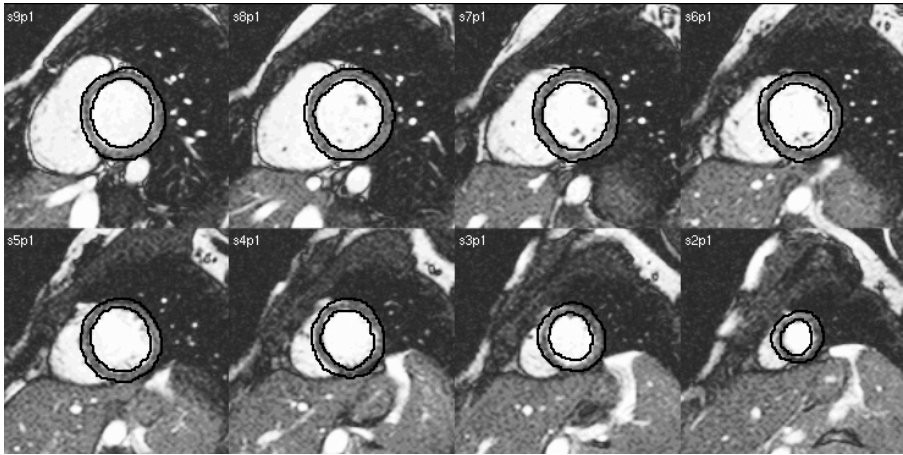


Figure 4.8: Typical contours generated by the 3D-ASM. Shown in an MRI acquisition.

4.5 Results

For all the MRI data sets, the 3D-ASM visually reached a stable state before the final iteration. From the initial CT data set of 25 patients, 2 matches were excluded from further quantitative evaluation due to matching failure. For the remaining 23 CT data sets, the 3D-ASM visually reached a stable state before the final iteration. For CT data, processing time for 100 iterations was 625 ± 141 seconds (min. 278 sec., max. 984 sec.). For MR, this was 157 ± 13.8 seconds (min. 140 sec., max. 175 sec.). Processing times were measured on an AMD Athlon XP, 1.8 GHz machine. A typical example of an automatic segmentation of a cardiac MRI data set is shown in Fig. 4.8.

4.5.1 Quantitative data

The ranges of the average and maximal distances between the manual and automatic contours are presented in Table 4.2. Figure 4.9 shows the percentage of accepted automatic contours as a function of the maximal allowed distance to the expert drawn

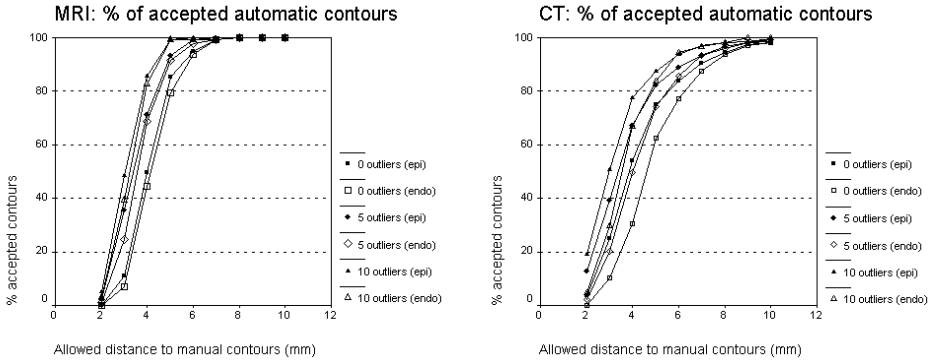


Figure 4.9: Average percentage of accepted contours, determined by the maximum allowed distance to the manually drawn contours along 100 radial lines and with 0%, 5% or 10% outliers. a) MRI studies. b) CT studies.

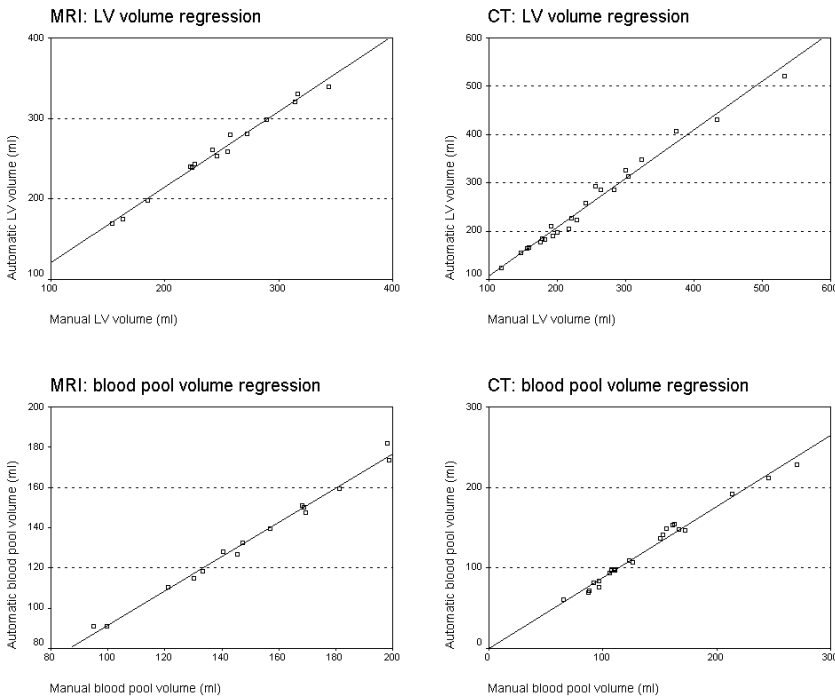


Figure 4.10: Regression of automatic volumes on manual volumes. a) LV volume from MRI studies b) LV volume from CT studies. c) Blood pool volume from MRI studies. d) Blood pool volume from CT studies.

Table 4.2: Ranges of the average and maximum distances measured per subject between manual and automatic contours (mm). Averages over all subjects are between brackets.

		CT	MRI
Average	endo	1.31 - 2.60 ($\mu = 1.85$)	1.34 - 2.05 ($\mu = 1.72$)
	epi	1.02 - 2.97 ($\mu = 1.60$)	1.27 - 1.85 ($\mu = 1.55$)
Maximum	endo	4.2 - 12.7	4.3 - 7.6
	epi	3.3 - 12.7	4.0 - 7.4

contours. The results for the volume regression analyses are shown in Figure 4.10. For the LV volume, the regression line for MRI was described by $y = 0.94x + 27.8$ ($r^2 = 0.99$), whereas for CT this line was described by $y = 1.01x + 7.48$ ($r^2 = 0.98$), with y the automatic and x the manual volume. For the blood pool, the regression formula for MRI was $y = 0.85x + 6.62$ ($r^2 = 0.99$) and for CT $y = 0.88x + 0.38$ ($r^2 = 0.98$).

The Bland-Altman analyses show a slight systematic overestimation of LV volume and a systematic underestimation of the blood pool volume, for both CT and MRI (see Fig. 4.11). Blood pool volume errors increase with increasing volumes (see Figs. 4.11(c), 4.11(d)).

4.6 Discussion

The results from the global point-to-point distance measures (see Table 4.2) show that, on average, segmentation results from our 3D-ASM do not deviate much from the manual contours. Average distances range between 1.02 mm and 2.97 mm for CT, and between 1.34 mm and 2.05 mm for MRI. For CT these distances translate to approximately 1.5-4.5 pixels, while for MRI these distances are on the order of 1 pixel or less. The maximal deviations of the automatic segmentation results from the manual contours are in the order of 5-18 pixels for CT, and in the order of 2-4 pixels for MRI. These numbers show that globally the automatic segmentations are very close to the manual ones, while localized larger deviations may occur. The average unsigned distances over all subjects observed in MRI compare favorably to those observed in the 3D-AAM by Mitchell *et al.* [31], i.e., 1.72 mm vs. 2.75 mm for the endocardium and 1.55 mm vs. 2.63 mm for the epicardium.

With respect to the results of the clinical contour quality studies, we observed the following. The chart in Fig. 4.10(a) shows that more than 79.3% of automatic MRI endocardial and epicardial contours do not deviate more than 5 mm anywhere from the manually drawn ones (~ 2.5 pixels). Moreover, the chart shows that epicardial contours are better than endocardial ones. This result did not show in the global evaluation from the point-to-point distances (Table 4.2). The percentage of accepted epicardial contours was 85.4% vs. 79.3% for endocardial contours. If 5% outliers were allowed, 93.2% of the epicardial contours were accepted, vs. 91.4% of the endocardial contours. This means that for MRI, in more than 91.4% of the contours little or no corrections to the automatic contours are necessary if localized contour positioning errors of approximately 2.5 pixels are considered acceptable.

For CT, also epicardial contours were better than the endocardial ones. Average percentage of accepted contours reached 75.1% at allowed deviations of 5 mm for epicardial contours, and 62.6% at allowed deviations of 5 mm for endocardial contours (see

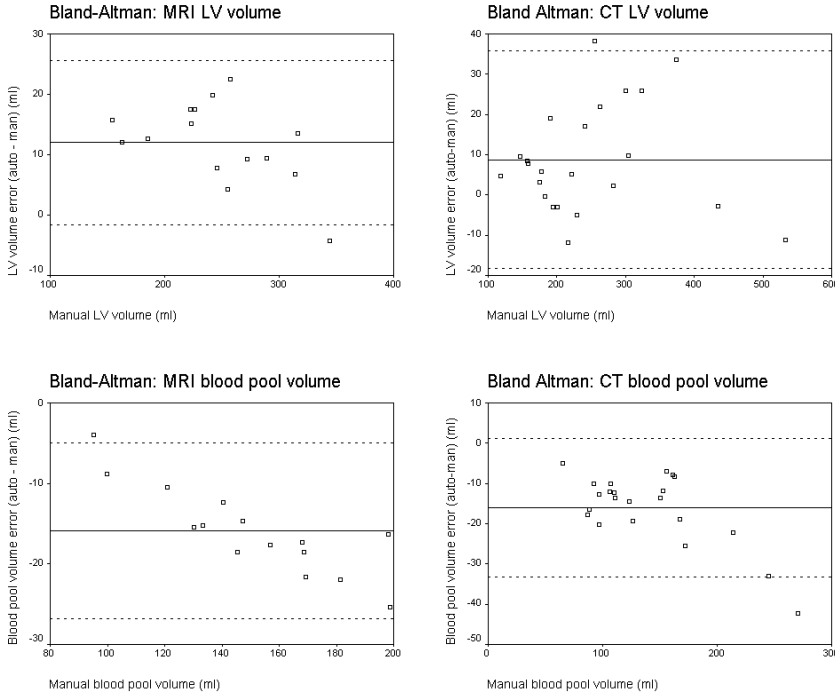


Figure 4.11: Bland-Altman analyses of automatic volumes and manual volumes. a) LV volume from MRI studies b) LV volume from CT studies. c) Blood pool volume from MRI studies. d) Blood pool volume from CT studies. The solid lines show the mean errors, the dashed lines show the mean $\pm 2\sigma$.

Fig. 4.10(b)). When allowing 5% outliers, these numbers are 82.4% and 74.1% respectively.

Classification of pixels into a number of tissue classes is based on a 3D volume around the surfaces of the model mesh. In MR, due to partial volume effects, the boundary between blood and myocardium tends to exhibit fuzziness. Either turbulent or slow blood flow in the vicinity of the myocardium causes inhomogeneities in the blood pool gray values. This may lead to misclassification in particular of pixels belonging to the blood pool, which was tackled by reserving multiple classes in the TSFIS for the blood pool tissue. In CT, with injection of a contrast agent, inhomogeneities in the blood pool as mentioned above, do not occur, and thus this is no issue.

However, with CT, tissues of other organs close to the heart appear slightly darker than myocardium, but brighter than air. Consequently, epicardial edges are either very fuzzy or absent in the inferior section. In slices closer to the apex, epicardial edges are sometimes absent altogether, increasing the possibility that deviations between automatic and manual contours exceed the 5 mm threshold. If no outliers are allowed and the deviation from manual contours is more than 5 mm in just one sample point, the contour is rejected completely. It has to be recognized that the acceptance criteria allowing only 5% outliers or less, are very strict, and consequently shows lower acceptance rates than might be expected based on the results of the point-to-point dis-

tances. To avoid misclassification, an extra tissue class may be added to the TSFIS to capture these structures.

Allowed deviations of 5 mm can seem very large in CT, with an in-plane resolution of about 0.5 mm/pixel. However, the CT short-axis data sets have been reconstructed from axial slices. The axial data sets had an in-plane resolution of about 0.5 mm/pixel, and a slice thickness of approximately 2 mm. Applying the model to the short-axis reconstructed CT data, which was the only CT data that was available to us, we were not able to benefit from the true resolution of the original CT scans. Thus, the high-resolution CT data, exhibiting a nearly isotropic nature, could not be exploited in this application of the 3D-ASM. Moreover, due to image reconstruction from axial data, in some cases reconstruction artifacts appeared. These artifacts were caused by displacement of the heart between subsequent axial slices.

Figures 4.10(a)-4.10(d) reveal good linear correlations between manually and automatically segmented volumes, both for blood pool volumes and LV volumes. Correlation coefficients for blood pool and LV volumes with expert derived volumes for both CT and MRI were very high, $r^2 \geq 0.98$. The high correlation between automatic and expert derived volumes is a consequence of the volumetric model-based approach, where local inaccuracies are averaged out. Thus, even in cases where the automatic segmentation might be locally inaccurate, still valuable information regarding volume-based parameters can be extracted. This observation also shows that our method is very robust. In cases where transitions between tissues are fuzzy, or visually not detectable, the model-based approach is able to generate local contour estimates, based on evidence from more reliable positions and the built-in statistical knowledge, that do not produce large volume errors.

Bland Altman analysis (Fig. 4.11) shows that both for CT and MRI LV volumes are overestimated slightly, but systematically, while blood pool volumes are underestimated systematically. Additionally, in the errors of automatic blood pool volume with respect to manual blood pool volume (Figs. 4.11(c)-4.11(d)) a negative trend appears. This is more apparent for MRI than for CT. Overestimation of the LV volumes can be a consequence of a number of causes:

- the presence of epicardial fat (CT):
in the CT data sets, the gray value of epicardial fat is in between those of myocardium and air, and closer to myocardium. In MR, epicardial fat appears as very bright. For CT, epicardial fat will be wrongly classified as myocardium, for MR as blood. In the latter case, due to the classification as another tissue than myocardium, this does not affect the segmentation. In CT however, it results in a "too large" ventricle.
- unreliable edge information at the inferior epicardial border (CT):
at the inferior epicardial border, edges between myocardium and other tissues are visually not detectable. In this location, it depends on the ability of the FI-system to generate reliable candidate points for the endocardial surface at the posterior section. An alternative approach may be the exclusion of the inferior epicardial border from the FI-system, calculating shape deformation based on a smaller set of (pseudo-) landmarks.

Underestimation of the blood pool can have the following causes:

- the presence of dark structures in the blood pool close to the papillary muscles: these dark structures fused together with the papillary muscles form big dark objects, that force the endocardial surface of the model to by-pass them on the inside, leading to the underestimation of contour area and consequently blood pool volume.
- locations with high curvature at the endocardial borders: in a few cases we studied, the endocardial borders have strong local curvatures, amongst others, due to local infarct wall thinning. The shape model is unable to follow the boundaries at these locations of high curvature. However, since the candidate positions for the model sample points generated during the model updates are located exactly at the borders, even at the locations of local wall thinning, these errors are probably due to shape deformation limitations in the model. A way to resolve these model-induced inaccuracies is by performing a subsequent, more local boundary detection step, such as dynamic programming.

The trend in blood pool volume error (Figs. 4.11(c)-4.11(d)) may be caused by the error sources mentioned above. For some patients, the manual segmented contours showed locations with high curvature, e.g., due to pathological wall thinning. If the model does not allow high curvature on its surfaces, either due to external deformation constraints or because these high curvatures have not been observed in the training set, errors in the automatic segmentation will occur. The contribution of those errors to the volume of the LV is proportional to the model scale. With a larger model (when matching on a larger heart) the errors will be larger (see Fig. 4.12), resulting in a trend in a Bland-Altman plot.

The curves in Figure 4.9 together with the very small average unsigned distances between manual and automatic segmentation results, give good confidence in the clinical usefulness of the 3D-ASM for automatic segmentation of the cardiac LV in the future. Since the 3D-ASM typically reaches a stable state after approximately fifty iterations, processing times can be reduced by a factor of 2, if a stop criterion based on stability in deformation, translation, rotation and scaling would be implemented. Other suggestions for convergence criteria are proposed by Kelemen *et al.* [35] depending on the size of the deformation of a surface. Cootes *et al.* [72] evaluate convergence by recording the number of times that the best found pixel along a search profile is within the central 50% of the profile. When a sufficient number of the points satisfies this, the algorithm is declared to have converged.

In the combined histograms of all image patches used in the FCM, plateaus were observed in the blood pool area for MRI and in the dark structures area (air, epicardial fat) for CT. Therefore, in application of the FCM to MRI, three classes, together representing the blood pool, were used to cover the flat part better than a single class could, thus diminishing classification errors. For the same reason, two classes were used to cover the dark structures plateaus in CT classification. The application of multiple classes for a single tissue strongly improved the segmentation results, decreasing the number of failures dramatically. Incorporation of such a-priori knowledge about parameters that vary among different modalities introduces a modality-dependence, which we strive to avoid. However by excluding a GLM, we managed to omit image modality dependence from the statistical training stage of the ASM. We only train the underlying PDM, such that no retraining to several new data sets was required to apply the model to CT and MRI. Nevertheless, to reliably segment unseen modalities and/or acquisition protocols, further generalization of the Fuzzy Inference system may

be required.

The resulting segmentation obtained by the 3D-ASM is sensitive to initial model pose, shape and scale. We observed that our model does not deform easily in the direction of the long-axis, which may be a cause for the sensitivity of the results to initial model pose parameters. This sensitivity does not imply that the model has a small lock-in range. Figure 4.13 shows the first and last state of the model in a particular match, where the model was not initialized close to the manual segmentation. This figure shows that even in such cases a good final segmentation can be obtained.

The effect of initialization on final segmentation results, however, was not investigated. Although it is an interesting topic in itself, the topic of this paper is the application of the 3D-ASM to different modalities, without any alterations in the statistical framework.

In this paper, we applied our 3D-ASM to both CT and MRI data sets. Both modalities show the same ordering of tissues with respect to their gray levels (gv), i.e., $gv_{blood} > gv_{myocardium} > gv_{air}$. This model can equally well be applied to other modalities/acquisition protocols, that do not exhibit the same ordering, provided that classification of tissues in the resulting images is feasible. In such cases, only the defuzzification and rules part (3) of the Inference step in the TSFIS (and perhaps the number of classes used in the FCM) needs adjustment. Because data sets acquired with 3D ultrasound may be hard to classify, segmentation using this approach may give difficulties, but this will have to be investigated in detail.

With respect to the definition of 3D point correspondence, in this research an application specific point correspondence was used. However, there are no fundamental barriers to replace this point correspondence definition with an automatically determined or optimized one, as it does not have a major influence on implementation of the other parts of our method: this is a topic of ongoing work.

4.7 Conclusion

The presented 3D-ASM is a powerful tool for automatic segmentation of the cardiac left-ventricular myocardium. In this paper, it has been shown that with minimal adjustments in tuning parameters of the model, a single model can achieve good quality contours for both CT and MRI data sets. This means that no labor-intensive retraining is required for application to different modalities, or different acquisition protocols, provided that the image data sets have a homogenous gray value distribution among different tissues.

Though further research is still required, the experiments presented here demonstrate that with one 3D-ASM with a model update generation step independent of absolute gray value distributions, acceptable automatic segmentations of the cardiac LV can be achieved for multiple modalities, and protocols, without statistical retraining of any kind.

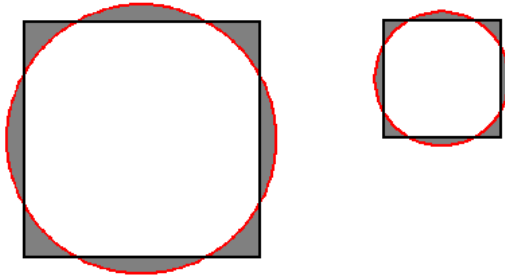


Figure 4.12: Example of area errors due to curvature constraints. Suppose that the square is the manual segmentation with high curvature and that the ellipse is a curvature constrained model. On the left, the errors marked in gray are larger than on the right due to a larger shape (and so a larger model), while the curvatures are the same in both cases.

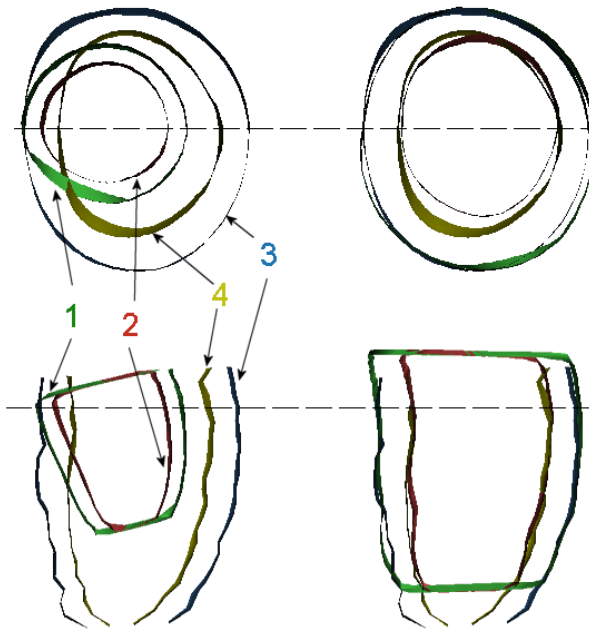


Figure 4.13: Initial (left) and final (right) state of the model in a match where initialization was far from the manual segmentation. Top: short-axis view; bottom: long-axis view. The dotted lines indicate the intersection plane of the other view. With respect to the numbers: 1 and 2 indicate the model's epicardial and endocardial surfaces, and 3 and 4 indicate the manual epicardial and endocardial surfaces respectively.