Automatic Construction of 3D-ASM Intensity Models by Simulating Image Acquisition: Application to Myocardial Gated SPECT Studies

Catalina Tobon-Gomez, Constantine Butakoff, Santiago Aguade, Federico Sukno, Gloria Moragas, and Alejandro F. Frangi, Senior Member, IEEE

Abstract—Active shape models bear a great promise for model-based medical image analysis. Their practical use, though, is underdetermined due to the need to train such models on large image databases. Automatic building of point distribution models (PDMs) has been successfully addressed and a number of autolandmarking techniques are currently available. However, the need for strategies to automatically build intensity models around each landmark has been largely overlooked in the literature. This work demonstrates the potential of creating intensity models automatically by simulating image generation. We show that it is possible to reuse a 3D PDM built from computed tomography (CT) to segment gated single photon emission computed tomography (gSPECT) studies. Training is performed on a realistic virtual population where image acquisition and formation have been modeled using the SIMIND Monte Carlo simulator and ASPIRE image reconstruction software, respectively. The dataset comprised 208 digital phantoms (4D-NCAT) and 20 clinical studies. The evaluation is accomplished by comparing point-to-surface and volume errors against a proper gold standard. Results show that gSPECT studies can be successfully segmented by models trained under this scheme with subvoxel accuracy. The accuracy in estimated LV function parameters, such as end diastolic volume, end systolic volume, and ejection fraction, ranged from 90.0% to 94.5% for the virtual population and from 87.0% to 89.5% for the clinical population.

Index Terms—Automatic model building, gated SPECT, intensity model, NCAT, SIMIND, 3D active shape model (ASM).

Manuscript received July 20, 2008; revised July 28, 2008. Current version published October 24, 2008. This work was supported in part within the framework of the CENIT-CTEAM Project funded by the Spanish CDTI-MITYC and in part by MEC TEC2006-03617 and ISCIII FIS2004/40676. Asterisk indicates corresponding author.

* A. F. Frangi is with the Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Universitat Pompeu Fabra, Barcelona 08003, Spain, and also with the Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), 08003 Barcelona, Spain (e-mail: alejandro.frangi@upf.edu).

I. INTRODUCTION

IN SPITE of the high technological developments in medical imaging systems for diagnostic cardiology, cardiac function is still mostly analyzed through visual assessment or manual delineation, which are time consuming, subjective, and error prone. This fact has generated the need for automated analysis tools to support diagnosis with reliable and reproducible image interpretation. However, the success of currently available commercial packages is modest and their use under-diffused.

On the one hand, automated delineation of the cardiac chambers from 3D and 4D image datasets is challenging. Recent surveys have pointed out the prevalence of model-based approaches to accomplish this task [1], [2]. Typically, they require a generic template which undergoes adaptation to fit specific image data. This strategy enables introducing a priori knowledge of the structure of interest into the segmentation process. In particular, active shape models (ASMs) [3] have been successfully employed in image segmentation [4], [5]. Unfortunately, construction of these models requires several training steps based on a target image database (ideally a rather extensive one). This is simply unachievable by sole manual processing on 4D datasets due to the huge amount of data involved. These steps include: (1) manual outlining of target boundaries, (2) consistent distribution of landmarks across sample shapes, (3) statistical shape decomposition yielding a point distribution model (PDM) [3], and (4) learning a statistical model of the intensity around the target object. Substantial efforts have been carried out to automatically construct PDMs by autolandmarking surface [6]–[8] or volumetric [9], [10] representations of already segmented structures. Some authors have shown techniques which circumvent the need for segmenting all sample volumes and work directly from the raw images [11], [12].

To the best of our knowledge, no work has attempted to automate the process of creating intensity models. This is precisely the focus of this work, which we use to complement our fully automatic ASM construction strategy initiated with the autolandmarking method by Frangi et al. [9], and more recently by Ordas et al. [12]. We show that it is possible to build a 3D-ASM, suitable for segmentation of gated single photon emission computed tomography (gSPECT) images, with a PDM previously built from a large database of cardiac computed tomography (CT) data [12]. The use of a virtual population provided access to known LV surfaces for training purposes and accuracy evaluation.

0278-0062/25.00 © 2008 IEEE
On the other hand, imaging simulators are currently a mature field of research providing tools for a variety of modalities: SPECT [13]–[15], CT [16], [17], ultrasound (US) [18], and magnetic resonance imaging (MRI) [19], [20]. Among them, SPECT simulators have the longest trajectory, hence they now offer straightforward tools for cardiac applications. This has motivated the use of gSPECT as a show case for the usage of our approach. Nonetheless, the underlying concepts regarding automatic building of statistical models can be applied to other major diagnostic imaging modalities.

Segmentation of the LV cavity from SPECT imaging is a challenging problem owing to limitations inherent to the modality (i.e., low-resolution, blurred boundaries, high noise levels, signal drops, absence of anatomical landmarks, etc.) [21]. Model-based post processing algorithms are quite widespread in clinical practice [22], [23]. Yet their quantifications are affected by intrinsic imaging drawbacks, specially in patients with small or hypertrophic hearts [24]. Similarly, less accurate calculations have been found in the presence of extracardiac activity, low-dose studies or severe perfusion defects [25], [26]. Hence, new approaches able to cope with these constraints are highly desirable. Deformable models [27] and level set based [28] algorithms are more sophisticated approaches previously applied to SPECT segmentation, giving promising results on simulated data. Still, further validation on real clinical cases is needed.

This paper is organized as follows. The theoretical background of ASMs is explained in Section II. The datasets used for our experiments are presented in Section III. A detailed description of the methodology for automatic construction of 3D-ASM intensity models is provided in Section IV. Section V presents the experimental setup of this work, followed by its results in Section VI. Section VII aims to further discuss the obtained results. Finally, the last section exposes the clinical contribution and outlook of our work.

II. BACKGROUND

A concise explanation of ASM is provided in the current section. An extended description can be found in Cootes et al. [3].

Basically, three main parts constitute the backbone of an ASM: a shape model, an intensity model, and a matching algorithm. The shape model (PDM) represents the shape variability of the object under study. For a 3D space, a linear PDM constructed from $n$ aligned shapes, $\{\mathbf{x}_i; i = 1, \ldots, n\}$, of $m$ landmarks each, $\{l_j = (l_{ij1}, l_{ij2}, l_{ij3}); j = 1, \ldots, m\}$, is a linear model defined by

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi \mathbf{b}$$  \hspace{1cm} (1)

where $\mathbf{x}$ is a $3m$-element vector obtained by concatenating all landmark coordinates in the form $(l_{i11}, l_{i21}, l_{i31}, l_{i12}, l_{i22}, l_{i32}, \ldots, l_{im1}, l_{im2}, l_{im3})$. Then, $\bar{\mathbf{x}}$ is the mean of aligned shapes in the training set, $\mathbf{b}$ is the shape parameter vector of the model, and $\Phi$ is a matrix whose columns are the principal components of the covariance matrix

$$S = \frac{1}{n-1} \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$  \hspace{1cm} (2)

Obtaining the $m$ 3D landmarks and their correspondence for all points on every dataset is not a trivial task. Our methodology was inspired by the method proposed by Frangi et al. [9]. Because of our particular application, a one chamber model (LV) was used. Such configuration is a subpart of our recently constructed whole heart model, trained from a high-resolution CT dataset [12]. Its training included 100 subjects in 15 temporal phases. Thus, 1500 sample volumes were considered in total.

Once the shape model has been established, the second component (intensity model) comes into action. It aims to grasp the intensity distribution typically found near the object’s boundaries. It does so by sampling the gradient of the intensity profiles along the perpendiculars to the mesh. From pixels sampled along each profile, the mean vector and covariance matrix are estimated and stored for later use during matching. An intensity model was calculated for each endocardial and epicardial wall of the 17 LV AHA’s segments [29]. Hence, a total of 33 regions were obtained, corresponding to 17 epicardial and 16 endocardial.

Finally, the third element (matching algorithm) has the role of deforming the mesh to match image data. Our approach is based on the sparse fitting method, SPASM, put forward by van Assen et al. [4]. We modified this technique by using an intensity model where each candidate point is obtained by selecting the minimal Mahalanobis distance between the sampled profiles and the mean profiles of the intensity model. Candidate points operate as deformation forces propagated to neighboring nodes with a weight function

$$w(\lambda, \omega) = e^{-||\lambda - \omega||^2/2\sigma_p^2}$$  \hspace{1cm} (3)

where $||\lambda - \omega||^2$ is the geodesic distance between nodes and $\sigma_p$ is the width of the normalizing Gaussian kernel. Deformation forces drive the mesh to a best-fit location after several iterations. The steps of the algorithm are illustrated in Algorithm 1.

III. MATERIALS

Two main datasets were used for this work: a virtual and a clinical population. The virtual population consisted of digital phantoms (see Section IV-A for details) and was considered for 3D-ASM intensity model training. Afterwards, it was employed to evaluate performance of the trained models by means of leave-one-out approach. Each case was segmented by a model trained with all cases but itself (in total $n-1$ cases).

The clinical population, on the other hand, was only used for performance evaluation. It included 20 subjects of which two were healthy, two hypertrophic, 11 infarcted, and five dilated. A rest gSPECT study and an MRI study were obtained for each subject with a mean interval of 53 days given no change in clinical condition.

Gated SPECT studies were acquired at a rate of eight frames per cardiac cycle. Patients were imaged one hour after administration of 99mTc-tetrofosmin using a Siemens ECAM SPECT system (Siemens Medical Systems, Erlangen, Germany) or an ADAC CardioEpic system (Philips Medical Systems, Best, The Netherlands) both with a double-detector at 90° with high resolution collimators. Sixty-four projections of a 64 × 64 matrix over 180° arc were obtained with a 6.60 mm/pixel resolution.
Fig. 1. Interpolated (top) and original (bottom) axial views of a virtual [(a), (b), (d)–(e)] and a clinical [(c), (f)] gSPECT study. They were reconstructed by means of OSEM [(a), (d)] and FBP [(b), (c), (e)–(f)].

Image data was reconstructed with filtered back-projection (see Fig. 1). MRI studies were acquired using a General Electric Signa CVi-HDx, 1.5-T scanner (General Electric, Milwaukee, WI). Datasets contained short-axis image stacks at 30 temporal phases. The slice thickness was 8 mm with an in-plane pixel resolution of 0.78 mm.

IV. METHODS

In the current section our methodology for automatic construction of intensity models for 3D-ASM is described thoroughly. For an overall view of the complete pipeline, refer to Fig. 2.

A. Digital Phantoms

To ensure a realistic representation of a clinical population, several anatomical parameters were modified in a random manner, as proposed by He et al. [30], resembling a normal distribution obtained from the Emory PET thorax model database [31]. The minimal population size, $n$, was calculated following the criteria exposed by Jain et al. [32]. In our case, close to twenty parameters were modified during patient generation, yielding $n_{\text{min}} = 200$. Detailed description of modified parameters follows.

1) Anatomical Variations: Aiming to include anatomical variations which induce usual attenuation artifacts (i.e., breasts or high diaphragms) [33], three main anatomical groups were implemented (see Fig. 3).

- Normal Subjects: Featuring males with a flat diaphragm and females with small breasts.
- Male Subjects with High Liver Dome: Half the male subjects present a high liver dome, creating strong edges which may attract segmentation algorithms.
- Female Individuals with Large Breasts: Breast size, position, and orientation were modified in order to represent possible attenuation effects.

Authorized licensed use limited to: UNIVERSITAT POMPEU FABRA. Downloaded on September 28, 2009 at 06:01 from IEEE Xplore. Restrictions apply.
1660 IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 27, NO. 11, NOVEMBER 2008

Fig. 4. General distribution of the virtual population, subdivided into anatomical groups. See Section IV-A for details.

In order to generate the population, eight representative individuals were chosen from the Emory PET thorax model database [31]. With these eight anatomical models, four male (M1, M2, M3, M4) and four female (F1, F2, F3, F4), a total of 208 subjects were created, for which half the males present a high liver dome and half the females were attributed large breasts. Fig. 4 presents a graph which illustrates the general distribution of the virtual population. Parameters used as NCAT input are summarized in Table I.

TABLE I
tORSO PARAMETERS OF FEMALE AND MALE SUBJECTS

<table>
<thead>
<tr>
<th>Gender</th>
<th>Model</th>
<th>Body</th>
<th>Ribcage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cm</td>
<td>cm</td>
<td>cm</td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>F1</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>31</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>F3</td>
<td>33</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>F4</td>
<td>32</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>M1</td>
<td>39</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>M2</td>
<td>37</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>M3</td>
<td>35</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>M4</td>
<td>38</td>
<td>28</td>
<td>27</td>
</tr>
</tbody>
</table>

 TABLE II
ANATOMICAL PARAMETERS FOR HEART VARIATION ACCORDING TO GENDER (ADAPTED FROM [30])

<table>
<thead>
<tr>
<th>Gender</th>
<th>Measure</th>
<th>Size</th>
<th>Orientation</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length</td>
<td>Ratio</td>
<td>Angle $\phi$</td>
<td>Angle $\psi$</td>
</tr>
<tr>
<td></td>
<td>cm</td>
<td></td>
<td>cm</td>
<td>cm</td>
</tr>
<tr>
<td>Female</td>
<td>7.4</td>
<td>3.20</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>Mean</td>
<td>0.9</td>
<td>0.50</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>SD</td>
<td>10.5</td>
<td>4.00</td>
<td>54</td>
<td>76</td>
</tr>
<tr>
<td>Max</td>
<td>5.7</td>
<td>2.44</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Min</td>
<td>8.3</td>
<td>3.17</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>Female</td>
<td>0.9</td>
<td>0.40</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Mean</td>
<td>11.6</td>
<td>4.32</td>
<td>41</td>
<td>73</td>
</tr>
<tr>
<td>Max</td>
<td>6.6</td>
<td>2.29</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2) Heart Variations: The heart of each subject was varied by modifying its length and left ventricular basal radius. Global position was altered by inducing different orientation angles and translations of the heart along posterior–anterior (P–A) and lateral (Lat) directions. Specific parameters are summarized in Table II.

3) Organ Uptake Ratios: Tracer uptake of organs differs from patient to patient. To mimic this physiological condition, heart, liver, lung, kidney, spleen, and background isotope uptake ratios were also modified in a random manner resembling a normal distribution of a typical clinical population [30]. Parameter values are displayed in Table III.

4) Phantom Generation: Each voxel phantom included activity and attenuation files for eight phases of a normal (1 s) cardiac cycle. Each set consisted of 98 slices of $64 \times 64$ pixels with a 6.25 mm isotropic voxel size. This low resolution matches the usual conditions present in our clinical studies.

Up to this point the anatomical models included a full thorax model that incorporates structures other than the heart, which are important for realistic gSPECT simulation. Aiming to extract LV true surfaces, higher resolution images with only the LV structure were generated. They consisted of 321 slices of $512 \times 512$ pixels each, with a 0.78 mm isotropic voxel size. Once true surfaces were extracted from these datasets, our 3D model was aligned to them using a similarity transformation through Procrustes Analysis [34]. Subsequently, nodes of the true surfaces acted as exact candidates to deform our mean shape using one iteration of the ASM algorithm. This process allowed warping the atlas model to all the training shapes in order to assure: 1) control over the distribution of clinical parameters in our training database of heart shapes based on published data, 2) the same number of nodes and mesh topology for all true LV surfaces, and 3) the inclusion of high intersubject and interphase variability during the matching process since the PDM is based on a larger database of real patient data.

B. Monte Carlo Simulation

In order to generate gSPECT studies for the virtual population, Monte Carlo simulation was employed using SIMIND code [13]. Details regarding the simulation setup are given below.

1) Collimator Parameters: SIMIND allows for modeling different types of collimators. A Siemens low-energy high-resolution (LEHR) collimator was chosen since it resembles our current clinical conditions [35]. Characteristics of such a collimator include: hexagonal shape, parallel hole collimator, radius of rotation of 20 cm, hole size of 1.24 mm, septal size of 0.90 mm, and thickness of 23.6 mm.

2) Projection Parameters: Noise free projections were obtained by simulating $10^7$ photon histories per projection. Sixty-four of them were obtained over a 180° arc, from 45° left posterior oblique to 45° right anterior oblique. Each projection consisted of a $64 \times 64$ matrix with 6.25 mm/pixel resolution. Energy resolution was set to 9% full-width-at-half-maximum (FWHM) at 140 keV and energy window threshold to 15% photopeak at 140 keV.

3) System Characterization: Ordered-subset Expectation Maximization (OSEM) reconstruction requires FWHM
parameters to be determined (see Section IV-C). This was accomplished by measuring point-sources at different distances from the collimator surface. The point source response was approximated to a symmetric Gaussian by means of nonlinear least squares fitting [36].

4) Image Generation: Simulations were run using grid computing on a cluster facility of 20 dual-processor dual-core SGI Teztrix 210/2, 3 Ghz/1333 Mhz, Intel Woodcrest processors. InnerGrid v5.0 (GridSystems, Palma de Mallorca, Spain) was employed as grid middleware. Distribution was achieved in the following manner. Each subject corresponds to eight digital phantom datasets (one for each cardiac phase), totalling 1664 digital phantoms (208 subjects × 8 time frames). Each dataset includes 64 projections, which were distributed to different nodes of the cluster such that one node will simulate only one projection of one digital phantom. The whole set of projections was then concatenated to obtain full projection volumes. This methodology allowed us to reduce computation time from 16 h to 48 min per subject. For the whole database it represented trading five months of calculations for about seven days.

C. Tomographic Reconstruction

Aiming to obtain datasets with different intensity features for our model training (see Section II) tomographic reconstruction after simulation was performed in two approaches: filtered backprojection (FBP) and ordered-subset expectation maximization (OSEM).

1) FBP Reconstruction: Reconstruction was performed with a Butterworth filter. Its cutoff frequency was visually inspected on a range from 0.30 to 0.80 pixels⁻¹ with step 0.2. Selected parameters were order 4 and cutoff frequency of 0.66 pixels⁻¹.

2) OSEM Reconstruction: Reconstruction was carried out using four subsets and 20 iterations. It also applied a quadratic penalty function using the four nearest neighbors of each pixel within a plane, along with the pixels adjacent to it on the slices above and below, as suggested by Fessler [36].

D. Postprocessing

Following reconstruction, images were automatically masked for truncation artifact removal. Subsequently, they were scaled to a 100 grey-level window, setting negative values to zero. Finally, they were saved in DICOM format in order to be processed by our 3D-ASM algorithm as a regular patient.

E. 3D-ASM Segmentation

Automatic segmentation of LV cavity was performed by means of 3D-ASM (see Section II). Implementation details are provided next.

1) ASM Parameters: A univentricular model of 2677 points (1835 for endocardium and 842 for epicardium) was used. The algorithm was set to run for 15 iterations or until the change in LV volume was not substantial between iterations (ΔVolume < 0.01 mL). New model instances were generated with 75% of the total shape variability. This constrain was imposed to obtain a smooth fit to match the sparse data obtained from SPECT imaging, as opposed to CT imaging which allows for finer details. Other ASM parameters are summarized in Table IV.

2) Dynamic Studies Segmentation: Cardiac dynamics add to our segmentation process yet another challenge: intensity profile variation per cardiac phase. The most intuitive scheme to approach this matter would be to obtain a model trained for each cardiac phase.

An alternative strategy is to perform ASM fusion [37], which has proven to be an effective technique for intensity model generation [38]. Under this methodology, only end diastolic (ED) and end systolic (ES) models were generated, since they represent the two most extreme circumstances on cardiac dynamics. Missing phases were obtained through a weighted fusion of ED and ES models. Weights used for each cardiac phase were set by the current heart phase index (LV contraction percentage) as logged by NCAT [39].

3) Model Initialization: We followed a very simple mechanism to roughly scale and position the mean shape of the model. The operator defines two epicardial points at the basal level and a third one at the apex. Corresponding anatomical landmarks of the mean shape were previously tagged by an experienced investigator. Consequently, the mean shape is aligned to the landmarks through a similarity transformation. The manual interaction required for this procedure lasts about 30 s. In complex cases (i.e., large perfusion defects) longer interaction may be required, up to 1.5 min, for a correct depiction of basal and apical planes.

For the virtual population, initialization points were extracted automatically from the true shapes, thus eliminating initialization bias for a better analysis of segmentation accuracy. The clinical database, instead, was initialized by an experienced investigator, hereafter referred to as Obs1.

4) Functional Analysis: Once the shape model is correctly matched to specific image data, LV volumes both in end diastole (EDV) and end systole (ESV) can be calculated. Ejection fraction (EF) can be derived from these measurements in order to evaluate systolic function of a patient.

V. EXPERIMENTAL EVALUATION

A. Segmentation Accuracy

- Idealized Versus Simulated Boundary Model: To evaluate the advantage of using advanced simulations during training, a comparison with two idealized boundary models was performed. The first model consisted of a step function (ST), ranging from zero to one corresponding to a normalized intensity profile. The second model located the boundary at the maximum gradient (GR) of a sampled profile, as initially proposed by Cootes et al. [3]. Both
virtual and clinical populations were segmented with these models.

- **True Versus Fitted Geometry:** Unsigned point-to-surface (P2S) errors were computed between the fitted meshes obtained with idealized and simulated boundary models and the gold standard LV surfaces. Mean ± SD values of all subjects in all temporal phases were computed.

- **Trained-tested Analysis:** To examine the influence of using the same reconstruction method both in training and segmentation stages, we performed an experiment combining trained-tested models. That is, a model trained with FBP reconstructed datasets was tested on an OSEM reconstructed dataset during segmentation, and vice versa. A Mann–Whitney U-test [40], with a 95% confidence interval, was carried out to determine statistical significance of the differences.

- **Clinical Dataset:** Location of LV borders in SPECT datasets is quite subjective due to the blurred nature of these images (see Fig. 1). However, to generate a proper gold standard for accuracy evaluation, LV contours were manually drawn according to a standard criterion: LV borders should be located at 40% of the maximum myocardial intensity. This value was obtained based on reported studies [41] and our clinical experience. In case of extensive perfusion defects, the clinical observer could modify the threshold down to 20%. Endocardial and epicardial border delineation of the LV, at ED, was performed by two observers (Obs1, Obs2) in two individual sessions (S1, S2). The resulting traces were used to: 1) evaluate intraobserver and interobserver variability and 2) obtain P2S errors of automatically segmented surfaces.

### VI. RESULTS

#### A. Quantitative

1) **Segmentation Accuracy:** Fig. 5 shows LV edges obtained with 3D-ASM for the ST, GR, FBP, and OSEM boundary models. Fig. 6 displays two clinical cases with severe perfusion defects. LV edges obtained with 3D-ASM for all boundary models are displayed as well. Corresponding true surfaces are included on both figures.
Table V shows the results for the trained-tested analysis and the idealized versus simulated boundary model analysis. The P2S errors of the segmentations performed with the idealized models are noticeably larger than the ones of the simulated boundary models. Endocardial errors were 28% larger than those of the FBP model and 20% larger than those of the OSEM model. Epicardial errors were 89% larger than those of the FBP model and 66% larger than those of the the OSEM model.

Subvoxel accuracy was obtained with our segmentation method for both reconstruction techniques (see Table V). For FBP reconstructed datasets, epicardial borders were segmented 35% more accurately than endocardial ones, while in OSEM reconstructed datasets the difference was 38%.

Fig. 7 displays the statistical significance evaluation of the trained-tested analysis. All compared groups generated significantly different P2S errors, except for endocardial errors of FBP-FBP versus OSEM-FBP and ST-FBP versus GR-FBP, and epicardial errors of ST-OSEM versus GR-OSEM.

Fig. 8 displays P2S errors for each cardiac phase, with ED being $t = 1$ and ES being $t = 5$. Endocardial errors obtained at ED were 21% larger with respect to ES for both FBP and OSEM reconstructed datasets. On the other hand, epicardial errors were 18% smaller at ED for FBP reconstructed datasets and only 3% lower for OSEM reconstructed datasets.

Table VI shows the results for the clinical population. P2S errors between 3D-ASM fitted shapes and manual delineations are also displayed. For endocardial errors, intraobserver and interobserver variabilities were not significantly different than those obtained automatically with the FBP and ST boundary models. The GR boundary model, instead, generated significantly higher P2S errors than intraobserver variability. They were also significantly higher than those of the FBP boundary model. Epicardial errors, on the other hand, were found to be significantly different for all schemes.

2) Sensitivity to Initialization: Table VII shows the results regarding initialization sensitivity for FBP and OSEM reconstructed datasets. For both of them, the added inaccuracy caused
by initialization error was 5% for endocardial borders and 8% for epicardial ones. Volume calculations presented an average error of 3.5 mL affecting the EF measurements in 4.7%. However, maximum errors came to be as large as 22 mm for accuracy measurements and 74 mL for volume calculations.

3) LV Function Analysis: Fig. 10 displays B&A plots of volume calculations for the virtual population. FBP reconstructed datasets produced EDV measurements with a 94.4% accuracy, ESV measurements with a 90.0% accuracy, and EF measurements with a 90.8% accuracy. For the OSEM reconstructed datasets, accuracy calculations were 94.5% for EDV, 90.2% for ESV, and 90.9% for EF. A further analysis of EF error relative to EDV is presented in Fig. 11.

For the clinical population, B&A plots are displayed in Fig. 12. 3D-ASM obtained accuracy levels of 89.5% for EDV, 87.0% for ESV, and 88.1% for EF. QGS measurements obtained accuracy levels of 81.7% for EDV, 83.5% for ESV, and 83.9% for EF. In concrete, the B&A plots for EF calculated with 3D-ASM displayed no bias and smaller variance than those of QGS.

Fig. 13 displays segmentation errors for the clinical population subgroups. Errors showed no obvious correlation to severity of perfusion defect. Only ESV of the none subgroup shows a high inaccuracy for both post processing algorithms. It must be noted that half the patients in this group (2 out of 4) presented hypertrophic LVs with collapsing walls at ES, hence the larger errors in ESV calculations.

B. Critical Analysis

1) Segmentation Accuracy: Idealized models demonstrated not to be robust enough for the segmentation task evaluated during this work. Fig. 14 illustrates this fact by displaying a bar
plot of the gradient profile averaged over all landmarks and all datasets of each population (i.e., $n_{\text{virtual}} = 208$ and $n_{\text{clinical}} = 20$). Position zero in the horizontal axis indicates the location of the boundary. Due to the absence of OSEM clinical datasets, only the FBP datasets are presented. Comparisons were performed against the corresponding gold standard which is represented with dark bars. Light bars represent the profile with respect to the best-fit boundary position according to the FBP, GR, and ST boundary models. It is interesting to observe that in all cases (virtual and clinical datasets) the actual best-fit profiles are more alike to the simulated profiles than to the idealized profiles. This is achieved in spite of the limitations of a simulated training set, which may not capture all the details of an actual clinical database. Similarly, the standard deviation of the difference between the gold standard and the simulated boundary models were smaller than those of the two idealized boundary models. In practical terms, it reduced P2S segmentation errors by at least 20% for endocardial borders and 66% for epicardial borders.

The trained-tested analysis showed that more accurate segmentation results are obtained when the same reconstruction method is used both in training and segmentation stages. Despite the fact that OSEM reconstruction allows for better definition of LV structures, endocardial borders are located with errors of the same magnitude as those obtained with FBP. We suspect that a substantial increase in image resolution is necessary before the apparent visual improvement of OSEM reconstructed datasets has a real impact on global quantitative parameters.

Overall decreased accuracy found on endocardial border segmentation is reasonable as the relative image resolution is lower for the inner surface of the LV. That is, the correct position of a large contour (epicardium) can be found more precisely than the position of a small contour (endocardium), given the same pixel size.

Greater P2S errors found at basal level are quite understandable since a correct depiction of LV basal plane is a well known complication of cardiac imaging postprocessing for most modalities [43]. SPECT images are specially challenging on this matter owing to the lack of commonly used anatomical landmarks such as the mitral valve or the left atria.

As can be observed in Fig. 9, P2S errors are larger at the inferoseptal basal segment. Because of the presence of the membranous septum, this region displays almost no tracer activity. Hence, during fitting the mesh is not actively deformed at this area of the LV wall. This is represented in the virtual phantoms as thinner septal structures. It is particularly noticeable at ED
where the difference in activity between the basal portion of the lateral wall and the basal portion of the septal wall is quite visible. At ES, though, due to thickening and shortening of the LV walls, the septum can be better defined at basal levels.

For the cardiac phase analysis, the larger epicardial P2S errors found at ES phase are natural (lower resolution and partial volume effect). However, the decrease in error observed for endocardial borders is counterintuitive. Visual inspection suggests this is caused by the higher segmentation inaccuracy at basal level, as mentioned above.

For the clinical studies, 3D-ASM errors for endocardial borders are comparable to interobserver variability. However, epicardial boundaries presented 20% larger errors than interobserver variability. This might be due to overestimation of wall thickness in places of extensive perfusion defects. Regardless of lack of data, a clinical observer may deduct a thinning of the LV walls caused by chronic infarcted myocardium. ASM, on the other hand, will try to conserve the wall thickness present on the remaining sampled data. It must be noted that intraobserver and interobserver variability under uncontrolled circumstances (i.e., without a standardized criterion) will most likely be larger than the ones measured during our experiments.

2) Sensitivity to Initialization: The evaluation of initialization sensitivity illustrated the extent of inaccuracy caused by initialization error. Yet, in average, this inaccuracy was rather small. The maximum errors revealed noticeable bias in case of very improper initialization points. However, in clinical dataset processing, initialization would be performed by a trained technician capable of efficiently and correctly defining basal and apical positions.

3) LV Function Analysis: For the virtual population, the scatter distribution of the B&A plots showed a dependency of the error on the LV volume. B&A plots also revealed that our algorithm tends to underestimate EDV, a tendency also present on QGS (see Table VIII). The most extreme case of overestimation was found for the largest heart. Yet its difference is within reported limits of discrepancy (30 mL from gold standard measurements) [51].

For ESV, a slight overestimation is revealed through the B&A plots, previously stated for QGS as well (Table VIII). For EF, the confidence intervals in the B&A plots are wider than those for EDV and ESV, probably caused by the higher dispersion observed on lower EF values. Note in Fig. 11 that many of the large discrepancies in EF calculations are located around small hearts (50 mL EDV). This parameter is known to be overestimated for this type of hearts when calculated from perfusion studies [52]. This is attributed to artificially increased counts in the LV cavity, complicating a proper calculation of ESV volumes.

For the clinical population, overall patterns of B&A plots were comparable to those of QGS. Calculated parameters showed less bias in underestimations. Smaller confidence intervals were found for 3D-ASM for all calculated parameters. Similarly, accuracy levels were higher than those obtained with QGS for all measured parameters.

No obvious correlation between perfusion defect severity and segmentation inaccuracy was found for our clinical database. Inaccuracy could be more related to low image quality or segmentation difficulty depending on pathology. For instance, the group with no perfusion defects was composed of hypertrophic patients and one dilated patient with left bundle branch block, both difficult cases to segment even for a clinical observer.

VII. DISCUSSION

A. Clinical Contributions

Our method obtained higher accuracy compared to QGS, one of the most widespread commercial packages. Although this result is obtained in a small population, this is quite encouraging for a simulation based approach since it bypasses the labor of clinical database collection and, furthermore, the underlying methodology is potentially applicable to other modalities.

The employed segmentation method could either be applied on transaxial slices or on reformatted short axis images. The use of the transaxial slices is preferable since time consuming operator assistance is required to define the LV long axis.

As can be concluded from previous works (see Table VIII) the tendencies of QGS for small hearts still need further review. Virtual populations with specific heart sizes may be useful for investigating this matter.
TABLE VIII
META ANALYSIS OF PUBLISHED WORKS COMPARING QGS POSTPROCESSING RESULTS AGAINST A GOLD STANDARD

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>N (FM)</th>
<th>Population</th>
<th>Gold Standard (GS)</th>
<th>Post Processing Software</th>
<th>Small Hearts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achtert et al.</td>
<td>1998</td>
<td>3 (n.a.)</td>
<td>Normal</td>
<td>MCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>111.0</td>
<td>mL 42.5 mL 61.9 mL 121.8</td>
<td>mL 50.0 mL 59</td>
</tr>
<tr>
<td>Tadamaru et al.</td>
<td>1999</td>
<td>16 (3/13)</td>
<td>Surgery</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>106.1</td>
<td>mL 54.9 mL 47.9 mL 112.1</td>
<td>mL 55.3 mL 50.8</td>
</tr>
<tr>
<td>Bavelaar-Croon et al.</td>
<td>2000</td>
<td>21 (7.14)</td>
<td>CAD MRI</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>151.0</td>
<td>mL 97.0 mL 43.0 mL 191.0</td>
<td>mL 114.0 mL 45.0</td>
</tr>
<tr>
<td>Nakajima et al.</td>
<td>2001</td>
<td>4 (n.a.)</td>
<td>Normal</td>
<td>Phantom</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>123.0</td>
<td>n.a. mL n.a. mL 131.0</td>
<td>mL n.a. mL n.a.</td>
</tr>
<tr>
<td>Nakajima et al.</td>
<td>2001</td>
<td>30 (10/20)</td>
<td>Mixed</td>
<td>GBP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>98.0</td>
<td>mL 54.0 mL 5.0 mL 103.0</td>
<td>mL 49.0 mL 7.0</td>
</tr>
<tr>
<td>Lipke et al.</td>
<td>2004</td>
<td>54 (15/39)</td>
<td>CAD MRI</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>122.0</td>
<td>mL 62.0 mL 52.2 mL 139.0</td>
<td>mL 60.0 mL 60.0</td>
</tr>
<tr>
<td>Lomsky et al.</td>
<td>2005</td>
<td>5 (n.a.)</td>
<td>Normal</td>
<td>NCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>113.0</td>
<td>mL 64.0 mL 45.6 mL 115.2</td>
<td>mL 44.8 mL 61.0</td>
</tr>
<tr>
<td>Schaefer et al.</td>
<td>2005</td>
<td>70 (16/45)</td>
<td>CAD MRI</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120.0</td>
<td>mL 60.0 mL 53.2 mL 137.0</td>
<td>mL 57.0 mL 60.6</td>
</tr>
<tr>
<td>Stegger et al.</td>
<td>2007</td>
<td>70 (16/45)</td>
<td>CAD MRI</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120.0</td>
<td>mL 60.0 mL 53.0 mL 137.0</td>
<td>mL 57.0 mL 61.0</td>
</tr>
<tr>
<td>Wu et al. [50]</td>
<td>2008</td>
<td>33 (n.a.)</td>
<td>CAD MRI</td>
<td>MRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n.a.</td>
<td>mL n.a. mL 40.2 mL n.a.</td>
<td>mL 40.1 mL n.a.</td>
</tr>
</tbody>
</table>

n.a. = Not available; Normal= Healthy hearts; Surgery= Patients who underwent coronary artery bypass surgery; CAD= Known or suspected coronary artery disease; Mixed= Common cardiomyopathies; MCAT= 3D dynamic cardiac-torsor phantom; MRI= Magnetic resonance imaging; Phantom= Cylindrical mathematical model; GBP= Gated blood-pool study; NCAT= 4D NURBS-based cardiac-torsor phantom; †= Underestimation; ‡= Overestimation.

B. Outlook
The feasibility of our approach has been illustrated in the context of one clinical application (viz. cardiac image analysis) and one specific imaging modality (viz. gSPECT). Nevertheless, the potential of this approach is much broader.

To start off, it can help decoupling the sample size requirements of building relevant statistics for the intensity models. Shape models could be built based on a high-resolution imaging modality (e.g., CT) and the derived PDM be sampled to generate a virtual population from which simulated images of other modalities can be produced (e.g., MR, SPECT, or US). Regarding sample size, only few real clinical images might be available for extreme anatomical variants (e.g., very small or very large hearts). However, they can be sampled uniformly when creating the virtual population for simulated data.

Another problem in learning intensity models directly from real images is related to the rapid evolution of most imaging technologies. Handling this problem would become simpler with our technique as we can regenerate the intensity models, as long as the employed simulator allows for it. The upgrades can be related to: 1) improvement of spatial resolution (i.e., smaller pixel size), 2) increase of temporal resolution (i.e., more frames per cycle), 3) development of better reconstruction techniques (e.g., iterative algorithms), 4) isotropic voxels (i.e., for MRI or CT), 5) variation on physical parameters used during acquisition (e.g., modification of MRI sequences), etc.

As the final advantage, we would like to mention that avoiding the need to use shapes derived from manually contoured shapes prevents expert dependency as the true boundary information is known by construction. Moreover, the possibility to build intensity models in every major modality based on a high-resolution PDM pave the way for handling more consistently multimodal datasets.

This approach, however, may present a number of disadvantages, depending on the realism and accuracy of the image acquisition simulator, such as computationally expensive processing, large amount of input parameters sometimes hard to determine, use of theoretical noise which may not resemble clinical conditions, etc.

VIII. Conclusion
This paper introduced the concept of using advanced imaging simulators to enable automatic creation of intensity models. Results show that gSPECT studies can be successfully segmented by models trained under this scheme with subvoxel accuracy. The accuracy in estimated LV function parameters range from 90.0% to 94.5% for the virtual population and from 87.0% to 89.5% for the clinical population. These results are within the intervals reported by other widely clinical segmentation tools.

Our future efforts along the generic approach we presented here will be to extend this technique to other imaging modalities. Efforts are underway to apply this approach to 3D US data [53] and we do not foresee any fundamental issues not to extend this technique to MRI and CT.

ACKNOWLEDGMENT
The authors would like to thank W. P. Segars and M. Ljungberg for their technical support on the use of NCAT [39] and SIMIND [13], respectively.

REFERENCES


