Fully Automatic Myocardial Segmentation of Contrast Echocardiography Sequence Using Random Forests Guided by Shape Model

Yuanwei Li, Chin Pang Ho, Matthieu Toulemonde, Navtej Chahal, Roxy Senior, and Meng-Xing Tang

***Abstract*— Myocardial contrast echocardiography (MCE) is an imaging technique that assesses left ventricle function and myocardial perfusion for the detection of coronary artery diseases. Automatic MCE perfusion quantification is challenging and requires accurate segmentation of the myocardium from noisy and time-varying images. Random forests (RF) have been successfully applied to many medical image segmentation tasks. However, the pixel-wise RF classifier ignores contextual relationships between label outputs of individual pixels. RF which only utilizes local appearance features is also susceptible to data suffering from large intensity variations. In this paper, we demonstrate how to overcome the above limitations of classic RF by presenting a fully automatic segmentation pipeline for myocardial segmentation in full-cycle 2-D MCE data. Specifically, a statistical shape model is used to provide shape prior information that guide the RF segmentation in two ways. First, a novel shape model (SM) feature is incorporated into the RF framework to generate a more accurate RF probability map. Second, the shape model is fitted to the RF probability map to refine and constrain the final segmentation to plausible myocardial shapes. We further improve the performance by introducing a bounding box detection algorithm as a preprocessing step in the segmentation pipeline. Our approach on 2-D image is further extended to 2-D+t sequences which ensures temporal consistency in the final sequence segmentations. When evaluated on clinical MCE data sets, our proposed method achieves notable improvement in segmentation accuracy and outperforms other state-of-the-art methods, including the classic RF and its variants, active shape model and image registration.**

***Index Terms*— Random forest, statistical shape model, contrast echocardiography, myocardial segmentation, convolutional neural network.**

Manuscript received May 19, 2017; revised August 25, 2017 and August 26, 2017; accepted August 27, 2017. Date of publication September 26, 2017; date of current version May 1, 2018. *(Corresponding author: Meng-Xing Tang.)*

Y. Li, M. Toulemonde, and M.-X. Tang are with the Department of Bioengineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: mengxing.tang@imperial.ac.uk).

C. P. Ho is with the Department of Computing, Imperial College London, London SW7 2AZ, U.K.

N. Chahal and R. Senior are with the Department of Echocardiography, Royal Brompton Hospital, London SW3 6NP, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2017.2747081

I. Introduction

MYOCARDIAL contrast echocardiography (MCE) is a cardiac ultrasound imaging tool that utilizes microbubbles as contrast agents. The microbubbles are injected intravenously and flow within the blood vessels. This can improve endocardial visualization and the assessment of left ventricle (LV) structure and function, complementing the conventional B-mode echo [1]. Furthermore, MCE can assess myocardial perfusion through the controlled destruction and replenishment of microbubbles [2]. Such perfusion information is useful for the diagnosis of coronary artery diseases (CAD) [1].

However, the analysis of MCE has been restricted to human visual assessment. Such qualitative assessment is time consuming and relies heavily on the experience of the clinician [3]. Automatic MCE quantification is desired because it is faster, more accurate and less operator-dependent. Quantification can involve the measurements of LV volumes, ejection fraction, myocardial volumes and thickness. It can also involve the assessment of myocardial perfusion by analyzing myocardial intensity changes over time. Myocardial segmentation is a widely used method that serves as an intermediate step to obtaining such MCE quantifications. However, manual segmentation is time-consuming and requires high level of expertise and training. The motivation of this paper is to develop an automatic approach for fast and accurate myocardial segmentation in MCE data. The resultant segmentation can additionally be used as input for subsequent tasks such as initialization for tracking algorithms [4].

Although there have been much work on segmentation in B-mode echocardiography, most of these methods do not work well on MCE. MCE utilizes contrast-enhanced ultrasound imaging technique to detect microbubbles by retaining nonlinear signals from microbubble oscillations while removing other linear signals from tissue. This is fundamentally different from B-mode echo which simply reflects and captures the linear tissue signals. This results in very different image appearances between MCE and B-mode echo. Specifically, one has to consider the following challenges when performing segmentation on MCE data.

- Intensity variations in the image due to ultrasound speckle noise, shadowing and attenuation artifacts, low

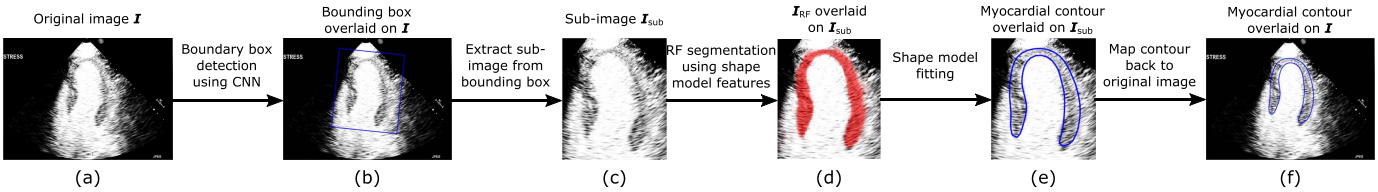

Fig. 1. Overall pipeline of our myocardial segmentation approach. The three main components of the approach are bounding box detection (a-b), RF segmentation (c-d) and shape model fitting (d-e).

signal-to-noise ratio, contrast changes over time during microbubble destruction and perfusion imaging [5].

- Geometrical variations in the pose and shape of the myocardium due to different apical chamber views, heart motion and probe motion. Each chamber view is acquired at a different probe position and captures a different 2D cross-section of the LV, resulting in variations of the 2D myocardial shape and orientation.
- Misleading intensity information such as 1) presence of structures (papillary muscle) with similar appearance to the myocardium, 2) weak image gradient information resulting in unclear myocardial border (especially epicardium).
- Speckle patterns in MCE are decorrelated due to the highly dynamic bubble signals as opposed to the static tissue speckle patterns in B-mode [6]. This poses challenges to some tracking algorithms which work by finding corresponding speckle patterns in different frames.

In this paper, we propose a fully automatic method that segments the myocardium on 2D MCE image. The method is based on random forest (RF) [7] and incorporates global shape information into the RF framework through the use of a statistical shape model that explicitly captures the myocardial shape. This provides stronger and more meaningful structural constraints that guide the RF segmentation more accurately. Our method takes the advantages of both the RF and the shape model in order to address the above challenges of MCE myocardial segmentation. The RF has strong local discriminative power and serves as a good intensity/appearance model that captures the large intensity variations of the MCE data. The shape model captures the myocardial shape variations of the MCE data and imposes a global shape constraint to guide the RF segmentation. This avoids inaccurate segmentation due to misleading intensity information on the MCE. Fig. 1 shows the overall pipeline of our myocardial segmentation approach. The main contributions of our proposed method are:

- A novel shape model (SM) feature is introduced which incorporates the shape model into the RF framework. The feature improves RF segmentation by generating smoother and more coherent RF probability map that conforms to myocardial structure and shape. (Fig. 1c-d)
- A shape model fitting algorithm is developed to fit the shape model to the RF probability map to produce a smooth and plausible myocardial contour. (Fig. 1d-e)

The work in this paper is an extension of a preliminary paper [8]. The new contributions are:

- A convolutional neural network (CNN) is employed to automatically detect a bounding box enclosing the myocardium. This pre-processing step removes any pose variations of the myocardium and improves the subsequent RF segmentation using SM features. The two-stage process of bounding box detection followed by shape inference is similar to [9]. (Fig. 1a-b)
- The approach is extended to full-cycle 2D+t MCE sequences in which a temporal constraint is imposed to ensure temporal consistency in the sequence segmentations.
- The proposed approach is evaluated on a larger dataset of 2D and 2D+t MCE data from 21 subjects.

The rest of this paper is structured as follows. In Section II, we discuss some existing approaches related to cardiac segmentation. In Section III, we describe the different components of our segmentation method and extend it to sequence segmentation. In Section IV, we describe the experiments conducted to evaluate the proposed approach. In Section V, we report the segmentation results of our approach and show that it outperforms the other state-of-the-art methods. We conclude the paper in Section VI and discuss some limitations and future directions.

II. Related Work

In this section, we give a brief overview of the segmentation methods used for cardiac ultrasound and highlight their limitations when applied to our MCE data. Current literature focus mainly on either LV segmentation [9]–[15] or myocardial segmentation [4], [16]–[19] in B-mode echocardiography. There are fewer of them addressing segmentation in MCE [3], [20]–[22].

Existing methods can be classified into two categories. The first category defines the segmentation task as a contour finding problem in which the optimal contour is found by an optimization procedure based on image information regularized by constraints applied on the contour. Two representative approaches in this category are the active contour [23] and active shape model (ASM) [24]. Active contour uses a parametric representation of the curves regularized by geometrical constraints. It is used to detect endocardial border in short-axis B-mode echo by Mishra et al [10] and Mignotte and Meunier [11]. Malpica *et al.* [21] perform myocardial segmentation in MCE sequences using a coupled active contour which imposes distance constraints between the epi and endocardial contours guided by inter-frame motion estimation derived from optical flow. However, active contour approaches depend on edge information which is unreliable on MCE image. Unclear epicardial border and the presence of papillary muscle often lead to spurious contour estimation.

Level sets adopted by Lin *et al.* [12] combines edge and region information for LV segmentation. However, region information is adversely affected by the large intensity variations in MCE. A recent paper by Pedrosa *et al.* [18] uses the B-spline explicit active surfaces framework for 3D B-mode myocardial segmentation and they compare different ways of coupling epi and endocardial segmentations.

ASM is another widely used approach in cardiac segmentation. It builds a statistical shape model from a set of training shapes by Principal Component Analysis (PCA). The model is then deformed to fit an image in ways similar to the variations observed in the training set. Butakoff *et al.* [19] uses an automatically constructed ASM for myocardial segmentation in 3D B-mode echo. Pickard *et al.* [22] uses ASM for MCE myocardial segmentation and they apply a specialized gradient vector flow field to increase its edge capture range. Bosch *et al.* [13] extends ASM by using a joint model of shape, appearance and motion to detect time-continuous LV endocardial contours over time-normalized 2D+t B-mode sequences. The advantage of ASM is that it can account for the myocardial shape variations in our MCE data and provide shape prior to help in cases where the myocardial border is unclear. However, the linear intensity model used by ASM is insufficient to characterize the appearance of MCE data which exhibit huge intensity variations. Therefore, a more powerful non-linear intensity model is required. Furthermore, some of the above methods require good manual initialization of the contour to prevent the solution from getting stuck in poor local minima.

The second category uses machine learning techniques to solve the segmentation task which is often cast as a pixel-wise classification problem. For example, Binder *et al.* [25] use an artificial neural network for pixel classification while Xiao *et al.* [26] combine maximum a posteriori and Markov random field methods for region-based segmentation. For the purpose of this paper, we focus on the RF [7] as the classifier due to its accuracy and computational efficiency. RF has been successfully applied to various segmentation tasks in the medical imaging field [27]–[29].

RF is well-suited for our MCE data because it effectively builds a non-linear appearance model based on local intensity regions. Therefore, it can better cope with the large intensity variations of the MCE data compared to the simple linear intensity model of ASM. However, the classic RF that utilizes only local appearance features has some limitations. First, the intensity information in MCE may be misleading and result in inaccurate segmentation. For instance, the RF can misclassify the papillary muscle as the myocardium because they have similar intensity. The epicardial boundary is also often difficult to identify based on intensity information alone. Second, the RF classifier assigns a class label to each pixel independently. Structural and contextual relationships between the pixel labels are ignored [28], [30]. This results in segmentation with inconsistent pixel labelling which leads to unsmooth boundaries, false detections in the background and holes in the regions of interests. Lastly, RF outputs an intermediate probability map which needs to be post-processed to obtain the final segmentation.

To overcome the above limitations of the classic RF, several works have incorporated local contextual information into the RF framework. Lempitsky *et al.* [17] demonstrates promising myocardial delineation on 3D B-mode echo by using image pixel coordinates as position feature for the RF in order to learn the shape of the myocardium implicitly. Verhoek *et al.* [4] further extends the method by using optical flow to propagate the single-frame RF segmentation for sequence tracking. Montillo *et al.* [28] introduces the entangled RF which uses intermediate probability predictions from higher levels of the tree as features for training the deeper levels. Kontschieder *et al.* [30] introduces the structured RF that incorporates structural relationships into the output predictions by predicting structured class labels for a patch region rather than an independent class label for each individual pixel. This is used by Oktay *et al.* [15] to extract a new boundary representation of the 3D B-mode echo data which is then utilized in a multi-atlas framework for LV segmentation. Milletari *et al.* [14] employs a Hough forests with implicit shape and appearance priors for the simultaneous detection and segmentation of LV in 3D B-mode echo. The Hough forest votes for the location of the LV centroid and the contour is subsequently estimated by using a code-book of surface patches associated with the votes. All the above works use contextual information to improve the RF segmentation. But they only impose weak structural constraint locally and do not explicitly learn the myocardial shapes from the MCE data.

III. Method

The overall pipeline for our myocardial segmentation method is summarized in Fig. 1. Given an input MCE image $\boldsymbol{I}$, a bounding box enclosing the myocardium is first detected using a CNN. Using the bounding box, a sub-image $\boldsymbol{I}_{\text{sub}}$ is cropped out from $\boldsymbol{I}$ and then rescaled to a fixed size. Next, an RF classifier with SM feature is used to predict a myocardial probability map $\boldsymbol{I}_{\text{RF}}$ for the sub-image. A statistical shape model is then fitted to $\boldsymbol{I}_{\text{RF}}$ to give a final myocardial contour which is subsequently mapped back to the original image space. For sequence segmentation, an additional constraint term is added to the shape model fitting step to ensure temporal consistency in the segmentations.

A. Bounding Box Detection

As we will show later (Section V-A), our SM features only work well on RF input images which do not contain significant pose variations (translation, scaling, rotation) in the myocardium. This requires the rigid alignment of myocardium in RF input images. This can be done by image registration [31] but the running time can be long. To this end, we employ a CNN to automatically detect a bounding box containing the myocardium (Fig. 1b). The bounding box is then used to extract a sub-image from the original image and the sub-image is rescaled to a fixed size. This ensures the RF input image $\boldsymbol{I}_{\text{sub}}$ is free from pose variations in the myocardium. CNN has been proven to be a good method in many object detection tasks [32]–[34] and it is also computationally efficient when implemented on GPUs. In our case,

TABLE I
CNN ARCHITECTURE FOR BOUNDING BOX DETECTION. C AND FC DENOTE CONVOLUTIONAL LAYER AND FULLY-CONNECTED LAYER RESPECTIVELY. KERNEL PARAMETERS ARE GIVEN BY [KERNEL HEIGHT × KERNEL WIDTH × NUMBER OF KERNELS / STRIDE]

Layer	Kernel parameters	Layer	Kernel parameters
1. C1	$[11 \times 11 \times 96 / 4]$	7. C5	$[3 \times 3 \times 256 / 1]$
2. Max pooling and LRN		8. Max pooling	
3. C2	$[5 \times 5 \times 256 / 1]$	9. FC6	$[1 \times 1 \times 4096 / 1]$
4. Max pooling and LRN		10. FC7	$[1 \times 1 \times 4096 / 1]$
5. C3	$[3 \times 3 \times 384 / 1]$	11. FC8	$[1 \times 1 \times 5 / 1]$
6. C4	$[3 \times 3 \times 384 / 1]$		

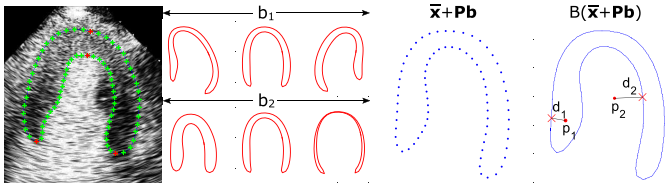

Fig. 2. (a) Manual annotations showing key landmarks in red and other landmarks in green. (b) First and second modes of variation of the shape model with b_i varying in the range of $\pm 2\sqrt{\lambda_i}$. (c) SM feature computation. Left: Landmarks x generated randomly by the shape model in (1) (*blue dots*). Right: SM feature values d_1 and d_2 measure the signed shortest distance from boundary B (*blue contour*) to pixel p_1 and p_2 respectively. d_1 is positive and d_2 is negative.

CNN can automatically learn a hierarchy of low to high level features to predict the myocardial pose in an MCE image accurately.

Similar to [9], we define the bounding box $\boldsymbol{B}$ using 5 parameters $(B_x, B_y, B_w, B_h, B_\theta)$ which represent its centroid (B_x, B_y), size (B_w, B_h) and orientation B_θ respectively. Bounding box detection is cast as a regression problem using CNN in which the values of the bounding box parameters are predicted. We use the CaffeNet architecture which is a slight modification of the AlexNet [35]. Table I shows the network architecture comprising 5 convolutional layers and 3 fully-connected layers together with the kernel parameters for each layer. Max pooling is always performed using a 3×3 kernel with stride $= 2$. Rectified linear units (ReLU) is the activation function applied after all the convolutional layers. Local response normalization (LRN) is performed after the max pooling operations at layer 2 and 4. Dropout layer is added after the fully-connected layers, FC6 and FC7. The final classification layer in the original architecture is replaced with a regression layer FC8 which contains 5 units that correspond to the 5 bounding box parameters. The network is trained using l_2 Euclidean loss between the predicted and ground truth bounding box parameters. During training, the original 460×643 image $\boldsymbol{I}$ is downsampled to a size of 256×358 pixels. Random cropping (227×227 pixels) and rotation (within $\pm 10°$) are then performed to prepare the training samples which are presented to the network in mini-batches of 32. The network is initialized with random weights from $(\mu, \sigma) = (0, 0.01)$. Optimization is performed using Adam algorithm with learning rate $= 10^{-4}$, momentum $= 0.9$ and weight decay $= 5 \times 10^{-4}$. The training is run for 18000 iterations. During testing, we extract 55 crops at uniform intervals from the downsampled test image and combine their predictions to give the final predicted bounding box. A sub-image is then extracted using this bounding box and it is rescaled to a size of 242×208 pixels to be used as input to the RF.

B. Shape Model Construction

In this section, we explain the construction of a statistical shape model of the myocardium which is required for the subsequent stages of our segmentation pipeline. Statistical shape model is extremely useful for representing objects with complex shapes. It captures plausible shape variations while removing variations due to noise. The 2D myocardial shape can be effectively represented using the point distribution model [24] which uses a set of landmarks to describe the shape. The model is built from a set of training shapes using PCA which captures the correlations of the landmarks among the training set. Each training shape is defined by a set of M landmarks. It consists of 4 key landmarks with the other landmarks uniformly sampled in between (Fig. 2a). The key landmarks are the two apexes on the epicardium and endocardium and the two endpoints on the basal segments. The point distribution model is given by:

$$\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b} \tag{1}$$

where $\boldsymbol{x}$ is a $2M$-dimensional vector $(x_1, y_1, ..., x_M, y_M)$ containing the x,y-coordinates of the M landmarks, $\bar{\boldsymbol{x}}$ is the mean landmark coordinates of the training shapes, $\boldsymbol{P} = (\boldsymbol{p}_1|\boldsymbol{p}_2|...|\boldsymbol{p}_K)$ contains K eigenvectors of the covariance matrix and each $\boldsymbol{p}_i$ is associated with its eigenvalue λ_i, $\boldsymbol{b}$ is a K-dimensional vector containing the shape parameters where each element b_i is bounded between $\pm s\sqrt{\lambda_i}$ to ensure that only plausible myocardial shapes are produced. s is the number of standard deviation from the mean shape. The value of K can be chosen such that the model can explain a required percentage p of the total variance present in the training shapes. The shape model is built from manual annotations represented in the cropped coordinate space of $\boldsymbol{I}_{\text{sub}}$. Since images extracted from bounding box are already corrected for pose variations, we do not need to rigidly align the training shapes prior to PCA. Fig. 2b shows the first and second modes of variation of our myocardial shape model.

C. RF With SM Feature

Myocardial segmentation using RF can be treated as a pixel-wise binary classification problem [17]. The RF takes the sub-image $\boldsymbol{I}_{\text{sub}}$ from Section III-A as input and predicts the class label (myocardium or background) of every pixel in that sub-image based on a set of features. An RF is an ensemble of decision trees. During training, each split node of the tree learns a binary feature test that best splits the training pixels into its two child nodes by maximizing the information gain. The splitting continues recursively until the maximum tree

depth is reached or the number of training pixels falls below a minimum. At this time, a leaf node is created and the class label distributions of the training pixels reaching that node is stored and used to predict the class probability of future unseen test pixels. During testing, the predictions from all the trees are averaged to give a final myocardial probability map.

In the classic RF, the binary feature test is based on appearance/intensities of local image regions. This is either the mean pixel intensity of a displaced feature box from the reference pixel or the difference of the mean pixel intensities of two such feature boxes [36]. We introduce an additional novel SM feature [8] that is derived from the shape model constructed in Section III-B). An SM feature is constructed by randomly sampling some values of $\boldsymbol{b}$ to generate a set of landmarks $\boldsymbol{x}$ using (1) (Fig. 2c left). A closed myocardial boundary B is then formed from the landmarks through linear interpolation. The SM feature is then given by the signed shortest distance d from the reference pixel $\boldsymbol{p}$ to the myocardial boundary B (Fig. 2c right). d is positive if $\boldsymbol{p}$ lies inside B and negative otherwise. The SM feature is in fact a signed distance transform of the myocardial boundary generated by the shape model. Each SM feature is defined by the shape parameters $\boldsymbol{b}$. During training, an SM feature is constructed by random uniform sampling of each b_i values in the range of $\pm s\sqrt{\lambda_i}$. The binary SM feature test, parameterized by $\boldsymbol{b}$ and a threshold τ, is written as:

$$t_{\mathrm{SM}}^{\boldsymbol{b},\tau}(\boldsymbol{p}) = \begin{cases} 1, & \text{if } \; D(\boldsymbol{p}, B(\bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b})) > \tau \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $D(.)$ is the function that computes d, B is the function that converts the landmarks to a closed boundary by linear interpolation. Depending on the binary test outcome, pixel $\boldsymbol{p}$ will either go to the left (1) or right (0) child node. During training, the RF learns the values of $\boldsymbol{b}$ and τ that best split the training pixels at the split node. The SM features explicitly impose a global shape constraint in the RF framework. The random sampling of $\boldsymbol{b}$ also allows the RF to learn plausible shape variations of the myocardium.

D. Shape Model Fitting

The RF probability map is an intermediate output that cannot be used directly for subsequent analysis and applications. Simple post-processing on the probability map such as thresholding and edge detection produce noisy segmentations with false positives and incoherent boundaries due to the pixel-based nature of the RF classifier. In this section, we again make use of the shape model by fitting it to the RF probability map [8]. This generates a final closed myocardial contour that is smooth and coherent. The shape model allows only plausible myocardial shape which improves the segmentation accuracy by imposing shape constraints that correct for some of the misclassifications made by the RF.

Shape model fitting is formulated as an optimization problem where we want to find the optimal values of the shape and pose parameters $(\boldsymbol{b}, \boldsymbol{\theta})$ such that the shape model best fit the RF probability map under some shape constraints. That is,

$$\min_{\boldsymbol{b},\boldsymbol{\theta}} \|\boldsymbol{I}_{\mathrm{RF}} - \boldsymbol{I}_{\mathrm{M}}(T_{\boldsymbol{\theta}}(\bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b}))\|^2 + \alpha \frac{1}{K} \sum_{i=1}^{K} \frac{|b_i|}{\sqrt{\lambda_i}}$$
$$\text{subject to} - s\sqrt{\lambda_i} < b_i < s\sqrt{\lambda_i}, \quad i = 1, \ldots, K. \tag{3}$$

The first term compares how well the model matches the RF probability map $\boldsymbol{I}_{\mathrm{RF}}$. $\boldsymbol{I}_{\mathrm{M}}(.)$ is a function which computes a binary mask from a set of landmarks generated by the shape model. This allows us to evaluate a dissimilarity metric between the RF probability map and the model binary mask by simply taking their sum-of-squared differences. The second term is a regularizer which imposes some shape constraints by keeping b_i's small so that the final shape does not deviate too much away from the mean shape. This term is also related to the probability of the given shape [37]. α controls the weighting given to this regularization term. Another shape constraint is imposed on the objective function by setting the upper and lower bounds of b_i to $\pm s\sqrt{\lambda_i}$ so that it can only vary within reasonable range similar to that of the shape model. The optimization is carried out using the pattern search algorithm [38] since the objective function is not differentiable due to the difficulty of representing the derivative of the $\boldsymbol{I}_{\mathrm{M}}$ function in mathematical form. The algorithm carries out global optimization which can be handled easily due to the small problem size (small number of shape and pose parameters). At the start of the optimization, we initialize each b_i and θ_i to zero.

E. Sequence Segmentation

We extend the above segmentation method for single 2D MCE image to 2D+t MCE sequences. The proposed method introduces temporal consistency to sequence segmentation by ensuring that the segmentation of the current frame does not differ too much from that of the previous frame. Specifically, an additional temporal constraint term is added to (3) in the shape model fitting step. The new objective function to minimize becomes

$$\min_{\boldsymbol{b},\boldsymbol{\theta}} \|\boldsymbol{I}_{\mathrm{RF}} - \boldsymbol{I}_{\mathrm{M}}(T_{\boldsymbol{\theta}}(\boldsymbol{x}))\|^2 + \alpha \frac{1}{K} \sum_{i=1}^{K} \frac{|b_i|}{\sqrt{\lambda_i}} + \beta \frac{1}{2M} \left\|\boldsymbol{x} - \boldsymbol{x}_{prev}\right\|^2$$
$$\text{subject to} \; - s\sqrt{\lambda_i} < b_i < s\sqrt{\lambda_i}, \quad i = 1, \ldots, K. \tag{4}$$

where $\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b}$ is the predicted landmark coordinates of the current frame and $\boldsymbol{x}_{prev}$ is the predicted landmark coordinates of the previous frame. The last term in (4) is the temporal constraint which computes the sum-of-squared differences between the landmark positions of the two adjacent frames. The term makes use of the segmentation from the previous frame as a reference and penalizes any segmentation of the current frame which deviates too much away from the reference. The approach uses the previous segmentation as a guide for subsequent segmentation and this ensures the myocardial segmentations throughout the sequence transit smoothly in time. The temporal term is normalized by the number of landmarks M and its influence is controlled by the weighting β. The segmentation of the previous frame is used as initialization for the current frame during optimization.

IV. Experiments

A. Data and Annotations

Our dataset consists of healthy subjects and CAD patients who are defined as those demonstrating ≥70% luminal diameter stenosis of any major epicardial artery by qualitative coronary angiography. There are a total of 21 subjects of which 10 did not demonstrate CAD, 5 had single-vessel disease and 6 had multi-vessel disease. MCE exams from these 21 subjects were used in this paper. The data were acquired using a Philips iE33 ultrasound machine (Philips Medical Systems, Best, Netherlands) and SonoVue (Bracco Research SA, Geneva, Switzerland) as the contrast agent. For each of the 21 exams, MCE sequences in the apical 2, 3 and 4-chamber views are acquired. From these sequences, we randomly selected 242 2D MCE images to form $Dataset1$ which comprises an approximately equal proportion of the three different apical views and different cardiac phases (end-systole (ES) or end-diastole (ED)). $Dataset1$ is split into a training and a test set with a ratio of 14:7 on a patient basis and a ratio of 159:83 on an image basis. In addition, we chose 6 out of the 7 test subjects in $Dataset1$ and from these 6 subjects, we randomly chose 12 2D+t MCE sequences (2 for each subject) which again comprises an equal proportion of the three apical views. We then randomly select and crop out one cardiac cycle from each of these 12 sequences to form $Dataset2$. On average, each temporally cropped sequence in $Dataset2$ consists of 22±4 frames and captures one complete cardiac cycle from ES to ES.

Manual annotations for the two datasets are done by an expert. For $Dataset1$, each image is manually annotated with a bounding box which encloses the myocardium. This is used for training and testing the CNN. A myocardial contour is also manually delineated for each image in $Dataset1$ and for every image of the 12 sequences in $Dataset2$. This is used for training and testing the RF. Subsequently, 4 key landmarks are manually identified on the myocardial contour as illustrated in Fig. 2a. 18 landmarks are uniformly sampled in between each pair of key landmarks to give a total of $M = 76$ landmarks that define each myocardial shape. The landmark annotations are used for shape model construction.

B. Evaluation Metrics

Segmentation accuracy is evaluated quantitatively using Jaccard index which measures the overlap between two contours, mean absolute distance (MAD) which measures the average point-to-point distance between two contours and Hausdorff distance (HD) which measures the maximum distance between two contours. In addition, clinical indices such as endocardial and myocardial areas are also computed from the segmentations and compared to the ground truth in terms of correlation, bias and standard deviation. Correlation and Bland-Altman plots are also presented. Paired t-test is used to test for significant differences at 5% significance level.

C. Implementation Details

The CNN for the bounding box detection is implemented in Caffe [39] and runs on a machine with one NVIDIA GeForce GTX 950 GPU. The parameters used for the CNN are described in Section III-A.

The shape model is constructed with parameter $K = 16$ $p = 99\%$ and $s = 2$. The same shape model is used for both the SM feature and the shape model fitting. For the SM feature, we observe that the segmentation results are insensitive to K, p and s. For the shape model fitting, these parameters have more significant influence on the results. Hence, we only tune their values for the shape model fitting and use the same values for the SM feature. To this end, only a single shape model needs to be constructed for the two tasks, making the approach more robust and generalizable.

For the RF, we use 20 trees with a maximum tree depth of 24. Further increasing the number of trees in the forest and the depth of each tree adds to the computational cost with no significant improvement in segmentation accuracy. All training images for the RF are pre-processed using histogram equalization to reduce intensity variations between different images. 10% of the pixels from the training images are randomly selected for tree training. For the shape model fitting, we empirically set $\alpha = 3000$ and $\beta = 10$ via cross-validation.

Unless otherwise stated, 3-fold cross-validation is applied on the training set of $Dataset1$ to optimize the above parameters. Once the optimal parameters are found, the entire training set of $Dataset1$ is used to learn a CNN model, an RF model and a shape model. Using these models, testing is then performed on 1) the test set of $Dataset1$ and 2) the entire $Dataset2$.

Using a machine with Intel Core i7-4770 at 3.40 GHz and 32 GB of memory, RF training takes 7.1 minutes for 1 tree. Given a test image, RF segmentation takes about 25.5s using 20 trees but tree prediction can be parallelized so that it takes 1.3s per tree. The bounding box detection takes 0.1s and the shape model fitting takes 7.8s. In total, our fully automatic algorithm takes around 9.2s to segment one image.

D. Comparisons With Other State-of-the-Art Approaches

We compare our proposed approach to ASM for static segmentation on $Dataset1$ and to image registration and optical flow for sequence segmentation on $Dataset2$.

We use a modified ASM [40] which selects a set of optimal features for the appearance model around each landmark point instead of using the Mahalanobis distance. The parameters for shape model construction are $K = 16$ and $M = 76$. The length of landmark intensity profile is 6 pixels. The search length for each landmark is 2 pixels and the number of search iterations is 40. The shape constraint is limited to $\pm 1.5 \times \sqrt{\lambda_i}$ and two resolution levels are used for matching. The ASM is initialized by manually placing the contour near the myocardium.

The non-rigid image registration is based on B-spline free-form deformations [31]. The error metric used is the sum-of-squared difference and two resolution levels are used. Smoothness penalty is set to 0.003 and B-spline control point spacing is set to 32 pixels. For each MCE sequence, the first frame is registered to all the other frames. A manual segmentation is performed on the first frame and it is propagated to the other frames in the sequence through the transformation fields found by registration.

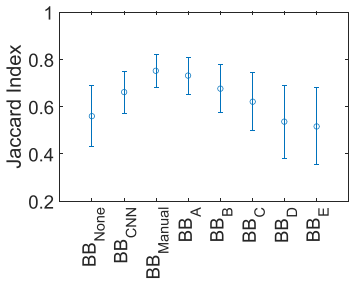

Fig. 3. Effect of bounding box detection on RF segmentation. Error bar denotes the standard deviation over Dataset1.

TABLE II
DETECTION ERROR OF THE BOUNDING BOX PARAMETERS ESTIMATED BY CNN AGAINST THE MANUAL GROUND TRUTH

Detection error	Mean±SD
δB_x (mm)	0.25±3.03
δB_y (mm)	-0.62±2.38
δB_w (mm)	0.45±4.83
δB_h (mm)	-0.15±5.15
δB_θ (°)	0.08±3.71

Optical flow is based on the algorithm described in [41]. It is a variational model based on the gray level constancy and the gradient constancy assumptions, together with a smoothness constraint. The weighting for the smoothness term is set to 0.05. A multi-resolution approach is also used with a downsampling ratio set to 0.75 and the width of the coarsest level set to 10 pixels. Optical flow motion fields are computed between consecutive frame pairs in the sequence and are used to propagate the manual segmentation on the first frame to all the other frames.

The above methods all require some manual user inputs as initialization and the algorithm parameters are optimized using grid search.

V. RESULTS

We evaluate different components of our segmentation method separately in the four sections below. Sections V-A to V-C evaluate the bounding box detection, the SM feature and the proposed static segmentation algorithm respectively using $Dataset1$. Section V-D evaluates the sequence segmentation method using $Dataset2$.

A. Bounding Box Detection

In this experiment, we evaluate the accuracy of the CNN bounding box detection algorithm against the manual ground truth. We measure the detection error as $\boldsymbol{B}_{detected} - \boldsymbol{B}_{groundtruth} = \delta\boldsymbol{B} = (\delta B_x, \delta B_y, \delta B_w, \delta B_h, \delta B_\theta)$. Table II shows the bias and standard deviation of the error for each bounding box parameter. All parameters show mean values close to zero, indicating small systematic bias. Localization uncertainty of (B_w, B_h) is higher than that of (B_x, B_y). This indicates the CNN is less accurate in determining the scale of the bounding box compared to its position. CNN also has high localization precision for predicting the bounding box orientation (B_θ). Qualitatively, the top row of Fig. 7 shows the manual and CNN-detected bounding boxes on some example MCE images.

In the next experiment, we study the effect of bounding box on the RF probability map. Specifically, we analyze three cases in which different input images are used for RF training and testing. BB_{None}: Original full size image; BB_{CNN}: Sub-image cropped from the bounding box detected by CNN; $\text{BB}_{\text{Manual}}$: Sub-image cropped from the manual bounding box; The RF probability maps obtained for these three cases are evaluated against the ground truth segmentations using Jaccard index as shown in Fig. 3. Our SM feature works under the assumption that the shape of interest in all the RF input images are aligned to a reference. The bounding box effectively performs this alignment by cropping out a sub-image that is free from any myocardial pose variations which leads to more accurate RF segmentation. BB_{None} has the worst segmentation accuracy because it does not account for any pose variations. BB_{CNN} improves the segmentation accuracy but does not perform as well as $\text{BB}_{\text{Manual}}$. This is due to possible inaccuracy in the CNN-detected bounding box and pose variations may not be completely removed. However, BB_{CNN} is fully-automatic while $\text{BB}_{\text{Manual}}$ requires the manual annotation of bounding box which makes the overall segmentation method semi-automatic. An additional advantage of the bounding box is that it removes irrelevant image regions and reduces the image size so that subsequent RF segmentation and shape model fitting is faster.

To further investigate the dependence of RF segmentation results on the bounding box detected, we conduct the following experiment. We perturb each ground truth bounding box $\boldsymbol{B}_{groundtruth}$ in the set $\text{BB}_{\text{Manual}}$ by adding a random error $\delta\boldsymbol{B}$ to it. The error of each bounding box parameter δB_i, where $i = x, y, w, h, \theta$, is sampled randomly from a zero-mean normal distribution with standard deviation σ_{B_i}. By setting different σ values, we create 5 sets of bounding boxes BB_{A}-BB_{E} with increasing amount of errors introduced to the ground truths. Specifically, we set $\sigma_{\boldsymbol{B}} = (\sigma_{B_x}, \sigma_{B_y}, \sigma_{B_w}, \sigma_{B_h}, \sigma_{B_\theta})$ for BB_{A} and set multiples of it, $2\sigma_{\boldsymbol{B}}, 3\sigma_{\boldsymbol{B}}, 4\sigma_{\boldsymbol{B}}, 5\sigma_{\boldsymbol{B}}$, for BB_{B}, BB_{C}, BB_{D} and BB_{E}. σ values of each bounding box parameter are chosen so that σ values of BB_{C} are the same as the standard deviation of the CNN detection errors reported in Table II. Fig. 3 shows that RF segmentation accuracy decreases with increasing perturbations from the ground truth bounding boxes. This confirms that the final RF segmentation result is heavily dependent on the accuracy of bounding box detection. Future work should therefore be directed at improving the bounding box detection since there is still a significant gap for improvement on RF segmentations between using the CNN-detected bound boxes and the manual ground truths.

B. Shape Model Feature

We study the effect of SM feature on RF segmentation by comparing it with the classic RF that uses simple intensity features [36] as well as RF that uses other contextual features such as the entanglement features [28] and position features [17]. We use the sub-images extracted from the manual bounding boxes $\text{BB}_{\text{Manual}}$ to train and test all the RFs in this section.

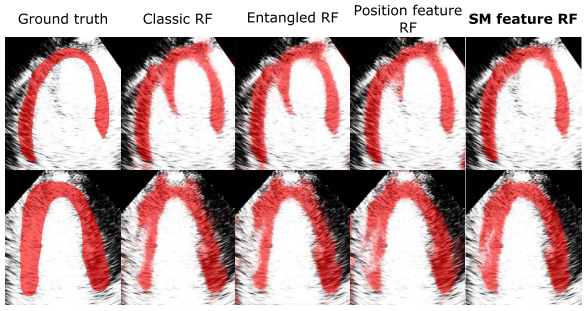

Fig. 4. Visual comparison of the RF probability maps obtained by different RF variants.

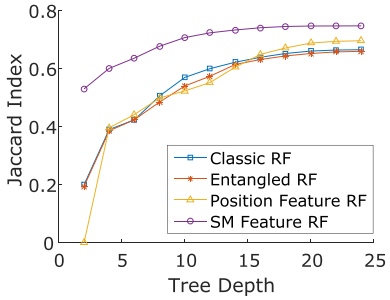

Fig. 5. Comparison of Jaccard indices of different RF variants vs tree depth.

Fig. 4 compares the probability maps generated by RFs using different features. Our SM feature produces smoother and more coherent probability map. Other RFs often misclassify low intensity structure in the LV cavity as myocardium (row 1) and the high intensity region in the myocardium as background (row 2). RF with SM feature can overcome these problems by imposing a global shape constraint.

Fig. 5 shows the segmentation performance of different RFs at different tree depths. Our SM feature RF outperforms the other RFs at all tree depths. This is because the SM feature captures the explicit geometry of the myocardium to guide the segmentation. A binary test of the SM feature in the split node partitions the image space using meaningful myocardial shape based on the shape model. Position feature [17] also learns the myocardial shape implicitly but its binary test partitions the image space using simple straight line which is less effective. To learn complex shape like the myocardium, many more position feature tests are needed compared to SM feature tests which learn the myocardial shape directly. This causes SM feature to be the more discriminative feature at lower levels of the tree, leading to better results than other RFs especially at lower tree depths.

C. Static Segmentation on Dataset1

We compare the following different approaches for the evaluation of static segmentation on *Dataset*1.
ASM: Active shape Model [40]
M1: BB_{CNN}+ SM feature RF + Canny edge detector
M2: BB_{CNN}+ SM feature RF + Shape model fitting
M3: BB_{CNN}+ SM feature RF + Shape model fitting (View specific)
M4: BB_{Manual}+ SM feature RF + Shape model fitting

Fig. 6a presents the segmentation accuracy results (Jaccard, MAD and HD) and Table III reports the evaluation on clinical indices such as endocardial area and myocardial area. ASM does not perform well because it uses a simple intensity profile model to search for the best landmark position. This model is not adequate for noisy MCE images. RF can provide a much stronger and discriminative intensity model. In M1, we apply our segmentation method using the CNN bounding box detection algorithm (BB_{CNN}) and the RF with SM features. However, we replace the last shape model fitting step with a Canny edge detector to obtain a binary edge map as the final segmentation. This gives better results than the ASM but the final segmentation is not regularized by global shape constraint. To improve the results further, shape model fitting is added in M2 to give our fully automatic segmentation approach. It combines the local discriminative power of RF with the global shape constraint imposed by the shape model. The fitting guides the segmentation in regions where the RF probability map has low confidence predictions. It also ensures the final segmentation is a smooth and coherent contour that represents plausible myocardial shape. The bottom row of Fig. 7 shows the final myocardial contours predicted by M2. The method is able to segment the myocardium accurately even in the presence of shadowing and attenuation artifacts which result in unclear epicardial border. In M3, we trained three separate RF models and shape models for the three different apical chamber views. There is a small improvement in results since each model learns more specifically the different anatomy of each view although a general model that includes all views (M2) is also quite robust. Since there are less training data for each view model in M3, we expect the results to improve with more data. In M4, we replace the CNN bounding box in M2 with manual bounding box (BB_{Manual}). This results in a semi-automatic approach which accurately removes any myocardial pose variations and achieves the best results in the final segmentation.

D. Sequence Segmentation on Dataset2

Fig. 6b presents the segmentation accuracy of the different approaches on *Dataset*2. Our proposed approach (N1-N3) achieves significantly more accurate segmentation results than image registration (R) and optical flow (OF) methods. Image registration and optical flow perform tracking by finding corresponding speckle patterns between frames and they are based on the constant intensity assumption. They do not perform well on MCE because MCE data exhibit high intensity variations and decorrelated speckle patterns due to the highly dynamic microbubbles.

N1 is our static segmentation method which uses (3) for shape model fitting. It is the same as M2 and both RF model and shape model are trained on *Dataset*1 which consists of only ES and ED frames. N2 is exactly the same as N1 except that the training is performed on *Dataset*2 which consists of frames in all cardiac phases. Segmentation results for N2 are obtained using leave-one-out cross-validation on the 6 subjects in *Dataset*2. Although N2 is trained on all cardiac phases and is expected to have a more representative shape model,

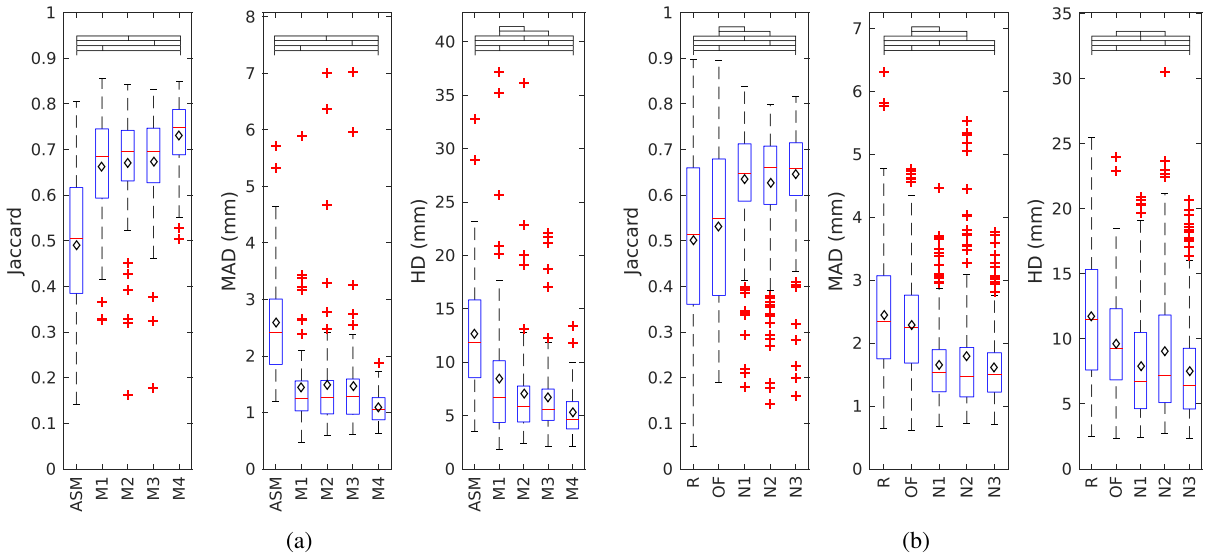

Fig. 6. Segmentation accuracy results of the various approaches for (a) *Dataset1* and (b) *Dataset2*. The red line and the black diamond marker represent the median and the mean respectively. The ends of the whiskers represent the lowest and highest data point within 1.5 times the interquartile range. Top black brackets indicate the difference between the two approaches is significant using the paired t-test.

TABLE III
CORRELATION COEFFICIENT, BIAS AND STANDARD DEVIATION BETWEEN GROUND TRUTH AND ESTIMATED CLINICAL INDICES FOR *Dataset1*. NOTE THAT IT IS NOT POSSIBLE TO OBTAIN CLINICAL INDICES FROM THE PARTIAL CANNY EDGE MAPS OF APPROACH M1

	Endocardial Area			Myocardial Area		
	corr	bias	std	corr	bias	std
	(val)	($\times 10^2$ mm^2)		(val)	($\times 10^2$ mm^2)	
ASM	0.72	-3.86	5.70	0.29	0.37	3.78
M2	0.90	-0.05	3.54	0.45	-0.04	2.71
M3	0.94	-0.22	2.89	0.51	-0.07	2.55
M4	0.99	-0.10	1.43	0.89	0.25	1.56

it is surprising to find that N1 has slightly more accurate results than N2. This could mean that the shape model trained from ED and ES frames alone is adequate for segmenting myocardial shapes over the full cardiac cycle. This is possible since the ED and ES shapes are at the extreme ends and all other shapes of the cardiac cycle can be found as intermediate transitions in between these two extremes. Another reason for the better performance of N1 could be due to N1 being trained on a more diverse dataset comprising 14 subjects while N2 is trained based on only 5 subjects during cross-validation. Increasing the number of training sequences from more subjects could potentially improve the results for N2.

N3 is our sequence segmentation method and it achieves the best results by extending N1 using (4) for shape model fitting. The temporal constraint allows N3 to produce temporally more consistent segmentations throughout the sequence with the most improvement reflected by the HD metric. The clinical indices of endocardial and myocardial areas are also computed for the automatic segmentations from N3 and compared to the manual segmentations in terms of correlation and Bland-Altman plots as shown in Fig. 8. It can be observed that

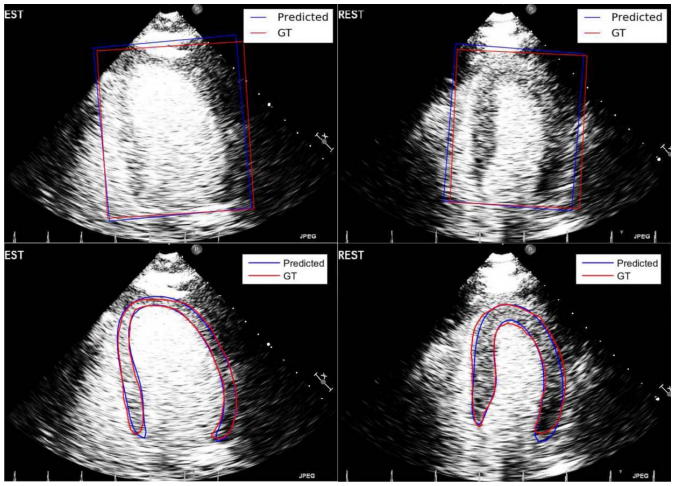

Fig. 7. Top: Visual comparison of the bounding boxes estimated by CNN (*blue*) and the ground truth (*red*). Bottom: Visual comparison of the final contours estimated by our fully automatic approach M2 (*blue*) and the ground truth (*red*).

our method results in overestimation for smaller areas and underestimation for larger areas. Fig. 9 shows the myocardial and endocardial area-time curves over one cardiac cycle. The curves are the mean across the 12 sequences in *Dataset2*. We observe that our proposed method produces curves that are similar to the ground truth. Visual segmentation results on some full MCE sequences using N3 can be found in the supplementary materials.[1]

VI. DISCUSSION AND CONCLUSION

We have presented a fully automatic approach for fast and accurate segmentation of myocardium in 2D MCE image.

[1]Downloadable supplementary materials are available at http://ieeexplore.ieee.org. This includes three AVI format movie clips which show the N3 segmentation results of three MCE sequences taken in the apical 2, 3 and 4 chamber view respectively.

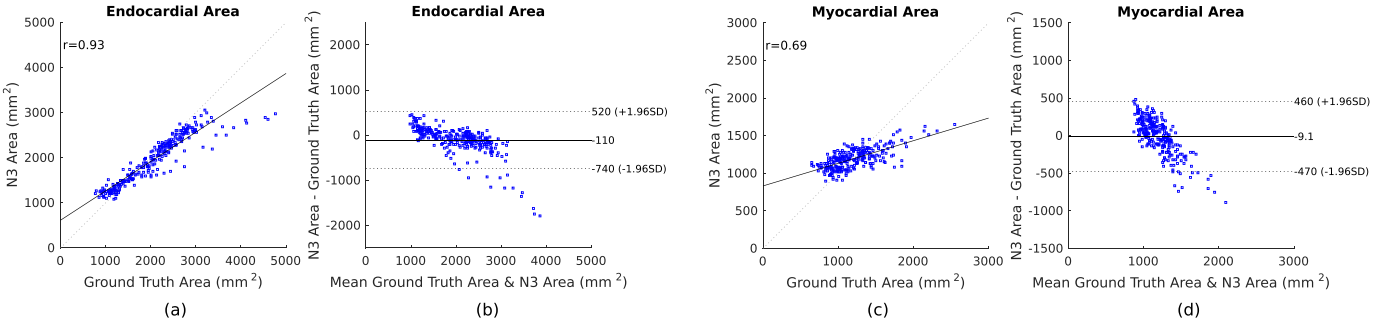

Fig. 8. Correlation plots (a,c) and Bland-Altman plots (b,d) of the endocardial and myocardial areas computed for the manual segmentations and estimated segmentations from N3. For BA plots, solid line: bias; dotted lines: limits of agreement ($\mu \pm 1.96\sigma$).

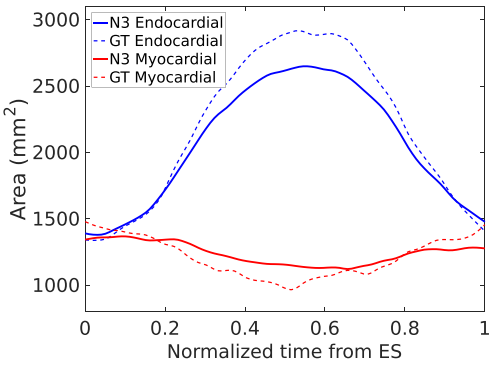

Fig. 9. Mean endocardial (blue) and myocardial (red) areas of 12 sequences over one cardiac cycle computed from ground truth (dash lines) and N3 segmentations (solid lines).

The proposed method uses a statistical shape model to guide the RF segmentation by imposing global shape constraints. This is done by first incorporating a novel SM feature into the RF framework and then fitting the shape model to the RF probability map to obtain the final myocardial contour. The SM feature outperforms other contextual features by producing more accurate RF probability map while the shape model fitting step further improves the final segmentation results by producing a smooth and coherent contour. Bounding box detection using CNN serves as an important image alignment step that improves the performance of subsequent RF segmentation using SM feature. The method is further extended to 2D+t MCE sequence which imposes temporal consistency in the final sequence segmentations. The overall segmentation method combines the advantages of both ASM and RF, and outperforms either of this method used alone.

Our proposed method is generic and can be applied to other image data containing large intensity variations where prior knowledge of shape becomes important in guiding the segmentation. Current study only limits our approach to contrast echocardiography data. As future work, we will evaluate our method on other datasets of B-mode echocardiography or even medical data of different modalities in order to test the robustness and generality of the approach. In addition, our sequence dataset *Dataset*2 is small and only based on 6 subjects. Increasing the training data on this set can allow us to train better models than the one derived from *Dataset*1 which comprises only ES and ED frames.

Extension to 3D data should also work in principle. In this case, we need to compute the shortest distance from a point to a surface for the SM feature which can increase computational cost. Shape model fitting can also take longer if 3D binary volume needs to be generated from the mesh surface of 3D shape model. Since this is the limiting step with the longest running time in our segmentation pipeline, future work will look at more efficient ways of optimizing this step in order to make our approach real-time. One way is to define the shape model on distance maps directly instead of the landmarks as described in [42]. This will save the computational cost of converting the contour into a distance map or binary mask during SM feature computation and shape model fitting.

We have added a simple temporal constraint term to induce temporally smooth sequence segmentation. More sophisticated tracking algorithms such as block matching and optical flow could be incorporated into our framework to ensure temporal consistency. This is done in [4] where the RF segmentation is propagated using optical flow. Finally, effort should also be directed at more accurate bounding box detection since we have shown that the final segmentations obtained using our approach is strongly dependent on this step.

Acknowledgment

The authors would like to thank Dr Robert Eckersley, Prof. Daniel Rueckert and other members from the BioMedIA group for their help and advice. The work of Y. Li was supported by the Imperial College President's Ph.D. Scholarships. The work of M.-X. Tang was supported by the Engineering and Physical Sciences Research Council, U.K. under Grant EP/M011933/1.

References

[1] R. Senior *et al.*, "Contrast echocardiography: Evidence-based recommendations by European association of echocardiography," *Eur. Heart J.-Cardiovascular Imag.*, vol. 10, no. 2, pp. 194–212, 2009.

[2] K. Wei, A. R. Jayaweera, S. Firoozan, A. Linka, D. M. Skyba, and S. Kaul, "Quantification of myocardial blood flow with ultrasound-induced destruction of microbubbles administered as a constant venous infusion," *Circulation*, vol. 97, no. 5, pp. 473–483, 1998.

[3] M. Ma, M. van Stralen, J. H. C. Reiber, J. G. Bosch, and B. P. F. Lelieveldt, "Left ventricle segmentation from contrast enhanced fast rotating ultrasound images using three dimensional active shape models," in *Proc. FIMH*, 2009, pp. 295–302.

[4] M. Verhoek, M. Yaqub, J. McManigle, and J. A. Noble, *Learning Optical Flow Propagation Strategies Using Random Forests for Fast Segmentation in Dynamic 2D & 3D Echocardiography*. Berlin, Germany: Springer, 2011, pp. 75–82.

[5] M.-X. Tang *et al.*, "Quantitative contrast-enhanced ultrasound imaging: A review of sources of variability," *Interface Focus*, vol. 1, no. 4, pp. 520–539, 2011.

[6] R. Kolar, R. Jirik, V. Harabis, K. Nylund, and O. H. Gilja, "Registration of ultrasound contrast images for perfusion analysis," in *Proc. IEEE Int. Ultrason. Symp.*, Sep. 2009, pp. 1251–1254.

[7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8] Y. Li, C. P. Ho, N. Chahal, R. Senior, and M.-X. Tang, "Myocardial segmentation of contrast echocardiograms using random forests guided by shape model," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Athens, Greece, Oct. 2016, pp. 158–165.

[9] S. K. Zhou, "Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram," *Med. Image Anal.*, vol. 14, no. 4, pp. 563–581, 2010.

[10] A. Mishra, P. K. Dutta, and M. Ghosh, "A GA based approach for boundary detection of left ventricle with echocardiographic image sequences," *Image Vis. Comput.*, vol. 21, no. 11, pp. 967–976, 2003.

[11] M. Mignotte and J. Meunier, "A multiscale optimization approach for the dynamic contour-based boundary detection issue," *Comput. Med. Imag. Graph.*, vol. 25, no. 3, pp. 265–275, 2001.

[12] N. Lin, W. Yu, and J. S. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Med. Image Anal.*, vol. 7, no. 4, pp. 529–537, 2003.

[13] J. G. Bosch *et al.*, "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1374–1383, Nov. 2002.

[14] F. Milletari, M. Yigitsoy, N. Navab, and S. Ahmadi, "Left ventricle segmentation in cardiac ultrasound using Hough-forests with implicit shape and appearance priors," in *Proc. MICCAI CETUS*, 2014, pp. 49–56.

[15] O. Oktay *et al.*, *Probabilistic Edge Map (PEM) for 3D Ultrasound Image Registration and Multi-atlas Left Ventricle Segmentation*. Cham, Switzerland: Springer, 2015, pp. 223–230.

[16] J. M. B. Dias and J. M. N. Leitao, "Wall position and thickness estimation from sequences of echocardiographic images," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 25–38, Feb. 1996.

[17] V. Lempitsky, M. Verhoek, A. Noble, and A. Blake, "Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography," in *Functional Imaging and Modeling of the Heart* (Lecture Notes in Computer Science), N. Ayache, H. Delingette, and M. Sermesant, Eds. Berlin, Germany: Springer, Jun. 2009, pp. 447–456.

[18] J. Pedrosa *et al.*, "Left ventricular myocardial segmentation in 3-D ultrasound recordings: Effect of different endocardial and epicardial coupling strategies," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 64, no. 3, pp. 525–536, Mar. 2017.

[19] C. Butakoff *et al.*, "Left-ventricular epi- and endocardium extraction from 3D ultrasound images using an automatically constructed 3D ASM," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 4, no. 5, pp. 265–280, 2016.

[20] G. Zwirn, R. Beeri, D. Gilon, and S. Akselrod, "Automatic endocardial-boundary detection in low mechanical-index contrast echocardiography," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2310–2322, Nov. 2006.

[21] N. Malpica, M. Ledesma-Carbayo, A. Santos, E. Pérez, M. A. García-Fernandez, and M. Desco, "A coupled active contour model for myocardial tracking in contrast echocardiography," in *Image Understanding and Analysis*. London, U.K.: Imperial College London, 2004.

[22] J. E. Pickard, J. A. Hossack, and S. T. Acton, "Shape model segmentation of long-axis contrast enhanced echocardiography," in *Proc. 3rd IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2006, pp. 1112–1115.

[23] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[24] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[25] T. Binder *et al.*, "Artificial neural networks and spatial temporal contour linking for automated endocardial contour detection on echocardiograms: A novel approach to determine left ventricular contractile function," *Ultrasound Med. Biol.*, vol. 25, no. 7, pp. 1069–1076, 1999.

[26] G. Xiao, M. Brady, J. A. Noble, and Y. Zhang, "Segmentation of ultrasound B-mode images with intensity inhomogeneity correction," *IEEE Trans. Med. Imag.*, vol. 21, no. 1, pp. 48–57, Jan. 2002.

[27] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images," *NeuroImage*, vol. 57, no. 2, pp. 378–390, 2011.

[28] A. Montillo, J. Shotton, J. M. Winn, J. E. Iglesias, D. N. Metaxas, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of CT images," in *Proc. IPMI*, 2011, pp. 184–196.

[29] J. Margeta, E. Geremia, A. Criminisi, and N. Ayache, "Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data," in *Proc. 2nd Int. Workshop Statist. Atlases Comput. Models Heart. Imag. Modelling Challenges (STACOM)*, 2011, pp. 109–119.

[30] P. Kontschieder, S. R. Bulò, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *Proc. ICCV*, 2011, pp. 2190–2197.

[31] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[32] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. ICLR*, 2014.

[33] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[34] C. Garcia and M. Delakis, "Convolutional face finder: A neural architecture for fast and robust face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1408–1423, Nov. 2004.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1106–1114.

[36] A. Criminisi *et al.*, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med. Image Anal.*, vol. 17, no. 8, pp. 1293–1303, 2013.

[37] D. Cristinacce and T. F. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognit.*, vol. 41, no. 10, pp. 3054–3067, 2008.

[38] C. Audet and J. E. Dennis, Jr., "Analysis of generalized pattern searches," *SIAM J. Optim.*, vol. 13, no. 3, pp. 889–903, 2002.

[39] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM Int. Conf. Multimedia (MM)*, Orlando, FL, USA, Nov. 2014, pp. 675–678.

[40] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.

[41] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, *High Accuracy Optical Flow Estimation Based on a Theory for Warping*. Berlin, Germany: Springer, 2004, pp. 25–36.

[42] A. Tsai *et al.*, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 137–154, Feb. 2003.