A Novel Model-Based 3D+time Left Ventricular Segmentation Technique

Stephen P. O'Brien*, Ovidiu Ghita, and Paul F. Whelan, *Senior Member, IEEE*

Abstract—**A common approach to model-based segmentation is to assume a top-down modelling strategy. However, this is not feasible for complex 3D+time structures such as the cardiac left ventricle due to increased training requirements, aligning difficulties and local minima in resulting models. As our main contribution, we present an alternate bottom-up modelling approach. By combining the variation captured in multiple dimensionally-targeted models at segmentation-time we create a scalable segmentation framework that does not suffer from the 'curse of dimensionality'. Our second contribution involves a flexible contour coupling technique that allows our segmentation method to adapt to unseen contour configurations outside the training set. This is used to identify the endo- and epi-cardium contours of the left ventricle by coupling them at segmentation-time, instead of at model-time. We apply our approach to 33 3D+time MRI cardiac datasets and perform comprehensive evaluation against several state-of-the-art works. Quantitative evaluation illustrates that our method requires significantly less training than state-of-the-art model-based methods, while maintaining or improving segmentation accuracy.**

Index Terms—**active shape model (ASM), spatio-temporal left ventricle dynamics, contour coupling, ASM optimisation.**

I. INTRODUCTION

CARDIOVASCULAR disease (CVD) is a major cause of death in the Western World. CVD is responsible for nearly half of all deaths in Europe [1], and approximately a third of all deaths in the USA [2]. Magnetic Resonance Imaging (MRI) is a popular minimally-invasive imaging modality used to diagnose CVD. The clinical MRI imaging modality produces 3D and 3D+time images of the heart. To diagnose CVD, doctors commonly assess the functionality of the left ventricle (LV) over the cardiac cycle. Quantitative indicators such as ejection fraction and cardiac output are used to assist in the assessment of CVD pathologies [3]. To derive these metrics, the two boundaries describing the LV (the endo- and epicardium) need to be extracted from a patient scan. Classically, this is done manually by doctors - however manual analysis is no longer feasible due to the quantity of data in 3D and 3D+time images. Moreover, manual annotation is prone to inter/intra observer variability which can affect the reliability and repeatability of CVD diagnosis. Therefore, an automatic segmentation solution is needed to address this clinical problem.

The challenges associated with cardiac LV segmentation arise from both the image domain and the physical structure of the heart. MRI images can suffer from significant intensity variation, as well as poor contrast in areas, most notable between the epicardium and the surrounding thoracic region. Motion artefacts caused, for example, by patient breathing during image acquisition degrades image quality and creates irregular shifts between anatomical structures in adjacent slices of the volumetric data. The spatial and temporal resolution of MRI images can be coarse relative to the physical size of the LV and Computer Aided Diagnosis (CAD) tools require either data interpolation or the development of flexible segmentation techniques.

The combination of these difficulties preclude any purely data-driven segmentation method from being a feasible choice for 3D+time LV segmentation. It is widely recognised that maximising the use of *a priori* knowledge of LV deformation is a pre-requisite for robust segmentation methods. However, capturing the complex spatio-temporal dynamics is made difficult by the significant anatomical variation of the cardiac structure itself - especially when pathologies are considered. Therefore, since model-based approaches commonly rely on a training set of expertly pre-segmented images, the quality, quantity and variation of these images becomes of paramount importance.

The vast majority of recent model-based approaches in this area have been based on a *top-down* modelling strategy. In this approach, one attempts to capture all of the desired variation of the segmentation target in one model, so the dimensionality of the model matches that of the segmentation target. For clarity, we refer to this as the *single-model* strategy, since only one model is used to capture the desired variation. With respect to cardiac segmentation, the work by Mitchell *et al.* [4] extended the standard 2D Active Appearance Model (AAM) - first introduced by Cootes *et al.* [5] - to a 3D implementation. Using a similar approach Bosch *et al.* [6] model temporal deformation in a 2D+time AAM-based segmentation of echocardiographic images. Kaus *et al.* [7], describe a 3D deformable model for LV myocardium segmentation. The authors present promising results but exclude datasets exhibiting large motion artefacts from their evaluation. This issue was later addressed by Stegmann *et al.* [8], [9] who use an unsupervised motion-correction scheme before segmentation. The authors concatenate the end-systolic and end-diastolic cardiac phases of the LV into a single AAM in an attempt to include temporal variation in their model. This technique is also used by Fritz *et al.* [10] who extended it further to consider the 3D surfaces of the left and right ventricles at both the end-systolic and end-diastolic cardiac phases as a single shape. The authors argue that if, for example, the right ventricle is poorly defined

Asterisk denotes corresponding author.
*S. P. O'Brien (email: stephen.obrien32@mail.dcu.ie), O. Ghita and P. F. Whelan are all with the Centre for Image Processing and Analysis, Dublin City University, Dublin 9, Ireland

compared to the LV, then the fitting process will be guided by the stronger image feature, thus providing a more robust overall segmentation. However, they do not indicate whether they weight the landmark fitting process accordingly or if they rely purely on the ASM approximation of the landmarks (further discussed in Section II-A2) to correctly guide the segmentation process. Another 3D-ASM LV segmentation technique is described by van Assen *et al.* [11] - also used in [12] and later developed in [13] - where the authors replace the standard landmark fitting stage in the ASM formulation with a fuzzy inference system. The authors report segmentation accuracy comparable to the state-of-the-art implementations evaluated in their study. In our implementation we also replace the standard ASM fitting process, as we will demonstrate that it is unreliable when applied to complex image environments. Finally, to alleviate the requirement on manual annotation of training samples, several papers have explored different approaches for the automatic generation of landmarks and point-correspondence [14], [15]. For a complete review of the application of model-based segmentation approaches see the recent work of Heimann *et al.* [16].

While largely successful as the research community transitioned from 2D to 3D segmentation, the single-model strategy is not easily generalisable to higher dimensional problems. This is reflected both by the notable lack of single-model 3D+time research and the criticisms recently expressed in a number of separate works [17]–[20]. The central issue is that as the dimensionality of the modelling problem rises there is a consequent rise in: the degrees of freedom, the size of the required training set, the aligning difficulties and the occurrence of multiple local minima in the resulting model. This is generally referred to as the 'curse of dimensionality' [17] (CoD). The reliance on training and the limited availability of large pre-annotated image databases undermines the practicality of single-model 3D+time segmentation methods. In fact, the only notable single-model 3D+time work presented in the literature, by Lorenzo-Valdés *et al.* [21], is heavily reliant on pre-segmented data. The authors' technique requires not only the left and right ventricles but also the background structures to be pre-segmented to build their 3D+time atlas. Naturally, their training data exhibits insufficient variation and Gaussian blurring was used to artificially enlarge model variation. Tolli *et al.* [22] also present a method for artificial enlargement of the variation in small training sets for construction of a 3D ASM of the four chambers of the heart. It should be noted that the authors also needed to greatly increase the allowed deformations of the ASM from the standard $\pm 3\sqrt{\lambda_i}$ to $\pm 5\sqrt{\lambda_i}$ (this issue is explained in detail in Section II-A1). More recently, Zhang *et al.* [23] argue against the artificial enlargement of the training set to maintain structural consistency in the resulting models. The authors extend the work of Mitchell *et al.* [24] using a hybrid ASM/AAM implementation, training both a 3D and a 3D+time model. The authors concede that they do not have enough training images to capture sufficient temporal variation in their 3D+time model, and so their 3D model must be used for a phase-by-phase final segmentation stage in order to avoid the over-constrained temporal variation in the 3D+time model. These works [21], [23] show the limitations,

most notably with respect to training requirements, in the single-model approach for 3D+time segmentation.

However, a recent trend in the literature [17], [19], [20] has gone some way in reassessing the approach to higher dimensional modelling and segmentation problems. Instead of creating a single model, subsections of the overall variation are captured in separate models which are then combined at segmentation-time. In this way, the full variation of the LV is still captured, but the dimensionality of each model is lower than an equivalent single-model approach and therefore has less training requirements, simpler aligning and is better able to encapsulate the statistical variation of the training set. Zambal *et al.* [19] describe a 3D left ventricle segmentation scheme, where the LV is modelled using a set of 2D Active Appearance Models (AAMs), each capturing a specific section of the ventricle. These models are interconnected with a separate 3D shape model that controls the global positioning and scale during segmentation. The authors report that their technique outperforms a standard 3D AAM by 11% in terms of segmentation accuracy. More importantly, the authors show that their method inherently copes with irregular translation between slices since each AAM can move freely in its own slice, while structural consistency is maintained using a separate 3D model. This is a benefit in cardiac segmentation where adjacent slices are often misaligned and this negates the need for a pre-aligning phase as required by the approach proposed by Stegmann *et al.* [8]. Previous work by our group, Lynch *et al.* [25], [26], describes a more sophisticated approach whereby a non-rigid model is used to encode *a priori* temporal deformations into a level-set formulation, with promising results presented. Andreopoulos *et al.* [17] recently presented a 3D+time cardiac segmentation method, which relies on two components. The first captures 3D shape variation of the left ventricle in a 3D AAM. The second part encompasses the temporal variation of the ventricle in a 2D+time ASM. The 2D+time ASM is used to refine the initial 3D AAM segmentation. Like Zambal *et al.* [19], the authors argue strongly against creating single 3D+time models due to the issues relating to the CoD. While their work does decompose the 3D+time problem into spatial and temporal models the authors still build a single-model 3D shape AAM. Therefore they are still faced with significant training requirements (and related issues) albeit to a lesser degree than Lorenzo-Valdés *et al.* [21]. To address this the authors extend the wavelet-based hierarchical ASM presented by Davatzikos *et al.* [27] to a 3D implementation, a method shown to perform well with limited training samples.

The aim of this paper is to further analyse the multi-model approach and bring the concept to its logical conclusion. As our first contribution, we decompose the 3D+time LV modelling problem into its component parts - i.e. shape, spatial and temporal variation. The models are then combined in a dimensionally scalable segmentation framework that unifies the variation in the separate models. In our second contribution, we further divide the shape variation into separate endo- and epicardium component models. Many recent studies [9], [13], [17] consider the endo- and epicardium interfaces as a single shape. However, this design choice further exacerbates

the issues related to the CoD. We show that by coupling the two contours at segmentation-time instead of model-time we increase the flexibility of our segmentation technique to cases outside the training set. We experimentally verify that our approach requires significantly less training and achieves comparable or better segmentation accuracy when compared with state of the art implementations.

II. METHOD

A. ASM Theory

We use the Active Shape Model, first developed by Cootes *et al.* [28], [29] to capture the shape, spatial and temporal variation of the LV. The ASM is chosen instead of the AAM as the majority of relevant 3D+time variation related to the LV is present in the shape variation. We do not consider the texture variation in the largely homogeneous LV blood pool and surrounding myocardium to be sufficient to warrant the use of an AAM.

1) ASM Construction: When building an ASM from a set of training shapes, the segmentation target must be descretised into N landmarks. A shape vector, $\mathbf{s}$, containing N landmarks is described as:

$$\mathbf{s} = [l_0, l_1, \ldots, l_i, \ldots, l_N] \tag{1}$$

where l_i is the ith landmark of the shape. It follows that the training set, $\mathbf{S}$, containing M training shapes is defined as:

$$\mathbf{S} = [\mathbf{s_0}, \mathbf{s_1}, \ldots \mathbf{s_M}] \tag{2}$$

The set of training shapes must then be aligned into a common coordinate system. As stated earlier, the complexity of this step is proportional to the dimensionality of the target shape. Since we use a bottom-up modelling strategy, we build a 2D shape model that is only trained on the basal-most slice of the end-diastolic volume. Therefore, to align our training samples we only need to remove translation between the shapes as we capture the scale variation in the ASM. Point-correspondence is established with an arc-length re-sampling method. After our training set ($\mathbf{S}$) has been aligned ($\mathbf{S}'$) we apply Principal Component Analysis (PCA) to extract the eigenvectors, ϕ, and eigenvalues, λ, describing the shape variation. According to the ASM formulation these are truncated to capture 95% of the training set variation. The truncated eigenvectors (Φ) are combined with the mean-shape of the training set ($\bar{\mathbf{s}}$) to generate the equation that governs ASM shape synthesis:

$$\mathbf{s} \approx \bar{\mathbf{s}} + \Phi\beta \tag{3}$$

Shape synthesis using (3) is constrained using the vector β_{limit}, where the ith component of β_{limit} is given by $\pm 3\sqrt{\lambda_i}$ [30]. Therefore, all shapes synthesized by the ASM are subject to: $|\beta| \leq \beta_{limit}$.

2) Standard ASM Fitting: The ASM is fitted to an unseen image according to an iterative two-stage process. First, each landmark updates it's location individually, using only local image information and has no interaction with surrounding landmarks or the ASM itself. This new set of landmarks (L') is then approximated by the ASM which keeps the shape

within the learned range of variation. This process repeats until convergence is reached.

There are, however, numerous flaws with this approach when applied to complex segmentation tasks, such as the segmentation of the LV. First, in noisy images, where contours are poorly-defined, individual landmarks can fit to spurious image areas. This will negatively influence the resulting ASM approximation and can affect the overall convergence. Second, since the landmarks are updated locally, L' can, and usually does, assume irregular shapes that are outside the learned variation of the ASM. In effect this creates a disconnect between the constrained nature of the ASM and the unconstrained nature of the landmark update process. This severely undermines the reliability of the ASM in noisy images. This shortcoming in the original ASM formulation has been previously dealt with in several works in the literature [10], [31]–[34]. To address this problem, we propose to replace the standard fitting process with a global ASM optimisation technique. This will be addressed in the next section of the paper.

B. ASM Optimisation

According to (3), shapes are synthesised by varying β within learned constraints (β_{limit}). We fit the ASM to an unseen image by direct manipulation of the β vector. This allows us to perform a *global* contour fit by creating an objective function that minimises the energy of the whole contour, thereby vastly improving the robustness of the fitting processes. As we train the endo- and epicardium separately (this issue will be further discussed in Section II-C) we require an objective function for each contour. To describe the objective functions for endo- and epicardium minimisation we first need to define some descriptive notation. A β vector generates a shape, $\mathbf{s}_\beta$, (i.e. $\mathbf{s}_\beta = \bar{\mathbf{s}} + \Phi\beta$) and delineates a corresponding region, Ω_β. The objective function operates on the image information around the contour: defined as Ω_β^+ and Ω_β^- for the outer and inner regions, respectively. The function $I(\Omega)$ returns a vector of grey-scale values inside the image region Ω. It follows that $\bar{I}(\Omega)$ and $I^\sigma(\Omega)$ return the mean and standard deviation, respectively, of the greyscale information in the region Ω.

Using this notation, we create an objective function (f) for minimising the endo- and epicardium in our 2D example (f_{endo} and f_{epi}, respectively). The endocardium objective function f_{endo} is defined as:

$$f_{endo}(\beta, T) = \left| \bar{I}\left(T\left(\Omega_\beta^+\right)\right) - \bar{I}\left(T\left(\Omega_\beta^-\right)\right) \right| \tag{4}$$

where T denotes a translation transformation, defined as: $T = [t_x, t_y]$. This objective function calculates the contrast between the inner and outer region of Ω_β. To fit the endocardium to an unseen image, f_{endo} is minimised according to:

$$\min_{|\beta| \leq \beta_{limit}} f_{endo}(\beta, T) \tag{5}$$

Equation (5) states that the endocardium's energy is minimised when there is a maximum contrast between the inner and outer regions of Ω_β. This is expected since there is high contrast between the left ventricular blood pool and the surrounding darker myocardium. By operating on image regions instead

of individual 1D landmark profiles our minimisation method is not susceptible to outliers [35] negatively influencing ASM convergence.

The epicardium is minimised similarly to the endocardium, however the difference between the standard deviations of the inner and outer region also needs to be maximised since we know that the area defined by the myocardium is relatively homogeneous compared to the region immediately outside the myocardium wall. Specifically, outside the myocardium there is significant intensity variation between the bright right ventricular blood pool and the very dark surrounding thoracic region. The addition of this extra term is required as the epicardium border is usually poorly defined in MRI images, therefore extra robustness is required at this segmentation stage. The epicardium objective function is formulated as follows:

$$
\begin{aligned}
f_{epi}\left(\beta, T\right) = & \left[\left|\bar{I}\left(T\left(\Omega_\beta^-\right)\right) - \bar{I}\left(T\left(\Omega_\beta^+\right)\right)\right| \right. \\
& + \left. \left|I^\sigma\left(T\left(\Omega_\beta^-\right)\right) - I^\sigma\left(T\left(\Omega_\beta^+\right)\right)\right|\right] \quad (6)
\end{aligned}
$$

Similarly to (5), the epicardium objective function is minimised by:

$$
\min_{|\beta| \leq \beta_{limit}} f_{epi}\left(\beta, T\right) \quad (7)
$$

Since each β parameter is associated with an eigenvector and since the eigenvectors Φ are ordered in terms of the percentage of variation they capture, we implement a coarse-to-fine optimisation scheme. Specifically, we use Brent's method [36] to minimise the ASM fitting errors for each β parameter sequentially, from largest eigenvector to smallest. We iterate this minimisation process until β converges, or until the difference between two successive iterations is below a small threshold. Notice that in (5) and (7) we can elegantly maintain the structural integrity of the ASM, whereby all synthesised shapes during fitting are within the learned variation, β_{limit}.

To highlight the difference in fitting accuracy we present a qualitative comparison between the standard Cootes ASM fitting formulation [29] and our segmentation method in Fig. 1. The Cootes fitting method illustrates the disconnect between the point search phase (points shown in red) and the corresponding ASM approximation (contour shown in green). In this image the influence of the papillary muscle negatively affects the quality of the ASM fit. This is a prime example of the weakness of the Cootes approach when applied to complex segmentation tasks. By contrast, our optimisation method is not influenced by the papillary muscle as it is treated as a local feature in our ASM minimisation process. Instead, our technique correctly optimises the ASM contour since the best global fit is found on the true endocardium border.

C. Contour Coupling

When learning the shape variation of the LV contours we construct separate endo- and epicardium shape models. Traditionally both contours are regarded as a single shape [9], [13], [17]), however this approach has several drawbacks.
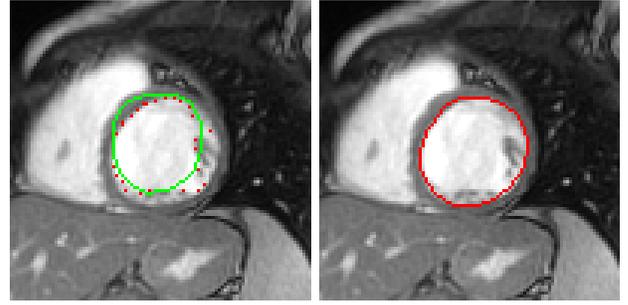
Fig. 1. On the left is the standard Cootes ASM fitting formulation. On the right is our ASM optimisation technique applied to the same image. All images in this paper are best viewed in colour.

First, training both shapes as one entity generates an ASM model that captures three different sources of variation simultaneously - i.e. the shape variation of the endocardium and epicardium as well as the spatial relationship between them. Second, as a result of embedding the spatial relationship between the contours inside the model, each endocardium sample is learned only in the presence of a corresponding epicardium sample, and vice versa. This limits the overall adaptability of the model since varying β, in this case, will vary both contours simultaneously. This issue is especially relevant in cases where there is only a small training set available to model the complex anatomical variation of the LV.

In our approach, we train separate endo- and epicardium models and derive their spatial relationship at segmentation time. In other words, we decouple the contours at model construction time - instead coupling them at segmentation time. Now that the endo- and epicardium shape models are trained independently, we can use a flexible data-driven coupling technique that adapts to unseen contour configurations *outside* the training set. This is an extremely useful property for LV segmentation as there is a substantial range of anatomical variation in both healthy and pathological patients, that cannot, practically, be modelled exclusively using training set samples.

We formulate the contour coupling technique as a multistage process. First, we segment the endocardium in the image. We use the endocardium region, Ω_{endo}, to drive the epicardium segmentation as it is the stronger feature of the two. We perform a radial search, starting from the centre of the shape, $C = \frac{1}{N+1} \sum_{i=0}^{N} l_i$ (where l_i is the ith landmark in the shape), by collating gradient information along a 1D profile of length L (Fig. 2). The result of the process used to construct the cumulated gradient profile is illustrated in Fig. 4. Using Ω_{endo}, we can derive a probability density function with parameters μ_{endo} and σ_{endo} describing the mean distance and standard deviation, respectively, between the endocardium contour and C. As we know that the epicardium is approximately concentric with the endocardium, we can *estimate* the probable distance of the epicardium from C, μ'_{epi}, as the next peak in the 1D gradient profile after μ_{endo} (Fig. 4). To take account for irregularities in the shape of the epicardium and to allow for error in the estimate of the epicardium we define the

region $\Omega_{\mu'_{epi}+3\sigma_{endo}}$ as the outer limit of the region of interest, in which, we allow the epicardium ASM to minimise (see Fig. 3). We have shown experimentally that $3\sigma_{endo}$ is a robust indicator of the variation in the epicardium, highlighting the close relationship between the contours. Therefore we define the region of interest as:

$$\Omega_{roi} = \Omega_{\mu'_{epi}+3\sigma_{endo}} \setminus \Omega_{endo} \qquad (8)$$
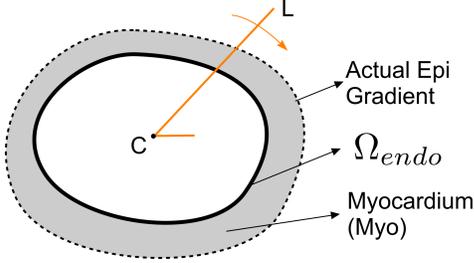
Fig. 2. Radial search for gradient data, after endocardium segmentation.
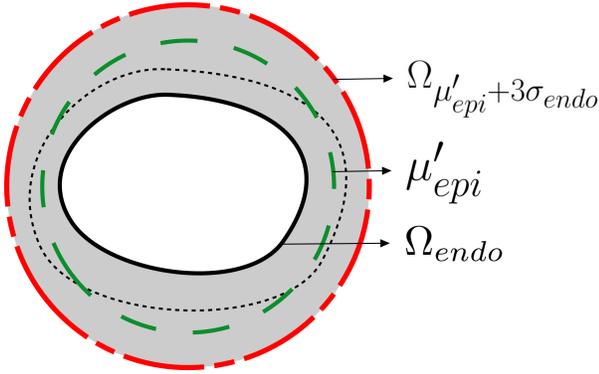
Fig. 3. Using the estimated mean epicardium location μ'_{epi} to derive the outer limit of the region of interest, $\Omega_{\mu'_{epi}+3\sigma_{endo}}$.

In this way, Ω_{roi} enforces structural consistency between the endo- and epicardium. The fact that the spatial relationship between the contours is enforced only at segmentation time (through Ω_{roi}) allows many more potential unseen contour configurations, outside the training set variation, to be successfully segmented. Overall, this improves the flexibility of our segmentation technique in comparison with methods that consider both contours as a single shape. Returning to the endo- and epicardium objective functions, (4) and (6), the contour coupling constraint is applied so that Equation (7) is subject to the extra constraint:

$$\begin{aligned} \mathbf{s}_{\beta_{epi}} &\subseteq \Omega_{roi} \\ &\subseteq \Omega_{\mu'_{epi}+3\sigma_{endo}} \setminus \Omega_{endo} \end{aligned} \qquad (9)$$

Fig. 5 illustrates the benefit of our coupling approach in contrast with an ASM that considers the endo- and epicardium as a single shape. The single endo/epi model tries to find the best trade-off between fitting the two contours simultaneously, resulting in a poor global fit. By contrast, since we train separate shape models and maintain structural consistency
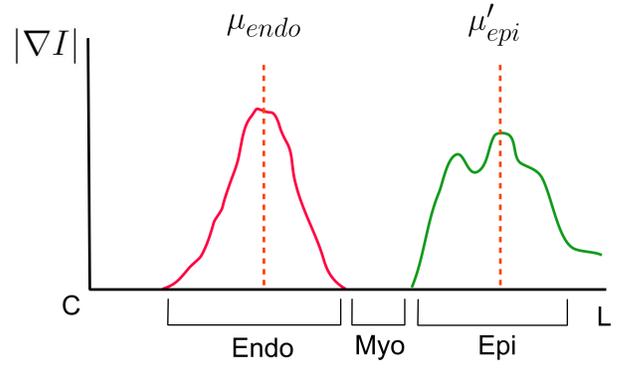
Fig. 4. Representation of epicardium estimation, derived from known endocardium border.

between them, our technique is more adaptable to the intrinsic properties of the endo- and epicardium contours in cardiac data.
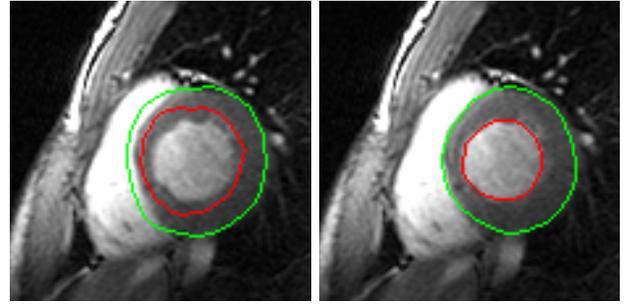
Fig. 5. On the left is an example of an ASM trained considering the endo- and epicardium as single shape. On the right is our coupling method applied to the same image.

D. Spatial and Temporal Modelling

As indicated earlier, our bottom-up strategy captures the spatial and temporal variation in separate models. Our hypothesis is that if we learn the temporal variation only from the basal-most slices and the spatial variation only from the end-diastolic volume we can combine the information from both models to segment any other slice in an unseen 3D+time dataset. As a result, we require only a fraction of a 3D+time dataset to be annotated for training compared with the single-model approach. For example, if a dataset has spatial and temporal dimensions z and t, respectively - our method requires $z+t-1$ slices to be annotated as opposed to $z \times t$ in the single-model approach. This has significant practical implications, since a standard 3D+time dataset requires between 6-8 hours for two cardiologists to annotate [23].

To capture the temporal variation we begin with the basal-most slice of shape annotations across the temporal axis, as in (10), where Y_i is the ith training sample in the training set $\mathbf{Y}$.

$$\begin{aligned} Y_i &= \left[\left[l_0^0, l_1^0, \ldots, l_N^0\right], \ldots, \left[l_0^M, l_1^M, \ldots, l_N^M\right]\right] &(10) \\ Y_i' &= \left[\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_M\right] &(11) \\ Y_i'' &= \left[\beta_0, \beta_1, \ldots, \beta_M\right] &(12) \end{aligned}$$

where N is the number of landmarks in a shape, and M is the number of shapes in a single training sample. Since we are operating with a set of 2D shapes, alignment of each training sample is achieved by removing translation (Y_i'). Point correspondence is again enforced with arc-length re-sampling for each shape in the set. Now, we can restate (10) by considering each set of landmarks as a shape, as in (11). Using Equation (3) we can map each shape into an equivalent β vector. This creates a compact and abstract slice description in β space that simplifies the description of each training sample (12). The abstract description of the temporal slice is now focused on the macro-variations of the shapes, instead of relearning the micro-variation of each landmark which has already been captured in the 2D ASM (3). PCA can now be applied to the aligned training set ($\mathbf{Y}''$). This creates a new ASM modelling strategy that is described in Equation (13), where the α parameter controls the variation of LV shapes in the temporal axis - thus controlling the *temporal profile* of the model. We use a similar process to construct a spatial ASM (14) where the ξ parameters controls the *spatial profile* of the spatial model.

$$\mathbf{t} \approx \bar{\mathbf{t}} + \Psi\,\alpha \tag{13}$$
$$\mathbf{z} \approx \bar{\mathbf{z}} + \Upsilon\,\xi \tag{14}$$

We construct α_{limit} and ξ_{limit} so that they are constrained according to their corresponding eigenvalues, in a method similar to that detailed in Section II-A1. Each element of $\mathbf{t}$ and $\mathbf{z}$ (t^k and z^k) is itself a β vector and directly maps into shapes denoted by, α^k and ξ^k, respectively. The objective functions of $\mathbf{t}$ and $\mathbf{z}$ (g and h, respectively) can therefore be defined as a summation of the corresponding 2D endo- and epicardium objective functions, as follows:

$$
\begin{aligned}
g_{endo}\left(\alpha, \mathbf{T}\right) &= \sum_{k=0}^{P} \left| \bar{I}\left(T^k\left(\Omega_{\alpha^k}^+\right)\right) - \bar{I}\left(T^k\left(\Omega_{\alpha^k}^-\right)\right)\right| \\
&= \sum_{k=0}^{P} f_{endo}\left(\alpha^k, T^k\right)
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
g_{epi}\left(\alpha, \mathbf{T}\right) &= \sum_{k=0}^{P} \Big[\left| \bar{I}\left(T^k\left(\Omega_{\alpha^k}^-\right)\right) - \bar{I}\left(T^k\left(\Omega_{\alpha^k}^+\right)\right)\right| \\
&\quad + \left| I^\sigma\left(T^k\left(\Omega_{\alpha^k}^-\right)\right) - I^\sigma\left(T^k\left(\Omega_{\alpha^k}^+\right)\right)\right|\Big] \\
&= \sum_{k=0}^{P} f_{epi}\left(\alpha^k, T^k\right)
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
h_{endo}\left(\xi, \mathbf{T}\right) &= \sum_{k=0}^{Q} \left| \bar{I}\left(T^k\left(\Omega_{\xi^k}^+\right)\right) - \bar{I}\left(T^k\left(\Omega_{\xi^k}^-\right)\right)\right| \\
&= \sum_{k=0}^{Q} f_{endo}\left(\xi^k, T^k\right)
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
h_{epi}\left(\xi, \mathbf{T}\right) &= \sum_{k=0}^{Q} \Big[\left| \bar{I}\left(T^k\left(\Omega_{\xi^k}^-\right)\right) - \bar{I}\left(T^k\left(\Omega_{\xi^k}^+\right)\right)\right| \\
&\quad + \left| I^\sigma\left(T^k\left(\Omega_{\xi^k}^-\right)\right) - I^\sigma\left(T^k\left(\Omega_{\xi^k}^+\right)\right)\right|\Big] \\
&= \sum_{k=0}^{Q} f_{epi}\left(\xi^k, T^k\right)
\end{aligned}
\tag{18}
$$

where P and Q refer to the number of temporal volumes and spatial slices in the dataset, respectively. The objective functions for g and h are minimised according to:

$$\min_{|\alpha| \le \alpha_{limit}} g_{endo}\left(\alpha, \mathbf{T}\right) \tag{19}$$

$$\min_{|\alpha| \le \alpha_{limit}} g_{epi}\left(\alpha, \mathbf{T}\right) \tag{20}$$

$$\min_{|\xi| \le \xi_{limit}} h_{endo}\left(\xi, \mathbf{T}\right) \tag{21}$$

$$\min_{|\xi| \le \xi_{limit}} h_{epi}\left(\xi, \mathbf{T}\right) \tag{22}$$

where Equations (20) and (22) are subject to the coupling constraints similar to (9) so that: $\mathbf{s}_{\alpha_{epi}} \subseteq \Omega_{roi}^g$ and $\mathbf{s}_{\xi_{epi}} \subseteq \Omega_{roi}^h$, respectively.

Fig. 6 illustrates the limits of the temporal model with respect to the scale variation in our evaluation datasets (further details can be found in Section III-A). The shaded area between the model limits (shown as the upper and lower lines) represent the scale variation in all datasets from basal to apex regions. As shown in Fig. 6, we can note that with a reduction in the training level the temporal model still maintains sufficient variation to capture the full range of LV variation in the datasets. The spatial model exhibits similar characteristics.

III. RESULTS

A. Data description

The datasets used for evaluation have been kindly made public by Andreopoulos *et al.* [17]. The database consists of 33 3D+time child MRI datasets - with ages ranging from 2 to 17 years old. All datasets were acquired using the FIESTA protocol and exhibit pathologies though not all related to the LV, and we can note that this data shows a considerable range of anatomical variation. Dataset resolution is 256×256 pixels in the XY plane, between 8 and 15 spatial slices and 20 temporal volumes. Pixel spacing ranges between 0.93 and 1.64 mm. All datasets have clinically validated annotations of the endo- and epicardium. For further details on the data used, please refer to Andreopoulos *et al.* [17].

We present a direct quantitative comparison of our method and to the method presented in [17] in our evaluation. As stated earlier, the authors go some way towards our approach by separating spatial and temporal variation. However, since they train a single-model 3D AAM it gives an ideal point of comparison between our approach and the single-model strategy. Therefore, while presenting these results it should be noted that our method requires between 83% and 89% less training
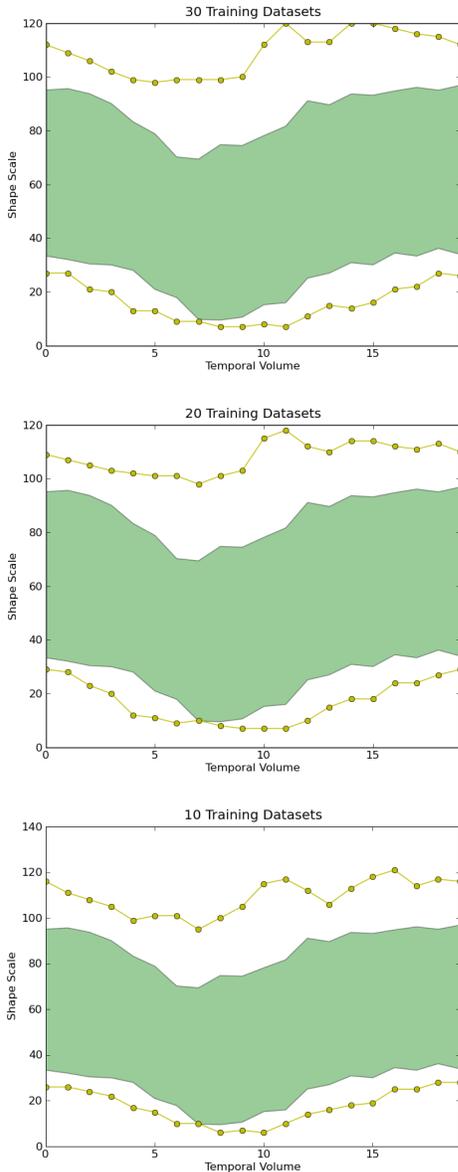
Fig. 6. Scale variation of the evaluation datasets (shaded region) shown with respect to the learned variation of the temporal model (upper and lower lines). The quantity of model training reduces from 30 to 20 to 10 datasets, from top to bottom.

than the method detailed in [17]. We will quantitatively show that we achieve comparable or better segmentation accuracy despite the significant difference in training requirements.

B. Testing Protocol

Our primary aim in this study is to analyse the performance of our method with reducing training. To do this we present a range of results with varying levels of training and testing. Specifically, we use X datasets for training our models, while the remaining $33 - X$ datasets are used for segmentation. This process is repeated 33 times for each value of X. Therefore, as the level of training decreases the level of testing increases. We do this to remove any bias in the choice of datasets for modelling at each stage. In total, we evaluate the performance

of the proposed algorithm based on 2178 3D+time dataset segmentations - i.e. 43560 volume segmentations.

To quantify segmentation accuracy we use the volumetric error metrics as described in [17] so the methods can be directly compared. We use mean volumetric point-to-curve error as well as absolute volume error. We do not require a complex manual initialisation procedure as in many single-model papers [13], [23] where multiple 3D model parameters need to be initialised or estimated from manual segmentations [17]. We need only a single xy seed point on the basal-most slice of the end-diastolic cardiac phase, providing an initial translation, T, to localise the LV blood pool in that slice only. This stage could, of course, be automated using a method similar to that described in [18]. As stated in Zambal *et al.* [19], because there is no rigid connection between shapes in our approach and since we minimise for both β and translation T (see Section II-B), we can implicitly compensate for irregular translation between adjacent slices. As a result, we do not require any pre-registration phase before segmentation, as in [8].

C. Quantitative Results and Discussion

The initial observation that can be drawn from Table I and II is that our method shows excellent stability as training is reduced. It should be remembered that the datasets being segmented in this case exhibit significant variation. Compared with Andreopolous *et al.* [17] we attain similar, or better, mean volumetric point-to-curve errors but achieve consistently tighter standard deviation measures. This is further reflected with the absolute volume errors in Table I and II. We feel this shows that our approach generates more consistent segmentations and is less likely to fit in spurious image areas than [17]. We expect the epicardium segmentation to be slightly less accurate than the endocardium since it is the more diffuse border of the two. Andreopolous *et al.* [17] appears to perform slightly better in terms of epicardium accuracy, again however, we show lower standard deviation in our results. We wish to draw attention also to the comparison (albeit indicative) against the results presented by Lorenzo-Valdés *et al.* [21]. Their approach can be considered the antithesis of our modelling philosophy. We feel our method compares very favourable against theirs, especially considering the fact that they test purely on a leave-one-out basis to show their method in a best-case scenario. It should also be noted that their models require significant Gaussian blurring to improve model variation.

We also compute regression values for the endo- and epicardium. Figs. 10 and 11 show the regression graphs for the endo- and epicardium, respectively, as training reduces. Table III shows the regression values for the full range of testing. We also show comparison against the recent variational approach of Ayed *et al.* [37] and as well as the previous work from our group, Lynch *et al.* [26], in Table IV. These results show that our method compares favourably with state of the art techniques and it should be noted that both papers [26], [37] aim for pixel-level segmentations, a scenario that favours the inclusion of the papillary muscles in the analysed 3D+time data.

TABLE I

ENDOCARDIUM SEGMENTATION RESULTS. MEAN REFERS TO THE AVERAGE VOLUMETRIC POINT-TO-CURVE ERROR BETWEEN OUR METHOD AND MANUAL ANNOTATIONS. VOLUME REFERS TO THE ABSOLUTE VOLUME DIFFERENCE BETWEEN OUR SEGMENTATION AND MANUAL ANNOTATIONS. (NOTE: THE RESULT FOR ZHANG *et al.* [23] IS AN AVERAGE OF THEIR RESULT FOR NORMAL AND PATHOLOGICAL DATASETS)

Endocardium Results		
Train - Test	Mean (mm)	Absolute Volume (cm^3)
32 - 1	1.42 ± 0.24	5.19 ± 2.65
30 - 3	1.47 ± 0.16	5.51 ± 1.79
25 - 8	1.50 ± 0.11	5.52 ± 0.77
20 - 13	1.55 ± 0.11	5.74 ± 1.02
15 - 18	1.63 ± 0.11	6.41 ± 1.10
10 - 23	1.87 ± 0.21	7.92 ± 1.68
Andreopoulous [17]	1.43 ± 0.49	6.10 ± 6.44
Indicative Results		
Kaus [7]	2.28 ± 0.93	–
Lynch [25]	1.24 ± 1.33	–
Lorenzo-Valdés [21]	1.88 ± 2.00	–
Zhang [23]	1.69 ± 0.38	–

TABLE II

EPICARDIUM SEGMENTATION RESULTS. MEAN REFERS TO THE AVERAGE VOLUMETRIC POINT-TO-CURVE ERROR BETWEEN OUR METHOD AND MANUAL ANNOTATIONS. VOLUME REFERS TO THE ABSOLUTE VOLUME DIFFERENCE BETWEEN OUR SEGMENTATION AND MANUAL ANNOTATIONS. (NOTE: THE RESULT FOR ZHANG *et al.* [23] IS AN AVERAGE OF THEIR RESULT FOR NORMAL AND PATHOLOGICAL DATASETS)

Epicardium Results		
Train - Test	Mean (mm)	Absolute Volume (cm^3)
32 - 1	1.78 ± 0.45	10.78 ± 6.81
30 - 3	1.80 ± 0.20	11.48 ± 4.40
25 - 8	1.72 ± 0.13	10.10 ± 2.05
20 - 13	1.82 ± 0.11	10.84 ± 1.39
15 - 18	1.84 ± 0.09	11.37 ± 1.45
10 - 23	1.98 ± 0.13	12.26 ± 1.69
Andreopoulous [17]	1.51 ± 0.48	9.82 ± 8.97
Indicative Results		
Kaus [7]	2.62 ± 0.75	–
Lynch [25]	–	–
Lorenzo-Valdés [21]	2.75 ± 2.62	–
Zhang [23]	1.89 ± 0.49	–

Finally, we show several qualitative examples of segmentations using our method, from a random selection of datasets and slices within those datasets. We show 3 collections of images for 30, 20 and 10 training samples respectively (Figs 7, 8 and 9). These images qualitatively illustrate the robustness of our approach, with limited training, in the presence of significant anatomical variation, as well as variation in image quality and intensity.

IV. CONCLUSION

We have presented a new model-based technique for 3D+time cardiac segmentation. Our main contribution is a bottom-up modelling strategy that divides shape, spatial and temporal variation into separate models. In this paper, we have

TABLE III

CORRELATION COEFFICIENTS BETWEEN MANUAL AND AUTOMATIC VOLUMES.

Volume Regression Results		
Train - Test	Endocardium	Epicardium
32 - 1	0.98	0.92
30 - 3	0.98	0.95
25 - 8	0.99	0.98
20 - 13	0.98	0.98
15 - 18	0.98	0.98
10 - 23	0.97	0.98
Andreopoulous [17]	0.95	0.97

TABLE IV

CORRELATION COEFFICIENTS BETWEEN MANUAL AND AUTOMATIC AREAS.

Area Regression Results		
Train - Test	Endocardium	Epicardium
32 - 1	0.97	0.93
30 - 3	0.96	0.93
25 - 8	0.96	0.94
20 - 13	0.95	0.93
15 - 18	0.95	0.93
10 - 23	0.91	0.91
Indicative Results		
Ayed [37]	0.94	0.96
Lynch [26]	0.82	0.89

shown that by coupling the endo- and epicardium contours at segmentation-time instead of at model-time, our method can adapt to unseen contour configurations outside of the training set. Finally, we replace the standard ASM fitting formulation with a global contour optimisation technique that is shown to be well suited to challenging imaging environments.

Compared with the single-model strategy our approach requires significantly less training and does not suffer from the 'curse of dimensionality'. Our technique has been thoroughly compared against several state-of-the-art works and has shown consistently comparable or better performance. Specifically, our technique requires between 83% and 89% less training than the method detailed in [17] while maintaining comparable segmentation accuracy. In future work we wish to demonstrate our contour coupling approach in scenarios with multiple spatially-related contours. For example, a natural extension is application in left and right ventricle segmentation. Moreover, we will investigate other application domains for our modelling approach to illustrate the generalisability of our overall modelling strategy for problems exhibiting complex spatio-temporal variation.

ACKNOWLEDGMENT
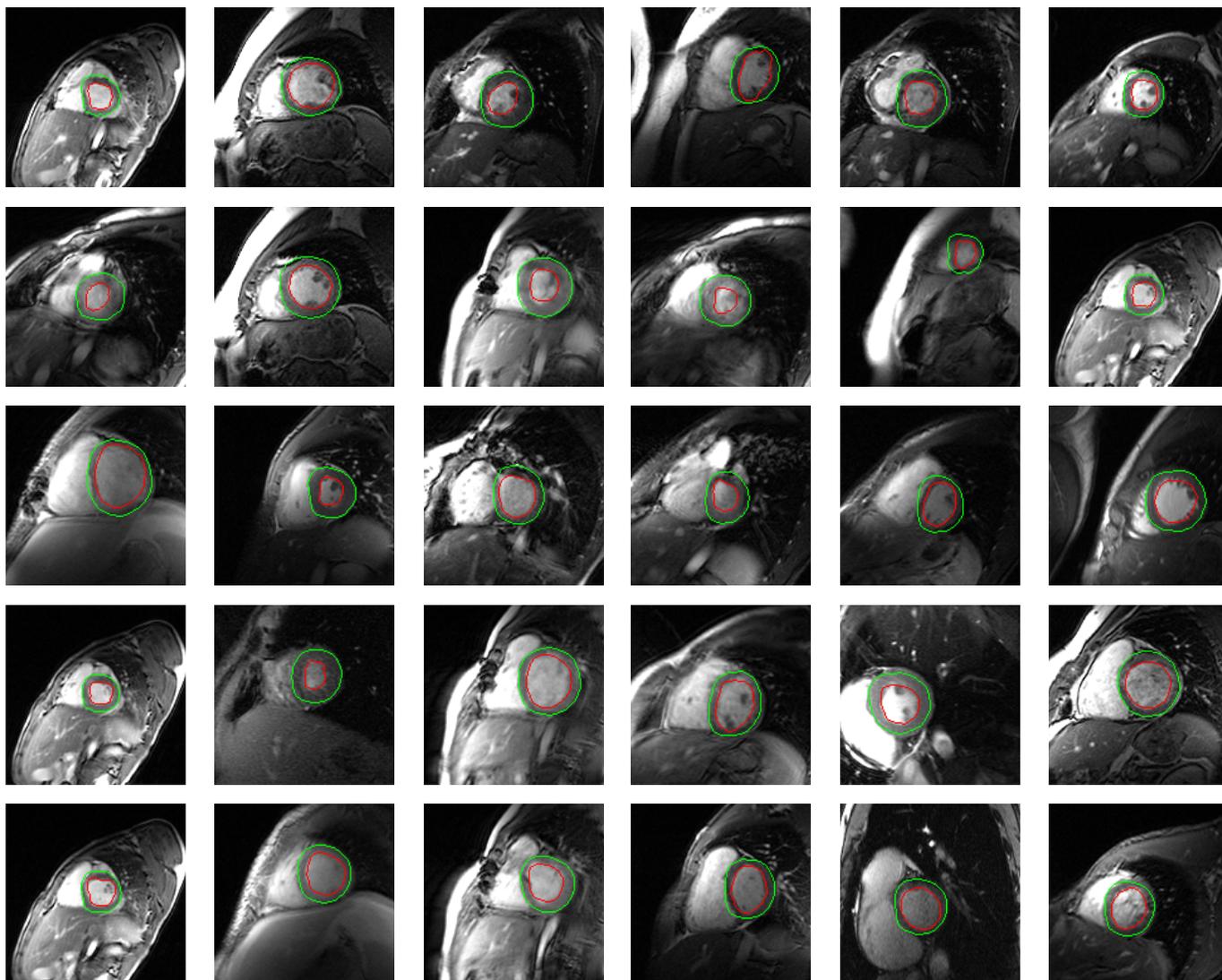
Fig. 7.   Sample Segmentations using 30 training datasets.
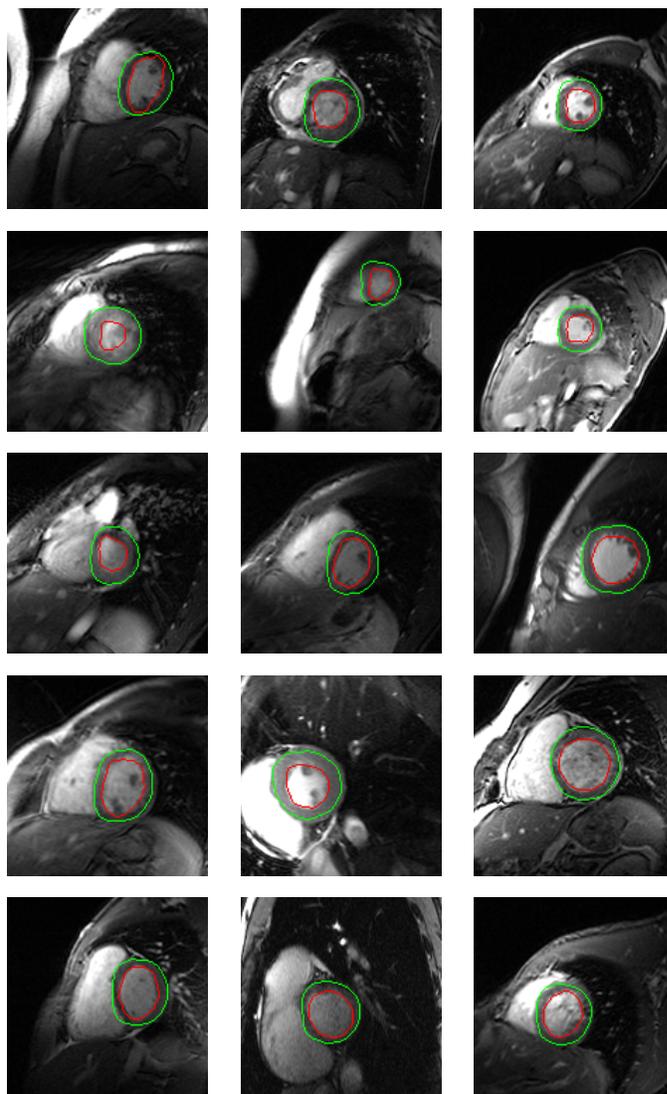
Fig. 8.   Sample Segmentations using 20 training datasets.

REFERENCES

[1] S. Allender, P. Scarborough, V. Peto, M. Rayner, J. Leal, R. Luengo-Fernandez, and A. Gray, "European cardiovascular disease statistics," *Brussels: European Heart Network*, 2008.

[2] D. Lloyd-Jones, R. Adams, M. Carnethon, G. De Simone, T. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund *et al.*, "American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Heart Disease and Stroke Statistics – 2009 Update: A Report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee," *Circulation*, vol. 119, no. 3, pp. e21–e181, 2009.

[3] A. Frangi, W. Niessen, and M. Viergever, "Three-dimensional modeling for functional analysis of cardiac images, a review," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 2–25, 2001.

[4] S. Mitchell, J. Bosch, B. Lelieveldt, R. van der Geest, J. Reiber, and M. Sonka, "3-D active appearance models: segmentation of cardiac MR and ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002.

[5] T. Cootes, G. Edwards, C. Taylor *et al.*, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[6] J. Bosch, S. Mitchell, B. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. Reiber, "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1374–1383, 2002.

[7] M. Kaus, J. Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 8, no. 3, pp. 245–254, 2004.

[8] M. Stegmann, H. Olafsdottir, and H. Larsson, "Unsupervised motion-compensation of multi-slice cardiac perfusion MRI," *Medical Image Analysis*, vol. 9, no. 4, pp. 394–410, 2005.

[9] M. Stegmann and D. Pedersen, "Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation," *Proceedings of SPIE Medical Imgage Processing*, vol. 5747, pp. 336–350, 2005.

[10] D. Fritz, D. Rinck, R. Dillmann, and M. Scheuering, "Segmentation of the left and right cardiac ventricle using a combined bi-temporal statistical model," *Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display. Proceedings of the SPIE*, vol. 6141, pp. 605–614, 2006.

[11] H. van Assen, M. Danilouchkine, F. Behloul, H. Lamb, R. Geest, J. Reiber, and B. Lelieveldt, "Cardiac LV segmentation using a 3D active shape model driven by fuzzy inference," in *Proc. MICCAI 2003*, 2003, pp. 533–540.

[12] H. van Assen, M. Danilouchkine, A. Frangi, S. Ordás, J. Westenberg, J. Reiber, and B. Lelieveldt, "SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data," *Medical Image Analysis*, vol. 10, no. 2, pp. 286–303, 2006.

[13] H. van Assen, M. Danilouchkine, M. Dirksen, J. Reiber, and B. Lelieveldt, "A 3-D active shape model driven by fuzzy inference: application to cardiac CT and MR," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 5, pp. 595–605, 2008.
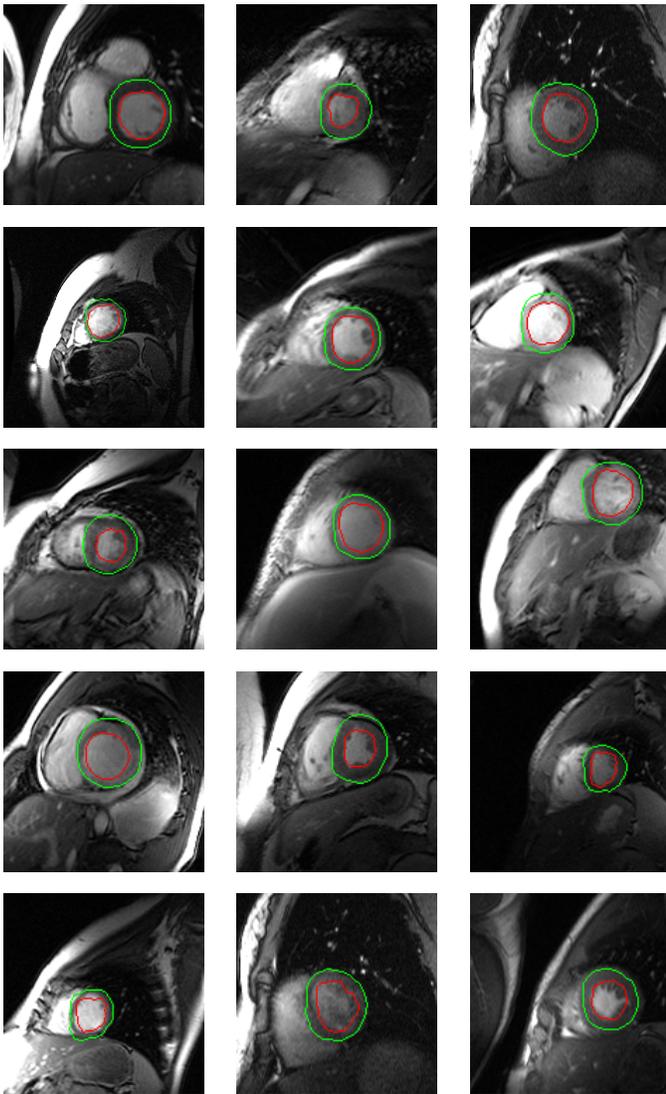
Fig. 9. Sample Segmentations using 10 training datasets.

Fig. 10. Volumetric Endocardium Regression Plots. All X and Y axes are manual volumes and automatic volumes, respectively. Both axes are measured in cm^3. Reading from left-to-right and from top-to-bottom the images correspond to evaluation using: 32, 30, 25, 20, 15 and 10 training datasets, respectively.

Fig. 11. Volumetric Epicardium Regression Plots. All X and Y axes are manual volumes and automatic volumes, respectively. Both axes are measured in cm^3. Reading from left-to-right and from top-to-bottom the images correspond to evaluation using: 32, 30, 25, 20, 15 and 10 training datasets, respectively.

[14] A. Frangi, D. Rueckert, J. Schnabel, and W. Niessen, "Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1151–1166, 2002.

[15] D. Rueckert, A. Frangi, and J. Schnabel, "Automatic construction of 3D statistical deformation models using non-rigid registration," in *Proc. MICCAI 2001*, 2001, pp. 77–84.

[16] T. Heimann and H. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, pp. 543–563, 2009.

[17] A. Andreopoulos and J. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

[18] M.-P. Jolly, H. Xue, L. Grady, and J. Guehring, "Combining registration and minimum surfaces for the segmentation of the left ventricle in cardiac cine MR images," in *Proc. MICCAI 2009*, vol. 5762, 2009, pp. 910–918.

[19] S. Zambal, J. Hladuvka, and K. Buhler, "Improving segmentation of the left ventricle using a two-component statistical model," *Lecture Notes In Computer Science*, vol. 4190, pp. 151–158, 2006.

[20] M. De Bruijne, B. Van Ginneken, M. Viergever, and W. Niessen, "Adapting active shape models for 3D segmentation of tubular structures in medical images," in *Proc. Information Processing in Medical Imaging*. Springer, 2003, pp. 136–147.

[21] M. Lorenzo-Valdés, G. Sanchez-Ortiz, A. Elkington, R. Mohiaddin, and D. Rueckert, "Segmentation of 4D cardiac MR images using a

probabilistic atlas and the EM algorithm," *Medical Image Analysis*, vol. 8, no. 3, pp. 255–265, 2004.

[22] T. Tolli, J. Koikkalainen, K. Lauerma, and J. Lötjönen, "Artificially enlarged training set in image segmentation," in *Proc. MICCAI 2006*, vol. 4190, 2006, pp. 75–82.

[23] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-D cardiac MR image analysis: left and right ventricular morphology and function," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 350–364, Feb 2010.

[24] S. Mitchell, B. Lelieveldt, R. van der Geest, H. Bosch, J. Reiver, and M. Sonka, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 5, pp. 415–423, 2001.

[25] M. Lynch, O. Ghita, and P. Whelan, "Segmentation of the left ventricle of the heart in 3-D+t MRI data using an optimized nonrigid temporal model," *IEEE Transactions on Medical Imaging*, vol. 27, no. 2, pp. 195–203, 2008.

[26] ——, "Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge," *Computerized Medical Imaging and Graphics*, vol. 30, no. 4, pp. 255–262, 2006.

[27] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 414–423, 2003.

[28] T. Cootes and C. Taylor, "Active shape models–smart snakes," in *Proc. British Machine Vision Conference*, 1992, pp. 266–275.

[29] T. Cootes, C. Taylor, D. Cooper, J. Graham *et al.*, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[30] T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," *Technical Report*, 2001.

[31] J. Lötjönen, S. Kivistö, J. Koikkalainen, D. Smutek, and K. Lauerma, "Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images," *Medical Image Analysis*, vol. 8, no. 3, pp. 371–386, 2004.

[32] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum, "Multiscale 3-D shape representation and segmentation using spherical wavelets," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 598–618, 2007.

[33] J. Liu and J. Udupa, "Oriented active shape models," *IEEE Transactions on Medical Imaging*, vol. 28, no. 4, pp. 571–584, 2009.

[34] B. Van Ginneken, A. Frangi, J. Staal, B. ter Haar Romeny, and M. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 924–933, 2002.

[35] K. Lekadir, R. Merrifield, and G. Yang, "Outlier detection and handling for robust 3-D active shape models search," *IEEE Transactions on Medical Imaging*, vol. 26, no. 2, pp. 212–222, 2007.

[36] R. P. Brent, *Algorithms for Minimization without Derivatives*. Prentice-Hall, 1973.

[37] I. Ben Ayed, L. Shuo, and R. Ian, "Embedding overlap priors in variational left ventricle tracking," *IEEE Transactions on Medical Imaging*, vol. 28, no. 12, pp. 1902–1913, 2009.