Segmentation Of B-Mode Cardiac Ultrasound Data by Bayesian Probability Maps

Abstract

In this paper we present a model for describing the position distribution of the endocardium in the 2 chamber apical long-axis view of the heart in clinical B-mode ultrasound cycles. We propose a novel Bayesian formulation, including priors for spatial and temporal smoothness, and preferred shapes and position. The shape model takes into account both endocardium, atrial region and apex. The likelihood is built using a statistical signal model, which attempts to closely model a censored signal. In addition, the use of a censored Gamma mixture model with unknown censoring point, to handle artefacts resulting from left-censoring of the in US clinical B-mode, is to our knowledge novel. The posterior density is sampled by the Gibbs method to estimate the expected latent variable representation of the endocardium, which we call the Bayesian Probability Map; the map describes the probability of pixels being classified as being within the endocardium. The regularization parameters of the model are estimated by cross-validation, and the results are compared against the 2 chamber apical model of Chen et al.

Keywords: Bayes, segmentation, cardiac, B-mode ultrasound, latent variable, censored data, mixture model, Gibbs sampling.
1. Introduction

Echocardiography is more accessible, mobile and inexpensive compared to other medical imaging techniques and has become a widely used diagnostic method in cardiology in recent years. Unfortunately ultrasound (US) images struggle with inherent problems which in large part stem from noise, and is often referred to as speckle contamination. Segmentation of organ structures in ultrasonic data cannot be performed without dealing with the effects of speckle, either by removing or suppressing it, or using the information that speckle provides (Noble and Boukerroui, 2006). In this paper we adopt the latter approach.

US noise patterns, or speckle noise, are characteristic for various types of tissue. The noise is dependent on many factors, such as spatial arrangement and density of scatterers within the resolution cell of the US system (Burckhardt, 1978; Wagner et al., 1983; Shankar, 2000). To address this problem, different statistical models have been proposed to model the distribution of speckle in envelope-detected US signal.

Since lossy coding, such as log-compression, is used to construct B-mode images, the models proposed differ from those that apply to uncompressed envelope detected radio frequency (RF) US data. In Tao et al. (2006), the Weibull, Normal and Log-Normal and Gamma distribution were evaluated for modeling clinical B-mode data, with the Gamma distribution being found to be the most flexible. The Rayleigh, Nakagami, Inverse Gaussian, Gamma, and K distributions were tested in Nillesen et al. (2008) for use as models for envelope detected cardiac RF US data. Gamma was found to be the most appropriate. Furthermore, the Fisher-Tippett distribution (Dutt, 1995) has
been proposed since it is an exact model of the log-compressed envelope in the case of fully developed speckle (high density of random scatterer within the resolution cell). The Log-Normal distribution has additionally been used to model speckle in ovarian (Zimmer et al., 2000), breast and cardiac (Xiao et al., 2002) clinical B-mode US.

Several works have focused on modeling the speckle noise as an integral part of the segmentation model, see Noble (2010); Noble and Boukerroui (2006) for an exhaustive survey. In Zhu et al. (2009), mixtures of Nakagami densities are used to model the blood and background distribution, and a mixture of shifted Rayleigh distribution is used in Zhu et al. (2007). In Paragios et al. (2005) the Exponential distribution is used to model the blood pool, while the Gaussian distribution is used for the tissue/walls. Many previous works use single component distributions, such as the Rayleigh (Dias and Leitao, 1996; Figueiredo et al., 2000), Shifted Rayleigh (Mignotte and Meunier, 2001), Gaussian (Boukerroui et al., 2003; Ashton and Parker, 1995; Rekeczky et al., 1999; Hansen et al., 2002), Gamma (Tao et al., 2003), and Beta (Martin-Fernandez and Alberola-Lopez, 2002) distribution. In this paper, we will consider 5 distributions for use in a mixture model of clinical B-mode data: the Gamma, Nakagami, Rayleigh, Log-Normal and Fisher-Tippett distribution.

In addition to log-compression, the clinical B-mode US data is often also subject to proprietary post-processing algorithms and quantization, which adversely affects the statistical properties of the signal. One specific artifact is left-censoring of data, which means that data is cut off at an (to the user) unknown amplitude. All amplitudes below this unknown level, are gathered
in the first bin of the B-mode image histogram. *Note that we use intensities and amplitude in this paper to denote the same quantity, since amplitudes are displayed as intensity values in the B-mode ultrasound image.* For display purposes, this has little effect, but it poses problems for statistical modeling of the censored data, as the first bin often will deviate sharply from the remainder of the data. In an image with areas of low signal return (i.e. signal dropout) this will result in a very high count in the first bin.

In the field of segmentation of ultrasonic data, the heart has long been a major subject, specifically the endocardium. Information about position of the endocardium may be used for determining ejection fraction, by comparing systolic and diastolic volume, and assessment of regional wall abnormalities of the heart; measures used in diagnosis of ischaemic heart disease. There are many approaches to solving the problem of segmentation of structures, e.g. the endocardium, in the left-ventricle of the heart. These include Active Contours (Mishra et al., 2003), Variational Level Set formulations (Yan and Zhuang, 2003; Lin et al., 2003; Chen et al., 2007), Bayesian (Boukerroui et al., 2003; Song et al., 2002; Xiao et al., 2000; Figueiredo and Leitao, 1992; Mignotte and Meunier, 1999), Deep Belief Networks (DBNs) (Carneiro et al., 2012), and variants of Active Appearance Motion models (Bosch et al., 2002; Mitchell et al., 2002) An excellent overview of a large part of existing methods is given in Noble and Boukerroui (2006). Many works concern the segmentation of the endocardium in the parasternal short axis view of the heart (see Klingler et al., 1988; Binder et al., 1999; Mishra et al., 2003; Mignotte and Meunier, 1999; Yan and Zhuang, 2003). The more challenging apical long-axis view, more subject to signal dropout and out-of-plane motion
(resulting in foreshortening, which affect the appearance of heart walls) has been considered in Bosch et al. (2002); Zhu et al. (2009); Boukerroui et al. (2003); Chen et al. (2007); Lin et al. (2003); Belaid et al. (2011); Carneiro et al. (2012); Dietenbeck et al. (2012). This view commonly suffers from partially missing heart walls, which have been dealt by use of shape prior formulations in (e.g., Bosch et al., 2002; Zhu et al., 2009; Chen et al., 2007; Lin et al., 2003; Carneiro et al., 2012; Dietenbeck et al., 2012) and spatio-temporal continuity (Bosch et al., 2002; Zhu et al., 2009; Boukerroui et al., 2003; Chen et al., 2007; Lin et al., 2003).

Lin et al. (2003) presents a multi-scale level set model, with shape prior, for segmentation of endocardium in 2 chamber long-axis apical view. The segmentation functional is a combination of an edge and Chan-Vese region based term. The signal model is Gaussian for higher scale levels, using interaction between scales aims to compensate for lack of validity of the Gaussian assumption for lower scales.

Chen et al. (2007) proposes a variational shape prior model, applied to the segmentation of the endocardium in 2-chamber apical view of the heart. The proposed segmentation functional contains a smoothness constraint, shape and intensity prior. The shape and intensity prior is learned from manually traced examples of the endocardium. The model makes no distribution assumption on the data, and thus estimated density function are non-parametric. The shape prior weight is optimized by selecting the weight which generates a contour with maximal mutual information between density functions computed over a narrow band close to the contour, and corresponding densities in a prior intensity profile.
Boukerroui et al. (2003) formulates a Bayesian MRF model, in which the conditional density distribution of the intensities is assumed Gaussian. It is argued that since the data has been decimated through the use of multi resolution discrete wavelet transform (DWT) pyramid, the Gaussian distribution model is approximatively valid. One of the objects of study is segmentation of endocardium in 4-chamber apical views. The posterior energy function consists of a data fidelity component and a smoothness constraint. The energy is minimised using the Iterated Conditional Mode algorithm. This model is an extension of Boukerroui et al. (1999); the main addition is an adaptive weight which balances the local and global statistics in the image. This model does not employ any shape priors.

Bosch et al. (2002) presents Active Appearance Motion Models (AAMM) which is an adaption of the Active Appearance Model (AAM) developed by Cootes et al. (1994, 1995), for segmentation of the endocardium in 4-chamber apical view. AAMM represents shape, appearance and motion, and is trained from manually trained contours. The B-mode ultrasound data used in Cootes et al. (1994, 1995) was subjected to nonlinear intensity normalisation, to compensate for pre- and post processing which results in offset and gain variations causing intensity ranges to differ greatly. The process results in the intensities being transformed into a form which follows a Gaussian distribution. Note that the motivation behind nonlinear intensity normalisation is similar to the one for our proposed formulation of a censored mixture model.

Mitchell et al. (2002) introduced a 3D Active Appearance Model. Previous AAM models (e.g., Mitchell et al., 2001) only exploited temporal co-
herence and are not a full 3D approach. The model was validated with the same data as in Bosch et al. (2002). The difference in results was attributed to the lack of nonlinear intensity normalisation in Mitchell et al. (2002).

Dietenbeck et al. (2012) uses a level set model, with hyperquadric shape prior to segment the myocardium in 4-chamber and 2-chamber apical views. The authors argue that a simple shape prior model makes their model view-independent. The considered US images are subjected to a Gaussian filter and then the classic Chan-Vese model is applied, and thus no distributional assumptions are made on the signal.

Carneiro et al. (2012) uses Deep Belief Networks (DBNs) and decoupling of the rigid and nonrigid classifiers to reduce the need of large training sets in the segmentation of the left-ventricle in 4-chamber apical views.

Finally, sparse coding and dictionary learning was recently applied to tracking of endocardial and epicardial contours in 4D apical long-axis data, in a level-set framework, by Huang et al. (2014).

We propose a novel Bayesian formulation, with a shape model which takes into account both endocardium, atrial region and apex, in combination with a statistical signal model, which attempts to closely model the censored signal, not through an approximation by e.g. Gaussian distribution as is the case in Lin et al. (2003); Boukerroui et al. (2003) or by using no distributional assumptions such as Dietenbeck et al. (2012). In this regard our model resembles Bosch et al. (2002), as described above. The use of a censored Gamma mixture model with unknown censoring point, to handle artefacts resulting from left-censoring of the data in US clinical B-mode is to our knowledge novel. In Zhu et al. (2007) a similar model is used to model RF
data (not subject to censoring and/or quantization) using a shifted Rayleigh distribution. The shift parameter is unique for each region in Zhu et al. (2007), while our censoring point is global, since data points are censored at the same intensity level regardless of where they are spatially located.

Then samples from the posterior distribution are acquired through Gibbs and the novel approach of drawing from translation distribution and the samples then form a probability map, which is the conditional mean of the samples. Using the mean of samples to produce a segmentation measure is also employed by Carneiro et al. (2012), in which an expected contour onto a PCA shape space. Instead of giving a single segmenting curve, the certainty of which may vary along the curve, our method provides a more versatile measure. Naturally there are other models which use Bayesian formulations, but there are no models which take our specific approach leading to a probability map. Of the aforementioned works, dealing with the apical view, only Boukerroui et al. (2003) employs a Bayesian framework, although we note that Carneiro et al. (2012) uses MAP estimation of parameters. There are works not dealing with the apical view in cardiac US segmentation, that employ Bayesian formulations such as Martin-Fernandez and Alberola-Lopez (2002) (kidney), Destrempes et al. (2009) (carotid arteries), Hansen et al. (2002) (intracoronary ultrasound).

The main steps of the proposed method are: 1) the hyperparameters of the posterior are estimated, through fitting of a censored Gamma mixture model onto image intensity data and cross-validation. 2) the posterior of the latent variables representing the encocardium is sampled; the samples are used to compute the conditional mean over the marginal posterior, which we
Figure 1: Two chamber apical long-axis view of the human heart. Apex region is indicated by dashed circle. Original image "Apical two chamber view" by CC Patrick J. Lynch and C. Carl Jaffe, Yale University, 2006 is used under a Creative Commons license.

The paper is organised as follows. In Sec. 2, the proposed Bayesian model is described, the likelihood formulation is described in 2.3, and the prior formulation in Sec. 2.5. The algorithm is described in Sec. 3. Experiments and comparison w.r.t. an interobserver study and the model of Chen et al. is presented in Sec. 4. In Sec. 5, advantages and limitations of the method are discussed. In Sec. 6 the proposed method is summarized and concluding remarks are made.

1.1. Differences to our previously published work

In this paper, we present a Bayesian model for determining the position distribution of the endocardium in the left-ventricular long-axis apical views in clinical B-mode US cycles. Earlier versions of this model are presented
Figure 2: The symbol \( u \) represents the \( T \) latent variable images \( I_u \) stacked into a single vector. Each \( I_u \) (top row) corresponds to an image in the US sequence (bottom row) of length \( T \), where \( z \) denotes the spatial coordinates. Three regions are defined: one signifies the endocardium, two atrial region and three background. The notation \( I_u \) is used in the latent variable prior, defined in Sec. 2.5.

In Hansson et al. (2009a) and Hansson et al. (2009b). There are major differences between the current model and the ones presented in Hansson et al. (2009a,b): Our signal model has changed significantly in this work: it now takes into account left-censoring of data instead of using a standard mixture model formulation.

The prior component formulation has been significantly simplified. The prior now only consists of a smoothing component and a shape prior, instead of the previous 5 component prior formulation. The atrium is explicitly mod-
Figure 3: Regions used in likelihood formulation: $R_{12}$ signifies endocardium and atrial region, while $R_3$ denotes the background.

eled in the current model, by assigning a latent variable class and shape prior for the atrial region. The atrium was previously modelled by a prior which decreased the likelihood of sampling the endocardial class, below a point in the image which was estimated using training data. The connectivity prior has been eliminated since in the new formulation the contour is discretized, eliminating the possibility of holes in the region. The Bernoulli prior in the previous model, which is a prior on the volume of the endocardium, has been transformed into a prior on proportions in the mixture model.

Previous versions did not include an apex sampling weight, as thus the Metropolis Hastings algorithm did not need to be employed, as it was possible to sample directly from the translation distribution (see Hansson et al., 2009b, Eq. 17, pg.1080). We also explain that we sample from a manifold $\mathcal{M}$ of simply connected non-intersecting regions; this was previously also done but not given explicit description.

Further additions in our latest work are: priors for mixture parameters,
selection of distribution model by use of the AMDL criterion, estimation of prior parameters by cross-validation, and comparative validation of the model and an inter variability study.

2. Model

We formulate a Bayesian model for determining the position of the endocardium in a clinical B-mode US cycle. In order to make this process more robust we incorporate prior knowledge of the position of the atrium into our model. Due to signal dropout the true anatomical form of the atrium is rarely visible and instead we model the low-intensity region connected to atrium, in the sequel referred to as the atrial region.

2.1. Summary of Model

We form the marginal posterior distribution of the latent variables representing the endocardium. The marginal posterior of the endocardium is computed by marginalising over the remaining latent variables representing atrial region and background. The posterior contains the following hyperparameters. In the likelihood they concern the mixture model of image intensities and apex region, and in the prior, shape prior and smoothness. Due to analytic intractability, in our formulation the hyperparameters are not marginalised out but estimated, and the estimates are substituted into the model. The estimation of the hyper parameters is described in Sec. 3. The conditional mean of the endocardial latent variables $u_{endo}$ summarises the marginal posterior and represents the Bayesian Probability Map, describing the position of the endocardium.
2.2. Posterior

Assume that there are $N$ pixels in a sequence of US images. For each pixel $i = 1, \ldots, N$, the latent variables $u_i$ and $w_i$ are defined. The latent variable $u_i$ indicates to which region pixel $i$ belongs, $u_i = 1$ indicates endocardium, $u_i = 2$ atrial region and $u_i = 3$ the background, see Fig. 2.

The motivation behind having two sets of latent variables $u_i$ and $w_i$ is that they correspond to different parts of the model. The data model considers two classes $R_{12}$ and $R_3$ for the latent variable $w_i$, corresponding to the endocardium and atrial region, and background. This is due to the fact that the endocardium and atrium have a similar intensity distribution. The variables $w_i$ thus correspond to the data model. The full model however uses three classes for the latent variables $u_i$: endocardium, atrium and background. These three are distinguished from each other through sampling on a manifold $\mathcal{M}$ of simply connected non-intersecting regions, as described in Sec. 3.4.

The latent variables are stacked into the vector $u$. The vector $u_{\text{endo}}$ of equal size to $u$ is defined by $u_{\text{endo}} = (\delta(u_1 - 1), \delta(u_2 - 1), \ldots, \delta(u_N - 1))$, where $\delta$ is the Kronecker delta. The elements of $u_{\text{endo}}$ are set to zero, if the associated latent variables do not belong to the endocardium.

In the same manner, $u_{\text{atr}}$ and $u_{\text{bkgr}}$, are defined by $u_{\text{atr}} = (\delta(u_1 - 2), \delta(u_2 - 2), \ldots, \delta(u_N - 2))$ and $u_{\text{bkgr}} = (\delta(u_1 - 3), \delta(u_2 - 3), \ldots, \delta(u_N - 3))$, respectively. Let $U$ and $U_{\text{endo}}$ be random vectors, corresponding to $u$ and $u_{\text{endo}}$, respectively. The latent variables $w_i$ are defined in the same way as $u_i$.

The problem of determining the position of the endocardium can thus be formulated as the determination of the marginal posterior distribution of
the latent variables, where the nuisance parameters $u_{\text{atr}}$ and $u_{\text{bkgr}}$ have been integrated out, or

$$P(U_{\text{endo}} = u_{\text{endo}}|x, \Delta) = \sum_{u_{\text{atr}},u_{\text{bkgr}}} P(U = u|x, \Delta)$$

$$\propto \sum_{u_{\text{atr}},u_{\text{bkgr}}} L(x|U = u, \gamma, \sigma_{\text{apex}}) P(U = u|\Lambda),$$

where $L$ is a likelihood function, $x = \{x_1, ..., x_N\}$ represent gray level intensities, corresponding to the pixels $i = 1, ..., N$, stacked into a single vector, and $\Delta = \{\Lambda, \sigma_{\text{apex}}\}$ are hyperparameters.

2.3. Likelihood

The likelihood is defined as

$$L(x|u, \gamma, \sigma_{\text{apex}}) = \prod_i \left[\gamma_i 1(u_i \in R_{12}) \times (1 - \gamma_i) 1(u_i \in R_3)\right] A(z_i|\sigma_{\text{apex}}),$$

where $u_i$ are the realisations of latent variables $U_i$ corresponding to $x_i$, $\gamma_i$ and $\sigma_{\text{apex}}$ are the likelihood hyperparameters, $1$ is an indicator variable, and $z_i$ is the spatial coordinate of pixel $i$. The endocardium and atrial region, and background is denoted by $R_{12}$ and $R_3$, respectively (see Fig. 3).

The hyperparameters $\gamma$ are modelled as the posterior class probabilities of the intensity mixture model with the help of the latent variables $w_i$, or,

$$\gamma_i = P(W_i \in R_{12}|x_i, \Theta, \pi) = \frac{\pi_{12} p(x_i|\Theta_{12})}{\pi_{12} p(x_i|\Theta_{12}) + \pi_3 p(x_i|\Theta_3)}$$

where the parameters $\Theta = \{\Theta_{12}, \Theta_3\}$ of the likelihood describe the mixture distributions $p(x|\Theta_{12})$ and $p(x|\Theta_3)$, and $\pi = \{\pi_{12}, \pi_3\}$, s.t. $\pi_{12} + \pi_3 = 1$ are
prior weights on the relative size of regions, obtained from training data, c.f. Sec. 3.1. It follows that

\[ 1 - \gamma_i = 1 - P(W_i \in R_{12}|x_i, \Theta, \pi) = P(W_i \in R_3|x_i, \Theta_3, \pi) . \tag{4} \]

The apex sampling weight \( A(z_i|\sigma_{\text{apex}}) \) is introduced to deal with the fact that the endocardium is problematic to model in the two-chamber apical long axis view. When the apex of the heart is highly curved, this may appear as thickening of the cardiac wall due to foreshortening (Deserno, 2011). This is a new addition to the models introduced in Hansson et al. (2009a) and
The apex sampling weight $A(z_i | \sigma_{\text{apex}})$ is defined as

$$A(z_i | \sigma_{\text{apex}}) = 1 - \frac{G_{\sigma_{\text{apex}}}(z_i - z_{\text{apex}})}{\max(G_{\sigma_{\text{apex}}}(z_i - z_{\text{apex}}))} ,$$

(5)

where $G_{\sigma_{\text{apex}}}$ is the 2-dimensional Gaussian density, with mean $z_{\text{apex}}$ (the apex coordinate of the endocardium) and covariance matrix $\sigma_{\text{apex}} \cdot I_{2 \times 2}$. The influence of the down-weighting term (5) decays according to Gaussian distribution with parameter $\sigma_{\text{apex}}$. By adjusting the parameter $\sigma_{\text{apex}}$ the amount of down-weighted intensities is controlled; this parameter is set through cross-validation. A Gaussian down-weighting was chosen since a soft symmetric down-weighting was desired, since no obvious asymmetric pattern of mis-
classified intensities in the apex was found. The formulation is clearly an approximation.

In the formulation of the likelihood, we need to take into account the left-censoring of the data, seen in Fig. 4, which is the result of pooling all the data equal and below an (unknown) amplitude \( a > 0 \) and assigning these the value \( b_1 \). Since we do not have access to the internal machine architecture and software, this amplitude \( a \) is unknown to the user. Typically \( b_1 \) will be e.g. 0, 1 or 6. This will affect the visual appearance only slightly, since the affected amplitudes are often in the low range, but this type of censoring must be taken into account when formulating a statistical model for the data. Thus we propose a censored mixture model with unknown censoring point for US data. Censored data models with unknown censoring limits commonly occur in survival analysis (see e.g., Klein and Moeschberger, 2003).

The censored mixture model \( p(x|\Theta, a) \) is defined as follows. Depending on how many bits are used in quantization and on the degree of nonlinearity, the number \( b_1 \) will change. For a very high bit count \( b_1 = 0 \), since the quantization error is then \( < 0.5 \). As the bit count decreases, the number \( b_1 \) will increase; this of course disregards any nonlinear binning which may affect the result. The number \( a + b_1 \) is thus the range over which the signal is censored.

The distribution of the censored amplitudes \( x \) is expressed as a censored mixture model of \( q \) Gamma densities,

\[
p(x|\Theta, a) = P_1(\Theta, a)\delta_0(x-b_1) + g(x|\Theta, a) ,
\]

(6)
where

\[ P_1(\Theta, a) = \sum_{j=1}^{q} \alpha_j \int_{0}^{a+b_1} f_{\text{gam}}(s|k_j, \theta_j) ds \]  

(7)

and

\[ g(x|\Theta, a) = \begin{cases} 
0, & \text{if } x \leq b_1 \\
\sum_{j=1}^{q} \alpha_j f_{\text{gam}}(x+a|k_j, \theta_j), & \text{if } x > b_1 
\end{cases} \]  

(8)

In (6) \( \delta_0 \) denotes the Dirac delta function, not to be confused with Kronecker delta used in Sec. 2.2.

The shape parameter of the gamma distribution is restricted s.t. \( k_j > 1 \). This is to avoid components which solely model the first bin (akin to a Dirac ”spike”), as we regard distributions with such components as non-informative.

Here the Gamma density function is denoted by \( f_{\text{gam}} \), and parameters \( \Theta = \{\Theta_{12}, \Theta_3\} \). The elements of \( \Theta \) correspond to the endocardium and atrial region, and the background. Thus \( \Theta_{12} = \{\alpha_j, k_j, \theta_j; j = 1, \ldots, q_{12}\} \) are the parameters of a \( q_{12} \) component censored mixture distribution for the endocardium and atrial region, and \( \Theta_3 = \{\alpha_j, k_j, \theta_j; j = q_{12} + 1, \ldots, q\} \) are the parameters for the \( q_3 = q - q_{12} \) components mixture for the background. The mixture parameters \( \alpha_j > 0 \) fulfill \( \sum_j \alpha_j = 1 \).

The mode of component \( j \) is denoted by \( m_j \). We assume that the sum of modes of components associated with region \( R_{12} \) is less than those associated with region \( R_3 \), i.e. \( \sum_{j=1}^{q_{12}} m_j < \sum_{j=q_{12}+1}^{q} m_j \).

2.4. Mixture Parameter Priors

The mixture parameter prior is defined as

\[ P(\Theta) \equiv P(k_1, \theta_1, \ldots, k_q, \theta_q) \]  

(9)
To achieve the best possible separation between components we impose priors on the shape parameters and scale parameters corresponding to the two regions $R_{12}$ and $R_3$. For the shape and scale parameters of the Gamma densities a uniform prior is defined,

$$P(k_1, \theta_1, \ldots, k_q, \theta_q) \propto \begin{cases} 1, & \text{if } k_j, \theta_j \in K_{12}; j = 1, \ldots, q_{12} \wedge k_j, \theta_j \in K_3; j = q_{12} + 1, \ldots, q \\ 0, & \text{otherwise} \end{cases},$$

(10)

where $K_{12}$ and $K_3$ are each a simplex, corresponding to region $R_{12}$ and $R_3$ constructed from shape and scale parameters estimates obtained from training data.

2.5. Latent Variable Priors

The priors on the latent variables are defined as

$$P(u|\Lambda) = P_{sd}(u|\lambda_{sd})P_{\text{shape|sd}}(u|\lambda_{\text{shape}}, \lambda_{\text{mean}}),$$

(11)

where $\Lambda = \{\lambda_{\text{shape}}, \lambda_{\text{mean}}, \lambda_{sd}\}$ are weighting parameters controlling the influence of the priors. The two components of the prior characterize different kinds of properties preferred. The spatial derivative $P_{sd}$ prior enforces spatial and temporal smoothness for latent variable images. Possible shape variations around the mean shape are described by $P_{\text{shape|sd}}$ constructed from eigenshapes of manually segmented images through $P_{\text{shape|sd}}$.

The sequence of images is divided into subsequences, to take the temporal variations of the endocardium into account, and so for each part of the sequence a corresponding set of eigenshapes and mean is used.
The prior component formulation has been significantly simplified compared to those contained in Hansson et al. (2009a,b). Specifically the atrium is explicitly modeled in the current model, by assigning a latent variable class and shape prior for the atrial region. The atrium was modelled in Hansson et al. (2009b) by a prior which decreased the likelihood of sampling the endocardial class, below a point in the image which was estimated using training data. The prior in the current submission is simpler in that it only consists of a smoothing component and a shape prior, instead of the previous 5 component prior formulation. The connectivity prior in Hansson et al. (2009a,b) is eliminated since in the new formulation the contour is discretized, eliminating the possibility of holes in the region. The Bernoulli prior in Hansson et al. (2009a,b), which is a prior on the volume of the endocardium, has been transformed into a prior on proportions in the mixture model.

Let $I_{1(u \in R_{12})}(z,t)$ be the 3-dimensional array of latent variables corresponding to an image in the US cycle, where endocardial and atrial region latent variables are set to one. In Fig. 2, the array $I_u(z,t)$ is illustrated. Here $z$ and $t$ are spatial and temporal coordinates, respectively (see Fig. 2). The spatial derivative prior is then given by

$$P_{sd}(u | \lambda_{sd}) \propto \exp\{-\lambda_{sd}||I_{1(u \in R_{12})} * f||_{L_1}\}, \quad (12)$$

where $f$ is a three dimensional Laplacian kernel and $*$ denotes convolution. By increasing $\lambda_{sd}$ our sampling is regularised towards smoothness.

Inspired by the Bicycle Chain model in Sommer et al. (2009), we build our shape prior as follows. For each image in the US cycle, there is a corresponding image containing the latent variables describing the endocardium, atrial region and background. The contours of the endocardium and atrial
regions are discretized in an equidistant manner. Let $z^e_h$ be the $h$:th spatial coordinate in the $xy$-plane, taken from an equidistant set of $n$ points from the boundary of the endocardium. For the atrial region the $k$:th spatial coordinate in a equidistant set of $m$ points is denoted by $z^a_k$. See Fig. 5 for illustration. Then $v = (z^e_1, ..., z^e_n, z^a_1, ..., z^a_m)$ jointly provides the location of the endocardium and atrial region. Now let $t_z$ and $\phi$ be translation and rotation parameters, respectively, that take $v$ as close as possible to the mean shape $\bar{v}$ of the training data, such that $||v - \bar{v}||_2$ is minimized, where $|| \cdot ||_2$ denotes the $L^2$-norm. This can be easily be extended to include scaling, however in our application this was not found necessary. The parameters are collected in the vector
\begin{equation}
\eta = (z^e_1, ..., z^e_n, z^a_1, ..., z^a_m, t_z, \phi) \in \mathbb{R}^{mn+3}.
\end{equation}

Now, let the shape prior be defined as
\begin{equation}
\begin{align*}
P_{\text{shape}|sd}(u|\lambda_{\text{shape}}, \lambda_{\text{mean}}) \propto \\
\prod_t \exp \left\{ -\lambda_{\text{shape}}(\eta_t - \bar{\eta}_t)^T((1 - \lambda_{\text{mean}})C^+_t + \lambda_{\text{mean}}I)(\eta_t - \bar{\eta}_t) \right\},
\end{align*}
\end{equation}
where $\eta_t$ and $C_t$ represent the joint shape of the endocardium and atrial region, $\bar{\eta}_t$ is the mean of all vectors $\eta_t$, and $C^+_t$ is the Moore-Penrose pseudoinverse (Penrose, 1955) of the covariance matrix of the training shapes at the time $t$, respectively.

Increasing $\lambda_{\text{shape}}$ makes the influence of the shape prior larger, while larger $\lambda_{\text{mean}}$ increases the influence of the mean shape in the shape prior.
Figure 6: Summary of the proposed algorithm. In the first step, distributional parameters are estimated from the cardiac cycle $x$. Then in the second step, the posterior distribution $P(u_{endo}|x; \Delta)$ is sampled and summarized by conditional mean $\hat{u}_{CM}^{endo}$.

3. Algorithm

Our algorithm for generating Bayesian Probability Maps contains two steps: the hyperparameters of the posterior are first estimated, after which the posterior is sampled and summarised, as illustrated in Fig. 6. In the first step, we fit the censored Gamma mixture model onto image intensity data, to obtain an estimate for the likelihood hyperparameters $\gamma$. The likelihood hyperparameter $\sigma_{\text{apex}}$ and the regularisation parameters $\lambda_{\text{sd}}, \lambda_{\text{shape}}$ and $\lambda_{\text{mean}}$ are estimated using cross-validation. In the second step, the posterior of the latent variables $u_i$ is sampled by Gibbs sampling (Geman and Geman, 1984); the samples are used to compute the conditional mean over the marginal posterior, which we refer to as a Bayesian probability map.
3.1. Estimation of Mixture Parameters

Assume that the log-compressed intensities $x = \{x_1, ..., x_N\}$ are independently censored at an unknown point $a$, and subsequently quantized (i.e. binned). Let $b_1$ denote the value of the first bin.

The parameters $\Theta$ of the mixture model (6) are obtained as the solution to the MAP (Maximum a Posteriori) problem

$$\hat{\Theta} = \arg \max_{\Theta,a} p(x|\Theta,a)P(\Theta).$$

(15)

We solve the MAP problem (15) by use of Expectation Maximization (EM) (Dempster et al., 1977). Thus the complete data likelihood of the model (6) is represented according to the latent variable model as

$$p(x, w^*|\Theta, a) = \prod_{j=1}^{q} \prod_{i=1}^{n_j} \left( \alpha_j \int_0^{a+b_1} f_{\text{gam}}(s|k_j, \theta_j) ds \right)^{\delta(W_i^* - j)} \times$$

$$\prod_{i=n_1+1}^{n} \left( \alpha_j f_{\text{gam}}(x_i + a|k_j, \theta_j) \right)^{\delta(W_i^* - j)} \times P(k_1, \theta_1, ..., k_q, \theta_q)$$

(16)

where $x$ is censored B-mode data, and $w^* = (w_1^*, ..., w_N^*)$ are latent variables which provide class membership assuming data independence, c.f. Sec. 2.2. Note that the latent variables $w^*$ are not related to the latent variables $w$ in Sec. 2.3.

On the E-step, we build the expected complete data log-likelihood, conditioned on the measured data and the previous parameter estimates. Specifically, index the censored intensities $x$ s.t.

$$i = \begin{cases} 1, ..., n_1 & \text{if } x_i = b_1, \\ n_1 + 1, ..., N & \text{if } x_i > b_1, \end{cases}$$

(17)
where $b_1$ is the gray level intensity at the first bin.

The expected complete data log-likelihood is then

$$\chi(\Theta, \Theta^{(n)}) = E_{w^*|x, \hat{\Theta}^{(n-1)}} \{ \log p(x, w^*|\Theta, a) \}$$

$$= \sum_{j=1}^{q} \sum_{i=1}^{n_1} \log \left( \alpha_j \int_0^a \frac{f_{\text{gam}}(s|k_j, \theta_j)ds}{P(W_i^* = j|\Theta^{(n)}, x_i = b_1)} \right) P(W_i^* = j|\Theta^{(n)}, x_i = b_1)$$

$$+ \sum_{j=1}^{q} \sum_{i=n_1+1}^{n} \log \left( \alpha_j f_{\text{gam}}(x_i + a|k_j, \theta_j) \right) P(W_i^* = j|\Theta^{(n)}, x_j > b_1)$$

$$+ \log P(k_1, \theta_1, ..., k_q, \theta_q),$$

(18)

where $P(W_i^* = j|\hat{\Theta})$ is defined analogously to Eq. (3). On the M-step, $\chi(\Theta, \Theta^{(n)})$ is maximized for $\Theta$ to obtain an update for the parameters. The steps are iterated until convergence.

Initial values for the EM algorithm is obtained by use of the $k$-means algorithm (MacQueen, 1967) using squared Euclidean distances. The data is divided into $q$ groups using k-means, and the Maximum Likelihood Estimate (MLE) is computed for each group. The EM algorithm is run with 50 initializations, and the value of $\Theta$ which produces the highest posterior value is adopted.

The weights $\pi_{12}$ and $\pi_3$ in (3) are estimated by the proportion of pixels inside and outside the region $R_{12}$ in the training data, respectively.

The parameter $\sigma_{\text{apex}}$ of the apex sampling weight $A(z_i|\sigma_{\text{apex}})$ is estimated using leave-one-out cross-validation, see Sec. 3.3.

3.2. Sampling of the Posterior

The posterior is sampled by alternating between conventional Gibbs sampling (Geman and Geman, 1984; MacKay, 2003) and sampling of latent vari-
able image translations. On the Gibbs sampling step, we draw the elements of the sample latent variable vector $u$ from the conditional distribution

$$P(u_i | u_1^{(r)}, \ldots, u_{i-1}^{(r)}, u_{i+1}^{(r-1)}, \ldots, u_N^{(r-1)}, x, \Delta)$$

$$= \left\{ P(u_i = k | u_1^{(r)}, \ldots, u_{i-1}^{(r)}, u_{i+1}^{(r-1)}, \ldots, u_N^{(r-1)}, x, \Delta) \right\}_{k=1}^3, \quad i = 1, 2, \ldots, N,$$

(19)

where $P$ is the posterior probability from Eq. 1. Then, to obtain sample vector $u^{(r+1)}$, we sample the distribution of translations which spatially move the latent variable image $I_u$. The details of the translation sampling step are as follows. We sample the conditional translation distribution

$$P(t | u, x, \Delta) \equiv P(u' | u, x, \Delta) \propto L(x | u') P(u' | \Lambda),$$

(20)

where the latent variable vector $u'$ is obtained from $u$ by spatially translating the latent variable image $I_u$ by $t$.

We sample (20) by the Metropolis Hastings algorithm (Metropolis et al., 1953) with proposal density

$$P_{\text{prop}}(t | u, x, \Delta) \propto$$

$$\left( \prod_{i=1}^N \gamma_i^1(u'_i \in \mathbb{R}_{12})(1 - \gamma_i)1(u'_i \in \mathbb{R}_3) \right) P(u' | \Lambda) \propto \prod_{i=1}^N \gamma_i^1(u'_i \in \mathbb{R}_{12})(1 - \gamma_i)1(u'_i \in \mathbb{R}_3),$$

(21)

where we have used the fact that the prior is invariant under translation and can be disregarded. The proposal distribution is an approximation of the conditional translation distribution (20), in that the apex sampling weight $A(z_i | \sigma_{\text{apex}})$ has been dropped in (21). Previous versions did not include an apex sampling weight, as thus the Metropolis Hastings algorithm did not need
to be employed, as it was possible to sample directly from the translation
distribution (see Hansson et al., 2009b, Eq. 17, pg.1080).

It follows that

$$\log P_{prop}(t|u, x, \Delta) = \sum_{i=1}^{N} 1(u'_i \in R_{12}) \log \gamma_i + 1(u'_i \in R_3) \log(1 - \gamma_i) + C , (22)$$

where $C$ is a constant, related to the latent variable priors $P(u'|\Lambda)$, which
does not depend on the translation. The sum in (22) represent correlations between the translated latent variable image and log probability densities; this is clear since $u'_i$ assumes all possible translations and thus corresponds to the moving image in (Gonzalez and Woods, 2002, chap. 4, Eq. 4.6-27). Hence, the logarithms of the conditional translation probabilities can be computed efficiently in the Fourier domain by the correlation theorem. By the Metropolis-Hastings rule for independent proposals, we draw the translation sample $t'$ from $P_{prop}(t|u^{(r)}, x, \Delta)$, the sample is accepted if

$$b = \frac{P(t'|u^{(r)})}{P(t=0|u^{(r)})} \cdot \frac{P_{prop}(t=0)}{P_{prop}(t')} \geq 1,$$

otherwise accept with probability $b$. If the translation is accepted, we set $u^{(r+1)} = T'_t(u^{(r)})$, where $T'_t$ denotes the translation operator.

3.3. Estimation of Hyperparameters

In this section we present priors for distribution parameters, estimation of prior parameters by cross-validation, and mixture model selection by use of the AMDL criterion.

For each cycle manually segmented in its entirety, the data is divided into
two sets corresponding to endocardium and atrial region, and background.
ML estimates of distribution parameters $\Psi_{12}$ and $\Psi_3$ are computed for each
region $R_{12}$ and $R_{3}$ by maximizing the likelihood

$$L_{\text{train}}(x) = L_{\text{train}}^{12}(x; c, c) \cdot L_{\text{train}}^{3}(x; d, d)$$  \hspace{1cm} (23)$$

where

$$L^{(c)}_{\text{train}}(x; e_1, c) = \prod_{i=1}^{c} \left( \int_{0}^{a} f(s|\Psi^{(c)}) ds \right) \cdot \prod_{i=e_1+1}^{c} (f(x + a|\Psi^{(c)})) ,$$  \hspace{1cm} (24)$$

and number of observations are $c + d$, where $c$ observations are taken from the endocardium and atrial region, and $d$ observations from the background. The number of censored observations in each region is given by $c_1$ and $d_1$. Note that without the censoring limit $a$, the two likelihood terms in (23) would be maximized independently. The parameters $\hat{\Psi}_{12}$ and $\hat{\Psi}_{3}$ are not to be confused with the distribution parameter $\Theta$, c.f. Sec. 2.3. The parameters $\hat{\Psi}_{12}$ and $\hat{\Psi}_{3}$ are used to construct a prior for $\Theta$.

The hyperparameters $K_{12}$ and $K_{3}$ of the uniform distribution prior (10) are defined by

$$K_{12} = [\min_{l}(\hat{k}_{12}^{l}), \max_{l}(\hat{k}_{12}^{l})] \times [\min_{l}(\hat{\theta}_{12}^{l}), \max_{l}(\hat{\theta}_{12}^{l})]$$  \hspace{1cm} (25)$$

and

$$K_{3} = [\min_{l}(\hat{k}_{3}^{l}), \max_{l}(\hat{k}_{3}^{l})] \times [\min_{l}(\hat{\theta}_{3}^{l}), \max_{l}(\hat{\theta}_{3}^{l})],$$  \hspace{1cm} (26)$$

where $\hat{k}_{12}^{l}$ and $\hat{\theta}_{12}^{l}$ are the ML estimates of region $R_{12}$ obtained by maximizing the joint likelihood (23) for training sequence $l$, and $\hat{k}_{3}^{l}$ and $\hat{\theta}_{3}^{l}$ are the corresponding ML estimates for region $R_{3}$.

The prior parameters $\Lambda = \{\lambda_{\text{shape}}, \lambda_{\text{mean}}, \lambda_{sd}\}$ are estimated by leave-one-out cross-validation over the training data set. Leave-out-one cross-validation is performed as follows. Each training sequence of images has a corresponding groundtruth segmentation sequence. In each round of cross-validation the
Table 1
Evaluated distribution models

<table>
<thead>
<tr>
<th></th>
<th>Gamma</th>
<th>Nakagami</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_{\text{gam}}(x</td>
<td>k, \theta) = x^{k-1}e^{-x/\theta}/\theta^k \Gamma(k)$,</td>
</tr>
<tr>
<td></td>
<td>$x \geq 0$ and $k, \theta &gt; 0$</td>
<td>$x \geq 0$ and $\mu, \omega &gt; 0$.</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$f_{\text{logn}}(x</td>
<td>\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi\sigma}}e^{-(\log x - \mu)^2/2\sigma^2}$,</td>
</tr>
<tr>
<td></td>
<td>$x \geq 0$ and $\sigma &gt; 0$</td>
<td>$f_{\text{rayl}}(x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x &gt; 0$ and $\mu \in \mathbb{R}, \sigma &gt; 0$.</td>
</tr>
<tr>
<td>Fisher-Tippett</td>
<td>$f_{\text{fish-tip}}(x</td>
<td>\mu, \sigma) = \exp{(x - \mu)/\sigma - \exp{(x - \mu)/\sigma}}, x \geq 0$ and $\mu \in \mathbb{R}, \sigma &gt; 0$</td>
</tr>
</tbody>
</table>

covariance matrix of shapes is computed, c.f. (14), from groundtruth segmentations, leaving out one training sequence. Using the estimated covariance matrix the algorithm is applied to the left out training sequence, and the optimal parameters are determined w.r.t. Average Distance (AD) from the algorithm segmentation to the groundtruth segmentation. This procedure is repeated for all samples, and the mean is taken as the parameter estimate.
Mean AMDL score for distribution models

<table>
<thead>
<tr>
<th>Dist. model</th>
<th>Gamma</th>
<th>Nakagami</th>
<th>Rayleigh</th>
<th>Fisher-Tippett</th>
<th>Log-Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td># comp.</td>
<td>$q_c = 2$</td>
<td>$q_c = 2$</td>
<td>$q_c = 2$</td>
<td>$q_c = 3$</td>
<td>$q_c = 2$</td>
</tr>
<tr>
<td></td>
<td>$q_d = 2$</td>
<td>$q_d = 2$</td>
<td>$q_d = 2$</td>
<td>$q_d = 1$</td>
<td>$q_d = 2$</td>
</tr>
<tr>
<td>Mean AMDL</td>
<td>$5.10 \cdot 10^6$</td>
<td>$5.20 \cdot 10^6$</td>
<td>$5.38 \cdot 10^6$</td>
<td>$5.35 \cdot 10^6$</td>
<td>$5.38 \cdot 10^6$</td>
</tr>
</tbody>
</table>

Estimation of Number of Mixture Components and Evaluation of Distribution Models

Several distribution models are considered for use in a mixture density, modeling endocardium and atrial region, and background gray level intensities in clinical B-mode US cardiac data, which have been reported as appropriate for modeling cardiac US gray levels in previous works (Nillesen et al., 2008; Tao et al., 2006). The ones considered here are the Gamma, Nakagami, Rayleigh, Fisher-Tippett and Log-Normal distribution, given in Table 1. The relationship between Nakagami and Gamma distribution is given by the connection

$$Y \sim \text{Gamma}(\mu_{\text{gam}}, \omega_{\text{gam}}) \text{ and } X \sim \text{Nakagami}(\mu_{\text{nak}}, \omega_{\text{nak}})$$

then

$$X^2 = Y(\mu_{\text{nak}}, \omega_{\text{nak}}/\mu_{\text{nak}}). \tag{27}$$

The Rayleigh distribution is similar in shape to Gamma and Nakagami, and was first proposed in Burckhardt (1978) for modeling US data. The log-normal distribution has been used for liver (Zimmer et al., 2000) and breast US (Xiao et al., 2002). In Dutt (1995) the Fisher-Tippet, or double-
exponential distribution, is shown to be most suitable for log-compressed
envelope detected RF data with fully developed speckle.

We use a maximum of four densities to model intensities in the US cycles,
thus 6 component density configurations for each distribution: \((R_{12}, R_3) =
(1,1), (1,2), (2,1), (2,2), (1,3)\) and \((3,1)\). In order to avoid densities which
only model the first bin in the histogram, as mentioned for the case of the
Gamma distribution in Sec. 2.3, the following constraints are imposed: \(\mu > 1\)
for the Nakagami distribution and \(\sigma < 0.5\) for the Log-Normal distribution.

We compute the MAP solution, c.f. (15), for each choice of component
configuration and distribution. Uniform priors for the parameters of each dis-
tribution is constructed in an analogous manner to the case of the Gamma
distribution, c.f. Sec. 3.3. We select the configuration of components and
distribution which produces the least mean Asymptotic MDL (AMDL) score,
see (e.g., Barron et al., 1998), taken over all training sequences. The Asymp-
totic Minimum Description Length (AMDL / BIC) principle selects the finite
mixture model with \(d\) free parameters, which minimizes the quantity

\[
-2 \log \mathcal{L}(\hat{\omega}) + d \log r ,
\]

where \(\mathcal{L}\) is the likelihood of data given model parameters, \(\hat{\omega}\) the MLE of the
model, and \(r\) is the number of observations.

3.4. Summarising the Posterior

To the \(r\):th sample latent variable vector \(u^{(r)}\) corresponds a sequence
\(I_{u}^{(s)}\) of latent variables images, where \(s = 1, ..., T\). In each latent variable
image \(I_{u}^{(s)}\), the regions corresponding to the endocardium and atrial region are
represented by two curves $C^{(s)}_{\text{endo}}$ and $C^{(s)}_{\text{atrium}}$. Now let $\mathcal{M}$ be the manifold of non-intersecting simple curves. Only samples $u^{(r)}$, whose corresponding curves lie on $\mathcal{M}$ are drawn. Thus only samples are drawn whose corresponding latent variable images contain two distinct contiguous regions, endocardium and atrium.

To characterize the posterior distribution of $P(u_{\text{endo}}|x; \Delta)$, we compute an estimate of the conditional mean over the posterior

$$E\{u_{\text{endo}}|x, \Delta\} \approx \frac{1}{M} \sum_r u^{(r)}_{\text{endo}} = (\hat{P}(U_i \in \text{endocardium}|x, \Delta))_{i=1}^N \equiv \hat{u}^{\text{CM}}_{\text{endo}}$$

of the endocardial latent variables $u_{\text{endo}}$, whose corresponding curve $C_{\text{endo}} \in \mathcal{M}$.

By the strong law of large numbers $\hat{u}^{\text{CM}}_{\text{endo}} \to E\{u_{\text{endo}}|x, \Delta\}$ when $M \to \infty$. The corresponding image $I_{\hat{u}^{\text{CM}}_{\text{endo}}}$ represents the Bayesian probability map.

4. Experiments

4.1. Material

The US data used in this paper consists of 28 single beat cardiac cycles of two-chamber (2C) apical long-axis views of the heart. The cycles were obtained, at the setting 1.4, 2.4 or 3.2 MHz, using the echocardiogram machines Philips Sonos 7500, Philips iE33 or GE Vivid 7, from consecutive adult patients referred to the echocardiography laboratory at Malmö University hospital, Sweden which has a primary catchment area of 250,000 inhabitants. Expert outlines of the endocardium in the cycles have been provided by the same hospital. We used a clinical database which was created independently
Figure 7: Left: Ultrasound image. Center: Bayesian Probability Map. Color indicates probability from 0 (blue) to 1 (red). Right: Ultrasound image with overlaid Bayesian probability map and ground truth endocardium (dashed white) of our study. For our study we required an ultrasound plane containing a view with at least partially visible endocardial walls (we refer to this as an in-plane view). In the case of out-of-plane views, causing the anterior wall to
disappear from view as seen in Fig. 8, the probability map \( \hat{u}_{\text{CM} \text{endo}} \) (c.f. Eq. 29) will contain a relatively large area of low probability \( (0 < \hat{u}_{\text{CM} \text{endo}} \leq 0.8) \). With a missing anterior wall the position distribution of the endocardium is multimodal, and thus there is a risk of a faulty segmentation. Thus the inferior wall will in this case also not be segmented correctly.

An inter- and intravariability study was performed on 14 randomly selected in-plane cycles, of which 11 were included in the testing and training of the algorithm. In addition to the 11 in-plane cycles retained from the inter- and intravariability study, a further 17 in-plane cycles were collected. This brings the total of in-plane cycles used in this paper to 28, each acquired from individual patients.

Out of the 28 cycles, 20 were manually segmented in their entirety. Due to time constraints, six frames were segmented in the remaining 8 datasets: three frames at end-diastole and end-systole, respectively.

By selecting 16 equidistant frames in the cycle the length of the single beat cycles was normalized to length 16, which was found to be sufficient to capture the temporal variations for all sequences; the maximum length of an unnormalized sequence was 52.

Of the in-plane cycles (containing a view of the endocardium) used for the intervariability study, four cycles were discarded for testing and training: three cycles were discarded due to destructive interference and reverberations from pathological structures in lung and one due to a distinctive rib-shadow, where large part of the data (falsely) exhibits tissue-like appearance. These defects do not affect the manual segmentation, but are out-of-scope for our proposed method. When gathering the remaining test and training data,
Figure 8: Left: Ultrasound image. Center: Bayesian Probability Map. Color indicates probability from 0 (blue) to 1 (red). Right: Ultrasound image with overlaid Bayesian probability map and ground truth endocardium (dashed white). Here the proposed algorithm fails, due to a faulty ultrasound plane. The anterior wall is hardly visible in the ultrasound image, which causes the faulty positioning of the probability map. The large area of low probability ($0 < \hat{u}_{CM}^{endo} \leq 0.8$) is a clear indication that this is indeed a very problematic cycle of US images.

two cycles were excluded due to incorrect zoom setting, and one cycle due to significant rib-shadow.

The atrial region was segmented by a trained non-expert in the training set. Only fully segmented cycles can be used in the construction of the shape model, in order to capture the spatio-temporal variations of the endocardium and atrial region, and thus the 13 training sets were randomly selected among the 19 fully segmented sequences. The remaining 15 datasets were retained as test data.

We chose to use segmentations of the most senior expert is used for the shape model, since only the most senior expert segmented all of the 20 sequences, while the remaining experts only fully segmented a subset. The random assignment of cycles for training and testing from the entire dataset is very important for robustness of the model, especially in this case where
our dataset is limited. Thus for this reason, while it would be preferable, it is not possible to include the 11 in-plane inter- and intraobserver cycles in the test partition of our data. Four intervariability cycles were assigned as test sets through random selection.

4.2. Quantitative Validation

We evaluated the proposed method on our test data, 15 cardiac cycles acquired from individual patients as described in Sec. 4.1. For each cycle 6 manually segmented frames were obtained, 3 end-diastolic and 3 end-systolic. We employed validation measures of cardiac 2D+T US segmentation methods used in published quality studies (Chen et al., 2007; Boukerroui et al., 2003; Bosch et al., 2002; Mitchell et al., 2002) : Contour Average Distance (Contour AD), Area Correlation, Area error and Corresponding Point Average Distance (CPAD). The ground truth segmentation was taken from a senior expert observer.

CPAD is the mean unsigned Euclidean distance between corresponding landmarks on ground-truth (see Fig. 9) and the segmentation obtained by the proposed method. In Bosch et al. (2002) it is argued to be a more robust way of measuring automatic segmentation results than Contour AD, which measures minimal distance of points on ground-truth to algorithm segmentation. Contour AD may not register significant deviations along the border, since it only considers the distance to the closest point on the corresp. contour, and not in which order the points have been placed. In our case, two of the landmark points (1 and 37) were placed at the mitral valve and the third (19) at the apex of the endocardium, and the remaining were equidistantly placed between these three, see Fig. 9. In the case of contour
AD the first and last landmark was used to define the region of interest; the apex point is not relevant in this case since we are considering the least distance from the contour to any point on the heart walls.

In addition to this score we determined the median and median absolute deviation (MAD) of each of 37 discretization point of endocardial contour. This was done to highlight the uncertainty of the segmentation at different part of the endocardium. Median absolute deviation (MAD) for the dataset $Y = \{y_1, \ldots, y_n\}$ of $n$ observations is defined as

$$\text{MAD}(Y) = \text{median}(y_i - \text{median}(Y)).$$

(30)

Median and MAD are attractive measures, since they are robust measures of central tendency and dispersion, respectively. For area measures the endocardial area was defined as area enclosed by the endocardial border.

The median, median absolute deviation, mean and standard deviation is used illustrate the results of the proposed method in Fig. 10 and Table 4 were computed over the final sample obtained using the proposed method.

4.3. Manual Inter- and Intraobserver Variability of Ground Truth Data

Manual segmentations of the endocardium were performed by three independent expert clinicians, using the software package ITK-Snap (Yushkevich et al., 2006). The interobserver study was limited in scope, and so the numbers in Table 3 are necessarily approximate. Note that among other works concerning segmentation of LAX sequences, the number of experts used is either one (Bosch et al., 2002; Mitchell et al., 2002; Boukerroui et al., 2003) or two (Jacob et al., 2001).
As described in Sec. 4.1, 14 sequences in-plane cycles were drawn randomly from our hospital database. For the interobserver study all of the cycles were segmented entirely by all three experts. In the intra-observer study 3 end-diastolic and 3 end-systolic frames were segmented by each clin-
Figure 11: Expert contour (red), Contour of final sample using proposed method (black). Figures (a), (b), (c) and (d) are taken from two cycles where the proposed approach performs favorably. Figures (e), (f), (g) and (h) are taken from two cycles where the proposed approach has difficulties. Figures (a), (c), (e) and (g) depict Systolic frames, while figures (b), (d), (f) and (h) are Diastolic frames.
This was repeated two times at an interval of six months between each round of segmentation. The results of this study are presented in Table 3. In the calculations of deviations from groundtruth, the segmentation of the senior expert clinician was used as baseline.

In order to check that the test set and inter-intravariability set are similar, we compute the ejection fraction for both sets. Ejection fraction is a measure to access the function of the heart, e.g., low EF indicates stiffness of the heart walls. Ejection fraction is computed by \( \frac{(\text{End-Diastolic Volume} - \text{End-Systolic Volume})}{\text{End-Diastolic Volume}} \). The volume is approximated using Method of Discs (or Modified Simpson’s formula): \( \text{Volume} = 0.85 \times \frac{\text{Endocardial Area}}{\text{Length Apex to Base of Endocardium}} \). A Lilliefors test confirms that both group are approximatively normal, and thus we can perform a two-sample t-test. The t-test gives a p-value of 0.1416, and therefore we conclude that there is no significant difference in population medians. Thus test and inter variability sets are both in-plane from patients of similar health status.

4.4. Estimated hyperparameters

Table 2 contains the least mean AMDL scores for the 5 considered distributions and 6 possible configurations of mixture components for the two regions \( R_{12} \) and \( R_3 \). The Gamma model provides the lowest (and thus best) mean AMDL score, with \( q_c = 2 \) for endocardium and atrial region, and \( q_d = 2 \) for background. Thus we model the US data as a four component \( (q = 4) \) Gamma mixture model. The Nakagami distribution has results similar to Gamma, and thus one could employ this instead of Gamma with minor changes to the end result. In the case of Fisher-Tippett the EM algorithm did
Table 3

Manual Inter- and intraobserver error.
Mean ± std (regular), Median ± MAD (italic).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Interobserver</th>
<th>Intraobserver</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPAD</td>
<td>3.67 ± 2.20 mm (6.08 ± 3.90 pxl)</td>
<td>2.47 ± 1.79 mm (4.12 ± 3.02 pxl)</td>
</tr>
<tr>
<td></td>
<td>4.31 ± 2.87 mm (7.40 ± 5.31 pxl)</td>
<td>3.16 ± 2.46 mm (5.29 ± 4.38 pxl)</td>
</tr>
<tr>
<td>Contour AD</td>
<td>2.80 ± 0.77 mm (4.85 ± 1.36 pxl)</td>
<td>1.55 ± 0.81 mm (2.62 ± 1.35 pxl)</td>
</tr>
<tr>
<td></td>
<td>3.02 ± 0.97 mm (5.13 ± 1.71 pxl)</td>
<td>1.87 ± 1.08 mm (3.14 ± 1.93 pxl)</td>
</tr>
<tr>
<td>Area Correlation coeff.</td>
<td>0.95</td>
<td>0.9547</td>
</tr>
<tr>
<td>Area Error (pxl²)</td>
<td>-3.1 ± 16.35 %</td>
<td>4.65 ± 13.19 %</td>
</tr>
</tbody>
</table>

Figure 12: Ultrasound segmentation is a difficult task, prone to inaccuracies. Here we display three expert segmentations (green, red, black) together with the segmentation (final sample) from proposed approach (yellow).

not converge in 5 instances. These results were not included in the calculation of the mean, since we do not wish to punish the Fisher-Tippett distribution inordinately. It may be possible that convergence could be obtained using a different initialization scheme.
When sampling the posterior, 60 discretization points for the endocardium and atrium \((n = m = 60)\), c.f. Sec. 2.5, was found to be sufficient to capture small sample variations. Leave-out-one cross-validation (c.f. Sec. 3.3) was performed on the training set, and the mean parameter estimate was computed over all cross-validation rounds, for parameters \(\lambda_{\text{shape}}, \lambda_{\text{mean}}, \lambda_{\text{sd}}\) and \(\sigma_{\text{apex}}\). The estimates \(\bar{\lambda}_{\text{shape}} = 1.15, \bar{\lambda}_{\text{mean}} = 0.16, \bar{\lambda}_{\text{sd}} = 0.6\) and \(\bar{\sigma}_{\text{apex}} = 43\) were obtained.

The simplices \(K_{12}\) and \(K_{3}\), constructed from shape and scale parameters estimated in (23), were found to be \(K_{12} = [1, 12] \times [0.022, 0.109]\) and \(K_{3} = [1.50, 7.12] \times [0.0562, 0.273]\). The small values for scale parameters \(\theta\) is due to normalization of the data. The prior weights on the relative size of regions, c.f. Sec. 3.3, are \(\pi_{12} = 0.32\) and thus \(\pi_{3} = 1 - \pi_{12} = 0.68\).

Significant left-censoring of data, more than 20 percent of total data observations in first histogram bin, was found in 16 out of 28 data sets. See Fig. 4 for an example of a left-censored US image. This motivates the use of our EM algorithm for left-censored data, cf. Sec. 3.1.

Initialization of latent variables \(u\) is given by the mean of the training shapes for each time instant. The positioning of the initial shape is achieved by drawing a sample from the translation distribution of the latent variables, c.f. Sec. 3.2, hence no manual initialization is needed.

4.5. Results

In Fig. 7 we display Bayesian Probability Maps formed from 50 samples; with a burn-in of 50 samples. The probability map spans colors from red to blue with degree of probability, of area being within the endocardium. Hence, red indicates the highest probability while no color signifies zero probability.
Table 4

Segmentation results for proposed method and method of Chen et al. (2007) on test data.

<table>
<thead>
<tr>
<th>Method</th>
<th>Ad hoc parameters</th>
<th>Auto-Interaction Evaluation</th>
<th>Performance</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Interaction type</td>
<td>Measure</td>
<td>Value</td>
</tr>
<tr>
<td>Chen et al. (2007)</td>
<td>(+/-)</td>
<td>Manual Training</td>
<td>CPAD</td>
<td>7.66 ± 3.34 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(11.40 ± 5.00 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.34 ± 4.86 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(12.38 ± 7.08 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Contour AD</td>
<td>4.88 ± 1.38 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7.04 ± 2.11 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.07 ± 1.86 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7.53 ± 2.60 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Area corr. coeff.</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Area error (pxl²)</td>
<td>2.92 ± 43.95 %</td>
</tr>
<tr>
<td>Proposed method</td>
<td>(-)</td>
<td>Manual Training</td>
<td>CPAD</td>
<td>3.36 ± 2.68 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(5.00 ± 4.17 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.43 ± 3.47 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.74 ± 5.44 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Contour AD</td>
<td>2.58 ± 0.85 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.66 ± 1.32 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.59 ± 1.07 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.91 ± 1.71 pxl)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Area corr. coeff.</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Area error (pxl²)</td>
<td>3.80 ± 16.56 %</td>
</tr>
</tbody>
</table>

**TABLE KEY:**
- **Evaluation:** QN(m) = quantitative (manual delineations).
- **Ad hoc parameters:** (-) none or sensitivity analysis is done, (+/-) yes and some attempt at sensitivity analysis, (+) yes and no sensitivity done.
- **Automation:** (+) full, (-) interactive guidance/correction.
- **Interaction type:** Actions required by user.
- **Performance:** Values given as mean (regular), mean ± standard deviation (regular) or median ± MAD (italic), excluding Area correlation coeff.
- **Validation type:** OB(m) = human observer (number of observers).
- **Validation number:** Number of subjects (number of images).

Note: We employ the same notation for properties of models as used in Noble and Boukerroui (2006).
Figure 13: Segmentation results for model of Chan et al. and proposed method. For Corresponding Point Average Distance (CPAD) and Average Distance (AD) measure, median ± MAD are shown, while for Area Error mean ± standard deviation are displayed.

It follows that the more varied in color the map is, the more uncertain it is. This is seen in the right-hand plots in Fig. 7, which in places have colorful border regions.

The obtained results have somewhat lower median and similar MAD than those obtained through our interobserver study. We have analysed the difference through a statistical test. First we performed a Lilliefors test for normality on the results from our proposed method and the interobserver study. In both cases the null-hypothesis of normality was rejected at significance level 0.05 (p-value 0.001). Thus we cannot use e.g. a two-sample t-test, since this assumes normality of the groups which are compared. Instead we employ a Wilcoxon rank sum test, which assumes as null-hypothesis that the
Figure 14: Segmentation results for frames 1, 5, 7, 10, 13 and 16 for seven cycles using proposed method. The seven cycles evaluated are those for which ground truth was available for the entire cycle. For Corresponding Point Average Distance (CPAD) and Average Distance (AD) measure, median ± MAD are shown, while for Area Error mean ± standard deviation are displayed.

medians of the two compared groups are equal. The hypothesis of equal medians was rejected at significance level 0.05, with p-value 0.021. Thus there is significant difference between medians: the median 3.36 mm of CPAD results of the proposed method significantly lower than the median 3.67 of the interobserver results. However the absolute difference between medians is rather small, in addition to the interobserver study being small, and so one needs to be careful when drawing inference from this result.

In Fig. 14 the tracking results (over 6 equidistant frames) are displayed for seven test cycles which were fully segmented. The median error is ap-
proximately the same over the cycles, while the MAD of the error is higher around the diastolic frames (1,13,16). In these frames the heart is in its expanded phase, the anterior wall is close to the edge of the scanning sector, and here the lateral resolution is reduced due to distortions introduced by ribs and soft tissue Savakus et al. (1982).

In Table 4 and Figure 13 the results for the algorithm of Chen et al. (2007) and the proposed method are shown. We choose to compare our approach with the model of Chen et al. (2007) because it is one of the few working considering the 2 chamber apical view, which also is our object of study. Furthermore connection between image statistics and hyper parameter very interesting, despite its limited experimental validation. The same training data used in the proposed model was used for the construction of the mean contour and the mean intensity image used in the algorithm; 13 end-diastolic and end-systolic frames where selected to build a prior model for each state. We compute results for the 90 frames used for testing the proposed method. Our strategy was to first select a high, medium and low weight for the shape parameter and then evaluate 10 parameters values around the parameter value producing highest value of a mutual information measure (called MIIG). On a standard laptop (Intel (R) Core Duo, 2.40 GHz, 4 GB RAM) the time to segment one frame was 1.5 hrs. The initialisation is very crucial, a very small deviation affects the end-result. As seen in Table 4 and Fig. 13, the model of Chen et al. (2007) the median result of the proposed method (3.36) is lower than the interval [4.32,11.00] (median ± MAD) of Chen et al.. This is also confirmed by measures of Average Distance, Area Error and Area Correlation Coefficient. In our judgement, this has several
reasons: the functional is too weakly connected to the actual ultrasound data. The only connection to the data is the mutual information term which is tuned in order to estimate the shape prior weight. The idea is attractive but in many cases we found that actually was no shape prior weight which could give a competitive result.

In Fig. 10 the median and median absolute distance (MAD) of distances from the final sample of the contours, to each expert point along the endocardium is displayed.

Running time on a standard laptop (same used as for the experiments for the model of Chen et al.) is 3.2 sec/sample. Total running time for a 16 slice sequence, where 150 samples are obtained of each slice, is approximatively 2.5 hours. The code is written in MATLAB with some elements in C++.

5. Discussion

To check the validity of our algorithm we compare our results against those produced on cycles from our database. When the proposed method is compared against the results of experts on material from our hospital database, the method performs quite well, see Table 4 and 3. When comparing against another model which treats segmentation of the endocardium in the 2 chamber apical view Chen et al. (2007), we see that our proposed approach performs better.

The inferior walls (point 1 – 18) are easier for the proposed algorithm to properly segment, as seen in Fig. 10, while there is an increase in mean distance of the segmentation to groundtruth of the anterior wall points 19 – 28. It is also evident that the variation of anterior wall points is larger. This
is typically the most difficult part of the endocardium to segment, as it is the region most prone to signal dropout.

Fig. 11 highlights situations where the proposed approach works well and less so. The most important factor in how well the algorithm will perform is how severe signal dropout is in the anterior wall. If the dropout is more pronounced, then the algorithm will have to rely more on priors which necessarily will cause the result to deviate more from the true chamber wall, as seen in Fig. 11g and 11f. One should also note that not only the algorithm is adversely affected by severe signal dropout; manual segmentation by expert will also be adversely affected. The extent of how much a single expert is affected by signal dropout is difficult to quantify, barring an extensive intra-expert study.

The parameters of the model, determined by cross-validation over 13 training samples, are not optimal for every setting. Having an adaptive scheme, where the parameters are set according to e.g. the current distribution of the sample, would be more desirable and would potentially increase the flexibility of the model. Increasing the training set may also yield a similar result, although one must take into account the effect of specific diseases on the ultrasound echo, which may make a cross-validated parameter estimate, even over a large sample, too approximate.

We have made our study with censored clinical B-mode US data, by using a censored mixture model, which does not assume any additional information about the signal, such as knowledge of dynamic compression parameter (Shankar, 2003). Furthermore the study is performed on the most challenging of views of the heart, namely two-chamber apical long axis view, as opposed
to the more commonly examined short-axis view.

The proposed method assumes that it is possible to separate US data into components corresponding to blood and background. If artifacts, e.g. rib-shadows, are very substantial, this assumption will be violated. It would be possible to include this data in the training data set, but since these data sets have substantially different appearance, it would skew parameter estimation in an unwanted manner; rib-shadow would require very large shape prior parameters.

To handle the problem of the heart falling out of view, the proposed method could be extended using the approach of Dietenbeck et al. (2012). Using this approach the sampling probability would be given solely by the shape prior outside a user-defined ROI, containing the endocardium.

In addition to segmentation of the endocardium, the proposed method could be extended with a 4:th region, modeling the myocardium, which lies between the endocardial and epicardial borders. This additional region would then also give the position of epicardium. This is a natural extension, since modeling the myocardium using prior knowledge of shape and statistical distribution has previously been used in several works such as Dias and Leitao (1996); Paragios (2003); Zhu et al. (2010); Huang et al. (2014) and Dydenko et al. (2006). Naturally this would imply that training data would have to be provided for this region, which would increase the manual segmentation work-load.

Finally, an interesting extension of the model would entail parallelization of the method, by minimising burn-in. This has the potential to drastically reduce computational load. If burn-in is minimal this means that we will
be sampling from a stationary distribution, and thus can generate samples in parallel. Convex multi-region variational methods, (e.g., Andrews et al., 2011), or graph-cut methods, (e.g., Vu and Manjunath, 2008), can provide initial samples which are close to the stationary distribution, and thus will minimize burn-in. One possible strategy could be using \( v \) number of runs of this type of initialization method to create \( v \) number of seed points. Using these seed points \( K \) number of say 100 samples can be generated in parallel, and the MAP estimate would then be picked from these \( K \).

Clinical relevance

This algorithm has been developed in collaboration with several clinicians in the field of echocardiography. The following uses of the algorithm has been suggested: The proposed Bayesian Probability Maps can be employed as a supporting tool in manual segmentation, highlighting areas of difficulty. The Bayesian Probability Maps would be displayed side by side with the raw images, as shown in Fig. 7; naturally excepting the ground-truth seen in these figures.

In addition to the use in manual segmentation, the algorithm could be used as a pre-classifier for manual segmentation, indicating how difficult a US cycle would be to segment. This could be achieved by measuring how much of the probability map consists of values in e.g. the interval \((0, 0.8] \). If a proportionally large part of the cycle consists of lower probabilities then it would be marked as difficult.
6. Conclusion

We have presented a novel Bayesian approach for describing the position distribution of the endocardium in the challenging 2 chamber left-ventricular long-axis view of the heart in US image cycles. A left-censored Gamma mixture model is introduced, to address the quantization error in the B-mode US data. The problem of determining the position distribution, is cast as a latent variable model, which represents the three regions endocardium, atrial region and background, for which the posterior density is estimated. The data likelihood of the endocardium and atrial region, and background, is described as a mixture of components of the censored gamma mixture model. The model is refined by incorporating priors for spatial and temporal smoothness, and preferred shapes, by using the principal components of manually segmented training shapes. Misclassification of the apex of the endocardium, due to foreshortening related to a highly curved apex of the endocardium, is handled by down-weighting of the likelihood in this region.

The posterior density is sampled by a Gibbs method to estimate the expected latent variable image of the endocardium, which we call the Bayesian Probability Map, since it describes the probability of pixels being classified as being within the endocardium. The algorithm is initialized by sampling the translation distribution of the latent variables, improving the convergence rate of the algorithm. The regularization parameters of the model are estimated by cross-validation, and the results are validated using a variety of validation measures, in order to compare with the model of Chen et al., and inter variability results.

Our experiments show the usefulness of the Bayesian Probability Maps
for the clinician since, instead of producing a single segmenting curve, it highlights the uncertain areas and suggests possible segmentations. Furthermore, an interobserver study shows that the results produced by the algorithm are comparable to expert segmentations and outperforming the results of Chan et al..

References


models-their training and application. Comput Vis Image Underst 61, 38–59.


Dydenko, I., Jamal, F., Bernard, O., D’Hooge, J., Magnin, I., Friboulet, D., 2006. A Level Set Framework with a Shape and Motion Prior for
Segmentation and Region Tracking in Echocardiography. Medical Image Analysis 10, 162 – 177.


56


