Tracking and Segmentation of the Endocardium of the Left Ventricle in 2D Ultrasound using Deep Learning Architectures and Monte Carlo Sampling

Jacinto C. Nascimento, Gustavo Carneiro and António Freitas Ph.D.

Abstract

The tracking and segmentation of the left ventricle (LV) of the heart from ultrasound data still deserves the attention in medical image community, being a commonly used method in practical clinical setup nowadays. The goal above stated can be formulated as a sequential state estimation problem, which is essential to the study of linear and non-linear dynamical systems. In this work, we present a particle filter based approach, rooted in Bayesian estimation and Monte Carlo procedures, for tracking and segmenting the LV. Two main ingredients characterize this formalism: the *prediction* that models the dynamic of the object in consecutive frames and *filtering* that collects information in the present frame. This methodology allows the computation of the expected segmentation value at the current time instant of the object, given all previous and current observations. Although the probability values of the filtering distribution can be straightforwardly computed, sampling from it is challenging, so this means that a *proposal* distribution is needed since it provides an easier way for sampling the filtering distribution. In this work, we present an algorithm for tracking and segmenting the LV in 2D Ultrasound data based on the observations made above. The contributions are as follows: (i) a new transition model (prediction) that combines different motion regimes presented in the systole and diastole phases of the cardiac cycle; (ii) a new observation model (filtering) built with a deep neural network, and (iii) a new proposal distribution for efficient sampling mechanism. The usefulness of our approach is evaluated using a database of disease cases and another dataset of normal cases, where both datasets present long axis views of the left ventricle. Using a training set comprising diseased and healthy cases, we show that our approach produces accurate results for tracking the endocardium. Also, we show that our method correlates well with inter-user statistics produced by four cardiologists.

I. Introduction

Automatic tracking and segmentation of the left ventricle (LV) endocardium of the heart is an important step towards the estimation of the heart condition, since its quantitative measurement is used as a crucial indicator of the cardiac health. Such tool is able to provide useful information to improve the workflow by increasing the patient throughput and decreasing the interuser variability. Moreover, it constitutes a way of measuring the dynamic behavior of the human heart, where the regional characterization of the heart wall motion is necessary to isolate the severity and extent of diseases such as ischemia. Other features, such as the ejection fraction of the left ventricle, the left ventricle mass of the myocardium and wall thickness and thickening constitute important information that can be accessed with such automatic procedure.

This is however, a difficult problem to be fully solved. Indeed, when developing an algorithm for tracking and segmenting the LV, several problems are encountered in ultrasound data. This usually comprises: (i) fast motion during systole (contraction) phase, (ii) low signal-to-noise ratio, (iii) edge dropout caused by motion, (iv) presence of shadows produced by the dense muscles, (v) specific properties and settings of the ultrasound machine, as well as (vi) anisotropy of the ultrasonic image formation [1]. In ultrasound images, the LV appearance is characterized by a dark region, representing the blood pool inside the chamber, enclosed by the endocardium, myocardium, and epicardium, which are roughly depicted by a

Jacinto C. Nascimento, is with the Institute for Systems and Robotics and with Informatics and Computer Engineering Department, Instituto Superior Técnico, Torre Norte, piso 6, Av Rovisco Pais 1, 1049-001, Lisboa, Portugal (e-mail: jan@isr.ist.utl.pt).

Gustavo carneiro is with the School of Computer Science, Australian Centre for Visual Technologies, University of Adelaide, North Terrace, Inkarni Wardli Building, Adelaide, SA 5005, Australia (e-mail: carneiro.gustavo@gmail.com).

Dr. António Freitas is with Department of Cardiology, Fernando Fonseca Hospital, Lisbon, Portugal (e-mail: aeffreitas@gmail.com).

brighter region. Also note that specific spatial texture and gray value distribution of each region vary substantially among different cases and even within each case. All the above mentioned issues impose obstacles when developing an automatic procedure for the LV segmentation in ultrasound data.

Most of the current methodologies formulate the problem of tracking as a state estimation problem, in which the expected segmentation is computed taking into account the previous and current observations over the space of segmentation parameters [2]. Under this formalism, the segmentation parameters constitute the state vector while the image represents the observations. The expected segmentation described above is computed using the filtering distribution, which calculates the probability of a possible segmentation given the previous and current observations. The computation of this expected value is not possible to be obtained analytically, given the high number of dimensions of the space of segmentation parameters and the non-Gaussianity of the distribution. As such, it is common to approximate this expected value using Sequential Monte Carlo (SMC) sampling techniques, meaning that only a few weighted samples (each sample representing a segmentation of the object) are needed to produce the expected value. The weights in the samples are computed using the observation and transition models, while the samples are obtained from sampling the filtering distribution [3]. Another usual problem is the difficulty in sampling this filtering distribution, which is solved by sampling another distribution, called the *proposal* distribution, that provides a reasonable approximation to the filtering distribution, but being much simpler to sample. Then the probability of the proposal distribution has to be taken into account when calculating the sample weights. Finally, using the samples and their respective weights, it is possible to compute the expected segmentation mentioned above.

II. Related work

Several trends characterize the related work for object segmentation. One such trend is represented by the active contours [4] whose geometry allows a broad shape coverage by employing a geometric representation that involves many degrees of freedom. The deformable model designation stems from the use of elasticity theory at the physical level within a Langrangian dynamics setting. In particular, deformable model acts as an elastic body that respond to applied forces and constraints, and coupled with it an energy is associated and modified as the deformable model moves in the image domain. In the Lagrangean setting, the deformation energy is associated to elastic forces that are internal to the model, which are called *internal forcers* (*i.e.* the prior). Under physics perspective, the external potential energy functions are defined in terms of data of interest in the image (*e.g.* boundary of the object to be segmented). These potential energies are associated to the *external forces* that are able to deform the model to fit the desired data. The energy of the deformable model is supposed to be minimal when the deformable model is located at boundary of interest (external energy) and has a shape which is supposed to be relevant considering the shape of the sought object (internal energy).

Though successful at several tasks [5], [6], problems regarding the initialization as well as the presence of outliers in the image motivated the development of level-set methods [7]. Level-set theory aims to exchange the Lagrangian formalization (used in active contours) and replace it with an Eulerian form. In this class of approaches, the initial valued partial differential equations control the front (*i.e.* boundary of the object) evolution, representing the boundary as the zero level-set instance. For the level-set class of approaches, the LV contour is represented by the zero-level in the signed distance function. In general the level-set based approaches can provide higher robustness against the initialization [6], [8]–[16] and robustness to sharp corners and cusps. Although level set have shown outstanding results in medical image applications, they face limitations when dealing with the prior knowledge defined in the optimization function regarding the LV boundary, shape and texture distribution.

The above issues motivated another line of research, namely the pattern recognition methods that involve the use of a database of annotated LV images (*i.e.*, a training set) to automatically build a model of the LV shape and appearance. One of the first examples of pattern recognition models are the active shape and appearance models [1], [17]–[19]. There are a few issues that affect these approaches: 1) need of

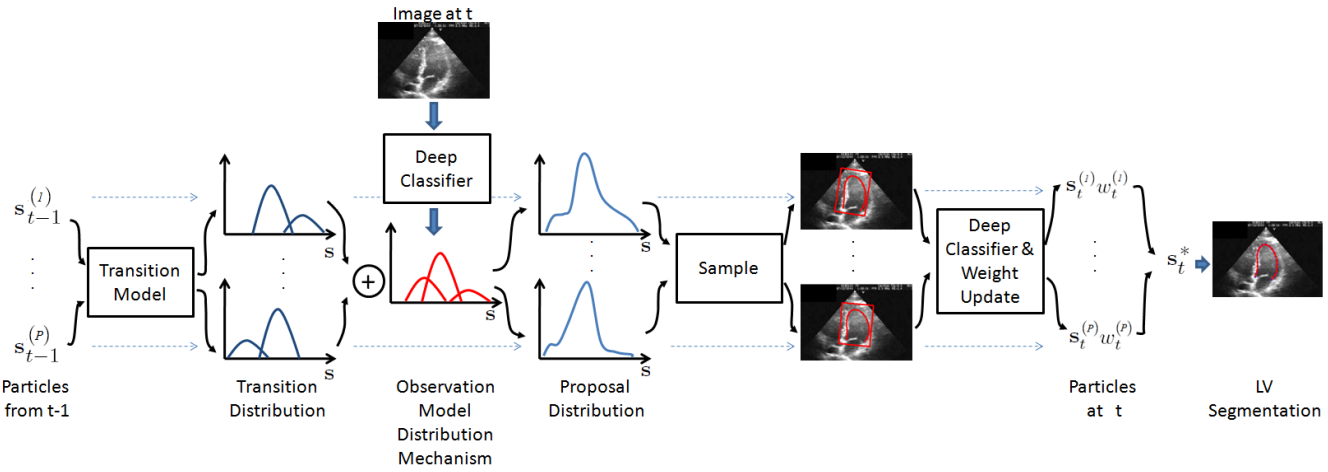
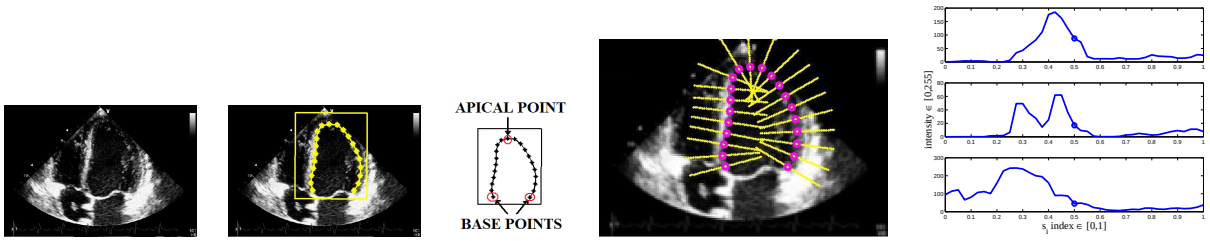


Fig. 1. Block diagram containing all steps of the tracking algorithm proposed in this chapter.

large annotated training set, 2) during inference, the initialization must be close to a local optimum, and 3) the Gaussian distribution assumption for the shape and appearance models constrain their capacity. The initialization problem has been successfully handled [20], [21], but the large training set is still an issue in the field.

Though receiving less attention, the transition model plays an important role in the computation of the filtering distribution, since it conveys information about the dynamics. The most usual transition model is the prediction estimated from the Kalman filtering [22]. However, the Gaussian assumption of the Kalman filter is not realistic, given the complex motion patterns of the heart that violates such assumption. More interesting transition models are built when providing more degrees of freedom to explain those motion patterns that are more likely to happen in practice. For instance, Sun et al. [23] introduce a transition model that is learned from training data using an information-theoretic criterion, but the lack of a prior distribution in the model imposes the need of a large training set to provide a reliable transition model. A related approach is proposed by Yang *et al* [24] consisting of a transition model that depends not only on the previous state vector, but also on all state vectors up to current time instant. As previously, this model is also automatically learned from training data and consists of a manifold describing the motion pattern of the heart. Models based solely on prior information [25] also seem inadequate given that there might be information present in training data that may not have been captured by the prior. The transition model proposed by Nascimento [26] consisting of a mixture of two models (one for systole and another for diastole) seems more adequate, and inspired us to implement our transition model. The main difference is that we use both a prior information on the motion patterns, assuming the existence of two cardiac cycles (*i.e.*, systole and diastole), and a learned model from data instead of a transition model containing only prior distributions [26].

Concerning now the tracking methods based on SMC sampling techniques, it is necessary to use a proposal distribution that approximates reasonably well the filtering distribution [3]. Senegas et al. [25] propose an SMC sampling method using a proposal distribution based only on the observation model, which does not take into consideration the transition model. Sun et al. [23] introduce a proposal distribution based only on the transition model, which also presents a limitation given that the observation model is not considered. The work that inspired our model was proposed by Okuma et al. [27], who proposed a tracking algorithm (*i.e.*, not LV tracking) combining discriminative classifier detections and particle filtering to track multi-target non-rigid objects. Notice however that the work presented herein contrasts with [27] in sense that we are now concerned with the precision of the segmentation, which is a mandatory requirement concerning the segmentation of the LV.



(a) (b) (c) (d) (e)

Fig. 2. (a) Original training image; (b) manual segmentation of the LV in yellow dots inside the rectangular patch (yellow lines); (c) representation of the canonical coordinate system for LV contour with the base and apical points highlighted and located at their canonical locations within the patch delimited in (b); (d) lines taken at orthogonal lines radiating from the points depicted in (b); (e) intensity profiles in three lines.

III. Contributions

In this chapter, we propose a new LV tracking algorithm based on SMC methods. See Fig. 1 for an illustration of the proposal. Our main contributions are the following:

1) *New transition model*: The transition model proposed in this chapter makes use of the prior information that at each time instant, the heart is either expanding (diastole) or contracting (systole). The deformation caused by these motion patterns are described by a linear transform, whose parameters are learned from the training data.
2) *New observation model*: the model is built with deep learning architectures which involves a statistical pattern recognition model, where we address the robustness to imaging conditions unseen in training data. In order to handle the robustness to imaging conditions, we move away from the use of boosting classifiers [21], and rely on the use of deep neural network classifiers [28]. The main advantage of deep neural networks is its ability to produce more abstract feature spaces for classification and to automatically generate optimum feature spaces directly from image data.
3) *New proposal distribution*: the proposal is inspired by the work of Okuma et al. [27], which combines the detection results from the deep learning architecture with the transition model. This combination provides precise segmentation, and robustness to imaging conditions and drifting.

IV. Statistical model of the segmentation algorithm

We assume a non-Gaussian state-space model, where the state sequence is a process represented by $\{k_t, \mathbf{s}_t | t \in \mathbb{N}\}$, where $k \in \{\text{systole}, \text{diastole}\}$ is a (discrete) label indicating the cardiac phase at t^{th} time instant, and $\mathbf{s} \in \mathbb{R}^{2N}$ denotes the contour representation with N key points. The above hybrid state is assumed to be an unobserved (hidden) Markov process with the initial state distribution represented by $p(k_0, \mathbf{s}_0)$ and the transition distribution that takes into consideration the previous cardiac phase and contour representation with $p(k_t, \mathbf{s}_t | k_{t-1}, \mathbf{s}_{t-1})$. The observations consist of the images $\{I_t | t \in \mathbb{N}^*\}$, which are conditionally independent given the process $\{k_t, \mathbf{s}_t | t \in \mathbb{N}\}$, with marginal distribution $p(I_t | k_t, \mathbf{s}_t)$. Also, we assume the existence of a training set $\mathcal{D} = \{(I, \theta, \mathbf{s}, k)_j\}_{j=1}^{M}$ containing M training images I of the ultrasound imaging of LV, the parameters of a rigid transformation $\theta = [\mathbf{x}, \gamma, \sigma] \in \mathbb{R}^5$ (position $\mathbf{x} \in \mathbb{R}^2$, orientation $\gamma \in [-\pi, \pi]$, and scale $\sigma \in \mathbb{R}^2$) that aligns rigidly the annotation points to a canonical coordinate system (see Fig.2), a respective manual annotation $\mathbf{s} \in \mathbb{R}^{2N}$ in this canonical coordinate system (this means that the contour has N two-dimensional points), and $k \in \{\text{systole,diastole}\}$, which denotes the phase of the cardiac cycle.

A. Overview of the method

The goal of the algorithm proposed herein is to produce a segmentation $\mathbf{s}_t$ for each frame I_t, where $t \in \{1, ..., T\}$ represents the time variable with T denoting the number of frames in the sequence. The

optimal contour at each time instant t is produced as follows:

$$\mathbf{s}_t^* = \int_{\mathbf{s}_t} \mathbf{s}_t p(\mathbf{s}_t|I_{1:t}) d\mathbf{s}_t, \tag{1}$$

The above integral is difficult to compute given the high space dimensionality of s. Thus, we resort to use the sequential importance resampling (SIR) algorithm [3] to estimate it. With SIR, the filtering distribution $p(\mathbf{s}_t|I_{1:t})$ is approximated with a set of P weights and particles $\{w_t^{(l)}, \mathbf{s}_t^{(l)}\}_{l=1}^P$, which can approximate the segmentation as [2]:

$$\mathbf{s}_t^* \approx \sum_{l=1}^{P} \mathbf{s}_t^{(l)} w_t^{(l)}, \tag{2}$$

with $w_t^{(l)} \approx p(\mathbf{s}_t^{(l)}|I_{1:t})$ and $\sum_{l=1}^P w_t^{(l)} = 1$. The mains steps of the SIR method are listed in Algorithm 1

Algorithm 1 SIR Algorithm.

1: **for** $t = 1$ to T **do**
2: sample $\{k_t^{(l)}, \mathbf{s}_t^{(l)}\}_{l=1}^P$
3: Update sample weights $\{\tilde{w}_t^{(l)}\}_{l=1}^P$
4: Normalize sample weights: $w_t^{(l)} = \frac{\tilde{w}_t^{(l)}}{\sum_l \tilde{w}_t^{(l)}}$ for $l \in \{1, ..., P\}$;
5: Compute effective number of particles $N_{eff} = 1/\sum_l (w_t^{(l)})^2$;
6: **if** $N_{eff} < K_{Neff} \times P$ **then**
7: re-sample by drawing P particles from current particle set proportionally to weight and replace particle, and set $w_t^{(l)} = 1/P$ for $l \in \{1, ..., P\}$
8: **end if**
9: Compute LV segmentation for I_t, *i.e.* how it is computed $\mathbf{s}_t^*$ in (2)
10: **end for**

One of the issues of SIR algorithm is that while it is easy to compute $p(\mathbf{s}_t^{(l)}|I_{1:t})$ for a contour $\mathbf{s}_t^{(l)}$, it is hard to sample from this distribution, so it is necessary to have a proposal distribution that approximates well this filtering distribution, and from which is relatively simple to sample. In the next sections we describe the main components of the proposed model comprising the transition model, the observation model and the proposal detailing precisely how do we compute the steps 2, 3 and 9 shown in Algorithm 1.

B. Transition model

One of our contributions is the definition of the transition model. The main assumptions are: the current cardiac phase k_t only depends on the previous cardiac phase k_{t-1} and the current LV contour $\mathbf{s}_t$ depends on the previous contour $\mathbf{s}_{t-1}$ as well as the previous cardiac phase k_{t-1}. Based on these assumptions, the transition model becomes:

$$p(k_t, \mathbf{s}_t|k_{t-1}, \mathbf{s}_{t-1}) = p(k_t|k_{t-1}) p(\mathbf{s}_t|k_{t-1}, \mathbf{s}_{t-1}), \tag{3}$$

where $p(k_t|k_{t-1})$ stands for the switching/staying of the cardiac phases at each time step. The second term in (3), is given as

$$p(\mathbf{s}_t|k_{t-1}, \mathbf{s}_{t-1}) = g(\mathbf{s}_t|f(\mathbf{s}_{t-1}, \mathbf{M}(k_{t-1})), \Sigma_{\mathbf{s}}), \tag{4}$$

where $g(.)$ represents a multivariate Gaussian density function, the function $f(.)$ produces an affine transformation of the contour $\mathbf{s}_{t-1}$; $\mathbf{M}(k_{t-1})$ is a linear transform of the contour $\mathbf{s}_{t-1}$ learned from the training data that allows an expansion (or contraction) according to the phase k_{t-1}; and $\Sigma_{\mathbf{s}}$ is the covariance of the annotations obtained during the training phase.

C. Observation model

The observation model is defined as the process of image formation assumed to be

$$p(I_t|k_t, \mathbf{s}_t) \propto p(k_t, \mathbf{s}_t|I_t)p(I_t), \tag{5}$$

where $p(I_t)$ is constant and the first term expanded as follows

$$p(k_t, \mathbf{s}_t|I_t) = \int_\theta p(k_t|\theta, I_t)p(\mathbf{s}_t|\theta, k_t, I_t)p(\theta|I_t)d\theta. \tag{6}$$

Equation (6) contains three terms: the rigid (affine) detection contained in $p(k_t|\theta, I_t)$, the non-rigid segmentation in $p(\mathbf{s}_t|\theta, k_t, I_t)$ and the the prior distribution of the rigid detection $p(\theta|I_t)$.

The first term in (6) represents the rigid detection that is computed using discriminative classifier. The classifier receives as input an image patch extracted from image I_t using θ and outputs the probability of $k_t \in \{$systole, diastole$\}$. Fig. 2-(b,c) illustrates how the image patch is extracted given the aligned base and apical points. The second term in (6), represents the non-rigid segmentation that is defined as follows:

$$p(\mathbf{s}_t|\theta, k_t, I_t) = \prod_{i=1}^{N} p(\mathbf{s}_t(i)|\theta, k_t, I_t), \tag{7}$$

where $p(\mathbf{s}_t(i)|\theta, k_t, I_t)$ represents the probability that the ith contour keypoint $\mathbf{s}_t(i) \in \mathbb{R}^2$ is located at the LV contour. Fig. 2-(d) shows that for each $i \in \{1, ..., N\}$ the possible $\mathbf{s}_t(i)$ is confined to points lying on a line orthogonal to the LV contour.

Finally, the third term in (6) is defined as $p(\theta|I_t) = g(\theta|\bar{\theta}, \Sigma_\theta)$, where $\bar{\theta}$ and Σ_θ are the mean and covariance values of the training set values for θ, and $g(.)$ denotes the multivariate Gaussian density function.

D. The Filtering distribution

From the transition model defined in Section IV-B, and the observation model detailed in Section IV-C we can now derive the solution to the filtering problem. Denoting the state and observation vectors up to time instant t as $\mathbf{s}_{0:t} \triangleq \{\mathbf{s}_0, ..., \mathbf{s}_t\}$ (similarly for $k_{0:t}$ and $I_{1:t}$), the solution is given by the following Bayesian recursion [2]:

$$\begin{aligned} p(k_t, \mathbf{s}_t|I_{1:t}) &= \frac{p(I_t|k_t, \mathbf{s}_t)p(k_t, \mathbf{s}_t|I_{1:t-1})}{p(I_t|I_{1:t-1})} \\ &= \frac{p(I_t|k_t, \mathbf{s}_t)\sum_{k_{t-1}} \int p(k_t, \mathbf{s}_t|k_{t-1}, \mathbf{s}_{t-1})p(k_{t-1}, \mathbf{s}_{t-1}|I_{1:t-1})d\mathbf{s}_{t-1}}{\sum_{k_t} \int p(I_t|k_t, \mathbf{s}_t)p(k_t, \mathbf{s}_t|I_{1:t-1})d\mathbf{s}_t}, \end{aligned} \tag{8}$$

where $p(k_t, \mathbf{s}_t|I_{1:t-1}) = \sum_{k_t} \int p(k_t, \mathbf{s}_t|k_{t-1}, \mathbf{s}_{t-1})p(k_{t-1}, \mathbf{s}_{t-1}|I_{1:t-1})d\mathbf{s}_{t-1}$. The first term (in the numerator of (8)) is the observation model given in (5), the second is the transition model (see (3)) whilst the third is the available estimate at the previous time step.

Notice that (8) computes the distribution of cardiac phases and LV contours at every single time instant without taking any option regarding the cardiac phases or contour values. This is an important issue, since it endows a robustness in the algorithm.

E. Deep Particle filtering

In the particle filtering setup, the posterior $p(k_t, \mathbf{s}_t|I_{1:t})$ in (8) is approximated with a finite set of P particles, as follows:

$$p(k_t, \mathbf{s}_t|I_{1:t}) \approx \sum_{l=1}^{P} w_t^{(l)} \delta(k_t - k_t^{(l)})\delta(\mathbf{s}_t - \mathbf{s}_t^{(l)}), \tag{9}$$

where $\delta(.)$ is the delta function. The particles $\{k_t^{(l)}, \mathbf{s}_t^{(l)}\}_{l=1}^{P}$ are sampled from a proposal distribution $q(.)$ with $(k_t^{(l)}, \mathbf{s}_t^{(l)}) \sim q(k_t, \mathbf{s}_t | k_{0:t-1}^{(l)}, \mathbf{s}_{0:t-1}^{(l)}, I_{1:t})$, where each particle is expressed as

$$\tilde{w}_t^{(l)} = w_{t-1}^{(l)} \frac{p(I_t|k_t^{(l)}\mathbf{s}_t^{(l)}) p(k_t^{(l)}, \mathbf{s}_t^{(l)} | k_{t-1}^{(l)}\mathbf{s}_{t-1}^{(l)})}{q(k_t^{(l)}, \mathbf{s}_t^{(l)} | k_{0:t-1}^{(l)}\mathbf{s}_{0:t-1}^{(l)}, I_{1:t})}, \tag{10}$$

where $\tilde{w}_t^{(l)}$ and $w_t^{(l)}$ represent the un-normalized and normalized weights, respectively.

In this work, the proposal distribution combines both prediction and filtering distributions as in [27]. The key difference is that we accomplish this task differently. More specifically, the proposal distribution takes into account a transition model (see (3)) for each particle and a detection provided by a discriminative classifier based on a deep belief network (DBN). This, of course, results in a mixture of Gaussians as follows

$$\begin{aligned} q(k_t, \mathbf{s}_t | k_{0:t-1}^{(l)}, \mathbf{s}_{0:t-1}^{(l)}, I_{1:t}) &= q(k_t, \mathbf{s}_t | k_{t-1}^{(l)}\mathbf{s}_{t-1}^{(l)}, I_t) \\ &= \alpha q_{\mathrm{DBN}}(k_t, \mathbf{s}_t | I_t) + (1-\alpha) p(k_t, \mathbf{s}_t | k_{t-1}^{(l)}, \mathbf{s}_{t-1}^{(l)}) \end{aligned} \tag{11}$$

with the detection term (first term in (11)) given by

$$q_{\mathrm{DBN}}(k_t, \mathbf{s}_t | I_t) = \sum_{\{\widetilde{k}_t, \widetilde{\mathbf{s}}_t\}} p(\widetilde{k}_t, \widetilde{\mathbf{s}}_t | I_t) g(\mathbf{s}_t | \widetilde{\mathbf{s}}_t, \Sigma_{\mathbf{s}}) p(k_t | \widetilde{k}_t) \tag{12}$$

where $\{\widetilde{k}_t, \widetilde{\mathbf{s}}_t\}$ stands for the top detections, *i.e.* the local maxima of the observation distribution defined in (6), $g(.)$ is the multivariate Gaussian density function with mean $\widetilde{\mathbf{s}}_t$ and covariance $\Sigma_{\mathbf{s}}$, and $p(k_t|\tilde{k}_t)$ is the transition between cardiac phases defined in (3). The parameter $\alpha \in [0, 1]$ in (11) is used to weight the contribution of the observation and transition models. Note that $\alpha = 0$ represents a distribution that takes into account only the transition model, while $\alpha = 1$ denotes a proposal distribution built based on the observation model only.

Note that the main advantage of this proposal distribution is that when the motion model fails, the observation model has a chance to recover the boundary based on the observations computed from the current image. For the LV tracking this is important because it is hard to obtain a faithful model of the LV motion. Nevertheless, the presence of the transition model is still quite important to deal with the detection and segmentation failures of the observation model.

F. Segmentation of the Left ventricle

Using the particle filter setup in (9) for the filtering distribution, we estimate the values for the state variables at each time step t as follows:

$$\mathbf{k}_t^* = \arg \max_{k \in \{systole, diastole\}} E_{p(k_t, \mathbf{s}_t | I_{1:t})}[k], \tag{13}$$

where $E_{p(k_t, \mathbf{s}_t | I_{1:t})}[k] \approx \sum_{l=1}^{P} w_t^{(l)} \delta(k - k_t^{(l)})$, and

$$\mathbf{s}_t^* = E_{p(k_t, \mathbf{s}_t | I_{1:t})}[\mathbf{s}_t | k_t^*] \approx \frac{1}{\sum_{l=1}^{P} w_t^{(l)} \delta(k_t - k_t^*)} \sum_{l=1}^{P} w_t^{(l)} \mathbf{s}_t^{(l)} \delta(k_t - k_t^*). \tag{14}$$

Given all the main ingredients of the statistical model in Sections IV-B-IV-F, we can now provide more details about the SIR in Algorithm 1. Thus, the sampling in step §2 is done by using (11); the weights update in step §3 is accomplished using (10), and finally the segmentation in step §9, is performed using (13), (14), to update $k_t^\star$ and $\mathbf{s}_t^\star$, respectively.

In the following section we provide details concerning the training and inference of the Deep belief network, for the rigid and non-rigid classifiers (6) in the observation model in (5).

V. Training and inference on deep belief networks

Deep belief networks (DBN) are artificial networks containing a large number of hidden layers and nodes, which allows the construction of models of large capability. However, the backpropagation algorithm for estimating the classifier parameters is limited, in sense that it only provides reliable estimates when the initial guess is close enough of the local optimum of the objective function to be minimized. Hinton et al. [28] found an efficient mechanism to surpass this difficulty using unsupervised training of multiple layers of restricted Boltzmann machines (RBM), which are represented by a hidden and a visible layer of stochastic binary units with connections only between layers (*i.e.*, no connections within layers).

After the parameters of several layers of RBMs are learned, the whole network is trained using backpropagation to adjust the weights to a local maximum for the regressor and classifier functions. The use of deep learning are grounded on the two following two main ideas [29]:

§ **1** An unsupervised generative model learns the process of the LV image generation, and
§ **2** a discriminative model that is trained on the above generative model.

A. Training the DBN

The training stage of the DBN comprises both the training of the rigid and non-rigid classifiers (see (6)). For training the rigid classifier (first term in (6)) we follow the strategy proposed by Carneiro *et al.* [30]. Basically, this method builds an image scale space $L(\mathbf{x},\sigma) = G(\mathbf{x},\sigma) * I(\mathbf{x})$, where $G(\mathbf{x},\sigma)$ is a Gaussian kernel, and $*$ is the convolution operator, and $I(\mathbf{x})$ is the input image, σ is the image scale parameter. In this methodology we have to define a set of images scales, say $\{\sigma_1,...,\sigma_Q\}$. More specifically, in the present approach we have three scales, *i.e.* $\sigma = \{4, 8, 16\}$, ($Q = 3$) and we train three classifiers separately. For training each classifier we generate a set of positive and negative training sets, that are defined based on a scale-dependent margin m_σ. The generation process of each training sample (positive or negative), can be viewed as function $\varphi(.)$ that receives a triplet containing the jth image I_j, scale σ_q and a set of affine parameter θ, and outputs the image patch $\mathcal{P}_{n_q\times n_q}$ of size $n_q \times n_q$ containing the LV, *i.e.* formally $\mathcal{P}_{n_q\times n_q} = \varphi(I_j,\sigma_q,\theta)$. Basically, the function $\varphi(.)$ comprises a scale operation on image I with σ_q and then perturbing by an affine transformation in $\theta \in \mathbb{R}^5$. This produces a patch of size $n_q \times n_q$, where n is a vector indexed by $q \in \{1,..,Q\}$. After this process, we finally subtract each patch pixel with its mean. This contrast normalization provides more robustness against brightness variations.

The difference between the generation positives and negatives are the ranges in which the affine transformation parameters in θ are defined. Thus, for positive samples, denoted herein as $\mathcal{P}os(k,q,j)$, and considering a training image I_j, an image scale σ and a cardiac phase $k \in \{\mathrm{diastole}, \mathrm{systole}\}$, the samples are randomly generated inside the following range

$$\mathcal{P}os(k,q,j) = \{\mathcal{P}_{n_q\times n_q} | \theta \sim \mathrm{Dist}(\mathcal{R}(\theta)), |\theta - \theta_\mathrm{j}| < \mathrm{m_p}, \mathrm{k_j} = \mathrm{k}\} \tag{15}$$

where $\mathrm{Dist}(.)$ is an uniform distribution, over the range $\mathcal{R}(\theta) = [\max(\{\theta_j\}_{j=1,..,M}) - \min(\{\theta_j\}_{j=1,..,M})] \in \mathbb{R}^5$

For the negative samples the range of parameters $\mathcal{R}$ is now larger and becomes as

$$\mathcal{N}eg(q,j) = \{\mathcal{P}_{n_q\times n_q} | \theta \sim \mathrm{Dist}(\mathcal{R}(\theta)), |\theta - \theta_\mathrm{j}| > 2\mathrm{m_p}\} \tag{16}$$

where m_p in (15),(16) is the margin between positive and negative samples that facilitates the training process by avoiding similar examples with opposite labels. This avoids overtraining in the classifiers.

Notice that, with the above procedure, we are able to train the rigid (affine) DBN. To accomplish this, we stack several hidden layers to reconstruct the input patches in the $\mathcal{P}os$ and $\mathcal{N}eg$ sets. This corresponds to unsupervised training (issue § **1** above). After this, three nodes are added to the top layer of the DBN to account for: (1) $p(k = \mathrm{systole}|\mathrm{I},\theta)$, (2) $p(k = \mathrm{diastole}|\mathrm{I},\theta)$, and (3) the formed patch $\mathcal{P}_{n_q\times n_q}$ does not contain the LV. Finally, the discriminative training on the generative model is to find the appropriate weights of the DBN (issue § **2** above).

TABLE I
LEARNED CONFIGURATION FOR THE DEEP BELIEF NETWORKS.

Affine Classifier						
σ	Visible Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Hidden Layer 4	Output Layer
4	196 (14×14 pix.)	100	100	200	200	3
8	49 (7×7 pix.)	50	100	-	-	3
16	16 (4×4 pix.)	100	50	-	-	3

Non-rigid Classifier						
σ	Visible Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Hidden Layer 4	Output Layer
4	41	50	50	-	-	1

The non-rigid regressor is trained only at the finest scale $\sigma = 4$ (second term in (6)). In the training process we build an orthogonal line radiating from each contour sample (see Fig. 2 (d)). This allows us to define the ith sample of the jth contour point $\mathbf{s}_j(i)$. The input of the regressor is the profile (gray levels) of the orthogonal line (see Fig. 2 (d) for an illustration), and returns the location in each normal line that is closest to the LV boundary.

VI. EXPERIMENTAL EVALUATION

In this section we first examine the data used for the experimental evaluation, the annotation procedure and the team involved. We next describe the learned configurations of the deep belief networks, for both rigid and non-rigid classifiers. Finally we perform a interuser statistics study comprising the *modified Williams index* the *Bland-Altman* and *Scatter plots* and a comparison with the state of the art.

A. Datasets and manual annotations

The data set used for the experiments comprises 20 sequences for training and testing (20 sequences from 20 subjects with no overlap), from which 16 present some kind of cardiopathy. According to the cardiologist's report[1], the following cardiopathies/abnormalities are considered:

§1. Dilation of the LV. The dilation can be classified in mild, moderate or severe;
§2. Presence of hypertrophy of the LV. The hypertrophy can also be classified into mild, moderate or severe;
§3. Wall motion abnormalities. The abnormalities can be classified as global (affecting all the LV segments), or localized (affecting some of the LV segments);
§4. Function of the LV. The function may be preserved, mild, depressed, or severe (*i.e.* dysfunction of the LV);
§5. Presence of valvular heart disease; and
§6. Presence of a pacemaker device.

The data set is splitted into two sets: $\mathcal{T}_1$ and $\mathcal{T}_2$. The set $\mathcal{T}_1$ contains 16 sequences presenting some cardiopathy as mentioned above and other two sequences from healthy subjects, we term these sequences as $\mathcal{T}_{1_{A\ldots R}}$. The other set comprises two healthy sequences $\mathcal{T}_{2_{A,B}}$.

We worked with four members of the cardiology services from Hospital Fernando Fonseca [2] who annotated five sequences in $\mathcal{T}_1$, they annotated roughly 20 frames in each sequence; one annotation in the systole and in diastole phases of the cardiac cycle. In the set $\mathcal{T}_{2_{A,B}}$ the head of the team provide us 40 annotations (20 for each sequence in $\mathcal{T}_2$). We access how the results of the proposed algorithm correlate with interuser variation in the set $\mathcal{T}_1$. Also, we perform a quantitative comparison between the estimates obtained in $\mathcal{T}_2$, thus, the results reported in five sequences of the LV.

[1]This was done in collaboration with Dr. António Freitas from Hospital Fernando Fonseca who detailed each of the sequences

[2]The annotations were possible thanks to the collaboration with the cardiology service in Hospital Fernando Fonseca headed by Dr. António Freitas

B. Training of the DBN

This section describes the two training phases of the DBN, comprising the afine (rigid) and non-rigid stages. For training rigid classifier, three separate classifiers are trained, each one at different scale, *i.e.* $q \in \{1, ..., Q\}$. To accomplish this, we generated 100 positive and 500 negative patches that will integrate the sets $\mathcal{P}os$ and $\mathcal{N}eg$, respectively (see (15),(15)). This initial training set, is further, divided 80% in $\mathcal{P}os$ and $\mathcal{N}eg$ for training and validation sets. The multiscale strategy mentioned in Section V-A is used in training and segmentation purposes at three different scale; $\sigma = \{16, 8, 4\}$. The dimension of the patches depends on the scale used. Thus, the original patch size is 56×56, but it decreases as the scale increases. More specifically, we used patches of dimension, 4×4, 7×7 and 14×14, for scales 16, 8 and 4, respectively. The validation set, is used to estimate the parameters of the DBN, namely: (i) number of visible, hidden and output layers; (ii) number of nodes for the visible, hidden and output layers. These sets of parameters are shown in Table I for both rigid and non-rigid classifiers. This was achieved using the annotations data set contained in $\mathcal{T}_1$.

The non-rigid classifier (see (7) and Fig. 2 (d,e)) is trained with the approach described in Section V-A. Each normal line has 41 pixel-length. In this training stage we also use the same positive set $\mathcal{P}os$, 80 samples for training and 20 for validation.

C. Error measures

To perform a quantitative assessment we use several error mesures that compute the mismatch between the estimated LV and the ground truth contours. Among possible choices for the quantitative study, we will use metrics that are common and widely known in the literature. More specifically, we use the Hammoude distance (HMD) [31], average error (AV) [26], mean absolute distance (MAD) [32], and average perpendicular error (AVP) between the estimated and ground truth contours, (see [26], [33] for a complete definition of the these measures).

D. Comparison with Inter-user Statistics

The robustness of the proposed approach against the interuser variability is assessed following the works of Chalana *et al.* [34] and Lopez *et al.* [35]. The measures used are the following: (i) *modified William index*, (ii) *Bland-Altman plot* [36], and (iii) *scatter plot*. These comparisons are performed on the diseased sets contained in $\mathcal{T}_{1,\{A,B,C\}}$, for which we have four LV manual annotations per image delineated by four different Cardiologists (Sec. V-A).

Fig. 3 illustrates the inter-variance of the delineations given by the four cardiologists. This figure shows that indeed the delineation of the LV contour is somehow subjective. Parts of the contour belonging to the apex and the right lateral part are the regions in which the discrepancy is larger.

For each sequence we have about 17 expertise annotations. Considering four cardiologists, this gives 51 manual annotation for each sequence, resulting in a total amount of 204 annotations in $\mathcal{T}_1$

In order to have a fair comparison, we train three separate DBN classifiers using the following training sets: 1) $\mathcal{T}_1 \setminus \mathcal{T}_{1,A}$, 2) $\mathcal{T}_1 \setminus \mathcal{T}_{1,B}$, and 3) $\mathcal{T}_1 \setminus \mathcal{T}_{1,C}$, where $\setminus$ represents the set difference operator. These three classifiers are necessary because when testing any image inside each one of these three sequences, we cannot use any image of that same sequence in the training process.

1) Modified Williams Index: To provide the results concerning the modified Williams index we start by considering the set $\{\mathbf{s}_{j,k}\}$, where $j \in \{1..M\}$ indexes the testing images, and $k \in \{0..U\}$ indexes the manual annotations, where the index $k = 0$ denotes the computer-generated contour (i.e., each one of the M images has U manual annotations). The function $D_{k,k'}$ measures the mismatch or disagreement between users k and k', which is defined as

$$D_{k,k'} = \frac{1}{M} \sum_{j=1}^{M} d_{-}(\mathbf{s}_{j,k}, \mathbf{s}_{j,k'}), \tag{17}$$

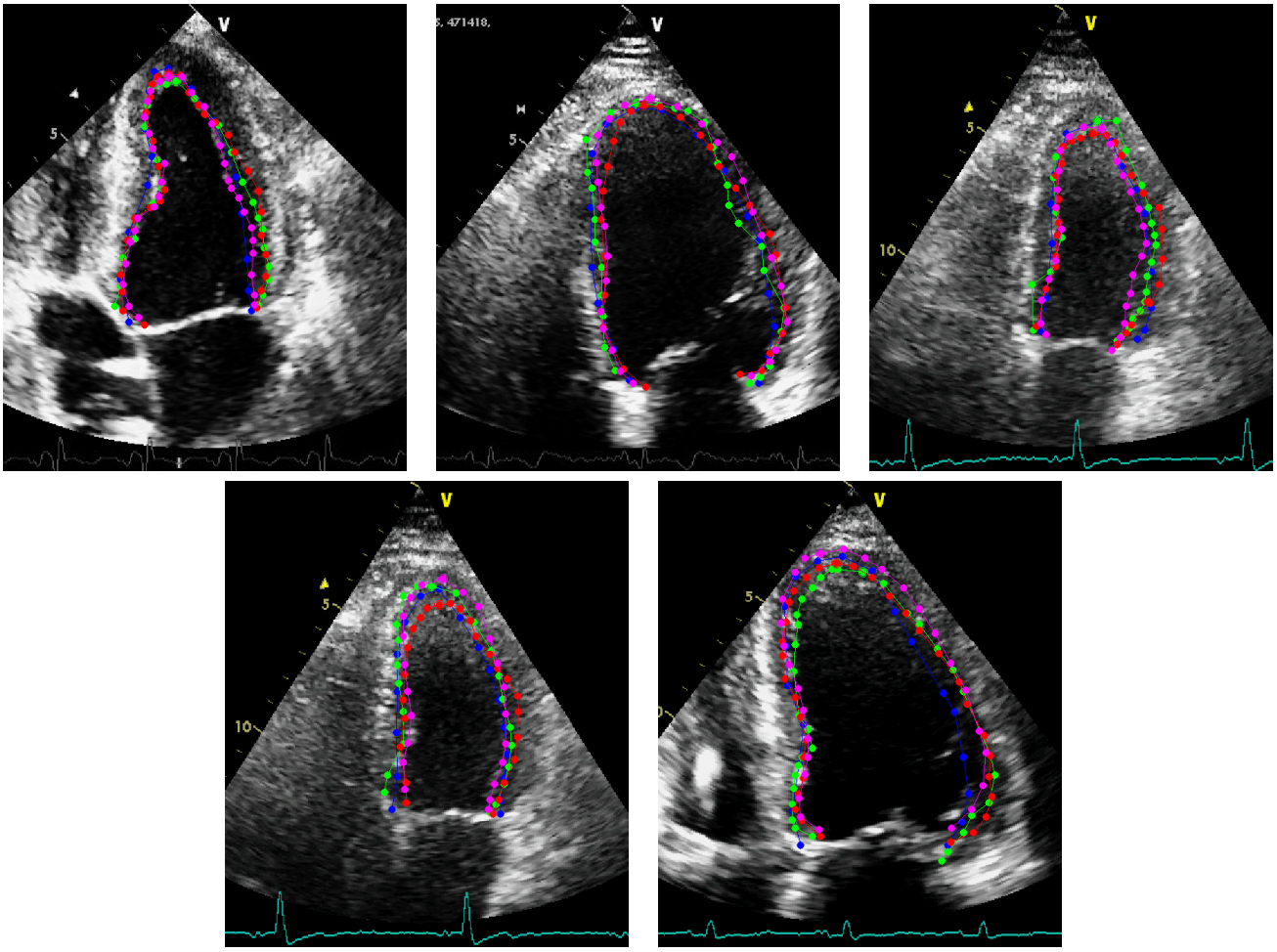


Fig. 3. In each image, four manual delineations of the LV contour are shown, where each color refers to the annotation of an expert.

where $d_{-}(.,.)$ is an error measure between two annotations $\mathbf{s}_{j,k}$, $\mathbf{s}_{j,k'}$. The measure can be, for instance, the average distance between the estimated contou and the ground-truth contour. The modified Williams index is defined as

$$I' = \frac{\frac{1}{U}\sum_{k=1}^{U}\frac{1}{D_{0,k}}}{\frac{2}{U(U-1)}\sum_{k}\sum_{k':k'\neq k}\frac{1}{D_{k,k'}}}. \tag{18}$$

A confidence interval (CI) is estimated using a jackknife (leave one out) non-parametric sampling strategy [34] as follows:

$$I'_{(.)} \pm z_{0.95}se, \tag{19}$$

where $z_{0.95} = 1.96$ represents 95^{th} percentile of the standard normal distribution, and

$$se = \left\{\frac{1}{M-1}\sum_{j=1}^{M}[I'_{(j)} - I'_{(.)}]\right\} \tag{20}$$

with $I'_{(.)} = \frac{1}{M}\sum_{j=1}^{M} I'_{(j)}$, and $I'_{(j)}$ is the Williams index (18) calculated by leaving image j out of computation of $D_{k,k'}$. A successful measurement for the Williams index is to have the average and confidence interval (19) close to one.

TABLE II
COMPARISONS IN THE SEQUENCES. EACH CELL SHOWS THE MEAN VALUE AND THE STANDARD DEVIATION IN PARENTHESES.

Sequence One			
Approach	Hamm.	Aver.	Hausd.)
Proposal	**0.17(0.04)**	3.3(0.8)	**19.1(2.2)**
[40]	0.18(0.06)	**3.2(0.8)**	20.0(2.6)
[26]	0.24(0.03)	4.8(0.9)	22.4(2.1)
Sequence Two			
Approach	Hamm.	Aver.	Hausd.
Proposal	**0.15(0.03)**	**2.9(0.5)**	**19.4(1.4)**
[40]	0.17(0.02)	3.0(0.5)	19.8(1.1)
[26]	0.24(0.03)	4.8(0.7)	20.2(1.4)

TABLE III
COMPARISON OF THE COMPUTER GENERATED CURVES TO THE USERS' CURVES WITH RESPECT TO ALL THE ERROR MEASURES FOR THREE SEQUENCES USING THE AVERAGE AND 0.95% CONFIDENCE INTERVAL (IN PARENTHESIS) OF THE WILLIAMS INDEX.

measure	d_{HMD}	d_{AV}	d_{MAD}	d_{AVP}
Average(CI)	$0.83\ (0.82, 0.84)$	$0.91\ (0.90, 0.92)$	$0.94\ (0.93, 0.95)$	$0.83\ (0.82, 0.84)$

2) Bland-Altman and Scatter Plots: Quantitative results to provide the Bland-Altman [36] and scatter plots from which we also compute the correlation coefficient and the p-value. To accomplish this we have: (i) the gold standard LV volume computed via an iterative process using the manual annotations [34]; (ii) the Cardiologists' LV volumes, and (iii) the computer generated LV volume. To estimate the LV volume from 2-D contour annotation we use the area-length equation [37], [38] with $V = \frac{8A^2}{3\pi L}$, where A denotes the projected surface area, L is the distance from upper aortic valve point to apex, and V is expressed in cubic pixels. The p-values are computed as follows: 1) compute several independent p-values from 3 samples, each taken from separate sequence; and then 2) combine the p-values using the Fisher's method into a single result by assuming independence among the p-values [39].

E. Experimental results

Fig. 4 shows the qualitative performance of the proposed approach at segmenting the LV. In green contour it is displayed the expertise annotation superimposed with the estimated contour in red. Each row corresponds to a different cardiac sequence, and all the sequences are contained in the set $\mathcal{T}_{1,\{A,B,C\}}$. It is seen that the proposed approach exhibits quite remarkable segmentations in these sequences.

We show a comparison with the state of the art in Table II using the error metrics Hammoude (Hamm), average (Aver.) and Hausdorff (Hausd.) In this study, we used the two test sequences $\mathcal{T}_{2_{A,B}}$ and compared with our previous methods [40], [26]. The approach in [40] only contains the observation model, which means that does not contain any dynamic model, only the static segmentation. In [26] a deformable based approach is proposed, which contains a transition model that is based on a switching mechanism. More specifically, at each time instant (each frame) the deformable model is able to switch the phase (*i.e.* $k_{=\text{systole}}$) two hypothesis are considered.

In terms of inter-user statistics, Table III shows the mean values as well as and confidence intervals of the Williams index defined in (18)-(19). This is done for all ultrasound sequences considered for the comparison with inter-user statistics. Fig. 5 shows the scatter and Bland-Altman plots. Concerning the scatter plot, the correlation coefficient between the users and gold standard is 0.99 with p-value$=$ 3.11×10^{-68} (left images images in the figure) and for the gold standard versus computer the correlation is 0.95 with p-value$=$ 1.9×10^{-4} (right image in the figure). In the Bland-Altman plots, the Inter-user plot

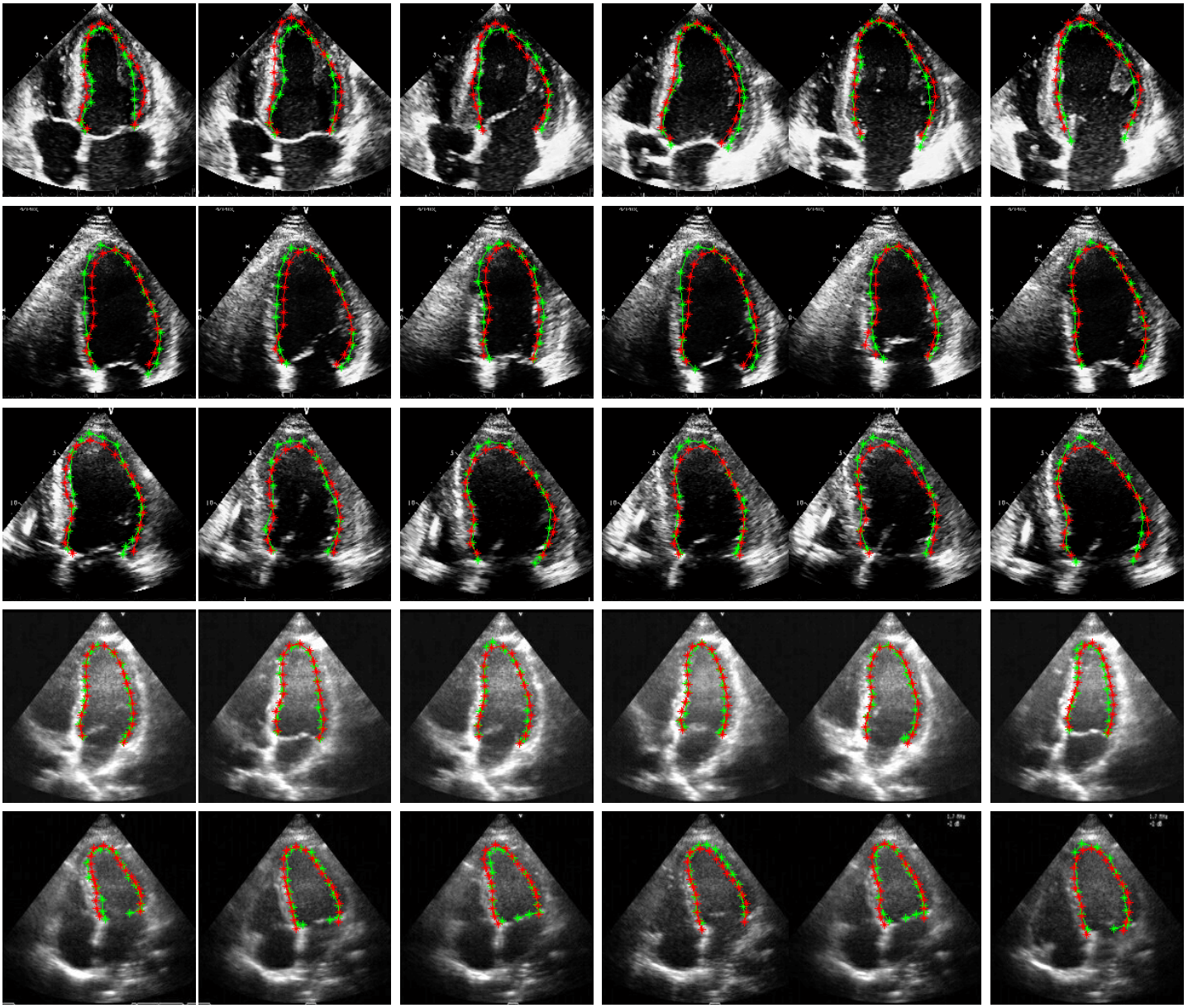


Fig. 4. LV contour estimates (red) are compared against the expertise ground-truth (green). Also it is shown the cardiac phase and the correspondent confidence in the top of each snapshot.

produced a bias of 4.9×10^4 with confidence interval of $[-5 \times 10^5, 5 \times 10^5]$, while the Gold vs Computer plot shows a bias of 1.8×10^5 and confidence interval of $[-4 \times 10^5, 7 \times 10^5]$.

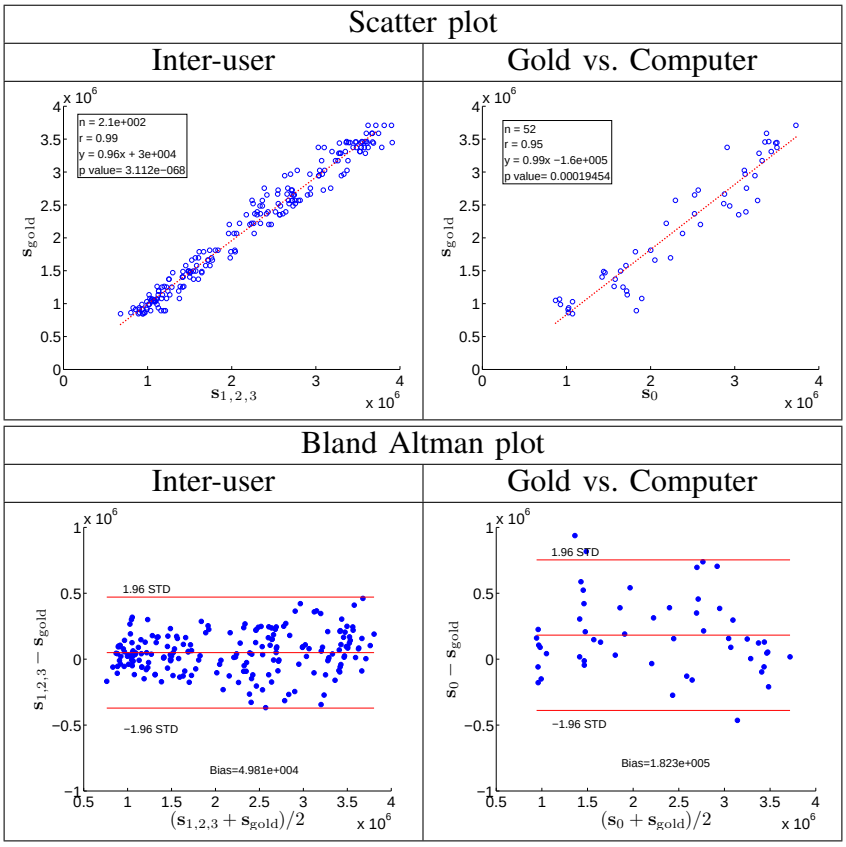


Fig. 5. Scatter plots with linear regression and Bland-Altman bias plots

VII. Conclusions and Further work

We have described an automatic algorithm for the segmentation and tracking of the LV in 2D ultrasound data. The contribution presented herein resides is in the ability to combine deep learning architectures with sampling importance resampling (SIR) techniques incorporating dynamical model. Under this framework, a novel transition and observation models are presented. In the dynamical model, there is no commitment to a specific heart dynamical regime, instead it combines the transition and observation results. Notice however, that there are some limitations concerning the training sets used. In particular, the present framework is still dependent on a rich training data set, being preferable to have several frames from distinct patients rather several frames from the same patient. Also, when generating the positive and negatives samples (in Section V-A eqs. (15),(16)) that are estimated from the training set, the range may not be enough to robustly represent all the variation that can happen with the left ventricle. In further work we plan to address the above issues by developing a semi-supervised based techniques to reduce such dependence. We also plan to apply this approach to other anatomies and other medical imaging techniques.

References

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

[2] A. Doucet, N. de Freitas, N. Gordon, and A. Smith, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[3] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE T. Sig. Proc.*, vol. 50, no. 2, pp. 174–188, 2002.

[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 4, no. 1, pp. 321–331, 1987.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Eur. Conf. Comp. Vis.*, 1998, pp. 484–498.

[6] N. Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 773–776, 2003.

[7] R. Malladi, J. Sethian, and B. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, pp. 158–175, 1995.

[8] O. Bernard, D. Friboulet, P. Thevenaz, and M. Unser, "Variational B-spline level-set: A linear filtering approach for fast deformable model evolution," *IEEE Trans. Imag. Proc.*, vol. 18, no. 6, pp. 1179–1991, 2009.

[9] C. Corsi, G. Saracino, A. Sarti, and C. Lamberti, "Left ventricular volume estimation for real-time three-dimensional echocardiography," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1202–1208, 2002.

[10] D. Cremers, S. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 335–351, 2006.

[11] E. Debreuve, M. Barlaud, G. Aubert, I. Laurette, and J. Darcourt, "Space-time segmentation using level set active contours applied to myocardial gated SPECT," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 643–659, 2001.

[12] Q. Duan, E. D. Angelini, and A. Laine, "Real time segmentation by active geometric functions," *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 223–230, 2010.

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

[14] M. Lynch, O. Ghita, and P. Whelan, "Segmentation of the left ventricle of the heart in 3-D+T MRI data using an optimized nonrigid temporal model," *IEEE Trans. Med. Imag.*, vol. 27, no. 2, pp. 195–203, 2008.

[15] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *International Journal of Computer Vision*, vol. 46, no. 3, pp. 223–247, 2002.

[16] A. Sarti, C. Corsi, E. Mazzini, and C. Lamberti, "Maximum likelihood segmentation of ultrasound images with Rayleigh distribution," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 52, no. 6, pp. 947–960, 2005.

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

[18] T. Cootes, C. Beeston, G. Edwards, and C. Taylor, "A unified framework for atlas matching using active appearance models," in *Information Processing in Medical Imaging*, 1999, pp. 322–333.

[19] S. Mitchell, B. Lelieveldt, R. van der Geest, H. Bosch, J. Reiber, and M. Sonka, "Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 415–423, 2001.

[20] D. Comaniciu, X. Zhou, and S. Krishnan, "Robust real-time myocardial border tracking for echocardiography: An information fusion approach," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 849–860, 2004.

[21] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta, "Databased-guided segmentation of anatomical structures with complex appearance," in *Conf. Computer Vision and Pattern Rec. (CVPR)*, 2005.

[22] D. Terzopoulos and R. Szeliski, *Tracking with Kalman Snakes*. MIT Press, 1993.

[23] W. Sun, M. Cetin, R. Chan, V. Reddy, G. Holmvang, V. Chandar, and A. Willsky:, "Segmenting and tracking the left ventricle by learning the dynamics in cardiac images," *Inf Process Med Imaging*, vol. 19, pp. 553–565, 2005.

[24] L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu, "3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers," in *CVPR*, 2008.

[25] J. Sénégas, T. Netsch, C. Cocosco, G. Lund, and A. Stork, "Segmentation of medical images with a shape and motion model: A Bayesian perspective," in *ECCV Workshops CVAMIA and MMBIA*, 2004, pp. 157–168.

[26] J. C. Nascimento and J. S. Marques, "Robust shape tracking with multiple models in ultrasound images," *IEEE Trans. Imag. Proc.*, vol. 17, no. 3, pp. 392–406, 2008.

[27] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *ECCV*, 2004.

[28] R. Salakhutdinov and G. Hinton, "Learning a non-linear embedding by preserving class neighbourhood structure," in *AI and Statistics*, 2007.

[29] G. Hinton. http://videolectures.net/nips09_hinton_dlmi/.

[30] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Trans. Med. Imaging*, vol. 27, no. 9, pp. 1342–1355, 2008.

[31] A. Hammoude, "Computer-assited endocardial border identification from a sequence of two-dimensional echocardiographic images," Ph.D. dissertation, University Washington, 1988.

[32] X. S. Zhou, D. Comaniciu, and A. Gupta, "An information fusion framework for robust shape tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 115–129, 2005.

[33] G. Carneiro and J. C. Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2592–2607, 2013.

[34] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Trans. Med. Imag.*, vol. 16, no. 10, 1997.

[35] C. Alberola-Lopez, M. Martin-Fernandez, and J. Ruiz-Alzola, "Comments on: A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Trans. Med. Imag.*, vol. 23, no. 5, pp. 658–660, 2004.

[36] J. Bland and A. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307–310, 1986.

[37] J. C. Reiber, A. R. Viddeleer, G. Koning, M. J. Schalij, and P. E. Lange, "Left ventricular regression equations from single plane cine and digital X-ray ventriculograms revisited," *Clin. Cardiology*, vol. 12, no. 2, pp. 69–78, 1996, kluwer Academic Publishers.

[38] H. Sandler and H. T. Dodge, "The use of single plane angiocardiograms for the calculation of left ventricular volume in man," *Amer. Heart J.*, vol. 75, no. 3, pp. 325–334, 1968.

[39] R. Fisher, "Questions and answers number 14," *The American Statistician*, vol. 2, no. 5, pp. 30–31, 1948.

[40] G. Carneiro, J. C. Nascimento, and A. Freitas, "Robust left ventricle segmentation from ultrasound data using deep neural networks and efficient search methods," in *IEEE Int. Symp. on Biomedical Imaging, from nano to macro (ISBI)*, 2010, pp. 1085–1088.