
Article:
Butakoff, C., Balocco, S., Sukno, F.M. et al. (4 more authors) (2016) Left-ventricular epi- and endocardium extraction from 3D ultrasound images using an automatically constructed 3D ASM. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 4 (5). pp. 265-280. ISSN 2168-1163

https://doi.org/10.1080/21681163.2014.910703

Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2014
http://dx.doi.org/10.1080/21681163.2014.910703

Left-ventricular epi- and endocardium extraction from 3D ultrasound images using an automatically constructed 3D ASM

C. Butakoff[a]*, S. Balocco[b], F.M. Sukno[a], C. Hoogendoorn[a], C. Tobon-Gomez[c], G. Avegliano[d] and A.F. Frangi[e]

[a]*Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain;* [b]*Department of Applied Mathematics and Analysis, Universitat de Barcelona, Spain; Computer Vision Center, Bellaterra, Spain;* [c]*Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK;* [d]*Cardiac Imaging Department, Instituto Cardiovascular de Buenos Aires, Buenos Aires, Argentina;* [e]*CISTIB Center for Computational Imaging and Simulation Technologies in Biomedicine, Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*

(*Received 13 December 2013; accepted 28 March 2014*)

In this paper, we propose an automatic method for constructing an active shape model (ASM) to segment the complete cardiac left ventricle in 3D ultrasound (3DUS) images, which avoids costly manual landmarking. The automatic construction of the ASM has already been addressed in the literature; however, the direct application of these methods to 3DUS is hampered by a high level of noise and artefacts. Therefore, we propose to construct the ASM by fusing the multidetector computed tomography data, to learn the shape, with the artificially generated 3DUS, in order to learn the neighbourhood of the boundaries. Our artificial images were generated by two approaches: a faster one that does not take into account the geometry of the transducer, and a more comprehensive one, implemented in Field II toolbox. The segmentation accuracy of our ASM was evaluated on 20 patients with left-ventricular asynchrony, demonstrating plausibility of the approach.

Keywords: ASM; cardiac segmentation; statistical model; shape model; 3D ultrasound; cardiac segmentation

1. Introduction

Ultrasound (US) is known to be the fastest, least expensive and least invasive modality for imaging the heart. Because of the 3D structure and deformation of the heart muscle during the cardiac cycle, analysis of irregularly shaped cardiac chambers or description of valve morphology using 2D images is inherently limited. Developments in 3D echocardiography started in the late 1980s (Woo 2009). During the last two decades it evolved from free-hand scanning, replaced later by mechanical scanning of several planes using a linear transducer, to matrix phased-array transducers that are able to acquire a 3D volume of the whole heart almost in real time (RT).

The appearance of this new modality brought in new challenges and the need for new analysis tools, many of which rely on correct segmentation of the myocardium. However, the presence of artefacts or having a poor window limits the accuracy of such tools. In fact, the suboptimal quality forced many studies to reject up to one-third of the data (Nikitin et al. 2006; Bellenger et al. 2000). An extensive survey of traditional approaches to US segmentation can be found in Noble and Boukerroui (2006), Frangi et al. (2005), Lelieveldt et al. (2006), Angelini, Homma et al. (2005) and Angelini, Jin et al. (2005). The classifications of approaches to modelling cardiac geometry can be found in Montagnat and Delingette (2001) and Frangi et al. (2005).

Let us start with approaches using explicit surface representation. To introduce a shape constraint on a segmentation algorithm, Hong et al. (2006) proposed to use a set of prototype shapes. The resulting shape is a Nadaraya–Watson kernel-weighted average of the prototypes. The authors propose to use 2D Haar-like features to detect the myocardium and require manual annotation in four-chamber view to reduce the search space of the optimisation algorithm. The alignment of the prototypes uses as well the known apical four-chamber view plane. To compute the 2D Haar-like features, the authors propose to cut the 3D volume into several long-axis slices. It might be useful to note that the segmentation accuracy is evaluated in voxels, and their conversion to millimetres is not straightforward.

A fully automatic registration-assisted segmentation approach with a wire mesh was proposed in Zagrodsky et al. (2005). Rigid registration based on mutual information is used to initialise the segmentation. External forces are generated using a 3D extension of the Sobel edge detector with clamped intensities to remove strong and weak edges. Subsequently the zone of influence of the edges is enlarged by a 3D extension of the generalised gradient vector field.

Ping et al. (2007) proposed an active contour approach on a multilevel cubic B-spline grid for segmenting both epi- and endocardium on echocardiographic images with contrast agent. The movement of the contour is performed by displacing vertices of the grid. The segmentation

*Corresponding author. Email: constantine.butakoff@upf.edu

combines fuzzy feature information and a multilevel free-form deformation model into the objective function that has to be minimised in order to obtain accurate segmentation. This is the only study on RT 3D (RT3D) contrast US data that demonstrates segmentation of the epicardium. The idea is to segment the endocardium and use it as a constraint for epicardium segmentation.

A segmentation of triplane echocardiograms by 2D-constrained active appearance motion models (AAMMs) (Sonka et al. 2001; Bosch et al. 2002) is presented in Hansegård, Urheim et al. (2007). The authors demonstrate a method of constraining AAMM to known positions of several points. The AAMM is used to learn the statistics of the shape, represented by contours in three planes, along the whole cardiac cycle, preserving the shape–time correspondence. This is achieved by concatenating the shapes in all temporal phases into a single vector. The points to constrain the AAMM are estimated by dynamic-programming-based active contours, and are searched for in the vicinity of the contour provided by the AAMM. Such a fusion of the AAMM and active contours in an iterative scheme demonstrates better segmentation accuracy than if these segmentation algorithms were used separately.

Finally, we briefly mention several works on cardiac wall tracking and dynamic active shape/appearance models. Orderud, Hansegård et al. (2007) and Hansegård, Orderud et al. (2007) use similar methodology: the cardiac contour is deformed by integrating the search for cardiac boundaries along contour normals into the extended Kalman filter. Essentially, the difference is that in Hansegård, Orderud et al. (2007) a shape prior based on principal component analysis (PCA) is incorporated into the framework. Another publication from the same group (Orderud et al. 2008) has shown how both epi- and endocardium can be coupled in the tracking framework. A classical correlation-based tracking was investigated in Crosby et al. (2009) and Duan et al. (2009). The first one applied normalised correlation to envelope detected beam data directly while a curious feature of the second one is that evaluation has been performed using the open-chest set-up (canine hearts). One more study, which could be of interest, is Gerard et al. (2002), where instead of using state space equations, the deformation of the normal left ventricle (LV) is learnt from tagged MR.

Among level-set-based approaches, it is worth to mention a work by Corsi et al. (2002), who modified the Malladi–Sethian equation for gradient-based image segmentation. They removed the inflationary term to avoid the propagation of the evolving surface beyond regions with missing boundaries (boundary leaking). Since the shape cannot inflate any more, the evolution has to start close to the true boundary. The authors suggested that initialisation by placing – five to seven points in five short-axis view slices is enough. Three years later a homogeneity-based active contour, integrated into a level-set framework, which does not use image gradient, was proposed by Angelini, Homma et al. (2005). The authors start by denoising the RT3D US image using brushlets. The idea behind the method is to deform the surface looking for an optimal partitioning of the voxels into homogeneous regions (inside and outside the surface). An interesting fact about the approach is that it extracts highly curved surfaces with minimal boundary leaking. Another level-set-based approach was proposed by Corsaro et al. (2006). It is an unconditionally stable 3D semi-implicit time discretisation scheme for solving the level-set formulation of the Riemannian mean curvature flow problem, which allows for fast image segmentation. Using the combination of finite element and finite volume methods, they achieved 10 times speed-up in comparison with the classical level-sets. Another important contribution is the elimination of the orientation effect (the authors show that the evolved surface that uses left or right oriented triangulation differs from the exact solution and propose their solution to that problem). The proposed framework appears to be robust to vanishing gradients, with a highly curved resulting surface just as in Angelini, Homma et al. (2005). On the downside there is no evaluation on a database of clinical cardiac images.

Many of the mentioned methods use a predefined shape model, which is matched to a 3D image. The benefit of using a predefined shape is that it simplifies establishing links between the model and the cardiac anatomy and allows to easily correlate data between different studies, patients or modalities. From the algorithmic point of view, imposing shape regularity constraints on the predefined model, instantiated to the data, would allow to robustly recover the correct shape even in the areas of ill-defined borders, which are typical for 3D US (3DUS). Another interesting problem not addressed by most papers is the segmentation of the epicardium. Being able to segment both epi- and endocardium could provide an interesting insight into myocardial deformation and wall thickening, which is already being measured in other imaging modalities.

In this paper we consider the problem of automatic construction of a 3D active shape model (ASM) (Cootes and Taylor 1992) and using it to segment the epicardium and endocardium of a cardiac LV in 3DUS images. Building an ASM would require constructing a model of plausible shape variations (a point distribution model or PDM) and a model of local image appearance in the vicinity of each shape point. These requirements lead to the necessity of having a database of images with delineated myocardial contours. In the imaging modalities, which provide high-quality images, such as multidetector computed tomography (MDCT), the process of obtaining shapes, corresponding to the images, can be automated through registration (Ordas, Oubel et al. 2007), but it is not easy to do in every modality. 3DUS images are rather noisy and have poor level of detail, which might render

any registration algorithm ineffective. Therefore, we cannot rely only on US data and so propose to automatically construct the PDM from MDCT images as in Ordas, Oubel et al. (2007). The local appearance model, on the other hand, can be automatically learned from synthetic US data. Having the LV geometry defined by the PDM allows generating a collection of plausible shapes and the corresponding 3DUS images. In this case, the cardiac boundaries will be in the positions given by the shape, and the size of the set is limited only by the available computational resources. The synthetic data were generated using a simplified model of US image formation, proposed in Yu and Acton (2002) and Bamber and Dickinson (1980) and extended to 3D, and a more comprehensive approach was implemented in Field II (Jensen and Svendsen 1992; Jensen 1996). The first of these two models, we shall call it FastGen for convenience, assumes a uniform point spread function (PSF), is very fast and takes seconds for generating one 3D image. The second one has a more realistic model of sound propagation, is computationally expensive and requires approximately 30 h on a single CPU (Intel Xeon 5140, 2.33 GHz). In this work we compare the segmentation accuracy of an automatically constructed ASM, trained on real data and on images generated using both US models.

The paper is structured as follows. We start by a brief description of ASM in Section 2, which is used for both providing an US image generator with the information about cardiac geometry and segmenting US images. The description of the datasets used for the evaluation of the automatically constructed ASM is presented in Section 3, followed by the evaluation itself in Section 4 and conclusions in Section 5. A brief description of US simulation approaches can be found in the appendices.

2. Active shape model

The linear ASM consists of a PDM and image intensity model. The PDM is constructed by applying PCA to a set of aligned shapes and retaining eigenvectors corresponding to a predefined percentage of shape variability (Cootes and Taylor 1992). Shapes are defined by landmark points placed along the contour of the object of interest. The shape alignment in 3D has been performed using unit-quaternions (Horn 1987). The learned shape variability can be modelled by varying $\mathbf{b}$ in the following equation:

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi\mathbf{b}, \tag{1}$$

where $\mathbf{x}$ is an n-vector, representing the shape, obtained by concatenating all the landmark coordinates into a single real-valued vector. If the landmarks have coordinates (x_i, y_i, z_i), the concatenated vector will be of the form $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots)^{\mathrm{T}}$. The n-vector $\bar{\mathbf{x}}$ is the mean of the aligned (by Procrustes analysis) shapes in the training set; the $n \times m$ matrix Φ is the eigenvector matrix. Controlling the retained variability (or number m of retained eigenvectors) controls the level of allowed shape deformation. If the retained variability (usually expressed as percentage of total variability) is low, the model will only provide for the most frequently appearing and strong shape deformations, as learnt from the training set.

To match the model to an image, local image appearance is learned from the same training set of image–shape pairs. For each image and every landmark of the shape, we sample the gradient amplitude along the shape normal at 11 equidistant positions 0.5 mm apart (voxel size). These samples are concatenated into a vector $\mathbf{g}_{ik}$, which we will call *profile*, providing a profile for ith landmark and kth image. The number of sampling positions is motivated by the analysis performed in Sukno et al. (2007). We also performed a test, varying the number of samples from 7 to 23, and the best accuracy was obtained for 11 samples. Longer profiles did not seem to achieve a significant improvement; however, they increase computational load.

To cope with variation in intensities, each profile is normalised by the sum of absolute values of its elements as in (2).

$$\tilde{\mathbf{g}}_{ik} = \mathbf{g}_{ik} \Big/ \sum_j |g_{ik}(j)|. \tag{2}$$

After normalisation, for each landmark and all images, the mean profile $\bar{\tilde{\mathbf{g}}}_i$ and covariance matrix $\mathbf{C}_i$ are estimated. The collection of such pairs for each landmark i constitutes an *intensity model*. During matching, each landmark i of the current shape estimate is displaced along the corresponding shape normal as to minimise the Mahalanobis distance (3) between the sampled $\tilde{\mathbf{g}}_i^{\mathrm{s}}$ and mean $\bar{\tilde{\mathbf{g}}}_i$ profile.

$$d_i\left(\tilde{\mathbf{g}}_i^{\mathrm{s}}, \bar{\tilde{\mathbf{g}}}_i\right) = \left(\tilde{\mathbf{g}}_i^{\mathrm{s}} - \bar{\tilde{\mathbf{g}}}_i\right)^{\mathrm{T}} \mathbf{C}_i^{-1} \left(\tilde{\mathbf{g}}_i^{\mathrm{s}} - \bar{\tilde{\mathbf{g}}}_i\right). \tag{3}$$

Normally, the best position for each landmark is found from a set of N_{c} candidate positions: the profile is sampled at N_{c} equidistant locations along the normal and the candidate corresponding to the smallest Mahalanobis distance is chosen.

After displacing all the landmarks, the resulting shape is constrained such that its i-th PCA parameter $\mathbf{b}_i$ belongs to the interval determined by the corresponding eigenvalues: $\left[-\beta\sqrt{\lambda_i}, \beta\sqrt{\lambda_i}\right]$. The value of β is typically set to 3 or established experimentally.

The PDM used in our experiments was constructed from a set of high resolution CT images as in Ordas, Oubel et al. (2007). The total PDM training set consisted of 100 individuals (age 58 ± 8 years, 56% men) for which the multi-slice computed tomography (MSCT) was performed for non-invasive evaluation of the coronary arteries. The population included 60 subjects who had normal coronary arteries, 20 patients with significant coronary artery

disease (CAD) without a history of previous myocardial infarction (MI) and 20 patients with CAD and a history of MI. Each scan includes 15 temporal phases with a voxel dimension of $0.4 \times 0.4 \times 2.0$ mm. Imaging was performed with a 64-row detector Toshiba Aquilion 64 system (Toshiba America Medical Systems, Tustin, CA, USA), operating at Clinica Creu Blanca (Barcelona, Spain). Between 80 and 100 ml of contrast material (Xenetic 350) at an injection rate of 5 ml/s was used. Rotation time ranged from 400 to 500 ms depending on heart rate, and tube voltage was 120 kV at 400–430 mA. Only one PDM was constructed from the shapes corresponding to all the temporal phases of all the subjects (The modes of variation of the model can be seen in Figure 1). The matching of the model was performed using 95% of retained variability corresponding to 71 mode, $N_c = 7$ candidates for landmark displacement and the regularisation parameter β equal to 3. The convergence was declared after 10 iterations in two resolutions. Finally, in some images a part of the myocardium was outside the field of view. No special provisions have been made in this case. Outside the field of view the voxels have zero intensity so the sampled profiles will also be zero everywhere, providing no information for finding the best position for any landmark that falls in that area of the image. The Mahalanobis distance will be the same for any candidate position in this case and such landmarks, therefore, will take their positions in the image based on the shape prior only.

3. Evaluation data-sets

3.1 Synthetic training and testing sets

To find the appearance of echocardiographic images we shall use two approaches:

- FastGen (Appendix 1), which assumes a uniform PSF, does not take into account the geometry of the transducer (assuming a 3D analogue of a linear probe) and is very fast.
- Field II (Appendix 2), which has a more realistic model of sound propagation, requires the complete specification of the probe (a 2D matrix phased array transducer) and is slow.

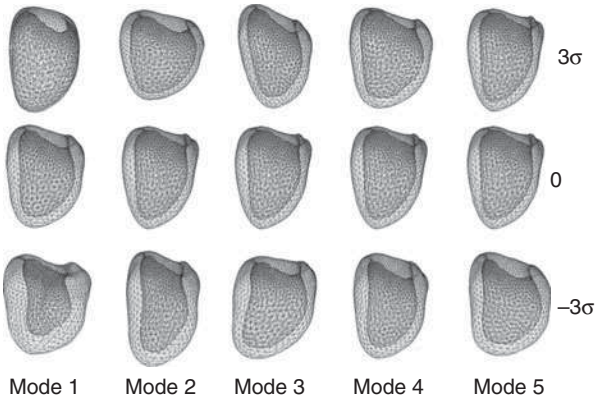

Figure 1. The first five modes of variation of the PDM evaluated at ±3 SDs.

3.1.1 The training sets

The FastGen generated training set for the intensity model consisted of 270 volumes. The following parameters were chosen to introduce variability in both shape geometry and speckle appearance pattern. Ten shapes corresponding to the random variations (within $\pm 1.5\sqrt{\lambda}$) of the first five principal components (PCs) of the PDM were generated. The standard deviation (SD) of Gaussian envelope over the sine wave in axial direction σ_z was taken equal to 0.30, 0.50 and 0.70 mm. The SDs of Gaussian envelope over sine wave in lateral and elevation directions (σ_x and σ_y) were taken equal to 0.50, 0.75, 1.00 mm. The SD of the Gaussian noise σ_n was taken equal to 0.75, 1.00 and 1.25. In order for the training set to have only noise pattern variation, the intensities were kept constant. Their values have been chosen empirically to have good visual contrast between tissues and are equal to 270 for the myocardium, 70 for the background and 60 for the blood pool. Due to the profile normalisation, the model trained on these data should be able to deal with images that have different tissue contrast.

3.1.2 The testing sets

The testing was performed on both synthetic and real data. The synthetic testing set was generated with an idea of providing images of different tissue contrast. It consisted of 260 volumes. To generate the synthetic images, 20 shapes corresponding to the random variations (within $\pm 1.5\sqrt{\lambda}$) of the first five PCs of the PDM were obtained. Noise variances have been chosen empirically to reproduce the real images as much as possible. They are: $\sigma_z = 0.30$ mm for axial direction, $\sigma_x = \sigma_y = 0.70$ mm for lateral and elevation and $\sigma_n = 1.20$ mm for the Gaussian noise. Intensities $t(x, y, z)$ (from 0 to 255) are $60, 75, \ldots, 255$ for myocardium; 50 for background (40 for the blood pool). So the myocardium contrast (difference in intensities between the myocardium and the background) varies from 10 to 205 in steps of 15. Two sample images can be seen in Figures 2 and 3.

The generation of the training and testing sets with Field II followed similar guidelines, taking into account its specifics. Since there were more parameters to set, the training set was slightly larger and consisted of 360 images, while the testing again contained 260 images. All the parameters are summarised in Tables 1 and 2.

The initial shape for the segmentation of synthetic images was the mean shape of the PDM (consisting of 2677

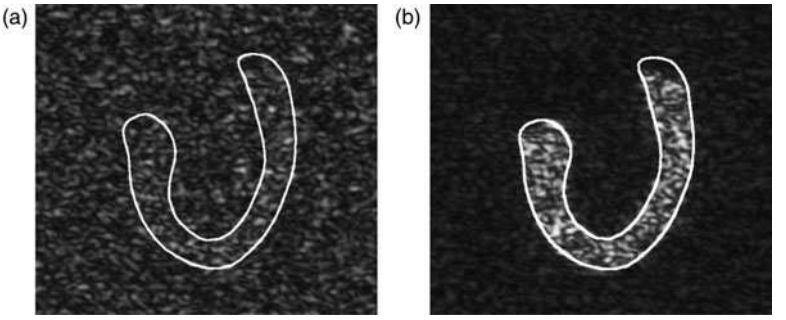

Figure 2. Sample images, created by FastGen, corresponding to the low (a) and high (b) intensity differences with superimposed shapes, which were used to generate these images.

points: 880 endocardium and 1797 epicardium) aligned to the ground-truth shape by a similarity transform.

3.2 In vivo *training and testing sets*

The *in vivo* set consisted of manually landmarked end-systole (ES) and end-diastole (ED) LV 3D volumes of 20 patients with left ventricular asynchrony (3 dilated, 5 with acute MI, 1 hypertrophic, 11 asymptomatic or ischaemic, acquired at Hospital Clinico San Carlos, Madrid and Centro Cardiovascular Sant Jordi, Barcelona). These patients are unrelated to those used for the PDM construction. The data were acquired using a Philips IE33 echograph (Philips Ultrasound, Inc., Andover, MA, USA) with X3-1 transducer and exported into accessible format using Philips QLAB v6.0 quantification software. The exported data are envelope-detected $224 \times 208 \times 208$ images with voxel size varying from $0.6 \times 0.6 \times 0.7$ mm to $1.0 \times 1.0 \times 0.7$ mm. Landmarking has been performed by a trained specialist. No inter- and intra-observer studies have been performed, but we adapted those reported in recent studies. Manual landmarking was performed using GIMIAS (Larrabide et al. 2009). ED and ES were detected visually by monitoring the aortic valve. Sample images can be seen in Figure 3.

Since the data-sets used to construct the PDM and those used for testing correspond to different patients, a natural question arises: whether the PDM is able to represent the shapes of the *in vivo* testing set. Figure 4 shows how the reconstruction error changes with the retained variance. Since it is typical to retain 95%, the corresponding error introduced by the PDM is 0.28 ± 0.06 mm, which is reasonably small and smaller than the voxel size.

Due to the small quantity of data, the experiments that involved *in vivo* training set used a leave-one-out cross

Table 1. Model parameters for FastGen.

Parameter	Range
Training	
σ_z (mm)	0.30, 0.50, 0.70
σ_x, σ_y (mm)	0.50, 0.75, 1.00
σ_n	0.75, 1.00, 1.25
Myocardium intensity	250
Blood pool intensity	60
Background intensity	70
Testing	
σ_z (mm)	0.30
σ_x, σ_y (mm)	0.70
σ_n	0.70
Myocardium intensity	60, 75, 90, … , 255
Blood pool intensity	40
Background intensity	50

Table 2. Model parameters for Field II.

Parameter	Range
General[a]	
Probe size (mm)	20×15
Field of view (°)	120
Speed of sound (m/s)	1540
Centre frequency (MHz)	3
Sampling frequency (MHz)	100
Focus depth (mm)	70
Probe matrix dimensions (mm)	15×20
Depth of cardiac apex (mm)	20
Piezo element matrix (elements)	32×64
Training	
Number of scan lines	40×40, 60×60, 80×80
Active elements	128, 256, 512
Number of scatterers ($\times 10^3$)	500, 1000, 1500, 2000
Myocardium intensity	250
Blood pool intensity	60
Background intensity	70
Testing	
Number of scan lines	80×80
Active elements	512
Number of scatterers ($\times 10^3$)	2000
Myocardium intensity	60, 75, 90, … , 255
Blood pool intensity	40
Background intensity	50

[a] Dimensions of each element were computed as the physical size of the probe divided by the number of elements in rows and columns. The distance between elements was computed as dimensions of an element divided by 1000. Most of the parameters are the same as in the examples coming with Field II.

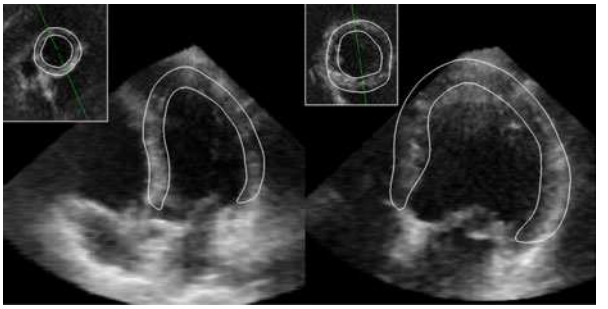

Figure 3. Real US images (from two patients) from our testing set with superimposed manual delineations.

validation (all the images of the same patient had been excluded).

The initial shape for the segmentation of *in vivo* images was the mean shape aligned by an affine transform to six points: four points on the endocardium in the basal plane, the centre of the aortic valve and the apex, as illustrated in Figure 5.

4. Experiments

4.1 Validation on synthetic data

To start, we wanted to verify that ASM trained on the synthetic data can successfully segment the synthetic data of different contrast produced by the same algorithm. For this purpose, FastGen and Field II-generated testing sets have been segmented by the ASM trained on the corresponding training sets. The results are shown in Figure 6. The segmentation algorithm is not completely invariant to the image contrast variation, as it could be expected due to intensity normalisation, and the images with higher tissue contrast have higher segmentation accuracy. Nevertheless, the ASM trained on Field II set appears to have slightly inferior performance on the low-contrast data.

The symmetrical point-to-surface error is defined for two shapes (surface meshes) s_1 and s_2 as:

$$\varepsilon(s_1, s_2) = \left[d(s_1, s_2) + d(s_2, s_1)\right]/2, \tag{4}$$

where

$$d(s_1, s_2) = \frac{1}{N}\sum_{i=1}^{N}\left\| s_1(i) - \underset{p\in s_2}{\operatorname{argmin}} \|s_1(i) - p\|^2 \right\| \tag{5}$$

is an asymmetric point-to-surface error, which is the mean distance between each vertex of the mesh s_1 and mesh s_2 ($s_1(i)$ refers to the ith vertex of the mesh s_1); $\|\cdot\|$ is the l^2 norm.

4.2 Evaluation on real images

In this section, we compare the ASM trained on real images and artificially generated images and see whether there is any benefit of using the latter. In the case, when real images had been used for training, a leave-one-out strategy was employed (training on 38 images, corresponding to 19 patients, and segmenting the ED and ES of the remaining patient). The segmentation accuracy is shown in Figure 7. The results of the volume and ejection fraction (EF) estimation by the best approach (the ASM trained on FastGen data) are summarised in Table 3. The accuracy was measured with respect to manually delineated contours. As one can note from the confidence intervals (CIs), the difference between the different training sets is not statistically significant (with 95% confidence level) although there seems to be an improvement when FastGen is used. There are of course several issues that are worth commenting, assuming that all the approaches perform equally well. It would be expected for Field II to be superior to FastGen due to being a more realistic model. Nevertheless, Field II has many

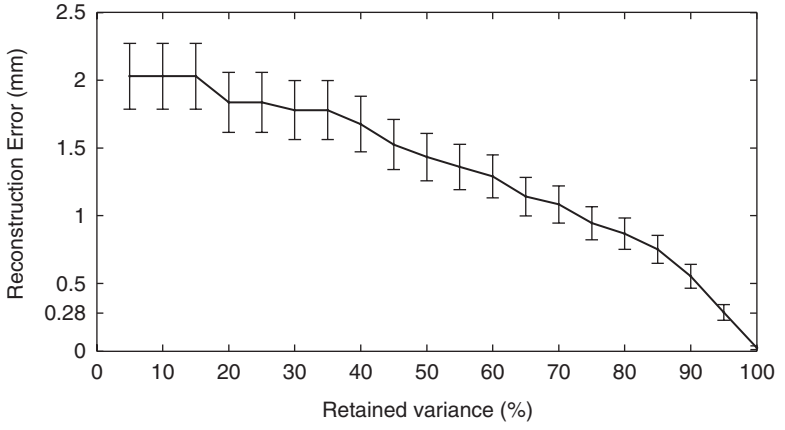

Figure 4. Average error, with 95% CIs, of reconstructing the manual shapes corresponding to the *in vivo* data using the PDM. The error corresponding to 95% retained variability is 0.28 ± 0.06 mm.

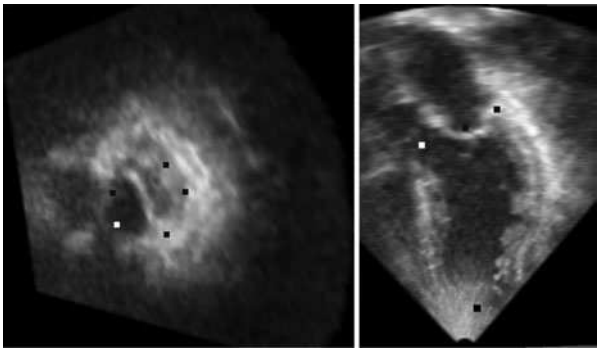

Figure 5. Short axis and long axis views of the points used for model initialisation superimposed on an image. The white point corresponds to the aortic valve and provides an orientation cue.

more parameters to tune and generating a representative training set would require sweeping over all of them (when the parameters of the equipment are unknown, which is the usual case). The latter would lead to an enormous database, but Field II in 3D is very computationally demanding. It takes about half an hour to generate one 3D volume on a 60-processor cluster fully dedicated to the task. Theoretically this would imply 1 week for a set of 280 volumes. Using real data for ASM training has its own downside in that it requires a lot of manual delineating in 3D, which is very time consuming. It is also very difficult to produce consistent delineations in every image of the sequence (like images corresponding to ED and ES) mostly due to the noisiness and rather poor quality of the images. Using FastGen with ASM eliminates the computational cost problem and the necessity of having manual delineations.

The Bland–Altman and scatter plots for the best approach (FastGen) are shown in Figures 8 and 9. The scatter plot suggests a nice linear relationship with high correlation between the measurements and ground truths. The slope of the fitted line is smaller than 1.0 and intercept is greater than 0.0 as is expected when there is no

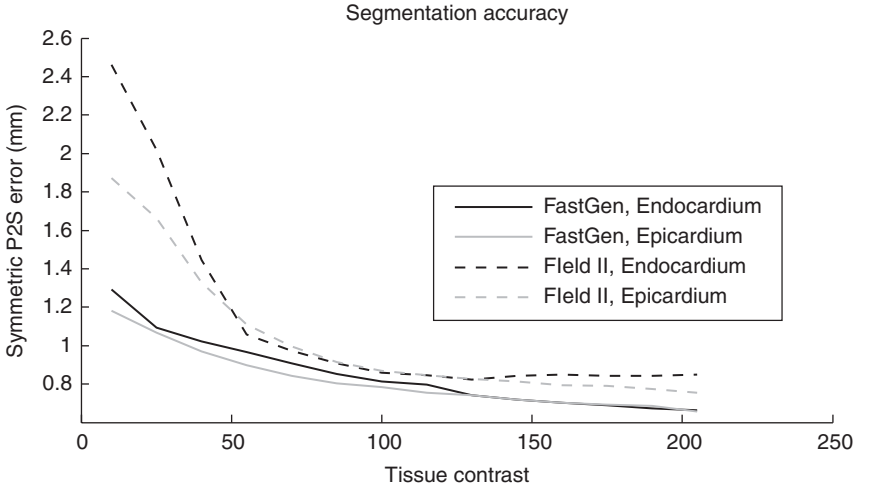

Figure 6. Accuracy of segmenting the simulated images with varying tissue contrast. The contrast is expressed as difference of scatterer amplitudes. The symmetric P2S errors for endocardium and epicardium are shown.

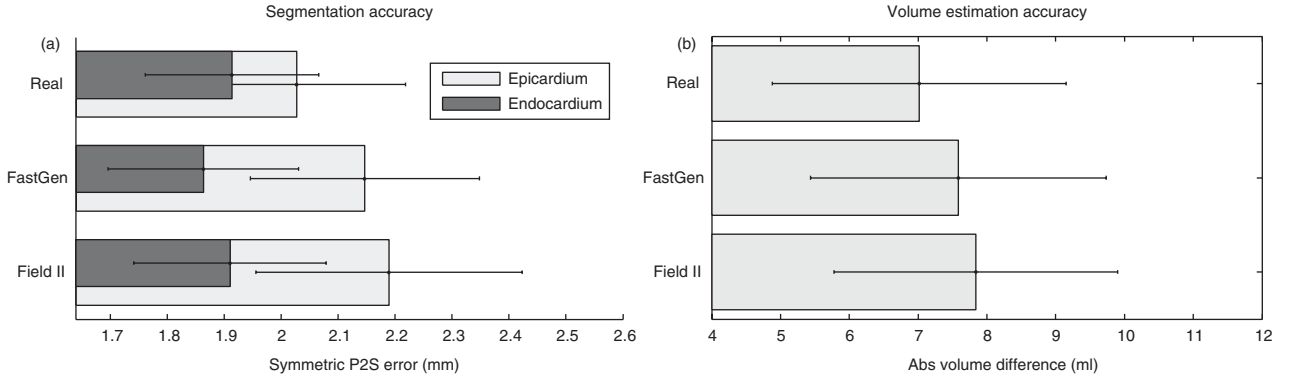

Figure 7. Accuracy of segmenting the real images by an ASM trained on different training sets in terms of (a) the mean symmetric point-to-surface error for endocardium and epicardium and (b) mean absolute volume difference for LV cavity. Error bars represent 95% CI of the mean.

Table 3. Evaluation results on the real data-sets by an ASM trained on FastGen data.

	LV volume (ml)	LV EF (%)
Mean difference	2.66 ± 2.19	−5.84 ± 2.93 (Relative: −12.6)
Limits of agreement	±20.24 (±5.54)	±18.55 (±2.59)
RMSE	10.54	10.92
	P2S error (mm)	
Epicardium	2.15 ± 0.20	
Endocardium	1.87 ± 0.16	

Note: Together with the measurements 95% CIs are shown.

relationship between the error of the volume estimation and the magnitude of the volume (Bland and Altman 2003). Some segmentation examples can be seen in Figure 10. What is interesting to note is that in some cases the LV did not fit into the field of view, which is normal for dilated hearts. In this case, the model simply could not find an edge to move to and the final shape was the result of matching the landmarks within the field of view and applying the shape model constraints.

An interesting question that can be answered using the image generation is how much data in the training set are actually needed for accurate segmentation. The size of the training sets in the experiments has been chosen from practical considerations, trying to make the Field II training set in a reasonable time. But since FastGen is really fast, we can use it to increase the size of the training set and see how it affects the segmentation accuracy. To generate larger training set with FastGen, we used the same parameters as before but with more shapes (40 different shapes in total). The biggest training set consisted of 1080 image–shape pairs. A total of 36 training sets have been generated: starting from four random shapes, all variations of all the parameters (27 images per shape), and randomly adding one shape per training set.

The results can be seen in Figure 11. As expected, the error on the artificial testing set is decreasing with the size of the training set and stabilises around 1000, but with acceptable results already around 500. Of course these numbers are bound to the training and testing sets used and after 1000 images the training fails to offer more information to the intensity model. On the other hand, the training set size does not seem to have much effect on real data, although very wide CI hampers drawing any conclusions from the plot. The possible reason could lie in the combination of insufficient image quality (which in general results in high interobserver variability and our results are not far from it) and simplified US model of FastGen. Nevertheless, although there is no evidence to support the use of more training data, the computational simplicity of the image generation algorithm allows to easily obtain large quantity of training data and, therefore, the segmentation algorithm can easily benefit from it.

Finally we would like to address the problem of initialisation. As was mentioned, we use six points. These six points allow to get a reasonably good initial guess for the shape position and orientation. Although the positions for these points are easily identifiable (and quick to place), we wanted to see how the accuracy of the algorithm will change if there is an initialisation error. To evaluate this effect, we randomly perturbed the initial shape by random translation, scaling and rotation. To make sure that a wide range of perturbations is covered, we picked 20 random perturbations for each of the following parameter ranges:

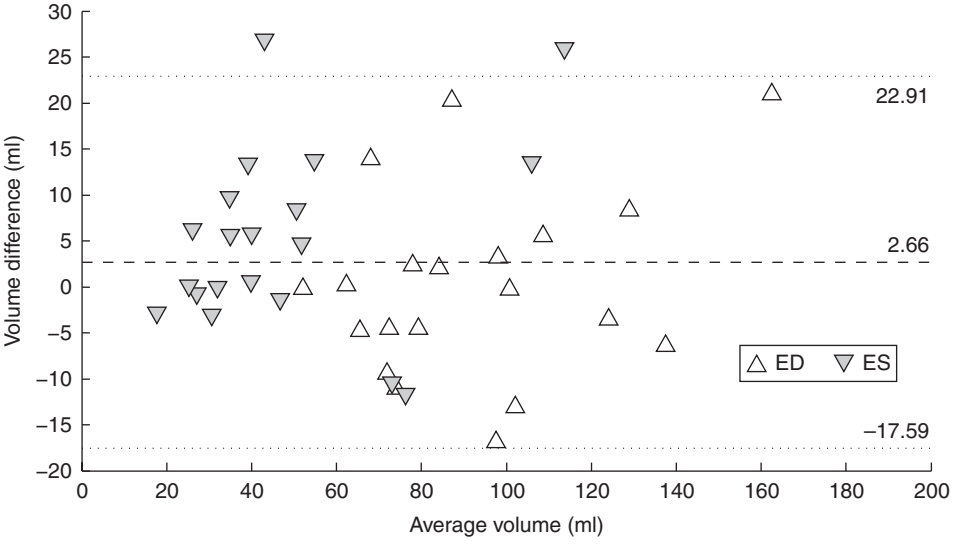

Figure 8. Bland–Altman plot of the volume estimation accuracy on the real images by an ASM trained on FastGen training set.

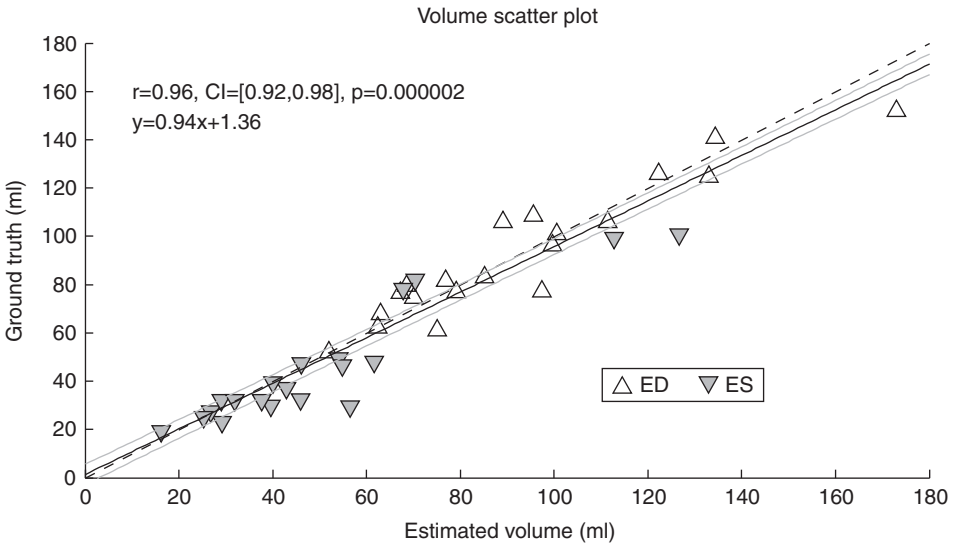

Figure 9. Comparison of volumes estimated by the proposed algorithm and ground truths. The solid black line is the fitted one, dashed line is the equality line, grey lines represent a 95% CI of the fit. The CI is 95% of the correlation coefficient.

- Translation along every axis of up to ±1mm
- Rotation about all the axes of up to ±3°, translation along every axis of up to ±3mm, scaling by a factor up to 0.1
- Rotation about all the axes of up to ±6°, translation along every axis of up to ±6mm, scaling by a factor up to 0.2
- Rotation about all the axes of up to ±9°, translation along every axis of up to ±9mm, scaling by a factor up to 0.3
- Rotation about all the axes of up to ±12°, translation along every axis of up to ±12mm, scaling by a factor up to 0.4

After the perturbation, the shape has been used as the initial shape for the proposed framework. Figure 12 shows how the segmentation accuracy changes with the surface-to-surface distance between the initial shape and the ground truth (the distances were averaged over the whole set for every displacement). For convenience, the ninth degree polynomial curves had been fit to the point clouds. As can be seen, the method is relatively insensitive to small displacements, but as they grow bigger, the accuracy drops and the model converges to local minima. It can also be noted that the algorithm showed higher robustness on the synthetic data-sets, but this is mostly because the images have better definition of myocardium and there are no other structures or artefacts present in the image.

It is natural then to start looking for the ways of improving the robustness. When dealing with segmentation of only LV in 3D echocardiographic images, placing the initial shape requires not only placing it in the neighbourhood of the correct contours (preferably) but

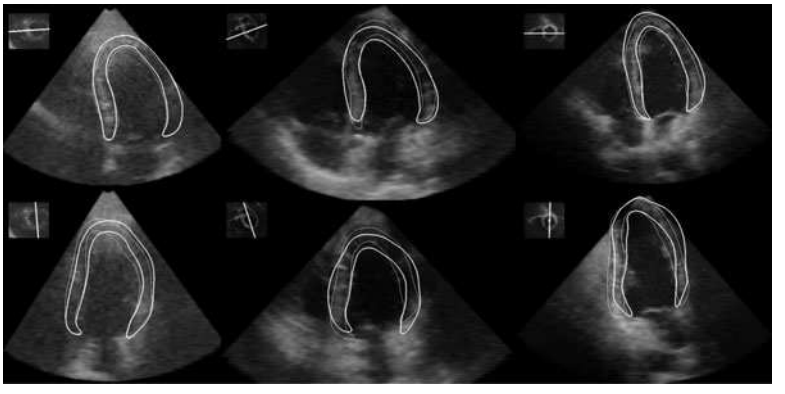

Figure 10. Segmentation examples. Three hearts, one per column, two perpendicular views per heart are shown. The cut planes chosen are shown in the thumbnail in the upper left corner and are aligned with the long axis. Thin line represents manual delineation while thick – automatic one, by the ASM trained on FastGen data.

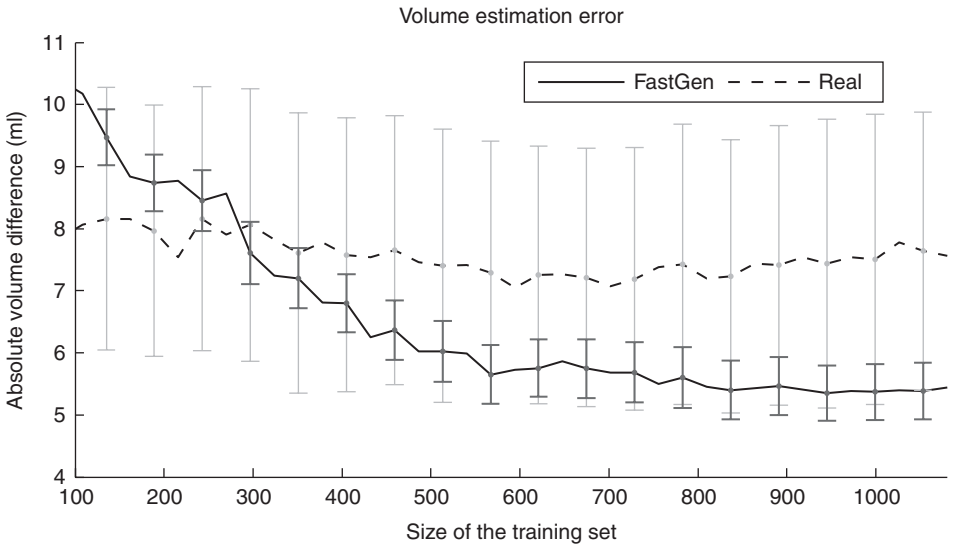

Figure 11. Segmentation accuracy measured in terms of mean absolute volume difference for different sizes of the FastGen training set evaluated on the real and FastGen generated test data. Error bars correspond to 95% CI of the mean.

also finding correct orientation (so that anatomical regions can be accurately determined from the shape without using the corresponding image). The latter, without considering more structures than the LV, is in a way ambiguous, as the only indicator of the orientation within the LV is the aorta (there are also papillary muscles, but they are usually not clearly identifiable). Therefore, if the automatic initialisation has to be attempted, the algorithm has to identify more structures than just the LV. One of the ideas could be to use the registration of any given 3D echocardiographic image to some template image (as in Zagrodsky et al. (2005)). But registration usually would take more time than the segmentation itself. There has also been very little research on the accuracy of registering 3D echocardiographic images corresponding to different patients. Nevertheless, it would be highly desirable to automate the process. Currently, a possible option of reducing the sensitivity to initialisation could be to use several initialisations as in Mitchell et al. (2002) at the cost of significant performance decrease.

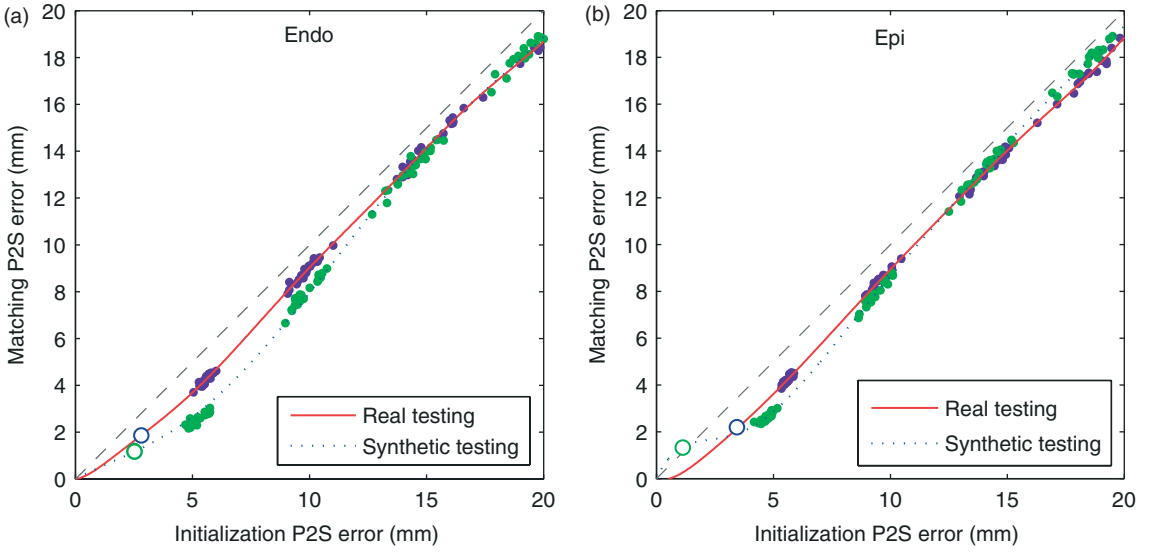

Figure 12. (Colour online) Sensitivity of initialisation for (a) endocardium and (b) epicardium matching evaluated on real and synthetic data. The dashed line corresponds to equal errors. The green and blue circles correspond to the specific segmentation outcomes (averaged over the whole set) on synthetic and real data, respectively. The dotted and solid curves are polynomials, fitted to the circles. The hollow circles show the position of the initialisation used in the experiments.

4.3 Comparison with other methods

To put our work in the context of other contemporary papers, Table 4 presents the results of several techniques recently reported in the literature. This table shows the mean with 95% limits of agreement (mean $\pm t \cdot$ SD, where t is taken from the Student's t table, whenever authors of these papers reported the accuracy as mean $\pm$ SD, we multiplied SD by the corresponding t). The value in the column *Volume* means that the authors did not classify volumes into end-diastolic volume (EDV)/end-systolic volume (ESV). An exception is the work by Angelini, Homma et al. (2005), who provided both estimated and ground truth volumes, which allowed us to compute the values we needed. In the latter case, in the column *Volume* we combined the results for ESV and EDV (the statistics were computed from both sets of volumes) in order to compare the result with other papers. It can be seen from the table that our results are consistent with those of the state-of-the-art algorithms (considering accuracy of EF estimation) and better in estimating the volumes themselves (on average). Nevertheless, in terms of EF, the results are still much worse than those corresponding to the intra- and interobserver variability reported in other studies. On the other hand, the accuracy of volume estimation, reported by other clinical studies, is also on the same level (Hung et al. 2007). For instance, Jenkins et al. (2004) reported the scatter of EDV at the level of $\pm$29 ml and ESV at the level of $\pm$18 ml for 50 patients (although the measurements were compared with those in MR and the error could accumulate) nevertheless, the EF was significantly smaller – 0 ± 7%, which indicates consistency in inaccuracies between ED and ES. In our case, the scatter of EF was rather large and the consistency has to be improved.

Let us go over all the approaches presented in Table 4 and start with the constrained AAMM fused with active contours and unconstrained AAMM of Hansegård, Urheim et al. (2007). The major disadvantage of the approach is that AAMM represents both the shape deformation at a certain time instant and the motion pattern. In order to build a representative model, the training set should have not only all the possible cardiac shapes but all the possible deformation patterns as well. This calls for a large training set with manual delineations in every temporal phase, which is difficult to obtain. An automatic generation of such a training set using an US image generator combined with realistic LV deformations, extracted from MDCT data, or using a biomechanical model as in Duan et al. (2007), would really benefit the approach. As an additional complication, combining AAMM with active contours requires tuning the parameters of each algorithm plus a coupling weight. On the positive side, AAMM allows for realistic modelling of both shape and cardiac deformation (and texture, of course, if needed).

The level-set-based algorithm of Angelini, Homma et al. (2005) segments the ventricle by partitioning the image into homogeneous regions instead of using the advection term with gradient information that is commonly encountered in the level-set methods. The approach depends on spatio-temporal brushlet denoising, which adds additional parameters to tune to the level-set-specific parameters. Since there is no prior, leaking is possible and the resulting boundary lacks smoothness. The advantage is that the algorithm does not require any training set and can be initialised anywhere in the image (the algorithm tracks only one closed blob centred around the centre of mass of the object of interest).

The wire mesh of Zagrodsky et al. (2005) employs 3D Sobel edge detector, which relies on median filtering. Although these two filters do not require any significant tuning, the median filter is probably not the best choice for US images, and many spurious edges might still remain on the image. The algorithm requires balancing internal and external forces and the execution speed is very low: 3 min for registration, 8 min for segmentation on a dual 1.7 GHz Pentium. On the other hand, the approach does not require a training set and the initialisation is automated through registration.

Finally, a modified Malladi–Sethian equation to avoid leaking of Corsi et al. (2002) is rather a typical level-set segmentation. It is automatic and no training set is required. It does not use any shape prior and leaking is prevented essentially by removing the advection term.

What we have proposed is an approach that uses explicit shape representation with automatically constructed models of the shape and local appearance. If we compare it with the above approaches the shape can be fairly complex; user interaction is reduced only to model initialisation during segmentation and tuning of ASM parameters (which can usually be left unchanged). For example, introducing epicardium into segmentation pipeline is much easier than in the above approaches, except for the AAMM (it would only require more work on manual delineations). It is done simply by merging meshes of both structures.

Another advantage of our approach lies in using multi-modal data. The shape model, automatically created from a large amount of MDCT images, provides an accurate shape model as opposed to the one that might have been constructed from US images, where the boundaries are poorly defined and have to be guessed. The generation of artificial US images, on the other hand, avoids having incorrect LV delineations in real data.

All the above approaches have been proposed for the envelope-detected US signal. Nevertheless, sometimes it is possible to obtain the original, raw, radio-frequency (RF) signal, which is potentially richer and contains more information. There were several publications reporting segmentation using this signal representation in 2D (Dydenko et al. 2003; Bernard et al. 2007; Yan et al. 2007) and 3D (Nillesen et al. 2009); however, the access to such signal is generally restricted due to the competition

Table 4. Summary of LV endocardium segmentation algorithms. The errors are given as mean $\pm t \cdot SD$ with the 95% CIs of the mean in the parentheses. t was chosen according to the number of images to provide 95% limits of agreement.

R eferences	Algorithm	Year	Modality	Dataset	Equipment	Gold standard	EDV (ml)	ESV (ml)	EF (%)	Vol. (ml)
	Our approach, automatically constructed ASM.		RT3D	20 CRT patients	Philips 1 E33 echograph with X3-1 transducer	Manually traced contours in RT 3D	0.1 ± 21.1 (± 4.7)	5.2 ± 21.1 (± 4.7)	− 5.8 ± 18.5 (± 2.9)	2.6 ± 20.2
Hansegård, Urheim et al. (2007)	DP-CAAMM, constrained AA M M fused with active contours	2007	Triplane US	36 adults (old myocardial infarction, valve disease, DCM, pulmonary hypertension, heart transplant, structurally normal hearts)	GE Vivid 7 with matrix-phased array transducer 3V	Manual landmarks in r RT 3D	− 3.1 ± 40.6 (± 6.8)	0.61 ± 26.4 (± 4.4)	− 1.3 ± 12.8 (± 2.1)	n/a
Hansegård, Urheim et al. (2007)	Unconstrained AAMM	2007	Triplane US				− 7.3 ± 40.6 (± 6.8)	− 2.5 ± 44.7 (± 7.4)	− 1.5 ± 22.3 (± 3.7)	n/a
Angelini, Homma et al. (2005)	Homogeneity- based active contour	2005	RT 3D	10 patients with pulmonary hypertension: 8 with primary PH, 2 secondary PH associated with congenital defects		Manually traced contours in RT 3D	16.1 ± 57.8 (± 18.3)	6.6 ± 39.7 (± 12.5)	0.6 ± 25.6 (± 8.1)	11.4 ± 45.8
Zagrodsky et al. (2005)	Wiremesh	2005	RT3D	10 patients (healthy and diseased)	Volumetries and Philips SON OS 7500	Manually traced contours in RT 3D	− 0.1 ± 49.3 (± 15.6)	− 4.2 ± 32.3 (± 10.2)	2.6 ± 21.1 (± 6.7)	− 2.1 ± 37.8
Corsi et al. (2002)	Modified Malladi-Sethian equation	2002	RT3D	20 patients	Volumetries	Manual volume measurements in M RIs	n/a	n/a	n/a	− 15.6 ± 41.1

			Inter observer variability				Inter observer variability		
References	Algorithm	Year	EDV	ESV	EF		EDV	ESV	EF
Hansegård, Urheim et al. (2007)		2007	13.0 ± 38.6 ml	9.9 ± 30.5 ml	− 1.7 ± 12.8%		N/a	N/a	N/a
Sugeng et al. (2006)		2006	11.2 ± 17.6%	14.2 ± 24.1%	10.5 ± 17.0%		3.9 ± 4.0%	5.6 ± 8.0%	5.6 ± 6.9%
Sugeng et al. (2006)	Adapted to our data		10.3 ± 15.6 ml	13.1 ± 21.4 ml	5.5 ± 8.5%		3.6 ± 3.6 ml	2.5 ± 3.4 ml	2.9 ± 3.5%

between different hardware manufacturers. In principle, the proposed approach could be extended to learn the statistics of RF signal in the neighbourhood of boundaries, but, in this case, the knowledge of exact specifications of the imaging system may be necessary.

5. Conclusions

In this paper, we proposed an approach to automatically construct an ASM for 3DUS segmentation. In contrast to the majority of 3DUS segmentation algorithms, our model includes both epi- and endocardium, providing simultaneous segmentation of a complete LV. The approach is based on the combination of automatically constructed shape model from MDCT and local appearance learned from data-sets automatically generated using two commonly accepted models of US. One of them, FastGen, is fast and assumes uniform PSF, while the other, Field II, is a more realistic, but linear, sound propagation model. It has been shown that although using synthetic images to train an ASM demonstrates a similar segmentation accuracy as training ASM from manually delineated images, the former are much faster to obtain if FastGen is used. It is also much easier to control the image quality of the generated data, while with real data the quality depends a lot on the patient and many of them do not have a good acquisition window.

Although the results of all the approaches are comparable, probably because of the simplified appearance model used by the ASM, the best segmentation results were obtained by FastGen, resulting in the volume estimation accuracy with limits of agreement across the whole population of 2.6 ± 20.2 ml. The average point-to-surface segmentation errors for epicardium and endocardium were 2.15 ± 0.20 and 1.87 ± 0.16 mm, respectively.

The synthetic 3DUS images have two clear benefits over the real ones: high quality and detail and an already solved correspondence problem between a shape and an image. The latter avoids any manual landmarking, which is error prone, complicated in 3D and shows high interobserver variability (approximately 10.3 ± 15.6 ml for EDV and 13.1 ± 21.4 ml for ESV) (Sugeng et al. 2006). On the other hand, it is very difficult to obtain a large set of good quality real 3DUS images, while by generating them artificially we can obtain a set of any size. The choice of the US propagation model is dependent on the segmentation algorithm. In the case of the classical linear ASM, the advanced model of Field II does not seem to improve the intensity model and FastGen can be beneficial in terms of time, computational resources and implementation effort.

The high computational cost of Field II and lack of accurate information about the employed US probe did not allow us to fully investigate the construction of an optimal synthetic training set. Still, the biggest drawback of the proposed methodology is its lack of consistency when segmenting the data of the same patient (as can be observed from EF errors), and it would benefit from borrowing some ideas from the tracking algorithms such as adaptation to the observed data. Automating the initialisation would also be convenient. Currently we have seen automatic initialisation based on registration and Hough transform. Registration, though, takes too much time and is unreliable in the US. Hough transform, on the other hand, combined with filtering and edge detection could be a viable approach as shown by van Stralen et al. (2007).

Acknowledgements

We would like to thank J.A. Jensen, from the Technical University of Denmark, for providing us with a 64-bit version of Field II. We are also grateful to various clinicians who contributed data to this study: F. Carreras and R. Leta of Clinica Creu Blanca (Barcelona); J.L. Zamorano and L. Perez-de-Isla from Hospital Clinico San Carlos (Madrid); M. Sitges from Hospital Clinic i Provincial de Barcelona (Barcelona); G. Avegliano and M. Petit from Centro Cardiovascular Sant Jordi (Barcelona).

Funding

This work was partially supported by MEC TEC2006-03617, CDTI CENIT-CDTEAM and TIN2009-14536-C02-01 grants.

References

Angelini ED, Homma S, Pearson G, Holmes JW, Laine AF. 2005. Segmentation of real-time three-dimensional ultrasound for quantification of ventricular function: a clinical study on right and left ventricles. Ultrasound Med Biol. 31(9):1143–1158.

Angelini ED, Jin Y, Laine AF. 2005. State-of-the-art of levelset methods in segmentation and registration of medical imaging modalities. In: Suri JS, Wilson D, Laxminarayan S, editors. Handbook of biomedical image analysis: registration models. Vol. 3. New York: Springer.

Bamber JC, Dickinson RJ. 1980. Ultrasonic B-scanning: a computer simulation. Phys Med Biol. 25:463–479.

Bellenger NG, Burgess MI, Ray SG, Lahiri A, Coats AJS, Cleland JGF, Pennell DJ. 2000. Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, radionuclide ventriculography and cardiovascular magnetic resonance. Are they interchangeable? Eur Heart J. 21:1387–1396.

Bernard O, Touil B, Gelas A, Prost R, Friboulet D. 2007. Segmentation of myocardial regions in echocardiography using the statistics of the radio-frequency signal. In: Proceedings of international workshop on functional imaging and modeling of the heart; 2007 Jun 7–9; Salt Lake City, UT, USA. Berlin, Heidelberg: Springer; p. 433–442.

Bland JM, Altman DG. 2003. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol. 22(1):85–93.

Bosch JG, Mitchell SC, Lelieveldt BPF, Nijland F, Kamp O, Sonka M, Reiber JHC. 2002. Automatic segmentation of

echocardiographic sequences by active appearance motion models. IEEE Trans Med Imag. 21(11):1374–1383.
Cootes T, Taylor C. 1992. Active shape models – smart snakes. In: Proceedings of British machine vision conference; 1992 Sep 22–24; Leeds, UK. London: Springer; p. 266–275.
Corsaro S, Mikula K, Sarti A, Sgallari F. 2006. Semi-implicit covolume method in 3D image segmentation. SIAM J Sci Comput. 28(6):2248–2265.
Corsi C, Saracino G, Sarti A, Lamberti C. 2002. Left ventricular volume estimation for real-time three-dimensional echocardiography. IEEE Trans Med Imag. 21(9):1202–1208.
Crosby J, Amundsen BH, Hergum T, Remme EW, Langeland S, Torp H. 2009. 3-D speckle tracking for assessment of regional left ventricular function. Ultrasound Med Biol. 35(3):458–471.
Duan Q, Moireau P, Angelini ED, Chapelle D, Laine AF. 2007. Simulation of 3D ultrasound with a realistic electro-mechanical model of the heart. In: Proceedings of international workshop on functional imaging and modeling of the heart; 2007 Jun 7–9; Salt Lake City, UT, USA. Berlin, Heidelberg: Springer; p. 463–473.
Duan Q, Angelini ED, Herz SL, Ingrassia CM, Costa KD, Holmes JW, Homma S, Laine AF. 2009. Region-based endocardium tracking on real-time three-dimensional ultrasound. Ultrasound Med Biol. 35(2):256–265.
Dydenko I, Friboulet D, Gorce JM, D'hooge J, Bijnens B, Magnin IE. 2003. Towards ultrasound cardiac image segmentation based on the radio frequency signal. Med Image Anal. 7(3):353–367.
Frangi AF, Niessen WJ, Viergever MA, Lelieveldt BPF. 2005. A survey of three-dimensional modeling techniques for quantitative functional analysis of cardiac images. In: Landini L, Positano V, Santarelli MF, editors. Advanced image processing in magnetic resonance imaging. Boca Raton, FL: CRC Press.
Gerard O, Billon AC, Rouet JM, Jacob M, Fradkin M, Allouche C. 2002. Efficient model-based quantification of left ventricular function in 3-D echocardiography. IEEE Trans Med Imag. 21(9):1059–1068.
Hansegård J, Orderud F, Rabben SI. 2007. Real-time active shape models for segmentation of 3D cardiac ultrasound. In: Proceedings of international conference on computer analysis of images and patterns, LNCS. Vol. 4673; 2007 Aug 27–29; Vienna, Austria. Berlin, Heidelberg: Springer; p. 157–164.
Hansegård J, Urheim S, Lunde K, Rabben SI. 2007. Constrained active appearance models for segmentation of triplane echocardiograms. IEEE Trans Med Imag. 26(10):1391–1400.
Hong W, Georgescu B, Zhou XS, Krishnan S, Ma Y, Comaniciu D. 2006. Database-guided simultaneous multi-slice 3D segmentation for volumetric data. In: Proceedings of European conference on computer vision, LNCS. Vol. 3954; 2006 May 7–13; Graz, Austria. Berlin, Heidelberg: Springer; p. 397–409.
Horn BKP. 1987. Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am Optic Image Sci Vis. 4:629–642.
Hung J, Lang R, Flachskampf F, Shernan SK, McCulloch ML, Adams DB, Thomas J, Vannan M, Ryan T, EAS. 2007. 3D echocardiography: a review of the current status and future directions. J Am Soc Echocardiogr. 20(3):213–233.
Jenkins C, Bricknell K, Hanekom L, Marwick TH. 2004. Reproducibility and accuracy of echocardiographic measurements of left ventricular parameters using real-time three-dimensional echocardiography. J Am Coll Cardiol. 44(4):878–886.
Jensen JA. 1996. Field: a program for simulating ultrasound systems. Med Biol Eng Comput. 34(Supplement 1), Part 1:351–353.
Jensen JA, Svendsen NB. 1992. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. IEEE Trans Ultrason Ferroelectr Freq Control. 39(2):262–267.
Larrabide I, Omedas P, Martelli Y, Planes X, Nieber M, Moya JA, Butakoff C, Sebastián R, Camara O, De Craene M, et al. 2009. GIMIAS: an open source framework for efficient development of research tools and clinical prototypes. In: Proceedings of international workshop on functional imaging and modeling of the heart; 2009 June 3–5; Nice, France. Berlin Heidelberg: Springer; p. 417–426.
Lelieveldt BPF, Frangi AF, Mitchell SC, van Assen HC, Ordas S, Reiber JHC, Sonka M. 2006. 3D active shape and appearance models in medical image analysis. In: Faugeras O, Paragios N, Chen Y, editors. Handbook of mathematical models of computer vision. New York: Springer; p. 471–486.
Mitchell SC, Bosch JG, Lelieveldt BP, van der Geest RJ, Reiber JH, Sonka M. 2002. 3-D active appearance models: segmentation of cardiac MR and ultrasound images. IEEE Trans Med Imag. 21(9):1167–1178.
Montagnat J, Delingette H. 2001. A review of deformable surfaces: topology, geometry and deformation. Image Vis Comput. 19(14):1023–1040.
Nikitin NP, Constantin C, Loh PH, Ghosh J, Lukaschuk EI, Bennett A, Hurren S, Alamgir F, Clark AL, Cleland JG. 2006. New generation 3-dimensional echocardiography for left ventricular volumetric and functional measurements: comparison with cardiac magnetic resonance. Eur J Echocardiogr. 7(5):365–372.
Nillesen MM, Lopata RGP, Huisman HJ, Thijssen JM, Kapusta L, de Korte CL. 2009. 3D Cardiac segmentation using temporal correlation of radio frequency ultrasound data. In: Proceedings of international conference medical image computing and computer assisted intervention, LNCS; 2009 Sep 20–24; London, UK. Berlin, Heidelberg: Springer; p. 927–934.
Noble JA, Boukerroui D. 2006. Ultrasound image segmentation: a survey. IEEE Trans Med Imag. 25(8):987–1010.
Ordas S, Oubel E, Leta R, Carreras F, Frangi AF. 2007. A statistical shape model of the heart and its application to model-based segmentation. In: Proceedings of SPIE. Vol. 6511; 2007 Feb 17; San Diego, CA. SPIE.
Orderud F, Hansegård J, Rabben SI. 2007. Real-time tracking of the left ventricle in 3D echocardiography using a state estimation approach. In: Proceedings of international conference medical image computing and computer assisted intervention, LNCS. Vol. 4791; 2007 Oct 29–Nov 2; Brisbane, Australia. Berlin, Heidelberg: Springer; p. 858–865.
Orderud F, Kiss G, Torp H. 2008. Automatic coupled segmentation of endo- and epicardial borders in 3D echocardiography. In: Proceedings of IEEE international ultrasonics symposium; 2008 Nov 2–5; Beijing. IEEE. p. 1749–1752.
Ping Y, Sinusas A, Duncan JS. 2007. LV segmentation from 3D echocardiography using fuzzy features and a multilevel FFD model. In: Proceedings of IEEE international symposium on biomedical imaging; 2007 Apr 12–15; Arlington, VA. IEEE. p. 848–851.

Sonka M, Lelieveldt BPF, Mitchell SC, Bosch JG, van der Geest RJ, Reiber JHC. 2001. Active appearance motion model segmentation. In: Proceedings of USF international workshop on digital and computational video; 2001 Feb 9; Tampa, FL, USA. IEEE. p. 64–68.

Sugeng L, Mor-Avi V, Weinert L, Niel J, Ebner C, Steringer-Mascherbauer R, Schmidt F, Galuschky C, Schummers G, Lang RM, et al. 2006. Quantitative assessment of left ventricular size and function: side-by-side comparison of real-time three-dimensional echocardiography and computed tomography with magnetic resonance reference. Circulation. 114(7):654–661.

Sukno FM, Ordas S, Butakoff C, Cruz S, Frangi AF. 2007. Active shape models with invariant optimal features: application to facial analysis. IEEE Trans Pattern Anal Mach Intell. 29 (7):1105–1117.

van Stralen M, Leung KYE, Voormolen MM, de Jong N, van der Steen AFW, Reiber JHC, Bosch JG. 2007. P2A-8 fully automatic detection of left ventricular long axis and mitral valve plane in 3D echocardiography. In: Proceedings of IEEE international ultrasonics symposium; 2007 Oct 28–31; New York, NY. IEEE. p. 1488–1491.

Woo J. 2009. A short history of the development of 3-D ultrasound in obstetrics and gynecology. Available from: http:www.ob-ultrasound.nethistory-3D.html

Yan P, Jia CX, Sinusas A, Thiele K, O'Donnell M, Duncan JS. 2007. LV segmentation through the analysis of radio frequency ultrasonic images. In: Proceedings of international conference information processing in medical imaging, LNCS. Vol. 4584; 2007 Jul 2–6; Kerkrade, The Netherlands. Berlin, Heidelberg: Springer; p. 233–244.

Yu Y, Acton ST. 2002. Speckle reducing anisotropic diffusion. IEEE Trans Image Process. 11(11):1260–1270.

Zagrodsky V, Walimbe V, Castro-Pareja CR, Qin JX, Song JM, Shekhar R. 2005. Registration-assisted segmentation of real-time 3-D echocardiographic data using deformable models. IEEE Trans Med Imag. 24(9):1089–1099.

Appendix A1: Fast image generation with FastGen

This US image generation method follows an approach originally proposed by Bamber and Dickinson (1980). It is assumed that the imaging system can be modelled by a linear, space-invariant PSF, which would correspond to a 3D analogue of a linear probe. Let $t(x,y,z)$ be an echogenicity model (an image with different intensity values corresponding to different tissues, see Figure A1 (b)) of the object being imaged (Figure A1(a)). The x, y and z are lateral, elevation and axial coordinates. First, scatterer distribution is modelled by multiplying the echogenicity model by a Gaussian white noise $G(\sigma_n;x,y,z)$ with zero mean and variance σ_n^2 (Figure A1(c)):

$$T(x,y,z) = t(x,y,z)\cdot G(\sigma_n;x,y,z). \tag{A1}$$

The 3D ultrasonic echo data-set $V(x,y,z)$ can then be obtained by a convolution

$$V(x,y,z) = h(x,y,z)*T(x,y,z), \tag{A2}$$

where

$$h(x,y,z) = h_1(x,\sigma_x)\cdot h_1(y,\sigma_y)\cdot h_2(z,\sigma_z), \tag{A3}$$

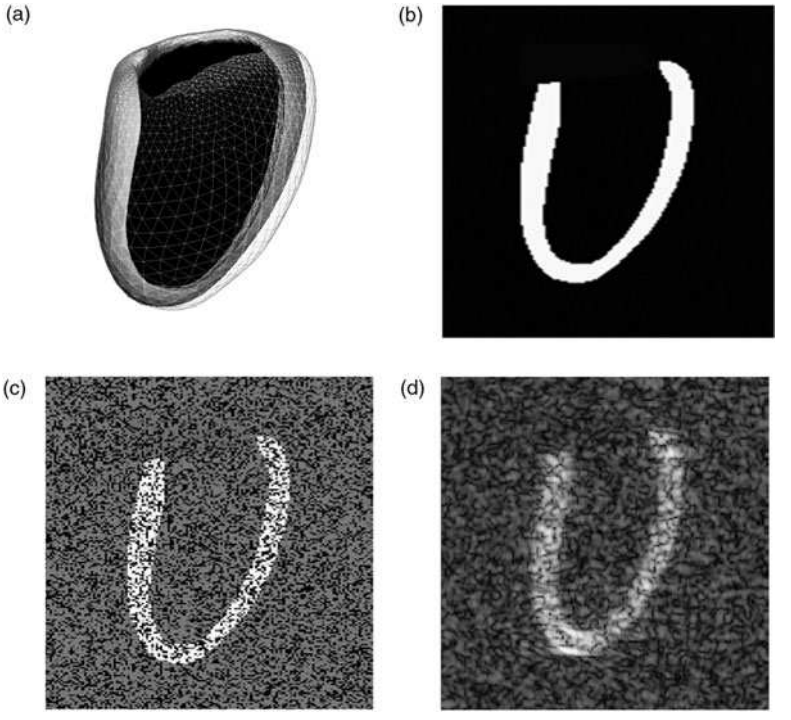

Figure A1. US image generation by FastGen. The shape (a) is used to generate an echogenicity model (b) where different voxel intensities represent different tissues. Subsequently Gaussian noise is added to introduce scatterer variations (c), the result is convolved with the PSF (8) and the envelope of the convolution result is computed (d).

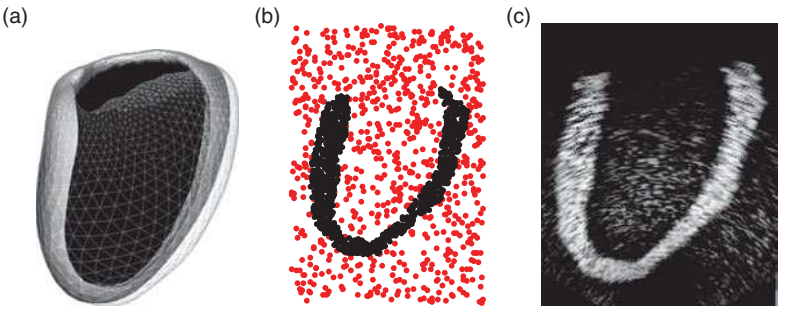

Figure B1. US image generation using Field II. The shape (a) is used to randomly populate a 3D space with point scatterers (b), each of them having an assigned value that characterises strength of response. The image is obtained by modelling sound wave propagation (c). The black points in (b) correspond to myocardium. Note that different densities of points have been used only for better visualisation.

$$h_1(u,\sigma_u) = \exp\left[-u^2/(2\sigma_u^2)\right], \tag{A4}$$

$$h_2(v,\sigma_v) = \sin\left(2\pi f_0 v/c\right)\exp\left[-v^2/(2\sigma_v^2)\right], \tag{A5}$$

where c is the speed of sound in a soft tissue (assumed 1540 m/s) and f_0 is the centre frequency of the transducer. $f_0 = 3$ MHz is used throughout this paper, being a typical frequency for cardiac imaging.

The image of the envelope-detected amplitude, $A(x,y,z)$ (shown in Figure A1(d)), is given by

$$A(x,y,z) = \left|V(x,y,z) + i\hat{V}(x,y,z)\right|, \tag{A6}$$

where $\hat{V}(x,y,z)$ is the Hilbert transform of $V(x,y,z)$ and i is the imaginary unit.

Appendix B1: Image generation using Field II

This approach relies on linear systems theory to find the US field for the cases of both pulsed and continuous wave. The latter is performed using the spatial impulse response, which, when the transducer is excited by a stimulus, modelled by the Dirac delta function, gives the emitted US field at a specific point in space as a function of time. The field for any kind of excitation can then be found by convolving the spatial impulse response with the excitation function. The modelling program allows for any transducer geometry (unlike that from the previous subsection) and apodisation, and uses a far-field approximation, to keep the process realistic and simple. The approach assumes a homogeneous bounded medium, where the pressure is sufficiently small to ensure linear wave propagation.

The method can be described using Huygens principle, where the impulse response is calculated from a summation of all spherical waves from the aperture area S as:

$$h(\mathbf{r}_1,t) = \int_S \frac{\delta(t-|\mathbf{r}_1-\mathbf{r}_2|c)}{2\pi|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}S, \tag{B1}$$

where $|\mathbf{r}_1-\mathbf{r}_2|$ is the distance from the transducer at position $\mathbf{r}_2$ to the field point $\mathbf{r}_1$, $\delta(t)$ is the Dirac delta function and c is the speed of sound. For a number of apertures (such as round piston, circular convex element and rectangular element), the calculation can be done analytically. However, it is not possible for a general aperture. Field II, therefore, divides the aperture into smaller mathematical elements to describe advanced shapes. Subsequently, any kind of linear US field can be calculated using spatial impulse responses. The emitted pressure field $p(\mathbf{r}_1,t)$ is given by

$$p(\mathbf{r}_1,t) = \rho_0 \frac{\partial v(t)}{\partial t} * h(\mathbf{r}_1,t), \tag{B2}$$

where ρ_0 is the density of the medium and $\partial v(t)/\partial t$ is the acceleration of the front face of the transducer. The received voltage signal for the pulse echo field is:

$$v_r(\mathbf{r}_1,t) = v_{\mathrm{pe}}(t) * f_m(\mathbf{r}_1) * h_{\mathrm{pe}}(\mathbf{r}_1,t), \tag{B3}$$

$$f_m(\mathbf{r}) = \frac{\Delta\rho(\mathbf{r}_1)}{\rho_0} - \frac{2\Delta c(\mathbf{r}_1)}{c}, \tag{B4}$$

where the scattering signal $f_m(\mathbf{r})$ arises from spatial variations in density $\Delta\rho(\mathbf{r}_1)$ and speed of sound $\Delta c(\mathbf{r}_1)$. Here $h_{\mathrm{pe}}(\mathbf{r}_1,t)$ is the two-way spatial impulse response, which is a convolution between the impulse response of the transmitting and receiving aperture. The impulse response $v_{\mathrm{pe}}(t)$ includes the excitation convolved with the transducer's electro-mechanical impulse response in both transmit and receive. For further details, please refer to (Jensen and Svendsen, 1992; Jensen, 1996).

The image generation process is similar to that of FastGen (see Figure B1) with an exception that the echogenicity model is not defined on a uniform grid anymore. It is created by populating the 3D space with randomly positioned point scatterers, each having an amplitude that characterises the strength of the reflected signal. Subsequently, Field II uses this echogenicity model together with a custom transducer definition to obtain the US image.