3D active shape modeling for cardiac MR and CT image segmentation
Assen, Hans Christiaan van

Citation

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).
SPASM: a 3D-ASM for Segmentation of Sparse and Arbitrarily Oriented Cardiac MRI Data

Reprinted from:
SPASM: a 3D-ASM for Segmentation of Sparse and Arbitrarily Oriented Cardiac MRI Data
Medical Image Analysis, 10(2): 286-303, ©2006, with permission from Elsevier
Abstract

A new technique (SPASM) based on a 3D-ASM is presented for automatic segmentation of cardiac MRI image data sets consisting of multiple planes with arbitrary orientations, and with large undersampled regions. Iterative updates were calculated using an update propagation scheme using a Gaussian weighting kernel. Thus, updates for the void data locations on the model surfaces were obtained. Feature point detection is performed using a Fuzzy Inference System, based on Fuzzy C-means clustering of image patches, extracted around the model surfaces. Model parameters were optimized on a computer cluster and the computational load distributed with grid computing.

SPASM was applied to sparse image data sets composed in three different ways, comprising images with different orientations (radial long-axis, short-axis, and 2- and 4-chamber views) and stemming from different MRI acquisition protocols. Segmentation performance was compared to manual segmentation. For all data configurations, (sub-)pixel accuracy was achieved. Performance differences between data configurations appeared to be statistically different ($p < 0.05$) for some, but not clinically relevant. Comparison to results from other 3D model-based methods showed that SPASM is comparable to or better than these other methods, but SPASM uses considerably less image data. Sensitivity of the segmentation performance to initial model placement proved to be limited within a range of position perturbations of approximately 20 mm in all directions.

7.1 Introduction

7.1.1 Purpose

Nowadays, cardiac MRI and CT are increasingly used for cardiac functional analysis in daily clinical practice. Both modalities yield dynamic 3D image data sets. With CT, images are acquired in an axial orientation and for cardiac analysis, usually short-axis views are reconstructed from the axial image data. With MRI, images can be acquired in any spatial orientation. Commonly used orientations are Short Axis (SA) views [85], and long-axis (LA) views (2-chamber and 4-chamber). These orientations are parallel or perpendicular to the LV long-axis, hence the partial-volume effect in visualizing the myocardial wall is minimal. The short-axis acquisitions consist of a full stack of typically 8 to 12 (parallel) slices covering the heart from apex to base. However, there is an ongoing debate on potential improvement of functional measurements by using long axis views or radially scanned long-axis image slices, since they appear to give better volume quantification due to better definition of the apex and base [4]. Also, depending on the structure of the image data, sometimes assumptions about ventricular geometry have to be made, especially for quantification from sparse radial stacks and long-axis views.

For quantitative analysis of cardiac function, typically a cardiologist or radiologist manually segments the images. After segmentation, measurements of global and regional functional parameters can be performed, such as wall thickening or wall thinning, LV blood pool volume and Ejection Fraction (EF). Due to the increasing amount of data, the amount of work for manually delineating the image data has become prohibitively large, and automated segmentation is highly desired.
7.1 Introduction

7.1.2 Background

Recent work has shown that integration of prior knowledge into medical image segmentation methods is essential for robust performance. Many recent methods utilize a statistical shape model, and the seminal work of Cootes [19, 28] on 2D Active Shape Models (ASMs)- and Active Appearance Models (AAMs) has inspired the development of 3D ASMs [26, 74] and, 3D AAMs [31]. These models capture the shape variations of a set of corresponding landmarks, and are fitted to the image data within statistically derived deformation limits. Along similar lines, 3D Spherical Harmonics (SPHARM) [35], 3D Statistical Deformation Models (SDMs) [18, 32, 38, 39] and 3D medial representations (m-reps) [44] have been described. These mainly differ from each other in the way the statistical shape variations are parameterized. Lorenzo-Valdés et al. [86] presented a cardiac 4D probabilistic atlas, which models the probabilities that a voxel belongs to LV, RV, myocardium or any of the surrounding structures. However, to fit all these mentioned statistical models to image data, the matching mechanisms are either based on a dense volumetric registration of an intensity model [18, 31, 32, 38, 39, 86] or on a dense set of updates along the model surface [26, 44, 74].

Many references to registration techniques used with dense models and sparse data sets can be found in Koikkalainen and Lötjönen [87]. They evaluate a number of rigid and surface-based registration techniques using sparse data and dense models for magneto- and electroencephalographic studies. Van ’t Ent et al. [88], however, applied a model based on spherical harmonics to the head for segmentation of scalp, skull and brain. A set of basis surfaces was extracted by means of Singular Value Decomposition on the parameter matrices of the spherical harmonic surface representations of the isotropic dense data sets. The model was applied to a few differently oriented MR slices for model-based recovery of the scalp, skull and brain surfaces. The surfaces of the model are composed of a combination of basis shapes, the coefficients of which are estimated by a combination of nonlinear and linear least squares fitting techniques. In [82], a densely sampled statistical shape model of the femur is matched to sparsely sampled image data for computer-assisted anterior cruciate ligament surgery. The method starts with finding the closest model point for each data point. The distances between the sparse data points and the closest model points are minimized by estimation of the parameters of the combined rigid and non-rigid transformations of the model. The minimization algorithm combines a simulated annealing technique with a downhill simplex algorithm. Both Koikkalainen [87] and Fleute [82] apply a statistical model, trying to find the modeled shape variation such that the sum of squared distances between the measured data points and the model is minimal. As a result, they have to solve the point-to-surface correspondence and update this correspondence for each iteration to produce a goodness-of-fit measure that is used to iteratively evaluate the parameter updates. This is achieved by propagation of the derived update information over the model surfaces.

Montagnat et al. [89] reported matching a model to 4D echocardiographic radially oriented imagery of one subject, based on zero-order statistics (mean shape) and controlled by internal regularization and external forces. The ultrasound imagery is preprocessed using anisotropic diffusion to remove edges due to noise. Radial trilinear data interpolation in sparse areas was performed to produce gray values in void data areas, which led to lower resolutions farther from the LV long-axis. The method was applied to one data set and lacked a ground truth, therefore no quantitative or statistical evaluation was reported and estimation of the segmentation performance could not be performed. Montagnat et al. applied their method to a radial stack of 9 slices, which
is the minimum number of slices required to achieve clinically acceptable results for the assessment of cardiac function according to Bloomer et al. [4]. This means that the model is not applied to a sparse data set. To our knowledge, recovery of the shape of increasingly sparsely sampled organs (in particular the heart) using a statistical shape model based on a dense mesh has not been reported yet. The same holds for systematic investigation of the influence of data sparsity on the reconstruction accuracy of organ surfaces, which will be performed in Chapter 8.

The main contribution of this work is the development of a 3D-ASM that removes the requirements on data sampling and:

- is applicable to sparsely sampled data sets without making assumptions about voxel isotropy or parallel slices (see Fig. 7.1),
- enables volumetric quantification from sparse data through a statistically driven shape interpolation instead of through simplified assumptions about ventricular geometry,
- is extensible to other modalities without retraining the shape model [65, 74].

To accomplish this we present SPASM, a 3D-ASM of the cardiac left ventricle (LV) that is able to segment sparsely sampled and arbitrarily oriented 3D cardiac LV data sets. The underlying statistical shape model was based on a 3D-atlas that was constructed using non-rigid registration [18, 75]. Matching of the model to sparse, arbitrarily oriented image data is accomplished through a deformable mesh that enables propagation of image updates over the model surface. This way, a statistically driven shape interpolation is realized. Independence of a trained gray level model is achieved through a Takagi-Sugeno Fuzzy Inference System (TSFIS) [61] for determining iterative model updates based on relative intensity differences [74].

Another main contribution is the extensive evaluation of the accuracy of cardiac surface reconstruction on different data configurations with different sparsity.

### 7.2 Methods

Active Shape Models were introduced by Cootes et al. [11, 19] and consist of a statistical shape model (often referred to as Point Distribution Model (PDM)) and a matching algorithm. The PDM is trained from a population of typical examples of the target shape, and models shape variability as a linear combination of a mean shape, i.e. a mean set of (pseudo-)landmarks, and a number of eigenshapes.

#### 7.2.1 Shape Modeling

To generate a statistical shape model, each shape in a training set is described using a set of sample points. For this, landmark points on expert segmentations are determined, being easily identifiable positions in the image data. Resampling of the manual segmentation between the landmark points yields the pseudo landmark points that build a shape vector

\[ \mathbf{x} = (x_1, y_1, z_1, ..., x_n, y_n, z_n)^T. \] (7.1)
Figure 7.1: (a) Radially oriented cardiac image stack, (b) a combined long-axis and short-axis data set, and (c) a short-axis image stack (only showing every second image).

(Pseudo) landmarks can also be determined in an automatic manner [18]. Shape differences can only be modeled if correspondence between points in different shapes has been established. The order of the sample point coordinates in the shape vector are determined by the point correspondence, and is therefore the same for all subjects.

From all shapes \( x_i \) in the training set (number of samples \( n_s \)) a mean shape vector

\[
\bar{x} = \frac{1}{n_s} \sum_{i=1}^{n_s} x_i \quad (7.2)
\]

and a covariance matrix

\[
S = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (x_i - \bar{x})(x_i - \bar{x})^T \quad (7.3)
\]

are calculated. From the covariance matrix \( S \), the eigenvectors and eigenvalues of the training set are computed using Principal Component Analysis (PCA). Depending on the amount of variation in the training set represented by the model, the eigenvectors \( \phi_i \) corresponding to the \( t \) largest eigenvalues are retained, represented by matrix \( \Phi = (\phi_1 | \phi_2 | ... | \phi_t) \). From this matrix and the mean shape, a shape \( x \) can be generated

\[
x = \bar{x} + \Phi b, \quad (7.4)
\]

with \( b \) a \( t \)-dimensional vector containing the model parameters

\[
b = \Phi^T (x - \bar{x}). \quad (7.5)
\]

By limiting the components of \( b \) within statistically derived boundaries (e.g., \( \pm 3\sigma \)) the resulting shape from Eq. (7.4) is expected to resemble the shapes seen in the training set. For an elaborate introduction to ASMs, the reader is referred to [11, 19, 72].
7.2.2 Atlas construction

A critical issue to achieve extension of PDMs to three and more dimensions is point correspondence: the landmarks have to be placed in a consistent way over a large database of training shapes, otherwise an incorrect parameterization of the object class would result. The methodology employed to automatically achieve this point correspondence of the heart was described in detail in [18]. The general layout of the method is to align all the images of the training set to a mean atlas (Fig. 7.2). The transformations are a concatenation of a global rigid registration with nine degrees of freedom (translation, rotation, and anisotropic scaling) and a local transformation using non-rigid registration. After registration of all samples to the mean shape, the transformations are inverted to propagate a topologically fixed point set on the atlas surface to the coordinate system of each training shape. While it is still necessary to manually segment each training image, this technique relieves from manual landmark definition. The method can easily be set to build either 1- or 2-chamber models; in this work we have used a 1-chamber model (see Fig. 7.3). To build the statistical shape model, the auto-landmarked shapes are aligned using Procrustes alignment [83]. PCA is performed on the remaining differences between equivalent landmarks on different shape samples. These differences are solely shape related.

7.2.3 Matching algorithm

The model described above was extended with a matching algorithm to apply it to image segmentation. A key design criterion behind this matching approach was applicability to data acquired with arbitrary image slice orientations, from different modalities (MR and CT), and to sparsely sampled data with arbitrary image slice orientations. This implies that:

- 2D image data alone is sufficient for updating the 3D model, to ensure applicability to arbitrarily oriented sparse data without assumptions about sampling density
- generation of update points is performed based on relative intensity difference to remove the dependency on training-based gray-level models.

To accomplish this, the landmark points are embedded in a triangular surface mesh. During the matching, this mesh is intersected by the image planes, generating 2D contours spanned by the intersections of the mesh triangles. To remove dependencies on image orientation or limited 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 projected onto the local surface normals. 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 Iterative Closest Point algorithm [59]. Successively, the parameter vector $b$ controlling model deformation is calculated. An adjustment to $b$ with respect to the previous iteration is computed, using both the candidate model state, $\hat{x}_{n+1}$, and the current model state, $x_n$

$$\hat{b}_{n+1} = b_n + \Delta b = b_n + \Phi^T (\hat{x}_{n+1} - x_n)$$  \hspace{1cm} (7.6)
with \( x_n \) representing the aligned current state of the mesh, and \( b_n \) representing the parameter vector describing the current shape of the model within the statistical bounds. The vector \( \hat{x}_{n+1} \) is the proposed model shape for the next iteration, and \( \hat{b}_{n+1} \) its shape parameter vector before statistical constraints have been applied.

In densely sampled data, a 3D data volume can be reconstructed that enables generation of a 3D update in each model landmark. However, in sparsely sampled data containing large undersampled regions, a (dense) 3D data volume cannot be reconstructed: interpolation between sparse image slices with different orientations (e.g., a radial stack of cardiac LA views) is non-trivial, if possible. In void locations, no information can be extracted from the image data to contribute to a new model instance. However, for the calculation of new model parameters, updates for the complete landmark set are required: setting updates of zero displacement would fixate the nodes to their current position, thus preventing proper model deformation.

### 7.2.4 Update propagation to undersampled surface regions

To overcome large void areas without update information, we propose a node propagation mechanism that distributes the updates from non-void update locations towards the void regions (see Fig. 7.4(a)). Such node updates are defined as (for node \( \omega \))

\[
v_{\omega,n+1} = \hat{x}_{\omega,n+1} - x_{\omega,n},
\]

(7.7)

where \( \hat{x}_{\omega,n+1} \) is the estimated position of node \( \omega \) for iteration \( n + 1 \) determined by the model update scheme (see Section 7.2.5) (i.e., before propagation, and after projection on the surface normal), and \( x_{\omega,n} \) is the position of node \( \omega \) for iteration \( n \).

During propagation, such source update vectors \( v_{\omega} \) are weighted in the receiving nodes \( \lambda \) according to the geodesic distance between source node and receiving node using a Gaussian-shaped kernel the width of which is defined by \( \sigma_p \) (see Fig. 7.4(b)). To avoid propagation of an update over the entire model surface, propagation is stopped when the geodesic distance exceeds a fixed threshold \( \chi \) (\( \chi \equiv 3\sigma_p \)). A Gaussian kernel has unity area under the curve, which is achieved by a normalization factor \((2\pi\sigma_p)^2\). However, in this paper, the weight of an update at its source location is normalized instead, omitting the normalization factor, leaving only the exponential term of the propagation kernel. As a result, the propagation weights for individual source node updates are defined as:

\[
w(\lambda, \omega) = \begin{cases} e^{-\frac{||\lambda - \omega||^2}{2\sigma_p^2}} & \text{if } ||\lambda - \omega|| \leq \chi, \\ 0 & \text{if } ||\lambda - \omega|| > \chi, \end{cases}
\]

(7.8)

where \( w(\lambda, \omega) \) is the weight at the location of the receiving node \( \lambda \), \( \omega \) is the source node, \( ||\lambda - \omega|| \) is the geodesic distance to the origin of the update, and \( \sigma_p \) is the width of the Gaussian kernel. Consequently, the weight at the source (\( \lambda = \omega \)) is unity.

If multiple paths exist from source node to receiving node, only the shortest path is used. Thus, a receiving node accepts propagated updates from any source only once. After all propagations stopped, a cumulative contribution from all source nodes \( \omega_m \)
7.2 Methods

Figure 7.2: Atlas construction: a set of final global ($T_g$) and local ($T_l$) transformations can take any sample shape of the training set to the atlas coordinate system (left). Once the final global and local transformations are obtained, they are inverted and used to propagate any number of arbitrarily sampled landmarks on the atlas, to the coordinate system of the original samples (right).

Figure 7.3: (a) Shape of the 1-chamber atlas constructed from MRI data. (b) Mean shape of the 3D-SPASM constructed from the atlas. Different colors indicate different sectors of the mean shape, that are assigned their own FCM operation within the model update scheme (see Section 7.2.5). The epicardial surface is represented with a wireframe, the endocardial surface is represented by a surface. (c) Apex view of the mean shape of the 3D-SPASM.

to each receiving node $\lambda$ is computed. Each receiving node has a list of weighted contributions from source nodes

$$v_{\omega_m} \cdot w(\lambda, \omega_m) \quad \forall \ m \in [0, M - 1]$$

where $M$ is the total number of update sources in the mesh. A total update per node is computed as the ratio between the sum of all contributions and the number of con-
Figure 7.4: (a) Propagation of single model updates. Update sources are indicated as large dots. Propagation from the update sources surrounded with a large circle is illustrated here. From a source, an update vector originates (short arrow). Updates are first propagated (longer solid arrows) to the nearest nodes in the mesh (surrounded with open squares). From the nodes, updates are propagated to adjacent nodes weighted with a Gaussian kernel. Secondary updates are indicated in dotted lines, tertiary updates in dashed lines. (b) Gaussian propagation of single model updates projected on the mesh. Top, no propagation, this shows the three update vectors from the first level triangle update (two times). One set of update vectors corresponds to the nodes surrounded with open squares in (a). In the middle, a propagation kernel sigma of 4 mm is used, and at the bottom sigma is 8 mm.

Contributing source nodes $K$. The total number of contributions from source nodes to a receiving node $\lambda$ is defined as

$$K_\lambda = \sum_{m=1}^{M} k_m \quad \text{where} \quad k_m = \begin{cases} 1 & \text{if } \|\lambda - m\| \leq \chi, \\ 0 & \text{if } \|\lambda - m\| > \chi. \end{cases}$$ (7.10)

The total update to node $\lambda$ is thus defined as

$$\mathbf{u}(\lambda) = \frac{1}{K_\lambda} \sum_{m=1}^{M} \mathbf{v}_{\omega_m} \cdot w(\lambda, \omega_m)$$ (7.11)

The complete update of the model is defined as the concatenation of all update vectors for single mesh nodes.
7.2 Methods

(a)

(b)

Figure 7.5: (a) Classified set of image patches from a radial image. (b) Classified set of image patches from a short-axis image. (A=LV blood pool, B=RV blood pool, C=myocardium, D=air, E=outside image patches)

\[ \mathbf{v} = (\mathbf{v}(1) \mid \ldots \mid \mathbf{v}(L)) \]  

(7.12)

with \( L \) the total number of nodes in the mesh. So, using Eq. (7.6), and acknowledging that

\[ \mathbf{v}_{n+1} = (\hat{x}_{n+1} - x_{n}) \]  

(7.13)

finally yields

\[ \Delta b = \Phi^T \mathbf{v}. \]  

(7.14)

7.2.5 Feature point detection using Fuzzy Inference

The mesh structure combined with the update propagation enables applications to sparsely and arbitrarily oriented data. To apply the model to different modalities without retraining, the matching algorithm should not employ any trained intensity model to generate the updates. Instead, we have developed an update mechanism based on a Takagi-Sugeno Fuzzy Inference System (TSFIS) [61], which uses Fuzzy C-Means clustering (FCM) on the gray values of a 3D volume patch surrounding the current instance of the model (see Fig. 7.5). As a result of the FCM three tissue clusters exist. For CT and multiple MRI protocols the relative ordering of the tissues with respect to their representation in gray values are the same: blood is brighter than myocardium which is brighter than air. This does not hold for black blood MRI, but it does for CT and the MRI protocols used here (see Section 7.3.1). During inference, the brightest cluster is assigned the label blood, the medium bright cluster is assigned the label myocardium and the darkest cluster is assigned the label air.

Due to surface coil effects inhomogeneities in the gray value distributions can exist. To avoid possible negative effects of inhomogeneities on the fuzzy clustering results, the ventricle shape is divided in multiple anatomical sectors (see Fig. 7.3) allowing for multiple clustering operations. Furthermore, each anatomical sector can be assigned
its own rule set in the inference part of the TSFIS. This approach has been described in [90], and can be summarized as follows:

1. **input**
   - extract a rectangular image patch at each intersection between the mesh and any 2D image. Gray values from the patches are pooled according to the sectors on the model surface.

2. **fuzzification**
   - Using multiple standard Fuzzy C-Means (FCM)[70] clustering operations for multiple anatomical sectors, gray values are classified based on the intensities of blood, myocardium and air.

3. **inference of model updates**
   - For each pixel, three fuzzy membership degrees (FMDs) result from fuzzy clustering, above. Based on the FMDs, a mesh update is inferred:
     
     (a) **defuzzification** (see Fig. 7.5)
     - for each pixel
     $$T(x, y) = \begin{cases} 
     \text{bloodpool} & \text{if } L(x,y) = \text{bright}, \\
     \text{myocardium} & \text{if } L(x,y) = \text{medium bright}, \\
     \text{air} & \text{if } L(x,y) = \text{dark},
     \end{cases}$$
     with $T(x,y)$ the tissue label and $L(x,y)$ the pixel gray value, both at coordinate $(x,y)$.

     (b) **transition inference**
     - **endocardial border** from outside to inside take the first transition from myocardium to blood pool
     - **epicardial border** from inside to outside take the first transition from myocardium to
       - any other tissue at the lung, anterior and posterior wall,
       - blood pool, at the septum.

In the inference part of the FIS, a crisp value per pixel is derived based on the fuzzy membership distributions of the tissue classes. First, all pixels with a gray value below a preset threshold value are considered to be *air*. The threshold value is determined as a proportion $p$ of the separation between the class centers of the darkest two clusters resulting from the FCM. Secondly, the labels of all other pixels are determined using membership degree threshold levels. If the membership degree for blood of a pixel is above the threshold level for blood, the pixel is assigned the label *blood*. If not, the pixel is assigned the label *myocardium* if the pixel's membership degree for myocardium is above the threshold level for myocardium. The same routine holds with respect to the air class. Pixels can be left unclassified. For the value of $p$ and the membership degree thresholds, see Table 7.1. Pixels are only classified if they clearly belong to one tissue class. If a pixel does not reach a preset minimum membership degree for any tissue class (see Tab. 7.1), it is not classified and not considered for inference. The patch size in the input part was selected such that multiple tissues were included in the patch.
7.3 Experimental setup

7.3.1 Test data and protocol

Data acquisition

A group of healthy volunteers was scanned with a 1.5T Gyroscan NT5 (Philips Medical Systems, Best, The Netherlands) MR scanner, using the Balanced Fast Field Echo (BFFE) and the Turbo Field Echo (TFE) protocols. For all scans and protocols, the QBody coil was used. A number of acquisitions with different slice orientations was performed during breath hold in end expiration. For 15 subjects, the following protocol was followed. First, a scout view and 2- and 4-chamber views were acquired. This was followed by acquisition of short-axis (SA) views with the BFFE protocol, yielding a stack of typically 10-12 parallel image slices per subject. Every image slice was acquired in a separate breath hold. Next, a radial scan (RAD) was performed comprising four LA image slices, with inter-slice angle of $45^\circ$ (see Fig. 7.6(a)). All four slices were acquired in the same breath hold with the TFE protocol (see Fig. 7.6(b)), thus avoiding breathing-induced slice shifts. Image slices had a $256^2$ matrix and covered a field-of-view of 300-400 mm. Slice thickness and slice gap for the SA acquisitions were 8 mm and 2 mm respectively; for the RAD TFE acquisitions, the slice thickness was 8 mm. For an additional set of 5 patients, 2-chamber, 4-chamber and SA views were acquired; radial scans were not available for these subjects. The data were used in different combinations of image orientations, a radial set (RAD, N=15), a multi-view set (MV, N=20), a full short-axis set (SA, N=15) and a partial short-axis set (SA-sub, N=15), which are explained in more detail in Section 7.3.2.

Quantitative evaluation

For segmentation performance comparison, LV contours were manually drawn in all data sets to serve as gold standard. The manual segmentations in the LV SA stacks were connected to form 3D surfaces. This was done by equiangular resampling of the manual contours and connecting the sample points from neighboring slices to form triangulated surfaces. To assess interobserver variability with respect to manual de-
Table 7.1: Optimal parameter set obtained through exhaustive search using grid-computing techniques. Values for the defuzzification and ASM parts are taken from [90]. Propagation parameters were optimized for the evaluation data set in this paper.

<table>
<thead>
<tr>
<th>Defuzzification (see Sec. 7.2.5, [90, 91])</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Air cut-off proportion ( p )</td>
<td>-0.20</td>
</tr>
<tr>
<td>Blood pool membership threshold</td>
<td>0.20</td>
</tr>
<tr>
<td>Myocardium membership threshold</td>
<td>0.05</td>
</tr>
<tr>
<td>Air membership threshold</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ASM [90, 91]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of modes of variation</td>
<td>60</td>
</tr>
<tr>
<td>Allowed variation per mode</td>
<td>( 2\sigma )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Propagation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian kernel width ( \sigma_p ) (Eq. (7.8))</td>
<td>RAD 4 planes 8 mm</td>
</tr>
<tr>
<td></td>
<td>MV 4 planes 6 mm</td>
</tr>
<tr>
<td></td>
<td>SA 11 planes 6 mm</td>
</tr>
<tr>
<td></td>
<td>SA-sub 4 planes 6 mm</td>
</tr>
</tbody>
</table>

Interlineation for the radial slices, contours on all 15 RAD data sets were manually drawn by two observers. The manual segmentation of one of the two observers was regarded as the gold standard. All error measures were calculated with respect to the gold standard thus defined. To quantify interobserver variability and the performance of the SPASM on the sparse image data, point-to-surface (P2S) error measurements were performed. Point-to-surface distances were measured from all node locations (points) on the segmentation results of SPASM to the closest locations on the surfaces of the manual segmentation (see Fig. 7.7(a)). Per subject, a mean and a maximum error are computed. Volumes were calculated by closing the model at the base (see Fig. 7.7(b)) and the manual shapes at both the apex and the base (see Fig. 7.7(c)). Inter-observer variability results are presented together with the final P2S segmentation errors achieved by other groups with their model-based cardiac segmentation techniques (see Tab. 7.2).

For the evaluation of SPASM, distances were measured from node locations (points) on the segmentation results of SPASM to the surface of the manual segmentation.

To test whether differences in the P2S errors were significantly different, t-tests were performed. Observations on the SA, SA-sub and RAD data sets and the interobserver variability were obtained for the same subjects, therefore these results were compared using a paired samples t-test. Comparison of the results achieved on the MV data set to those from any other data set was performed with an independent samples paired t-test, because the majority of the subjects were present in only one of the data sets. SPSS version 11.0 (SPSS, Chicago, IL) was used as statistical software.

**Breath hold correction**

Image data sets are acquired during breath hold in end expiration. Thus, displacement of the cardiac structures between acquisition of different slices is minimized. However, small differences in expiration level can still exist, leading to shifted tho-
7.3 Experimental setup

Table 7.2: Point-to-surface distances from automatic results for other 3D automatic cardiac segmentation methods (all values are $\mu \pm \sigma$ in mm).

<table>
<thead>
<tr>
<th></th>
<th>endocardium</th>
<th>epicardium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lötjönen et al. [32]$^1$</td>
<td>$2.01 \pm 0.31$</td>
<td>$2.77 \pm 0.49$</td>
</tr>
<tr>
<td>Mitchell et al. [31]$^2$</td>
<td>$2.75 \pm 0.86$</td>
<td>$2.63 \pm 0.76$</td>
</tr>
<tr>
<td>Kaus et al. [26]</td>
<td>$2.28 \pm 0.93$</td>
<td>$2.62 \pm 0.75$</td>
</tr>
<tr>
<td>Lorenzo-Valdés et al. [86]$^3$</td>
<td>$1.88 \pm 2.00$</td>
<td>$2.75 \pm 2.62$</td>
</tr>
</tbody>
</table>

racic and cardiac structures in the images.

To correct for possible misalignment of SA image slices due to acquisition in separate breath holds (and consequently difference in inspiration level), in case of matching on the RAD or the SA stacks, manual contours from the radial TFE stack were used. For each SA slice a displacement was calculated by alignment of the center of the manual endocardial contour to the center of the eight intersection points of this SA slice with the four manual endocardial RAD contours.

For the data used for matching to the MV combination of image slices, no RAD images were present. This conforms to the clinical practice, as these are not routinely acquired. Since the 2- and 4-chamber view intersection almost coincides with the long-axis of the LV, the centers of gravity of the manual SA LV endocardial contours were first translated in plane to this intersection line. Next, as a refinement step, the centers of gravity were translated along the lines of intersection between the 2-chamber plane and SA planes, causing the profile of short-axis manual delineations to best match the profile of the manual delineation of the LV endocardium in the two-chamber view.

Matching initialization

Initialization of the model in the target data set was performed as follows. Initial pose and scale were calculated from manual delineations on two image slices from the SA acquisitions, the most basal and most apical ones. Due to the rotational symmetry of the model with respect to the long-axis and because of the sectorization, special attention was paid to initialize the model such that the myocardial sectors corresponded to the approximately correct anatomical locations in the image data. For this, manual segmentations of the right ventricle in the same slices were used.

Parameter settings for the membership thresholds for the FIS used to define model updates at locations where the model is intersected by image planes were taken from

---

$^1$These are the best obtained results by Lötjönen et al. with a 4-chamber ASM based on a probabilistic atlas. Other models were built using normalized mutual information, landmark probability distribution, PCA, and ICA.

$^2$Mitchell et al. compute errors of the automatic segmentation results slightly different than in this work. Mitchell et al compute (2D) distances in the image slices along lines perpendicular to the centerline between automatic and manual segmented contours. This does not guarantee shortest point-to-curve or point-to-surface distances, and may thus overestimate errors with respect to the method used in this paper and by Lötjönen et al.

$^3$Lorenzo-Valdés et al. , like Mitchell et al. , compute errors in a 2D in-plane manner for three mid-ventricular slices. These are the best obtained results by Lorenzo-Valdés et al.
previous work [90]. Best settings for the update propagation parameter ($\sigma_p$) and the maximally allowed variability per deformation mode were determined in the evaluation data sets in an exhaustive search in the parameter space using a grid computing middleware called InnerGrid (GridSystems SA, Palma de Mallorca, Spain) over 50 Intel Xeon processors, and taking P2S error measures of the final state of the model with respect to manual segmentation as criterion for evaluation. The optimal settings for the parameters are listed in Table 7.1. For the different combinations of image slice orientations, different settings for the propagation kernel width ($\sigma_p$) were found to give the best results.

7.3.2 Matching experiments

Matching experiments were designed to show that SPASM is able to segment different combinations of image data with arbitrary orientations, and to compare the performance of SPASM to that of other methods found in literature. The following image slice combinations were used for quantitative evaluation:

**Radial stack (RAD)** SPASM was fitted to four images acquired in a radial orientation with the TFE protocol (see Fig. 7.8(a)).

**Multi-view combination (MV)** SPASM was fitted on a combination of 2 SA image slices (one apical and one mid-ventricular slice) combined with a 2-chamber- and a 4-chamber view (see Fig. 7.8(b)).

**Short-axis stack (SA)** SPASM was fitted to a full short-axis stack (11 slices, see Fig. 7.8(c)).

**Short-axis stack sub set (SA-sub)** To make a fair comparison between the matching performance on the RAD and the MV data sets (with four planes) and on the SA data set, all 330 combinations of 4 planes out of the full set of 11 were tested (see Fig. 7.8(d)).

To assess the sensitivity to initial position, additional experiments were carried out while varying the $x$, $y$, and $z$-coordinates of the model’s initial position. Offsets were added to $x$- and $y$-coordinates ranging between -20 mm and +20 mm, and with steps of 5 mm. For the $z$-coordinates, offsets ranged between -10 mm and +10 mm with a step of 10 mm. Because this experiment was performed for the radial data set (15 subjects), the multi-view data set (20 subjects), and the short-axis data set (15 subjects, and using the full stack) in total this experiment entailed 12,150 runs of SPASM. This sensitivity to the initial placement was tested in an exhaustive search on a computer cluster with 85 processors plus a grid-computing middleware called InnerGrid (GridSystems SA, Palma de Mallorca, Spain). The test variable was the resulting P2S error measure, both for the epicardial surface and for the endocardial surface.

7.4 Results

The optimal parameters with respect to the defuzzification, model deformation and update propagation are shown in Table 7.1. Optimal defuzzification and model deformation parameters were taken from [90]. The propagation update parameters were
Figure 7.7: (a) Point-to-surface distances are measured from points on the model to the closest points on the manual surface. (b) Closed model for volume estimation (c) Closed manual shape for volume estimation.

Figure 7.8: (a) RAD view with 4 planes. (b) Multiple view with 2 SA views and a 2-chamber and a 4-chamber view. (c) All 11 SA planes. (d) Combination of 4 SA planes. This combination of 4 planes gave the best overall segmentation result of all 330 possible combinations of 4 SA planes out of 11.
7.4 Results

(a) Best result (NT22771), average P2S: endocard 1.62 mm, epicard 1.54 mm

(b) Typical result (NT22821), average P2S: endocard 1.95 mm, epicard 2.27 mm

(c) Worst result (NT22897, average P2S: endocard 3.20 mm, epicard 3.73 mm

Figure 7.9: Final segmentation result of some of the subjects in the test population. (a-c) Results shown on slice 1 through 4 of a RAD TFE stack, respectively.

Results from the P2S evaluations between the final model instances (on the RAD data set, the MV data set and the data sets constructed by the different SA plane combinations) and manual SA segmentations are presented in Table 7.3. The results for the SA-sub data set only reflect the combination of four planes that gave the best segmentation results. The optimal configuration of 4 SA planes is shown in Figure 7.8(d). T-tests were performed on the point-to-surface errors comparing all data combinations to the interobserver variability and to the results achieved with SA-full. Endocardial and epicardial volumes were calculated and compared for all combinations to those achieved with SA-full. Comparison to manual volumes was not performed, because the manual volumes did not contain the apex, while the automatic volumes did. Results of the t-tests are presented in Table 7.4. Based on the results of the P2S distance measurements, one subject was excluded for the RAD test due to huge error values for all the runs in the tuning process. For the MV test, one subject was excluded due to a mismatch of the model resulting from spurious edges confusing the feature point detection. For the SA data sets, for all 15 subjects good results were obtained.

Final segmentation results for three of the subjects from the RAD data set are shown on all four slices in Figure 7.9. These are the best, an intermediate, and the worst results. In Figure 7.10 distributions of the P2S errors over the model surfaces are shown.
Table 7.3: Point-to-surface distances from points on the automatically segmented surfaces to the manually segmented surfaces, measured per subject, and averaged over the total population (14 subjects for RAD, 15 for SA, 19 subjects for MV), and volumes averaged over the populations. All values are $\mu \pm \sigma$, P2S errors in mm and volumes in ml; endo means endocardium, and epi means epicardium.

<table>
<thead>
<tr>
<th></th>
<th>Average P2S errors (mm)</th>
<th>Average volumes (ml)</th>
<th>planes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>endo</td>
<td>epi</td>
<td>endo</td>
</tr>
<tr>
<td>SA Manual</td>
<td>1.27 ± 0.30</td>
<td>1.14 ± 0.29</td>
<td>112.2 ± 26.1</td>
</tr>
<tr>
<td>RAD</td>
<td>2.24 ± 0.54</td>
<td>2.83 ± 0.78</td>
<td>132.5 ± 18.5</td>
</tr>
<tr>
<td>SA</td>
<td>1.97 ± 0.54</td>
<td>2.23 ± 0.46</td>
<td>122.0 ± 27.3</td>
</tr>
<tr>
<td>SA-sub</td>
<td>2.14 ± 0.58</td>
<td>2.46 ± 0.50</td>
<td>125.2 ± 26.3</td>
</tr>
<tr>
<td>MV Manual</td>
<td>131.9 ± 31.5</td>
<td>227.1 ± 52.1</td>
<td>127.5 ± 28.1</td>
</tr>
<tr>
<td>MV</td>
<td>2.02 ± 0.93</td>
<td>2.29 ± 0.53</td>
<td>137.5 ± 24.1</td>
</tr>
</tbody>
</table>

Table 7.4: Point-to-surface errors from all data sets compared to the interobserver variability and to automatic results on SA. Volumes compared to to automatic volumes on SA. Numbers are $p$-values. Asterisks mark statistical significance at the 0.05 confidence level.

<table>
<thead>
<tr>
<th></th>
<th>RAD</th>
<th>MV</th>
<th>SA-sub</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>point-to-surface errors compared to interobserver variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endocardium</td>
<td>.000*</td>
<td>.003*</td>
<td>.000*</td>
<td>.000*</td>
</tr>
<tr>
<td>epicardium</td>
<td>.000*</td>
<td>.000*</td>
<td>.000*</td>
<td>.000*</td>
</tr>
<tr>
<td>point-to-surface errors compared to SA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endocardium</td>
<td>.078</td>
<td>.861</td>
<td>.005*</td>
<td></td>
</tr>
<tr>
<td>epicardium</td>
<td>.027*</td>
<td>.757</td>
<td>.002*</td>
<td></td>
</tr>
<tr>
<td>volumes compared to SA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endocardium</td>
<td>.065</td>
<td>.665</td>
<td>.020*</td>
<td></td>
</tr>
<tr>
<td>epicardium</td>
<td>.000*</td>
<td>.566</td>
<td>.011*</td>
<td></td>
</tr>
</tbody>
</table>

Color-coded for the best, worst and typical segmentation results on the RAD, MV, SA, and SA-sub data configurations. For a comparison of performance on MV to both the interobserver variability and to the results achieved with SA independent samples t-tests were used. For all comparisons with respect to the other data configurations paired t-tests were used. The results from these comparisons can be found in Table 7.4. Results for the sensitivity of the SPASM to initial model placement are given in Figure 7.11 for the radial data set, in Figure 7.12 for the multi-view data set, and in Figure 7.13 for the short-axis data set. In these figures, columns indicate either epicardial results or endocardial results, rows indicate different $z-$displacements. Within charts, different data series represent different $y-$displacements, while the offsets in $de x-$direction are plotted on the horizontal axes.

Computation times for SPASM very much depend on the width of the propagation kernel and on the number of images slices used. Approximate computation times ranged from 40 seconds ($\sigma_p=3$ mm) to five minutes ($\sigma_p=10$ mm) both on the SA-full data set.
7.5 Discussion

In this paper, a new 3D active shape model, coined SPASM, is presented that is able to automatically segment cardiac MRI image data sets consisting of multiple planes with different orientations, and having large undersampled regions. This is achieved by the incorporation of an update propagation scheme, that causes image updates to be propagated over the surface. The method requires that breathing induced shifts between slices are corrected in a preprocessing step, and that the initialization of the model in the data set is within a certain range of the LV, which is discussed in Section 7.5.2. Both slice shift correction and initialization can be done automatically very well, but since those were not the topics of this work, we performed them (partly) manually.

7.5.1 Segmentation performance

From the segmentation performance results on the different data sets used in this chapter (see Tabs. 7.3 and 7.4) it can be seen that:

- on the SA data set (11 planes) SPASM works significantly better than on the SA-sub data set (four planes) for both the endocardium and the epicardium,
- for the epicardium, segmentation performance on the SA and the MV data set is significantly better than on the RAD data set. However, for the endocardium there is no significant difference,
- for all other data configurations, performances are comparable.
- segmentation performances on the SA data set and the MV data set in particular, are almost the same (epicardium: $2.23 \pm 0.46$ mm and $2.29 \pm 0.53$ mm respectively, endocardium: $1.97 \pm 0.54$ mm and $2.02 \pm 0.93$ mm respectively). Differences between these two techniques are not statistically significant (epicardium: $p=0.757$, endocardium: $p=0.861$).

Segmentation performance on the MV, SA, and SA-sub data sets was slightly better than on the RAD data (see Tab. 7.3). This may be caused by the different acquisition protocol used for the RAD data set. This data set was acquired with the TFE protocol which yields images of less quality due to the limited acquisition time. With the TFE protocol, four slices were acquired in one breath hold, i.e., 25-30 seconds. The images in the SA data set (acquired with the BFFE protocol) were acquired in 15-20 seconds per slice (full stack approximately 220 seconds, i.e., 11 times 20 seconds plus the time between slice acquisitions).

Performance on the MV data set (BFFE) was slightly better than on the SA-sub data set, possibly because of a better definition of the apex. In the SA (and SA-sub) data set the apex was fuzzy, because of partial-volume effects. In the selection of image slices for the SA-sub data set, the best combination of four planes was defined by two apical slices, one mid-ventricular slice and one basal slice. Apex definition in LA views is better than in SA views [4]. Possibly, by using two apical slices in the SA-sub data set, this feature is compensated for.

From the t-tests (see Table 7.4), it appears that the accuracy with respect to surface localization achieved on the MV data configuration (4 planes) does not differ significantly from that achieved when SPASM was applied to SA (11 planes). However, P2S errors for all data configurations were significantly larger than the interobserver variability from the full SA data set. This may be caused by the inclusion of the apex in the
Figure 7.10: P2S errors encoded in color for best, typical and worst segmentation results.
Figure 7.11: Sensitivity to initial displacements. RAD data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the radial data set. Different curves show different displacements in the AP-direction ($y$–direction). Plots on different rows show different displacements in the direction of the long-axis ($z$–direction), (a,b) $z=-10$ mm (c,d) $z=0$mm (e,f) $z=10$ mm. The displacements in the LV-RV direction (in mm) are plotted along the $x$–axes of the plots.
model, whereas this was not included in the manual segmentations. As a result, large errors are observed near the apical regions of the model (see Figs. 7.7(a) and 7.10). Volumes could not be compared to manual volumes directly, because the model included the apex, which the manual shapes did not. This leads to systematic volume differences. From the volumetric t-tests (see Table 7.4), volumes were not significantly different to SA-full from MV and for the endocardium also from RAD. Thus, a limited set of differently oriented images in combination with the model-based approach provides an accurate and efficient way of reconstructing the endocardial and epicardial LV surfaces and yields accurate quantification of LV volumes.

Although there are some significant differences between the performances on different data sets, these differences are so small that they are not clinically relevant. Global functional parameters like stroke volume and ejection fraction are calculated from the endocardial segmentation. For the endocardium, all differences in average P2S errors are less than 0.28 mm, and volume differences between SA and SA-sub, which are significantly different, are on the order of 3ml.

### 7.5.2 Sensitivity to initial model placement

Observed sensitivities to initial model placement show that initialization of the model is important, but within a range of 15-20 mm in all directions effects in final segmentations are minimal (see Figs. 7.11-7.13). This can be concluded from the following observations:

- For the RAD data set, flat curves (equivalent performance) can be observed for over 15-30 mm in the $x-$direction (axis in LV-RV direction) within 20 mm displacements in the $z-$direction (direction of the long-axis). Displacements in the $y-$direction (AP-direction) of 15-20 mm do not have a big influence on final segmentation performance (see Fig. 7.11).

- For the MV data set, the global pattern is comparable to the radial data set sensitivities. In all directions, SPASM is slightly more sensitive to initial placement in this data set than in the RAD data set.

- For the SA data set, sensitivity in the $z-$direction (long-axis direction) is a little worse than in the RAD data set. Sensitivity with respect to the $x-$ and the $y-$directions (i.e., parallel to the plane orientations) is smaller.

These results are in agreement with expectations based on the composition of the data sets: small sensitivity in $x-$ and $y-$directions when the images are parallel to the $xy-$plane (SA data set), and the best resolution is in the $x-$ and $y-$directions. Small sensitivity with respect to the long-axis direction is found in the RAD data set, the data set in which the best resolutions are in the long-axis direction. In one other direction (perpendicular to the long-axis), which changes from plane to plane, there is a high resolution too. This may lead to smaller overall sensitivity to initial placement in the RAD data set, however, the in-plane resolution is worse for the RAD data set than for the other two data sets, because of the TFE protocol with which the images in this data set are acquired. The SA data set used in these sensitivity experiments is a full stack, therefore, a smaller sensitivity to initial model placement can be expected. In all plots in Figs. 7.11, 7.12 and 7.13 asymmetry in the sensitivity with respect to initial displacement in $x-$direction (pointing from the RV to the LV) is apparent. Larger sensitivity is observed when the model is displaced towards the RV than when it is
displaced towards the lung. Possible cause is the lack of air pixel samples in the population to be classified, leading to an unbalanced partitioning and misclassification. When shifted towards the LV, only proportions of the tissues present in the population deviate, but tissues do not disappear from the population.

For the use in a clinical environment the tools for automated cardiac MRI planning \[92, 93, 94\] can be utilized for initialization purposes. Those tools permit determination of the LV axis spatial orientation with an accuracy of 7 degrees \[93, 94\]. Danilouchkine et al. \[94\] observed that with this accuracy, the appearance of the heart is not different from manual scan planning. However, accurate automatic localization of the heart position remains a problem \[92, 93\] and often solved by scanning more slices then necessary to guarantee complete coverage of the LV. Hence, we only investigated initialization changes based on translations in the \(x-, y-\) and \(z-\)directions, assuming that the spatial orientation can be determined with an accuracy sufficient for automated SPASM initialization.

### 7.5.3 Protocol independence

In previous work \[74, 90, 91\], we demonstrated that the 3D-ASM was able to segment CT-data and MRI-data without any adaptations to the statistical component of the method. This was achieved by the implementation of the Fuzzy Inference System instead of a statistical Gray Level Model.

In this work, the results demonstrated that application to different MRI acquisition protocols without any adaptation to the SPASM yields good results. We applied SPASM to image data acquired with both the TFE and the BFFE protocol, without retraining the model. Differences in the value for the propagation kernel \(\sigma\) (see Tab. 7.1) between the RAD data set and the other data sets were solely caused by differences in the sampling density of the data sets.

### 7.5.4 Limitations

In the SA data set, the heart can be displaced between different slice acquisitions, due to different breath hold levels. Although this displacement is minimized by acquisition during end expiration, correction of slice positions was necessary for all but the RAD data set. The choice of which correction method should be applied, if at all, is a topic for debate by itself \[33, 95, 96, 97, 98\].

Segmentation errors of all methods are substantially larger than the interobserver variability (see Tabs. 7.2, 7.3). This may be caused by too rigid statistical constraints on the allowed deformation of statistical shape models in general. A relaxation of the shape constraints in the final phase of the matching may further improve results.

### 7.5.5 Comparison to other work

Performance of SPASM has been compared to results from other methods reported in literature \[26, 31, 32\] (see Tab. 7.2). In all data set configurations used, the accuracy of SPASM is comparable to these other methods. The precision of SPASM with respect to the different data sets is comparable to that achieved by Lötjönen et al. \[32\] and Lorenzo-Valdés et al. \[86\], and comparable to or better than the precision achieved by the methods of Mitchell et al. \[31\] and Kaus et al. \[26\]. However, in the three sparse data configurations, SPASM was applied to four image slices, whereas the other models used a stack of 7-10 SA slices \[26, 31, 86\] and a combination of two stacks (4-5 SA plus 4-7 LA slices) \[32\]. It has to be noted, however, that we acquired new data sets,
Figure 7.12: Sensitivity to initial displacements. MV data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the multi-view data set. Different curves show different displacements in the AP-direction ($y$-direction). Plots on different rows show different displacements in the direction of the long-axis ($z$-direction), (a,b) $z=-10$ mm (c,d) $z=0$ mm (e,f) $z=10$ mm. The displacements in the LV-RV direction (in mm) are plotted along the $x$-axes of the plots.
Figure 7.13: Sensitivity to initial displacements. SA-full data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the short-axis data set. Different curves show different displacements in the AP-direction ($y$-direction). Plots on different rows show different displacements in the direction of the long-axis ($z$-direction), (a,b) $z=-10$ mm (c,d) $z=0$ mm (e,f) $z=10$ mm. The displacements in the LV-RV direction (in mm) are plotted along the $x$-axes of the plots.
so we did not use exactly the same data sets that were used in [26, 31, 32, 86]. Although the other models were not applied to data with different orientation than the training data, they may be applied to different orientations and sparsity as well. However, this poses considerable challenges. First, shape-based interpolation [99, 100] is ill-posed for radial LA data sets like RAD. Second, it may be difficult to estimate the required intensity distributions [86] from sparse data, because sparse configurations may not contain all structures defined in the atlas. Last, robust registration of a 3D intensity template to a sparse data set is not trivial.

Like the approaches in [87, 82], our approach is based on fitting a statistical model of the anatomical organ to its sparse representation. However, unlike the other approaches, SPASM does not require correspondence computation during model matching, thanks to the propagation of the derived update information over the model surfaces.

Furthermore, SPASM is the only method that can be applied to a set of arbitrarily oriented (including radially) and very sparsely sampled data sets. In our data, SA-sub has a gap of 30 mm between slices 2 and 5, and completely lacks information from the four most basal slices (i.e. 40 mm), the RAD-2a and RAD-2b data sets have even larger void regions.

Due to the fact that SPASM includes the complete apex, and the manual shapes do not, distances observed at the apex are "artificially" large. This can be seen in Figure 7.7(a), and Figure 7.10 shows that in many cases errors close to the apex are larger than at other locations. Consequently, the errors reported in Table 7.3 overestimate the true segmentation errors, i.e., regarding the locations where both automatic and manual surface are present.

### 7.6 Conclusions

The presented SPASM proved to be a powerful tool for the automatic segmentation of sparse cardiac data sets with multiple image orientations. The incorporation of an update propagation scheme and a Fuzzy Inference System enabled application of SPASM to multi-protocol cardiac sparse data sets with a segmentation performance that is better than or comparable to other 3D model-based segmentation methods operating on a full data set with parallel image planes. SPASM does not require predefined image orientations and image slices with equal orientations as present in the training data. Because SPASM does not include a statistical gray level model, it is applicable to data sets from different MRI acquisition protocols without adaptations to the statistical framework or the update determination scheme (FIS).

Performance on the multi-view data set was comparable to performance on the SA data set (4 slices vs. 11 slices respectively). This means that possibly not as many as 11 slices are required for accurate cardiac LV segmentation results, provided that a combination of different image orientations is used, and a dense statistical shape model and update information propagation are employed.

Sensitivity to initial model placement proved to be acceptable within a range of approximately 20 mm. Further extension of this capture range and automatic initial model placement are topics of further research. In addition, further evaluation of SPASM with respect to the minimally required cardiac data sampling density and different combinations of LA and SA image slice orientations is ongoing.