Medical Image Analysis 36 (2017) 147–161

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automated annotation and quantitative description of ultrasound videos of the fetal heart

Christopher P. Bridge[a,*], Christos Ioannou[b], J. Alison Noble[a]

[a] Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom
[b] Fetal Medicine Unit, John Radcliffe Hospital, Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 26 March 2016
Revised 17 November 2016
Accepted 18 November 2016
Available online 19 November 2016

Keywords:
Ultrasound
Fetal
Heart
Cardiac
View detection
Rotation-invariant
Random forests
Particle filter

ABSTRACT

Interpretation of ultrasound videos of the fetal heart is crucial for the antenatal diagnosis of congenital heart disease (CHD). We believe that automated image analysis techniques could make an important contribution towards improving CHD detection rates. However, to our knowledge, no previous work has been done in this area. With this goal in mind, this paper presents a framework for tracking the key variables that describe the content of each frame of freehand 2D ultrasound scanning videos of the healthy fetal heart. This represents an important first step towards developing tools that can assist with CHD detection in abnormal cases. We argue that it is natural to approach this as a *sequential Bayesian filtering* problem, due to the strong prior model we have of the underlying anatomy, and the ambiguity of the appearance of structures in ultrasound images. We train classification and regression forests to predict the visibility, location and orientation of the fetal heart in the image, and the viewing plane label from each frame. We also develop a novel adaptation of regression forests for circular variables to deal with the prediction of cardiac phase. Using a particle-filtering-based method to combine predictions from multiple video frames, we demonstrate how to filter this information to give a temporally consistent output at real-time speeds. We present results on a challenging dataset gathered in a real-world clinical setting and compare to expert annotations, achieving similar levels of accuracy to the levels of inter- and intra-observer variation.

1. Introduction

Congenital heart disease (CHD) is one of the most common defects affecting infants at birth and covers a range of specific issues that affect the normal function of the heart. The established method for *in utero* detection of CHD is antenatal ultrasound screening of the fetal heart. Typical screening procedures are conducted at a gestational age of 18–22 weeks and involve the use of a two dimensional (2D) ultrasound transducer to examine visually the development and function of the different structures (Carvalho et al. (2013)). Unfortunately, detection rates of CHD vary widely due to a number of different factors including the training of the sonographer (Pézard et al. (2008); Allan (2000)), the nature of the defect, and the affluence of the region (Hill et al. (2015)).

A recent survey in the United States of America suggested that one of the key factors that limits the diagnosis rate is that many forms of CHD cannot be identified from a four-chamber view alone (Hill et al. (2015)). Recent guidelines (Carvalho et al. (2013)) have also emphasised the importance of using a number of different *viewing planes*, in addition to the common four-chamber view, in order to increase the rate of diagnosis of certain types of CHD.

Analysis of clinical fetal cardiac ultrasound videos is a challenging task, even for humans, for a number of reasons. Firstly, the indistinct appearance of anatomical structures in ultrasound images makes image interpretation difficult. This is compounded by variations in contrast levels and imaging parameters, as well as the presence of imaging artefacts such as speckle, shadowing and enhancement. In fetal cardiac videos (unlike adult echocardiography), the heart may take up only a small fraction of the screen and its location in the image can change due to motion of the probe and/or the fetus during scanning. The orientation of the fetus relative to the direction of the propagation of sound is also unknown and potentially variable. The appearance of the heart changes significantly throughout the cardiac cycle, and there may also be fetal motion in the direction perpendicular to the imaging plane that may cause the appearance to change or cause the heart to disap-

* Corresponding author.
E-mail addresses: christopher.bridge@eng.ox.ac.uk (C.P. Bridge), christos.ioannou@ouh.nhs.uk (C. Ioannou), alison.noble@eng.ox.ac.uk (J.A. Noble).

http://dx.doi.org/10.1016/j.media.2016.11.006
1361-8415/

pear altogether. Furthermore, while scanning, a sonographer will often review the different viewing planes of the fetal heart in relatively quick succession.

Computer-aided methods have the potential to improve detection rates of CHD, but little previous work has been carried out towards this aim. The focus of this paper is the general problem of automatically estimating key information of interest from videos of the healthy fetal heart during acquisition within a standard screening scan. This represents a critical first step in an image processing pipeline and could support good-quality acquisition and assist an operator in interpretation. Furthermore it provides a basis for further work towards automatic quantification and diagnosis of abnormal hearts.

In order to have a thorough and useful description of the state of the healthy heart at a given point in time, we must estimate the key parameters including its visibility, position and orientation in the image as well as the current viewing plane and the position in the cardiac cycle. Our approach is to pose this problem as an inference problem using *sequential Bayesian filtering*. There are a number of reasons for this choice. Sequential Bayesian filtering techniques allow a probabilistic belief over the 'state' of a 'system' (in our case the heart is the system and the state is its position, orientation, viewing plane and cardiac cycle position) to be updated on-line – and often in real-time – using all the observations that have been made so far. In particular, they naturally account for the uncertainty in individual observations made from the images, and balance them against a prior model of how the 'system' behaves in order to enforce temporal consistency. This is particularly important in this setting, where the information in each frame is often relatively weak or ambiguous due to the difficulty in interpreting ultrasonic reflection patterns, while the temporal model of heart behaviour over a number of frames is comparatively strong.

The outline of the remainder of the paper is as follows. Having reviewed related literature in Section 2, we formally define our problem in Section 3 and outline our proposed model in Section 4, with key components described in Sections 5 and 6. In Section 7 we describe the evaluation of the model on a dataset of fetal heart videos captured in a clinical setting. We present results in Section 8 and concluding remarks in Section 9.

2. Related work

To the best of our knowledge, this is the first work to attempt to automate analysis of fetal cardiac ultrasound videos. Previous authors have successfully performed view detection in images obtained from *adult* echocardiographic images using a variety of techniques (Agarwal et al. (2013); Wu et al. (2013); Zhou et al. (2006); Park et al. (2007); Qian et al. (2013); Kumar et al. (2009); Ebadollahi et al. (2004)), while others have had success in automatic recognition of other fetal structures in images (Carneiro et al. (2008); Rahmatullah et al. (2012); Namburete et al. (2013); Yaqub et al. (2012)) and, more recently, in videos (Maraci et al. (2014); Chen et al. (2015)). Finally, some work has attempted to estimate a more detailed description of the adult heart in echocardiographic data in the form of boundaries (Nascimento and Marques (2008); Carneiro and Nascimento (2013); Yang et al. (2008)).

2.1. View detection in adult echocardiagraphy

Several approaches to view detection in adult echocardiography make use of global image properties in order to deduce the view label. For example, Agarwal et al. (2013) use a histogram of oriented gradients (HOG) descriptor on the whole image, broken into four non-overlapping blocks. This can distinguish between two very different views (long axis and short axis) with a support vector machine (SVM) classifier. Wu et al. (2013) employ a similar method, using 'GIST' descriptors (Oliva and Torralba (2001)) in 16 image blocks instead of HOG descriptors. Zhou et al. (2006) use a multi-class classifier based on LogitBoost and rectangular filters ('Haar-like' filters) in order to distinguish between apical two-chamber and four-chamber views. Such global methods are not well-suited to fetal echocardiography because they assume a relatively consistent layout of frames, but in fetal imagery the position and orientation of the heart is unknown. Also, in our application, only small areas of the fetal images are relevant to view classification, and the rest of the image is taken up by the fetal abdomen and the womb.

This is overcome, to some extent, in the work of Park et al. (2007), which builds on the work in Zhou et al. (2006) by adding a left ventricle detection stage, which is then used to position the multi-class view classifier in the image. However, this relies upon the appearance of the left ventricle being fairly consistent between views, and there is unfortunately no such guarantee of consistency in the fetal views of interest to us. Furthermore, although it solves the problem of unknown position it does not solve the problem of unknown orientation.

Other methods rely on first detecting keypoints in the frame. Qian et al. (2013) detect space-time interest points in the video stream and describe them using a 3D scale-invariant feature tranform (SIFT) descriptor (in the two spatial dimensions plus time). Similarly, Kumar et al. (2009) detect interest points using the SIFT keypoint detector in the motion magnitude image, and describe them using local histograms of motion magnitude and intensity. In both cases, the extracted descriptors are quantised according to a pre-trained codebook, and an SVM classifier is used on the codebook histogram for classification. Such approaches are also unlikely to be effective in fetal imagery for the same reasons as the global methods. It is also difficult to estimate other information such as position, orientation and cardiac phase information from the frames using this approach.

Ebadollahi et al. (2004) first use the grey-scale symmetric axis transform (GSAT) to detect the "blobs" that are potential heart chambers. They then connect them in a Markov Random Field (MRF) graph structure in order to label the chambers and hence deduce the view label. This approach depends on reliable detection of chambers, and the authors showed that accuracy dropped dramatically when chamber detection was not reliable, as is likely to be the case in fetal imaging where structures other than the heart are visible.

2.2. Structure detection in fetal ultrasound imagery

Several authors have used ensemble methods that combine weak classifiers based on rectangular block filters to detect particular structures in fetal ultrasound imagery. For example Carneiro et al. (2008) used a Probabilistic Boosting Tree and rectangular filters to detect a number of different structures including the fetal head, abdomen and femur. Rahmatullah et al. (2012) and Namburete et al. (2013) used similar features and an Adaboost classifier to detect abdominal and cerebral landmarks in fetal images, and Yaqub et al. (2012) used random forest classifiers with rectangular filters for cerebral structures. We draw on these works by using random decision forests for detection of and discrimination between the different fetal heart views. However since rectangular block filters do not deal well with unknown orientations, we have instead chosen to use a alternative set of rotation invariant features (see Section 6.1).

One approach to fetal ultrasound *video* analysis is that of Maraci et al. (2014), who model the frames in short video sequences as the output of a linear dynamical system, and construct a SVM classifier based on kernels between the model parameters in order to detect subsequences containing structures of interest. This