Localizing Cardiac Structures in Fetal Heart Ultrasound Video

Christopher P. Bridge[1], Christos Ioannou[2], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, University of Oxford, Oxford, UK,
[2] Fetal Medicine Unit, John Radcliffe Hospital, Oxford, UK

Abstract. Recently, a particle-filtering based framework was proposed to extract 'global' information from 2D ultrasound screening videos of the fetal heart, including the heart's visibility, position, orientation, view classification and cardiac phase. In this paper, we consider how to augment that framework to describe the positions and visibility of important cardiac structures, including several valves and vessels, that are key to clinical diagnoses of congenital heart conditions in the developing heart. We propose a partitioned particle filtering architecture to address the problem of the high dimensionality of the resulting state space. The state space is partitioned into several sequential stages, which enables efficient use of a small number of particles. We present experimental results for tracking structures across several view planes in a real world clinical video dataset, and compare to expert annotations.

1 Introduction

Prenatal screening for congenital heart disease (CHD) is typically performed using a two-dimensional (2D) ultrasound examination in the second trimester to check for various structural and functional anomalies. However, because this is specialist work requiring detailed knowledge of fetal cardiac anatomy, detection rates are highly dependent on the sonographer's experience [6].

In this work, we develop automated methods to localize key anatomical structures in freehand video footage gathered from a screening session. This could be used, for instance, to feed back live information to a sonographer performing the scan, used to develop training tools, or used as the basis of further automated processes for diagnosis and quantification of CHD.

A few recent works have looked at automatically extracting information from fetal screening ultrasound video streams [2, 3, 1]. Chen et al. [3] used a combination of a convolutional neural network (CNN) and a temporal recurrent neural network to detect standard planes from fetal scans. Baumgartner et al. [1] also used a CNN to detect various views of the fetus and coarsely localize a variety of structures with a bounding box. In previous work focused on the fetal heart [2], we used a particle filtering approach in order to capture the temporal structure of the footage when estimating key variables in a robust, probabilistic manner. In this paper, we build on the method presented in [2] and extend the particle filtering framework to track a number of important cardiac structures.

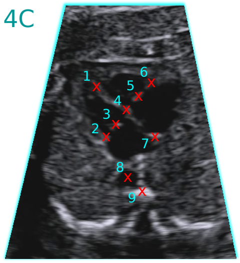
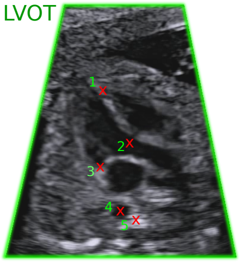
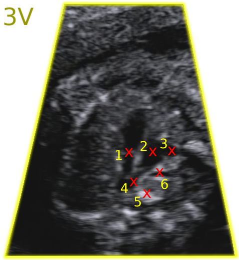
Four Chamber View (4C)	**Left Ventricular Outflow Tract view (LVOT)**	**Three Vessels View (3V)**
		
1. Apex. 2. Mitral Valve End. 3. Mitral Valve Center. 4. Crux Cordis. 5. Tricuspid Valve Center. 6. Tricuspid Valve End. 7. Base. 8. Descending Aorta (Center). 9. Spine.	1. Apex. 2. Aortic Valve. 3. Mitral Valve End. 4. Descending Aorta (Center). 5. Spine.	1. Pulmonary Valve. 2. Ascending Aorta (Center). 3. Superior Vena Cava (Center). 4. Descending Aorta (Center). 5. Spine. 6. Trachea (Center).

Table 1. Structures of interest

Though the approach is general, to place the current work in context we focus on the same three views of the heart as [2]: the *four chamber* view (4C) showing the two atria and two ventricles, the *left ventricular outlfow tract* view (LVOT), showing the aorta leaving the left ventricle, and the *three vessels* view (3V) showing the pulmonary artery, aorta and superior vena cava. Within these views, we have selected a number of anatomical structures of interest (Table 1).

2 Partitioned Particle Filters

We first review the framework of [2] and then describe how we have extended it. The particle filtering architecture in [2] tracks a *state* that captures 'global' characteristics of the heart at each frame t, specifically the heart's visibility $h_t \in \{0, 1\}$, image location of the heart center $\mathbf{x}_t \in \mathbb{R}^2$, heart orientation $\theta_t \in [0, 2\pi)$, viewing plane classification $v_t \in \{4\mathrm{C}, \mathrm{LVOT}, 3\mathrm{V}\}$, a circular variable $\phi_t \in [0, 2\pi)$ tracking the progress of the cardiac cycle, and $\dot{\phi}_t$ the rate of change of this cardiac phase variable with respect to t.

This set of variables of heterogeneous types is grouped into the *state tuple*, $\mathbf{s}_t$. It is also assumed that the scale of the heart, represented by the radius r, is known approximately at test time.

The *filtering distribution*, $p\left(\mathbf{s}_t \mid \mathbf{z}_{0:t}\right)$, over these variables at each frame, conditioned on image evidence $\mathbf{z}_t$, is represented by a finite number, N, of particles,

$\mathbf{s}_t^{(i)}$, $i = 0, 1, \ldots, N-1$ with corresponding weights $w_t^{(i)}$. At each time step, a new state value for each particle is sampled from a *prediction potential* $\psi(\mathbf{s}_t \mid \mathbf{s}_{t-1})$, which is a distribution over the state value at time $t+1$ given the state at time t. Then each sample is reweighted according to an *observation potential* $\omega(\mathbf{s}_t, \mathbf{z}_t)$ that reflects the compatibility of the state hypothesis represented by the particle with the observed image, and can be any non-negative function of its arguments. The observation potentials are learned using the random forests algorithm to perform classification and regression based on rotation-invariant features (RIFs) calculated from the image [4].

To extend [2] to structure tracking we extend the state tuple to contain variables relating to the locations of specific structures of interest. As a result, the localization procedure for the structures is able to use and influence the predictions of the global variables. However, this results in a very high dimensional state space. Particle filters typically do not perform well in such high dimensional spaces because a very large number of particles is needed to adequately cover the space and maintain a good approximation to the true filtering distribution [5].

This problem can be overcome by grouping the state variables into *partitions*, which can then be operated on in sequence [5]. We refer to MacCormick and Isard [5] for a rigorous explanation, but intuitively a state vector/tuple can be partitioned if both the following conditions apply:

1. The prediction potential for the variables in a given partition is (or may be assumed to be) independent of variables in *later* partitions (but may be conditioned on values for variables in *earlier* partitions). This means that the prediction step may take place for each partition before the updated values for variables in later partitions are known.
2. The observation potential for the variables in a given partition is (or may be assumed to be) independent of the variables in *later* partitions (but may consider the variables in *earlier* partitions). Therefore the particles may be reweighted and resampled according to each observation potential in turn.

The key insight into the advantage of the partitioned particle filter is that by operating on the partitions in sequence, the particles are guided into the peaks in the filtering distribution of each partition in turn. Consequently, a partitioned particle filter may make more efficient use of a small number of particles and operate in high dimensional spaces with a reasonable number of particles.

Although the two criteria are quite restrictive, the particle filter in [2] may be naturally broken into three partitions: one containing the visibility h_t, location $\mathbf{x}_t$, and view v_t; a second partition containing the cardiac phase ϕ_t and phase rate $\dot{\phi}_t$; and a third containing the orientation θ_t. The independence assumptions made in [2] mean that the two criteria are satisfied with no alterations to the observations or prediction potentials. The classification forests are independent of the cardiac phase variable by virtue of their training on a dataset containing heart examples from across the cycle. Furthermore both the classification forests and phase regression forests are orientation invariant as a result of using RIFs.

The filter presented in [2] may therefore be reformulated into three partitions to give the filtering architecture in Fig. 1. We introduce the shorthand subscript

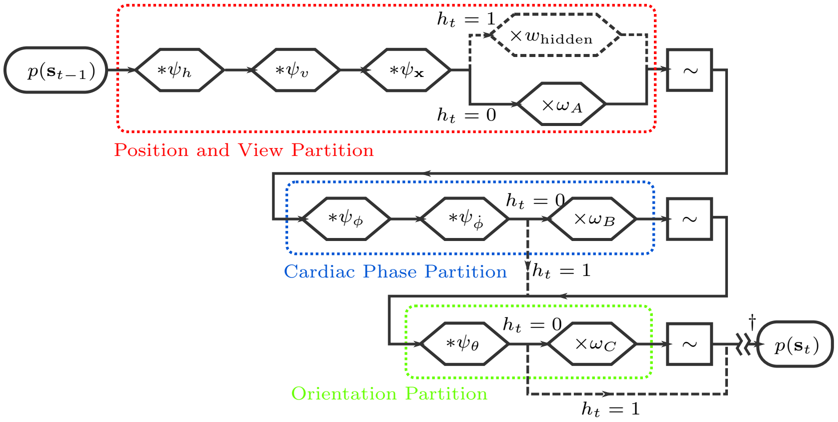


Fig. 1. Partitioned reformulation of filter architecture for tracking 'global' heart variables. The form of this diagram follows the convention used by [5] and shows the sequence of operations performed on the particle set within a single timestep. The round edged boxes represent distributions in the form of particle sets. The hexagonal boxes represent operations on each particle in the particle set: the '$*$' represents convolving the particle set with the prediction potential and the '$\times$' represents multiplying the particle weights by the observation potential. The square box containing the '$\sim$' represents the resampling operation across the entire particle set. The black dotted lines indicate routes taken only by 'hidden' particles (those with $h = 1$) and the colored dotted boxes contain the operations within a single partition (one color per partition). The break at the † symbol marks the location where extra stages are added for structure tracking in §4.

notation $\psi_h(\cdot)$, $\psi_v(\cdot)$, $\psi_{\mathbf{x}}(\cdot)$, $\psi_\theta(\cdot)$, $\psi_\phi(\cdot)$, $\psi_{\dot{\phi}}(\cdot)$ for the prediction potentials relating to the six variables, which are as defined in [2]. The observation potentials $\omega_A(\cdot)$, $\omega_B(\cdot)$, and $\omega_C(\cdot)$ are identical to $\psi_a(\cdot)$, $\psi_b(\cdot)$, and $\psi_c(\cdot)$ respectively from [2], and are based on random forests using RIFs. In §4, we will extend this architecture to use an additional partition for each structure of interest.

3 A Fourier Model For Structure Trajectories

Due to the nature of the cardiac cycle, over a short time interval the positions of the structures are likely to be close to periodic. Furthermore, an estimate of the cardiac phase variable, ϕ_t is available from the output of the global variable prediction. Rather than estimate a structure's position over the cardiac cycle in each frame independently, a Fourier model is used to capture this behavior.

In this model, the position of structure $a \in \mathbb{N}_0$ (where a is an index variable indexing the various structures (Table 1)) in the image at time t is described by the 2D column vector $\mathbf{q}_{a,t} \in \mathbb{R}^2$ containing the x and y components, i.e. $\mathbf{q}_{a,t} = \left[q_{a,t,1}, q_{a,t,2}\right]^T$, where $q_{a,t,1} \in \mathbb{R}$ is the x-component and $q_{a,t,2} \in \mathbb{R}$ is the y-component. Firstly, this is expressed relative to the heart center position, $\mathbf{x}_t$, orientation, θ_t, and scale, r, to give the *relative position vector* $\mathbf{p}_{a,t} \in \mathbb{R}^2$, where the two are related by:

$$\mathbf{q}_{a,t} = r\mathbf{R}_{[\theta_t]}\mathbf{p}_{a,t} + \mathbf{x}_t \tag{1}$$

where $\mathbf{R}_{[\theta_t]} \in \mathbb{R}^{2\times 2}$ is the 2D rotation matrix through angle θ_t.

The relative position vector $\mathbf{p}_{a,t}$ is calculated from the current value of the cardiac phase variable, ϕ_t, by assuming a truncated Fourier series approximation:

$$\mathbf{p}_{a,t} = \begin{bmatrix} c_{a,1,1} & c_{a,2,1} \\ c_{a,1,2} & c_{a,2,2} \\ c_{a,1,3} & c_{a,2,3} \\ c_{a,1,4} & c_{a,2,4} \\ c_{a,1,5} & c_{a,2,5} \\ \vdots & \vdots \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ \cos\phi_t \\ \sin\phi_t \\ \cos 2\phi_t \\ \sin 2\phi_t \\ \vdots \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} \mathbf{c}_{a,1} \; \mathbf{c}_{a,2} \end{bmatrix}^T \cdot \boldsymbol{\phi}_t \tag{3}$$

Given a short sequence of frames (covering a few cardiac cycles), the coefficients in the column vectors $\mathbf{c}_{a,1}$ and $\mathbf{c}_{a,2}$ may be found using a simple regularized least squares approach, where a prior variance λ is placed on the values of the coefficients with the exception of the zero order coefficients ($c_{a,1,1}$, and $c_{a,2,1}$), which together encode the mean position of structure over the whole cycle.

4 A Filtering Architecture for Structure Localization

We now show how the partitioned architecture in Fig. 1 can be extended to track structures. The basic idea is to include one new partition in the particle filter for each structure, with the partitions for the structures belonging to each of the three views grouped into one 'path' through the filter. This is shown in Fig. 2.

Each partition is identified by the index, a, of the corresponding structure. The Fourier model from §3 is used and the structure's position is assumed to be fixed given the coefficient vectors $\mathbf{c}_{a,1,t}$ and $\mathbf{c}_{a,2,t}$. However, these coefficient vectors are allowed to vary gradually over time. For notational convenience, the vectors are combined into a single coefficient vector $\tilde{\mathbf{c}}_{a,t} \in \mathbb{R}^{d_a}$. Additionally, there is a binary visibility variable $g_{a,t}$ for the structure that indicates whether the structure is visible or hidden due to the being obscured by an imaging artifact or being located off the edge of the image. The $g_{a,t}$ and $\tilde{\mathbf{c}}_{a,t}$ variables for each strucure are incorporated into the state tuple $\mathbf{s}_t$. We now define the prediction and observation potentials in Fig. 2.

4.1 Structure Visibility Prediction Potential, $\psi_{g_a}(\mathbf{s}_t \mid \mathbf{s}_{t-1})$

The visibility prediction potential for each structure's visibility variable operates in exactly the same way as that for the heart visibility in [2]. There is a fixed

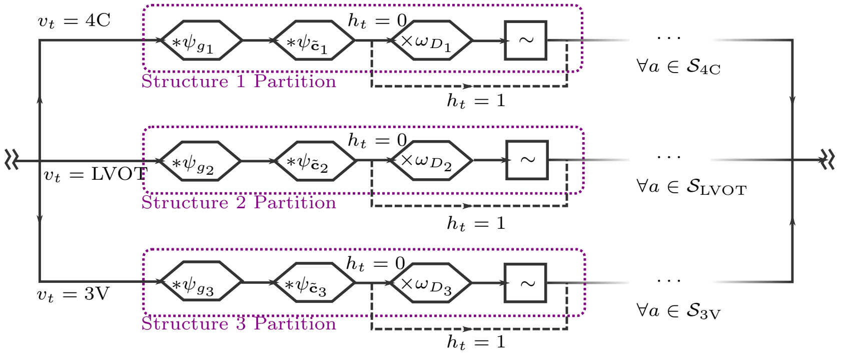


Fig. 2. Filter architecture extension for tracking structures. This is added to the architecture in Fig. 1 by inserting it at the dagger '†' symbol. The three paths through the filter relate to the three views (4C, LVOT, and 3V) of the heart. Structures with indices 1, 2 and 3 are shown as belonging to the three different views, but this is just an illustrative example and may not be the case in practice. $\mathcal{S}_{4\text{C}}$, $\mathcal{S}_{\text{LVOT}}$, and $\mathcal{S}_{3\text{V}}$ are the sets of structure indices in the three views (Table 1).

probability $p_{h \to v}$ of moving from hidden to visible and vice versa $p_{v \to h}$. These are chosen to give a certain fraction of hidden particles at equilibrium.

4.2 Structure Position Prediction Potential, $\psi_{\tilde{\mathbf{c}}_a}(\mathbf{s}_t \mid \mathbf{s}_{t-1})$

At training time, a mean vector $\tilde{\boldsymbol{\mu}}_a \in \mathbb{R}^{d_a}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}_a \in \mathbb{R}^{d_a \times d_a}$ is calculated for the coefficient vector $\tilde{\mathbf{c}}_{a,t}$, assuming a multivariate Gaussian distribution. The prediction potential is assumed to be a linear transition followed by additive Gaussian noise on the centered coefficient vector to allow the coefficients to vary smoothly during the video, i.e. of the form

$$(\tilde{\mathbf{c}}_{a,t+1} - \tilde{\boldsymbol{\mu}}_a) = \mathbf{A}\,(\tilde{\mathbf{c}}_{a,t} - \tilde{\boldsymbol{\mu}}_a) + \mathbf{G}\mathbf{n}_t \tag{4}$$

In order to ensure that the limiting distribution of the resulting Markov chain is the same as the prior distribution, the update matrix is set $\mathbf{A} = \alpha\mathbf{I}$ where $\alpha \in [0, 1]$, the covariance of the noise vector $\mathbf{n}_t$ is set to be $\mathbf{Q} = \tilde{\boldsymbol{\Sigma}}_a$, and the noise scaling is set to be $\mathbf{G} = \gamma\mathbf{I}$, where $\gamma \in [0, 1]$ and $1 = \alpha^2 + \gamma^2$.

4.3 Observation Potential, $\omega_{D_a}(\mathbf{s}_t, \mathbf{z}_t)$

The observation potential finds a score for the likelihood of structure a appearing at location $\mathbf{q}_{a,t}$ in the image. This uses a random forest classifier trained on the chosen structures and a background class, using the same rotation invariant features as the view classification forest, and the observation potential for a given structure is the posterior probability classification score for that structure at the relevant location. Hidden particles are given a small fixed score Ω_{hidden}.

5 Experiments and Results

We validated the proposed approach on the clinical dataset of fetal heart scanning videos used in [2], containing 91 videos from 12 subjects. We followed a similar leave-one-subject-out cross-validation in which all learned parameters are trained over 11 subjects, and the model is evaluated on the 12th subject. We used the manual annotations of the 'global' variables of interest from [2], and extended these to include structure locations. Pre-trained models from [2] were used for the observation and prediction potentials for the 'global' variables.

The additional models to train for each cross-validation fold included the $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$ parameters for each structure and the structure forest, $\omega_{D_a}(\cdot)$. The $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$ parameters were fitted to sequences of one cardiac cycle in length cut from the videos. All possible such sequences in the training set were used. The structure forests ($\omega_{D_a}(\cdot)$) were trained using 5000 patches containing each structure, and an equivalent number of randomly-selected background patches. The image features shared by all forest models form an RIF feature set with $J = 4$ radial profiles, maximum rotation order $K = 2$, and Fourier expansion order $M = 2$ (see [4] for more details). Only features from the central two radial profiles were used in the structures model, so that the effective patch size of the structures detection forest is half the heart radius, $r/2$.

For testing, we used the following parameter values: $p_{h \to v} = 0.35$, $p_{v \to h} = 0.15$ (giving an equilibrium with 0.3 of the total number of particles hidden), $\Omega_{\text{hidden}} = 0.05$, $\gamma = 0.3$, $\lambda = 1$. All random forest classification, phase regression, and structure localization models used 16 trees with a maximum depth of 10. The order of the Fourier models (§3) was set to 3.

Figure 3 shows the localization error for the structures when point estimates of position are found from the particle set via mean-shift. Those structures whose location is clearly defined by image features (such as the valve centers and vessels) are generally well localized, whereas the most poorly localized structures are those whose location in the image is ambiguous (e.g. the ends of the valves, the base and the apex). We observed that errors in the heart orientation significantly increased the average localization error. The average computation time per frame was 39.5 ms (25 frames per second) on a desktop PC (Intel i7-3770 3.40 GHz, 8 threads, 32 GB RAM), suggesting that this approach is well-suited for real-time applications[3]. Examples can be found in the supplementary video.

6 Conclusions

In this paper we have presented a fast method for fully automated tracking of anatomical structures in ultrasound videos of the fetal heart and presented results on a clinical dataset. Future work will include understanding behavior in the presence of heart abnormalities and the application of this work in support of sonographers in scanning and automated diagnosis.

[3] Our C++ implementation is available at https://github.com/CPBridge/fetal_heart_analysis

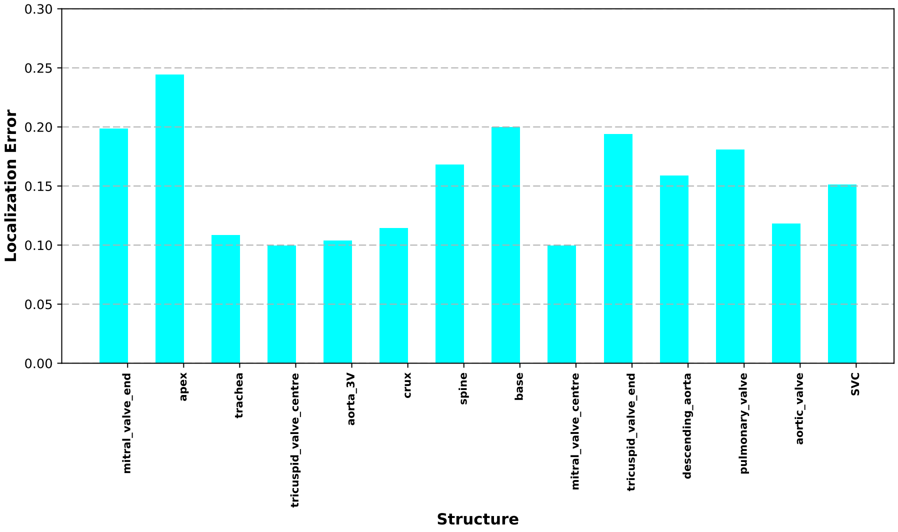


Fig. 3. Mean distance errors between estimated location and ground truth locations for each structure over all videos. The distance is normalized by the heart radius r.

References

1. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D.: Real-Time Detection and Localisation of Fetal Standard Scan Planes in 2D Freehand Ultrasound. arXiv abs/1612.05601 (2016), `http://arxiv.org/abs/1612.05601`
2. Bridge, C.P., Ioannou, C., Noble, J.A.: Automated Annotation and Quantitative Description of Ultrasound Videos of the Fetal Heart. Medical Image Analysis 36, 147–161 (Feb 2017)
3. Chen, H., Dou, Q., Ni, D., Cheng, J.Z., Qin, J., Li, S., Heng, P.A.: Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks. In: MICCAI 2015, Lecture Notes in Computer Science, vol. 9349, pp. 507–514. Springer International Publishing (2015)
4. Liu, K., Skibbe, H., Schmidt, T., Blein, T., Palme, K., Brox, T., Ronneberger, O.: Rotation-Invariant HOG Descriptors Using Fourier Analysis in Polar and Spherical Coordinates. International Journal of Computer Vision 106(3), 342–364 (2014)
5. MacCormick, J., Isard, M.: Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracking. In: Vernon, D. (ed.) ECCV (2). Lecture Notes in Computer Science, vol. 1843, pp. 3–19. Springer (2000)
6. Pézard, P., et al.: Influence of ultrasonographers' training on prenatal diagnosis of congenital heart diseases: a 12-year population-based study. Prenatal Diagnosis 28(11), 1016–1022 (2008)