ELSEVIER

SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data

Hans C. van Assen [a,1], Mikhail G. Danilouchkine [a,2], Alejandro F. Frangi [b],
Sebastián Ordás [b], Jos J.M. Westenberg [a], Johan H.C. Reiber [a,*], Boudewijn P.F. Lelieveldt [a]

[a] *Division of Image Processing, Department of Radiology, Leiden University Medical Center, Albinusdreef 2,
PO Box 9600, 2300 RC, Leiden, The Netherlands*
[b] *Computational Imaging Laboratory, Department of Technology, Universitat Pompeu Fabra, Barcelona, Spain*

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. Model landmark positions are updated in a two-stage iterative process. First, landmark positions close to intersections with images are updated. Second, the update information is propagated to the regions without image information, such that new locations for the whole set of the model landmarks are obtained. Feature point detection is performed by a fuzzy inference system, based on fuzzy C-means clustering. Model parameters were optimized on a computer cluster and the computational load distributed by grid computing. SPASM was applied to image data sets with an increasing sparsity (from 2 to 11 slices) comprising images with different orientations and stemming from different MRI acquisition protocols.

Segmentation outcomes and calculated volumes were compared to manual segmentation on a dense short-axis data configuration in a 3D manner. For all data configurations, (sub-)pixel accuracy was achieved. Performance differences between data configurations were significantly different ($p < 0.05$) for SA data sets with less than 6 slices, but not clinically relevant (volume differences < 4 ml). Comparison to results from other 3D model-based methods showed that SPASM performs comparable to or better than these other methods, but SPASM uses considerably less image data. Sensitivity to initial model placement proved to be limited within a range of position perturbations of approximately 20 mm in all directions.
© 2005 Elsevier B.V. All rights reserved.

Keywords: SPASM; Active shape model; ASM; 3D; Sparse; MRI; CT; Segmentation; Cardiac; Short-axis; SA; Long-axis; LA; Radial

1. Introduction

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 (van Rossum et al., 1987), and long-axis (LA) views (2-chamber and 4-chamber).

* Corresponding author. Tel.: +31715261117; fax: +31715266801.
E-mail addresses: h.c.v.assen@tue.nl (H.C. van Assen), m.danilouchkine@erasmusmc.nl (M.G. Danilouchkine), alejandro.frangi@upf.edu (A.F. Frangi), sebastian.ordas@upf.edu (S. Ordás), j.j.m.westenberg@lumc.nl (J.J.M. Westenberg), j.h.c.reiber@lumc.nl (J.H.C. Reiber), b.p.f.lelieveldt@lumc.nl (B.P.F. Lelieveldt).
[1] Present address: Biomedical Image Analysis, Biomedical Engineering, Technical University Eindhoven, Eindhoven, The Netherlands.
[2] Present address: Department of Cardiology, Erasmus Medical Center, Rotterdam, The Netherlands.

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 8–12 (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 (Bloomer et al., 2001). 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.

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 (Cootes et al., 1995, 2001) on 2D active shape models (ASMs) and active appearance models (AAMs) has inspired the development of 3D ASMs (van Assen et al., 2003; Kaus et al., 2004), and 3D AAMs (Mitchell et al., 2002). 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 (Kelemen et al., 1999; van't Ent et al., 2001), 3D statistical deformation models (SDMs) (Rueckert et al., 2003; Frangi et al., 2002; Chandrashekara et al., 2003; Lötjönen et al., 2004a), and 3D medial representations (m-reps) (Yushkevich et al., 2003) have been described. These mainly differ from each other in the way the statistical shape variations are parameterized. Lorenzo-Valdés et al. (2004) 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 (Mitchell et al., 2002; Rueckert et al., 2003; Frangi et al., 2002; Chandrashekara et al., 2003; Lötjönen et al., 2004a; Lorenzo-Valdés et al., 2004) or on a dense set of updates along the model surface (van Assen et al., 2003; Kaus et al., 2004; Yushkevich et al., 2003).

Many references to registration techniques used with dense models and sparse data sets can be found in (Koikkalainen and Lötjönen, 2004). 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. (2001), 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 (Fleute et al., 1999), 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 and Lötjönen (2004) and Fleute et al. (1999) 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.

Montagnat et al. (2003) reported matching a model to 4D echocardiographic imagery of a limited test set, based on zero-order statistics (mean shape). The model deformation was governed by internal regularization and external forces (like active surface models). Trilinear interpolation was applied to reconstruct the volumetric LV texture from radially acquired data samples. The update information was generated for each vertex of the model mesh by scanning the reconstructed volumetric LV texture normal to the mesh at the given vertex. The method was assessed on a 9-slice radial stack of in vivo US images. In this respect the experimental data can be regarded as dense, allowing reliable texture reconstruction. However, the accuracy of the proposed method was not assessed, and no application was reported to truly sparse 2- or 4-slice radial stacks, frequently used in cardiac MRI imaging.

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.
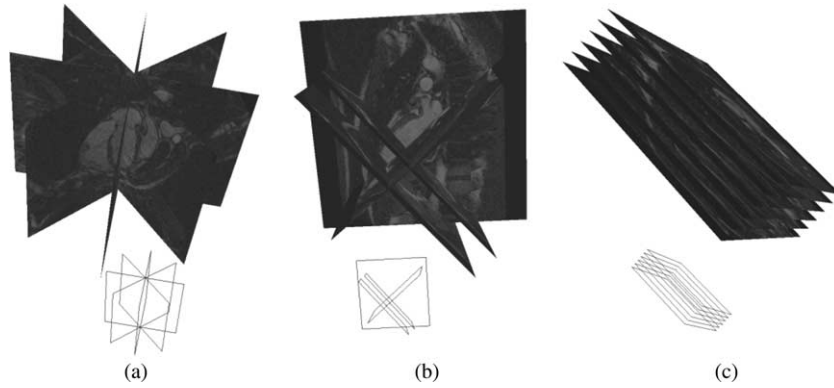
Fig. 1. (a) Radially oriented cardiac image stack, (b) a combined long-axis and short-axis data set, and (c) a short-axis image stack (only showing every second image).

The main contribution of this work is the development of a 3D-ASM that:

- is applicable to sparsely sampled data sets without making assumptions about voxel isotropy or parallel slices (see Fig. 1),
- enables volumetric quantification from sparse data through statistically driven shape interpolation instead of through simplified assumptions about ventricular geometry,
- is extensible to other modalities/protocols without retraining the shape model (van Assen et al., 2003). This increases flexibility with respect to new MR scanning protocols, which we regard as essential for application of the method in clinical practice.

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 (Frangi et al., 2002; Ordas et al., 2004). 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) (Takagi and Sugeno, 1985) for determining iterative model updates based on *relative* intensity differences (van Assen et al., 2003; Ordas et al., 2005b).

In addition, an important contribution of this work is the extensive evaluation of how the accuracy of cardiac surface reconstruction is influenced by the sparsity of the data.

2. Methods

Active shape models were introduced by Cootes et al. (1992, 1995) 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 popula-

tion 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.

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 pseudo-landmark points that build a shape vector

$$\mathbf{x} = (x_1, y_1, z_1, \ldots, x_n, y_n, z_n)^{\mathrm{T}}. \tag{1}$$

(Pseudo-)landmarks can also be determined in an automatic manner (Frangi et al., 2002).

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

$$\bar{\mathbf{x}} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{x}_i \tag{2}$$

and a covariance matrix

$$\mathbf{S} = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}} \tag{3}$$

are calculated. From the covariance matrix $\mathbf{S}$, 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 ϕ_i corresponding to the t largest eigenvalues are retained, represented by matrix $\mathbf{\Phi} = (\phi_1 | \phi_2 | \cdots | \phi_t)$. From this matrix and the mean shape, a shape $\mathbf{x}$ can be generated

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi b}, \tag{4}$$

with **b** a *t*-dimensional vector containing the model parameters

$$\mathbf{b} = \mathbf{\Phi}^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}}). \tag{5}$$

By limiting the components of **b** within statistically derived bounds (e.g., $\pm 3\sigma$) the resulting shape from Eq. (4) is expected to resemble the shapes seen in the training set. For an elaborate introduction to ASMs, the reader is referred to (Cootes et al., 1992, 1995; Cootes and Taylor, 2004).

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 (Frangi et al., 2002). The general layout of the method is to align all the images of the training set to a mean atlas (Fig. 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 onto the atlas surface to the coordinate system of each training shape. While it is still necessary to manually segment each training image, this technique reliefs 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. To build the statistical shape model, the auto-landmarked shapes are aligned using Procrustes alignment (Gower, 1975). PCA is performed on the remaining differences between equivalent landmarks on different shape samples. These differences are solely shape related.

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 (ICP) algorithm (Besl and McKay, 1992). Successively, the parameter vector **b** controlling model deformation is calculated. An adjustment to **b** with
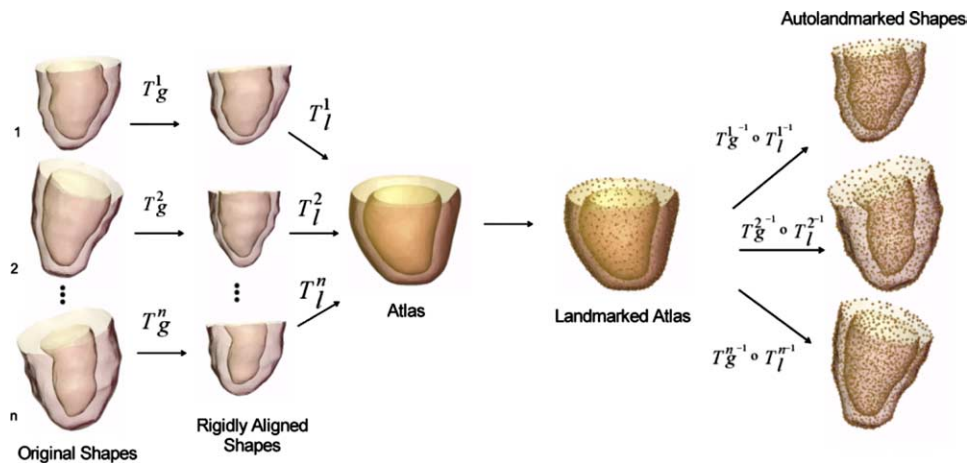

Fig. 2. Atlas construction: a set of final global (T_g) and local (T_l) transformations can take any sample shape of the training set to the atlas coordinate system (left). Once the final global and local transformations are obtained, they are inverted and used to propagate any number of arbitrarily sampled landmarks on the atlas, to the coordinate system of the original samples (right).

respect to the previous iteration is computed, using both the candidate model state, $\hat{\mathbf{x}}_{n+1}$, and the current model state, $\mathbf{x}_n$

$$\hat{\mathbf{b}}_{n+1} = \mathbf{b}_n + \Delta\mathbf{b} = \mathbf{b}_n + \mathbf{\Phi}^{\mathrm{T}}(\hat{\mathbf{x}}_{n+1} - \mathbf{x}_n) \qquad (6)$$

with $\mathbf{x}_n$ representing the mesh after alignment by ICP to the proposed model state, and $\mathbf{b}_n$ representing the parameter vector describing the current shape of the model within the statistical bounds. The vector $\hat{\mathbf{x}}_{n+1}$ is the proposed model shape for the next iteration, and $\hat{\mathbf{b}}_{n+1}$ 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 (see Fig. 3). 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.

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. 4(a)). Such node updates are defined as (for node ω)

$$\mathbf{v}_{\omega,n+1} = \hat{\mathbf{x}}_{\omega,n+1} - \mathbf{x}_{\omega,n}, \qquad (7)$$

where $\hat{\mathbf{x}}_{\omega,n+1}$ is the estimated position of node ω for iteration $n+1$ determined by the model update scheme (see Section 2.5) (i.e., before propagation, and after projection on the surface normal), and $\mathbf{x}_{\omega,n}$ is the position of node ω for iteration n.
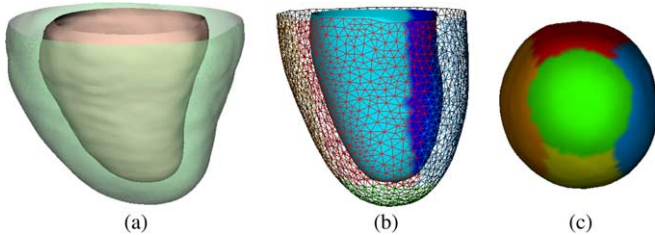

Fig. 3. (a) Shape of the 1-chamber atlas constructed from MRI data. (b) Mean shape of the 3D-SPASM constructed from the atlas. Different colors indicate different sectors of the mean shape, that are assigned their own FCM operation within the model update scheme (see Section 2.5). The *epicardial* surface is represented with a wireframe and shows three sectors (two sectors are hidden at the backside); the *endocardial* surface is represented by a surface and has two sectors. (c) Apex view of the mean shape (*epicardium*) of the 3D-SPASM showing all five epicardial sectors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
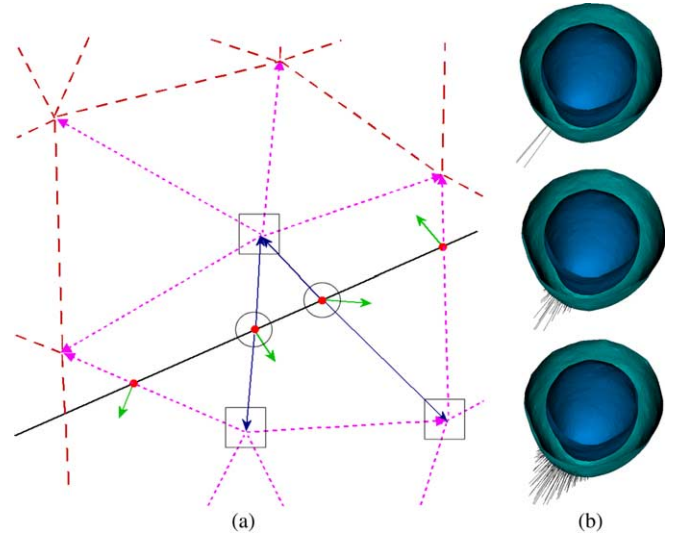

Fig. 4. (a) Propagation of single model updates. Update sources are indicated as large dots. Propagation from the update sources surrounded with a large circle is illustrated here. From a source, an update vector originates (short arrow). Updates are first propagated (longer solid arrows) to the nearest 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, $\sigma_p = 4$ mm, and at the bottom $\sigma_p = 8$ mm.

During propagation, such source update vectors $\mathbf{v}_\omega$ are weighted in the receiving nodes λ according to the geodesic distance between source node and receiving node using a Gaussian-shaped kernel the width of which is denoted by σ_p (see Fig. 4(b)). To avoid propagation of an update over the entire model surface, propagation is stopped when the geodesic distance exceeds a fixed threshold χ ($\chi \equiv 3\sigma_p$). A Gaussian kernel has unity area under the curve, which is achieved by a normalization factor $(2\pi\sigma_p)$. 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:

$$w(\lambda,\omega) = \begin{cases} e^{-\frac{\|\lambda-\omega\|^2}{2\sigma_p^2}} & \text{if } \|\lambda-\omega\| \leqslant \chi, \\ 0 & \text{if } \|\lambda-\omega\| > \chi, \end{cases} \qquad (8)$$

where $w(\lambda,\omega)$ is the weight at the location of the receiving node λ, ω is the source node, $\|\lambda-\omega\|$ is the geodesic distance to the origin of the update, and σ_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 stop, a cumulative contribution from all source nodes ω_m to each receiving node λ is

computed. Each receiving node has a list of weighted contributions from source nodes

$$\mathbf{v}_{\omega_m} \cdot w(\lambda, \omega_m) \quad \forall m \in [1, M], \tag{9}$$

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 contributing source nodes K. The total number of contributions from source nodes to a receiving node λ is defined as

$$K_\lambda = \sum_{m=1}^{M} k_m, \quad \text{where } k_m = \begin{cases} 1 & \text{if } \|\lambda - m\| \leqslant \chi, \\ 0 & \text{if } \|\lambda - m\| > \chi. \end{cases} \tag{10}$$

The total update to node λ is thus defined as

$$\boldsymbol{v}(\lambda) = \frac{1}{K_\lambda} \sum_{m=1}^{M} \mathbf{v}_{\omega_m} \cdot w(\lambda, \omega_m). \tag{11}$$

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

$$\boldsymbol{v} = (\boldsymbol{v}(1) \mid \cdots \mid \boldsymbol{v}(L)) \tag{12}$$

with L the total number of nodes in the mesh. So, using Eq. (6), and acknowledging that

$$\boldsymbol{v}_{n+1} = \hat{\mathbf{x}}_{n+1} - \mathbf{x}_n \tag{13}$$

finally yields

$$\Delta \mathbf{b} = \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{v}. \tag{14}$$

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) (Takagi and Sugeno, 1985), 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. 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 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 into multiple anatomical sectors (see Fig. 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 (Ordas et al., 2005b), 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 a single standard fuzzy C-means (FCM) (Bezdek, 1981) clustering operation for each anatomical sector, 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. 5) for each pixel

$$T(x, y) = \begin{cases} \text{blood pool} & \text{if } L(x, y) = \text{bright}, \\ \text{myocardium} & \text{if } L(x, y) = \text{medium bright}, \\ \text{air} & \text{if } L(x, y) = \text{dark} \end{cases}$$

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 **a** any other tissue at the lung, anterior and posterior wall, **b** 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
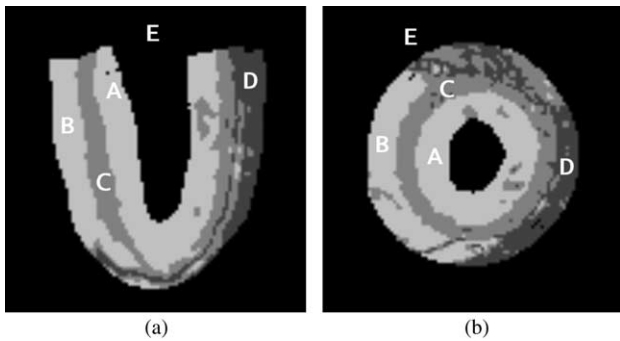

Fig. 5. (a) Classified set of image patches from a radial image. (b) Classified set of image patches from a short-axis image (A = LV blood pool, B = RV blood pool, C = myocardium, D = air, E = outside image patches).

Table 1
Optimal parameter set obtained through exhaustive search using grid-computing techniques

Defuzzification (see Section 2.5, Ordas et al., 2005a,b)			
Air cut-off proportion p			−0.20
Blood pool membership threshold			0.20
Myocardium membership threshold			0.05
Air membership threshold			0.50
ASM (Ordas et al., 2005a,b)			
Number of modes of variation			60
Allowed variation per mode (β)			2σ
Propagation	Data (Section 3.2)	# Planes	σ_p
Gaussian kernel width σ_p (Eq. (8))	RAD-4	4	8 mm
	RAD-2a	2	3 mm
	RAD-2b	2	7 mm
	MV-4	4	6 mm
	SA-full	11	6 mm
	SA-4—SA-10	4–10	6 mm

Values for the defuzzification and ASM parts are taken from (Ordas et al., 2005b). Propagation parameters were optimized for the evaluation data set in this paper.

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 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 Table 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.

3. Experimental setup

3.1. Test data and protocol

3.1.1. 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° (see Fig. 6(a)). All four slices were acquired in the same breath hold with the TFE protocol (see Fig. 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 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, radial sets (RAD, $N = 15$), a multi-view set (MV, $N = 20$), and short-axis sets (SA, $N = 15$), which are explained in more detail in Section 3.2.

3.1.2. 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 delineation 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 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(a)). Per subject, a mean error was computed. Volumes were calculated by closing the model at the base (see Fig. 7(b)) and the manual shapes at both the apex and the base (see Fig. 7(c)). Final P2S segmentation errors achieved by other groups with their model-based cardiac segmentation techniques are presented in Table 2.

To test whether differences in the P2S errors were significantly different, t-tests were performed. Observations on the SA, SA-4 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. Com-
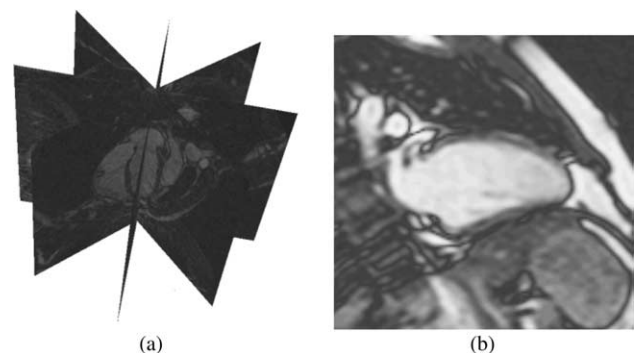

(a)   (b)

Fig. 6. (a) Radial cardiac image stack. (b) Radial slice acquired with the turbo field echo (TFE) protocol.
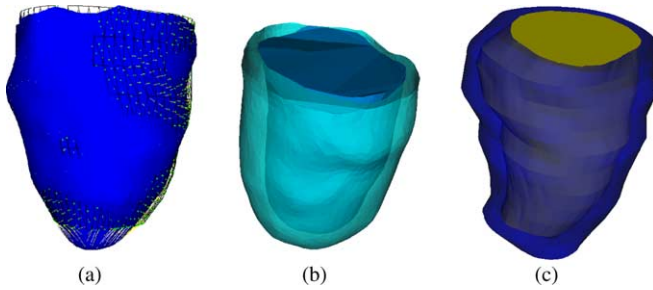
Fig. 7. (a) Point-to-surface distances are measured from points on the model to the closest points on the manual surface. (b) Closed model for volume estimation. (c) Closed manual shape for volume estimation.

parison of the results achieved on the MV data set to those from any other data set was performed with an independent-samples *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 for statistical testing.

3.1.3. 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 thoracic 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.

3.1.4. 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 previous work (Ordas et al., 2005b). Best settings for the update propagation parameter (σ_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 described in Section 3.3, 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 1. For the different combinations of image slice orientations, different settings for the propagation kernel width (σ_p) were found to give the best results.

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 the literature. Moreover, SPASM was tested on a range of data configurations seeking the sparsity limit, i.e., to find out how many slices are required for accurate quantification, without producing significantly different results from the complete SA data set, SA-full.

Different slice combinations were created and used for evaluation:

RAD-4. Full data set of four images acquired in a radial orientation with the TFE protocol and inter-slice angle of 45° (see Fig. 8(a)).

Table 2
Point-to-surface distances from automatic results for other 3D automatic cardiac segmentation methods (all values are $\mu \pm \sigma$ in mm)

	Endocardium	Epicardium
Lötjönen et al. (2004a)[a]	2.01 ± 0.31	2.77 ± 0.49
Mitchell et al. (2002)[b]	2.75 ± 0.86	2.63 ± 0.76
Kaus et al. (2004)	2.28 ± 0.93	2.62 ± 0.75
Lorenzo-Valdés et al. (2004)[c]	1.88 ± 2.00	2.75 ± 2.62

[a] 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.
[b] Mitchell et al. compute automatic segmentation errors 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.
[c] 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.
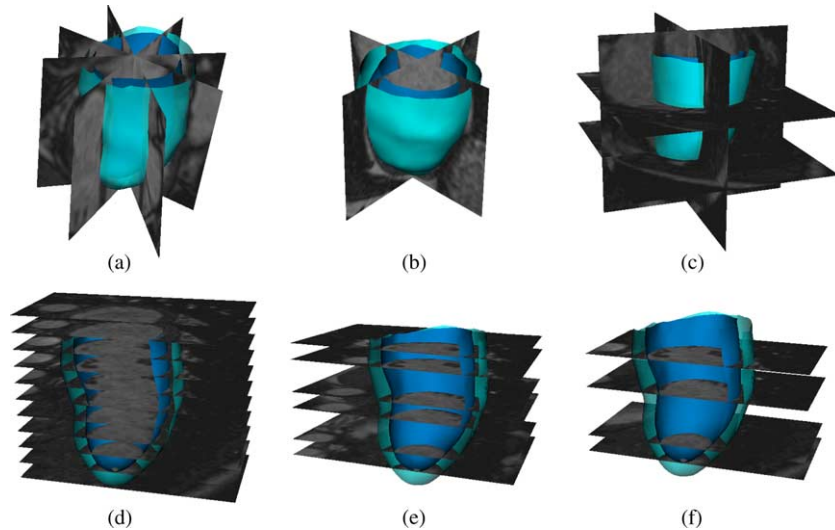
Fig. 8. (a) RAD view with 4 planes. (b) RAD view with 2 planes. (c) Multiple view with 2 SA views and a 2-chamber and a 4-chamber view. (d) All 11 SA planes. (e) Combination of 6 SA planes. This combination of 6 planes gave the best overall segmentation result of all 462 possible combinations of 6 SA planes out of 11. (f) 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.

RAD-2a. Sub-set of slices 0 and 2 from RAD-4, with inter-slice angle of 90° (see Fig. 8(b)).

RAD-2b. Sub-set of slices 1 and 3 from RAD-4, with inter-slice angle of 90°.

MV-4. Multi-view combination of an apical and a mid-ventricular SA slice, and a 2-chamber- and a 4-chamber view (see Fig. 8(c)).

SA-full. Full short-axis stack (11 slices, see Fig. 8(d)).

SA-4, SA-5, SA-6, SA-7, SA-8, SA-9, SA-10. Short-axis stack sub sets of 4, 5, 6, 7, 8, 9, and 10 SA slices. All possible combinations of 4, 5, 6, 7, 8, 9, and 10 SA slices were tested (see e.g., Fig. 8(e) and (f)).

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 and $+20$ mm, and with steps of 5 mm. For the z-coordinates, offsets ranged between -10 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 the cluster described in Section 3.3. The test variable was the resulting P2S error measure, both for the epicardial surface and for the endocardial surface.

3.3. Model tuning using a grid middleware

For the tuning of parameters involved in the model matching, a computer cluster running a grid middleware platform was used. The Grid middleware platform is the InnerGrid Nitya developed by GridSystems (Mallorca,

Spain), running on a 45-node dual Xeon (2.8 GHz CPU, 2 GB RAM) cluster, under Linux RedHat 9, accessible through a web interface (desktop portal) and from an API. As a whole, the cluster represents more than 103 GFlops of computing power. Each node in the cluster provides two agents to the grid, therefore a total of 90 agents were available.

The main components of the system are the server, which is the coordinator of the grid, and the agents, i.e., the pieces of software installed in all the machines federated in the grid. Therefore, the topology of the middleware is that of a server connected to a large group of agents that can be compiled for several platforms. The server is thus capable of sending the corresponding application, suitable for each operating system.

A tuning of the parameters that affect the appearance model and the β parameter was performed. The remaining parameters (χ and σ_p) were fixed at values beforehand expected to be not far from optimal. Parameters for six different models were tuned (Ordas et al., 2005b). The process lasted 1.5 days and yielded approximately 64,800 results. The optimal set of parameters of this first step were used in a second run for tuning the other two parameters that affect the shape model. This second run lasted 0.9 days and produced 5670 result files. In both runs, the post-processing was done in a local machine but could also be performed in the grid server itself, as a post-processing task. In addition, the propagation kernel width (σ_p) and the maximum allowed deformation per mode (β) were tuned for best performance on the different sparse evaluation data sets used here. This process required 540 additional runs of the model.

The data set used for model tuning consisted of 30 subjects. Fifteen were short-axis scans of healthy volunteers acquired with the balanced fast field echo protocol (BFFE) using the QBody coil on a Philips Gyroscan NT Intera,

1.5T MR scanner (Philips Medical Systems, Best, Netherlands). The other 15 studies included patients acquired using a General Electric Signa CV/i, 1.5T scanner (General Electric, Milwaukee, USA) with the FIESTA scan protocol. Selected patients suffered from common cardiac diseases: myocardium infarction (10), hypertrophy (2), and pericarditis (3).

4. Results

The optimal parameters from the tuning with respect to the defuzzification, model deformation and update propagation are shown in Table 1. Segmentation performance results with respect to seeking the sparsity limit, optimal values for σ_p for the range of data configurations, and

Table 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

	Average P2S errors		Average volumes		σ_p	Planes
	endo	epi	endo	epi		
SA-full manual	1.27 ± 0.30	1.14 ± 0.29	112.2 ± 26.1	188.6 ± 35.7		0–10
RAD-4	2.24 ± 0.54	2.83 ± 0.78	132.5 ± 18.5	243.0 ± 35.0	8	0–3
RAD-2a	2.32 ± 0.41	3.07 ± 0.77	137.8 ± 20.6	257.9 ± 41.3	3	0, 2
RAD-2b	2.57 ± 0.76	3.19 ± 0.93	131.3 ± 18.3	247.8 ± 38.8	7	1, 3
SA-full	1.97 ± 0.54	2.23 ± 0.46	122.0 ± 27.3	219.3 ± 41.3	6	0–10
SA-4	2.14 ± 0.58	2.46 ± 0.50	125.2 ± 26.3	223.0 ± 40.6	6	1, 2, 5, 7
SA-5	2.08 ± 0.54	2.39 ± 0.44	124.7 ± 25.9	222.0 ± 39.4	6	1, 2, 4, 5, 7
SA-6	1.97 ± 0.52	2.30 ± 0.50	123.4 ± 25.2	220.3 ± 39.0	6	1, 2, 4, 5, 7, 8
SA-7	1.99 ± 0.47	2.25 ± 0.40	122.0 ± 26.7	218.8 ± 40.3	6	1–3, 5, 7–9
SA-8	1.99 ± 0.47	2.25 ± 0.40	122.0 ± 26.7	218.8 ± 40.3	6	1–3, 5, 7–10
SA-9	1.95 ± 0.50	2.25 ± 0.44	122.1 ± 26.9	219.6 ± 39.5	6	0–5, 7–9
SA-10	1.97 ± 0.54	2.23 ± 0.46	122.0 ± 27.3	219.3 ± 41.3	6	0–9
MV-4 manual			131.9 ± 31.5	227.1 ± 52.1		0–3
MV-4	2.02 ± 0.93	2.29 ± 0.53	127.5 ± 28.1	229.1 ± 41.6	6	0–3

All values are $\mu \pm \sigma$, P2S errors in mm and volumes in ml; endo means endocardium, and epi means epicardium.
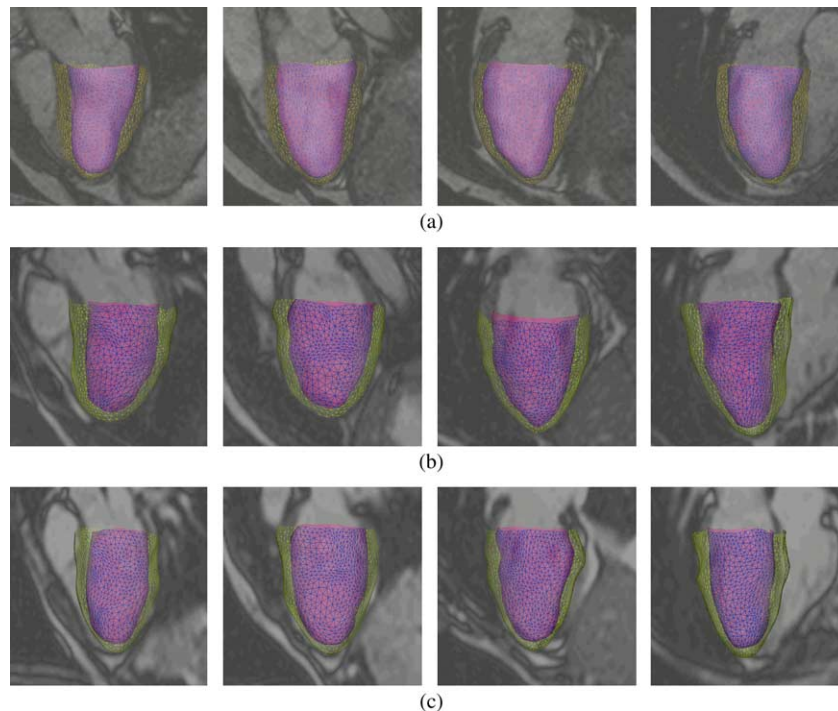

Fig. 9. Final segmentation result of some of the subjects in the test population. (a)–(c) Results shown on slices 1–4 of a RAD TFE stack, respectively. (a) Best result (subj. 1), ave. P2S: endocardium 1.62 mm, epicardium 1.54 mm. (b) Typical result (subj. 3), ave. P2S: endocardium 1.95 mm, epicardium 2.27 mm. (c) Worst result (subj. 6), ave. P2S: endocardium 3.20 mm, epicardium 3.73 mm.

the best slice combinations per data set are shown in Table 3. The results for the SA-X data sets only reflect the combinations of X planes that gave the best segmentation results. The optimal data configuration for SA-6 is shown in Fig. 8(e), and for SA-4 in Fig. 8(f). *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. Endo- and epicardial volumes were calculated and compared for all combinations to those achieved with SA-full. Comparison to manual volumes was not performed by means of statistical testing, because the manual volumes did not contain the apex, while the automatic volumes did.

Based on the results of the P2S distance measurements, for the MV test, one subject was excluded due to a mis-

match of the model for all settings for σ_p. For the SA and RAD tests, for all 15 subjects good results were obtained. However, due to an unrecoverable misalignment between a manual SA segmentation and an automatic segmentation on one RAD data set, one subject in the RAD configurations was excluded from further evaluation. Final segmentation results for three of the subjects from the RAD data set are shown on all four slices in Fig. 9. These are the best, an intermediate, and the worst results. In Fig. 10 distributions of the P2S errors over the model surfaces are shown color-coded for the best, worst and typical segmentation results on the RAD-4, MV-4, SA-4, SA-6 and SA-full data configurations. For a comparison of performance on MV-4 to both the interobserver variability and to the results achieved with
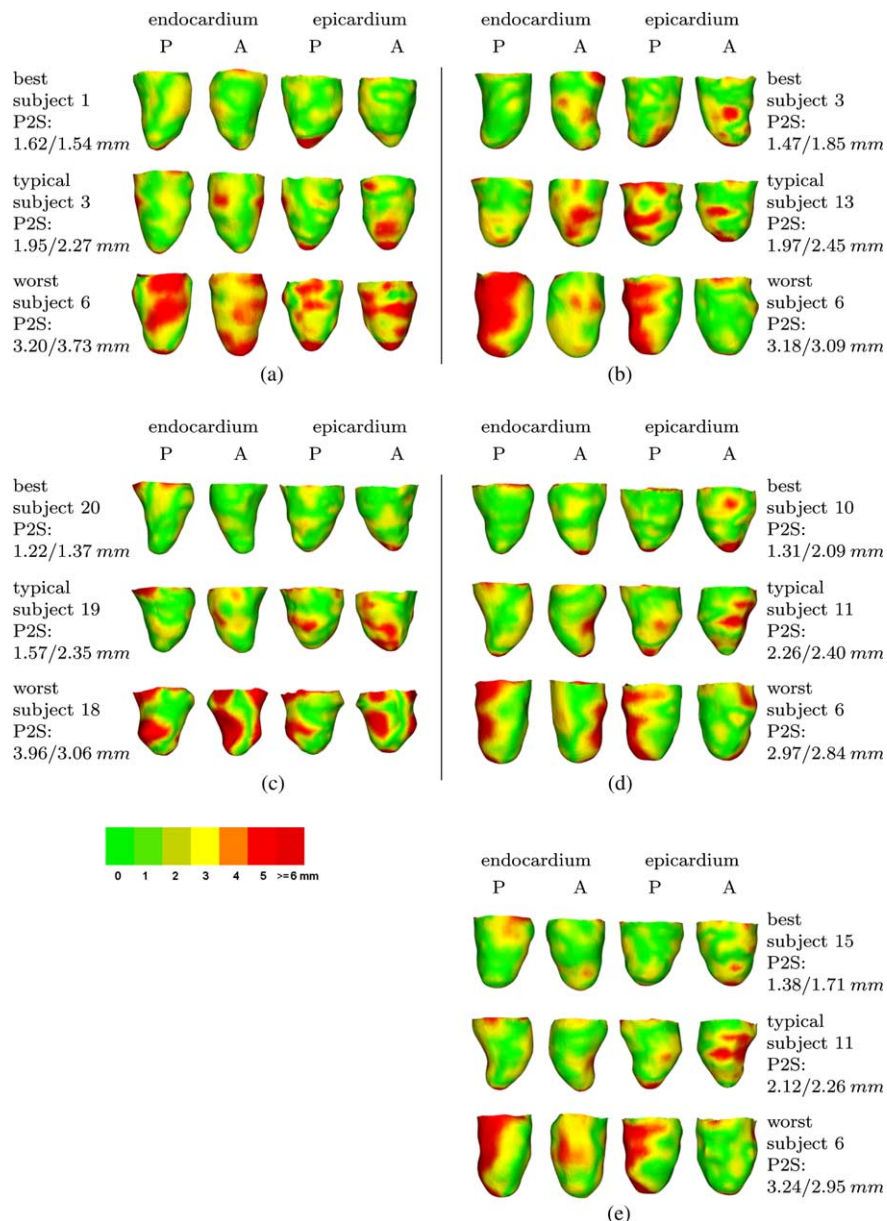

Fig. 10. P2S errors encoded in color for best, typical and worst segmentation results. (a) RAD-4, P2S *endocardium/epicardium*, (b) SA-4, P2S *endocardium/epicardium*, (c) MV-4, P2S *endocardium/epicardium*, (d) SA-6, P2S *endocardium/epicardium*, (e) SA-full, P2S *endocardium/epicardium*.

Table 4
Significance of point-to-surface errors from all data sets compared to the interobserver variability and to SA-full

	RAD			MV	SA						
	4	2a	2b	4	4	5	6	7	8	9	Full
Point-to-surface errors compared to interobserver variability											
Endocardium	.000*	.000*	.000*	.003*	.000*	.000*	.000*	.000*	.000*	.000*	.000*
Epicardium	.000*	.000*	.000*	.000*	.000*	.000*	.000*	.000*	.000*	.000*	.000*
Point-to-surface errors compared to SA-full											
Endocardium	.078	.049*	.001*	.861	.005*	.006*	.946	.695	.695	.471	
Epicardium	.027*	.003*	.002*	.757	.002*	.009*	.102	.691	.691	.502	
Volumes compared to SA-full											
Endocardium	.065	.002*	.135	.665	.020*	.038*	.171	.982	.982	.751	
Epicardium	.000*	.000*	.001*	.566	.011*	.108	.520	.639	.639	.757	

Volumes compared to SA-full. Numbers are *p*-values. Asterisks mark statistical significance at the 0.05 confidence level. Due to equal results to SA-full, *p*-values cannot be calculated for SA-10 (see Table 3).
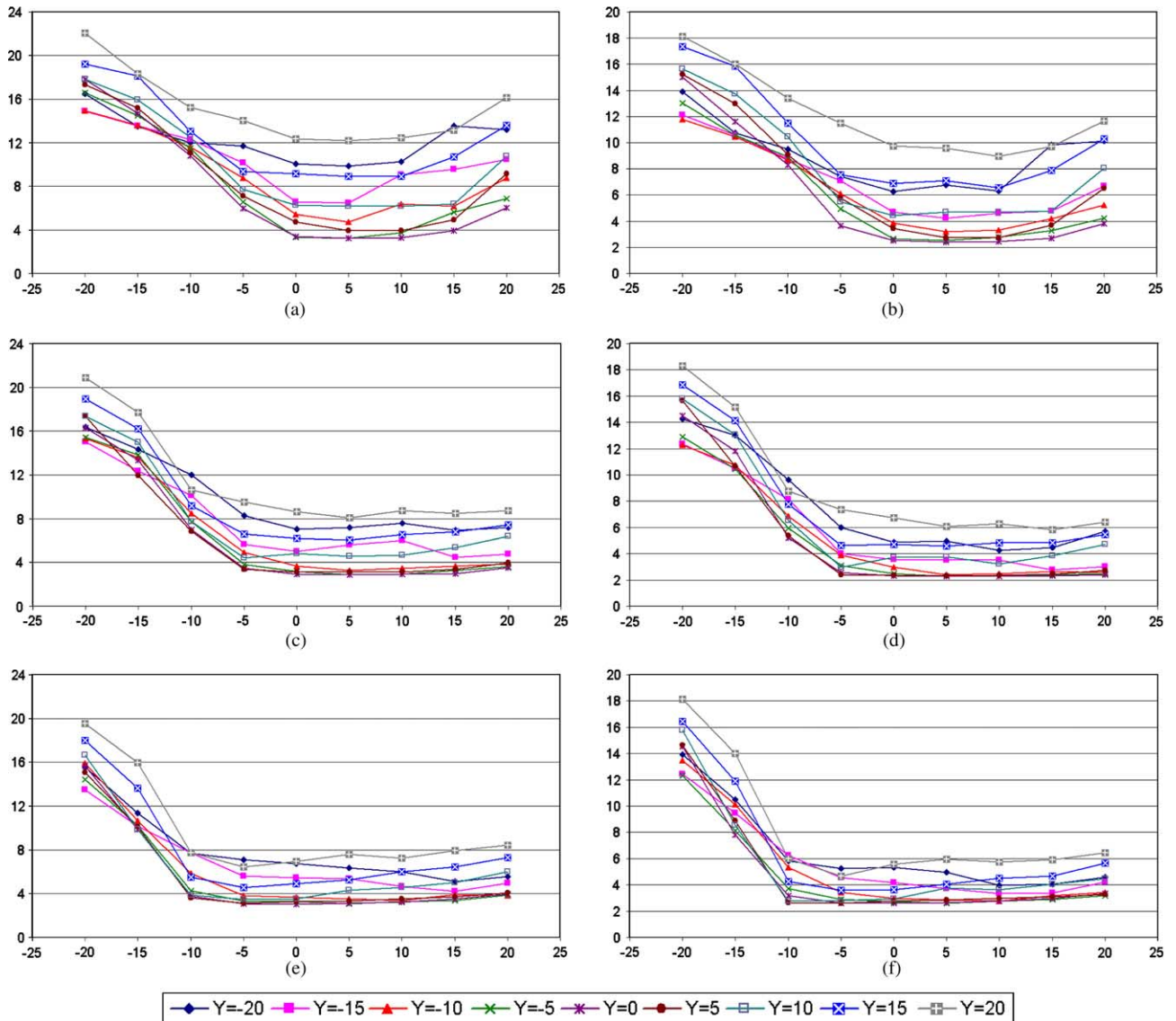

Fig. 11. Sensitivity to initial displacements. RAD data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the radial data set. Different curves show different displacements 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 are plotted along the *x*-axes of the plots.
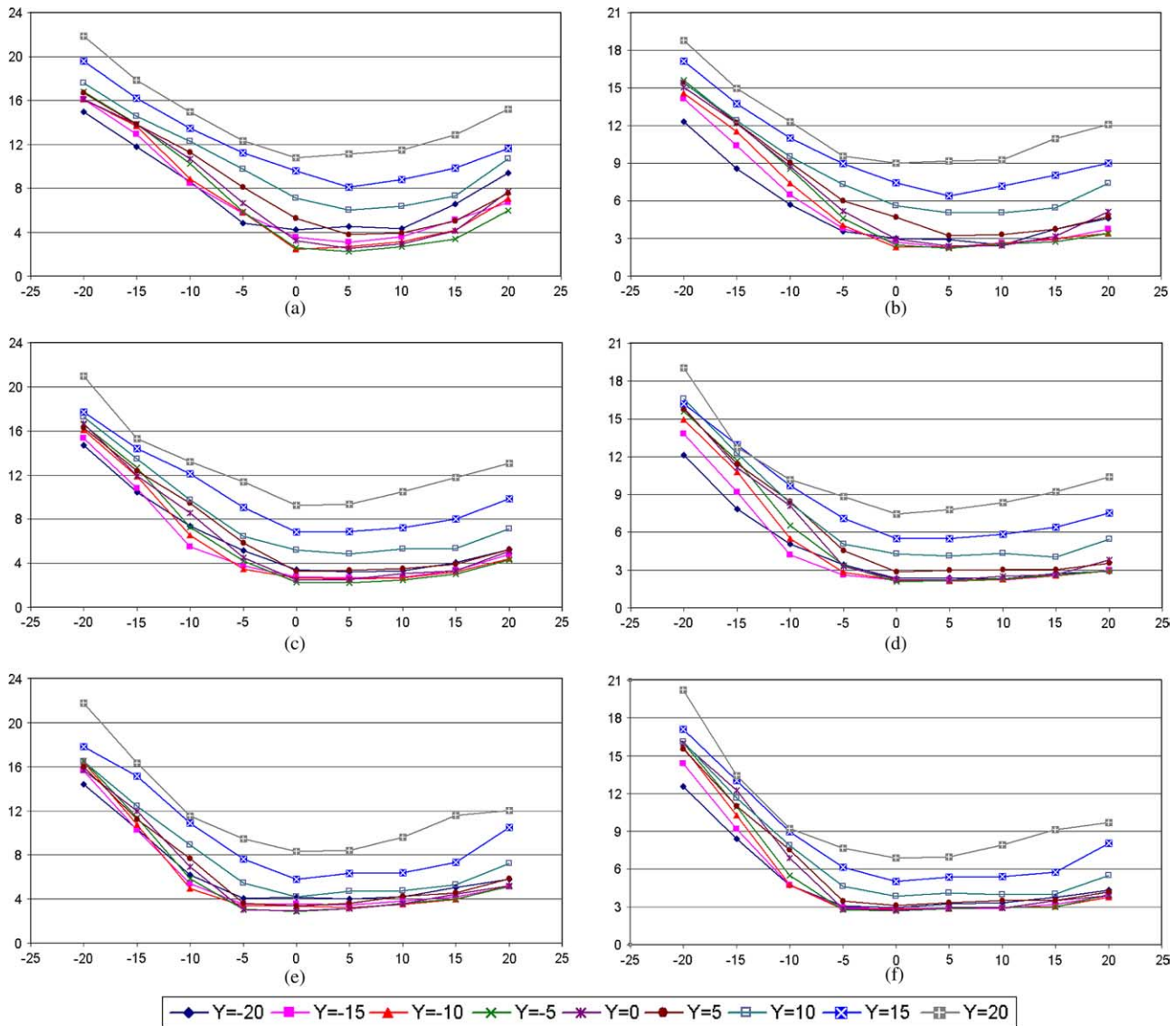
Fig. 12. Sensitivity to initial displacements. MV data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the multi-view data set. Different curves show different displacements 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 are plotted along the x-axes of the plots.

SA-full 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 4.

Results for the sensitivity of SPASM to initial model placement are given in Fig. 11 for the radial data set, in Fig. 12 for the multi-view data set, and in Fig. 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 the 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 s

($\sigma_p = 3$ mm) to 5 min ($\sigma_p = 10$ mm) on a standard desktop PC, both on the SA-full data set.

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
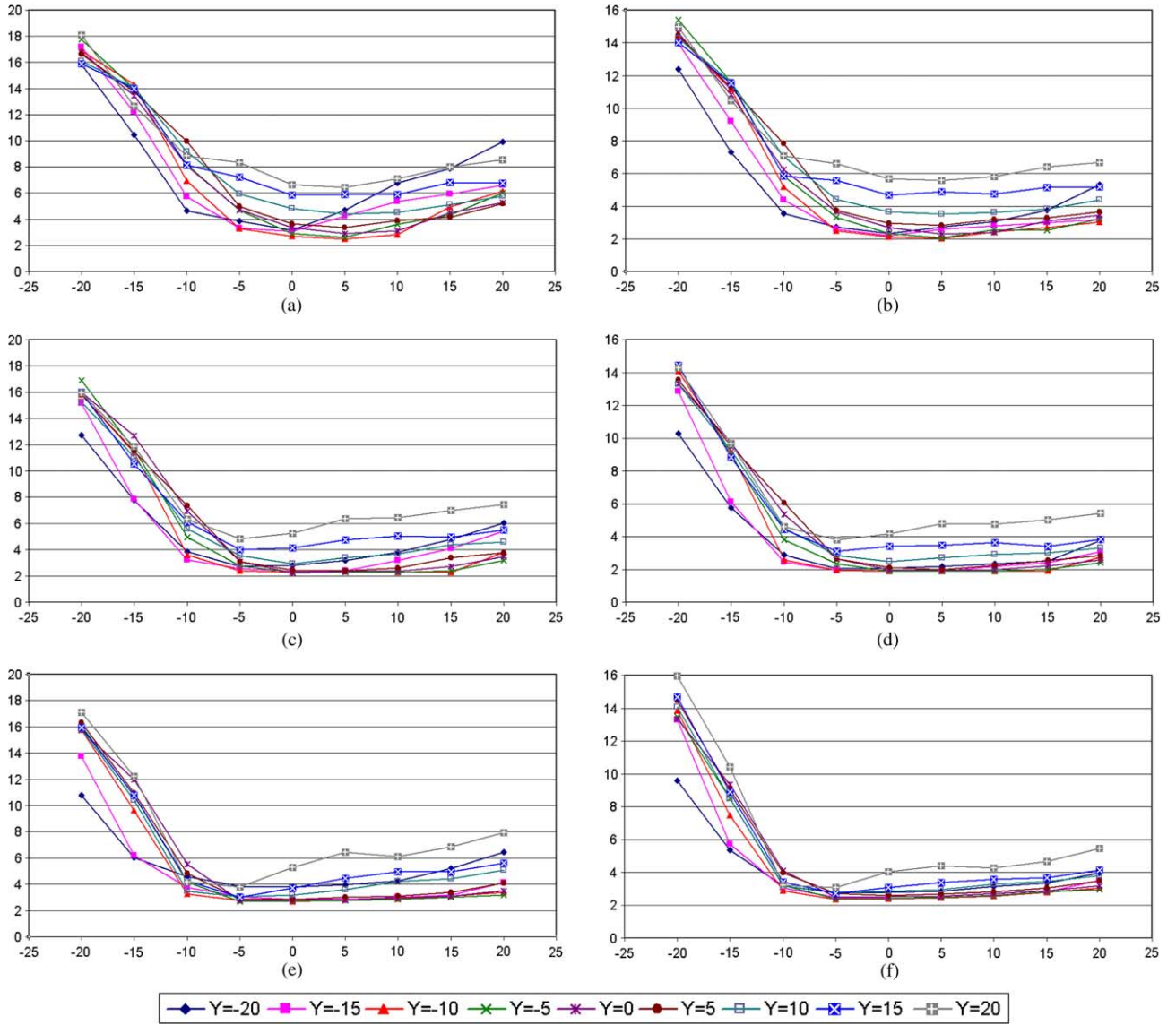
Fig. 13. Sensitivity to initial displacements. SA-full data set. Mean final segmentation errors (vertical axis in mm) for both the epicardium (a,c,e) and endocardium (b,d,f) as a result of perturbations in initial position of the model in the short-axis data set. Different curves show different displacements 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 are plotted along the x-axes of the plots.

discussed in Section 5.2. Both slice shift correction and initialization can be done automatically very well, but since those were not the topics of this paper, we performed them (partly) manually.

5.1. Segmentation performance

From the segmentation performance results on the different data sets used in this paper (see Tables 3 and 4), it can be seen that:

- on the SA data set (11 planes) SPASM works significantly better than on the SA-4 data set (four planes) for both the endocardium and the epicardium, However, for 6 slices and more, there are no significantly different results to SA-full.

- for the epicardium, segmentation performance on the SA-full data set is significantly better than on the RAD-4 data set, however, for the endocardium there is no significant difference.
- segmentation performances on the SA-full data set and the MV-4 data set, in particular, are almost the same (epicardium: 2.23 ± 0.46 and 2.29 ± 0.53 mm, respectively, endocardium: 1.97 ± 0.54 and 2.02 ± 0.93 mm, respectively). Differences between these two data configurations are not statistically significant (epicardium: $p = 0.757$, endocardium: $p = 0.861$).
- almost the same conclusions can be drawn from the statistical volume evaluations.
- with respect to point-to-surface errors from interobserver variability, the model performs worse for all data configurations.

Segmentation performance on the MV, SA-full, and SA-X data sets was slightly better than on the RAD data (see Table 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 s. The images in the SA data set (acquired with the BFFE protocol) were acquired in 15–20 s per slice (full stack approximately 220 s, i.e., 11 times 20 s plus the time between slice acquisitions).

Performance on the MV data set (BFFE) was slightly better than on the SA-4 and SA-5 data sets, possibly because of a better definition of the apex. In the SA data sets the apex was fuzzy, because of partial-volume effects. In the selection of image slices for the SA data sets, the best combination of planes always included at least two apical slices. Apex definition in LA views is better than in SA views (Bloomer et al., 2001). Possibly, by using at least two apical slices in the SA data sets, this feature is compensated.

From the *t*-tests (see Table 4), it appears that the accuracy with respect to surface localization achieved on the MV-4 (4 planes) data configuration and that achieved on SA-6 (6 planes) do not differ significantly from that achieved when SPASM was applied to SA-full (11 planes). However, P2S errors for all data configurations were significantly larger than the interobserver variability from the full SA data set (SA-full). This may be caused by the inclusion of the apex in the 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(a) and 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 4), volumes were not significantly different to SA-full from 6 SA slices and more and from MV-4. 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.

Global functional parameters like stroke volume and ejection fraction are calculated from the endocardial segmentation. Although there are some significant differences for SA-4 and SA-5 with respect to SA-full, these differences are so small (on the order of 3 ml for both endocardial and epicardial volumes), that they are not clinically relevant. Even with SA-4 and SA-5, we believe that quantification of volumes is clinically useful.

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 segmen-

tations are minimal (see Figs. 11–13). 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. 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. 11–13 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 (Lelieveldt et al., 2001; Jackson et al., 2004; Danilouchkine et al., 2005) can be utilized for initialization purposes. Those tools permit determination of the LV axis spatial orientation with an accuracy of 7° (Jackson et al., 2004; Danilouchkine et al., 2005). Danilouchkine et al. (2005) 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 (Lelieveldt et al., 2001; Jackson et al., 2004) and is 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.

5.3. Protocol independence

In previous work, we demonstrated that the 3D-ASM was able to segment CT-data (van Assen et al., 2003) and MRI-data (Ordas et al., 2005b) 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, and demonstrated on MR and CT images (Ordas et al., 2005b; van Assen et al., 2003).

In this work, the results demonstrated that application to different MRI acquisition protocols without any adaptation to 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 σ_p (see Table 1) between the RAD data set and the other data sets were solely caused by differences in the sampling density of the data sets.

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 (Dornier et al., 2003; Bidaut and Vallée, 2001; Stegmann et al., 2005; Stegmann and Pedersen, 2005; Lötjönen et al., 2004b).

Segmentation errors of all methods are substantially larger than the interobserver variability (see Tables 2 and 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.

In general, models of higher dimensionality require more training data sets to derive relevant shape statistics. The fact that we did not study the effect of the number of training samples on the model statistics, given the high dimensionality of the model, may be regarded a limitation.

5.5. Comparison to other work

Performance of SPASM has been compared to results from other methods reported in literature (Kaus et al., 2004; Mitchell et al., 2002; Lötjönen et al., 2004a) (see Table 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. (2004a) and Lorenzo-Valdés et al. (2004), and comparable to or better than the precision achieved by the methods of Mitchell et al. (2002) and Kaus et al. (2004). However, in three of the tested data configurations, SPASM was applied to four image slices, whereas the other models used a stack of 7–10 SA slices (Kaus et al., 2004; Mitchell et al., 2002; Lorenzo-Valdés et al., 2004) and a combination of two stacks (4–5 SA plus 4–7 LA slices) (Lötjönen et al., 2004a). It has to be noted, however, that we acquired new data sets, so we did not use exactly the same data sets that were used in (Kaus et al., 2004; Mitchell et al., 2002; Lötjönen et al., 2004a; Lorenzo-Valdés et al., 2004).

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 (Raya and Udupa, 1990; Grevera and Udupa, 1996) is ill-posed for radial LA data sets like RAD-4, RAD 2a and Rad 2b. Second, it may be difficult to estimate the required intensity distributions (Lorenzo-Valdés et al., 2004) 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 (Koikkalainen and Lötjönen, 2004; Fleute et al., 1999), 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-4 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 Fig. 7(a), and Fig. 10 shows that in many cases errors close to the apex are larger than at other locations. Consequently, the errors reported in Table 3 overestimate the true segmentation errors, i.e., regarding the locations where both automatic and manual surface are present.

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-full data set (4 slices vs. 11 slices, respectively). Also, performance on 6 or more SA slices was comparable to performance on SA-full (11 slices). This underlines the robustness of SPASM with respect to slice sparsity, and shows that our method is not very sensitive to missing data.

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.

References

Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. 14 (2), 239–256.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Bidaut, L.M., Vallée, J.P., 2001. Automated registration of dynamic MR images for the quantification of myocardial perfusion. J. Magn. Reson. Imaging 13, 648–655.

Bloomer, T.N., Plein, S., Radjenovic, A., Higgins, D.M., Jones, T.R., Ridgway, J.P., Sivananthan, M.U., 2001. Cine MRI using steady state free precession in the radial long axis orientation is a fast accurate method for obtaining volumetric data of the left ventricle. J. Magn. Reson. Imaging 14, 685–692.

Chandrashekara, R., Rao, A., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D., 2003. Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In: Taylor, C.J.A.N. (Ed.), Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol. 2732. Springer, Berlin, pp. 599–610.

Cootes, T.F., Cooper, D., Taylor, C.J., Graham, J., 1992. A trainable method of parametric shape description. Image Vision Comput. 10 (5), 289–294.

Cootes, T.F., Cooper, D., Taylor, C.J., Graham, J., 1995. Active shape models – their training and application. Comput. Vision Image Understanding 61 (1), 38–59.

Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active appearance models. IEEE Trans. Pattern Anal. 23 (6), 681–685.

Cootes, T.F., Taylor, C.J., 2004. Statistical models of appearance for computer vision. Technical Report, Imaging Science and Biomedical Engineering, University of Manchester. URL: http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf.

Danilouchkine, M.G., Westenberg, J.J.M., Reiber, J.H.C., Lelieveldt, B.P.F., 2005. Accuracy of short-axis cardiac MRI automatically derived from scout acquisitions in free-breathing and breath-holding modes. Magn. Reson. Mater. Phys. 18 (1), 7–18.

Dornier, C., Ivancevic, M.K., Thévenaz, P., Vallée, J.P., 2003. Improvement in the quantification of myocardial perfusion using an automatic spline-based registration algorithm. J. Magn. Reson. Imaging 18, 160–168.

Fleute, M., Lavallée, S., Julliard, R., 1999. Incorporating a statistically based shape model into a system for computer-assisted anterior cruciate ligament surgery. Med. Image Anal. 3 (3), 209–222.

Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J., 2002. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. IEEE Trans. Med. Imaging 21 (9), 1151–1166.

Gower, J., 1975. Generalized procrustes analysis. Psychometrika 40, 33–50.

Grevera, G.J., Udupa, J., 1996. Shape-based interpolation of multidimensional gray-level images. IEEE Trans. Med. Imaging 15, 881–892.

Jackson, C.E., Robson, M.D., Francis, J.M., Noble, J.A., 2004. Computerized planning of the acquisition of cardiac MR images. Comput. Med. Imag. Grap. 28, 411–418.

Kaus, M.R., von Berg, J., Weese, J., Niessen, W.J., Pekar, V., 2004. Automated segmentation of the left ventricle in cardiac MRI. Med. Image Anal. 8 (3), 245–254.

Kelemen, A., Szekely, G., Gerig, G., 1999. Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans. Med. Imaging 18 (10), 828–839.

Koikkalainen, J., Lötjönen, J., 2004. Reconstruction of 3-D head geometry from digitized point sets: an evaluation study. IEEE Trans. Inf. Technol. B 8 (3), 377–386.

Lelieveldt, B.P.F., van der Geest, R.J., Lamb, H.J., Kayser, H.W., Reiber, J.H.C., 2001. Automated observer-independent acquisition of cardiac short-axis MR images: a pilot study. Radiology 221 (2), 537–542.

Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Elkington, A.G., Mohiaddin, R., Rueckert, D., 2004. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Med. Image Anal. 8, 255–265.

Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K., 2004a. Statistical shape model of atria, ventricles and epicardium from short- and long-axis mr images. Med. Image Anal. 8 (3), 371–386.

Lötjönen, J., Pollari, M., Kivistö, S., Lauerma, K., 2004b. Correction of movement artifacts from 4-D cardiac short- and long-axis MR data. In: Barillot, C., Haynor, D.R., Hellier, P. (Eds.), MICCAI 2004, Lecture Notes in Computer Science, vol. 3217. Springer, Berlin, pp. 405–412.

Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M., 2002. 3D active appearance models: segmentation of cardiac MR and ultrasound images. IEEE Trans. Med. Imaging 21 (9), 1167–1178.

Montagnat, J., Sermesant, M., Delingette, H., Malandain, G., Ayache, N., 2003. Anisotropic filtering for model-based segmentation of 4d cylindrical echocardiographic images. Pattern Recogn. Lett. 24 (4–5), 815–828.

Ordas, S., Boisrobert, L., Bossa, M., Laucelli, M., Huguet, M., Olmos, S., Frangi, A.F., 2004. Grid-enabled automatic construction of a two-chamber cardiac PDM from a large database of dynamic 3D shapes. In: IEEE International Symposium of Biomedical Imaging, pp. 416–419.

Ordas, S., van Assen, H.C., Boisrobert, L., Laucelli, M., Puente, J., Lelieveldt, B.P.F., Frangi, A.F., 2005a. Statistical modeling and segmentation in cardiac MRI using a grid computing approach. In: Sloot, P.M.A., Hoekstra, A.G., Priol, T. (Eds.), European Grid Conference, Lecture Notes in Computer Science, vol. 3470. Springer, Berlin, pp. 6–15.

Ordas, S., van Assen, H.C., Puente, J., Lelieveldt, B.P.F., Frangi, A.F., 2005b. Parametric optimization of a model-based segmentation algorithm for cardiac MR image analysis: a grid-computing approach. In: Solomonides, T., Mcclatchey, R., Breton, V., Legré, Y., Norager, S. (Eds.), from Grid to Healthgrid, Proceedings of Healthgrid 2005, Studies in Health Technology and Informatics, vol. 112. IOS Press, Oxford, pp. 146–156.

Raya, S., Udupa, J., 1990. Shape-based interpolation of multidimensional objects. IEEE Trans. Med. Imaging 9, 32–42.

Rueckert, D., Frangi, A.F., Schnabel, J.A., 2003. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. IEEE Trans. Med. Imaging 22 (8), 1014–1025.

Stegmann, M.B., Ólafsdóttir, H., Larsson, H.B.W., 2005. Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. Med. Image Anal. 9 (4), 394–410.

Stegmann, M.B., Pedersen, D., 2005. Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation. In: Fitzpatrick, J.M., Reinhardt, J.M. (Eds.), Proceedings of the SPIE. Medical Imaging, vol. 5747, San Diego, CA, pp. 336–350.

Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybernet. 15 (1), 116–132.

van Assen, H.C., Danilouchkine, M.G., Behloul, F., Lamb, H.J., van der Geest, R.J., Reiber, J.H.C., Lelieveldt, B.P.F., 2003. Cardiac LV segmentation using a 3D active shape model driven by fuzzy inference. In: Ellis, R., Peters, T. (Eds.), MICCAI 2003,

Lecture Notes in Comput Science, vol. 2878. Springer, Berlin, pp. 533–540.

van Rossum, A.C., Visser, F.C., Van Eenige, M.J., Valk, J., Roos, J.P., 1987. Oblique views in magnetic resonance imaging of the heart by combined axial rotations. Acta Radiol. 28, 497–503.

van't Ent, D., de Munck, J.C., Kaas, A.L., 2001. A fast method to derive realistic BEM models for E/MEG source reconstruction. IEEE Trans. Bio-Med. Eng. 48 (12), 1434–1443.

Yushkevich, P., Fletcher, P.T., Joshi, S., Thall, A., Pizer, S.M., 2003. Continuous medial representations for geometric object modeling in 2D and 3D. Image Vision Comput. 21 (1), 17–27.